DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
242

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
242

LAURIS KAPLINSKI

The application of oligonucleotide hybridization model for PCR and microarray optimization

UNIVERSITY OF TARTU
PRESS
1632

Institute of Molecular and Cell Biology, Faculty of Science and Technology, University of Tartu, Estonia

Dissertation was accepted for the commencement of the degree of *Doctor Philosophiae* in bioinformatics on July 10th, 2013 by the Council of the Institute of Molecular and Cell Biology, University of Tartu.

Supervisor: Prof. Maido Remm, PhD
Institute of Molecular and Cell Biology,
University of Tartu, Estonia

Opponent: Alexander E. Pozhitkov, PhD
University of Washington, Seattle, USA

Commencement: Room 105, 23B Riia street, Tartu, on 11th of September 2013, at 10.00.

Publication of this thesis is granted by the Institute of Molecular and Cell Biology, University of Tartu and by the Graduate School in Biomedicine and Biotechnology created under the auspices of European Social Fund.

The research in this thesis was funded by the SLIC-513771 EU grant, by Estonian Ministry of Education and Research (grants SF0180027s10, SF0182649s04 and SF0180142), by Enterprise Estonia (Grant EU19730) and by EU (European Regional Development Fund through the Estonian Centre of Excellence in Genomics).

European Union
European Social Fund

Investing in your future

ISSN 1024–6479
ISBN 978–9949–32–337–1 (print)
ISBN 978–9949–32–338–8 (pdf)

University of Tartu Press
www.tyk.ee

TABLE OF CONTENTS

LIST OF ORIGINAL PUBLICATIONS

I. Kaplinski L, Andreson R, Puurand T, Remm M (2005). MultiPLX: automatic grouping and evaluation of PCR primers. Bioinformatics 21(8): 1701–2.

II. Kaplinski L, Scheler O, Parkel S, Palta P, Toome K, Kurg A, Remm M. (2010) Detection of tmRNA molecules on microarrays at low temperatures using helper oligonucleotides. BMC Biotechnol. 10:34

III. Scheler O, Kaplinski L, Glynn B, Palta P, Parkel S, Toome K, Maher M, Barry T, Remm M, Kurg A (2011) Detection of NASBA amplified bacterial tmRNA molecules on SLICSel designed microarray probes. BMC Biotechnology 11:17

IV. Kaplinski L, Möls M, Möls T, Remm M. Predicting failure rate of Multiplex PCR (submitted)

Author's contributions:

Ref. I: created the grouping algorithm, wrote the software implementation and was responsible for drafting the manuscript.

Ref. II: implemented the probe selection algorithm, wrote the software, designed the detection probes, performed data analysis and was responsible for drafting the manuscript.

Ref. III: conducted this study, designed the chaperon probes, performed data analysis and was responsible for drafting the manuscript.

Ref. IV: conducted this study, performed data analysis and was responsible for drafting the manuscript

LIST OF ABBREVIATIONS

NN	nearest neighbor
NP	nondeterministic polynomial time
T_m	melting temperature
tmRNA	transfer-messenger RNA
PCR	polymerase chain reaction
ROC	receiver operating characteristic
ΔG	free energy change

INTRODUCTION

Nucleic acid based technologies are becoming increasingly important in each year at fields as diverse as medical diagnosis, nanoscale engineering and information storage. Although the basis of nucleic acid structure was determined over half a century ago, the study of more complex structures and interactions is ongoing. The practically infinite variability of the primary structure of nucleic acids makes the case-by-case study of their properties practical only for few important biological molecules, such as tRNAs and rRNAs. For the overall majority of cases, scientists rely on statistical models that predict with varying success the important properties of nucleic acid structure and interactions. The greater the importance of nucleic acid-based technologies, the more precise the models need to be.

The first part of this thesis gives a brief overview of the thermodynamic and kinetic properties of nucleic acids hybridization. The second part concentrates on the nearest-neighbor thermodynamic model that is currently the state-of-the-art method for calculating hybridization parameters. The third part gives an overview of the application of thermodynamic model on the design of PCR primers, with special attention paid to multiplex PCR and microarray hybridization probes.

The research part of this dissertation covers the following topics: 1) development of efficient method for grouping primer pairs in multiplex PCR, 2) determination of the most important primer-specific factors that cause the failure of multiplex PCR, 3) development of an automated procedure for designing an exhaustive set of probes for the detection of bacterial RNA, and 4) development of helper oligonucleotides for breaking the secondary structure of tmRNA.

REVIEW OF LITERATURE

The digital code

Nucleic acids form the most basic "digital code" of life. The unique property of a forming stable perfectly aligned double-helical structure, discovered by James Watson and Francis Crick, is the basis of both the storage and transmission of genetic information in living organisms.

In addition to information storage and transmission, nucleic acids, or more specifically RNA, have many metabolic and control functions in cells. Although the complexity of nucleic acid sequences is many orders of magnitude lower than the complexity of amino acid sequences of comparative length (Levitt 2009), they can nevertheless fold into complex 3-dimensional structures that have specific interactions within cell metabolic system. Like for proteins, the exact 3-D structure of a nucleic acid fragment is determined by both its primary structure and environment, where it was synthesized and operates.

Chemical basis of the digital code

In many biological applications, it is sufficient to think of nucleic acid sequences as a purely digital code built on a 4-letter alphabet. Nevertheless, it is important to understand, that its "digital" nature is determined by the chemical properties of the nucleic acids. These properties determine the specificity of chemical reactions – replication, transcription and translation – that result in copying or interpreting the genetic information.

In the context of information processing, the most important of those chemical properties are:

a) The polar structure of the nucleic acids. The chains are composed of negatively charged hydrophilic pentose phosphate "backbone", and neutral or positively charged hydrophobic bases. Although the negative charges of the phosphate groups repel each other thereby destabilizing the double helix, the containment of hydrophobic bases and hydrogen bonds in the internal region stabilizes the structure. The double-helix is also stabilized by mono- and divalent cations that partially neutralize the repulsive negative charges of the phosphates (Levitt 2009; Cisse, Kim, and Ha 2012)
b) The hydrogen bonds between aligned bases from 2 antiparallel strands. The preferred pairings are such that the aligned bases "fit" with each other, forming 2 (A/T) or 3 (G/C) hydrogen bonds. These bonds are stronger than in solution due to the hydrophobic nature of the inner part of dihelical structure (Levitt 2009).
c) π-stacking, van der Waal's forces and hydrophobic interactions between adjacent aromatic rings. Nucleic acid chains, even in the absence of double-helical structure, often take such a form that the planes of aromatic bases are stacked parallel to each other and partially overlap. In this configuration, the

interactions between p-electrons between adjacent bases are strongest and stabilize the structure of the whole nucleic acid sequence (Borer et al. 1974; Levitt 2009).

Hybridization of nucleic acids

The 3 above-mentioned chemical features contribute to the spontaneous formation of double-helical structure between 2 single-stranded nucleic acid chains or 2 regions of the same chain. This process, known as hybridization, is the basis of the formation of most of the ordered structures of nucleic acids, such as the near-perfect double helix of genomic DNA or clover-leaf structure of tRNA.

The strength and speed of spontaneous hybridization of specific nucleic acid sequence depends mostly on 3 factors, namely enthalpy, entropy and steric effects.

The enthalpy (the "pure" energetic effect) of hybridization (ΔH). This is the potential energy change during the formation of double helical structure, mostly caused by the formation of various bonds – hydrophobic, aromatic and hydrogen bonds and electromagnetic interactions between charged parts of molecule. In practically all environments, the double-helical structure is energetically advantageous for semi-complementary strands of nucleic acids and the double helix with its perfectly aligned Watson Crick pairs has the highest change of enthalpy (Levitt 2009).

The entropy effect (i.e. change in the degree of freedom of configurations) of hybridization (ΔS). The double helical structure always has a much lower entropy than random distribution of nucleic acid fragments or chains in solution. The more ordered the alignment in double helix, the lower is the entropy and the smaller is the probability of nucleic acid chain randomly taking this configuration (Levitt 2009). The statistical effect of entropy to the actual hybridization is strongly temperature-dependent.

Steric effects. Nucleic acid chains being macromolecules, the formation of certain configurations may be improbable or impossible simply due to their size. One very common effect is steric blocking of the hybridization with the inner part of single-stranded molecule due to its 3-D structure.

For the actual calculation of the properties of chemical reactions, including nucleic acid hybridization at constant temperature, it is usually sufficient to use the composite free energy value (ΔG, Gibbs' free energy).

$$\Delta G = \Delta H - T\Delta S$$

Gibbs' free energy corresponds to the actual amount of energy that can be extracted or is needed to perform certain reaction or transition. Its value determines the equilibrium constant of a reaction – if ΔG of a reaction is negative, the equilibrium is shifted towards products; if it is positive, towards reagents. As ΔG contains both enthalpy and entropy components, it is also temperature-dependent.

The hybridization of nucleic acids is a complex process that depends on the type of molecules (DNA or RNA) and their primary sequence. Longer nucleotide chains hybridize in zipper-like fashion, starting from short "seeds". The hybridization of shorter oligonucleotides is usually modeled on a one-step process, although even in this case it seems to start from a contiguous perfectly aligned seed of 7 base pairs (Cisse, Kim, and Ha 2012). The dissociation seems to start from the ends of hybridized oligonucleotides (Wienken and Baaske 2011).

For the complementary molecules with the identical primary sequence, the hybridization strength normally increases in the following order: DNA/DNA < DNA/RNA < RNA/RNA; but there are exceptions, as certain Watson-Crick or mismatched pairs and motifs have very different pairing strengths in different hybridization types (Sugimoto, Nakano, and Katoh 1995; T. Xia et al. 1998).

Steric effects are the cause of long nucleotide sequences not forming perfectly paired hybridized duplexes. Although the perfect duplex between such strands would be globally most energetically advantageous, its spontaneous formation would require extremely improbable random formation of exactly paired duplex between 2 long sequences and unwinding of the helical structure of single-stranded molecules. Instead, certain regions of single stranded molecules form locally hybridized double-helical structures that represent a local energy minimum and block the formation of global alignment (Levitt 2009). As the composition of locally hybridized regions is dependent on the folding history of nucleotide chain, the secondary structure of long DNA and RNA chains is unpredictable (Chavali et al. 2005).

Steric effects become the stronger the longer nucleic acid chains are. In living cells topoisomerases are needed to clear out steric blockage of nucleotide chains during replication.

Hybridization is stochastic process; there are no objective factors that force a nucleic acid fragment to search out and hybridize only with its "correct" partner. Any 2 fragments can associate and form hybridized pair, and any 2 hybridized fragments can dissociate. What differs between the "correct" and "incorrect" hybridizations are their energetic effects. Usually the ΔG of hybridization is lowest (i.e. most energetically favorable) for the hybridization between perfect Watson-Crick complementary sequences and highest (i.e. energetically unfavorable) between certain sterically incompatible sequences. As the spontaneous processes are statistically favored towards the state with lower ΔG, perfect or near-perfect hybridized states are normally favorable and eventually prevail when the hybridization process runs long enough.

Hybridization is also a reversible process; at any temperature above absolute zero, both the hybridization and dissociation processes can take place between arbitrary fragments of nucleic acids. The speed of both processes, as expressed by the reaction rates of annealing (K_{ann}) and melting (K_{melt}), are determined by the change in free energy (ΔG) during hybridization and the concentrations of free and hybridized nucleic acid fragments. The balance of hybridized and free

states as expressed by the equilibrium constant K_d is determined by the ratio of corresponding reaction speeds – and thus depends on the ΔG of hybridization.

Hybridization and melting go through different intermediary stages, as indicated by the fact that the effects of internal mismatches in oligonucleotide sequences are different for annealing and melting. While the dissociation speed is highest when mismatches occur in the middle of sequence, the association speed has a step-like dependence on the number of consecutive perfectly aligned nucleotides. This is in accordance with the melting models, where dissociation proceeds over the intermediate form where only the middle of the sequence is in the hybridized state (Wienken and Baaske 2011). Hybridization, on the other hand starts with a seed of certain minimal length in any region of nucleotide chain. Seven contiguous base-pairs increases the speed of the annealing of DNA or RNA by several orders of magnitude compared to alignments with shorter perfectly aligned spans (Cisse, Kim, and Ha 2012).

The positions of mismatches relative to each other in oligonucleotide sequence also affect the hybridization equilibrium. Mismatches closer than 5 bp have a smaller effect on the stability of hybridization than mismatches with wider spacing. Also mismatches close to the ends of oligonucleotide chain have less effect (Hadiwikarta et al. 2012).

The term 'melting temperature' (T_m) is often used to characterize certain oligonucleotides, such as PCR primers and detection probes. It corresponds to the temperature at which the rates of both hybridization and dissociation are equal, and thus exactly half of the oligonucleotides are in the hybridized and half in the free state. The ΔG of hybridization reaction is roughly 0 at this temperature, because the equilibrium depends on the concentrations of the hybridizing oligonucleotides in addition to the free energy of helix formation.

In many practical applications of oligonucleotide hybridization like PCR, the exact value of equilibrium constant K_d can be ignored as long as it is sufficiently high or low. For example, if the temperature is sufficiently below T_m, the ΔG of hybridization is significantly negative and the equilibrium is strongly shifted towards the hybridized state; and vice versa, i.e. if the temperature is sufficiently above T_m, most nucleic acids are in free state.

While such approach is often sufficient for experimental work, there are other cases where it may not be possible to guarantee a sufficient temperature difference from T_m. In such cases more precise calculation of ΔG value is beneficial, as it allow the experimenter to take into account the actual equilibrium ratio and find optimal reaction parameters. One such example is the discrimination, either by PCR or hybridization array, between closely related organisms (e.g., bacterial strains). These often differ only by few nucleotides. Thus primers or probes designed to be complementary to the DNA sequence of one organism hybridize only slightly more strongly to their intended target than to control group (Figure 1). To achieve the maximal possible discriminatory power, annealing should take place at the precise temperature at which the difference between actual detectable hybridization to target and nontarget is maximum. Another example is the discrimination between many target mole-

cules in a single fluorescence channel by the differences in the melting temperatures of their corresponding hybridization probes (Fu, Miles, and Alphey 2012), i.e. where the precise prediction of melting curves of hybridization probes is essential.

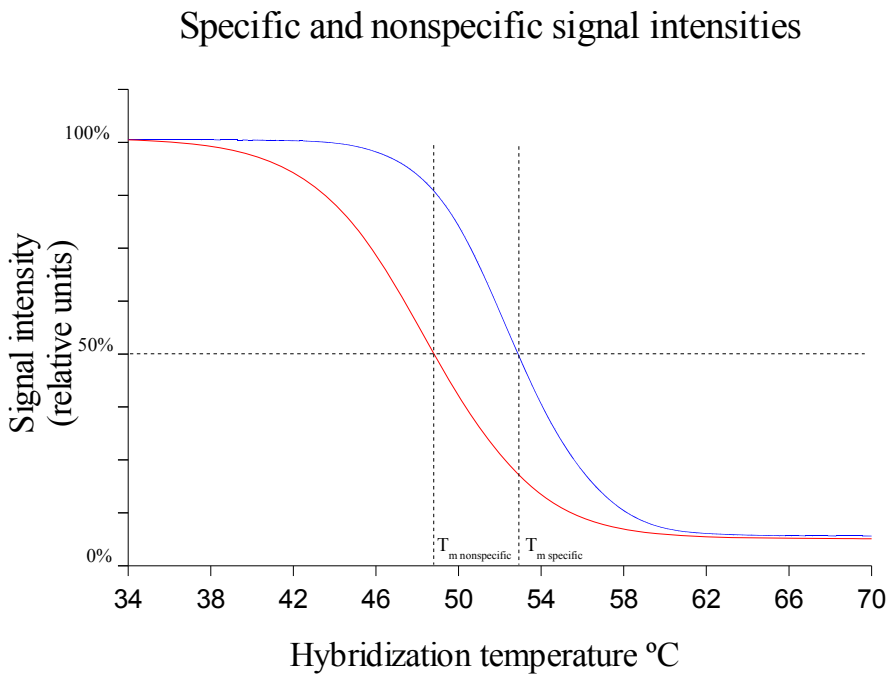


Figure 1: Theoretical melting curve of a perfect hybridization and nonspecific hybridization with one mismatch of the same probe. Blue – melting curve of perfect hybridization. Red – melting curve of nonspecific hybridization. X axis – annealing temperature. Y axis – relative signal intensity (proportional to the amount of hybridized probes). The melting temperatures of both specific and non-specific hybridization are marked with vertical lines.

Certain non-perfect alignments, e.g. terminal G/T (and especially G/U in case of RNA) mismatches, are only slightly disadvantageous (ΔG in the range of 0.2 kcal/mol) compared to perfect matches in aqueous solution. A similar effect is also seen for the antiparallel combination of G/A and A/G mismatches, which may be very difficult to discriminate by only hybridization because the equilibrium constants are very similar for specific and nonspecific hybridization. But if subsequent PCR reaction is used and thus DNA strands are bound to the active center of DNA polymerase, the ΔG difference between perfect and mismatched pairing of a terminal nucleotide is amplified by an order of magnitude (Petruska et al. 1988).

The concentration of mono- and divalent cations has a strong effect on the hybridization of nucleic acids. Higher salt concentrations make all hybridized states, both specific and non-specific, more stable, possibly by shielding negative charges of phosphate backbones.

High-level structure of nucleic acids

The best-known structure of nucleic acids is the linear double-helix, of which there are 2 alternative configurations – the A and B forms that differ by the angle of the bases and the distance of phosphate backbones (Levitt 2009). Double helices are prevalent in genomic DNA, being formed during the synthesis of nucleotide strands and staying in the hybridized state for most of the time. On the other hand, long single-stranded nucleic acid chains normally do not form double helix even when their complementary chain is present because such a configuration has extremely low entropy (and thus also an extremely low probability of spontaneous hybridization).

Nevertheless, nucleic acid chains can spontaneously form complex structures due to partial self-complementarity of the different regions of the same molecule. The self-complementary parts of the molecule tend to form double-helical stems that are connected by single-stranded loops. Both local and distant parts of the molecule can hybridize, creating complex 3-D configurations that tend to stabilize the conformation of the molecules (Levitt 2009). The 3-D configuration itself is highly dependent on both the primary structure of the nucleic acid chain and its folding history. There is often no single prevalent configuration but many variants. The actual configurations of nucleic acids are also sensitive to changes in primary sequence as a single mutation can make it to fold in a completely different configuration. This phenomenon has been used for the detection of biological variants based on their structural properties instead of sequence differences (Dong et al. 2001).

It should be noted that all the properties of nucleic acid fragments are ultimately determined by their primary structure and local environment. For short oligonucleotides, the primary structure is all that needs to be known to predict precisely their hybridization behavior and secondary structure. For longer nucleotide chains, the secondary structure becomes dependent on the folding history of the molecule, and thus it is not strictly predictable from the primary structure. Nevertheless, the stability of various global and local conformations is determined by primary sequence (Breslauer et al. 1986; Dong et al. 2001).

The detection of the secondary structure of nucleic acids is complex process. For short sequences, the most energetically favorable configurations can be determined computationally. The most popular program for secondary structure prediction is UNAFold (Markham and Zuker 2008). Computational methods usually only predict the overall energetic potential of different structures and cannot estimate the relative abundance of actual configurations because the

latter depend on folding history. For native molecules, the most precise method is often by using sequence alignment with a phylogenetically close molecule of known structure. For new molecules without known reference structures various physical and chemical methods are used with varying success, for example X-ray crystallography and enzymatic cleavage. Often many methods, both experimental and computational, have to be combined to determine the true structure of nucleic acids (Dong et al. 2001).

Methods for calculating the hybridization parameters of nucleic acids

The oldest and frequently used method for estimating the melting temperature of 2 nucleotide strands with complementary sequences is by their GC content. It is practical for long nucleotide chains where it is not possible to use precise thermodynamic calculations. In long molecules, the statistical effect averages out the effects of local context. For short oligonucleotides, this method is imprecise because the local context influences the strength of G/C and A/T pairings considerably (Borer et al. 1974). The widely used formula is:

$$T_m = 81.5 + 16.6 \log\left\{\frac{[Na^+]}{1.0 + 0.7[Na^+]}\right\} + 0.41(\%G + C) - \frac{500}{D} - P$$

where D is duplex length and P is the percentage of mismatching positions (Wetmur 1991).

A similar method was traditionally used in probe design – hybridization probes had fixed length, a GC content within specified range, and a certain minimum number of mismatches with non-target sequences. By fulfilling these conditions, it is expected that a given oligonucleotide hybridizes strongly with a target sequence and does not hybridize with non-target sequences at annealing temperature, i.e. the hybridization is specific.

The G/C and A/T match and mismatch counting only takes into account the effect of hydrogen bonds between Watson-Crick pairs and ignores stacking effects between neighboring bases. Thus it is not precise, and in the worst cases the estimated T_m may be significantly inaccurate. This may result, for example, in poorly or nonspecifically annealing PCR primers.

