3. Results and Discussion

Oligonucleotide fingerprinting(OFP) is a powerful method in genome research, which is used for arrayed DNA libraries characterisation(Meier-Ewert et al., 1998), (Poustka et al., 1999; Radelof et al., 1998). However the experimental approach has changed unsignificantly from the originally proposed (Lehrach et al, 1990) and realised (S. Meier-Ewert, 1995). The main problem of the current technology derives mainly from the hybridisation characteristics of oligonucleotide probes: low stability in duplexes and lack of predictable hybridisation behaviour. The OFP method relies on radioactive labelling and detection, so that the results are dependent on the specific activity of probe and time of exposure.

The aim of this work was to establish sensitive non-radioactive labelling methods of short oligonucleotides and apply the DNA analogues PNA (peptide nucleic acid) oligonucleotides to improve the hybridisation properties of probes.

3.1 Description of model systems

Prior to the use of PNA in a large scale the criteria for comparison with the conventionally used DNA probes were to be established. The most important criteria of oligonucleotide hybridisation are the specificity and reproducibility. In the following part I describe the characterisation of specificity and reproducibility of DNA oligonucleotides labelled radioactively.

The experiments were performed using the following model systems : 1) fully characterised, complex DNA targets of >1000 bp length, spotted on nylon membrane at high density and 2) oligonucleotide targets immobilised on the nylon membrane. In this work the following terminology is used: the *target* is fixed on the membrane and the labelled *probe* is in the solution.

3.1.1 Model DNA array

The model DNA array consisted of a set of DNA clones, characterised by hybridisation with 82 radioactively labelled DNA oligonucleotides. This DNA set included 700 redundant genomic clones with average inserts size of 1.5Kb. The clones were sequenced from both ends and comprised a contig of 100Kb of completely known sequence (Radelof et al., 1998)). The system of complex DNA targets of high similarity is the most appropriate system to assess reproducibility of hybridisation. This well-characterised library was used also for the clustering analysis based on oligonucleotide hybridisation and then for comparison of oligofingerprinting data with sequence- and clones'

physical overlap information. Another advantage of this small library of low complexity is that for a model experiment a limited number of oligonucleotide probes was sufficient for characterisation of the library by oligofingerprinting.

The clone inserts were amplified by PCR in 384-well plates. They were then arrayed on nylon membrane (Hybond N+) using a robotic machine (Q-Bot, Genetix, UK) which utilises a 384-pin gadget. Diameter of the pins is 250µm. The smaller pin diameter DNA spots facilitates a higher surface density of the spotted DNA and thus sharper signals that are the main requirement for fluorescent detection. Each clone was spotted in duplicates and the hybridisation signal was counted as a mean value of two duplicates. The difference in signal was due to pin shape variance which introduced difference in amounts of DNA in the spots.

The estimation of DNA amount transferred by a pin is calculated as volume of a hemisphere:

$$\mathbf{V = 2/3\ pr^3}$$

where r is the radius of the pin

This corresponds to about 4 nl in single transfer or 40 nl for 10-times spotting. Experiments with spotting radioactively labelled DNA confirm that the total transferred volume is about 100 nl of DNA.

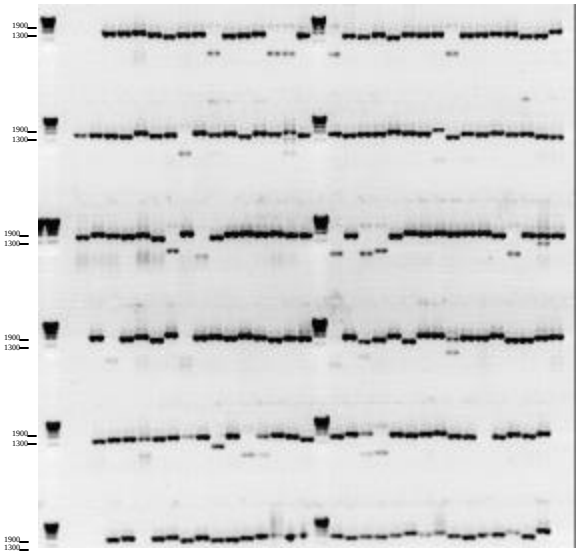
The hybridisation signal depends on the mole-amount of the target DNA. Assuming the yield of a PCR reaction as 1 µg in 30 µl (30ng/µl as estimated in the EthBr-stained agarose gel) the total amount of PCR(insert size 1500 bp) can be calculated as

$$\mathbf{C = (30ng/\mu l) / (1500 * 0.6ng/pmol) \sim 30\ fmol/\mu l}$$

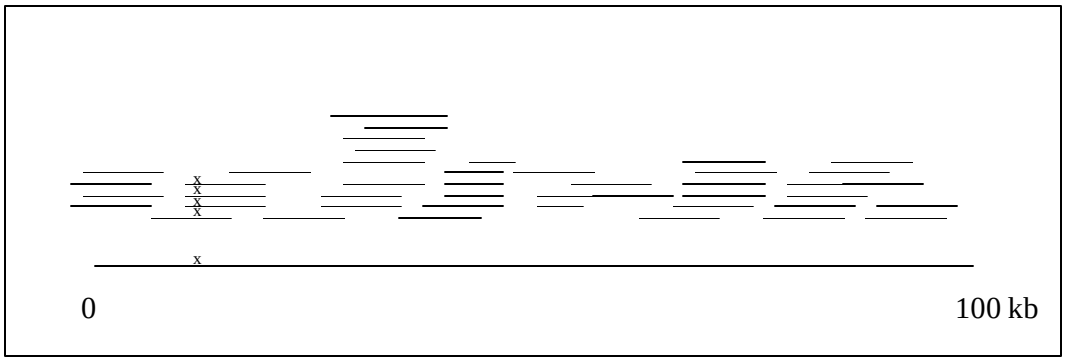
The total amount of DNA transferred to membrane is estimated as 2.6 fmol. The density of DNA in a spot of membrane is thus 2.6 fmol /π*$(250)^2$ μm^2 or ca 13 fmol/mm^2.

Figure 1. A. Agarose gel of PCR-products of the control clones inserts shows an almost equal yield of PCR and similar clone length (insert size 1200-1700 bp). 1 µl of a 30 µl PCR reaction is loaded in each lane. The first lane shows a DNA length marker (λ *Bst EII* digest). B. The scheme of the DNA clone contig. The contig of 100Kb consists of the overlapping clones. The clone collection is very redundant since many regions are represented by more than one clone. The nucleotide position (X) is present in 4 overlapping clones (clones which "cover" this position).

A.



B.



3.1.2 Short oligonucleotides as an immobilised target

A system of immobilised targets a few nucleotides longer than the probe reflects the specificity of duplex formation better than more a complex target system, for which it is difficult to discriminate the signal for the matching duplex over several mismatched duplexes (Drmanac et al., 1990).

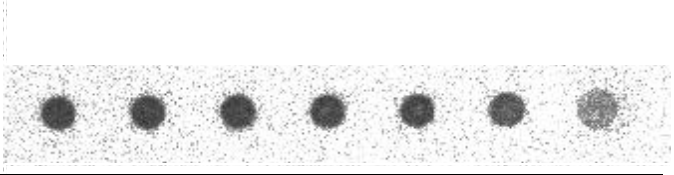
The model system used here consisted of 15-nt-long oligonucleotides containing each of the four nucleotide substitutions in central position:

5'-cagggXtttcccagt, X = A, C, G, or T

which were blotted and UV fixed onto membrane. This target system was sufficient to estimate many important parameters, such as hybridised and detected amount of DNA, as well as specificity of oligonucleotide hybridisation and sensitivity of detection method This system provided an easy control whether alternative labelling methods affected sequence specificity of hybridisation. The discrimination between targets, i.e. specificity of hybridisation, was detected by the difference in signal intensities between perfect and mismatched duplexes.

Initially the membrane binding capacity for short oligonucleotides was estimated to prove that the results obtained in this system would be adequate to results obtained with arrayed DNA clones. The constant amount of 10fmol of radioactively labelled oligonucleotide was mixed with 3-fold dilutions (from 30fmol to 30pmol) of the unlabelled oligonucleotide of the same length and blotted to a membrane (see Methods). The membrane was washed and the amount of bound to membrane radioactivity was monitored after 2.5 hour exposure to a phosphorscreen and quantified. As data from 3 independent experiments showed, the signal has reached a saturation at the 10 and 30 pmol of unlabelled oligonucleotide were spiked with the constant amount of radioactively labelled oligonucleotide (Figure 2).

Figure 2. An image of the constant amount of labelled oligonucleotide mixed to the correspondent amounts of unlabelled oligonucleotide and blotted onto membrane. The upper row in the table indicates the amount of unlabeled oligonucleotide used for blotting. The row below shows the counts for each dot.



pmol	0.03	0.1	0.3	1	3	10	30
counts	1142	1145	1127	1025	994	683	239

It can be concluded that blotting of short oligonucleotides is limited to about 10 pmol on a surface of 14 mm^2. It appeared that after pre-hybridisation washing with a buffer which contains 7.5% sarcosinate only 1/50 of oligonucleotide which has been used for blotting remained on the membrane, as it was revealed by the reduced counts on the membrane compared to unwashed membrane. The loss factor was taken into account to determine the sensitivity of detection. The final density of oligonucleotide bound to Hybond membrane if 10 pmol of oligonucleotide was blotted, is estimated as 20 $fmol/mm^2$. This is comparable to 13 $fmol/mm^2$ of DNA spotted robotically as described above. It means that the evaluation of sensitivity made for immobilised oligonucleotides is comparable to the system where the spotted PCR products are used.

3.2 Characteristics of oligonucleotide hybridisation with DNA clones

There are at least two major factors which influence the hybridisation signal: the number of complementary sites and the number of mismatched (partially complementary) sites in the long DNA target, where oligonucleotide can also hybridise. A detailed analysis of oligonucleotide hybridisation results reveals that the hybridisation behaviour of oligonucleotides is complex (Drmanac et al., 1990). Generally a hybridisation result is called „specific“ if the positive in hybridisation clone contains a complementary site for an oligonucleotide and “false positive”, if the complementary sequence is not present. Similarly, a hybridisation is called “false negative” if the complementary sequence is present but there was no hybridisation signal detected.

To assess specificity, the correlation of the presence of an oligonucleotide match in a clone and a hybridisation signal with the same clone was analysed using the data on hybridisation of 82 different oligonucleotides. An important question to answer was whether the clones of the similar sequence hybridised identically with any given oligonucleotide in the list. This is further called the *reproducibility* of the hybridisation.

In this chapter, a new method to evaluate the specificity and reproducibility of the hybridisation of a single oligonucleotide to a large number of clones is described. A hybridisation pattern unique for an oligonucleotide is

visualised by integrating of the hybridisation intensities of the large number of redundant in sequence clones. These patterns can be then compared for different oligonucleotides or probes of different chemistry. The programs mentioned in this section were written by Christoph Wierling and Ralf Herwig.

3.2.1 Specificity of hybridisation

The hybridisation data of 82 oligonucleotides on the control sequenced library was available from the shotgun clone pre-selection project (Radelof et al., 1998), so that for each of 700 clones a fingerprint (result of sequential hybridisations of a clone to number of oligonucleotides) was known.

The fingerprint data were grouped for clones of overlapping sequences according to the degree of the overlap. The clones had approximately 1000 bp of common sequence for each member of the group and 200-500 bp of overhangs on the either ends (as shown in the example of 12 clones of the same sequence on the Figure 3, A). The grouped fingerprints are shown as the matrix of clones (in rows) and hybridised oligonucleotides (in columns) on the Figure 3, B. Each line is oligonucleotide fingerprint of one clone. The grey levels of hybridisation intensity are scaled to four grades according to hybridisation intensities (the black is the highest).

Figure 3, B shows that the fingerprints are similar but not at all uniform for the clones, because the probes did not hybridise reproducibly to each clone in the overlap. The measure of reproducibility for a probe can be defined as the *ratio* of clones from group hybridised *equally* to a given probe. Since most of oligonucleotides should not hybridise to a given sequence, the positively hybridising oligonucleotide are the most important. If “reproducibility” is set as 50% positively hybridising clones from all clones (6 clones from 12 in the example on the Figure 3) then the oligonucleotides 4, 21, 26 in the example above can be considered as reproducibly positive, because they hybridised to more than 6 clones. The differences in physical overlap could account for different hybridisation results. However the fingerprints of the shortest (clones 10-12) and the longest (1 and 2) clones were nearly identical (Figure 3, A).

In order to access the specificity of each oligonucleotide hybridisation the experimental fingerprints were compared to the clones sequence. The GCG program ((GCG), G.C.G. Wisconsin Package Version 9.0 (Madison, Wisc.) was used to find which oligonucleotides were present or absent in each clone. The difference of hybridisation resulting from true presence/absence of an oligonucleotide in the sequence can be a characteristics of the hybridisation specificity of this oligonucleotide, especially for those which are reproducibly positive. Each oligonucleotide in the fingerprint can be classified as positive, false positive, negative and false negative.

The relevance to the sequence information for reproducibly positive probes for the example above is shown on Figure 3, B. It indicates a large proportion of reproducibly false positive oligonucleotides: 6 from total 14. Similar results were obtained with other clusters.

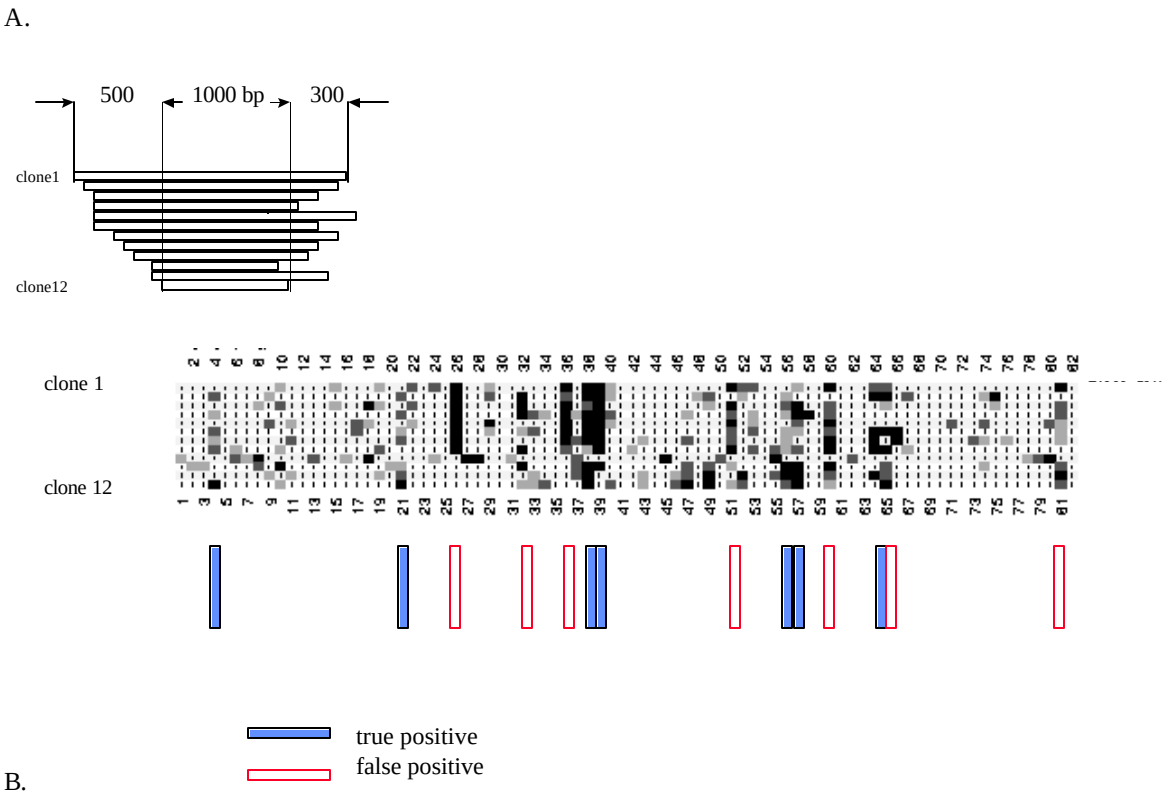


Figure 3. Analysis of grouped fingerprints of 12 clones. (A) The scheme of 12 overlapping clones. The clones had minimally 1000 bp of common (identical) sequence and additional sequences of 100-500 bp on either end. (B) The grouped fingerprints of the clones. If an oligonucleotide was positive for more than 6 clones the hybridisation was considered as "reproduced". The search with GCG Program was performed to check whether the sequences contained the correspondent 8-mers. The bars under the grouped fingerprints show the presence (then it's called "true positive") or the absence (then it's called "false positive") of the reproducibly positive oligonucleotides in the sequence.

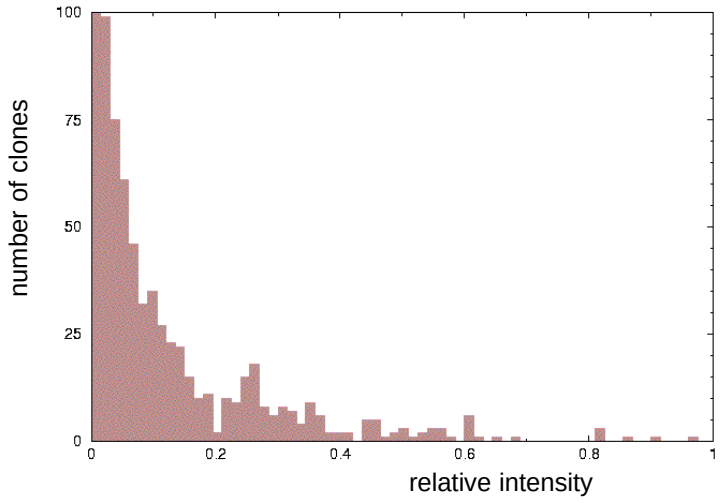
3.2.2 Analysis of hybridisation of a single oligonucleotide

In order to access the reproducibility of a single oligonucleotide hybridisation with a set of redundant clones it was necessary to set criteria for the „positive" hybridisation. Presumably, the clones containing a perfect site for an oligonucleotide should be reproducibly positive in hybridisation.

Ideally the hybridisation results should be assigned to a 1 or 0 value. In such a binary system 1 corresponds to positive and 0- to a negative hybridisation. A binary result would be a characteristic of absolute discrimination between perfectly matched and unrelated sites for an oligonucleotide in the clones' sequences.

In a real experiment all the values between 0 and 1 are obtained. Figure 4 shows the distribution of the scaled to 0-1 hybridisation signal intensities for 700 clones(Results1.1) hybridised to a 10-mer oligonucleotide. The picture illustrates that 70% of all clones had non-zero intensity. From these positive clones the majority had very low signal intensities: below 0.2 in a 0-1 scale.