Although not designed for the evaluation of hybridization strength BLASTN (Altschul et al. 1990) is a popular tool for estimating the strength and specificity of nucleic acid hybridization, especially the number of nonspecific hybridization sites in full genomes. The main benefits of BLASTN compared to more precise methods is the speed of the alignment search and the possibility of evaluating longer sequences where the nearest-neighbor thermodynamic algorithm does not work. BLASTN is designed for calculating the evolutionary distance between 2 sequences and uses 4x4 substitution matrix of transition

probabilities for scoring alignments. By modifying the substitution matrix, it is possible to make BLASTN estimate the true non-specific hybridization probability of nucleotides with higher precision that with the default evolutionary matrix (Eklund et al. 2010).

Nearest neighbor thermodynamic calculation

In theory, knowing ΔG of hybridization at given temperature makes it possible to calculate precisely the equilibrium constant of hybridization, and thus find the optimal nucleotide sequences and parameters for experiment. In practice, the calculation of true ΔG is extremely complicated or impossible because of the entropy contribution of the steric effects caused by the secondary or tertiary structure of nucleic acids. Thus a simplified model of ΔG calculation is used, most commonly the nearest neighbor algorithm (NN algorithm). It gives very precise results for short oligonucleotide fragments that hybridize in almost a single-step reaction and do not form strong secondary structures. NN algorithm is not appropriate for longer sequences (≥ 40bp).

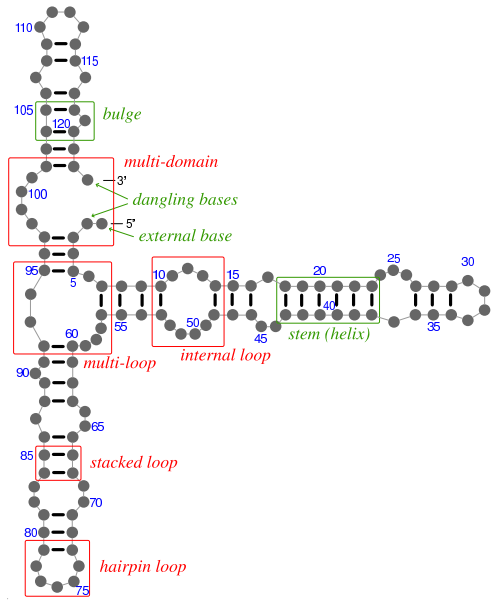
The NN method takes into account both the hydrogen bonds between bases of opposite strands and stacking interactions between adjacent bases of the same strand, generalizing these into a single set of thermodynamic values (ΔH and ΔS) that characterize hybridized dinucleotide (Borer et al. 1974). While calculating the hybridization energy of oligonucleotide sequence, it is necessary to sum the entropy and enthalpy or free energy values of each dinucleotide pair in the hybridized sequence (i.e. move along the sequence one nucleotide a time with the 2 nucleotide-wide window). The contribution of each single base pair into global enthalpy and entropy (and ΔG) is distributed between 2 values – one of the left-side dinucleotide and the other the right-side dinucleotide. The thermodynamic parameters of the full sequence is the sum of the contribution of all dinucleotide pairs (Breslauer et al. 1986).

The NN algorithm can be used not only for calculating the hybridization energy of perfect reverse complementary strands, but also strands with mismatches. For the latter, energy and entropy values of mismatched dinucleotide pairs have to be known.

In addition to the contribution of paired dinucleotides, the NN method has been extended to take into account other features (SantaLucia and Hicks 2004):

- The contribution of single terminating nucleotide pair
- Terminal mismatches
- Dangling ends (i.e., where one nucleotide chain is longer than other)
- Gaps (i.e., the 2 strands are not perfectly aligned but one forms loop)
- Internal loops (i.e., a certain inner part of alignment is not hybridized state)
- Coaxial stacking (i.e., the hybridization of 2 distinct middle regions of nucleic acid chain to a single complementary template)

Examples of structural features in RNA secondary structure are shown on Figure 2.



Example of a pseudoknot-free secondary structure containing all elementary structures (Andronescu M.R. 2000)

Figure 2. Structural features of RNA secondary structure.

For example, the ΔG of internal loop can be calculated using the following formula (SantaLucia and Hicks 2004):

$$\Delta G_{total} = \Delta G_{loopN} + \Delta G_{asymmetry} + \Delta G_{left\ mismatch} + \Delta G_{right\ mismatch}$$

where ΔG_{loopN} is constant, $\Delta g_{asymmetry}$ depends on the difference between the lengths of both sides of the internal loop, and $\Delta G_{left\_mismatch}$ and $\Delta G_{right\_mismatch}$ are the sequence-dependent contributions of mismatches at the ends of the loop.

The NN method can be extended to next-nearest-neighbor (NNN) method, which uses trinucleotides as elementary blocks. Although in some situations it gives even better predictions than the NN model, the gains are small compared to the difficulties in determining the correct parameters for mismatched tri-nucleotides (Owczarzy et al. 1999; SantaLucia and Hicks 2004).

Nearest neighbor thermodynamics can be successfully used for DNA/DNA, DNA/RNA and RNA/RNA interactions when the relevant dinucleotide hybridization parameters are known. For the ones involving RNA, this method is particularly important because the energetic contributions of nucleotide and dinucleotide pairs in the case of RNA interactions vary much more than in DNA/DNA interactions. For example, G/U mismatch in double-stranded RNA is almost as energetically favorable as the A/U Watson-Crick pair. Thus the probes designed by counting matches and mismatches may work very poorly unless the specific properties of RNA are taken into account.

By predicting the ΔH and ΔS of hybridization reaction, the NN thermodynamic model is also the most precise method for estimating the T_m of oligonucleotides. The most widely used formula also takes into account the concentration of cations and the nature of oligonucleotides:

$$T_m\,primer = \frac{\Delta H}{\Delta S + R * \ln\left(C_T / F\right)} + 12.5 \log\left[Na^+\right] - 273.15$$

where ΔH and ΔS are the enthalpy and entropy for helix formation respectively, R is the molar gas constant (1.987 cal/K° mol), and C_T is the total molar concentration of the annealing oligonucleotides. F describes the "symmetry" of oligonucleotide concentration. It is 1 for self-complementary oligonucleotides and 4 for non self-complementary oligonucleotides in equimolar concentration. The formula uses the concentration of sodium; if other cations are used in the buffer their concentrations can be converted to equivalent Na^+ concentration with empirical formulas (Stellwagen, Muse, and Stellwagen 2011; SantaLucia, Allawi, and Seneviratne 1996).

In most cases NN thermodynamics predicts very well the energy and T_m of oligonucleotide hybridization (Breslauer et al. 1986; Rychlik, Spencer, and Rhoads 1990; Chavali et al. 2005; Hughesman, Turner, and Haynes 2011). But it is important to keep in mind that the commonly used NN thermodynamic model treats ΔH and ΔS values as temperature-independent. This is incorrect, overestimating oligonucleotide T_m at high temperatures. At low temperatures the NN method also loses precision because it is unable to take into account the potential secondary and tertiary structure of single-stranded oligonucleotides (Hughesman, Turner, and Haynes 2011).

A more precise estimation of T_m can be done by using the experimentally determined heat capacity change:

$$T_m = \frac{\Delta H°(T_{ref}) + \Delta C_p (T_m - T_{ref})}{\Delta S°(T_{ref}) + \Delta C_p \ln(T_m / T_{ref}) + R ln(C_T / 4)}$$

where $\Delta H°(T_{ref})$ and $\Delta S°(T_{ref})$ are the thermodynamic values calculated with the temperature independent NN parameters, determined at the reference temperature (usually at 37°C) and ΔC_p is the heat capacity change (~ 42 cal $mol^{-1}K^{-1}$ for a single nucleotide pair). This formula gives a significantly better estimation of oligonucleotide T_m at temperatures up to 70°C, but still loses precision at low temperatures (Hughesman, Turner, and Haynes 2011).

The NN method is also the most widely used and precise method for predicting the possible secondary structures of oligonucleotide strands. It can give precise prediction about the free energy (and thus stability) of different structures (Borer et al. 1974), but since it relies on local energy minimization, it cannot always predict the relative abundance of different structures, especially

if competition between different hybridizations is involved (Chavali et al. 2005).

NN thermodynamic calculations give accurate results only if the hybridization behaves as a one-step reaction. This mostly holds true only for nucleotide sequences of <40 base-pairs. Longer sequences hybridize in a multi-step process in which initially small double-stranded "seed" is formed. This "seed" elongates from the initial hybridization point in 2 directions, given that the flanking regions of both strands are complementary (Levitt 2009). As the reaction kinetics of initial hybridization and elongation are quite different, the NN method cannot be used to calculate hybridization properties of long nucleotide chains.

Considerable effort has gone into determining the correct NN hybridization parameters. The most complete unified hybridization model and set of NN parameters was published (SantaLucia 1998). Although the most common cases have been solved with great accuracy, the work remains on-going, e.g. for the ΔH and ΔS temperature dependence (Hughesman, Turner, and Haynes 2011), unnatural bases, secondary structure variants and microarray-aligned probes (Hooyberghs, Van Hummelen, and Carlon 2009). The web-based database project NNDB (http://rna.urmc.rochester.edu/NNDB) collects recently published NN parameters together with tutorials and references about their application (Turner and Mathews 2010).

Many programs have been devised for calculating hybridization parameters (ΔH, ΔS, ΔG and T_m) of nucleic acids, usually in the context of some other task (PCR primer design, hybridization probe design). **MELTING** is an extensible specialized tool for calculating hybridization properties; it supports all major types of hybridization (DNA/DNA, DNA/RNA and RNA/RNA), allows the specification of the concentrations various cation types, denaturing agents and unnatural bases (inosine and azobenzene). It uses NN thermodynamic model for short oligonucleotides and several published empirical formulas for longer sequences where NN is unreliable (Dumousseau et al. 2012). **Visual-OMP** uses very precise physical model of oligonucleotide hybridization and multi-stage equilibrium algorithms for the design and evaluation of PCR primers and probes (SantaLucia 2007).

Practical applications of nucleic acid hybridization prediction

PCR primer design

Polymerase Chain Reaction (PCR) primers need to have certain desirable properties:

- They have to hybridize strongly at target sites at the annealing step
- They should not hybridize non-specifically at the annealing step
- They should completely dissociate at the heating step

- They should not form strong dimers with themselves or opposite primers
- They should not form strong hairpin loops

All these requirements are statistical, being related to average primer behavior. Some primers almost always hybridize non-specifically, some form weak dimers and so on. The important point is that the strength and frequency of incorrect hybridizations have to be low to ensure that only the correct reaction is carried out to a significant extent. Thus the efficiency of PCR relies fundamentally on the thermodynamic properties of oligonucleotide hybridization.

It is crucial to avoid mispriming in the first cycles of PCR as these will be amplified at a similar concentration to the actual products. On the other hand, artefacts such as partial products and chimaeras, formed at the later stages of PCR, will remain in low concentration compared to the product (Kanagawa 2003).

In common applications, the most important parameters of the PCR primer pair are melting temperature and secondary structure formation.

Primer T_m has to be in an optimal range depending on the PCR conditions (annealing temperature) and the T_m of template DNA (Mitsuhashi 1996); too low an annealing temperature (relative to the melting temperature of primers and product) can lead to the formation of false products, and too high an annealing temperature decreases the PCR yield (Rychlik, Spencer, and Rhoads 1990). Currently the best algorithm for T_m calculation is the NN thermodynamic model (Chavali et al. 2005).

PCR primers are normally 20–24 nucleotides long, i.e. well below the length threshold of the applicability of the NN model. Thus the NN algorithm can be used to predict precisely their hybridization strengths and eliminate candidate primers with wrong T_m. The most common all-purpose primer design software Primer3 uses this approach. NN thermodynamics also allow the precise calculation of the hybridization strength of the 3' sub-region of primer, which is the most important in PCR specificity (Mitsuhashi 1996; Miura et al. 2005).

The NN model cannot be used to calculate the melting temperature of product since it is not applicable to long DNA sequences (Rychlik, Spencer, and Rhoads 1990).

Primer secondary structures, such as hairpins and primer dimers lower the success rate of PCR. NN thermodynamics is also used to calculate the strengths of possible primer dimers and hairpins, and eliminate those that form too strong structures. Among these structures, the most unwanted ones are those that involve 3'-ends of primers (Mitsuhashi 1996). Although published NN thermodynamic models allow also the estimation of the free energy of different secondary structures of primers (SantaLucia and Hicks 2004), they are not always directly correlated with PCR success. One has to take into account that primer-product hybridization competes with the secondary structures of both primers and products, and thus the real effect on PCR success depends on the precise equilibrium between the different states (Chavali et al. 2005).

Depending on the PCR application, there often are additional requirements for primers like constraints to their length, need for specific tag at one end and so on.

Good primers should also avoid extreme GC content and certain nucleotide patterns, such as single nucleotide and CA repeats that are easily enforceable in primer design software (Yamada, Soma, and Morishita 2006; Mallona, Weiss, and Egea-Cortines 2011). The synthesis of a new DNA strand starts from the 3'-end of the primer and thus it is thought that the nucleotide composition at the 3'-end strongly influences the success rate of PCR. Both too high and too low GC content at the 3'-end is detrimental to PCR specificity (Mitsuhashi 1996). Certain nucleotide triplets (WSS and SWS) are over-represented and some (WCG) underrepresented in the 3'-ends of successful primers (Onodera and Melcher 2004).

Predicting the strength and number of nonspecific hybridization sites is more complicated, because not only has the actual sequence of interest to be taken into account (PCR product with flanking regions), but the sequence of the whole DNA that may be present in PCR tube. This may, for example, be the whole human genome and/or several bacterial genomes.

A perfect match between the whole primer and template DNA is not needed for the PCR reaction to start. Hybridization is a probabilistic process and even weak non-specific hybridizations can start polymerase reaction as long as the 3'-end of the primer is in a hybridized state (Mitsuhashi 1996; Miura et al. 2005; Yamada, Soma, and Morishita 2006; Kalendar, Lee, and Schulman 2011). As the primer sequence is incorporated into such false product, subsequent PCR cycles with this product work at an efficiency similar to the proper product. Thus it is often advisable to test not only the whole primer for uniqueness in source genome, but also shorter subsequences, especially the 3'-end (Mitsuhashi 1996; Andreson, Möls, and Remm 2008).

On the other hand, if the hybridization between primer and template is strong but the 3' end of alignment contains mismatches, the probability of a successful start of the replication is many orders of magnitude lower (Petruska et al. 1988). But it is known that certain DNA polymerases have exonucleolytic (proof-reading) activity and can remove mismatched nucleotide from the 3'-end of the primer and then start the elongation (Skerra 1992).

To design the most specific primers, the balance between primer melting temperature and annealing temperature has to be kept in the optimal range to maximize specific and minimize the non-specific hybridization ratio. The true hybridization ratio of primers for a given temperature can be predicted from the ΔG for both specific and nonspecific hybridization. Unfortunately, calculating the hybridization ΔG for all nonspecific positions in the genome is a computationally costly process (Yamada, Soma, and Morishita 2006). One way to speed up the test for non-specific hybridization is to find the shortest 3' subsequence that gives significant probability of mispriming, and test for the uniqueness of such a subsequence. The precise length of such subsequences

(SDSS – specificity determining subsequence) varies and can be calculated using NN thermodynamics (Miura et al. 2005).

The simplest method to avoid primers that may give too many non-specific hybridizations with the genome is to test the source sequence for motifs that are repeated in the whole genome. Often this is done in the form of masking repeats, i.e. marking in the source sequence the regions that are locally similar to known repeating motifs and should be excluded from primer design. It is expected that the remaining parts of the sequence are locally sufficiently unique, so that even if primers designed to such regions have few nonspecific binding sites in genome, the probability of any 2 of these sites being close enough together to start a false reaction is very low.

A common and most widely used method is masking known repeating motifs in the human genome (or other known genomes). The most widely used program for this is RepeatMasker, which uses precompiled libraries of repeated sequences.

A more precise method is to find the abundance of different oligonucleotide motifs in genome and mask all those words that have too many identical or similar motifs in the genome. Many algorithms and programs have been developed over the years, but only a few of them are applicable to routine PCR primer design (Andreson, Möls, and Remm 2008; Morgulis et al. 2006).

The most popular PCR primer design tool is **Primer3**. It can design both PCR primers and (optionally) hybridization probes, taking into account many experimentally determined factors in addition to the NN model (Untergasser et al. 2012).

Pythia uses a small set of thermodynamic parameters to estimate the PCR efficiency and precomputed index of non-specific hybridization sites with specified thermodynamic stability (Mann et al. 2009).

pcrEfficiency is a Web-based tool that uses statistical model to predict the efficiency of PCR reactions (Mallona, Weiss, and Egea-Cortines 2011).

RexPrimer uses Primer3 as the primer design backend and uses a set of databases of genomic variance (SNPs, indels, pseudogenes) to avoid designing primers to non-unique regions (Piriyapongsa et al. 2009).

Although the in-silico primer design algorithms have evolved considerably in last 20 years they cannot guarantee the success of PCR using designed primers. Thus if failure is not an option, for example, if the primer pair is used for diagnostic purposes, experimental verification of primers cannot be avoided (Carter et al. 2010).

Primer evaluation is a time-consuming process if either a large number of primers is required or an extensive test for nonspecific hybridization is needed. Thus for certain common tasks, such as human gene expression analysis, prebuilt primer databases are created (Wang et al. 2012).

Multiplex PCR

Multiplexing several PCR reactions into single tube is a very popular method to save time, chemicals, source DNA, and provide internal control of reaction success. Multiplex PCR has been successfully used, for example, in bridging gaps in shotgun sequencing (Sorokin et al. 1996; Tettelin et al. 1999), mutation and indel detection (Edwards and Gibbs 1994; Shen et al. 2010), pathogen detection (Elnifro et al. 2000; Strommenger et al. 2003; Pinar et al. 2004; Gardner et al. 2009; Shi et al. 2012), quantitative analysis of GM food (Rudi, Rud, and Holck 2003; Jinxia et al. 2011) and genotyping (Edwards and Gibbs 1994).

The number of concurrent PCR reactions in a single tube (multiplexing level) is usually in the low 10's (Edwards and Gibbs 1994; Nicodème and Steyaert 1997; Tettelin et al. 1999; Syvänen 2005) but much larger levels can be used in specific protocols (Fredriksson et al. 2007).

The average quality of a PCR is lower in a multiplex environment and is negatively correlated with the multiplex level (Edwards and Gibbs 1994; Henegariu et al. 1997; Rudi, Rud, and Holck 2003; Shum and Paul 2009). The precise causes are not well understood (Kanagawa 2003), but some probable factors can nevertheless be outlined.

Nonspecific hybridization of primers to source DNA that creates false products

Although this can happen with single-plex PCR as well, its effect on multiplex PCR may be much stronger simply because there are more different primer sequences in the reaction (Sorokin et al. 1996; Rudi, Rud, and Holck 2003; Shum and Paul 2009). Each primer-primer combination can potentially initiate the synthesis of a false product if the nonspecific hybridization sites of those primers fall close enough together and have correct directionality. The number of possible combinations of primer pairs in a multiplex environment is proportional to 2^n, where n is the number of PCR reactions in single tube (Syvänen 2005).

Although primer design programs can successfully exclude those that give significant probability of forming a false product between the primers of the same PCR, they cannot predict the false product between primers from different PCR.

Formation of primer-primer dimers

Most primer design programs test for dimers between the primers in the same PCR and exclude pairs that hybridize too strongly with each other. But in a multiplex environment dimers between primers from different reactions can form (Edwards and Gibbs 1994; Nicodème and Steyaert 1997; Syvänen 2005). Although the hybridization strength of primer dimers is normally low, the concentration of primers is initially very high compared to the concentration of target DNA. Thus even weak hybridizations can start a polymerase reaction

(Brownie 1997). Among all possible hybridizations between PCR primers the ones involving the 3'-end are considered the most dangerous. It is proposed that the ΔG of hybridization which includes 3'-end of both primers should be kept below 6 kcal/mol (Rachlin et al. 2005b).

Nonspecific interactions between primers and products

As the concentration of all products increases with time, so does the probability of nonspecific interactions between any primer and product. Normally primers are tested so as not to form too strong alignments with their own product, but interactions with the product of other PCR are possible in a multiplex environment.

Differences between PCR efficiency

Different PCR primer pairs have different efficiency (i.e. amplification speed), which introduces bias to multiplex PCR. PCR drift is variation in the product yield that is unbiased and most probably caused by the random fluctuations in the first few PCR cycles that subsequently become amplified (Kanagawa 2003). PCR selection is the biased difference between speeds of PCR reactions caused by the properties of primer or product sequences. Faster reactions may outcompete slower reactions in multiplex environment by depleting reagents. Also, if quantitative measurements are needed, the bias will cause either over- or underestimation of the specific product.

Primers with similar melting temperatures have more similar rates of amplification. Thus to achieve a more uniform amplification, it is useful to limit the maximum difference between the primer melting temperatures and product lengths in groups (Edwards and Gibbs 1994; Nicodème and Steyaert 1997; Elnifro et al. 2000; Syvänen 2005). Also calibrating primer concentrations of individual PCR reactions can be used to make reaction rates uniform (Jinxia et al. 2011). The relative speed of PCR reactions is dependent on the concentration of primers and products. In the final stage of PCR there is a strong tendency towards a 1:1 ratio of products (Kanagawa 2003).

There is greater probability of the occurrence of heteroduplexes and chimeras in multiplex PCR. A heteroduplex forms if 2 different products partially hybridize with each other. A chimera is a linear conglomerate of partial products of the same or different PCR. As these artefacts happen with higher probability in the later stages of PCR when the concentration of products is high and primers are depleted, it is advisable to terminate multiplex PCR earlier than single-plex (Kanagawa 2003).

Although not related to PCR failure, it is sometimes necessary to impose additional constraints to product length in a multiplex group. If gel electrophoresis is used to detect the result of PCR, it is necessary to limit the minimum difference between any product lengths in a multiplex group (Pinar et al. 2004).

For these reasons, it is often challenging to achieve high multiplex levels. Difficulty of distributing PCR primers into multiplexing groups is critically dependent on the average compatibility between any 2 primer pairs and the

multiplexing level required. For each compatibility and group size combination, there exists computational “phase transition", i.e. the maximum practical group size from which, upwards, the size of necessary calculations for finding usable groups starts to increase exponentially (Rachlin et al. 2005b).

Finding the optimal groups for PCR multiplexing is NP complete problem, but simple heuristic algorithms can achieve a near optimal solution (Nicodème and Steyaert 1997).

In many cases, the experimental conditions, such as the concentration of primers, enzyme and Mg^{++}, annealing and extension times, and temperatures, can be adjusted to increase the efficiency of multiplex PCR (Henegariu et al. 1997; Edwards and Gibbs 1994; Strommenger et al. 2003; Carter et al. 2010). Increasing PCR cycle lengths guarantees that all parallel reactions have time to be completed and reduces the differences between product yields (Shum and Paul 2009). These approaches have to be adjusted for each experimental situation. Thus they are time-consuming and cannot replace the stringent primer design and grouping (Elnifro et al. 2000). Using chemically modified bases in primers can help to suppress false products and make amplification speeds more uniform (Shum and Paul 2009).

Creating optimal or near-optimal multiplex groups is a multi-objective problem (Rachlin et al. 2005a; Lee, Shin, and Zhang 2007). Simultaneously there is a need to consider the number of groups, multiplexing levels and the quality or inclusiveness of PCR. Although the achievable grouping level depends on the list of primers, it is practical to separate primer design and grouping to make the algorithm simpler. Given the initial set of chemically similar primers, the multi-objective evolutionary algorithm (MOEA) performs good grouping with a reasonable performance (Lee, Shin, and Zhang 2007). For a fixed coverage, the maximum achievable multiplexing level is proportional to log(N) where N is the number of PCR reactions (Rachlin et al. 2005b).

Several tools and algorithms have been developed for multiplex primer design.

MultiPLX calculates the primer-dimer strengths using NN thermodynamic model and uses greedy algorithm to distribute a set of primer pairs into optimal groups based on their compatibility. In addition to hybridization strengths it can use user-specified compatibility values that makes it very flexible (Kaplinski et al. 2007).

MuPlex integrates primer design and multiplex group creation for the design of SNP genotyping primers using various criteria for multi-objective optimization (Rachlin et al. 2005a). Among the factors to be considered are T_m difference, the strengths of primer dimers, the number of non-specific hybridization sites, group sizes and SNP coverage.

PrimerStation uses an algorithm that predicts the actual primer hybridization ratio in liquid solution at annealing temperature, using the thermodynamic model. Good primers for multiplexing are required to have a hybridization ratio of >0.99 for specific binding and <0.05 for non-specific binding. As the calculation of precise ΔG is computationally costly, PrimerStation includes

a precomputed database of potential primer candidates with good specificity (Yamada, Soma, and Morishita 2006).

The MPP algorithm allows the design of multiplex primers to large number of diverse targets where multiple sequence alignment (MSA) is not practical or possible (Gardner et al. 2009).

One method for reducing potential primer-dimer formation is to minimize the number of distinct primers needed for multiplexed PCR. In this case, certain primers are chosen such that they participate in many individual PCR reactions, but each pair of primers can be amplified only in a single reaction. An implementation of such primer selection algorithm is **G-POT** (Konwar et al. 2005).

The average probability of primer compatibility is also crucial to multiplex PCR design. Thus for high multiplexing levels, specific experimental techniques that increase primer compatibility with each other are needed, for example, using universal primers (Rachlin et al. 2005b).

Another option to increase the success rate of multiplex PCR is to customize PCR protocol in such a way that only the first cycles use longer specific primers and the main amplification involves universal primer sequences. This method gives more uniform amplification speeds, allowing using it for quantitative PCR (Rudi, Rud, and Holck 2003). As the universal primers are specifically designed to be not complementary with any region of source DNA and to not form dimers the formation of nonspecific products is also suppressed. But even in this case, the initial specific primers have to be designed taking into account all possible cross-reactions (Brownie 1997; Shi et al. 2012).

One method for increasing the detection rate of multiplex PCR products is to selectively pick only correct products from the amplified mixture. Correct products have terminal sequences composed of primers from single pair, whereas most false products have terminal sequences from different pairs. Specific ligation probes can be designed that hybridize to both ends of the product sequences, generating circular products that can be further amplified (Fredriksson et al. 2007).

Although considerable interest for multiplex PCR exists, and many algorithms are created for designing optimized multiplex groups, there is no foolproof method. Thus in most cases, there is no alternative to experimental validation of multiplex groups (Henegariu et al. 1997; Elnifro et al. 2000).

Hybridization probe design

Hybridization based technologies have become popular methods for rapid and accurate analysis of biological samples. They have been successfully used for the detection of pathogenic viruses and bacteria (Wilson, Strout, and DeSantis 2002), analyzing species composition of environmental and medical samples (Guschin et al. 1997), quantitative analysis of GM food (Rudi, Rud, and Holck 2003) and for many other types of analyzes.

Ideally hybridization probes need the following properties (Kaderali and Schliep 2002; Harrison et al. 2013; X.-Q. Xia et al. 2010)

- They should hybridize strongly with the target sequence at the annealing temperature
- They should not hybridize with other sequences that may be present in the solution (non-target sequences) at annealing temperature
- The should not form hairpin loops

Compared to PCR primers, there are a few differences in the behavior of hybridization probes.

First they are usually attached to some kind of solid surface or gel matrix. This means that their movements are constrained and the hybridization energy formula applicable to solution is not precise, although there is a clear correlation between ΔG values obtained for probes in solution and on surfaces (Hooyberghs, Van Hummelen, and Carlon 2009; Hadiwikarta et al. 2012). How big is the difference between the thermodynamic properties of the oligonucleotides in solution and the identical nucleotides attached to surface depends on the technology used and is hard to derive analytically (J. Fuchs et al. 2010). Thus it is advisable to experimentally determine the correction factor for ΔG and T_m values calculated using the solution formula or to determine separate sets of NN parameters for immobilized oligonucleotides (Hooyberghs, Van Hummelen, and Carlon 2009; X.-Q. Xia et al. 2010).

During hybridization, the local concentration of hybridized probes increases much more rapidly than in solution and the electrostatic repulsion between the negatively charged phosphate chains affects the hybridization kinetics (J. Fuchs et al. 2010). The uneven distribution of probes influences the hybridization behavior (Vainrub and Pettitt 2011). The surface has limited accessibility and mass transport of target molecules to probes may become the limiting step in hybridization kinetics (J. Fuchs et al. 2010).

Fluorescent labeling, often used in microarray experiments, affects the hybridization properties of oligonucleotides. Immobilization of probes to surface also reduces the difference in the intensities of fluorescent signals of specific and nonspecific hybridization compared to solution hybridization (Hooyberghs, Van Hummelen, and Carlon 2009).

If hybridization probes are used for discriminating between closely related target molecules, such as the sequences from 2 bacterial strains, the precise calculation of hybridization properties becomes extremely important. Such sequences often differ only by few nucleotides, and consequently their hybridization energies and T_m values are very close (Dong et al. 2001; Kaderali and Schliep 2002). It is not sufficient to simply design highest-affinity probes to target sequence because these probes may have considerable affinity with the control sequence, thereby giving false positive signals (Bernhard Maximilian Fuchs et al. 1998). To achieve maximum discriminatory power, the ΔG of target hybridization has to be as negative, and ΔG of control hybridization, as positive as possible at the precise annealing temperature. In such a case the equilibrium

of probe-target interaction is mostly in the hybridized state and the equilibrium of probe-control interaction is mostly in the free state (Figure 1).

Unlike PCR, microarray hybridization occurs at a constant temperature. Thus hybridization probes have to compete with the secondary structure of target molecules that may exist at that temperature. If this secondary structure is too strong, hybridization to probes may be suppressed because their complementary regions in target molecule are either directly or indirectly blocked (Ratushna, Weller, and Gibas 2005).

If the hybridization energy (ΔG) of the secondary structure of target molecule is higher (ΔG is more negative) than ΔG between probe and target, the secondary structure out-competes hybridization probes, and a given target will not hybridize at all. If ΔG of the secondary structure is higher (i.e. hybridization is weaker) than between the target and the probe, the probes will eventually hybridize, but the process may be very slow. The cause of this is that at every moment most target molecules have secondary structures, and the regions complementary to probes are blocked. For hybridization to occur the target molecule has first to loosen and then come to the proximity of a hybridization probe. Both of these processes have a low probability and thus the combination of 2 has even lower (Dong et al. 2001).

The blockage of hybridization by the secondary structure of a target molecule is more prevalent in RNA targets because their secondary structure is stronger than that of DNA (T. Xia et al. 1998). Certain native RNA molecules (e.g., rRNA and tmRNA), have strong evolutionary conserved structures that make their hybridization challenging (Ratushna, Weller, and Gibas 2005; Yilmaz, Okten, and Noguera 2006). This is especially true if the hybridization has to take place at lower temperatures, at which RNA secondary structure is very stable (Peplies, Glockner, and Amann 2003).

There are several approaches to break RNA secondary structure and make their hybridization possible. The simplest and most used method is either to increase hybridization temperature or use chemicals that lower the T_m (i.e. lower the free energy of nucleic acid hybridization). Both of these affect not only the secondary structure formation, but also probe-target hybridization energy and T_m, but as this hybridization is normally stronger than the secondary structure of target, the overall hybridization speed increases (Ratushna, Weller, and Gibas 2005; Small et al. 2001).

Another method is breaking the target molecule into smaller parts. Although the effective concentration of the target is lowered this way, since breakage sometimes occurs in the complementary region, the remaining fragments have much weaker structures and the hybridization is more effective (Pozhitkov et al. 2006; Ratushna, Weller, and Gibas 2005).

If the precise secondary structure of the target molecule can be determined, it is possible to design hybridization probes with this taken into account (Ratushna, Weller, and Gibas 2005). As the secondary structure is often critically dependent on few mutations in primary sequence such probes may have great discriminatory power even under non-stringent conditions. For

example, probes that extend both sides of stem-loop structures can be designed. These probes can only hybridize on the correct structure, and are discriminatory even at temperatures as low as 4°C (Dong et al. 2001).

Another method is using small specific helper nucleotide molecules (chaperones) that hybridize to certain regions of the target molecule. They block the complementary or semi-complementary regions inside the target and thus the secondary structure cannot form. Although chaperones themselves block certain regions in the target, other regions become available for hybridization (Peplies, Glockner, and Amann 2003; Small et al. 2001; B. M. Fuchs et al. 2000; Kaplinski et al. 2010).

Compared to hybridization probes, chaperones are much more capable of competing with native secondary structure for 2 reasons. First, they may be applied at a temperature higher than the hybridization (annealing) temperature when the secondary structure is much weaker. If the temperature is subsequently lowered, a strong secondary structure in the target molecule cannot form. Second, unlike bound probes, chaperones are added to the solution at much higher concentration than that of the effective hybridization probes; thus they can interact with target molecules much more freely compared to hybridization probes that are attached to a solid matrix and are in low concentration.

There are many tools available for microarray probe design, often tailored for specific applications and technologies. **Oligoarray** is a widely used program using the NN thermodynamic model to choose probes with precise hybridization T_m (Rouillard, Zucker, and Gulari 2003). **ChipD** allows the design of thermodynamically similar probes that uniformly cover the whole length of source DNA sequence (Dufour et al. 2010). **OligoWiz** can use sequence annotation information, such as exon-intron structure in design process (Wernersson and Nielsen 2005). **Teolenn** is built as an extensible framework that allows users to prioritize between different quality scores (Jourdren et al. 2010).

PRESENT INVESTIGATION AND DISCUSSION

Aims of the present study

The main goal of the present study has been to investigate how to apply the well-known models of oligonucleotide hybridization to high-level experimental procedures of multiplex PCR and microarray hybridization.
The specific aims were:

1. To develop an efficient algorithm and implementation for estimating primer-primer and primer-product interactions in multiplex PCR and automatic grouping of PCR primers, based on the worst-case interaction strengths.
2. To find the main primer sequence-related factors that affect the PCR success rate in the multiplex PCR environment.
3. To develop an efficient method for the design of DNA hybridization probes for closely related RNA targets.
4. To study the effect of RNA secondary structure on hybridization efficiency and to develop a method to increase the RNA/DNA hybridization at low temperatures.

Method for automatic distribution of PCR into multiplex groups (REF. I)

Multiplexing PCR is a powerful method for lowering the cost of experiments. While primer-design software can predict the occurrence of primer-dimers and false products between the primers of single PCR, it cannot take into account the possibility of unwanted hybridizations between the primers and products of different PCR amplified together. The complexity of finding a compatible set of primer pairs to be amplified together grows exponentially, depending on the desired number of reactions in a single tube. We developed an automated tool, **MultiPLX** that calculates the strengths of different primer-dimers and primer-product interactions between all primer combinations in a given list of primer pairs using the NN thermodynamic model, and distributes these into optimized groups.

MultiPLX Algorithm

MultiPLX calculates 8 different parameters (scores) for each combination of 2 primer pairs.

1. Maximum binding energy (minimum ΔG) of 2 primers including 3'-ends of both primers
2. Maximum binding energy of 3'-end of one primer with any region of another primer
3. Maximum binding energy of any region of different primers
4. Maximum binding energy of 3'-end of one primer with any region of a PCR product

5. Maximum binding energy of any region of a primer with any region of a PCR product
6. Maximum product length difference between compared PCR primer sets
7. Minimum product length difference between compared PCR primer sets
8. Maximum difference in primer melting temperatures between compared PCR primer sets.

A schema of the hybridization types corresponding to the first 3 parameters is shown in Figure 3.

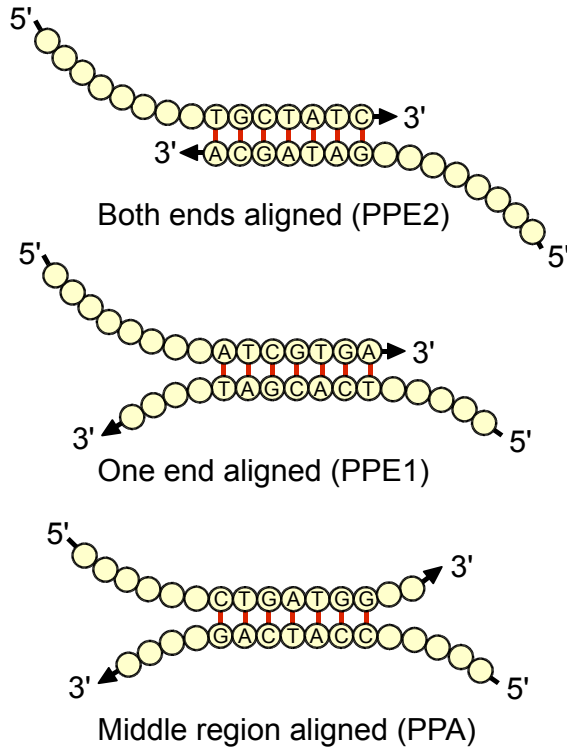


Figure 3. Schematic diagram of different primer dimer variants whose hybridization energies have been calculated by the MultiPLX algorithm. The variants differ in whether both, one or neither 3'-ends are in the hybridized state.

MultiPLX uses the NN thermodynamic method for the calculation of both the strength of local hybridizations between primers and products, and calculating the melting temperature of primers. By default, it uses published entropy and enthalpy values (Kaderali 2001) but the user can also submit a custom thermodynamic table if a newer or better adjusted version is available.

In addition to scores calculated by MultiPLX, the grouping algorithm can also include one additional user-specified score value. This allows the incorporation of the results of other primer evaluation and compatibility tests into the

grouping; for example taking into account the probabilities of the false products caused by non-specific hybridizations of primers from different pairs.

The calculated scores are used to find optimized group layout by keeping the scores between the members of any single group below specified cut-off values. MultiPLX uses a 2-tier approach by first distributing primer pairs into groups by a greedy algorithm and then equalizing group sizes by moving compatible primer pairs from big to smaller groups. If no more primers can be moved into smaller groups, but groups remain unequal size, it can perform an additional "shuffling" step by exchanging primers randomly between groups while preserving the compatibilities in the group.

Determining the factors that predict the failure of multiplex PCR (REF IV)

The success rate of PCR is lower in multiplex PCR, but the causes of this phenomenon are not well understood. We carried out a series of experiments to determine which parameters of primers are most strongly correlated with the failure of PCR in a multiplex environment. We were especially interested whether a) certain primer pairs disturbed (poisoned) other pairs that would normally work well, and b) whether certain combinations of otherwise good primer pairs were failing.

We created 8000 4-plex groups from primer pairs that worked flawlessly in previous single-plex control experiments. The reaction conditions were exactly the same as in the single-plex experiment, but roughly 1 in 4 reactions failed in multiplex. We created the best logistic model for predicting failure of multiplex PCR using primer pair parameters as independent variables. The factors incorporated into the best model would then correspond to the properties of primer pairs causing PCR failure in the multiplex experiment.

The analysis shows that the most important factors causing the failure of multiplex PCR are GC content of primers and the number of non-specific hybridization sites on the template DNA. Both of these factors also influence the quality of single-plex PCR in a similar way (Andreson, Möls, and Remm 2008). The more non-specific hybridization sites PCR primers had, the higher the probability of PCR failing in the multiplex environment. The detrimental effect of non-specific hybridization to PCR quality was thus similar, but much stronger, than in single-plex PCR.

We also found that the number of non-specific hybridization sites affected both the outcome of the same PCR and of other PCR in the same multiplex group (Figure 4). Primer pairs with a high number of non-specific hybridization sites caused the failure of other PCR in the same tube in addition to failing themselves. Thus these primers in particular should be avoided in multiplex experiments, and only high-quality primers should be used.

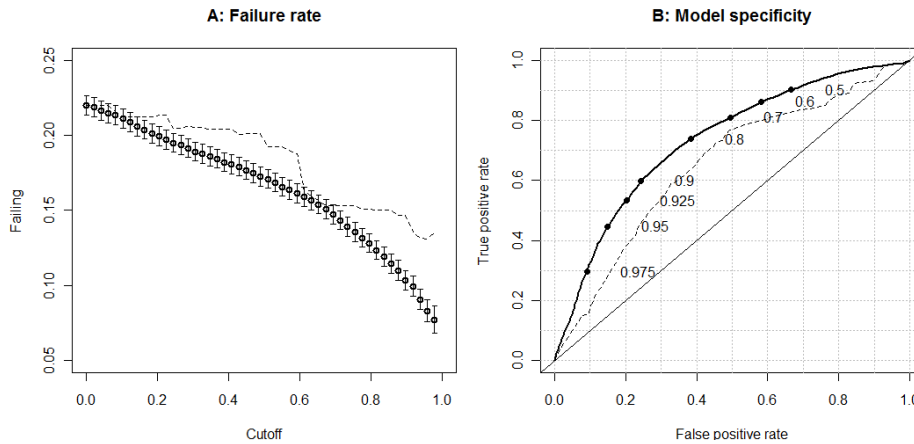


Figure 4. Improvement of multiplex PCR quality after applying PCR group quality model. Left: improvement of average PCR success rate after elimination of all primer pairs with predicted failure rate above a certain cut-off value. X-axis: applied cut-off value. Y axis: the ratio of failing PCR reactions. Dashed line: comparable failure rate of a model that ignores other group members. Right: ROC plot of model prediction. X-axis: the ratio of remaining failing primer pairs (of all failing pairs) after eliminating those with the prediction below the cut-off value. Y-axis: the ratio of remaining working primer pairs (of all working pairs). Dashed line: comparable plot of a model that ignores other group members. Cut-off values from 0.5–0.975 are marked on the ROC curve.

Our model indicates that single primer properties are significantly more important than interactions between primers from different pairs in predicting the failure rate of multiplex PCR. The interactions between primers had no statistically significant influence on the failure rate in our 4-plex experiments.

SLICSel (REF III)

Using bacterial tmRNA for species and strain discrimination poses specific challenges. First the tmRNA sequences of closely related species are very similar, making the number of possible discriminating probes limited. Second, tmRNA is known to fold into a strong secondary structure such that only certain parts of the whole molecule are exposed. And third, RNA/DNA hybridization is much stronger and less specific than DNA/DNA hybridization.