Figure 4. Distribution of hybridisation intensities for 700 clones hybridised with the DNA probe NTGGAGCTGN. The probe was hybridised to an array of control clones and the results were evaluated as described in Methods. To produce this histogram the signal values for each clone were normalised by dividing by the mean value and then transformed to the 0-1 scale(Methods). The histogram is built with a bin width of 0.03.



It is likely that the weakly hybridising clones are the "false positives" since their signal intensity is at a very low, background level. In order to cut off such "false" signals and to define a "positive" signal it was necessary to set a threshold.

For the data processing the following normalisation of hybridisation intensities was used: 1) ranking of intensities on a 0-1 scale, which makes the data from different experiments comparable, (Clark et al., 1999) 2) setting of a threshold of the ranked "positive" intensities arbitrarily to 0.9 similarly to the algorithm (Herwig et al., 1999) which calculates the fingerprint's similarity. The values below this threshold were set to zero.

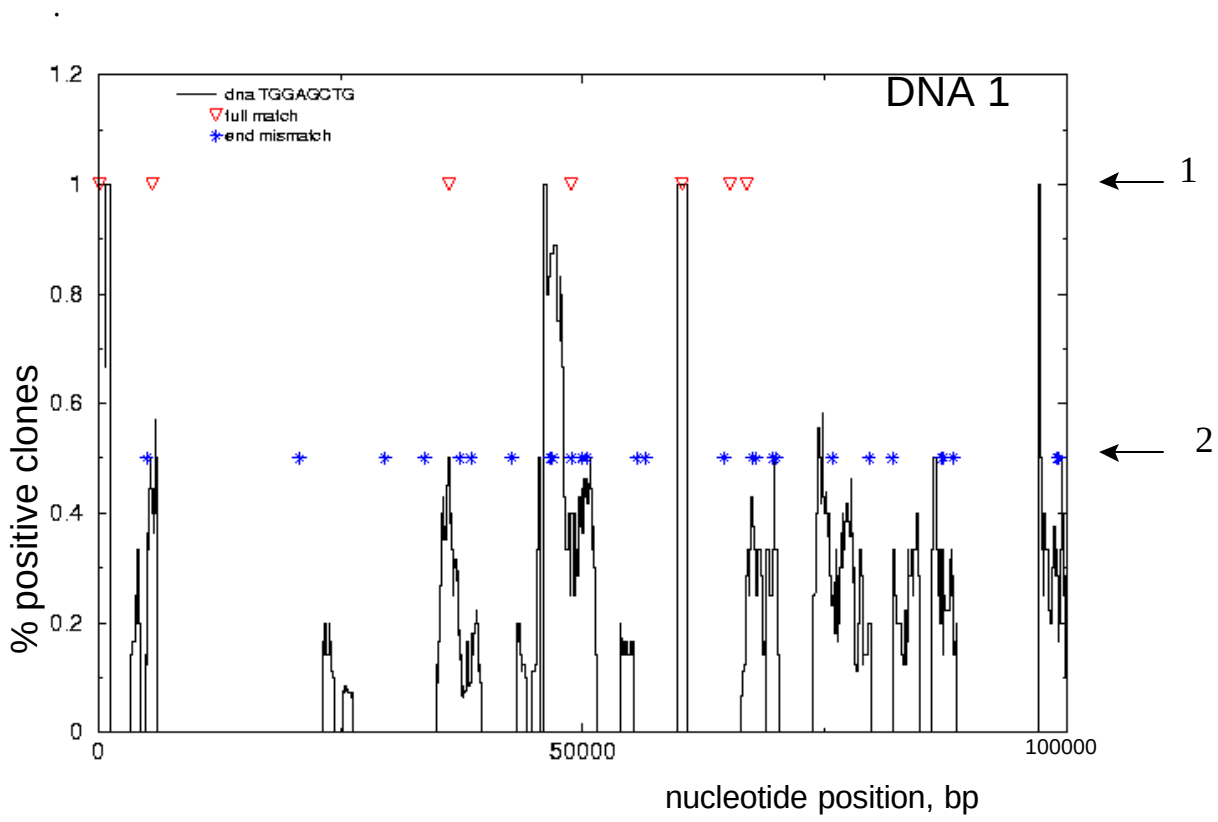
In this work a new method to analyse the hybridisation pattern of the oligonucleotide was introduced. Due to the redundant character of the library, most of the nucleotide positions in the sequence of the 100Kb long contig were covered by several clones, so it was possible to monitor hybridisation of an oligonucleotide to the clones in any overlap. The ranked hybridisation intensity values above the threshold were set to 1, integrated and divided by the sum of clones for each nucleotide position. The resulting number of "positive" clones was then plotted.

The Y values are characteristic of the reproduced positive hybridisation in the corresponding region. The peaks correspond to the regions in contig where most of the clones from an overlap were scored positive. As the value approaches a maximum of 1 (or 100%) this means that, e.g. 20 clones from 20 in the overlap were positive in hybridisation. The number of positive clones varied from 20% to 100% (values from 0.2 to 1 on the Y-axes). The number of total clones in the overlaps correspondent to the peaks varied significantly. For example, for NTGGAGCTGN it varied from 1 to 20 clones in the overlap with an average of 6-9 clones per overlap. As a consequence of unequal spatial representation of different regions in this library there can be an overestimation of reproducibility, such as in the case of a single clone with no overlapping clones. Taking this into account, a value of 20% is too low, so the threshold of reproducible "positive" hybridisation was set to 50% of clones.

An example of a hybridisation pattern of the oligonucleotide NTGGAGCTGN is shown (Figure 5). Interestingly, there were many more "positive" peaks than complementary sites for this probe (7 sites). Some of the peaks corresponded to the sites of a terminal mismatch (7 matched nucleotides out of 8), but there were overlapping clones which hybridised reproducibly (Y value of >0.5) even though they did not contain any match.

The value of 100% reproducibility (which means that all clones in an overlap were "positive" as judged from their ranked intensities) can correspond to 10 clones scored positive from total 10 in the region, or to 1 from 1. The latter case might be as well a false hybridisation.

Figure 5. The pattern of hybridisation of the radioactively labelled DNA oligonucleotide (NTGGAGCTGN). Hybridisation to 700 arrayed clones was performed and evaluated as described in Methods. The clones form a contig of total of 100Kbp. Along the Y-axes the proportion of the positive clones from the total number of clones covering the correspondent nucleotide position is shown. The value of 1 equals 100%. The clone was considered positive if its signal intensity was above 0.9 in the ranked to the scale 0-1. The coordinates for fully matched sites (red) and terminal mismatches(blue)are shown inside the diagram along the X-axes (marked with arrows 1 and 2, correspondingly).



3.3 Development of non-isotopic method to label oligonucleotide probes

Large scale experiments such as multiple hybridisations increase the demand for the development of non-radioactive methods which are more suitable for automation. The use of fluorescent dyes could speed up the process, however, most of fluorescent dyes have the limitation of a low sensitivity of detection which delays their use. The alternative detection methods of short probes hybridisation, are chemiluminescence or chemifluorescence which, as previously described, rely on a three-step protocol preceding the signal amplification with an enzyme: 1) blocking of the membrane after hybridisation of a biotin- or digoxigenin-labelled probe, 2) incubation with enzyme-streptavidin or enzyme-antibody conjugates and 3) washing off of the conjugate.

Here I describe the development of a protocol which results in quasi covalent labelling of a short oligonucleotide with an enzyme via biotin-Streptavidin link. The biotinylated 10-nt-long oligonucleotides probes were labelled with Alkaline Phosphatase and their hybridisation properties were tested.

3.3.1 Properties of the oligonucleotide directly coupled to an enzyme

The hybridisation signals of the target dilution series were quantified for the standard „3-step“ detection, directly labelled probe and the radioactive method. The conditions (probe concentration, etc.) of a chemiluminescent method were chosen similar to that of a radioactive method. The sensitivity was essentially the same for all three methods using the same concentration of probe using 12 hours of exposure for radioactivity or in 15-30 min for chemiluminescence (Figure 6). Taking into account losses of the target oligonucleotides of about 40-fold the minimum level of detection was less than 0.1fmol/mm^2.

The dependence between the amount of DNA blotted and the hybridisation signal was also tested. Enzyme-coupled probes were hybridised to six replicas of a dilution series of target and the mean signal values were plotted against target DNA amounts (Figure 7). The signal value is a function of the exposure time. Therefore, different CCD exposure times were examined. The highest signal to noise ratio was obtained using exposure time of 3 to 10 min. In independent experiments using different 10-nt long oligonucleotides a linear response between the hybridisation signal and the amount of DNA detected was observed over two orders of magnitude of target dilution. The average of 6 values of hybridisation signals for each concentration was plotted in double logarithmic scale (k=0.99 for 9 min of exposure). The regression line was calculated using the least square method.

Sequence specific hybridisation with enzyme coupled probes is shown on Figure 8. When probes specific for either of the target oligonucleotides were used at low stringency conditions of hybridisation, such as temperature of +4°C and high salt (0.8M Na^+), still considerable discrimination was achieved. The ratio of perfect to mismatch signals

was at least 10:1 at higher amounts of spotted DNA as it was quantified on a Lumi-Imager, and it increased dramatically with dilution, similarly to the results achieved by radioactive method. The probe complementary to each of the 4 DNA targets was used to control presence of targets on the membrane.

Figure 6. Comparison of sensitivity of Str-AP oligonucleotide hybridisation detection with the standard 3-step protocol and autoradiography with ^{33}P. The probes TGGGAAAGCC and TGGGAAAACC were hybridised to the oligonucleotides 5'-CAGGG**X**TTTCCCAGT (**X**=A,G,C or T) . The DNA amounts stated are the amounts used for blotting. The chemiluminescent images were obtained in 15 min of exposure as compared to 12-24 hour required for exposure with ^{33}P labelled oligomer.

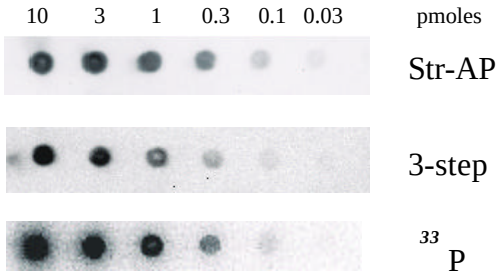


Figure 8. Specificity of hybridisation of Str-AP probes with blotted dilution series of four oligonucleotide targets. Three filter replicas with 3-fold dilution series were used. These images were produced using X-Ray film (Kodak), exposure time of 3min. (A) shows hybridisation of probe complementary to target C, (B) hybridisation with degenerate probe complementary to each target was performed in order to examine whether all targets were blotted equally, (C) hybridisation of the probe complementary to target T. Images produced with digital imaging were identical to the ones developed using conventional film.

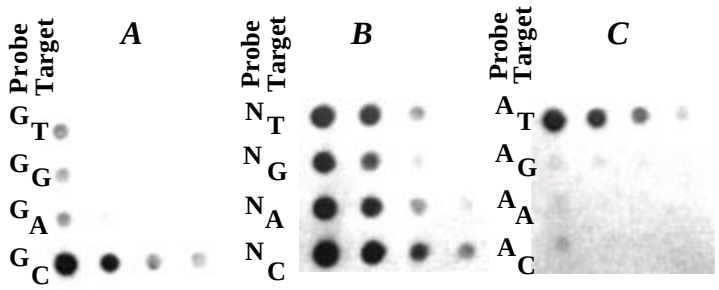
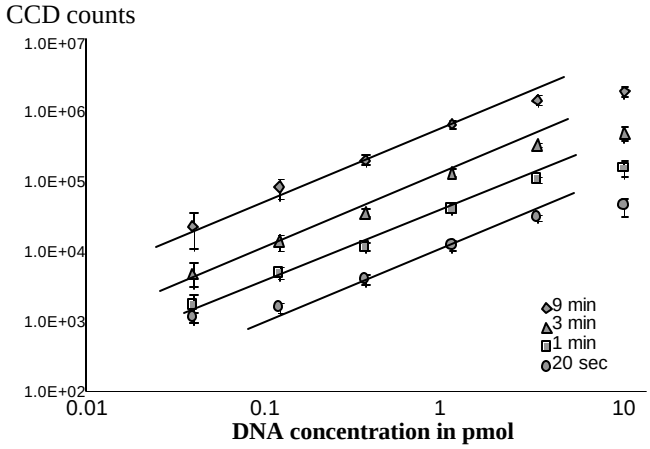


Figure 8. Relation between hybridisation signal and amount of DNA blotted. Six replicas of 3-fold dilution series were used to identify linearity of detection. The average of six value is plotted in double logarithmic scale (k=0.99 for 9min of exposure) The regression was calculated by the least square method.



3.4 Discussion

This thesis was focused on a new experimental approaches to the oligonucleotide fingerprinting. The problem of existing approach is that due to the low frequency of hybridisation it is necessary to hybridise a few hundred oligonucleotides in order to provide enough information for DNA characterisation. The use of shorter probes could overcome this drawback, however short oligonucleotides are very unstable, so the aim was to test whether DNA analogues PNA provide a solution to this problem.

Another goal was to substitute radioactive labelling of the probes with the adequately sensitive and simple method. The possible alternative were non-direct fluorescence and chemiluminescence. The experimental set up was to test non-radioactive detection first on DNA oligonucleotides and then apply it to DNA analogues. However, little was known on hybridisation properties of PNA with DNA attached to the nylon membrane.

In order to set up the criteria for PNA/DNA comparison as a first step the data on radioactively labelled oligonucleotide probes hybridisation were analysed. The data were obtained on a model library of redundant overlapping clones which was ordered in a contig. The size of the library was about 1/100 of a library usually analysed by oligofingerprinting (Meier-Ewert et al., 1998). The sequences of used clones here were fully known and had the same length of an insert.

The data obtained with DNA oligonucleotides provided a standard for evaluation for the hybridisation with DNA analogues and non-radioactive methods of detection. The detailed analysis of the hybridisation fingerprints allowed to estimate the reproducibility of interaction of short oligonucleotides with DNA clones and provided more insight into their hybridisation behaviour. The specificity of oligonucleotide hybridisation, the ability to distinguish between complementary and non-complementary DNA sequence, appeared a less important criteria.

3.4.1 Specificity and reproducibility of oligonucleotide hybridisation

The information on the clone's identity depends on the reproducibility and hybridisation intensity and its variability between different clones. It was emphasised earlier (Milosavljevic et al., 1995) that besides the mere presence of the complementary site the hybridisation intensity is affected by sequence features of clones.

The analysis of radioactive hybridisation of 82 DNA oligonucleotide probes with a redundant library performed here concludes, that perfect specificity of oligonucleotide hybridisation to the complex DNA is not achieved under conditions used. It can be illustrated by the fact that 70% of clones hybridise with an 8-mer probe as they have non-zero intensity of signal. Figure 5 shows the distribution of complementary sites for oligonucleotide in the sequence (7 sites in the 100Kbp sequence) and the rough estimation gives a much lower percentage of clones which should be positive. Since each such site complementary for an oligonucleotide can be represented by up to 10-20 clones, in maximum there should be no more than 10-20% positive clones, if only clones with complementary sites hybridised. The obtained distribution of signal intensities with the maximum in the low signal values suggests that there is an

interaction (association) of short oligonuclotide with DNA clones on the background level. This interaction is most likely unstable, which results in a low signal.

The intensity of hybridisation signal of an oligonucleotide with a DNA target can be described by the following equation (1)

$$\mathbf{I = N_p * I_p + N_m * I_m + I_b} \qquad \mathbf{(1)}$$

where

I- signal intensity

I_p- intensity of signal detected from interaction with the fully matched site

I_m- intensity of interaction with the mismatched site

I_b-(background) signal from non-specific interaction with(any)DNA

N_p,N_m, -number of corresponding sites for a clone

It is obvious to assume that mismatched duplexes and non-specific interactions are less stable than a perfect one, so the maximal intensity is obtained from a perfect match: $I_p > I_m > I_b$ (2)

The false positives (strong hybridisation signals with no perfect match in the sequence) can be explained by this model as the $N_{mismatch} >> N_{perfect}$ situation. The abundance of mismatched sites in the target results in a higher I value. The weak, background DNA-DNA interactions impact grows with an increased length of the target DNA, because the probability to find a mismatch or a partial match in the sequence becomes higher. The repeated sequences may contain such sites in increased concentration, so if the repeat unit is partially complementary to an 8-mer oligonucleotide probe, the hybridisation signal increases. Nevertheless it was reported that repeated sequences did not affect the oligofingerprinting (Poustka et al., 1999).

It was impossible to classify the impacts of various mismatches and mismatch position(for example “terminal”) by analysis of the available data. There was no strong correlation (M.Steinfath, personal communication) between false positive signals in hybridisation and occurrence of ”end-mismatch” sites (7 continuos bases in the 8-mer core). Also an attempt to estimate the influence of certain mismatch pairing (such as G:T known as the most stable (Maskos and Southern, 1993) was not successful either.

Among all the signals obtained in hybridisation with a single oligonucleotide, the highest signals are considered to belong to the clones with fully matched sites. The remaining signals are presumably from hybridisation with multiple single mismatched sites. More than one mismatch in a 8-mer duplex should in theory be very unstable as show results on hybridisation with short targets immobilised on membrane.

Considering the model above (Equation 1) it was necessary to define the threshold for a ”perfect” hybridisation so that is was distinguished from a background level hybridisation. However, the term ”positive” should be used

instead of ”perfect” because the most of signals did not correspond to a perfect site in the target DNA. The criteria of ”positive” hybridisation was set arbitrarily as the 10% top intensities (or a threshold of 0.9 in 0-1 scale).

The data processed in this way was used to compare hybridisation patterns and to evaluate the reproducibility and specificity (if any) of hybridisation based on hybridisation intensities from clones. Knowing the number of overlapping clones in each region it is possible to calculate the ratio of positive ones in any overlap. The clone was scored as ”positive” if it had an intensity at the top 10% of intensities as above. Reproducibility was estimated for each nucleotide position in the contig by calculating which portion of clones covering this position were positive. When more than 50% of clones of the same sequence were positive, such hybridisation was reproducible for these sequence. By such a calculation a single hybridisation event is ”profiled” along the clone assembly of 100Kb.

Simultaneously the peaks of reproducibly hybridising sequences are easily correlated with the positions of perfect matches for oligonucleotide. Using a threshold for a ”positive” of 0.9 for the ranked intensities the correlation between positions of matched sites and positions of clones reproducibly positive was the best. That means that the highest signals were more frequently obtained in hybridisation with complementary clones.

The main application of such diagrams is to assess the specificity of the hybridisation of one probe taking into account data on 700 hybridisations. Interestingly, as shown on Figure 5, often the sequences were strongly and reproducibly positive even they though did not contain a perfect match for an 8-mer oligonucleotide probe. Otherwise no positive clones were found for regions containing a complementary site. This may be referred to problem of “false negatives” which are also reproduced. The likely reason of different reproducibility of hybridisation of clones of same sequence is that local structure of DNAs might influence the availability of hybridisation sites for each single clone on the membrane differently. The diagrams were used further as method to compare hybridisation patterns of probes of different chemistry.

3.4.2 Enzymatic labelling of oligonucleotides using biotin-Streptavidin link

Compared to isotopic labelling of the probes a non-radioactive method of detection could reduce the cost and time needed for analysis. Unfortunately, most of the existing fluorophores have low signal to background ratio for nylon membranes, a typical support for DNA immobilisation. The detection should rely on high signal-to-noise ratio and possibility to re-use membranes with spotted DNA. Alkaline-Phosphatase based detection systems proved to be very sensitive and flexible to variety of substrates existing for this enzyme (Karger et al., 1992).

Although the methods for direct conjugation of enzyme with DNA probe have greatly simplified detection, the labelling of very short DNA probes, such as 10-nt-long, is incompatible with existing protocols. It is difficult to perform reliable purification of short oligonucleotides during intermediate chemical modifications.

Here the procedure of coupling of biotinylated 10-nt-long oligonucleotides with Alkaline Phosphatase conjugated with streptavidin (Str-AP) via biotin-streptavidin link, and hybridisation of the modified probe with no

further purification compared to earlier protocols (Tizard et al., 1990),(Alves et al., 1988) was introduced. This method also avoids chemical modifications of DNA along with intermediate purifications. Biotin is incorporated via synthesis at the 5'-end of the DNA and serves as a uniform label for coupling of any commercially available streptavidin-enzyme conjugates.

Since hybridisation of oligonucleotides to a complex DNA does not give definite results in terms of specificity, in order to characterise the enzymatic labelling of oligonucleotide the system of immobilised oligonucleotides was used. The sensitivity of the chemiluminescent detection method was sufficient to detect amounts of DNA adequate to the DNA array produced robotically, which corresponds to density of ca 15 fmol/mm^2. The new method was as sensitive as the radioactive detection method and showed to be identically sensitive compared to the three-step ”indirect” protocol. Using the quantitative evaluation it was shown that the enzyme amplified signal can be used in methods which require quantitative hybridisation detection.

The enzyme labelling by coupling is very efficient for short oligonucleotides and essentially sequence independent. The optimal molar ratio of streptavidin to biotinylated oligonucleotide for probe coupling was found to be 1.5:1. The direct coupling of oligonucleotide included the blocking of streptavidin free sites with the excess of free D-biotin prior to hybridisation. This was found to reduce non-specific background significantly. Although it is not clear how many streptavidin sites are available for biotin binding in a single Str-AP molecule, taking into account a non-cooperative character of biotin binding, it is likely that streptavidin binds a single biotinylated oligonucleotide instead of the theoretically possible four (Gonzalez et al., 1997). The proposed labelling method can be extended for different commercially available streptavidin- or even anti-digoxigenin antibodies-conjugated reporter groups allowing thus the flexibility of detection.

3.5 Short PNAs (peptide nucleic acid) handling

In order to use PNA as hybridisation probes it was necessary to synthesise them and introduce modifications and the labels in order to be able to detect their hybridisation. Little is known about behaviour of PNA as probes in hybridisation on DNA arrayed on membrane, so I intended to test direct as well as indirect fluorescent methods of detection.

PNA is a DNA analogues which by its chemical and physical properties is lying between DNA and peptides. In the synthesis the protection groups are similar to those which used in polypeptide synthesis, nevertheless the synthesis of PNA is easier and can be performed on adjusted machines commonly used for oligonucleotide synthesis. In the scope of this work PNAs were supposed to be post-synthetically labelled and characterised. The usual methods to analyse PNAs are HPLC (High Pressure Liquid Chromatography) and mass-spectrometry (Butler et al., 1996).Here I describe an alternative and a much more simple method for characterisation of PNAs and their modifications.

3.5.1 PNA synthesis, modification, deprotection and purification

The 8-mer probe sequences were designed according to the oligonucleotide set used in OFP (Herwig et al., 1999).

PNA was synthesised on a Universal support resin (Perseptive BioSystems). This PEG-PS support are derivatised with an amine linkage. The carboxy-terminal monomer is coupled to this linkage during the first synthesis cycle. Single coupling mode for monomers was used. Deprotection of bases and cleavage from the support is made in a single step of TFA-cresol treatment. Unmodified PNAs were stored in 0.1%TFA, but with the blocked aminogroup at low concentration also in water solutions. Several 8-mer PNA of same core with and without the linker were synthesised to test their hybridisation properties. According to the Perseptive BioSystems Manual PNAs with the linker perform better due to better solubility.

A new protocol combining synthesis on a support and modification without deprotection and using of spin columns provides quantitative purification from biotinylation modification reagents. Instead of inorganic salt buffers DIPEA (N,N-Diisopropylethylamine) was used to maintain the high pH necessary for activation of aminogroup of PNA which is easily removable by extraction with n-butanol.

For Fluorescein labelling an amino-reactive succinimide ester of carboxyfluorescein which yields carboxamide was used because it is more resistant to hydrolysis and has a higher conjugation rate than the isothionate derivatives.

To couple PNA with biotin the modification agent biotin succinimide ester -XX with the 14-atom spacer was used in order to enhance the availability of PNA then in hybridisation.

Most of the PNA were quantitatively purified on an HPLC C18 RP column because preliminary analysis of crude products showed that there were by-products or truncated PNAs which are better to avoid in subsequent hybridisations. Unlabelled 8-mer PNA eluted typically at 30-35% of the solvent B and PNA modified with biotin eluted at 37-40% of B-solvent.

3.5.2 SDS PAGE as new method for PNA handling

In addition to the existing HPLC and MALDI analysis a novel simple and reliable method to check the quality of PNA synthesis and subsequent modifications was designed.

Electrophoresis which is a typical method for oligodeoxyribonucleotides cannot be applied to PNA because of the uncharged nature of peptide-like backbone. The amino-group on the N-terminus of PNA is protonated under common conditions of electrophoresis and is not sufficient for mobility in agarose or polyacrylamide gel-electrophoresis (PAGE): PNAs stay in the loading well and slowly diffuse into the upper chamber. Nevertheless it was possible to resolve PNA by electrophoresis when an artificial net negative charge on the PNA was created by binding of SDS (sodium dodecyl sulphate).

Preliminary results with discontinuous SDS gel-electrophoresis (with the porosity of the gel varied from 12% to 20%T) had shown that PNA can be separated by SDS-PAGE. Oligodeoxyribonucleotides migrate at these conditions with the salt front (not shown). Since gels without concentrating zone gave better result they were mostly used afterwards. The buffer system thus was changed to Tris-borate-EDTA (TBE). We have found that the binding of SDS to PNA is strong enough to allow using ordinary PAGE while SDS was only added to the loading buffer and to both chambers. As shown on Figure 1, the resolution was improved with lowering of gel percentage. The increased smearing in lines 2 and 4 12 % was due to these 10-mer PNAs contained all four bases in terminal positions.

A reasonable separation was achieved when the gel was run for 30-40min at a constant current. Extended gel running times generated diffuse bands. Increased amounts of SDS in the loading buffer also resulted in diffuse bands, whereas increased glycerol concentration gave more concentrated bands (not shown). The results were independent of the amount of the PNA loaded (up to 3-5 µg per band).

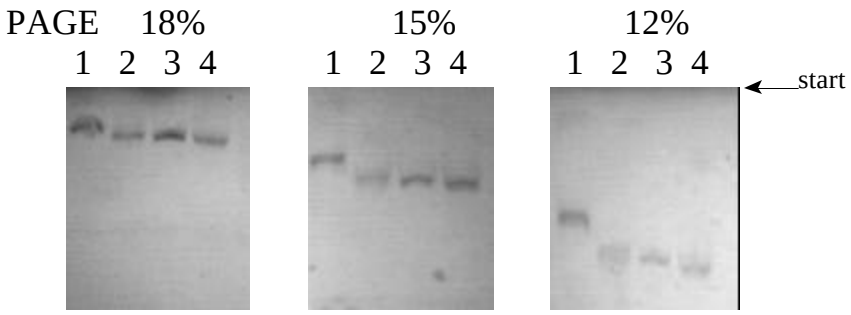


Figure 1. Dependence of mobility and separation from polyacrylamide percentage. The same four PNAs were run for 30 min in 18%, 15% and 12% PAAG. PNAs were loaded into the gel slots of 8mm and run in 1xTBE/0.1% SDS. PNA were 1) 8mer CACTCCCT, 2)10mer NCACTCCCTN; 3) 8mer TGCTCCTG, 4)10mer NTGCTCCTGN. The detection was performed by UV shadowing.

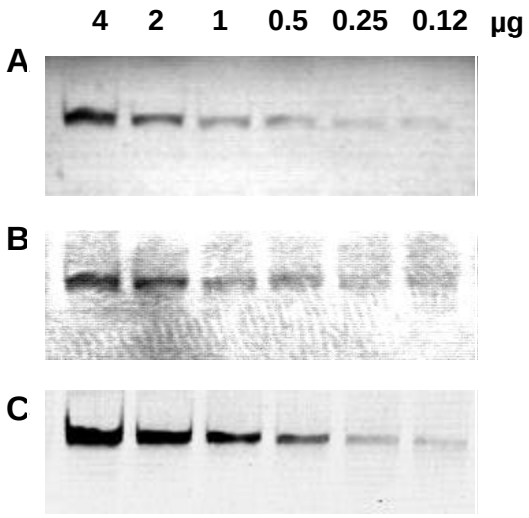


Figure 2. Comparison of different methods of PNA detection in PAGE. A two-fold dilution series of PNA #1340 from 4µg/µl to 0.1µg/µl was prepared in water. Dilutions were mixed with loading buffer and run on a 15%T gel in 1xTBE/0.1%SDS at a constant 20mA for 15 min. Different methods of detection have been applied: (A) UV-shadowing, (B) Coomassie staining and (C) SYPRO staining, according to Methods and Materials.

Different methods to detect PNA in the gel (Figure 2) were compared. The traditional oligonucleotide visualisation by shadowing on a fluorescent background (Sambrook et al., 1989) worked also for PNA. The nucleobase component of PNA absorbs UV-light and PNAs appear as a dark spot on a fluorescent background plate. This method provided quick and satisfactory detection of 0.5-1 µg of PNA per band. SDS-binding dyes could be also used to detect PNA. By the standard Coomassie-blue protein staining 1-2 µg of PNA were detected in two hours. The most sensitive alternative is fluorescent staining with SYPRO requiring an hour. The method allowed detection of 0.1 µg (around 50pmol) of PNA per band (Figure 2, C) which is at least an order of the magnitude less sensitive compared to amount of protein detected by this method (Haugland, 1996). This can be explained by diffusion of the PNA on gel during the staining procedure and a non-optimised detection system. Below, when the quantities of PNA analysed were within detectable range we have used the less sensitive but more direct detection method of UV-shadowing.

3.5.2.1 Dependence of PNA mobility on the nucleobase content and the length of the molecule

Mobility of proteins in SDS PAGE is proportional to the molecular weight of the protein with rare exceptions (Makowski and Ramsby, 1997). Interaction with SDS causes denaturation and unfolding of proteins which then bind constant amount of the detergent per mass-unit. This creates a net-negative charge proportional to the mass. By contrast, the mobility of PNA octamers, as seen in Figure 1, is very different and shows no correlation with the molecular weight (Table2).

From Figure 3 showing separation of four PNAs of the same core sequence except for one base it can be concluded that the difference in mobility is based on the nucleobase content of the PNAs. The data obtained suggest the dependence of mobility on the nucleobase content in the order G>T>C>A (PNA-G is faster than PNA-T, etc.). This observation is also confirmed by results shown in Figure 4. From left to the right (following the increasing of the mobility) the nucleobase content of PNAs changes from G/T- to A/C-rich as indicated by PNA numbers (see legend for Table 1). As seen form PNA code numbers PNAs with a higher sum of the G+T (first two digits correspond to G/T content) migrate more quickly than PNAs with a high sum of the A+C (last two digits stay for A+C). The T-rich PNA #1601 (G+T=7, A+C=1) runs more quickly than a C-rich one #0251 (G+T=2, A+C=6). Nevertheless the rule is not that

simple and there were some exceptions, for example #2123 (G+T=3, A+C=5) should have run more slowly than #2321 (G+T=5, A+C=3).

Table 2. Sequences, nucleotide content, molecular weight and mobility order of PNAs used. The name of PNA reflects the correspondent content of G,T,C and A monomers in the oligomer, e.g. in PNA #0251 is CACTCCCT that corresponds G=0, T=2, C=5, A=1. All sequences are given from N to C- terminus and sorted according to increasing G+T value. The L indicates hydrophilic spacer on the N- terminus (Methods and Materials). The molecular weights were calculated using the following values for monomers A(275 Da), G(291 Da), T(266 Da), C(251 Da).

PNA	SEQUENCE	G	T	C	A	MW
#N0251N	NCACTCCCTN	0	2	5	1	2079+
#N2330N	NTGCTCCTGN	2	3	3	0	2150+
#0251	CACTCCCT	0	2	5	1	2079
#0242	CTCACCAT	0	2	4	2	2103
#1142	CAGCCACT	1	1	4	2	2128
#1340	TTCTCCTG	1	3	4	0	2125
#2123	CACAGTGA	2	1	2	3	2192
#2321	TCACTGTG	2	3	2	1	2174
#2330	TGCTCCTG	2	3	3	0	2150
#3230	GCTGCTGC	3	2	3	0	2175
#5102	AGGGAGTG	5	1	0	2	2288
#1601	TGTTATTT	1	6	0	1	2178
#4102	AGGGAGT	4	1	0	2	1997
#L-2240	CCTGCTGC	2	2	4	0	2282
#L-2330	TGCTCCTG	2	3	3	0	2297
#L-3104	AATGAGGA	3	1	0	4	2403
#L-4112	TGGAGCAG	4	1	1	2	2395
#L-4310	TGCTGGTG	4	3	1	0	2337
#L-1610	TTGTTTTC	1	6	1	0	2302
#L-PNA-A	CCACAGCC	1	0	5	2	2260
#L-PNA-G	CCGCAGCC	2	0	5	1	2276

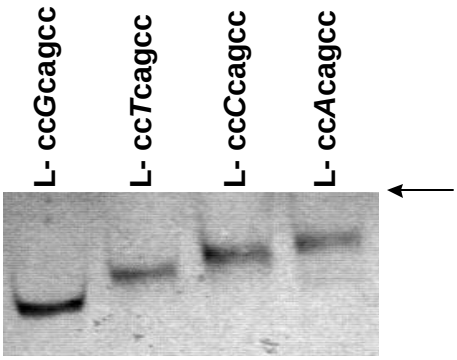


Figure 3. The influence of a single base change on mobility of PNA in PAGE. Four PNAs (L-ccNcagcc) differing by one nucleotide at a position 3 (indicated in capitals from the N-terminus) were run on a 12%T gel. Each lane contains 1µg of PNA. Arrow indicates the loading well positions.

All used PNA but three were octamers with or without modifications. The PNA #4102 is a heptamer, #N0251N and #N2330N are degenerated octamers, where N stays for all four nucleic bases. The molecular weight of the degenerated PNAs is defined only by the core octamer and it is indicated by "+".

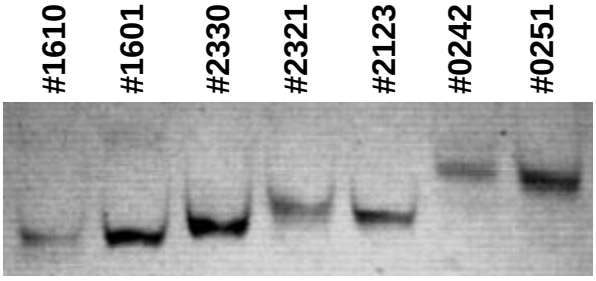


Figure 4. Mobility of seven PNA in 12%T PAGE. One microgram of each PNA in a 2µl volume were loaded on the gel (gel well of 7mm) and run at 120 V for 60 min with 1xTBE/0.1% SDS in both chambers. Detection is done by UV-shadowing.

The dependence of mobility on nucleobase content suggests similarity between PNA and DNA but dependence of mobility on the length of PNA is reversed compared to DNA. For DNA the longer it is the slower it

runs, but for PNA it is opposite (Figure 5). A seven nucleobases PNA migrated slower than a PNA of the same core with one additional nucleobase (Figure 5B, lanes 1 and 2) and 10-nucleobase-long degenerated PNA runs more quickly than its core octamer PNA (Figure 5B, lanes 3-6).

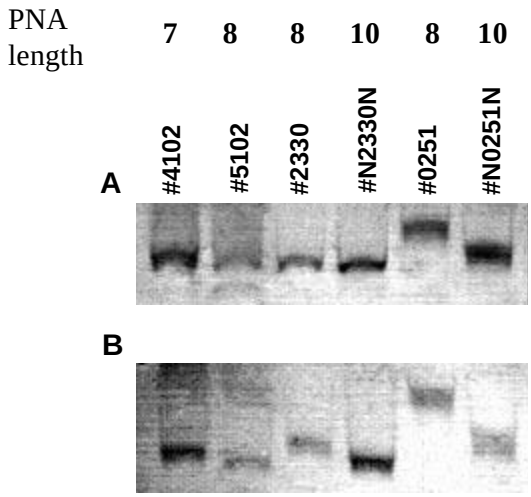


Figure 5. Mobility of PNA vs. PNA length. The mobility is inversely proportional to length in a pairwise comparison. The PNAs #4102 and #5102 have the same 7-nt-long core sequence but #5102 is longer by 1 base. #N2330N and #N0251N are longer than the corresponding #2330 and #0251 by 2 bases. PNA were run for 20 min on 15%T PAGE (A) or on 10%T PAGE (B). Two different running conditions were chosen to maximise resolution and minimise diffusion differences for degenerated 10-mer PNA. The figure shows, in addition, that G-rich PNA #5102 runs quicker than its C-rich complementary PNA #0251.

3.5.2.2 SDS separation of PNA modified with fluorescein and biotin

Comparative analysis of modified and non-modified PNA was performed using an electrophoretic system as described above. Modification should block the positive charge of the amino-group and thus might influence mobility on the gel.

Three octamer PNA of the same sequence TGCTCCTG, but modified with either biotin, fluorescein or an amino-linker exhibited the different mobility under PAGE as shown on Figure 6. Fluorescein modified PNA showed the lowest mobility (lane 1, Fig. 6) as do fluorescently modified oligonucleotides (Urdea et al., 1988). In contrast, modification with biotin that usually reduces DNA mobility (Urdea et al., 1988) increases the mobility of PNA, possibly by blocking of the only positive charge of PNA. Addition of the linker also increased PNA mobility compared to the unmodified molecule (lane 2 and 3, Figure 6) although the amino-end charge was unchanged. This may be caused by a similar

mechanism as the increased mobility of the PNA with increased length. The difference in mobility between unmodified PNA and biotin-PNA was greater than the difference in mobility between L-PNA and biotin-L-PNA.