To design the probes with maximum discriminatory power, the precise hybridization energies of the probes has to be calculated with all possible target and control sequences, and ensured that all targets hybridize with ΔG below, and all controls with ΔG above zero. The bigger the ΔG difference between the hybridization of a probe with 2 RNA molecules (ΔΔG), the higher will be the ratio of fluorescent signal intensities. Thus to get optimal detection probes, one

has to keep the worst case ΔΔG between any target and control hybridization above a certain cut-off value.

In certain cases, it is useful to have probes that are specific to multiple target sequences – for example, this applies with a single probe that has to detect the presence of any one of multiple pathogenic strains of a bacterium, with the non-pathogenic strains being control group. In this situation, it is important to keep the hybridization levels of probe with any of these alternate targets as close as possible. This can be ensured by keeping the ΔΔG of any 2 target hybridization below certain cut-off values.

We developed custom application SLICSel to do an exhaustive search among all possible hybridization probes with specified melting temperatures and choose the ones that match the abovementioned criteria. The probes picked by SLICSel are designed to discriminate between 2 sets of sequences, the target and the control sets. The ΔΔG cut-offs of target and control hybridization can be specified by the user, among other relevant parameters, including annealing temperature, desired probe length and so on.

Both DNA/DNA and DNA/RNA thermodynamic tables are included with SLICSel. It also allows the user to specify a custom table when very specific reaction conditions are needed.

SLICSel also accepts degenerate nucleotide symbols in both target and control sequences. In both instances the worst case value is used – the highest ΔΔG for target and lowest ΔΔG for control hybridization.

SLICSel has 3 unique properties: a) it does search by brute-force and lists all probe candidates that have at least the user-provided sensitivity and specificity; b) it designs probes that selectively detect any one sequence from a list of target sequences in the environment that may contain any or all of the control sequences; and c) it allows custom thermodynamic values to be used.

To verify our algorithm, we designed 97 probes to Streptococcus pneumoniae tmRNA with ΔΔG varying from 0 to –11 kcal/mol and probe length from 9 to 26 bp. Five other bacterial tmRNA sequences were used as controls: *S.pyogenes, S.agalactiae*, GrC/G *Streptococcus*, *K.pneumoniae* and *M.catharralis*. All tmRNA molecules were synthesized in vitro and hybridization experiment were done by microarray.

The majority of designed probes had good specificity to *S.pneumoniae* tmRNA. Of the 463 hybridization events only 20 had false positive signals that were stronger than 10% of the true positive signal level. The relative intensity of non-specific signals was inversely proportional to ΔΔG. By setting ΔΔG cut-off to 2 kcal/mol, only 6 false positive hybridization events remained, and no such event was detected by setting ΔΔG cut-off to 4 kcal/mol (Figure 5).

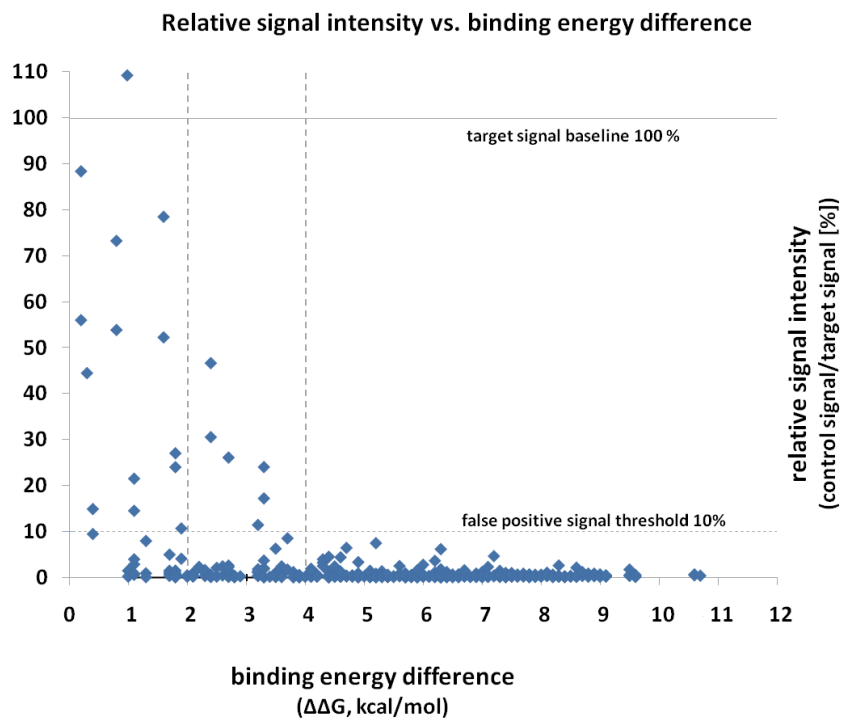


Figure 5. Relative intensity of false positive signals depending on the hybridization energy difference (ΔΔG) between specific and strongest nonspecific hybridization (Scheler et al. 2011).

Chaperones (REF II)

Detection of RNA molecules with DNA hybridization probes works well at temperatures above 45°C. Hybridization signals are weak at lower temperatures, although according to NN thermodynamics, the hybridization should be stronger at lower temperatures.

We performed a series of experiments with the microbial detection microarray using different hybridization times and verified that the weak fluorescent signals were caused by a slow hybridization speed. At hybridization temperatures above 42°C, signal strengths reached plateau in < 4 h, and at 34°C they did not reach equilibrium even after 12 h.

The slow hybridization speed may be caused by strong secondary and tertiary structures of the tmRNA molecule. We assessed the possibility of improving microarray hybridization signals by designing short complementary DNA oligonucleotides, which will potentially block secondary and tertiary structure forming regions in tmRNA and thus keep the target molecule accessible to microarray probes (Figure 6).

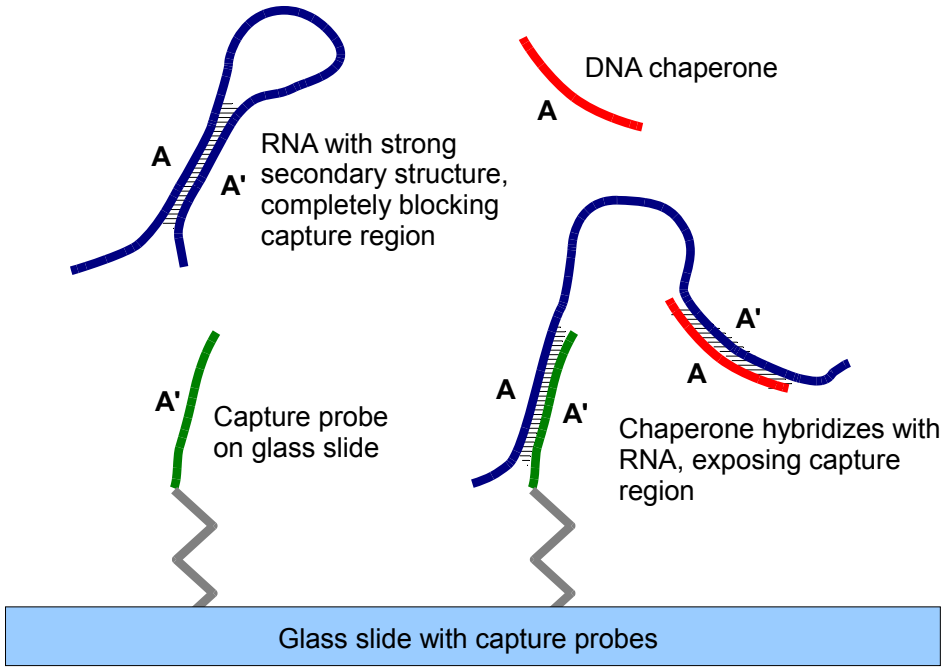


Figure 6. The working principle of DNA chaperones. Chaperones block the secondary structure formation and expose other parts of the molecule by hybridizing to the complementary region of RNA.

We designed 6 different helper oligonucleotides (chaperones), each complementary to one predicted functional domain of *S.pneumoniae* tmRNA. Chaperones were applied at elevated temperature, the solution was cooled and hybridization carried out.

We used previously designed detection microarray with 97 specific probes that covered most of the length of S.pneumoniae tmRNA. This allowed us to measure the changes in signal strength (i.e. hybridization speed) simultaneously at different regions of target molecule.

In the first experiment we applied all 6 chaperones simultaneously.

As expected, chaperone oligonucleotides suppressed these signals which had probes partially overlapping with chaperone region. All other signals increased on average, confirming our hypothesis. There were also visible peaks in the relative signal strength, suggesting regions that were almost completely blocked in the “native” state but that became well-exposed after introducing the chaperones (Figure 7).

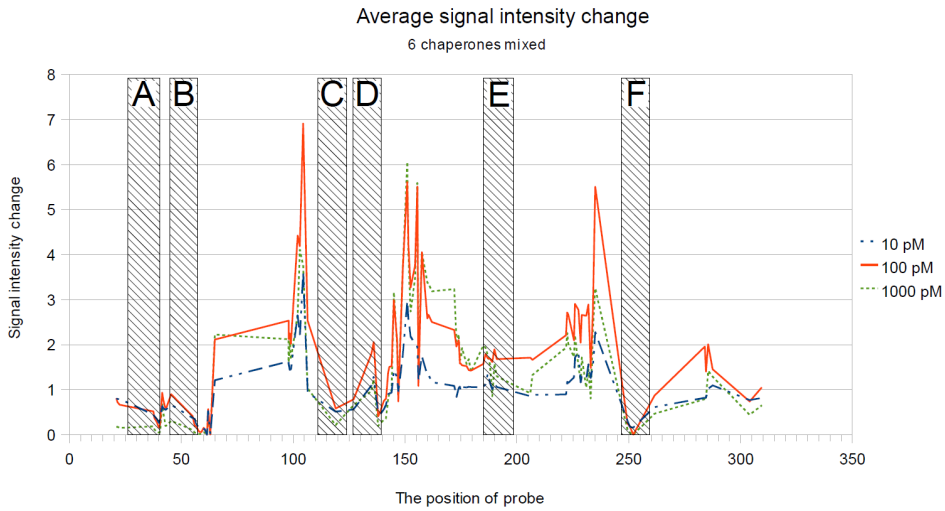


Figure 7. Relative change in the signal intensities of all 97 tmRNA-specific probes after applying the mix of all 6 chaperones. X axis – position of probe midpoint on tmRNA molecule. Y axis – relative change in signal intensity of the experiments with chaperones compared to the untreated base. (Kaplinski et al. 2010)

In the next experiment chaperones were applied one-by-one, allowing us to study, which regions of tmRNA would become accessible by applying a specific chaperone.

As in the first experiment, the signals whose probes were overlapping with chaperones were suppressed. For each chaperone signals increased significantly in some other areas of tmRNA. For chaperone F, these regions were the immediate neighboring areas, suggesting the existence of a hairpin loop. For chaperones A and E, the regions with improved signal strengths were separated, suggesting the existence of more complex tertiary structure.

CONCLUSIONS

In this investigation the oligonucleotide hybridization model was applied to multiplex PCR primer design and grouping, and to microarray hybridization.

1. We designed an efficient algorithm to partition existing PCR primer pairs into multiplex groups based on the minimum required compatibility between primers, and implemented it as a software package MultiPLX. The compatibility is defined as keeping the primer-primer or primer-product dimer hybridization energy below a specified cut-off value. MultiPLX calculates the hybridization energies using the NN thermodynamic model.
2. We analysed statistically which sequence-related properties of PCR primers are most strongly correlated with the failure of PCR in the 4-plex environment. The most important factors were primer GC content and the number of nonspecific hybridization sites of primers to genome. The more non-specific hybridization sites primer had the lower was the average success rate of the multiplex PCR. Also primers with many hybridization sites not only performed poorly themselves, but disturbed the PCR of other primer pairs in the same tube.
3. We implemented the program SLICSel that designs hybridization probes by exhaustively searching specified target sequences. By using the NN thermodynamic model, SLICSel picks all potential hybridization probes with the correct melting temperature and specified minimal difference between the specific and nonspecific hybridization strengths. We demonstrated experimentally that using a ΔG cut-off of 4kcal/mol at the annealing temperature was sufficient to ensure probe specificity in a wide range of probe lengths and GC content.
4. We designed a set of 6 specific helper oligonucleotides (chaperones) that hybridized to a tmRNA molecule and prevented the formation of secondary structure. By using these helper oligonucleotides, we could increase the hybridization speed at 34°C up to 4-fold. We also demonstrated how the hybridization of chaperone on one part of the target molecule can increase signals in distant parts of the molecule.

REFERENCES

Altschul, S F, W Gish, W Miller, E W Myers, and D J Lipman. 1990. "Basic Local Alignment Search Tool." *Journal of Molecular Biology* 215 (3) (October 5): 403–10. doi:10.1016/S0022-2836(05)80360-2.

Andreson, Reidar, Tõnu Möls, and Maido Remm. 2008. "Predicting Failure Rate of PCR in Large Genomes." *Nucleic Acids Research* 36 (11) (June 1): e66. doi:10.1093/nar/gkn290.

Borer, PN, B Dengler, I Tinoco, and OC Uhlenbeck. 1974. "Stability of Ribonucleic Acid Double-stranded Helices." *Journal of Molecular Biology* 86 (4): 843–853.

Breslauer, KJ, R Frank, H Blöcker, and Marky LA. 1986. "Predicting DNA Duplex Stability from the Base Sequence." *Proceedings of the National Academy of Sciences* 83 (11): 3746–3750.

Brownie, J. 1997. "The Elimination of Primer-dimer Accumulation in PCR." *Nucleic Acids Research* 25 (16) (August 15): 3235–3241. doi:10.1093/nar/25.16.3235.

Carter, Ian W.J., Catriona Halliday, Theo P Sloots, Todd M Pryce, Ian D Kay, Gerald B Harnett, Glenys R Chidlow, and Philip M Giffard. 2010. "PCR Methodology." In *PCR for Clinical Microbiology*, ed. Margret Schuller, Theo P. Sloots, Gregory S. James, Catriona L. Halliday, and Ian W.J. Carter, 11–47. Dordrecht: Springer Netherlands. doi:10.1007/978-90-481-9039-3.

Chavali, S, A Mahajan, R Tabassum, R Maiti, and D Bharadwaj. 2005. "Oligonucleotide Properties Determination and Primer Designing: a Critical Examination of Predictions." *Bioinformatics* 21 (20): 3918–3925.

Cisse, II, H Kim, and T Ha. 2012. "A Rule of Seven in Watson-Crick Base-pairing of Mismatched Sequences." *Nature Structural & Molecular Biology* 19 (6): 623–627.

Dong, F, HT Allawi, T Anderson, BP Neri, and VI Lyamichev. 2001. "Secondary Structure Prediction and Structure-specific Sequence Analysis of Single-stranded DNA." *Nucleic Acids Research* 29 (15): 3248–3257.

Dufour, Yann S, Gary E Wesenberg, Andrew J Tritt, Jeremy D Glasner, Nicole T Perna, Julie C Mitchell, and Timothy J Donohue. 2010. "chipD: a Web Tool to Design Oligonucleotide Probes for High-density Tiling Arrays." *Nucleic Acids Research* 38 (Web Server issue) (July 1): W321–5. doi:10.1093/nar/gkq517.

Dumousseau, Marine, Nicolas Rodriguez, Nick Juty, and Nicolas Le Novère. 2012. "MELTING, a Flexible Platform to Predict the Melting Temperatures of Nucleic Acids." *BMC Bioinformatics* 13 (1) (January): 101. doi:10.1186/1471-2105-13-101.

Edwards, MC, and RA Gibbs. 1994. "Multiplex PCR: Advantages, Development, and Applications." *Genome Research* 3 (4): S65–S75.

Eklund, Aron C, Pia Friis, Rasmus Wernersson, and Zoltan Szallasi. 2010. "Optimization of the BLASTN Substitution Matrix for Prediction of Non-specific DNA Microarray Hybridization." *Nucleic Acids Research* 38 (4) (March 1): e27. doi:10.1093/nar/gkp1116.

Elnifro, Elfath M, Ahmed M Ashshi, Robert J Cooper, and Paul E Klapper. 2000. "Multiplex PCR: Optimization and Application in Diagnostic Virology." *Clinical Microbiology Reviews* 13 (4): 559–570.

Fredriksson, Simon, Johan Banér, Fredrik Dahl, Angela Chu, Hanlee Ji, Katrina Welch, and Ronald W Davis. 2007. "Multiplex Amplification of All Coding Sequences Within 10 Cancer Genes by Gene-Collector." *Nucleic Acids Research* 35 (7) (January 1): e47. doi:10.1093/nar/gkm078.

Fu, G, A Miles, and L Alphey. 2012. “Multiplex Detection and SNP Genotyping in a Single Fluorescence Channel.” *PloS One* 7 (1): e30340.

Fuchs, B. M., F. O. Glockner, J. Wulf, and R. Amann. 2000. “Unlabeled Helper Oligonucleotides Increase the In Situ Accessibility to 16S rRNA of Fluorescently Labeled Oligonucleotide Probes.” *Applied and Environmental Microbiology* 66 (8) (August 1): 3603–3607. doi:10.1128/AEM.66.8.3603-3607.2000.

Fuchs, Bernhard Maximilian, Gunter Wallner, Wolfgang Beisker, Ines Schwippl, Wolfgang Ludwig, and Rudolf Amann. 1998. “Flow Cytometric Analysis of the In Situ Accessibility of Escherichia Coli 16S rRNA for Fluorescently Labeled Oligonucleotide Probes.” *Appl. Envir. Microbiol.* 64 (12) (December 1): 4973–4982.

Fuchs, J, J-B Fiche, A Buhot, R Calemczuk, and T Livache. 2010. “Salt Concentration Effects on Equilibrium Melting Curves from DNA Microarrays.” *Biophysical Journal* 99 (6) (September 22): 1886–95. doi:10.1016/j.bpj.2010.07.002.

Gardner, Shea N, Amy L Hiddessen, Peter L Williams, Christine Hara, Mark C Wagner, and Bill W Colston. 2009. “Multiplex Primer Prediction Software for Divergent Targets.” *Nucleic Acids Research* 37 (19) (October 1): 6291–304. doi:10.1093/nar/gkp659.

Guschin, DY, BK Mobarry, D Proudnikov, DA Stahl, BE Rittmann, and AD Mirzabekov. 1997. “Oligonucleotide Microchips as Genosensors for Determinative and Environmental Studies in Microbiology.” *Appl. Envir. Microbiol.* 63 (6) (June 1): 2397–2402.

Hadiwikarta, WW, JC Walter, J Hooyberghs, and E Carlon. 2012. “Probing Hybridization Parameters from Microarray Experiments: Nearest-neighbor Model and Beyond.” *Nucleic Acids Research* 40 (18): e138.

Harrison, Andrew, Hans Binder, Arnaud Buhot, Conrad J Burden, Enrico Carlon, Cynthia Gibas, Lara J Gamble, et al. 2013. “Physico-chemical Foundations Underpinning Microarray and Next-generation Sequencing Experiments.” *Nucleic Acids Research* 41 (5) (January 9): 2779–2796. doi:10.1093/nar/gks1358.

Henegariu, O, N A Heerema, S R Dlouhy, G H Vance, and P H Vogt. 1997. “Multiplex PCR: Critical Parameters and Step-by-step Protocol.” *Biotechniques* 23 (3): 504–11. doi:10.1007/BF00928712.

Hooyberghs, J, P Van Hummelen, and E Carlon. 2009. “The Effects of Mismatches on Hybridization in DNA Microarrays: Determination of Nearest Neighbor Parameters.” *Nucleic Acids Research* 37 (7) (April 1): e53. doi:10.1093/nar/gkp109.

Hughesman, CB, RFB Turner, and C Haynes. 2011. “Correcting for Heat Capacity and 5′-TA Type Terminal Nearest Neighbors Improves Prediction of DNA Melting Temperatures Using Nearest-neighbor Thermodynamic.” *Biochemistry* 50 (13): 2642–2649.

Jinxia, Ao, Li Qingzhang, Gao Xuejun, Yu Yanbo, Li Lu, and Zhang Minghui. 2011. “A Multiplex Nested PCR Assay for the Simultaneous Detection of Genetically Modified Soybean, Maize and Rice in Highly Processed Products.” *Food Control* 22 (10) (October): 1617–1623. doi:10.1016/j.foodcont.2011.03.018.

Jourdren, Laurent, Aurélie Duclos, Christian Brion, Thomas Portnoy, Hugues Mathis, Antoine Margeot, and Stéphane Le Crom. 2010. “Teolenn: An Efficient and Customizable Workflow to Design High-quality Probes for Microarray Experiments.” *Nucleic Acids Research* 38 (10) (June 1): e117. doi:10.1093/nar/gkq110.

Kaderali, L. 2001. “Selecting Target Specific Probes for DNA Arrays”. Universität Köln.