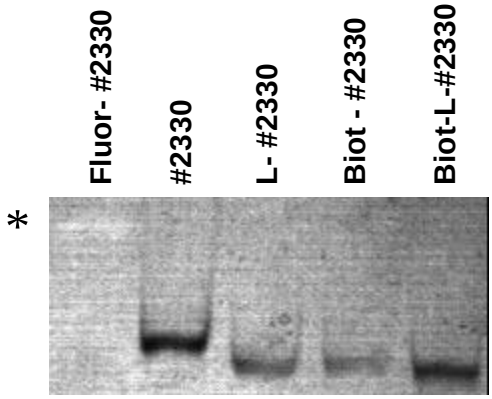


Figure 6. Mobility of PNA with different modifications on the N-terminus. Unmodified PNA #2330 has been labelled either with fluorescein (Fluor-#2330) or biotin (Biot-#2330). In addition, the PNA L-#2330 containing linker (see Methods) and the same PNA labelled with biotin Biot-L-#2330 have been used. UV-shadowing reveals fluorescently labelled PNA as a bright rather then the dark spot in comparison to the other PNAs. The asterisk (*) indicates the position of a fluorescent spot.

3.5.2.3 Detection of PNA, PNA/DNA duplexes and DNA on the same gel

SDS gel system also allowed to access hybridisation properties of unlabelled PNA. Using PAGE it was possible to detect PNA, PNA/DNA and DNA within the same gel. Two PNAs, differing by one nucleotide and complementary to either of 19-nt long DNA oligodeoxyribonucleotides (5'-TTTGGCTGC/TGGAGGGAGTG) were hybridised at 50µM concentration with both perfect and mismatched targets for 10 minutes at ambient temperature in 2 µl of 20mM Tris-HCl pH 8.0. After hybridisation the samples were mixed with loading buffer and loaded on the gel. No increase in gel temperature was observed during the electrophoresis.

The duplex formation can be identified by UV-shadowing, revealing PNA, DNA and PNA/DNA complexes as separated distinct bands (Figure 7, left). Under these conditions DNA migrated faster (closer to the dye front) than PNA or PNA/DNA as described before with non-SDS gel system (Perry-O'Keefe et al., 1996). After staining with SYPRO (Figure 7, right), free DNA was not detected.

The PAGE method permits to characterise hybridisation behaviour of PNA. Figure 7 shows that PNA-G was able to hybridise to both complementary and mismatched targets, although the intensity of the mismatched

duplex was lower. No shift was observed for PNA-G after reaction with non-complementary DNA oligonucleotides under the same hybridisation of oligonucleotides.

The mismatched hybridisation of PNA-G may be explained by higher stability of the G:T mismatch. Melting temperatures of mismatched pair PNA-G: DNA-T was 40°C compared to PNA-A:DNA-C of 37°C. Sensitive SYPRO staining, revealed some dissociation of weakly interacting DNA and PNA in the mismatched G:T heteroduplex band which appears as weak compared to the G:C and A:T bands. SYPRO staining revealed a very weak band at a position corresponding to that expected for a mismatched heteroduplex for PNA-A (gel line A:*T*), whereas UV shadowing detected a duplex for PNA-A only with its complementary DNA.

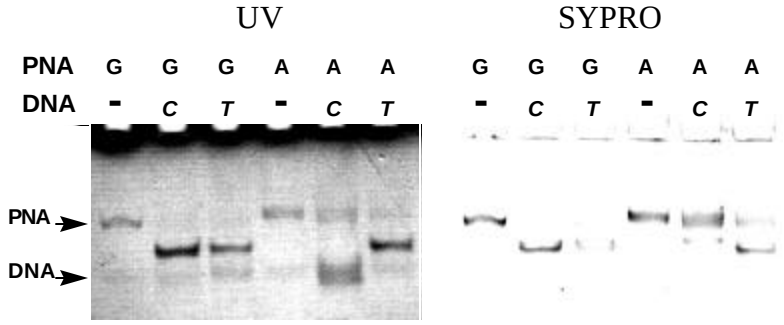


Figure 7. Gel electrophoresis PNA/DNA hybridisation assay. PNA-G and PNA-A (L-ccGcagcc and L-ccAcagcc) were hybridised with two 19-nt-long oligodeoxyribonucleotides, DNA-*C* and DNA-*T*. Each of the oligonucleotides forms perfect duplex with one PNA and mismatched with another. Hybridisation reactions were loaded on 15%PAGE and run for 20min in 1x TBE/0.1%SDS. buffer. Detection was performed by UV shadowing (left panel) and by SYPRO staining (right panel). Using SYPRO staining only PNA and PNA/DNA are detectable. The arrows indicate positions of the gel corresponding to either PNA or DNA (in the A/*C* position). The bands in between correspond to PNA/DNA duplexes.

3.5.3 Discussion

In contrast to DNA oligonucleotides analysis and detection of PNAs has only been performed using methods like HPLC and MALDI, because PNA, unlike DNA, does not show any mobility in the conditions of slab gel electrophoresis. No methods were available which do not require labelling of PNA would allow to detect PNA and PNA-DNA duplex simultaneously.

Here it is shown that PNA, after incubation with SDS, can be successfully separated by polyacrylamide gel electrophoresis. Similar to binding of hydrophobic parts of proteins, SDS seems to bind to PNA nucleobases. Nevertheless, mobility of PNA in the gel does not follow rules common to proteins and DNA.

Unlike oligonucleotides or peptides, for a given PNA sequence, longer PNA migrates faster under PAGE conditions than a truncated version (Figure 4), which may be explained by a number of factors. SDS binding to the PNA units increases negative charge of the 10-mer as compared with the 8-mer. The extra charge from SDS should increase mobility of longer PNA only if the size of resulting SDS/PNA micelles remains almost constant. The results support this hypothesis for short PNAs (<10-mer) revealing that mobility of PNA is proportional to the length of PNA molecules, so it is preferable to use the term length rather than molecular weight.

Evidence for the binding of SDS to nucleobases of PNA is given by the dependence of mobility in the gel on the nucleobase content of PNA for a given length, which shows more pronounced effects for shorter PNAs G>T>C~A. An effect of the base-specific mobility has been described for oligodeoxyribonucleotides (Bowling et al., 1991; Frank and Koster, 1979) where the trend follows the polarity order of nucleobases C>A>G~T. (C-rich oligonucleotides were faster than T-rich ones).

In earlier work it was observed (Hoffmann et al., 1998) that the retention times of different PNAs on reverse phase HPLC also showed base-content dependency. In HPLC the mobility gradient was determined as C>A>T>G, explainable by different degrees of base ionisation under highly acidic conditions (Hoffmann et al., 1998).

PNA mobility in gels shows an opposite trend that is GT-rich sequences migrate faster than CA-rich. This cannot be attributed to differential ionisation (Schmidt et al., 1996) because under PAGE conditions (pH~8) the nucleobases remain uncharged. It is plausible that non-polar side groups of nucleotides of the PNA molecules mediate binding of SDS with different efficiencies. We speculate that interactions of nucleobases with SDS is similar to the kind of interactions between nucleobases with acrylamide polymers. However, interaction with acrylamide decreases the mobility of DNA, whereas interaction between SDS and PNA in contrast provides strong negative charges and therefore increases mobility. This effect may explain the opposite trend in base-specific mobility between DNA (C>A>G~T) and PNA (G~T>A>C) for PNA.

The influence of charge gained by SDS binding to PNA on mobility was further examined with experiments using N-terminal modifications of PNA at the N-terminus, which should result in an increase of total negative charge. Modification with biotin leads to protection of the free aminogroup and results in increase mobility. However, an additional interaction of SDS with biotin cannot be ruled out. Addition of a linker that retains a N-terminal group, is

probably analogues to addition of another base. Even though this linker should increase solubility of PNA (manufacture's manual), it appears to more efficiently bind to SDS causing faster migration under electrophoresis. The observation that biotin-modified PNA shows increased mobility compared to PNA with a carbohydrate linker indicates that the "charge effect" is stronger than the "length effect".

Modification of a single PNA using two groups - biotin and/or linker - also caused increased mobility. Interestingly for PNA including the linker, the difference in mobility between the biotin modified and non-modified PNA was less pronounced than that for PNA without the linker. This may indicate that beyond a certain PNA size, mobility will not increase because of a different structure or geometry of PNA/SDS micelles. The stronger effect of the gained charge may be reversed and PNA will show length dependence of mobility similar to DNA fragments.

These data demonstrate that SDS PAGE can be a tool for PNA characterisation. Although gel analysis of PNA is not as informative or sensitive as mass-spectroscopy or HPLC, it is sensitive enough to detect traces of by-products and overall failure of synthesis. In fact, it opens a new field of analysis for non-charged DNA-analogues that may be further optimised by applying capillary or micro-channel systems (Yao et al., 1999). In addition, the gel system can be an inexpensive alternative for purification of PNA after synthesis.

SDS-PAGE also provides a tool for pilot examination of hybridisation behaviour, which can be effectively performed for several PNA species in parallel. Usually PNA/DNA duplex formation is detected by a mobility shift with DNA labelled as a probe. The labelled DNA fragments anneal with PNA and the PNA/DNA complexes are separated from free DNA.

So far, in-gel analysis of hybrid PNA duplexes required the use of reporter compounds since most of traditional DNA intercalating agents do not work for a PNA/DNA hetero-duplex (Armitage et al., 1998). Using less than 50pmol of PNA, PNA/DNA hetero-duplexes can be detected using our SDS approach without any introduction of fluorescent labels. Simultaneous detection of PNA, DNA and PNA/DNA on one gel eliminates the need for indirect comparisons such as the detection of only free DNA and its complex with PNA. Our approach should facilitate kinetics studies of PNA/DNA hetero-duplex formation. The specificity of interaction and single mismatch discrimination was demonstrated using this assay. The possible drawback of method described here is that the temperature dependence of hybridisation cannot easily be monitored since re-association occurs very quickly if samples were transferred to ambient temperature, i.e. during loading on gel.

The influence of SDS on formation (as well as mobility) of PNA/DNA duplexes during gel electrophoresis is yet unclear. Possibly, SDS changes the specificity of PNA/DNA interaction by introducing a negative charge to PNA.

3.6 PNA as probes for oligonucleotide fingerprinting

The uncharged DNA analogues, PNAs (peptide nucleic acid) are the most suitable alternative to DNA oligonucleotides to use in OFP. PNAs possess many advantageous properties when compared to DNA. PNA/DNA duplexes have higher thermodynamic stability than DNA/DNA duplexes of equal length and sequence. Higher stability in hybridisation allows the use of shorter probes than conventionally used octamer DNA, for example, 7-mers. The possibility of better mismatch discrimination in PNA/DNA duplexes (Orum et al., 1993) allows a more specific hybridisation.

Until now most of the data on PNA hybridisation was obtained in systems where PNAs were hybridised with comparable concentrations of DNA oligonucleotides in solution (Schwarz et al., 1999), (Giesen et al., 1998) or else PNA were attached to supports and hybridised with DNA (Igloi, 1998),(Weiler et al., 1997). In OFP the DNA target is attached to a solid support, and the PNA probe is used in a large excess compared to the target. In this work the hybridisation properties of short PNAs in such a system are described for the first time.

3.6.1 PNA hybridisation detection methods

PNAs containing biotin on the 5'-end were labelled with Alkaline Phosphatase and detected chemiluminescently as it has been described for oligonucleotides, below. The 8-mer PNA-enzyme conjugates were much more stable because they could be hybridised at +25°C temperature (Figure1) whereas 10-mer DNA-enzyme conjugates hybridises only at +8°C (not shown).

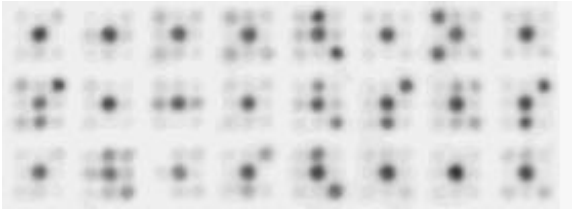


Figure 1. Hybridisation of an 8-mer PNA of the sequence CCACAGCC labelled with Alkaline Phosphatase to an array of DNAs at room temperature for 30 min. Hybridisation was detected with chemiluminescent substrate

CPD-Star by Lumi-Imager as in (Guerasimova et al., 1999). The DNA clones were spotted in a 3x3 pattern in duplicates with a central guide dot of salmon sperm DNA at a concentration of 100ng/µl.

In further experiments to label PNA the same method of labelling via biotin-streptavidin link with fluorescent protein Phycoerythrin was used, so that enzymatic signal amplification as with the Streptavidin-AP label and additional incubation steps with the Alkaline Phosphatase substrate were avoided. The fluorescent signal was detected with a custom-built laser scanning device (Nyarsik, in preparation)

3.6.1.1 Comparison of Phycoerythrin and Fluorescein labels sensitivity

For the following experiments the Phycoerythrin protein label was attached to the biotinylated PNA AGGGAGTG, via a Streptavidin-biotin link. Fluorescein was introduced via the amino-terminus of an aliquot of the same PNA. Hybridisation was performed under the same conditions, i.e. 30 min of hybridisation at room temperature and 30 min washing.

For signal quantification it was necessary to immobilise a dilution series of DNA on a membrane. Robotic spotting suffers from the drawback of pin-to-pin differences which result in uneven transfer of liquid onto the membrane and is thus not optimal for quantification. To produce a reliable DNA dilution series a piezo-ink jet system of liquid micro-delivery (Eickhoff, 1998) was used. With this system the solutions are delivered onto a membrane by dispensing 1, 2, 4, ... etc. drops per position.

The PCR products from 10 clones were pooled and purified using a Qiagen PCR purification kit to a final concentration of 100ng/µl, fixed on the membrane (3x5cm) and used for hybridisation.

The fluorescent signal was detected with 20µm resolution scanner (Nyarsik, in preparation) using an Argon-ion excitation laser (488nm for excitation) and by-pass filters of 570nm for Phycoerythrin and 520nm for Fluorescein. The excitation wavelength is at the maximum of absorption for both fluorophores: λ phycoerythrin = 498nm, λ fluorescein = 495nm (Haugland, 1996). The final value for each DNA dot was taken as the mean from 3 values. The signal measurements were taken using different laser intensities (90%, 80% and 70%) to prevent saturation of the signal at the highest intensity. The maximum resolution of the PhosphorImager of 25 µm used to detect signals from radioactively labelled probe on phosphor storage screen was not enough to detect signals even after an exposure of 48 hours (not shown) so the accuracy of dilution could not be confirmed with a control radioactive hybridisation.

The conclusion of the experiment was that PNA coupled with Phycoerythrin at 2nM concentration was sufficient to gain hybridisation signals of the same intensity as obtained with 2µM of Fluorescein-labelled PNA. This experiment clearly shows that Phycoerythrin is 3 orders of magnitude more sensitive than Fluorescein as a label.

However, calculation of curve-fitting coefficients, k, as not equal to 1, showed there was no linearity in the detection. (Figure 2).

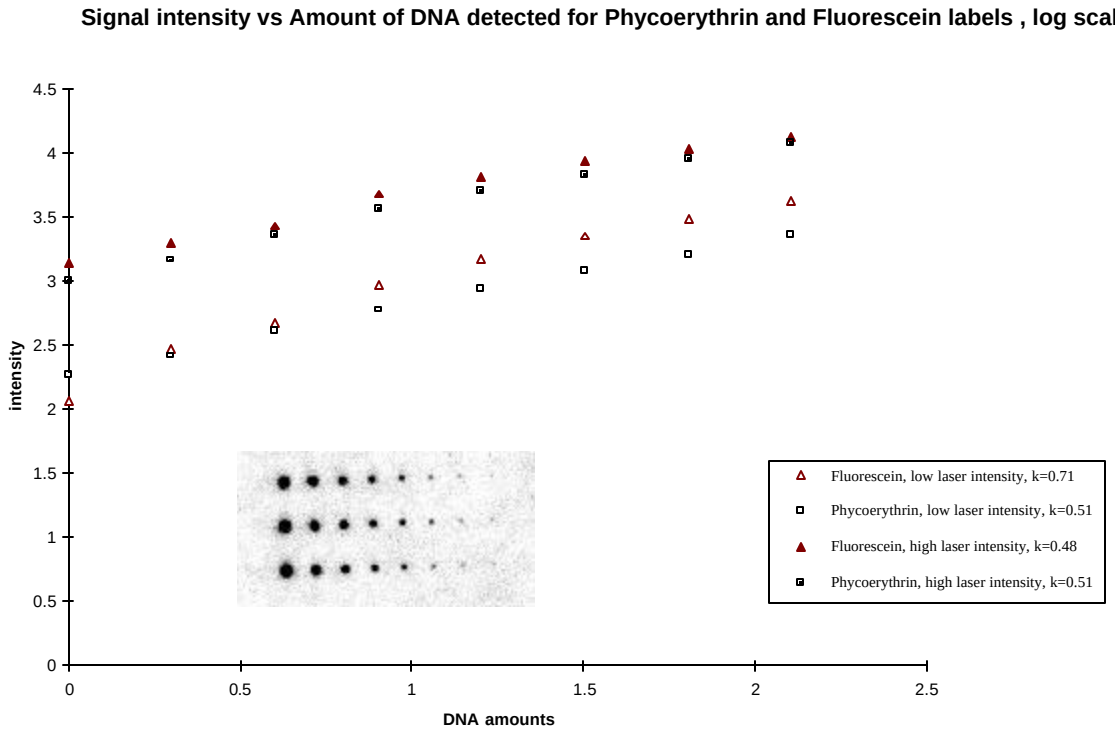


Figure 2. Hybridisation signal of Fluorescein- and Phycoerythrin-labelled probes versus amounts of DNA (taken as number of drops in each dot) plotted on a logarithmic scale. DNA dilutions were transferred to membrane and hybridised with Phycoerythrin-labelled PNA (2nM) and Fluorescein labelled PNA (2µM), the image from PNA-Fluorescein hybridisation is shown in the inset. The signal values were taken at the highest (closed symbols) and the lowest (open symbols) laser intensities for both dyes and plotted versus DNA amounts. The signals were quantified using TINA (Raytest) software. The curve fitting coefficients, k, were calculated by the least square method (Excel 7.0).

Indirect detection, where the biotinylated PNA is first hybridised, washed and then, in a separate step, detected by incubation with Phycoerythrin-Streptavidin conjugate, was also performed. Signal quantification showed that here was also no linear dependence between the amount of DNA and the signal. However, to provide similar sensitivity, the concentration of conjugate should be ca. 50 times more than used in direct coupling and hybridisation.

3.6.1.2 Reproducibility of detection by fluorescent scanning

The reproducibility of PNA hybridisation and detection was examined by repeated hybridisations of the same probe. The PNA probes were labelled and hybridised 3-5 times to the same filter copy containing 700 control clones spotted with the gridding robot (as described in Methods).

The diagram on Figure 3 shows a scatter plot of absolute intensity values from 2 subsequent hybridisations. The signal value for each spot was plotted against the corresponding signal in the second experiment. The correlation coefficient of the two data sets was 0.98. As seen on the Figure 3 the largest deviation was observed for low intensity values. Similar results were obtained with the other data sets.

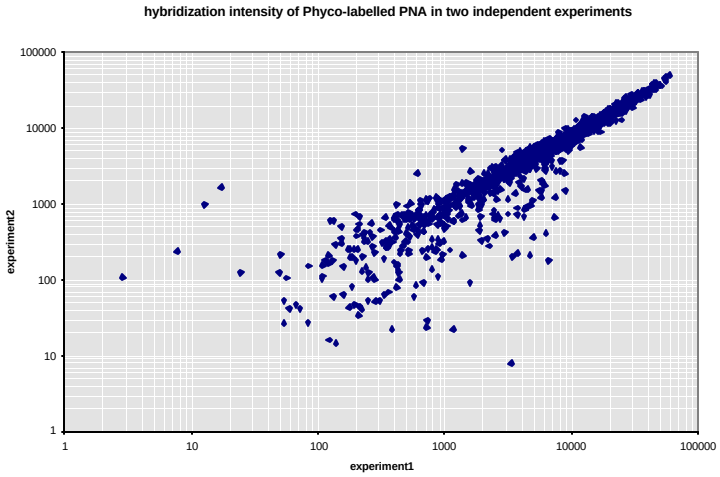


Figure 3. Scatter plot of the intensities obtained in hybridisation of Phycoerythrin labelled PNA in two independent experiments using the same DNA filter copy, shown in logarithmic scale.

3.6.2 Hybridisation properties of PNA conjugated to Phycoerythrin

The conditions of PNA hybridisation were tested using the arrayed sequenced clones (Results1.1). The concentration of PNA was found to be a critical parameter: below 20nM the signal intensity after long incubations was low (not shown) due to a low association rate.

3.6.2.1 Salt and detergent concentration influence

It is known that PNA hybridises to DNA at a low ionic strength and that an increase in ionic strength impairs selectivity of interaction with DNA (Tomac et al., 1998). The dependence of hybridisation signals on salt and detergent concentration was examined by variation of these parameters. The starting concentration was taken from conventional experiments with oligonucleotides. As a detergent in the hybridisation buffer the sodium N-lauroyl sarcosinate was used (Drmanac et al., 1990) because it can be used at low temperatures, where SDS precipitates.

Identical filter replicas were hybridised in buffers with gradual increases of NaCl concentration from 5 to 500 mM and a constant 7.5% sarcosinate. No change in hybridisation signal was observed. This is in contrast to data reported earlier on impaired specificity of PNA hybridisation in solution (Peffer Nancy et al.,).

The hybridisations were also performed in buffers with different sarcosinate concentrations (7.5%, 5%, 2.5%, 1% and 0.5%) and a constant 5 mM concentration of NaCl. Changing sarcosinate concentration also influences the salt concentration since sarcosinate exists in a form of sodium salt ((MW=300 g/mol) is ca. 0.25 mM Na^+). No signal reduction was observed, though a slight increase in background signal was observed if the concentration of detergent was below 1%.

The conclusion was drawn that in the system used NaCl concentration cannot be used to influence specificity of hybridisation. In further experiments detergent was used at 7.5% concentration in order to prevent possible protein-protein interactions and unspecific binding to membrane. The salt concentration used was 5 mM NaCl.

3.6.2.2 Temperature dependence of specificity

The hybridisation temperature of oligonucleotides is usually chosen to be lower than the melting temperature (T_m). Melting behaviour and stability of duplexes are functions of the nucleotide content of an oligomer (Wetmur, 1991), (Breslauer et al., 1986). For 8-10mer DNAs the optimal temperature of hybridisation was estimated as below 20°C and conventionally 4-10°C has been used to assure stability in the hybridisation.

PNAs were expected to be much more stable and to hybridise generally at a higher temperature than DNAs of corresponding lengths. To test this, hybridisations with Phycoerythrin-coupled PNAs were performed at different temperatures in order to find the optimal temperature range.

The first observation made was that adjusting the hybridisation temperature had a much more significant effect than an increase in time and temperature of washing, and resulted in better performance of PNAs in hybridisations. If a PNA was hybridised for 10-30 min, the dissociation of probe increased with the increased time and/or washing temperature. After hybridisation at +25°C and washing for 30 min at +25°C, +35°C or +45°C, the specificity increased, as expected, but the intensity of all signals decreased.

The influence of hybridisation temperature was tested by hybridisation of identical filter replicas at +25°C, +35°C +45°C and +50°C for 30min using the PNA probe AATGAGGA[1] (with a single linker). This probe gave a high number of “false” positive signals (or so called background signals) when hybridised at +25°C (Figure 4, A). Change in signal intensity for the same clones with the increase of temperature is shown on Figure 4, B. The best result was achieved at +45°C, at +50°C the hybridisation signal was considerably lost. Notably, Phycoerythrin fully retained its fluorescence during incubation at +35°C or at +45°C even after 16 hours of hybridisation.

The ratio of the mean perfect to the mean background signal changed at different hybridisation temperatures from 2:1 at 25°C to a ratio of 5:1 at +45°C. This indicates that at low specificity conditions (low temperature) the signals from non-perfect hybridisation were on average high and comparable to the signals from clones which contained a complementary sequence.

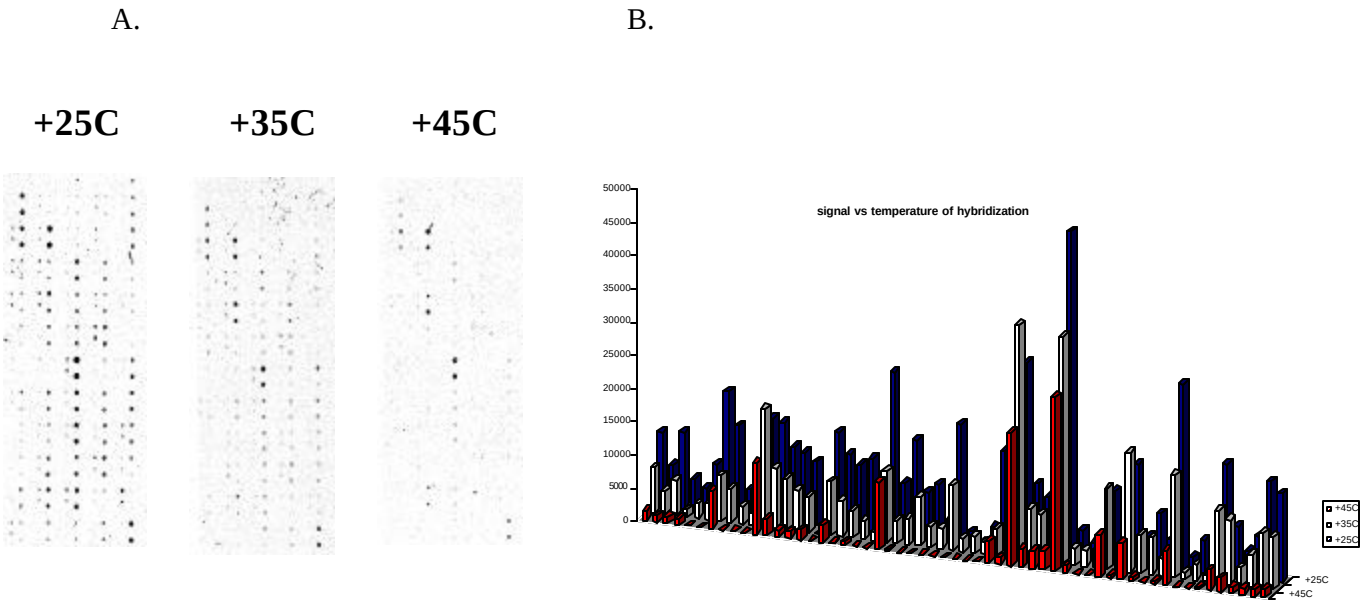


Figure 4. A. Fragment of images containing the same section of a DNA array. 3 filter replicas of a DNA array hybridised with a Phycoerythrin labelled PNA at different temperatures. When hybridised at +45°C only a few positives were still present. **B**. The signal intensity values for 60 randomly chosen clones (ca 10% of all clones). The graph shows the drop of signal to zero in 80% of these clones as the temperature was raised from 25°C (blue bars) to 45°C (red bars), which reflects increased selectivity of hybridisation.

All PNAs were hybridised at different temperatures, starting from +25°C, to define optimal conditions for each single PNA. If a PNA did not give any signal in the hybridisation at +25°C, it was tested at +8°C, a temperature which is commonly used for DNA oligonucleotides. The optimal temperature was defined as the highest temperature, where the signals were detectable and the ratio of the highest signals to average background signal was >2:1. The results are summarised in Table 2. Results from several 7-mer PNAs tested showed very high stability of these probes, such as AGGGAGT (hybridisation temperature +45°C) or AATGAGG (+35°C)

Interestingly, the optimum hybridisation temperature for a PNA coupled to Phycoerythrin differed considerably from estimations based on the T_m of an unmodified PNA. For example, according to the measured (Methods) T_m of an 8-mer PNA, CCGCAGCC[1] (G+C=7), of 57°C or of a PNA CCACAGCC[1] (G+C=6), with a single base change G:A, of 49°C the hybridisation temperature should be at least +35°C. However at this temperature no signal was observed for either PNA and the empirical hybridisation optimum for both PNAs was found to be +8°C.

The empirical temperature optima did not correlate with those determined by enthalpy ΔH (enthalpy was determined using the nearest -neighbour method (Breslauer et al., 1986)) which reflects thermodynamic stability. For example, for AGTGGCTG and CCTCCTGC (optimal hybridisation temperature +8°C) the ΔH was 55.8 and 60.1 kJ/mol, correspondingly, which was higher (thus presumably higher the stability) than for PNA AGGAGTAA (hybridisation optimum +45°C) with ΔH calculated as 53.8 kJ/mol.

As seen from Table 2, there is a correlation between purine (A+G) content of the PNAs and the temperature optimum of hybridisation: the higher the purine content, the higher the hybridisation temperature optimum.

The length of linker arm did not seem to have an effect (e.g. hybridisation temperature for CCTCCTGC[2] (A+G=1) of +8°C and TGGAGCTG[2] (A+G=5) of +35°C. The exception from 20 PNA analysed was the PNA AGTGGCTG (A+G=5, hybridisation optimum +8°C), which however did not have additional linkers.

Table 2. Hybridisation temperature optima for PNAs of different nucleobase content. ([1]) is used to mark PNAs with single spacer arm linker, ([2]) for PNAs with double linker, the other PNAs did not have a spacer between the PNA and the biotin group.

sequence of PNA	Optimal hybridisation temperature	A+G content
aggagtaa[2]	+45°C	7
ccaggagg[2]	+45°C	6
aatgagga[1]	+45°C	7
gacgagg[2]	+35°C	6
ctggagca[2]	+35°C	5
aatgagg[2]	+35°C	6
ggagcagc[2]	+35°C	6
tggagctg[2]	+35°C	5
tgctggtg[1]	+35°C	4
tggagcag[1]	+35°C	6
ctggaag[2]	+25°C	5
gaagcaga[2]	+25°C	7
cctcctgc [2]	+8°C	1
tcctcctg [1]	+8°C	1
ccgcagcc[1]	+8°C	2
agtggctg	+8°C	5
cagccact	+8°C	3
tgctcctg[2]	+8°C	2
cctgggca[2]	+8°C	4
tggctcctg[2]	+8°C	3

3.6.2.3 Dependence of rate of PNA hybridisation on time and temperature

The same PNA probes were hybridised at temperatures of +8°C and +35°C for different times (10min, 30min, 90 min and 3 hours) with identical DNA filters. The hybridisation was stopped at specified times by washing in an excess of hybridisation buffer for 30 min. The results for the +8°C hybridisation showed that after 10 min of hybridisation the signal was very low, however, after 30 min of hybridisation at +8°C the mean background to perfect signal ratio was comparable to that of hybridisation for 10 min at +35°C. This shows the dependence of the reaction rate on temperature, i.e. the rate is higher at a higher temperature.

Comparison of the time course at different temperatures tested has shown that after longer incubation the mean background to perfect signal ratio increased considerably. Figure 5 shows that after 10 min of hybridisation at +35°C the signal intensities for clones in the row marked with an arrow were the same for most of the clones in this row, but after 3 hours of hybridisation certain clones have considerably higher signals and a total, background hybridisation was reduced.

The equilibrium (no further increase in signal) for both temperatures tested was reached in 3 hours of hybridisation so this time was used in further experiments as standard. At hybridisation times up to 5-16 hours the signal intensities were essentially the same.

A. B.

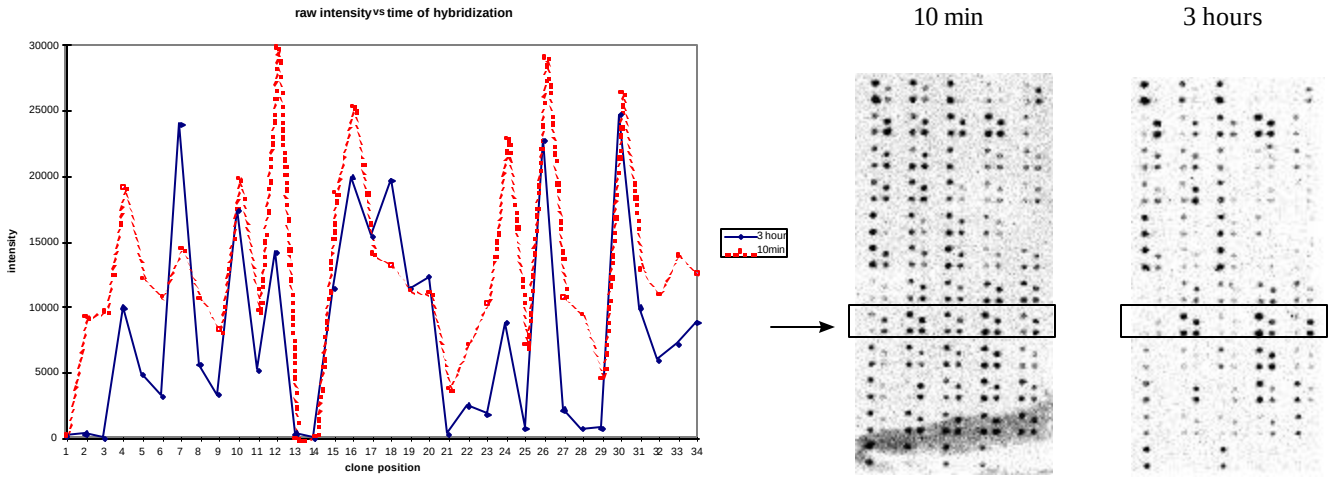


Figure 5. Dependence of signal of PNA hybridisation on time. The scheme (left panel) shows the change in absolute intensity for 70% (20 from 34) of randomly chosen 34 clones as hybridisation time increased from 10 min

(dotted line) to 3 hours (solid line). On the right panel are shown identical filter replicas of the DNA array which were hybridised with Phycoerythrin labelled PNA for the specified times and then washed and detected simultaneously. Each block of 4 dots contained 2 different DNA clones in duplicates. The arrow points out changes in signal intensities from relatively equal signals obtained from each clone (image of 10 min hybridisation) to brighter signals from specific clones (3 hour hybridisation) and considerably reduced signals from other clones.

3.6.3 Sequence specificity of PNA hybridisation

3.6.3.1 Control hybridisation of complementary PNAs

The high number of non-specific positive signals make it necessary to evaluate the sequence specificity of PNA hybridisation. Special probes were designed to provide an internal control for the specificity of PNA/DNA interaction. Two complementary PNAs (AGTGGATG and CATCCACT) were hybridised separately and the patterns of their hybridisation were compared. Sequence specific hybridisation should result in two identical patterns of hybridisation. The graphs used to visualise the hybridisation patterns also allow the correlation of the sites of the perfect match for probes and positions of reproducibly positive clones.

Two methods of integrating hybridisation values were used. In the first method the absolute signal values as obtained after image processing were summed for each nucleotide position and plotted. (Figure 6, A). As seen from the figure, one of the PNAs, PNA1, had higher signal intensities. This method however did not provide information on the proportion of clones hybridised in each group of redundant clones.

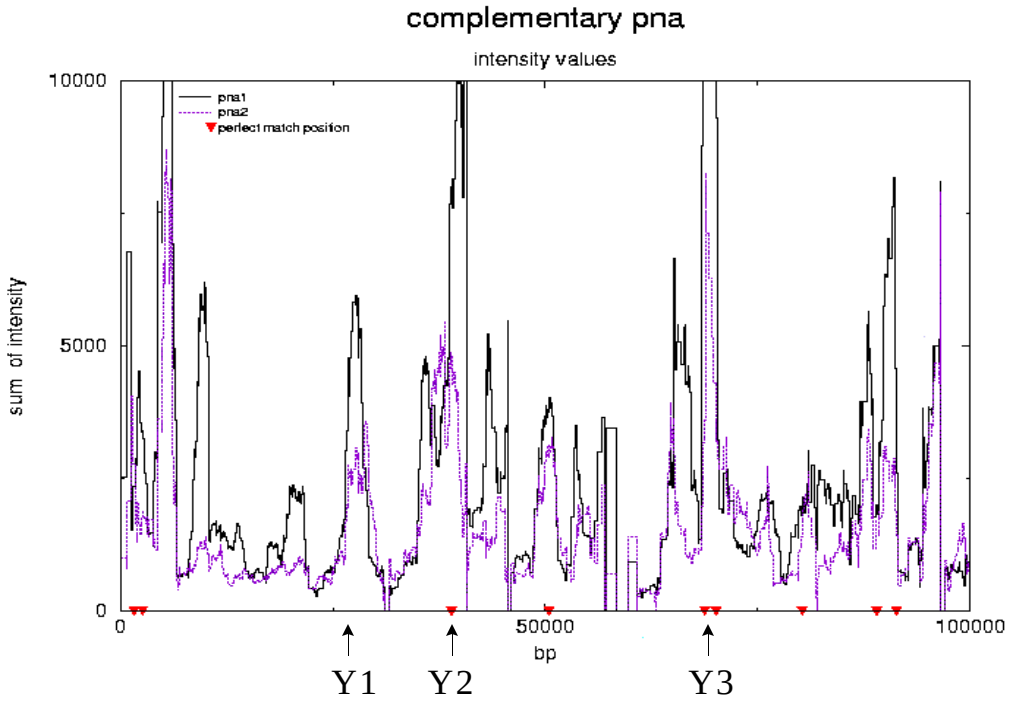
In the second method a rank from 0 to 1, as in (Methods), was assigned to each hybridisation signal value and then only these clones with intensities above a threshold of 0.9 were taken into account (Figure 6, B).