Kaderali, L., and A. Schliep. 2002. “Selecting Signature Oligonucleotides to Identify Organisms Using DNA Arrays.” *Bioinformatics* 18 (10) (October 1): 1340–1349. doi:10.1093/bioinformatics/18.10.1340.

Kanagawa, Takahiro. 2003. “Bias and Artifacts in Multitemplate Polymerase Chain Reactions (PCR).” *Journal of Bioscience and Bioengineering* 96 (4) (January): 317–23. doi:10.1016/S1389-1723(03)90130-7.

Kaplinski, Lauris, Reidar Andreson, Tarmo Puurand, and Maido Remm. 2007. “MultiPLX: Automatic Grouping and Evaluation of PCR Primers.” Ed. Anton Yuryev. *Bioinformatics* 21 (8): 1701–1702.

Kaplinski, Lauris, Ott Scheler, Sven Parkel, Priit Palta, Kadri Toome, Ants Kurg, and Maido Remm. 2010. “Detection of tmRNA Molecules on Microarrays at Low Temperatures Using Helper Oligonucleotides.” *BMC Biotechnology* 10 (1): 34.

Konwar, KM, II Mandoiu, AC Russell, and AA Shvartsman. 2005. “Improved Algorithms for Multiplex PCR Primer Set Selection with Amplification Length Constraints.” In *Proc. 3rd Asia-Pacific Bioinformatics Conference (APBC)*, 41–50.

Lee, IH, SY Shin, and BT Zhang. 2007. “Multiplex PCR Assay Design by Hybrid Multiobjective Evolutionary Algorithm.” In *Evolutionary Multi-Criterion Optimization*, 376–385. Springer.

Levitt, M. 2009. “Folding of Nucleic Acids.” *Polymerization in Biological Systems, Ciba Foundation Symposium* 7: 147–171.

Mallona, Izaskun, Julia Weiss, and Marcos Egea-Cortines. 2011. “pcrEfficiency: a Web Tool for PCR Amplification Efficiency Prediction.” *BMC Bioinformatics* 12 (1) (January): 404. doi:10.1186/1471-2105-12-404.

Mann, Tobias, Richard Humbert, Michael Dorschner, John Stamatoyannopoulos, and William Stafford Noble. 2009. “A Thermodynamic Approach to PCR Primer Design.” *Nucleic Acids Research* 37 (13) (July 1): e95. doi:10.1093/nar/gkp443.

Markham, Nicholas R, and Michael Zuker. 2008. “UNAFold: Software for Nucleic Acid Folding and Hybridization.” *Methods in Molecular Biology (Clifton, N.J.)* 453 (January): 3–31. doi:10.1007/978-1-60327-429-6_1.

Mitsuhashi, M. 1996. “Technical Report: Part 2. Basic Requirements for Designing Optimal PCR Primers.” *Journal of Clinical Laboratory Analysis* 10 (5) (January): 285–93. doi:10.1002/(SICI)1098-2825(1996)10:5<285::AID-JCLA9>3.0.CO;2-7.

Miura, Fumihito, Chihiro Uematsu, Yoshiyuki Sakaki, and Takashi Ito. 2005. “A Novel Strategy to Design Highly Specific PCR Primers Based on the Stability and Uniqueness of 3’-end Subsequences.” *Bioinformatics (Oxford, England)* 21 (24) (December 15): 4363–70. doi:10.1093/bioinformatics/bti716.

Morgulis, Aleksandr, E Michael Gertz, Alejandro A Schäffer, and Richa Agarwala. 2006. “WindowMasker: Window-based Masker for Sequenced Genomes.” *Bioinformatics (Oxford, England)* 22 (2) (January 15): 134–41. doi:10.1093/ bioinformatics/ bti774.

Nicodème, P, and JM Steyaert. 1997. “Selecting Optimal Oligonucleotide Primers for Multiplex PCR.” *Proc Int Conf Intell Syst Mol Biol* 5: 210–213.

Onodera, Kenji, and Ulrich Melcher. 2004. “Selection for 3’ End Triplets for Polymerase Chain Reaction Primers.” *Molecular and Cellular Probes* 18 (6): 369–372.

Owczarzy, R, P M Vallone, R F Goldstein, and A S Benight. 1999. “Studies of DNA Dumbbells VII: Evaluation of the Next-nearest-neighbor Sequence-dependent Interactions in Duplex DNA.” *Biopolymers* 52 (1) (January): 29–56. doi:10.1002/(SICI)1097-0282(1999)52:1<29::AID-BIP2>3.0.CO;2-K.

Peplies, J., F. O. Glockner, and R. Amann. 2003. “Optimization Strategies for DNA Microarray-Based Detection of Bacteria with 16S rRNA-Targeting Oligonucleotide

Probes." *Applied and Environmental Microbiology* 69 (3) (March 1): 1397–1407. doi:10.1128/AEM.69.3.1397-1407.2003.

Petruska, J, MF Goodman, MS Boosalis, LC Sowers, C Cheong, and I Tinoco. 1988. "Comparison Between DNA Melting Thermodynamics and DNA Polymerase Fidelity." *Proceedings of the National Academy of Sciences* 85 (17): 6252–6256.

Pinar, A, N Bozdemir, T Kocagöz, and R Alacam. 2004. "Rapid Detection of Bacterial Atypical Pneumonia Agents by Multiplex PCR." *Central European Journal of Public Health* 12 (1): 3–5.

Pozhitkov, Alex, Peter A Noble, Tomislav Domazet-Loso, Arne W Nolte, Rainer Sonnenberg, Peer Staehler, Markus Beier, and Diethard Tautz. 2006. "Tests of rRNA Hybridization to Microarrays Suggest That Hybridization Characteristics of Oligonucleotide Probes for Species Discrimination Cannot Be Predicted." *Nucleic Acids Research* 34 (9) (January 17): e66. doi:10.1093/nar/gkl133.

Rachlin, John, C Ding, C Cantor, and S Kasif. 2005a. "MuPlex: Multi-objective Multiplex PCR Assay Design." *Nucleic Acids Research* 33 (Web Server Issue): W544–W547.

Rachlin, John, Chunming Ding, Charles Cantor, and Simon Kasif. 2005b. "Computational Tradeoffs in Multiplex PCR Assay Design for SNP Genotyping." *BMC Genomics* 6 (1) (January): 102. doi:10.1186/1471-2164-6-102.

Ratushna, Vladyslava G, Jennifer W Weller, and Cynthia J Gibas. 2005. "Secondary Structure in the Target as a Confounding Factor in Synthetic Oligomer Microarray Design." *BMC Genomics* 6 (1) (January): 31. doi:10.1186/1471-2164-6-31.

Rouillard, Jean-Marie, Michael Zucker, and Erdogan Gulari. 2003. "OligoArray 2.0: Design of Oligonucleotide Probes for DNA Microarrays Using a Thermodynamic Approach." *Nucleic Acids Research* 31 (12) (June 15): 3057–3062. doi:10.1093/nar/gkg426.

Rudi, K, I Rud, and A Holck. 2003. "A Novel Multiplex Quantitative DNA Array Based PCR (MQDA-PCR) for Quantification of Transgenic Maize in Food and Feed." *Nucleic Acids Research* 31 (11): e62.

Rychlik, W, WJ Spencer, and RE Rhoads. 1990. "Optimization of the Annealing Temperature for DNA Amplification in Vitro." *Nucleic Acids Research* 18 (21): 6409–6412.

SantaLucia, John. 1998. "A Unified View of Polymer, Dumbbell, and Oligonucleotide DNA Nearest-neighbor Thermodynamics." *Proceedings of the National Academy of Sciences* 95 (4) (February 17): 1460–1465. doi:10.1073/pnas.95.4.1460.

SantaLucia, John. 2007. "Physical Principles and visual-OMP Software for Optimal PCR Design." *Methods in Molecular Biology (Clifton, N.J.)* 402 (January): 3–34. doi:10.1007/978-1-59745-528-2_1.

SantaLucia, John, Hatim T Allawi, and P Ananda Seneviratne. 1996. "Improved Nearest-neighbor Parameters for Predicting DNA Duplex Stability." *Biochemistry* 35 (11) (March 19): 3555–62. doi:10.1021/bi951907q.

SantaLucia, John, and Donald Hicks. 2004. "The Thermodynamics of DNA Structural Motifs." *Annual Review of Biophysics and Biomolecular Structure* 33 (January 12): 415–40. doi:10.1146/annurev.biophys.32.110601.141800.

Scheler, Ott, Lauris Kaplinski, Barry Glynn, Priit Palta, Sven Parkel, Kadri Toome, Majella Maher, Thomas Barry, Maido Remm, and Ants Kurg. 2011. "Detection of NASBA Amplified Bacterial tmRNA Molecules on SLICSel Designed Microarray Probes." *BMC Biotechnology* 11 (1): 17.

Shen, Z, W Qu, W Wang, Y Lu, and Y Wu. 2010. "MPprimer: a Program for Reliable Multiplex PCR Primer Design." *BMC Bioinformatics* 11 (1): 143.

Shi, Hui, Quoclinh Trinh, Wentao Xu, Baiqiang Zhai, Yunbo Luo, and Kunlun Huang. 2012. "A Universal Primer Multiplex PCR Method for Typing of Toxinogenic Pseudomonas Aeruginosa." *Applied Microbiology and Biotechnology* 95 (6) (July 20): 1579–1587. doi:10.1007/s00253-012-4277-8.

Shum, J, and N Paul. 2009. "Chemically Modified Primers for Improved Multiplex Polymerase Chain Reaction." *Analytical Biochemistry* 388 (2): 266–272.

Skerra, Arne. 1992. "Phosphorothioate Primers Improve the Amplification of DNA Sequences by DNA Polymerases with Proofreading Activity." *Nucleic Acids Research* 20 (14) (July 25): 3551–3554. doi:10.1093/nar/20.14.3551.

Small, J., D. R. Call, F. J. Brockman, T. M. Straub, and D. P. Chandler. 2001. "Direct Detection of 16S rRNA in Soil Extracts by Using Oligonucleotide Microarrays." *Applied and Environmental Microbiology* 67 (10) (October 1): 4708–4716. doi:10.1128/AEM.67.10.4708-4716.2001.

Sorokin, A, A Lapidus, V Capuano, N Galleron, P Pujic, and S D Ehrlich. 1996. "A New Approach Using Multiplex Long Accurate PCR and Yeast Artificial Chromosomes for Bacterial Chromosome Mapping and Sequencing." *Genome Research* 6 (5) (May 1): 448–453. doi:10.1101/gr.6.5.448.

Stellwagen, Earle, Joseph M Muse, and Nancy C Stellwagen. 2011. "Monovalent Cation Size and DNA Conformational Stability." *Biochemistry* 50 (15) (April 19): 3084–94. doi:10.1021/bi1015524.

Strommenger, B., C. Kettlitz, G. Werner, and W. Witte. 2003. "Multiplex PCR Assay for Simultaneous Detection of Nine Clinically Relevant Antibiotic Resistance Genes in Staphylococcus Aureus." *Journal of Clinical Microbiology* 41 (9) (September 4): 4089–4094. doi:10.1128/JCM.41.9.4089-4094.2003.

Sugimoto, N, S Nakano, and M Katoh. 1995. "Thermodynamic Parameters to Predict Stability of RNA/DNA Hybrid Duplexes." *Biochemistry* 34 (35): 11211–11216.

Syvänen, AC. 2005. "Toward Genome-wide SNP Genotyping." *Nature Genetics* 37: S5–S10.

Tettelin, H, D Radune, S Kasif, H Khouri, and SL Salzberg. 1999. "Optimized Multiplex PCR: Efficiently Closing a Whole-genome Shotgun Sequencing Project." *Genomics* 62 (3): 500–507.

Turner, Douglas H, and David H Mathews. 2010. "NNDB: The Nearest Neighbor Parameter Database for Predicting Stability of Nucleic Acid Secondary Structure." *Nucleic Acids Research* 38 (Database issue) (January 1): D280–2. doi:10.1093/nar/gkp892.

Untergasser, Andreas, Ioana Cutcutache, Triinu Koressaar, Jian Ye, Brant C Faircloth, Maido Remm, and Steven G Rozen. 2012. "Primer3 – new Capabilities and Interfaces." *Nucleic Acids Research* 40 (15) (August 1): e115. doi:10.1093/nar/ gks596.

Vainrub, Arnold, and B Montgomery Pettitt. 2011. "Accurate Prediction of Binding Thermodynamics for DNA on Surfaces." *The Journal of Physical Chemistry. B* 115 (45) (November 17): 13300–3. doi:10.1021/jp208141g.

Wang, Xiaowei, Athanasia Spandidos, Huajun Wang, and Brian Seed. 2012. "PrimerBank: a PCR Primer Database for Quantitative Gene Expression Analysis, 2012 Update." *Nucleic Acids Research* 40 (Database issue) (January 1): D1144–9. doi:10.1093/nar/gkr1013.

Wernersson, Rasmus, and Henrik Bjørn Nielsen. 2005. "OligoWiz 2.0--integrating Sequence Feature Annotation into the Design of Microarray Probes." *Nucleic Acids Research* 33 (Web Server issue) (July 1): W611–5. doi:10.1093/nar/gki399.

Wetmur, James G. 1991. “DNA Probes: Applications of the Principles of Nucleic Acid Hybridization.” *Critical Reviews in Biochemistry and Molecular Biology* 26 (3-4): 227–259.

Wienken, CJ, and P Baaske. 2011. “Thermophoretic Melting Curves Quantify the Conformation and Stability of RNA and DNA.” *Nucleic Acids Research* 39 (8): e52.

Wilson, WJ, CL Strout, and TZ DeSantis. 2002. “Sequence-specific Identification of 18 Pathogenic Microorganisms Using Microarray Technology.” *Molecular and Cellular Probes* 16 (2): 119–127.

Xia, T, J SantaLucia, M E Burkard, R Kierzek, S J Schroeder, X Jiao, C Cox, and D H Turner. 1998. “Thermodynamic Parameters for an Expanded Nearest-neighbor Model for Formation of RNA Duplexes with Watson-Crick Base Pairs.” *Biochemistry* 37 (42) (October 20): 14719–35. doi:10.1021/bi9809425.

Xia, Xiao-Qin, Zhenyu Jia, Steffen Porwollik, Fred Long, Claudia Hoemme, Kai Ye, Carsten Müller-Tidow, Michael McClelland, and Yipeng Wang. 2010. “Evaluating Oligonucleotide Properties for DNA Microarray Probe Design.” *Nucleic Acids Research* 38 (11) (June 1): e121. doi:10.1093/nar/gkq039.

Yamada, T, H Soma, and S Morishita. 2006. “PrimerStation: a Highly Specific Multiplex Genomic PCR Primer Design Server for the Human Genome.” *Nucleic Acids Research* 34 (Web Server Issue): W665–W669.

Yilmaz, L Safak, Hatice E Okten, and Daniel R Noguera. 2006. “Making All Parts of the 16S rRNA of Escherichia Coli Accessible in Situ to Single DNA Oligonucleotides.” *Applied and Environmental Microbiology* 72 (1) (January 1): 733–44. doi:10.1128/AEM.72.1.733-744.2006.

SUMMARY IN ESTONIAN

Oligonukleotiidide hübridisatsioonimudeli rakendamine PCR-i ja mikrokiipide optimeerimiseks

Nukleiinhapped on põhiliselt tuntud kui "digitaalse" geneetilise koodi kandjad kõigis teadaolevates elusrakkudes. Nukleiinhapete monomeeride (nukleotiidide) võimalik järjestus ühes DNA ahelas on praktiliselt ilma kitsendusteta ja mõjutab kaksikhelikaalse molekuli keemilisi ja füüsikalisi omadusi ainult vähesel määral. Siiski võivad nukleiinhapped, eelkõige üheahelaline RNA, moodustada spetsiifilisi kolmemõõtmelisi struktuure, mis võivad rakkudes omada metaboolseid või juhtfunktsioone.

Kaksikhelikaalse struktuuri spontaanne moodustumine antikomplementaarsete lõikude vahel samas DNA või RNA molekulis või erinevate molekulide vahel (hübridiseerumine) on aluseks nukleiinhapete kasutamisele biotehnoloogias, näiteks DNA polümeraasi ahelreaktsioonis (PCR) ja hübridisatsioonikiipidel. Paljudes hübridisatsiooni rakendustes käsitletakse lühikesi nukleiinhapete fragmente "digitaalsete" koodidena mis tunnevad spetsiifiliselt ära vastavaid antikomplementaarseid fragmente. Täpsema, universaalsema ja kvaliteetsema tulemuse saamiseks on aga vaja tunda hübridisatsiooni keemilisi aluseid ning käsitleda hübridisatsiooni kui tasakaalulist ja stohhastilist protsessi. Sellisel juhul on võimalik nii spetsiifilise kui mittespetsiifilise hübridisatsiooni tugevust (sulamistemperatuuri T_m) täpselt ennustada, kasutades hübridisatsiooni termodünaamilist mudelit. Kaasajal on kõige täpsemaks selliseks mudeliks lähima naabri mudel (Nearest Neighbour Model), mis arvestab nii antikomplementaarsete lämmastikaluste vaheliste vesiniksidemete kui kõrvutiasuvate lämmastikaluste vaheliste Wan de Waals'i jõudude mõju.

Nii PCR-is kui detektsioonikiipidel on sagedaseks probleemiks mittespetsiifiline hübridisatsioon. Ehkki see on üldjuhul nõrgem kui spetsiifiline hübridisatsioon, on potentsiaalseid mittespetsiifilisi seondumiskohti sageli palju rohkem. PCRi korral võib mittespetsiifiline hübridisatsioon põhjustada praimerite dimeeride ja valeproduktide teket. Detektsioonikiipide korral põhjustab mittespetsiifiline hübridisatsioon valepositiivseid signaale. Hübridiseerumise termodünaamiline mudel võimaldab hinnata täpselt valeseondumiste sagedust teatud temperatuuril ja seega vältida PCR praimereid või detektsiooniproove, millel on suur tõenäosus mittespetsiifiliseks hübridisatsiooniks.

Käesoleva doktoritöö üheks eesmärgiks oli täiustada multipleks-PCR (mitme praimeripaari koosamplifitseerimine) metoodikat luues algoritmi ja programmi praimerite automaatseks grupeerimiseks nii, et minimeerida nendevaheliste ebasoovitavate interaktsioonide võimalusi. Töö raames loodi programm MultiPLX, mis arvutab lähima naabri termodünaamilise mudeli abil kõikvõimalike interaktsioonide hübridisatsioonienergiad ning jaotab praimeripaarid automaatselt gruppidesse nii, et nendevaheliste interaktsioonide tugevus jääks alla etteantud nivoo. Lisaks võimaldab MultiPLX arvestada grupeerimisel produktide pikkusi ja praimerite sulamistemperatuure. Grupeerimiseks kasutab

MultiPLX kiiret lähendusmeetodit paigutades praimeripaarid alguses esimesse sobivasse gruppi. Järgmise sammuna toimub gruppide suuruste ühtlustamine praimeripaaride ümbertõstmise ning vahetamise teel.

Teiseks eesmärgiks oli uurida millised PCR praimerite primaarjärjestuse omadused on seotud multiplex PCR ebaõnnestumisega. Selleks teostati 8000 4-pleks katset kasutades eelenevalt ühe praimeripaariga (single-plex) PCR katsetes verifitseeritud praimeripaare inimese genoomi juhuslike lõikude amplifitseerimiseks. Kokku ebaõnnestus PCR multipleks katsetes 22,6% reaktsioonidest.

Esimese etapina tehti kindlaks, et praimeripaaride mittetöötamine multipleks katsetes ei ole juhuslik, vaid ebaõnnestumise tõenäosus varieerub erinevate praimeripaaride vahel. Seejärel analüüsiti erinevaid praimeripaaride ja produktide primaarjärjestusega seotud parameetreid laiendatud lineaarse statistilise mudeli (GLM) abil. Kõige olulisemad PCR ebaõnnestumisega korrelatsioonis olevad parameetrid olid praimerite GC sisaldus ning praimerite mittespetsiifiliste seondumiste arv genoomile. Lisaks leiti, et praimeripaarid, millel on suur mittespetsiifiliste seondumiste arv, häirivad ka teiste samas grupid amplifitseeritavate praimeripaaride reaktsioone.

Kolmandaks eesmärgiks oli luua kiire algoritm ja programm täieliku hübridiseerimisproovide komplekti loomiseks geneetiliselt lähedaste bakteriliikide eristamiseks tmRNA primaarjärjestuse põhjal. Töö käigus loodi programm SLICSel, mis arvutab lähima naabri termodünaamika põhjal välja kõigi võimalike oligonukleotiidide potentsiaalsed seondumistugevused sihtmärk- ja kontrolljärjestustele etteantud temperatuuril. Nende hulgast on võimalik automaatselt valida alamhulk mille korral seondumistugevuste erinevus spetsiifilise ja mittespetsiifilise seondumise vahel on suurem etteantud lävest.

SLICSel programmi abil disainitud proove testiti eksperimentaalselt bakteriaalse tmRNA detekteerimiseks. Leiti, et spetsiifilise ja mittespetsiifilise seondumistugevuse ΔG erinevus 4 kcal/mol on piisav valepositiivsete signaalide vältimiseks.