The graphs produced by either method show high similarity of profiles for the complementary PNAs AGTGGATG (PNA1) and CATCCACT (PNA2). For both PNAs there are highly reproducible regions with no match to the PNA sequence, such as peak Y1. Some stronger peaks (Y2 and Y3) corresponded to perfectly matched sites (Figure 6,A). This means that there is sequence specificity in hybridisation.

The patterns of hybridisation of the two PNAs are more similar if the comparison of ranked intensities is examined. The difference in the patterns, seen as non-matching peaks (Figure 6,B) corresponded to the low reproducibility of positive hybridisation (less than 0.3 or 30% positive clones from total number of clones in an overlap), and all these are false positives (no complementary match). A plausible explanation for the differences between the two probes is the different stability of corresponding mismatched base pairs which are more stable in the

case of one PNA but turn into a very destabilised pair in hybridisation of its complement (e.g. G:T mismatch in one case and A:C in the other).

A.



B.

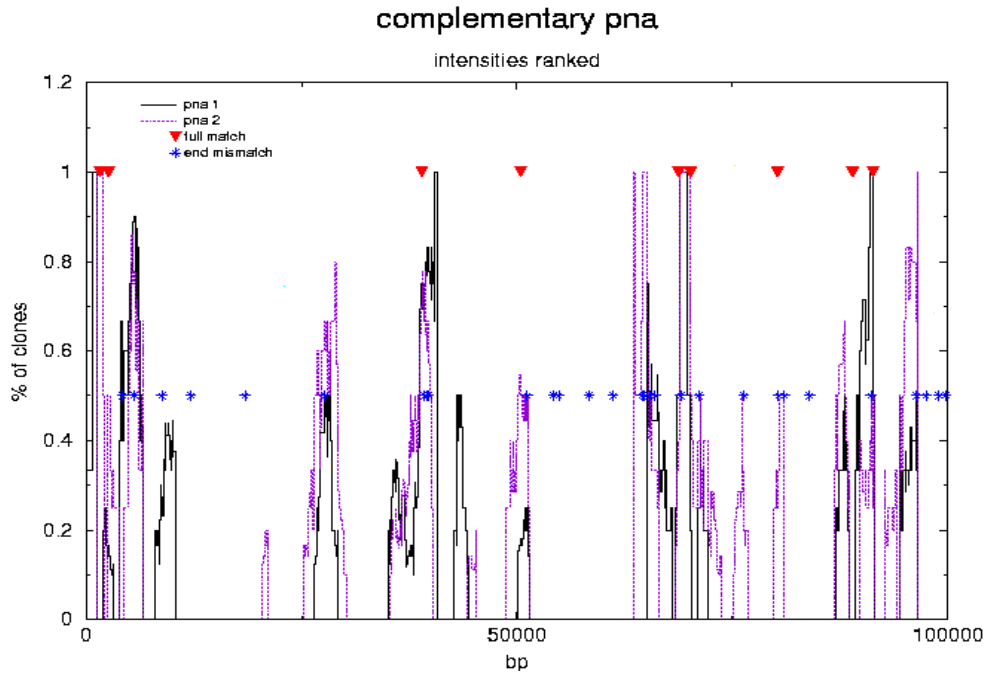


Figure 6. Distribution of positive clones for two separate hybridisations of complementary PNAs. To generate the plots the integrated signal intensity values(A) or ranked intensities(B) obtained in hybridisations were summed

as described in Methods. The hybridisation pattern of PNA1, GAGTGGAT, is shown with a solid line (black), for PNA2, CATCCACT, with a dotted line (purple). The X-axes show the nucleotide position in the contig, for which either a sum of intensity values from all clones containing this nucleotide is shown on the Y axes(A) or for the ranked intensity values above a threshold are summed(B). The sum of intensities or ranks is divided by the number of clones in the region.

The sites of the perfect (complementary)matches for the both PNAs are shown with red symbols. The sites where the intensities of signals were considerably different between PNAs are shown as Y1, Y2 and Y3 (A).

3.6.3.2 Hybridisation of PNA in parallel orientation

In liquid phase, PNAs are described as hybridising in parallel orientation (amino-terminus of PNA to the 5'-end of DNA) in addition to the usual anti-parallel (amino-PNA to the 3' of DNA) although with less efficiency (Peffer Nancy et al., 1997). To assess this property in the system of the membrane bound DNA two PNA sequences, amino'-GACGAGGT(PNA1) and its reverse sequence, amino'-TGGAGCAG (PNA2), were used. According to this, a complementary sequence to the PNA2 3'-accacgtc-5' should be complementary to PNA1, GACGAGGT, in parallel orientation.

Scheme 1. Hybridisation of PNA in anti- and parallel orientation to the same DNA target.

```
5'-TGGAGCAG   PNA2, antiparallel
3'-TGGAGCAG   PNA1, parallel
3'-accacgtc   DNA
```

Diagrams of positive clone distribution should show similar profiles if there were equal hybridisation in the parallel orientation.

PNA1, GACGAGGT, had no fully matched sites in the classic, antiparallel, orientation in the entire assembly of 100Kbp, thus it is called "negative". Nevertheless, there were positive clones in hybridisation with this PNA at a relatively high (+35°C) hybridisation temperature. The reproducibly positive clones' peaks only partially correspond to the "reversed" sequence complementary sites (Figure 7). The diagram shows similar distribution of the peaks (13 common peaks from a total of 21) for two probes, though their coincidence with the perfect sites for both probes is low. It can be concluded from the patterns of hybridisation, that the same clones hybridised to the PNAs in either orientations. The data obtained agree with the less efficient hybridisation of PNA in a parallel orientation described earlier, which is seen on the plots as the lower reproducibility of positive hybridisation to the "negative" probe.

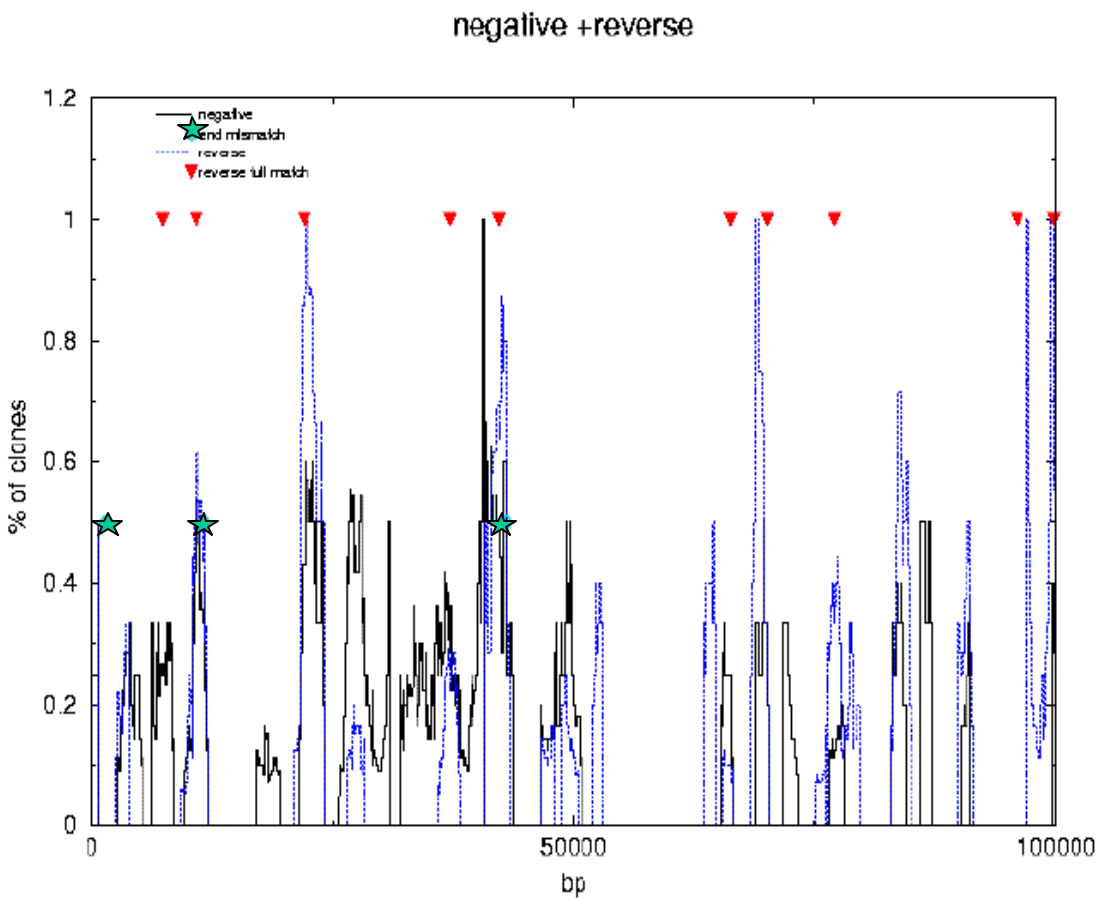


Figure 7. Comparison of profiles obtained by hybridisation of the ”negative” PNA, GACGAGGT, (solid line) and its reverse sequence, TGGAGCAG, (dotted line). The ”negative” PNA had no perfect match in the entire 100 Kbp sequence of the contig and only 3 end-terminal mismatch sites (shown as green symbol). The full match sites for the reverse sequence are shown in red.

3.6.3.3 Patterns of hybridisation of different PNAs

The PNAs TGGAGCAG, CAGCAGCA and AGTGGCTG were hybridised to the control clones under highly stringent temperature conditions. Plots of the distribution of positive clones were generated using the same threshold of “positive” hybridisation, 0.9. Hybridisation to different PNAs results in different distributions of the ”positive” peaks in the contig (Figure 8). The overlap in their patterns is insignificant.

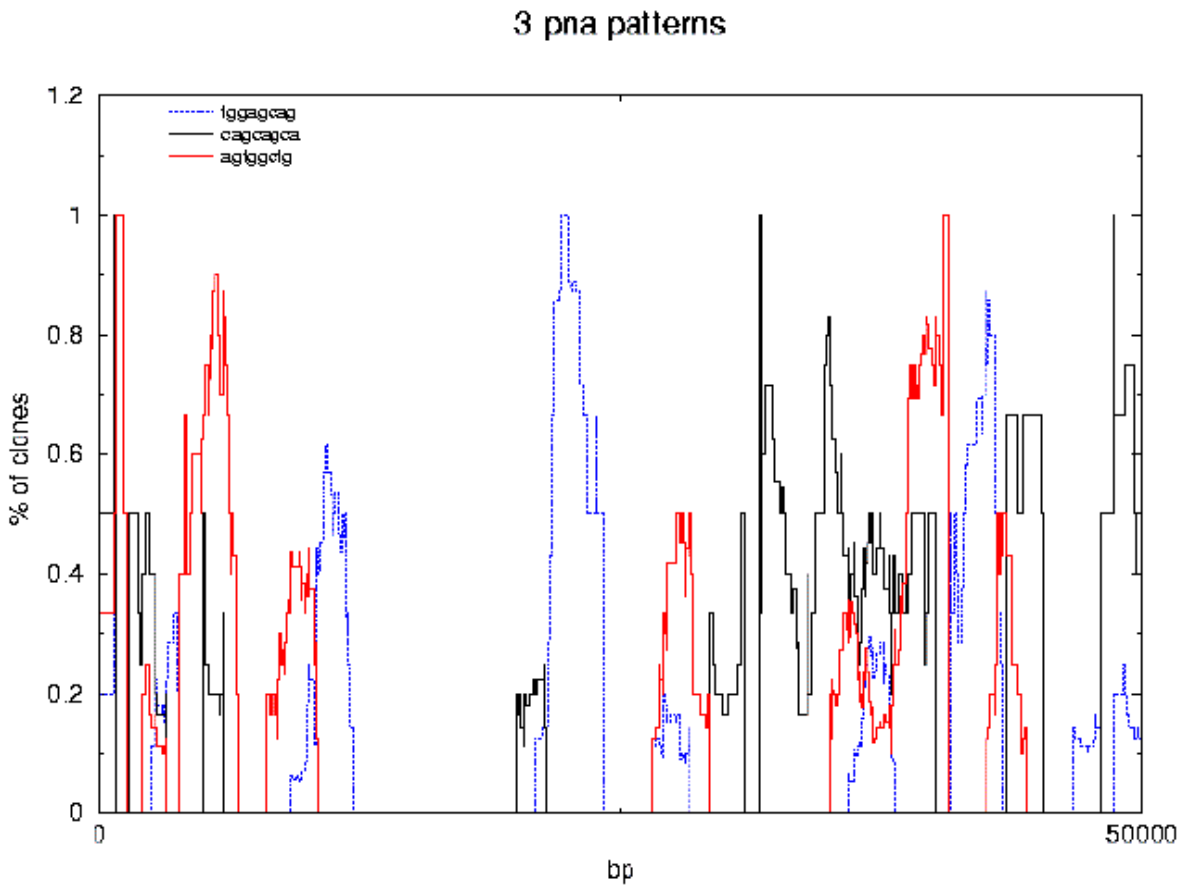


Figure 8. The patterns of hybridisation of 3 PNAs: TGGAGCAG (blue), CAGCAGCA (black) and AGTGGCTG(red). Only data for a 50Kb contig are included. Y values of 0.5 to 1 correspond to hybridisation of 50-100% of the clones which overlap that point on the X-axes. The hybridisation patterns are different.

Interestingly, the patterns of two PNAs with a single nucleotide difference (tggagcAg and tggagcTg) showed striking similarity (Figure 9). This figure shows that the distribution of peaks correlates with the positions of perfect match sites for both PNAs and that the patterns are very similar.

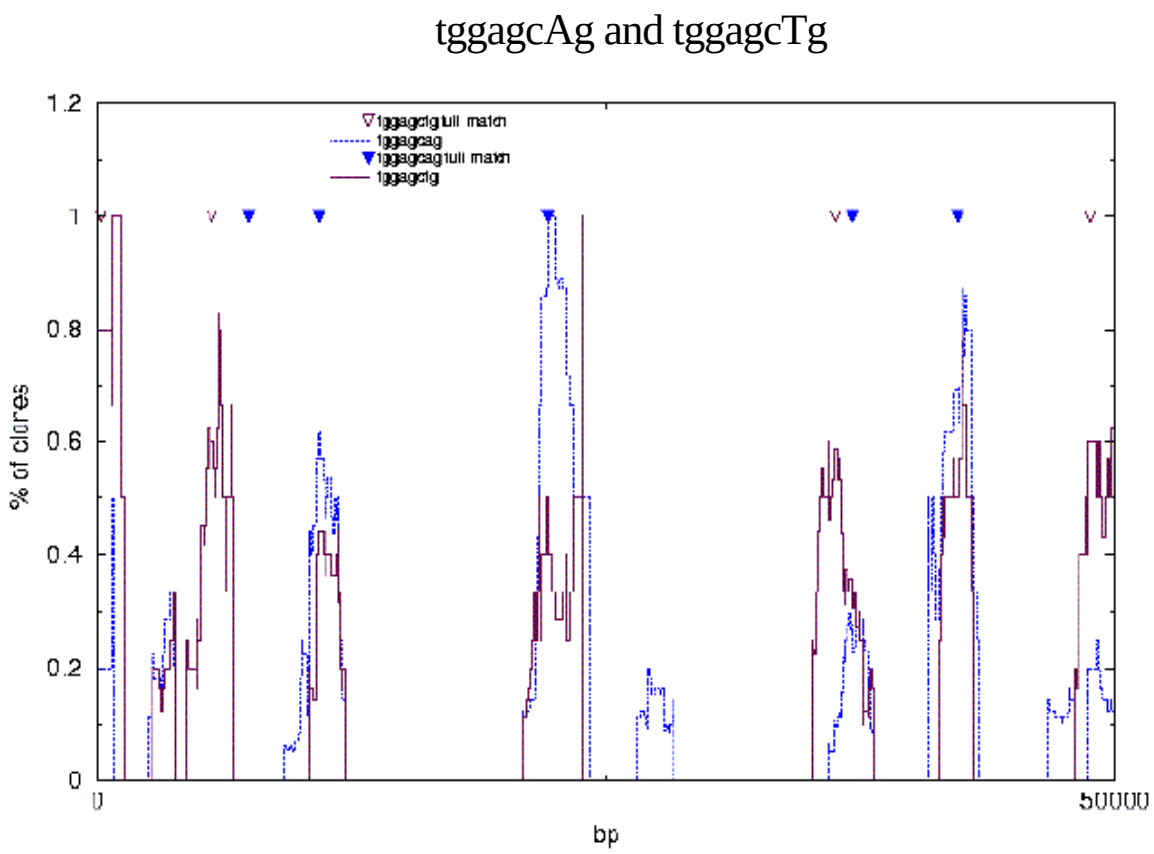


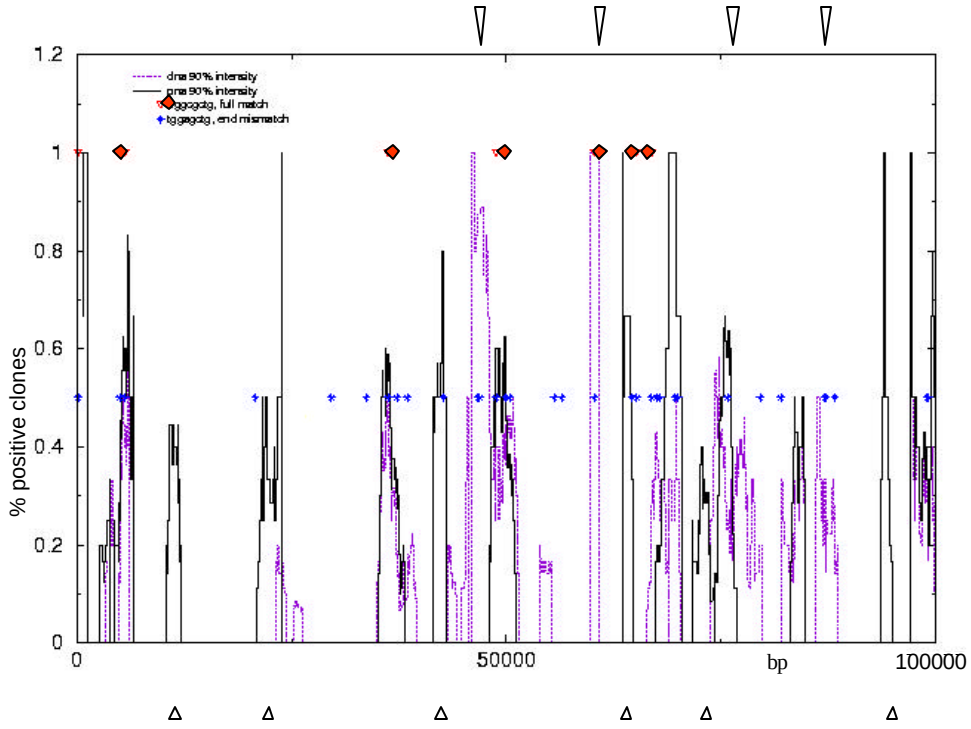
Figure 9. Comparison of the hybridisation profiles of PNAs which differ by a single base: PNA1 tggagcAg (brown, solid line) and PNA2, tggagcTg (blue, dotted line). The corresponding perfect match sites are shown above. For clarity, the figure shows the profile of only 50 Kb of the contig.

3.6.4 Comparison between hybridisations of PNA and DNA oligonucleotides of the same sequence

In the hybridisations an 8-mer PNA and a 10-mer DNA of same 8-mer core (for example, PNA tggagctg and DNA: NtggagctgN) were used. The DNA hybridisation was detected radioactively and the PNA fluorescently. The signals were evaluated using the TINA imaging software(Methods). Both hybridisation patterns were compared on the same diagram which shows the positive clones and perfect site distributions (Figure 10). The positive clones were chosen with a threshold of ranked intensities 0.9.

The sites where PNA peaks did not match the DNA peaks are marked outside the graph. 3 peaks were unique for DNA, and there were 6 sites where the clones hybridised only to the PNA. In the case of the DNA probe a unique

peak was more specific and corresponded to a perfect match site. However, hybridisation with a PNA probe was more reproducible as judged by the % of positively hybridised overlapping clones (peak height) as >50% in most cases.



peaks unique for a DNA probe

peaks unique for a PNA probe

Figure 10. The distribution of positive clones shown for PNA (solid) and DNA (dotted line) probes of the same 8-mer core sequence. The graph shows the percentage of positive clones from the total number of clones in a particular nucleotide position, X-axes). Outside the diagram, the positions of peaks unique for the DNA probe (top) or for the PNA probe (bottom) are shown. The positions of full matches (red, closed diamonds) and the terminal mismatches are shown inside the diagram.

The detailed distribution of positive clones within the overlaps are shown in Figure 11 for both probes. The positive clones were chosen based on a signal intensity threshold of 0.7. Often the clones from the same overlap positive for the PNA probe were negative for the DNA probe, and vice versa. On average, more clones were positive for the PNA probe than for the DNA probe. In some regions the clones which did not contain a perfect match hybridised with both the PNA and the DNA.

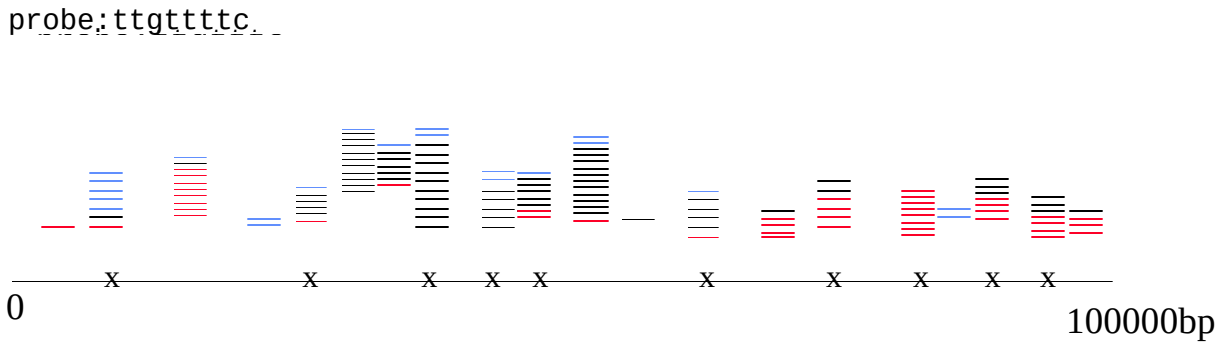


Figure 11. A scheme showing the distribution of clones which were positive in hybridisation with a 10-mer DNA and an 8-mer PNA of the same core sequence (TTGTTTTC). The DNA was labelled radioactively, the PNA - fluorescently. The positive clones were grouped by their position in the contig. The clones positive for the PNA are shown in blue, clones positive for the DNA in red and clones positive for both PNA and DNA are black. The sites containing positions of fully matched 8-mer sequence are indicated by an ”X”.

3.6.5 Results of clustering using data from hybridisations with a restricted number of PNA probes

The data obtained with fluorescently labelled PNA was used to cluster clones with the algorithm used to cluster DNA hybridisation data (Herwig et al., 1999).

Clustering means grouping the clones based on their oligonucleotide fingerprints, so it is easy to assess the accuracy of clustering because the positions (and thus, sequences) of all clones in the contig are known. An example of the clustering output is shown on Figure 12. According to the clones' coordinates, clones 1 to 4 overlap and the last clone (clone 5) belongs to a different fragment and thus has not been clustered correctly.

Figure 12. An example of a 5 members cluster where one clone was falsely grouped. In the second column there are the clone names (according to MPI nomenclature), in the third is the lower (l) nucleotide coordinate in the contig (the first nucleotide position) and in the fourth is the upper (u) nucleotide position. These coordinates indicate that the clones overlap. The clone marked with an asterisk (*) does not belong to this overlap, as seen from its coordinates, i.e. a false member in the cluster.

	Clone name	lower coordinate	upper coordinate
1	MPI_KK2_1E6	l=54205	u=55241
2	MPI_KK2_1M3	l=53910	u=55244
3	MPI_KK2_1M15	l=53786	u=54951
4	MPI_KK2_1A10	l=53768	u=55298
5	MPI_KK2_1O5*	l=25212	u=26327

For comparison the clustering was performed using data from 17 and 30, PNA or DNA oligonucleotides. Since the analysed library was small and redundant in sequence, this number was enough to assess the accuracy of clustering. The data from 17 DNA oligonucleotides of the same 8-mer core as PNA hybridised radioactively could be commpared to results obtained with the PNAs. Also the clustering was performed using random 30, 50 or 70 DNA probes.

The clustering results obtained with 17 PNA, 17 DNA, 30 PNA and 30 DNA probes are summarised in Table 3. It includes the data on entirely correct clusters or clusters with only one false member where the clusters had more than 2 (3 or more) and >3 (4 or more) clones per cluster.

Clustering with the17 PNA probe set resulted in a higher percentage of correct clusters than with the identical 17 DNA set, especially for the case where clusters with one false member were included. The same results are obtained if 30 PNA and 30 DNA data is compared. The results of clustering with 17 PNA, 17 DNA and 30 DNA showed that there no fully correct clusters of size >8 members were found (the same was true for clustering with the

70 DNA probes, Herwig, personal communication). With 30 PNA there were at least 5 such clusters. The percentage of correct clusters of 4-6 clones/cluster was much higher with PNA probe set than with the corresponding number of DNA probes.

Number of probes	Correct clusters		One false member		number of clusters		
	>2	>3	>2	>3	>2	>3	total
PNA17	**22%**	**18%**	**39%**	**35%**	61	40	125
DNA17	11%	10%	25%	25%	63	47	113
PNA30	**54%**	**47%**	**72%**	**68%**	62	44	110
DNA30	51%	44%	68%	59%	77	47	132

Table 3. Distribution of correct clusters obtained with PNA or DNA probes. Correct clusters are those which contain only physically overlapping clones. The percentage of fully correct clusters and clusters with one false member is shown for clusters containing 3 or more (>2), and for 4 or more (>3) clones/cluster.

The clustering of redundant clones should result in normalisation of the library, so that the clones can be pre-selected for sequencing. A simulation of clone pre-selection was performed by taking one clone from each cluster of 2 or more members. The clones were placed in ascending order according to their nucleotide positions. When this procedure was performed with the clusters obtained with 17 DNA probes, it resulted in 25 Kb of continuos sequence comprising 43 clones. For PNA17 the 32 Kb continues sequence could be reconstructed with 40 clones, so the pre-selection with PNA probes was more effective.

3.6.6 PNA hybridisation and detection by MALDI-TOF-MS

The feasibility of multiplexing OFP by the use of PNA octamers as hybridisation probes and MALDI-TOF-MS as the detection method which requires no labelling at all, was evaluated using a set of synthetic DNAs and the synthesised PNA probes.

3.6.6.1 Analysis of PNAs by MALDI MS.

Several matrices were examined for detection of PNAs including sinapic acid (SA), 2,5-dihydroxybenzoic acid (DHB) or α-cyano-4-hydroxycinamic acid (CHCA). The best results were obtained with CHCA. CHCA was dissolved in 50% ACN/0.1% TFA. For sample preparation, 0.5 µl analyte solution were mixed with 0.5 µl matrix solution on the stainless steel MALDI sample support. After solvent evaporation at ambient temperature, the crystalline samples were analysed in a Bruker Scout 26 Reflex II MALDI mass spectrometer. The molecular masses determined for the PNA oligomers purified by HPLC corresponded to the full-length product with little or no contamination by intermediate synthesis products.

Figure 14 shows the detection of 0.1 pmol of a single PNA (1 µl of PNA at 0.1pmol/µl was mixed with matrix solution) and 0.01 pmol of the same PNA (PNA at 0.1pmol/µl was diluted 1:10 and 1 µl was taken for the analysis). The detection limit for a single PNA was found to be 10-50 fmol.

3.6.6.2 Mass spectrometric representation of PNA mixtures

Different analytes have different detection sensitivities in MALDI-TOF-MS, and the strength of the recorded signals is interconnected in case when analyte mixtures are analysed. To explore how this affects PNA detection, the mixture of 5 PNA oligomers were analysed by MALDI-TOF-MS. The recorded spectra showed that in case of PNA mixtures at the same concentration (according to UV adsorption measurements), the signal intensities (peak heights) varied up to 40 % (Figure15, A). Analysis of PNAs differing in sequence in one position in equal proportions, showed that the peaks of bigger intensity were obtained for PNA-C and PNA-G (among PNA with the core sequence CCNCAGCC, N=A, C, T or G). The PNA octamer TGCTCCTG yielded the weakest signal. Previously it was shown that T-rich PNAs were discriminated the most in MALDI-TOF-MS (I. Gut, personal communication).

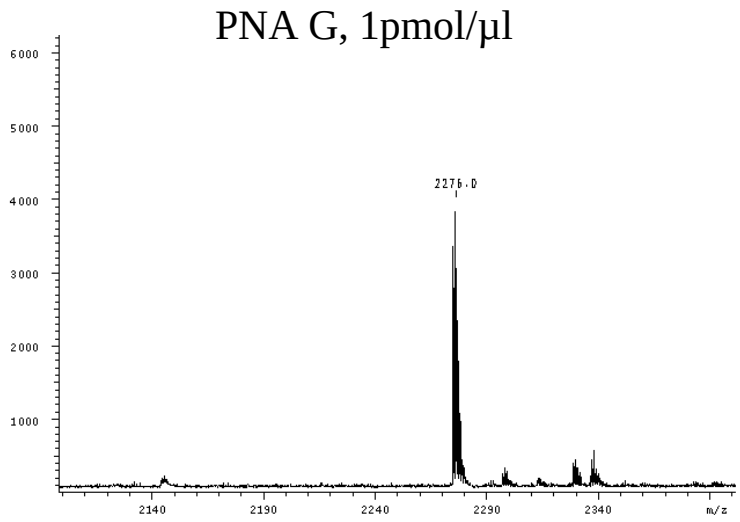
3.6.6.3 Multiplex hybridisation of PNA to the synthetic DNA on magnetic beads

Initial experiments towards multiplex hybridisation and detection of PNAs were performed on biotinylated DNA attached to magnetic beads via biotin-streptavidin complex formation. The use of such a system required optimisation of the hybridisation conditions and careful examination of specific versus non-specific interactions.

Hybridisation specificity was explored by the use of pooled 8-mer PNAs differing at one nucleotide position. The masses of the PNAs were PNA-C (pna1, 2235), PNA-T (pna2, 2250), PNA-A (pna3, 2259 Da) and PNA-G (pna4, 2275). Two DNA sequences were used as target containing the perfect site for the same PNA4 but in different orientations. To these target DNA molecules, a PNA probe positive for both strands, PNA5 (2296 Da), was hybridised together with other PNAs (Figure 15). In all experiments, the PNA probes used yielded good discrimination of the correct versus mismatched hybridisation: PNA-G (perfect) over other PNAs (Figure16, B). When the same mixture of PNAs was hybridised to the DNA target which contained the perfect site for PNA-G in 5'- to 5'-end orientation, the result was essentially the same as in case of usual, 3'- to 5'-end orientation (Figure16, C).

Figure 14. The PNA detection limit. A. 1 pmol of PNA CCGCAGCC. B. 0.1 pmol of the same PNA.

A.



B.

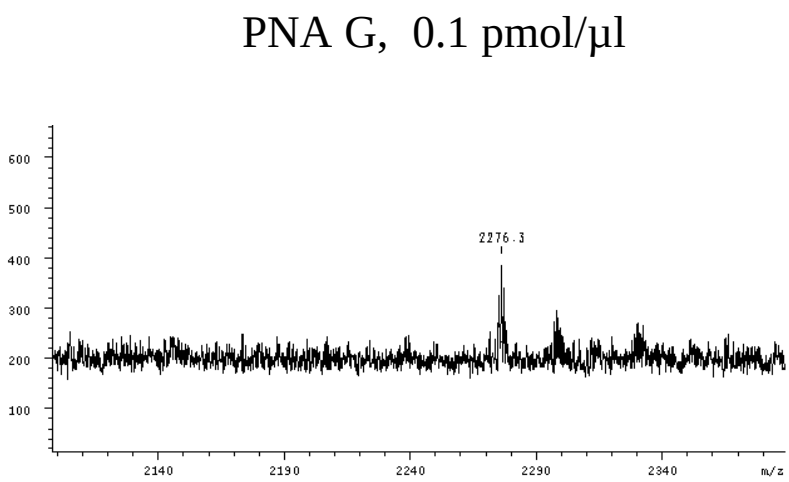
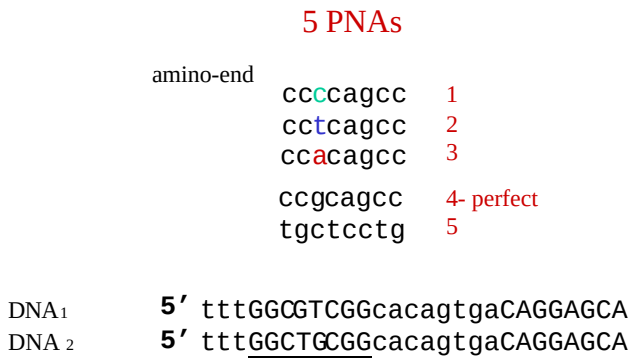


Figure 15. The scheme of used PNAs and DNA targets. Each DNA target oligonucleotide is complementary to PNA5 and PNA4 (underlined for “antiparallel” orientation, DNA_2).



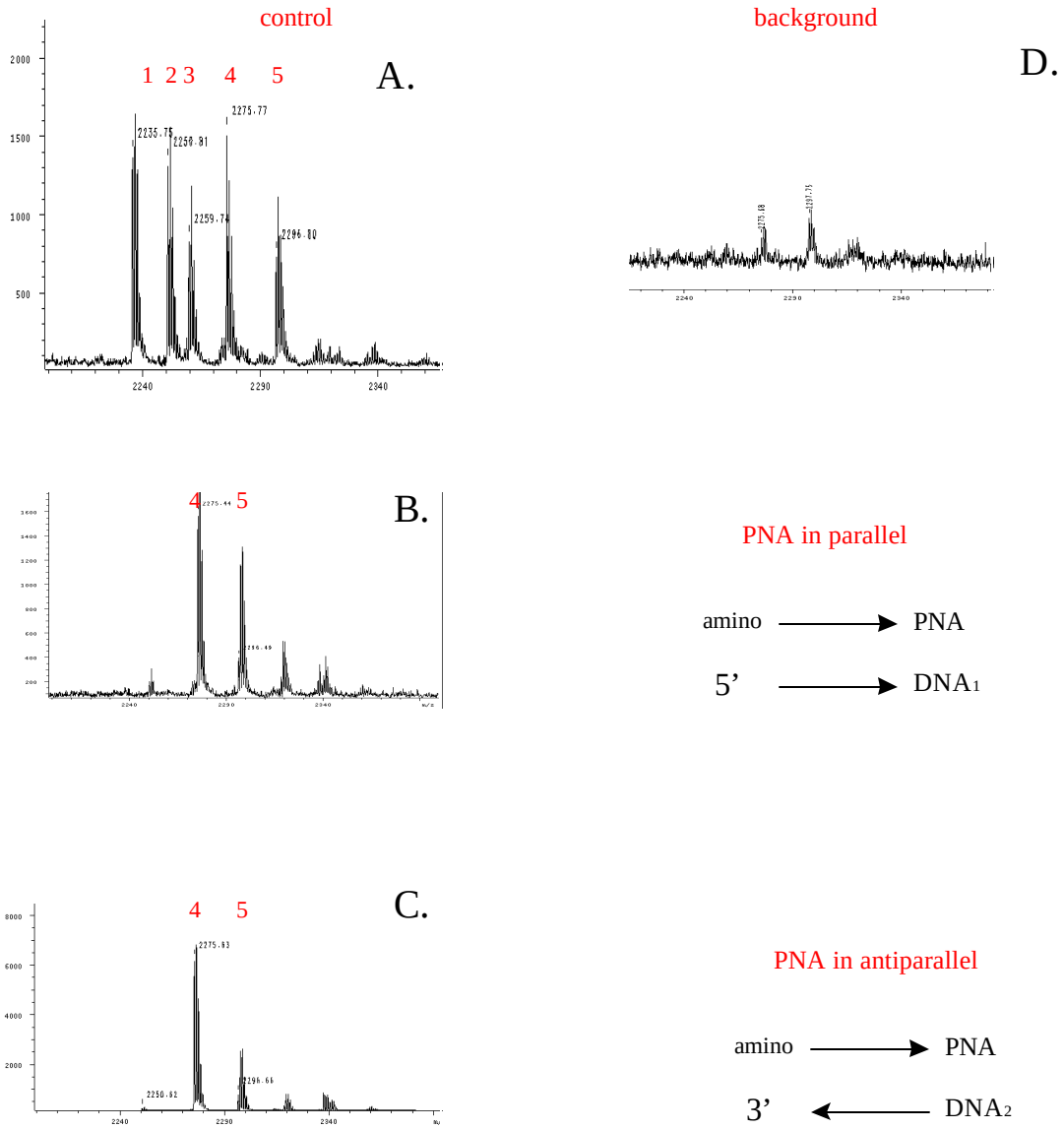


Figure 16. Hybridisation of 5 pooled PNA probes. (A) A mixture of five PNA octamers. Four PNA (1,2,3,4) differed only at the position 3 and the PNA5 was of unrelated sequence. DNA oligonucleotides were bound to magnetic beads via streptavidin-biotin link and hybridised to mixture of PNAs. (B). Hybridisation to DNA1 which contains the target site in parallel orientation to PNA4. (C) Hybridisation to DNA2 which contains the target site in anti-parallel orientation to the PNA4 (D) “mock hybridisation”. PNAs were incubated with magnetic beads without a DNA target, washed and extracted from beads.