Neljandaks eesmärgiks oli leida metoodika bakteriaalse tmRNA signaalide intensiivsuse tõstmiseks detektsioonikiibil kui hübridiseerumine toimub madalal temperatuuril. Tehti kindlaks, et madal signaalide intensiivsus oli tingitud aeglasest hübridiseerumisest, mida tõenäoliselt põhjustas tmRNA tugev sekundaarstruktuur. Töö käigus disainiti kuus spetsiifilist oligonukleotiidi (chaperoni), mis hübridiseerusid tmRNA erinevate regioonidega ning eeldatavalt pidid blokeerima sekundaarstruktuuri moodustumise. Selliselt töödeldud tmRNA hübridiseerumissignaalid olid madalatel temperatuuridel märgatavalt tugevamad. Oodatavalt blokeerisid chaperonid need signaalid, mille proovid hübridiseerusid chaperonidega kattuvatele regioonidele. Seevastu tõusid teiste signaalide intensiivsused, sealhulgas nii nende, mille proovid asusid chaperoni lähedal, kui ka nende, mis asusid tmRNA molekuli kaugetes regioonides.

ACKNOWLEDGEMENTS

First, I would like to thank my supervisor Prof. Maido Remm for his guidance, trust and support during my studies. He mastered perfectly the “leading from behind” leadership style by always giving us, doctoral students, maximum freedom possible to pursue our own goals, while helping us to stay focused whenever we felt lost.

I am grateful to my late B.Sc supervisor Dr. Veljo Sild, who first introduced me to computational modeling of biochemistry, a field which later became my main scientific interest.

I would like to thank all co-authors of the papers which this dissertation is based on.

Special thanks go to Reidar Andreson, Ott Scheler and Priit Palta for their helpful hand in designing and interpreting the results of experiments, and to Märt and Tõnu Möls for their help in statistical analysis of data.

I am grateful for the collective of our department for countless hours of enlightening and inspiring discussions.

Last but not least I want to thank my wonderful family Kersti, Minea, Jerzy Edvard, Karl Aleksander and Frank Artur without whom I either would have completed this dissertation many years earlier or not at all.

PUBLICATIONS

I

Kaplinski L, Andreson R, Puurand T, Remm M (2005).
MultiPLX: automatic grouping and evaluation of PCR primers.
Bioinformatics 21(8): 1701–2.

BIOINFORMATICS APPLICATIONS NOTE Vol. 21 no. 8 2005, pages 1701–1702
doi:10.1093/bioinformatics/bti219

Genome analysis

MultiPLX: automatic grouping and evaluation of PCR primers

Lauris Kaplinski[1,2], Reidar Andreson[1], Tarmo Puurand[1] and Maido Remm[1,*]

[1]Department of Bioinformatics, University of Tartu, Tartu, Estonia and [2]Estonian Biocentre, Riia str. 23, Tartu 51010, Estonia

Received on October 11, 2004; revised on November 25, 2004; accepted on December 10, 2004
Advance Access publication December 14, 2004

ABSTRACT

Summary: MultiPLX is a new program for automatic grouping of PCR primers. It can use many different parameters to estimate the compatibility of primers, such as primer–primer interactions, primer–product interactions, difference in melting temperatures, difference in product length and the risk of generating alternative products from the template. A unique feature of the MultiPLX is the ability to perform automatic grouping of large number (thousands) of primer pairs.

Availability: Binaries for Windows, Linux and Solaris are available from http://bioinfo.ebc.ee/download/. A graphical version with limited capabilities can be used through a web interface at http://bioinfo.ebc.ee/multiplx/. The source code of the program is available on request for academic users.

Contact: maido.remm@ut.ee

INTRODUCTION

Polymerase chain reaction (PCR) is widely used in different areas of science and biotechnology. Grouping of PCR primer sets for amplification in a single tube (multiplexing) provides substantial savings in terms of time, chemicals and, most importantly sample materials. Thus, multiplexing is a powerful way for optimizing the cost of genetic analysis and some procedures of PCR are patented (Piepenbrock *et al*., 2004), which further demonstrates its importance. While a lot of effort has been put into the experimental optimization of multiplex PCR conditions (Henegariu *et al*., 1997; Zangenberg *et al*., 1999), little attention has been paid to the prediction of primer compatibility in multiplex reactions. Smaller multiplex groups are often composed manually and verified by trial and error, but this kind of approach is suboptimal for large datasets. Although there are software programs available for testing the compatibility of PCR primer pairs for multiplexing (Rychlik, 1995; Singh *et al*., 1998; Vallone and Butler, 2004), these do not perform automatic grouping and are unable to handle large datasets automatically. We have developed a program, MultiPLX, for automatic primer compatibility testing and grouping. It can also be used for the compatibility evaluation of the existing primer groups. The program is written in standardized C++ for maximum performance and compatibility. We have built binaries for Linux (×86 compatible processors), Windows and Solaris operating systems. Most of the functionality can be used through the web-based interface also.

*To whom correspondence should be addressed.

ALGORITHM AND IMPLEMENTATION

The workflow of MultiPLX is divided into two main tasks: (1) calculation of compatibility scores for existing PCR primers and (2) grouping of PCR primers based on the cut-off scores. By scores, we are referring to all measurable parameters that may affect the compatibility of different PCR primer sets. Currently, MultiPLX is capable of calculating eight different pre-defined score types:

(1) maximum binding energy (deltaG) of two primers including 3′ ends of both primers
(2) maximum binding energy of 3′ end of one primer with any region of another primer
(3) maximum binding energy of any region of different primers
(4) maximum binding energy of 3′ end of one primer with any region of PCR product
(5) maximum binding energy of any region of primer with any region of PCR product
(6) maximum product length difference between compared PCR primer sets
(7) minimum product length difference between compared PCR primer sets
(8) maximum difference in primer melting temperatures between compared PCR primer sets.

The scores 1–5 are calculated by exhaustive evaluation of all possible gapless alignments and calculation of their deltaG values. The algorithm allows mismatches of any length. The thermodynamic calculations are performed using nearest-neighbor approximation. The default values of enthalpy and entropy of the dinucleotide pairs are the same as used in the program PROBESEL (Kaderali and Schliep, 2002), alternative set of enthalpy and entropy values from the Primer3 program is also available. The concentrations of monovalent salts, Mg^{2+} and DNA can be changed from command line. In addition to the predefined score types, MultiPLX allows the use of a user-specified score, which can be imported into the program. For example, the ability of primers from different PCR primer sets to generate additional PCR products from the human genomic DNA, as calculated by auxiliary program GT4MULTIPLX (http://bioinfo.ebc.ee/gt4multiplx/) can be imported as a custom score into MultiPLX.

Compatibility scores are calculated for all possible pairwise combinations of PCR primer sets. MultiPLX allows in-depth examination of all default scores, including the listing of alignments between primers and their products from different PCR sets. Each score

Table 1. Running times of different stages of MultiPLX software with 10, 100, 1000 and 10 000 randomly generated primer pairs on a Pentium4 2.8 GHz processor on Mandrake Linux 9.2, 1 GB RAM

Number of PCR primer sets to be grouped	Calculation of primer–primer interactions (h:min:s)	Calculation of primer–product interactions (h:min:s)	Calculation of groups (h:min:s)	Average multiplexing level (average number of primer pairs in groups) with 'normal' grouping stringency
10	<00:00:01	<00:00:01	<00:00:01	3.3
100	00:00:02	00:00:46	<00:00:01	12.5
1000	00:04:39	01:14:26	00:00:13	29.4
10 000	07:51:27	129:16:31	00:22:25	49.3

Primer lengths are in the range of 18–22 nt, product lengths in the range of 200–300 bp, average primer GC% is 50. Grouping was performed using default options, without optimization.

type has a default cut-off value to determine whether two given PCR primer sets are compatible and thus can be amplified in the same group. Compatibility cut-off values for all scores can be adjusted by the user either as generic stringency criteria (low, normal, high) or as exact numerical values. Default compatibility cut-offs are generating approximately N3 plex (high stringency), 6–10 plex (normal stringency) or 15–20 plex (low stringency) groups of 100 PCR primer sets.

Automatic grouping of primers is based on compatibility scores and corresponding cut-off values, mentioned above. Two primer sets are considered incompatible if at least one of the compatibility scores exceeds the predefined cut-off value. Two grouping algorithms are implemented.

Algorithm 1. Number of friends (default).

(1) Sort PCR primer sets by the number of compatible primer pairs they have.

(2) In a sorted list of primers, try to insert each primer set into existing groups. If this is not possible, create a new group and move the primer set there.

(3) Repeat until all PCR primer sets have been grouped.

(4) Report the result—list of primer sets associated with their group numbers.

Algorithm 2. Random grouping.

The second algorithm is based on random grouping.

(1) Reorder PCR primer sets randomly.

(2) Try to group each primer set into existing groups. If this is not possible, create a new group and move the primer set there.

(3) Repeat until all PCR primer sets are grouped.

(4) Record the number of groups and group number for each primer.

(5) Repeat the steps 1–4 N times (default value of N is 10 000).

(6) Report the result that generated the lowest number of groups.

The random grouping approach typically generates groups of very different sizes. To make the number of primers in each group more uniform an extra optimization step can be invoked. Optimization tries to move PCR sets from larger groups into smaller ones and swap elements randomly between the groups, until all group sizes differ by no more than one primer pair. None of the grouping algorithms guarantees the optimal solution (lowest possible number of groups). However, the first algorithm typically gives smaller number of groups with shorter computation time. Randomized grouping may be desirable in situations where non-deterministic grouping solution is preferred.

MultiPLX can also be used for the evaluation of existing PCR groups against the calculated score values. Groups where some primer pairs break cut-off rules will be listed alongside the problematic primers. Using this list it is possible to examine, which primers do not fit together and need to be replaced. This option can be helpful for the evaluation of the existing PCR groups with problematic results.

The computation time required by the program for different numbers of primers is shown in Table 1.

ACKNOWLEDGEMENT

This work was supported by the Estonian Ministry of Education and Research grant no. 0182649s04.

REFERENCES

Henegariu,O., Heerema,N.A., Dlouhy,S.R., Vance,G.H. and Vogt,P.H. (1997) Multiplex PCR: critical parameters and step-by-step protocol. *Biotechniques*, **23**, 504–511.

Kaderali,L. and Schliep,A. (2002) Selecting signature oligonucleotides to identify organisms using DNA arrays. *Bioinformatics*, **18**, 1340–1349.

Piepenbrock,C., Adorjan,P., Rujan,T. and Schmitt,A. (2004) Method for amplification of nucleic acids of low complexity. European patent no. WO2004015139.

Rychlik,W. (1995) Selection of primers for polymerase chain reaction. *Mol. Biotechnol.*, **3**, 129–134.

Singh,V.K., Mangalam,A.K., Dwivedi,S. and Naik,S. (1998) Primer premier: program for design of degenerate primers from a protein sequence. *Biotechniques*, **24**, 318–319.

Vallone,P.M. and Butler,J.M. (2004) AutoDimer: a screening tool for primer-dimer and hairpin structures. *Biotechniques*, **37**, 226–231.

Zangenberg,G., Saiki,R.K. and Reynolds,R. (1999) Multiplex PCR: Optimization guidelines. In Innis,M., Gelfand,G.H. and Sninsky,J.J. (eds), *PCR Application: Protocols for Functional Genomics*. Academic Press, London, pp. 73–94.

II

Kaplinski L, Scheler O, Parkel S, Palta P, Toome K, Kurg A, Remm M. (2010)
Detection of tmRNA molecules on microarrays
at low temperatures using helper oligonucleotides.
BMC Biotechnol. 10:34

Kaplinski *et al. BMC Biotechnology* 2010, **10**:34
http://www.biomedcentral.com/1472-6750/10/34

BMC Biotechnology

RESEARCH ARTICLE Open Access

Detection of tmRNA molecules on microarrays at low temperatures using helper oligonucleotides

Lauris Kaplinski*[1,2], Ott Scheler[2,3], Sven Parkel[2,3], Priit Palta[1,2], Kadri Toome[3], Ants Kurg[2,3] and Maido Remm[1,2]

Abstract

Background: The hybridization of synthetic *Streptococcus pneumoniae* tmRNA on a detection microarray is slow at 34°C resulting in low signal intensities.

Results: We demonstrate that adding specific DNA helper oligonucleotides (chaperones) to the hybridization buffer increases the signal strength at a given temperature and thus makes the specific detection of *Streptococcus pneumoniae* tmRNA more sensitive. No loss of specificity was observed at low temperatures compared to hybridization at 46°C. The effect of the chaperones can be explained by disruption of the strong secondary and tertiary structure of the target RNA by the selective hybridization of helper molecules. The amplification of the hybridization signal strength by chaperones is not necessarily local; we observed increased signal intensities in both local and distant regions of the target molecule.

Conclusions: The sensitivity of the detection of tmRNA at low temperature can be increased by chaperone oligonucleotides. Due to the complexity of RNA secondary and tertiary structures the effect of any individual chaperone is currently not predictable.

Background

Over the last decade microarrays have quickly found applications in microbial diagnostics, for detecting different pathogenic viruses, bacteria and other microbes [1] or for analyzing species composition in environmental and medical samples [2]. Also, many different biosensor technologies based on nucleic acid hybridization have been developed and proposed for quick and cost effective "in-the-field" detection and identification of diseases, pathogens or contaminants [3,4].

The most common target molecule for diagnostic and phylogenetic studies is 16S rRNA (or corresponding gene). It was used in the 1970s [5] and continues to be the most widely-used marker for discriminating bacterial species [2,6]. The advantages of ribosomal small subunit RNA are its presence in all species in high copy numbers and the different evolutionary rates of different regions of 16S rRNA, making various taxonomic studies possible [7,8]. Nevertheless, alternative marker molecules [9-12] have to be considered in case 16S rRNA is not suitable for precise detection and distinguishing between closely related species [13].

One interesting novel marker that has shown great potential in molecular diagnostics is the tmRNA transcript of bacterial *ssrA gene*. In living cells, tmRNA is present in relatively high copy numbers (around 1000 molecules per cell) [14,15] and is responsible for assisting ribosomes during translation when protein synthesis stalls. tmRNA molecules contain regions of species-specific sequence heterogeneity and can therefore be successfully used as markers for bacterial diagnostics [16,17].

Nucleic acid hybridizations in microbiology and molecular diagnostics have been performed at various temperatures ranging from 4°C and RT to around 40°C or even higher (50°C and above). It is suggested in previous studies, that the hybridization of complex target molecules is hindered below 42°C, leading e.g. to low signal intensities and bad probe specificity [18]. Low temperature hybridization is of great interest for emerging technologies, such as membrane biosensors, where the denaturation of membranes and proteins have to be avoided and "laboratory-on-chip" and embedded solutions, where maintaining different compartments with varying temperature can be complicated and costly. Modern oligonucleotide

* Correspondence: lauris@ebc.ee
[1] Department of Bioinformatics, Institute of Molecular and Cell Biology, University of Tartu, Riia 23, Tartu, 51010, Estonia
Full list of author information is available at the end of the article

design tools allow the hybridization affinity and specificity of local regions to be estimated quite precisely at different temperatures. One of our main goals was to develop a hybridization method that would be suitable for use below 37°C.

Several difficulties can arise in the detection of bacterial RNA by hybridization. Target RNA degradation has to be prevented and nonspecific hybridizations with wrong targets avoided. The latter is rather difficult on highly conserved RNA molecules [19]. Strong secondary structures can block the hybridization sites inside the molecule and thus prevent hybridization almost completely or retard it significantly [20,21]. The secondary structure of RNA is much stronger than that of the corresponding DNA [22] and the detection of RNA is more difficult [23]. It is suggested that secondary structure may be the main reason why hybridization-based detection fails at room temperature [18,21] and it has to be disrupted, or (in the case of synthetic molecules) its formation has to be minimized, to gain access to the target regions of the RNA molecule [21,24]. The latter is especially crucial in the case of rRNA and tmRNA molecules as they both fold into complex secondary and tertiary structures.

For certain applications it is also important to be able to estimate the relative or absolute abundances of different bacterial strains or species quantitatively. Quantification of hybridization poses additional challenges, especially if the process is too slow to reach equilibrium.

Several approaches to improve the efficiency of hybridization have been described. The hybridization temperature can be increased. It is predicted that while about 70% of a 70-mer cDNA molecule is inaccessible at 42°C, only 30% of DNA and 50% of RNA remains inaccessible at 65°C [20]. Some authors have suggested that a higher temperature increases hybridization specificity, but other authors have found no such effect [23]. Designing probes for specific exposed areas of the molecule also increases the hybridization efficiency [20]. Measuring or predicting the effect of secondary structure is difficult [25,26], especially as the parts of molecule that do not form double-helical stems can themselves be blocked by higher-level structures [20]. Cleaving target molecules to smaller fragments is one widely-used option; it can expose most hybridization sites that are normally blocked [20,25].

Alternatively, specific helper oligonucleotides (chaperones) can be added to the hybridization solution to increase hybridization efficiency [18,23,27]. These molecules bind to target molecules and block specific or nonspecific intramolecular interactions that cause secondary structure formation. Chaperones are specific to certain target molecules and they also increase the specificity of hybridization. Chaperones can also be marked with fluorophores or other detectable markers, solving the problem of detecting hybridized intact RNA [23,27]. It is reported that chaperones immediately side-to-side with hybridization probes are most effective in increasing the effectiveness of hybridization at low temperature [28]. This has been explained by the prevention of hairpin structure formation and by the effect of base stacking between capture probe and chaperone [27].

In this study we evaluated the effect of short DNA helper oligonucleotides (chaperones) on the hybridization of synthetic *Streptococcus pneumoniae* tmRNA molecules to DNA microarray probes. The practical objective was to find an improved protocol for detecting bacterial species by tmRNA hybridization. The theoretical objective was to elucidate the effect of the complex structure of a longer RNA target molecule on hybridization kinetics. In addition, a practical use for chaperones as a interesting novel tool in secondary structure analysis was demonstrated.

Results

Weak hybridization signals at low temperature

To determine the signal intensity and specificity over a range of temperatures, we performed hybridization experiments with synthetic *S. pneumoniae* tmRNA at temperatures ranging from 34°C to 72°C with 4°C steps. At temperatures below 42°C the relative signal intensities were lower than expected from the theoretical melting curves (Figure 1). On average, the decrease in relative signal intensity was more apparent for longer probes, but still clearly present even for probes only 9-10 nucleotides long.

The specificity of hybridization at 34°C was determined by analyzing signals from 21 probes on the same microar-

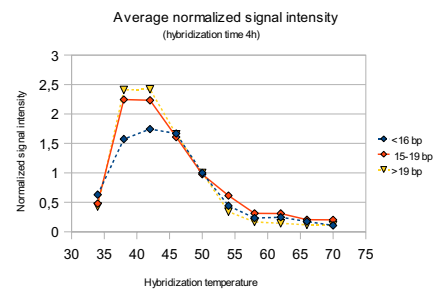


Figure 1 The average signal intensities over the full range of hybridization temperatures. The values are normalized by the signal intensity at 50°C to eliminate differences between the absolute signal values. <16 bp - the average signal intensity of probes shorter than 16 nucleotides. 15-19 bp - the average signal intensity of probes from 15 and 19 nucleotides long. >19 bp - the average signal intensity of probes at least 20 nucleotides long.

ray, designed to the tmRNA sequences of other bacteria. We excluded 4 nonspecific probes, that had detectable false signals at 46°C, possibly due to errors in probe design process. The signal intensities of the remaining nonspecific probes were at least 200 times lower than the average intensity of the specific probes at and below 46°C.

To determine whether the weak signals at low temperatures were caused by slow hybridization or by a shift in equilibrium towards the secondary structure of the target molecules, a series of experiments were performed at 34°C by varying the hybridization time from 2 h to 12 h. The signal intensities increased with time but did not reach a plateau even after 12 h at 34°C. This indicated that hybridization did not reach equilibrium and the low signal intensities were probably caused by a slow rate.

Chaperones increase the hybridization intensity at low temperatures

To determine whether the hybridization intensity at low temperatures can be increased by disrupting the tmRNA secondary structure, we added a mixture of six chaperone oligonucleotides in equal concentrations to the tmRNA solution before hybridization. Three different total chaperone concentrations were tested: 10, 100 and 1000 times the molar concentration of tmRNA. In all cases we recorded significant changes in the intensities of individual signals, which increased 2-3 times on average, although signals were strongly suppressed in several parts of the tmRNA molecule. We observed the highest increase of signal intensity with a relative concentration of chaperones to tmRNA = 100:1. At a relative concentration of 1000:1, the average signal strength was lower than at the 100:1 concentration. For all subsequent experiments we used a chaperone:tmRNA ratio of 100:1.

As expected, the signal intensities increased most markedly for probes that overlapped no chaperone. Most signals of overlapping probes were suppressed by chaperones, although few were higher.

To determine the effect of chaperones on probe specificity, we compared the signals of the *S. pneumoniae* probes with the signals of probes designed for other bacterial species. If the 4 probes, that had detectable false signals at 46°C, were removed, the difference between specific and nonspecific signals was more than 300-fold.