3.6.7 Discussion

Detection of PNA hybridisation by the fluorescent scanning

During the work of this thesis a method for labelling PNAs and detect them in hybridisations to immobilised DNAs was established. Since PNAs are lacking a phosphate backbone they cannot be labelled radioactively by kinase reaction in the same way as DNA oligonucleotides. A method for conjugation of a desoxynucleoside 3'-phosphate to the amino-terminus of a PNA and subsequent phosphorylation of the 5'-end of the latter has been described previously (Kozlov et al., 1998), however, it requires several post-synthetic chemical modifications. On the other hand the application of fluorescent dyes on nylon membranes is limited by the high background fluorescence of the membrane itself and a relatively low signal. As an alternative strategy I applied a method that was developed earlier for the labelling of biotinylated oligonucleotides with an enzyme via a biotin-streptavidin link (Results and discussion 3.3.1). This way the hybridisation signal can be detected chemiluminescently.

PNA-oligonucleotides turned out to have several advantages over DNA-oligonucleotides. First of all, they allowed not only detection by chemiluminescence but also by direct fluorescence. PNAs as short as 8 or 7-nt long were successfully labelled with both Alkaline Phosphatase (total MW=100 kDa) and the more bulky fluorescent protein Phycoerythrin (MW=300 kDa) via Streptavidin conjugates and then they were used directly in hybridisations. Direct labelling with a fluorophore avoids possible artefacts of enzymatic amplification of the signal and the number of steps in the detection procedure is reduced. The signals were detected by a line scanning device which was developed in-house. The protein Phycoerythrin was chosen as a fluorescent label because of its excellent fluorescent characteristics compared to other fluorophores, especially those with low molecular weight (see Introduction 1.7). The only drawback is the large molecular weight of this protein. An attempt to use as hybridisation probes the 10-mer DNA oligonucleotides conjugated to Streptavidin-Phycoerythrin was not successful even though a similar labelling with Alkaline Phosphatase worked well. This is probably caused by a larger protein which destabilises the duplex. The probable limitation of a bulky tag can also be concluded from the fact, that PNAs containing one or two spacer linker arms (which when added to 14 spacer arm of biotin resulted in total in a 20-26 carbon-atom spacer between oligonucleotide and protein) gave better results in hybridisation than PNAs with no spacer. It is likely that better distancing of the protein results in better accessibility of the target on the membrane for hybridisation.

The labelling method proposed here uses the advantage of both, Phycoerythrin's sensitivity and the improved signal to noise ratio after direct coupling of the probe. Comparison between the secondary detection (hybridisation of PNA alone and then incubation with the Phycoerythrin-Streptavidin conjugate) and the direct method (hybridisation of a PNA coupled to Phycoerythrin) showed that the direct method was more sensitive. Importantly, the complex of PNA with Phycoerythrin was stable and fully retained its fluorescence during incubation at +45°C for several hours.

The comparison of the detection sensitivity of PNAs conjugated with protein and PNAs labelled with fluorescein showed that PNAs labelled with Phycoerythrin were 50 times more sensitive than PNAs with fluorescein. Moreover the fluorescence of Phycoerythrin was stable after several scanning reads with the laser whereas the signal with fluorescein faded after several scannings.

The quantitative analysis of the signal dependence on the amount of DNA bound to the filter did not show linearity for both, Phycoerythrin and Fluorescein. The possible explanation for this might be the high local density of fluorophores which might cause quenching of the fluorescence. However, a linear signal response would only be needed if the analysis relied on exact amount of DNA bound to the membrane. In current image analysis there is no robust procedure to determine exactly the amount of DNA bound to the filter.

The reproducibility of PNA hybridisation was examined by subsequent rounds of hybridisation and detection on the same DNA filter. The data from 3 experiments demonstrated a significantly low deviation of signal intensities from the mean. Importantly fluorescence-based detection resulted in better spatial resolution of signals than radioactive methods and did not suffer from overshining (overlap between closely situated signals) which improved the subsequent image analysis.

As a prospective, this technique can be expanded to a simultaneous hybridisation of PNA labelled with two different sensitive fluorophors emitting at different wave lengths and a two-colour detection. This would result in a 2-fold increased throughput.

Hybridisation properties of PNAs

In order to optimise conditions for the PNA hybridisation on DNA arrays, the parameters of the hybridisation reaction were examined. A critical parameter for the detection appeared to be the concentration of the PNA probe (20nM), while changing of the ionic strength and detergent was tolerable. The kinetics study showed that even though the association rate was very high (signals were detectable within the course of 10 min of hybridisation) the optimal time of incubation was three and more hours. The experiments with the PNA AATGGAGA showed that at the conditions of a moderate stringency (low temperature) the re action was characterised by initial unspecific interaction with most of the DNA targets (Figure 5, 10min of hybridisation) and uniformly low signal intensity. With the increase (from 10 min to 1.5-3 hours) of the reaction time the ratio of perfect to unspecific signal increased from 2:1 to 5-10:1. Probably PNAs dissociated from the non-complementary DNA quicker than from the DNA which contained perfect match sequence.

For oligonucleotides the hybridisation signal depends on the thermodynamic stability of the duplex formed, and thus, on the temperature of hybridisation (Wetmur, 1991). The common experience with short DNA probes suggests

that 6-10 nt long oligonucleotides should be hybridised at temperatures below +20°C. The hybridisation temperature of Phycoerythrin-conjugated PNA was assayed in a range from +8°C to +50°C and was found to be dependent on the nucleobase content of the PNA. For example, hybridisation of PNA TCCTCCTG was not detectable at temperatures higher than +8°C, whereas PNAs AGGAGTAA or CCAGGAGG could be hybridised at +40-45°C. Some of the 7-mer PNA tested were particularly stable, because they could be hybridised at +35-45°C. These results confirm that the change of a backbone charge from negative to neutral considerably improves the stability of an oligonucleotide in a duplex.

The temperature dependence of PNA/DNA duplex stability was described for the in-solution systems to correlate with the pyrimidine content of an oligomer (Giesen et al., 1998), because pyrimidine-rich PNA form triplexes, and thus melting occurs at a higher temperature. However, it seems unlikely that PNA coupled with a such large protein tag can perform strand invasion into DNA fixed on the membrane. Here the observation was made, that the higher the purine content (A and G) of the PNA, the higher was the stability and temperature of hybridisation (Table1). Apparently the optimal hybridisation temperature of Phycoerythrin-PNA cannot be deduced from the T_m measured in solution, nor a correlation was found with the calculated enthropy of a given PNA. For example, for the PNA CCGCAGCC (A+G=3, a low purine content) the signal was only detected when it was hybridised at +8°C, although the T_m in solution was found to be 57°C and calculated enthropy was higher, than the enthropy of the PNA AATGGAGA (A+G=7), which could be hybridised at +45°C.

The dependency on the purine content of the duplex stability is somewhat surprising. The purine-rich PNAs are described as prone to aggregation (manufacturer's manual) which might result in the unspecific hybridisation. However, here the hybridisation of such purine-rich PNAs was sequence dependent and the pattern of hybridisation was similar to that of the DNA oligonucleotide of the same 8-mer core (Figure10).

PNA hybridisation specificity (defined as the ratio of a perfect to unspecific hybridisation signal) can be improved via increasing the stringency of hybridisation by increase of the hybridisation temperature. The increase of specificity with higher temperature of hybridisation was observed for several PNAs, however, 100% specificity (which can be measured by full correlation of complementary sites with the positive clones) was not achieved. This implies that there are other than complementarity factors which account for a stable interaction of PNA with DNA on the membrane.

In contrast to the expected high discrimination ability of PNAs, an observation can be made that there was no strong correlation between perfectly matched sites in the DNA target and a strong "positive" hybridisation signal. Nevertheless the experiments confirm that PNA /DNA interaction is sequence specific. For example, the two PNAs differing by one nucleotide base or the 2 complementary PNA had very similar patterns of hybridisation within the pair. PNAs of different sequence had hardly overlapping patterns of hybridisation. Importantly DNA clones of the same sequences hybridised reproducibly (with the same efficiency) to most of PNA probes tested.

Comparison between PNA and DNA probes

The analysis of the hybridisation patterns derived from PNA and DNA oligonucleotides of the same sequence showed striking similarity. It is therefore likely that the signal specificity of the interaction between complex DNA and the PNA oligocnucleotide is *nucleobase* dependent and the backbone charge cannot significantly influence it.

The diagrams showing the hybridisation pattern also display the percentage of overlapping clones of similar sequence, as defined by a signal intensity threshold. The number of peaks (frequency of hybridisation) and their height (the reproducibility) can be a criteria for comparison of PNA and DNA. The results show that the reproducibility was better for PNA-hybridisations (Figure11). Also PNAs were better as probes because the number of peaks (thus frequency of hybridisation) was as a rule higher with PNA oligonucleotides (e.g. 20 with a PNA versus 16 for a DNA (Figure 10)). The most likely explanation is the higher stability of a PNA probe in a duplex compared to a DNA oligonucleotide probe.

The speculation, that PNAs hybridise in both, the anti-parallel and parallel orientation hardly explains the observed difference in hybridisation patterns. Nevertheless, the possibility of interaction in the other direction cannot be excluded, as follows from the hybridisation of a ”negative” PNA (Figure 8). This PNA had *no match* in 5’-to 3’-end (antiparallel orientation) in the entire contig and the comparison with hybridisation patterns of its reverse sequence showes considerable similarity (Figure 8). The parallel 5’-end PNA to 5’-end DNA hybrids are described as less stable. However, no strong correlation of positions of reproducibly hybridised clones with sites containing complementary to the probe sequence in anti- or the parallel orientation was found.

The difference in hybridisation patterns of overlapping clones when probed with PNA and DNA oligonucleotides of the same sequence (Figure 11) can be explained by different accessibility of the DNA target which is the double stranded molecule (PCR product) in different ionic strength (low ionic strength for PNA and high for DNA). Also the PNA oligo is coupled to a bulky Streptavidin-Phycoerythrin protein complex, which may account for the difference in interactions with membrane bound DNA. The differences between patterns of DNA and PNA hybridisation mostly occurred in the regions containing no perfect match.

Importantly the clustering of clones, i.e. grouping based on fingerprint similarity, obtained by hybridisations of PNAs proved to be more accurate than with the equivalent set of DNA probes.

MALDI detection of PNA multiplex hybridisation

The next generation of OFP methods should ideally not require any labelling of the probes, and the clone characterisation should be performed within minutes. Label-independent detection of hybridised probes is possible by the use of MALDI-TOF-MS. Instead of a radioactive isotope or fluorescent dye, the hybridised probe molecules are detected directly and identified by their molecular mass. This approach paves the way for efficient multiplexing (I. Gut, 1998), (Griffin et al., 1997), (Ross et al., 1997). In MALDI-TOF-MS, complex mixtures of analyte molecules are resolved and detected simultaneously on a microsecond time scale. Compared to DNA and RNA oligomers, the higher ion stability of PNA oligomers provides the necessary high detection sensitivity.

Analysis of PNA mixtures

Compared to DNA and RNA, molecular ions of PNA are considerably more stable under the conditions relevant to Matrix-Assisted Laser Desorption/Ionisation Time-Of-Flight Mass Spectrometry (MALDI-TOF-MS). Their stability can be compared to the stability of unmodified peptide molecular ions. As a consequence PNA ions have long lifetime, which improves their detection sensitivity. The second consequence is that peptide matrices such as 2,5-dihydroxybnzoic acid (DHB) or α-cyano-4-hydroxycinamic acid (CHCA) can be used. Compared to 3-hydroxypicolinic acid, the most efficient matrix for the analysis of DNA and RNA samples, these matrices provide significantly higher ion yields. The price for the gain are higher internal energies of the generated molecular ions. These energies cause spontaneous decomposition of nucleic acid molecular ions, whereas peptide and PNA molecular ions are still detected intact. In summary, the detection sensitivity for PNA oligomers is expected to be one to two orders of magnitudes higher than for oligonucleotides.

PNAs lack the highly acidic phosphodiester groups of nucleic acids, and therefore are preferably detected as positively charged molecular ions. In MALDI of PNA oligomers, mainly singly protonated molecular ions $(M+H)^+$ are formed rendering the analysis of mixtures straightforward. Besides these ions, double protonated molecular ions $(M+H)^{2+}$, and metal-cation adduct ions $(M+Me)^+$ can be detected as satellite peaks. $(M+H)^{2+}$ ions rarely exceed 20% of the abundance of the corresponding singly protonated species. Adduction of Na^+ and K^+ instead of H^+ is frequently observed, recognised in the mass spectra by a constant mass shift of 22 or 38 Da, respectively. The abundance of this alternative ionisation mode depends on the amounts of metal cations in the analyte solution. Since this ionisation mode degrades the detection sensitivity and hampers mixture analysis, the metal cations should be removed as much as possible from the sample prior to the mass spectrometric analysis. Here the formation of $(M+Me)^+$ ions was suppressed by addition of ammonium acetate or ammonium citrate to the sample solution. In this case, an excess of ammonium ions competes for adduction yielding $(M+NH_4)^+$ molecular ions. During the desorption/ionisation event, however, these ions dissociate to NH_3 and $(M+H)^+$. Thus, adduction of NH_4^+ results in proton transfer, and no signal is lost in $(M+NH_4)^+$ satellite peaks.

Multiplex hybridisation and detection of PNA probes

The performed experiments and obtained results demonstrate the feasibility of multiplexing OFP by the use of PNA octamers as hybridisation probes, and MALDI-TOF-MS for the detection of hybridised PNA probes (Figure 16). With respect to sample throughput and total analysis costs, it would be advantageous to perform the hybridisation reactions, washing cycles, and detection of the hybridised probes on the same solid support in high density. This approach requires that the PCR amplified cDNA inserts are immobilised directly onto MALDI sample supports. These supports are made of stainless steel and expose a surface not well suited for DNA immobilisation. It has been shown, however, that a thin coating of epoxysilane-modified glass provides the necessary interface. This coating is compatible with MALDI-TOF-MS, and the PCR amplified target DNA can be immobilised onto the modified surface by the use of a primer that carries an amino-linker at the 5'-terminus (Kuchrzhak, 1999).

In this work, biotin-streptavidin complex formation was used to immobilise the target DNA to magnetic beads. The paramagnetic characteristic of these particles simplifies the necessary automation and parallel processing using commercially available pipetting robots. The obtained results demonstrate that PNA octamers hybridise correctly to oligonucleotides bound to magnetic beads in both, parallel and anti-parallel orientation, and that under these conditions a single mismatch is reliably discriminated. It is important that conditions that allow only specific hybridisation can be found, since MALDI TOF does not provide quantitative detection.

For multiplexing, an obvious challenge is the selection of PNA octamers that show similar hybridisation yields, and differ in molecular mass by at least 6 Da to be well separated by MALDI-TOF-MS. This condition requires special design of PNA probe mixtures, since PNAs can be quite redundant in their sequence, and thus their mass. For example, in the set of 82 oligonucleotide sequences which are commonly used in OFP there are only 43 different masses, and among those only 26 corresponded to unique sequence and the rest were the groups of oligonucleotides of different sequence of same base composition, thus the same mass (Table 1). So the multiplex probe set should be designed in a way that probes of the same mass are used in different hybridisation pools or a chemical modification can be used to generate variability of masses. The easiest way is the post-synthetic modification with the linker (MW of 146Da). Alternatively, different aminoacids can be introduced in the first cycle of the synthesis.

Another important point of multiplexing that has to be considered, is the maximal number of probes that can be hybridised and detected simultaneously. This question relates to the sensitivity of MALDI detection and hybridisation efficiency of PNA. Concerning probable overlaps in the target DNA sites where different PNA hybridise to, there may be competition for binding to these sites between probes, however in an experiment with four PNA-oligos which could potentially hybridise to the same site (Figure 15), the correct PNA could be detected unambiguously. Also in the case where target sites for different PNA oligos were next to each other (or overlapping)

the hybridisation of any PNA was not affected. The maximal number of PNAs analysed in a mixture was 15 so far, however, even if this will finally be the limit, it means a 15-fold multiplication of the assay. The future direction in this work will be the use of PNA mixtures of 20-30 oligomers to hybridise on DNA clones amplified by PCR.

Table 1. PNA sequences which correspond to the conventionally used for OFP DNA probe set sorted by their mass.

#	sequence	mass, Da	oligo#	sequence	mass, Da
1.	ccctgccc	2081	42.	tgctgtgt	2206
2.	tttcttcc	2101	43.	aggacctg	2209
3.	ctcaccat	2104	44.	ctggagca	2209
4.	cctcagcc	2105	45.	aagaacaa	2210
5.	cagcctcc	2105	46.	ctgggctg	2216
6.	ccctccag	2105	47.	agggacca	2218
7.	cctttgct	2126	48.	aagaaaat	2225
8.	agctcacc	2129	49.	tgctggtg	2231
9.	cctgctgc	2136	50.	ctggagaa	2233
10.	ccctgctg	2136	51.	atgagcag	2233
11.	ccagaacc	2138	52.	ccaggagg	2234
12.	cagacacc	2138	53.	ggccaagg	2234
13.	tttctctg	2141	54.	tgctggag	2240
14.	cctggcca	2145	55.	tggagctg	2240
15.	cagccctg	2145	56.	agtggctg	2240
16.	ctgcagcc	2145	57.	tggagcag	2249
17.	ctgctgct	2151	58.	agctggag	2249
18.	cttcctgg	2151	59.	ctggagga	2249
19.	tgctcctg	2151	60.	tggggctg	2256
20.	ttgttttc	2156	61.	gctggtgg	2256
21.	cctggtca	2160	62.	tggagaaa	2257
22.	catcctgg	2160	63.	aatgagga	2257
23.	ctcctgga	2160	64.	tgaagaag	2257
24.	cagcctga	2169	65.	agcaggag	2258
25.	tcactgtg	2175	66.	ggagctgg	2265
26.	cctgggct	2176	67.	gtggtggt	2271
27.	gctcctgg	2176	68.	tggagaag	2273
28.	tggccctg	2176	69.	tgggatgg	2280
29.	ctgctggc	2176	70.	gaggagaa	2282
30.	cctggctg	2176	71.	aggaagag	2282
31.	agcccaag	2178	72.	ggagaaga	2282
32.	tgttattt	2180	73.	gaggaaga	2282
33.	cagcctgg	2185	74.	gaagagga	2282
34.	tgctgctg	2191	75.	aggagaag	2282
35.	ctgtgctg	2191	76.	agggagtg	2289
36.	ttctggaa	2199	77.	gaaggagg	2298
37.	agctgctg	2200	78.	ggaggaag	2298
38.	tgcagctg	2200	79.	gaggagga	2298
39.	ctgagctg	2200	80.	aggaggag	2298
40.	ctactggg	2200	81.	ggaggtgg	2305
41.	ctggggcc	2201	82.	tgggggag	2305