Chaperones increase signal intensities in distantly located regions of the target molecule

To determine which regions of tmRNA were affected by the presence of all six chaperones, we arranged the signals by the probe midpoint position on tmRNA. Three clearly outstanding regions of increased signal strength were apparent around nucleotide positions 100, 150 and 240 of tmRNA (Figure 2). These regions fall outside the chaperone hybridization areas on tmRNA.

By hybridizing the tmRNA in the presence of individual chaperones and arranging the signals by probe midpoint position, we determined the regions most strongly affected by individual chaperones (Figure 3). Chaperone F amplified the signal intensities most strongly in the region (230-240) close to the chaperone hybridizing site (247-260), while chaperones A and E amplified the signals of distant regions. Chaperone A, binding to region 32-46, amplified signals 50 bp towards the 3' end of the tmRNA and also at the 3' end of the molecule. Chaperone E, binding to region 187-201, amplified signals 50 bp towards the 5' end. All chaperones strongly suppressed the signals of overlapped probes, except chaperone E.

We also calculated the probability of hybridization of each nucleotide in tmRNA using sFold and compared it with our hybridization diagrams. There was no significant correlation between the regions of greatly amplified signal strength and the regions of high probability of nucleotide pairing.

Discussion

Ideally, hybridization can be viewed as an equilibrium between free and bound target molecules. The relative abundances of molecules in both states are determined by the difference between rates of duplex formation and dissociation. These rates in turn are determined by the free energy changes in the corresponding processes. Signal intensity is directly proportional to the number of molecules hybridized and is thus determined by the difference in free energy change between hybridization and dissociation. Equilibrium is usually preferred for hybridization because the signal strength is greatest, since the maximum numbers of target molecules are bound to the probes. Also, as hybridization rates may differ among molecules, only the equilibrium state guarantees that actual signal intensities correlate with the concentrations and hybridization affinities of the target molecules. If equilibrium is not reached, a rapidly but weakly hybridizing target may give a stronger signal than a slowly but strongly hybridizing one, so its relative concentration is overestimated.

At low temperature hybridization experiments, two important factors can influence the signal intensities. Both of them are strongly influenced by secondary structures of the target molecule. First, the equilibrium can change and the probe cannot compete with the secondary structure any more [21,26]. Second, the concentration of accessible target configurations is lower at low temperature due to the increased stability of tightly packed secondary structures. The hybridization rate is proportional to the concentration of accessible target molecules and is therefore much slower at low temperatures [26].

If the actual rate of hybridization is low, we may not reach equilibrium during the experiment. In that case the

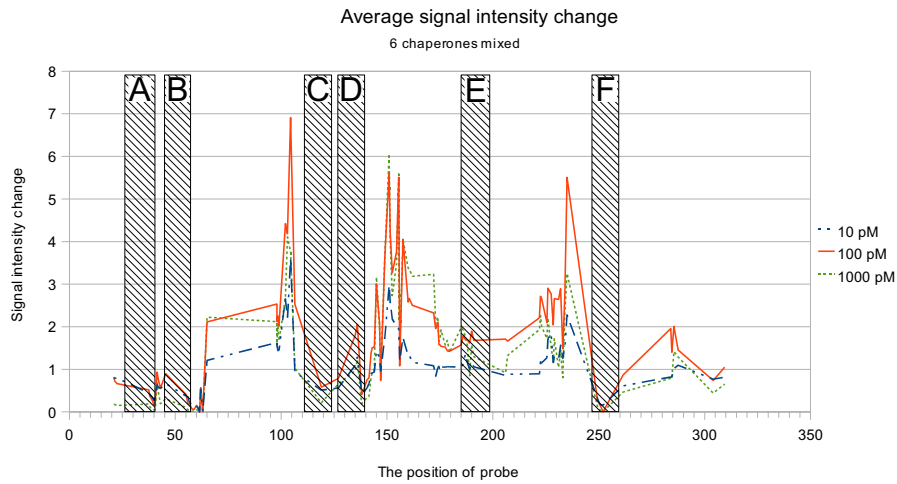


Figure 2 The average relative signal intensity of all probes at different chaperone concentrations. The signal intensities are arranged by the probe midpoint position on tmRNA. The hybridization regions of all chaperones are marked by shaded rectangles.

signals are initially low but increase over time. As we detected such behavior in our experiment, we concluded that the low signal intensities at low temperature are at least partially caused by slow hybridization.

By hybridizing with the target molecule, chaperones block the formation of at least some secondary or tertiary structure variants and thus increase the rate of hybridization at low temperatures [21,24]. Although it is shown, that designing chaperone specifically to the the region that forms intramolecular bonds increases hybridization signals significantly [28], this is not always possible.

We were unable to establish any correlation between the pairing probability of individual nucleotides in tmRNA, calculated with sFold, and the relative amplification of the signals in the presence of chaperones. This can be explained by the effect of tmRNA tertiary structure and the non-equilibrium state. The hybridization probability calculated by sFold only takes direct intramolecular pairings between nucleotides into account; it does not consider the blocking of potential hybridization sites by the globular structure of the molecule. Also, the probabilities of different conformations are calculated by absolute free energy levels, not taking into account the kinetics of secondary structure formation. Some conformations with low free energy may form very slowly, simply because they have to cross an unfavorable intermediate state.

When probe hybridization is blocked by the formation of a hairpin-like structure, the chaperone for the immediate neighborhood should work best because it does not allow the hairpin to form. It has been demonstrated that designing chaperones for the immediate neighborhood should work best if the accessibility of the region is hindered by the formation of secondary structure elements such as hairpins [28].

We were able to see such an effect with chaperone F, where the hybridization profiles of all probes suggest the presence of a hairpin. The chaperone binds to region 248-261 and the signals in the immediately preceding region (230-245) are strongly amplified in its presence. Thus, one can infer that without treatment, these regions are probably hybridized, forming a hairpin. Indeed, the ΔG plot generated by mFold suggests that there is a local energetic minimum between those regions. Nevertheless, it is important to notice that according to mFold this hairpin was not present in the most stable molecule conformation, as determined by the global energy minimum. Also, many other places with similar local energy profiles did not show a similar hybridization pattern. Thus, a hairpin, even if energetically favorable, may often not be the prevailing secondary structure pattern.

A high-level (tertiary) structure of RNA may also form by intramolecular hybridization between distant parts of a longer sequence. Thus the sites that affect the accessibility of a certain part of the molecule may be spatially separated from it. In that case, designing a chaperone for the immediate proximity of the probe may not work,

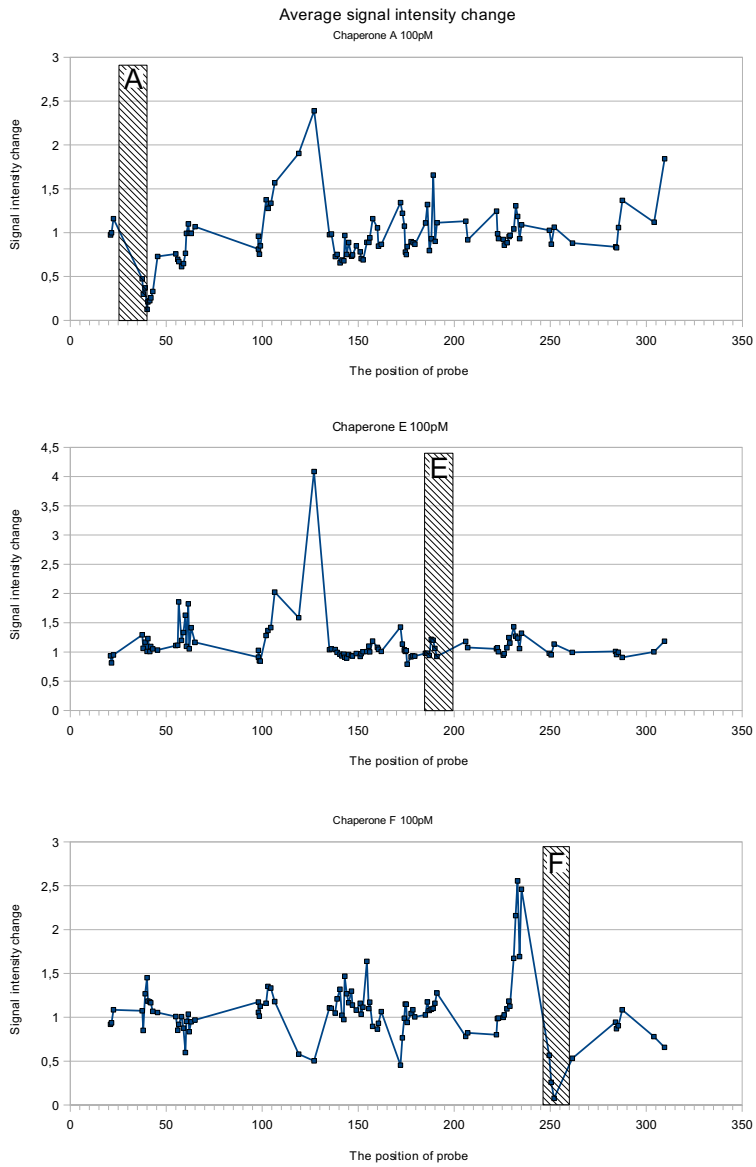


Figure 3 The relative signal intensity change in the presence of different chaperones. The signal intensities are arranged by the probe midpoint position on tmRNA. The shaded area is the hybridization region of the chaperone on tmRNA.

because it either cannot hybridize because of blockage by the three-dimensional structure, or fails to make the neighboring site accessible because it is blocked by some other region. In that case, the best results should be obtained by a chaperone hybridizing specifically to the tertiary structure-forming region, so the tertiary structure cannot form and the conformation of the molecule is loosened. Such chaperones can potentially increase the probe hybridization rate in many regions of the molecule.

In the current experiments, the effect of tertiary structure is suggested by the fact that a single chaperone was able to amplify signals in different regions of the tmRNA molecule (chaperones A and D) and different chaperones amplified signals in the same region (chaperones A, D and E). Such an effect was also suggested by the generic increase of signal intensities of all non-overlapping probes, irrespective of location, in the presence of chaperones.

As expected, chaperones almost completely block the signals of the overlapping probes. Nevertheless, this effect is not absolute, as seen with chaperone E. The reason is still unclear. One possibility is weak chaperone binding, so the higher affinity capture probes outcompete it during hybridization. However, the chaperone has to have a sufficiently high affinity to outcompete the secondary structure of tmRNA successfully. This contradiction may indicate that at least in some cases, the factor limiting hybridization is not secondary but tertiary structure. Although intramolecular double-stranded regions are energetically weak, they fold the tmRNA molecule in such a way that some parts of it are not easily accessible. In such a case, the energetically much stronger target-probe hybridization is slow because the probability that a probe will hit its target area is very low, especially as the probes are immobilized on a surface. Chaperones were applied at a higher temperature at which there was no tertiary structure. Also, the relative concentration of chaperone molecules was much higher, both because they were applied in abundance and because they were free in solution. Thus, the chaperones could hybridize to any region, and if the temperature was lowered, the tertiary structures either did not form or were much weaker.

Conclusion

We thus conclude that while the hybridization of tmRNA can sometimes be relatively slow at low temperatures, it can be significantly increased by using specific helper oligonucleotides (chaperones). The exact effect of certain helper nucleotides on the strength of the signal of certain capture probes depends on many factors, including probably the three-dimensional structure of the target molecule. The effect is not always local, meaning that a chaperone in the immediate proximity of the capture probe may not increase the signal strength, while one at a distant location might. As the structures of denatured nucleotide sequences cannot be precisely predicted at present, experimental verification of chaperones is necessary.

Methods

Bacterial strain and ssrA

The pCR®II-TOPO vector (Invitrogen, Carlsbad, CA, USA) containing *Streptococcus pneumoniae* ATCC 33400 tmRNA encoding *ssrA* was obtained from Dr. Barry Glynn, National University of Ireland, Galway, Ireland. The tmRNA gene was positioned into the vector under the transcriptional control of the T7 promoter sequence.

In vitro RNA transcription

The *ssrA*-containing vector was linearized in 1× buffer R using HindIII restriction endonuclease (both reagents from Fermentas UAB, Vilnius, Lithuania). The reaction was carried out for 60 min at 37°C, followed by 15 min enzyme inhibition at 65°C. *S. pneumoniae* tmRNA was transcribed *in vitro* using 25 ng linearized vector and 20 U T7 RNA polymerase according to the manufacturer's recommendations. Briefly, final 1× reaction buffer contained 2 mM ATP, 2 mM CTP, 2 mM GTP and 1 mM UTP; 30 U RiboLock™ ribonuclease inhibitor was added to prevent possible RNA degradation. Aminoallyl-UTP (aaUTP) was added to 1 mM final concentration, making the final UTP:aaUTP ratio 1:1. A reaction volume of 25 μl was achieved by adding DEPC-treated water. All the reagents were purchased from Fermentas UAB. The transcription reaction was continued for 120 min at 37°C. *In vitro* synthesized RNA was purified using a Nucleotide® RNA CleanUp Kit (Macherey-Nagel GmbH, Düren, Germany) according to the manufacturer's protocol. A final 60 μl of the material eluted was dehydrated in an RVC 2-25 CD rotational vacuum concentrator (Martin Christ GmbH, Osterode am Harz, Germany).

Fluorescent labeling of RNA

Extra amine groups were incorporated into the tmRNAs during *in vitro* transcription by adding aaUTP. The amine-modified RNA was further labeled with the monoreactive fluorescent dye Cyanine™ 3-NHS (Cy3) (Enzo, Farmingdale, NY, USA). Cy3 (50 nmole) was diluted in 2 μl DMSO (Applichem, Darmstadt, Germany) and added to tmRNA diluted in 7 μl 0.1 M Na_2CO_3 (pH 9.0). The mixture of RNA and dye was incubated at room temperature for 60 min and the remaining excess Cy3 label was quenched by adding 3.5 μl 4 M H_2NOH. After the coupling reaction, 35 μl 100 mM sodium acetate was added to neutralize the solution. The labeled RNA was purified with a NucleoSpin® Kit and dehydrated in an RVC 2-25 CD concentrator.

Streptococcus pneumoniae-specific microarray

Capture probes on the custom-made *S. pneumoniae*-specific microarray were designed using SLICSel 1.0 software http://bioinfo.ut.ee/slicsel. SLICSel is a program for designing specific oligonucleotide probes for detecting and identifying microbes. To ensure maximal probe specificity, SLICSel uses the nearest-neighbor thermodynamic model to calculate hybridization affinities for the intended target and non-target sequences. The microarray consisted of three *Streptococcus* family-specific and 94 *S. pneumoniae species*-specific probes covering almost the full length of the 335 nucleotide tmRNA molecule. Probe length varied between 9 and 26 nucleotides (average 16), melting temperature (Tm) between 53°C and 60°C (average 58°C) and binding energies (ΔG) with complementary tmRNA were predicted to be between -17 kcal/mol and -30 kcal/mol (average -23 kcal/mol) at 45°C and in 50 mM salt [Additional file 1]. In addition, 21 probes specific for other bacterial tmRNA sequences were designed to test the specificity of hybridization. The other bacteria included five further members of the *Streptococcus* family (Groups A, B, C, D and G), *Klebsiella pneumoniae* and *Moraxella catarrhalis*.

All the probes designed were tested using Mfold http://mfold.bioinfo.rpi.edu[29,30] to exclude those with potential secondary structures, and MegaBlast http://www.ncbi.nlm.nih.gov/blast/megablast.shtml[31] to eliminate possible cross-hybridization with unwanted targets, including tmRNAs from other species, bacterial DNA/RNA and human genomic DNA or RNA sequences. Three extra control probes with complementary fluorescent targets (spikes) were designed for normalization. Microarray probes with 5'amino modifications and C6 spacers were diluted in 100 mM $Na_2CO_3/NaHCO_3$ (pH 9.0) to 50 μM final concentration and spotted on to SAL-1 Ultra microarray slides in Asper Biotech Ltd., Tartu, Estonia. Each slide contained four datapoints because two identical subgrids were spotted with duplicate spots.

Chaperone design

tmRNA molecules fold into complex structures of pseudoknots, tRNA-like regions and mRNA-like regions. A set of helper oligonucleotides ("chaperones") was designed with SLICSel to reduce the difficulty of hybridizing certain inaccessible regions of tmRNA. Six different chaperones were designed to bind to predicted secondary structure regions in *S. pneumoniae* tmRNA and prevent those intramolecular interactions (Figure 4). The complete set of chaperones is shown in Table 1. All the microarray probes, *spike-s* and chaperones used in the current work were ordered from Metabion, Mariensried, Germany.

Microarray experiment

One pmol of labeled RNA was resuspended in 80 μl microarray hybridization buffer (6× SSC; 0.5% SDS and 5× Denhardt's solution) together with *spike*-s (0.25 nM each). The hybridization mixture was heat-denaturated at 95°C for 5 min and snap-cooled on ice. The RNA hybridization melting curve was obtained by incubating for 4 h at temperatures ranging from 34°C to 70°C. Helper oligonucleotide experiments were conducted at 34°C with 10, 100 or 1000 pmol of chaperones added (making the ratios of tmRNA:chaperone ratios in solution = 1:10; 1:100; 1:1000); an equal amount of RNAse-free water (Macherey-Nagel) was added to the controls without chaperones. The effect of each helper oligonucleotide, individually and in a mixture of all six chaperones, was investigated. To determine the effect of hybridization time, RNA with chaperones was hybridized on to the microarray slides for 2-12 h. All hybridization experi-

Table 1: Helper oligonucleotides (chaperones) used in the current study and their characteristics.

	Position	Length	Sequence 5'-3'	GC%	Tm	ΔG
ChpA	33-47	15	AGTCGCAAAATATGC	40	53,8	-21,0
ChpB	52-64	13	GTTTACGTCGCCA	53,8	54,2	-20,0
ChpC	116-129	14	CCTGCTGGTTTTTA	42,9	56,9	-19,6
ChpD	131-143	13	CAAATCGGGTCAC	53,8	54	-20,1
ChpE	188-202	15	TAGACAAGGCTTAAT	33,3	54,6	-20,5
ChpF	248-261	14	CCCTCGACACATAA	50	62,5	-21,6

Tm and ΔG are calculated by program SLICSEL.

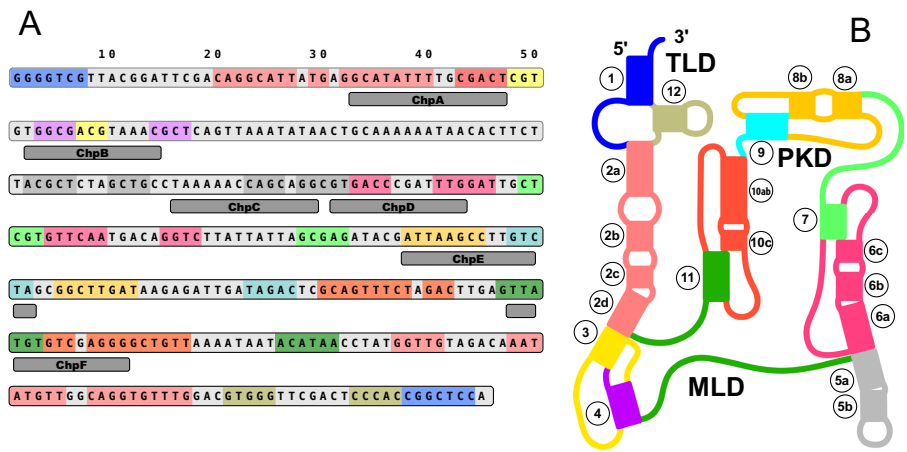


Figure 4 tmRNA structure and chaperone positions. **(A)** *S.pneumoniae* R6 tmRNA sequence [GenBank:NC_003098.1] with predicted helices highlighted in color and chaperone positions marked. Prediction and coloring according to tmRNA website [34]. **(B)** *E.coli* tmRNA secondary structure [35]. Highlight colors of helices are identical to panel A.

ments were conducted using an automated HS-400 hybridization station (Tecan Austria, Grödig, Austria). The complete hybridization protocol at 34°C is shown in Table 2. Prewash solution: 6× saline sodium citrate (SSC), 0.5% sodium dodecyl sulphate (SDS). Wash 1: 2× SSC. 0.03% SDS. Wash 2: 1× SSC. Wash 3: 0.2× SSC. All wash solutions and the prewash solution were warmed to 42°C. After hybridization, the slides were scanned using an Affymetrix 428 scanner (Affymetrix, Santa Clara, CA, USA), λ = 532 nm. Raw signal intensity data were analyzed using Genorama™ BaseCaller software (Asper Biotech). Microarray signals were rescaled to co-analyze data from different arrays by equating the average of the spike-specific signals from each microarray.

Secondary structure prediction

An RNA/DNA folding package mFold was used to determine the tmRNA secondary structure. All degenerate nucleotides in the *S. pneumoniae* tmRNA sequence were substituted with N. All folding parameters were kept at

Table 2: Microarray hybridization protocol used in an automated HS-400 hybridization station.

		Temp. C°	Duration	Repetitions
1	Prewash	85	Wash: 60 s; Soak: 30 s	1
2	Probe injection	34		1
3	Hybridization	34	High agitation	1
4	1. wash	23	Wash: 90 s; Soak: 30 s	3
5	2. wash	23	Wash: 90 s; Soak: 30 s	3
6	3. wash	23	Wash: 90 s; Soak: 30 s	3
7	Slide drying	23	90 s	1

default values. The ten most energetically advantageous structures were calculated for visual analysis of possible common motifs. The full-sequence energy diagram was also calculated. In addition, the probability of each nucleotide of tmRNA being in the hybridized state was calculated using the RNA analysis package sFold http://sfold.wadsworth.org[32,33] with default values.

Additional material

Additional file 1 List of microarray probes. List of all microarray probes (97 *S.pneumoniae* specific and 21 controls) used in hybridization.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
LK conceived and designed the study, analyzed and interpreted the results and wrote the manuscript. OS designed and was responsible for carrying out the microarray experiments and drafted the manuscript. SP participated in microarray experiments and helped to draft the manuscript. PP developed and tested probe selection algorithms and designed probes and chaperones. KT carried out the microarray experiments and performed Genorama analysis. AK conceived of the study, participated in its design and helped to draft the manuscript. MR provided financial and administrative support and participated in the design of the study.
All authors have read and approved the final manuscript.

Acknowledgements
This work was funded by the SLIC-513771 EU grant, the EU European Regional Development Fund, through the Estonian Centre of Excellence in Genomics, and grants SF0180026s09 and SF0180142 from the Estonian Ministry of Education and Research. We would like to thank Dr. Barry Glynn from the National Centre for Biomedical Engineering Science for providing us the the pCR®II-TOPO vector with *ssrA*.

Author Details
[1]Department of Bioinformatics, Institute of Molecular and Cell Biology, University of Tartu, Riia 23, Tartu, 51010, Estonia, [2]Estonian Biocentre, Riia 23b, Tartu, 51010, Estonia and [3]Department of Biotechnology, Institute of Molecular and Cell Biology, University of Tartu, Riia 23, Tartu, 51010, Estonia

Received: 23 September 2009 Accepted: 28 April 2010
Published: 28 April 2010

References
1. Wilson WJ, Strout CL, DeSantis TZ, Stilwell JL, Carrano AV, Andersen GL: **Sequence-specific identification of 18 pathogenic microorganisms using microarray technology.** *Mol Cell Probes* 2002, **16:**119-127.
2. Guschin DY, Mobarry BK, Proudnikov D, Stahl DA, Rittmann BE, Mirzabekov AD: **Oligonucleotide microchips as genosensors for determinative and environmental studies in microbiology.** *Appl Environ Microbiol* 1997, **63:**2397-2402.
3. Baeumner AJ: **Biosensors for environmental pollutants and food contaminants.** *Anal Bioanal Chem* 2003, **377:**434-445.
4. Teles FRR, Fonseca LR: **Trends in DNA biosensors.** *Talanta* 2008, **77:**606-623.
5. Woese CR, Fox GE: **Phylogenetic structure of the prokaryotic domain: the primary kingdoms.** *Proc Natl Acad Sci USA* 1977, **74:**5088-5090.
6. Desantis TZ, Stone CE, Murray SR, Moberg JP, Andersen GL: **Rapid quantification and taxonomic classification of environmental DNA from both prokaryotic and eukaryotic origins using a microarray.** *FEMS Microbiol Lett* 2005, **245:**271-278.
7. Amann R, Fuchs BM, Behrens S: **The identification of microorganisms by fluorescence in situ hybridisation.** *Curr Opin Biotechnol* 2001, **12:**231-236.
8. Woese CR: **Whither microbiology? Phylogenetic trees.** *Curr Biol* 1996, **6:**1060-1063.
9. Anthony RM, Brown TJ, French GL: **Rapid diagnosis of bacteremia by universal amplification of 23S ribosomal DNA followed by hybridization to an oligonucleotide array.** *J Clin Microbiol* 2000, **38:**781-788.
10. Fukushima M, Kakinuma K, Kawaguchi R: **Phylogenetic analysis of Salmonella, Shigella, and Escherichia coli strains on the basis of the gyrB gene sequence.** *J Clin Microbiol* 2002, **40:**2779-2785.
11. Kobayashi M, Urata T, Ikezoe T, Hakoda E, Uemura Y, Sonobe H, Ohtsuki Y, Manabe T, Miyagi S, Miyoshi I: **Simple detection of the 5S ribosomal RNA of Pneumocystis carinii using in situ hybridisation.** *J Clin Pathol* 1996, **49:**712-716.
12. Roth SB, Jalava J, Ruuskanen O, Ruohola A, Nikkari S: **Use of an oligonucleotide array for laboratory diagnosis of bacteria responsible for acute upper respiratory infections.** *J Clin Microbiol* 2004, **42:**4268-4274.
13. Yamamoto S, Harayama S: **PCR amplification and direct sequencing of gyrB genes with universal primers and their application to the detection and taxonomic analysis of Pseudomonas putida strains.** *Appl Environ Microbiol* 1995, **61:**1104-1109.
14. Chauhan AK, Spirion D: **The gene for a small stable RNA (10Sa RNA) of Escherichia coli.** *Mol Microbiol* 1989, **3:**1481-1485.
15. Lee SY, Bailey SC, Apirion D: **Small stable RNAs from Escherichia coli: Evidence for the Existence of New Molecules and for a New Ribonucleoprotein Particle Containing 6S RNA.** *J Bacteriol* 1978, **133:**1015-1023.
16. O'Grady J, Sedano-Balbas S, Maher M, Smith T, Barry T: **Rapid real-time PCR detection of Listeria monocytogenes in enriched food samples based on the ssrA gene, a novel diagnostic target.** *Food Microbiol* 2008, **25:**75-84.
17. Schönhuber W, Le Bourhis G, Tremblay J, Amann R, Kulakauskas S: **Utilization of tmRNA sequences for bacterial identification.** *BMC Microbiol* 2001, **1:**20.
18. Peplies J, Glockner FO, Amann R: **Optimization strategies for DNA microarray-based detection of bacteria with 16S rRNA-targeting oligonucleotide probes.** *Appl Environ Microbiol* 2003, **69:**1397-1407.
19. Fuchs BM, Wallner G, Beisker W, Schwippl I, Ludwid W, Amann R: **Flow Cytometric Analysis of the In Situ Accessibility of Escherichia Coli 16S rRNA for Fluorescently Labeled Oligonucleotide Probes.** *Appl Environ Microbiol* 1998, **64:**4973-4982.
20. Ratushna VG, Weller JW, Gibas CJ: **Secondary structure in the target as a confounding factor in synthetic oligomer microarray design.** *BMC Genomics* 2005, **6:**31.
21. Yilmaz LS, Ökten HE, Noguera DR: **Making All Parts of the 16S rRNA of Escherichia coli Accessible In Situ to Single DNA Oligonucleotides.** *Appl Environ Microbiol* 2006, **72:**733-744.
22. Xia T, SantaLucia J Jr, Burkard ME, Kierzek R, Schroeder SJ, Jiao X, Cox C, Turner DH: **Thermodynamic Parameters for an Expanded Nearest-Neighbor Model for Formation of RNA Duplexes with Watson-Crick Base Pairs.** *Biochemistry* 1998, **37:**14719-14735.
23. Small J, Call DR, Brockman FJ, Straub TM, Chandler DP: **Direct detection of 16S rRNA in soil extracts by using oligonucleotide microarrays.** *Appl Environ Microbiol* 2001, **67:**4708-4716.
24. Chandler DP, Newton GJ, Small JA, Daly DS: **Sequence versus Structure for the Direct Detection of 16S rRNA on Planar Oligonucleotide Microarrays.** *Appl Environ Microbiol* 2003, **69:**2950-2958.
25. Pozhitkov A, Noble PA, Domazet-Loso T, Nolte AW, Sonnenberg R, Staehler P, Beier M, Tautz D: **Tests of rRNA hybridization to microarrays suggest that hybridization characteristics of oligonucleotide probes for species discrimination cannot be predicted.** *Nucleic Acids Res* 2006, **34:**e66.
26. Yilmaz LS, Noguera DR: **Mechanistic Approach to the Problem of Hybridization Efficiency in Fluorescent In Situ Hybridization.** *Appl Environ Microbiol* 2004, **70:**7126-7139.
27. Liao JC, Mastali M, Li Y, Gau V, Suchard MA, Babbitt J, Gornbein J, Landaw EM, McCabe ER, Churchill BM, Haake DA: **Development of an advanced electrochemical DNA biosensor for bacterial pathogen detection.** *J Mol Diagn* 2997, **9:**158-168.
28. Fuchs BM, Glöckner FO, Wulf J, Amann R: **Unlabeled Helper Oligonucleotides Increase the In Situ Accessibility to 16S rRNA of Fluorescently Labeled Oligonucleotide Probes.** *Appl Environ Microbiol* 2000, **66:**3603-3607.

29. Mathews DH, Sabina J, Zuker M, Turner DH: **Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure.** *J Mol Biol* 1999, **288:**911-940.
30. Zuker M: **Mfold web server for nucleic acid folding and hybridization prediction.** *Nucleic Acids Res* 2003, **31:**3406-3415.
31. Zhang Z, Schwartz S, Wagner L, Miller W: **A greedy algorithm for aligning DNA sequences.** *J Comput Biol* 2000, **7:**203-214.
32. Ding Y, Chan CY, Lawrence CE: **RNA secondary structure prediction by centroids in a Boltzmann weighted ensemble.** *Rna* 2005, **11:**1157-1166.
33. Ding Y, Lawrence CE: **A statistical sampling algorithm for RNA secondary structure prediction.** *Nucleic Acids Res* 2003, **31:**7280-301.
34. Williams KP: **The tmRNA Website: invasion by an intron.** *Nucleic Acids Res* 2002, **30:**179-182.
35. Burks J, Zwieb C, Muller F, Wower I, Wower J: **Comparative 3-D modeling of tmRNA.** *BMC Mol Biol* 2005, **6:**14.

doi: 10.1186/1472-6750-10-34
Cite this article as: Kaplinski *et al.*, Detection of tmRNA molecules on microarrays at low temperatures using helper oligonucleotides *BMC Biotechnology* 2010, **10**:34

III

Scheler O, Kaplinski L, Glynn B, Palta P, Parkel S,
Toome K, Maher M, Barry T, Remm M, Kurg A (2011)
Detection of NASBA amplified bacterial tmRNA molecules
on SLICSel designed microarray probes.
BMC Biotechnology 11:17

Scheler et al. BMC Biotechnology 2011, 11:17
http://www.biomedcentral.com/1472-6750/11/17

BMC Biotechnology

METHODOLOGY ARTICLE Open Access

Detection of NASBA amplified bacterial tmRNA molecules on SLICSel designed microarray probes

Ott Scheler[1,2*], Lauris Kaplinski[2,3], Barry Glynn[4], Priit Palta[2,3], Sven Parkel[1,2], Kadri Toome[1], Majella Maher[4], Thomas Barry[4], Maido Remm[2,3], Ants Kurg[1,2]

Abstract

Background: We present a comprehensive technological solution for bacterial diagnostics using tmRNA as a marker molecule. A robust probe design algorithm for microbial detection microarray is implemented. The probes were evaluated for specificity and, combined with NASBA (Nucleic Acid Sequence Based Amplification) amplification, for sensitivity.

Results: We developed a new web-based program SLICSel for the design of hybridization probes, based on nearest-neighbor thermodynamic modeling. A SLICSel minimum binding energy difference criterion of 4 kcal/mol was sufficient to design of *Streptococcus pneumoniae* tmRNA specific microarray probes. With lower binding energy difference criteria, additional hybridization specificity tests on the microarray were needed to eliminate non-specific probes. Using SLICSel designed microarray probes and NASBA we were able to detect *S. pneumoniae* tmRNA from a series of total RNA dilutions equivalent to the RNA content of 0.1-10 CFU.

Conclusions: The described technological solution and both its separate components SLICSel and NASBA-microarray technology independently are applicative for many different areas of microbial diagnostics.

Background

The *ssrA* gene which encodes the tmRNA molecule has been identified in all known bacterial phyla [1,2]. The term tmRNA describes the dual "transfer" and "messenger" properties of this RNA molecule. In bacteria, the function of the tmRNA molecules is to release ribosomes that have become stalled during protein synthesis and to tag incomplete and unnecessary peptides for proteolysis. A typical tmRNA is between 300-400 nucleotides in size and is present in cells in relatively high copy number around 1000 copies per cell [3]. tmRNA molecules contain both conserved as well as variable regions between different species; complementary 3' and 5' ends fold together into a tRNA like structure that permits the entry to the ribosome when needed. Proteolysis-coding mRNA part and structural domains usually make up for the rest of the molecule. All those characteristics make the tmRNA transcript (and its *ssrA* gene) a suitable tool as a target marker molecule for phylogenetical analysis and species identification in microbial diagnostics. Over the last 10 years tmRNA and its corresponding gene have been used for species identification in several methods including fluorescence in situ hybridization (FISH) detection of specific bacteria [4], real-time PCR [5] and real-time NASBA [6] analysis of food and dairy contaminants and pathogen detection using biosensors [7]. Combining the capabilities of tmRNA for species identification with DNA microarray technology offers the potential to investigate samples simultaneously for large numbers of different target tmRNA molecules. DNA microarrays have found several practical applications in microbial diagnostics such as composition analysis and species identification of different environmental and medical samples as well as in microbial diversity investigation [8-10]. Depending on the experiment setup and specific probe design, precise detection of one specific microbe [11] or more complex analysis of microbial taxa can be performed [12]. The design of microarray probes for the detection of bacterial RNA poses unique challenges, because certain RNA/DNA or RNA/RNA mismatches have almost as strong binding affinity as matches [13]. The nearest-neighbor

* Correspondence: ott.scheler@gmail.com
[1]Dept. of Biotechnology, Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia
Full list of author information is available at the end of the article

thermodynamic modeling (NN) approach should therefore be used to calculate the hybridization affinities (ΔG) of probes [14-16]. The hybridization on microarray surface is more complex then hybridization in solution and the NN model should include surface and positional parameters for more accurate modeling [17,18]. Although there are many recent studies of surface hybridization thermodynamics [19], the exact hybridization properties of microarray probes cannot be precisely modelled and experimental verification is still needed [20,21]. A common feature of many microarray analysis protocols is that the nucleic acid sequences of interest are amplified and labeled prior to the hybridization experiment. Hybridization protocols may involve labeled cDNA [22], cRNA [23] or (RT-)PCR products [24]. RNA molecules can also be amplified by Nucleic Acid Sequence Based Amplification (NASBA) [25]. Although not as common as RT-PCR, NASBA is less prone to genomic DNA contamination and therefore more suitable for applications where the testing of microbial viability is important [26]. Several methods have recently been published that describe different NASBA product labeling methods for the purpose of microarray hybridization. These methods include the dendrimer-based system NAIMA [27], biotin-streptavidin binding assisted labeling [28] and aminoreactive dye coupling to aminoallyl-UTP (aa-UTP) molecules in NASBA products [29]. In this report we present a complete technological solution for detection of low amounts of bacterial tmRNA molecules. We describe a new software program, SLICSel, for designing specific oligonucleotide probes for microbial diagnostics using nearest-neighbor thermodynamic modeling and evaluate SLICSel by testing the specificity of the designed tmRNA specific probes. Finally we demonstrate the sensitivity of these probes using a molecular diagnostics method that combines tmRNA amplification by NASBA with microarray-based detection [29]. Using this approach we were able to specifically detect *S.pneumoniae* tmRNA in the amount that corresponds to a single bacterium or less in the presence of 4000-fold excess of other bacterial tmRNA.

Methods

SLICSel program for probe design

The nearest-neighbor thermodynamic (NN) modeling of probe hybridization strength with target (specific hybridization) and control (nonspecific hybridization) nucleotide sequences at exact annealing temperature is used as design criterion of the SLICSel program. The previously published empirical formula was used to adjust the calculated thermodynamic values to the actual annealing temperature and salt concentration [15]. No surface and positional effects were added to the model to keep it universal and not bound to specific technology. We also expect that NN parameters on surface, although slightly different, are in correlation with the ones in solution [19].

Bacterial strains

Streptococcus pneumoniae ATCC 33400 (*S.pneumoniae*), *Streptococcus pyogenes* ATCC 12344 (*S.pyogenes*), *Klebsiella pneumoniae* ATCC 13883 (*K.pneumoniae*), *Moraxella catarrhalis* ATCC 25238 (*M.catarrhalis*) were obtained from DSMZ (Braunschweig, Germany); *Streptococcus agalactiae* (*S.agalactiae*) and Group C/G streptococcus (GrC/G) from University College Hospital (Galway, Ireland). Bacterial strains were grown in Brain Heart Infusion Broth (Oxoid, Hampshire, UK). Total RNA extraction and CFU counting is further described in the Additional file 1.

Microarray design

We used the *S.pneumoniae* tmRNA molecule as the main specific target molecule, while tmRNAs from other bacteria were used as non-specific controls. The custom made microarray for SLICSel validation experiments contained 97 probes covering the whole *S.pneumoniae* tmRNA sequence. For NASBA-microarray experiment, the 25 best performing probes were selected and additional control probes specific to *S.pyogenes*, *S.agalactiae*, *M.catarrhalis* and *K.pneumoniae* (three for each) were also added. The precise probe list and microarray manufacturing have been described previously [30] and customization for the current article is described in the Additional file 1.

In vitro tmRNA synthesis for validation experiment

For *in vitro* transcription of tmRNA *ssrA* genes of *S. pneumoniae*, *S.agalactiae*, *S.pyogenes*, Group C/G streptococcus, *M.catarrhalis* and *K.pneumoniae* were inserted in the pCR® II-TOPO vector (Invitrogen, Carlsbad, CA, USA) under the transcriptional control of either T7 or SP6 promoter sequence. tmRNA molecules were transcribed from vector as described previously [30] with minor alterations. The complete protocol is available in the Additional file 1.

NASBA amplification experiment

A series of experiments were performed to determine the detection capability of NASBA in combination with microarray hybridization. A NASBA primer pair (see the Additional file 1) was designed to amplify a 307 nucleotide tmRNA product using *S.pneumoniae* total RNA as a template. The T7 promoter was added to the forward primer in order to generate a sense strand of the RNA molecule. Three different amounts of *S. pneumoniae* total RNA were added to the NASBA reactions: 1 pg,

100 fg and 10 fg, corresponding to 10, 1 and 0.1 CFU, respectively. An equal volume of NASBA water (included in EasyQ kit) was added to control experiment without any *S. pneumoniae* total RNA. NASBA reactions were performed with NucliSENS EasyQ Basic kit v2 (bioMerieux bv, Boxtel, NL) according to manufacturer's instructions but with addition of aminoallyl-UTP (aa-UTP) as described previously [29]. Final concentration of aa-UTP (Epicentre, Madison, WI, USA) used in the reaction was 1 mM. EasyQ kit was used for 96 NASBA amplifications instead of the original 48 by halving all of the manufacturer suggested reagent volumes. In experiments with background RNA 10 pg of *S.pyogenes, S. agalactiae, M.cattarhalis* and *K.pneumoniae* total RNA were added, making the RNA excess ratios of each control to target RNA 10:1, 100:1 and 1000:1, respectively. Following amplification, tmRNA was purified using a NucleoSpin® RNA CleanUp Kit and vacuum dried using RVC 2-25 CD rotational vacuum concentrator (Martin Christ GmbH, Osterode am Harz, Germany).

Labeling of aa-UTP modified RNA and microarray hybridization

Extra amine groups of aa-UTP modified tmRNA molecules were labeled with the monoreactive fluorescent dye Cyanine™ 3-NHS (Cy3) (Enzo, Farmingdale, NY, USA) as described previously [30]. For the SLICSel validation experiments, 300 ng of *in vitro* synthesized target or control RNA was hybridized onto microarray. In NASBA experiments all of the amplified material was used in the subsequent microarray hybridization. In both cases vacuum dried RNA was resuspended in 80 µl of hybridization buffer and hybridized for 4 hours on the microarray in an automated HS-400 hybridization station (Tecan Austria, Grödig, Austria) at 55 C°. Complete hybridization protocol and reagents are shown in the Additional file 1. After hybridization, the slides were scanned using an Affymetrix 428 scanner (Affymetrix, Santa Clara, CA, USA), λ = 532 nm. Raw signal intensity data was analyzed using Genorama™ BaseCaller software (Asper Biotech, Estonia).

Results

Probe design software

SLICSel was used to design hybridization probes for all bacterial species in the experiment. It uses a brute-force algorithm that finds all theoretically acceptable probe sequences. All designed probes are guaranteed to have at least specified minimum difference ($\Delta\Delta G_{control}$) between the binding energies (ΔG) of specific and non-specific hybridization and at most specified maximum binding energy difference ($\Delta\Delta G_{target}$) between the binding energies of the hybridization with different target sequences. The algorithm also accepts degenerate nucleotides in sequences; in which situation the worst-case variant is used (strongest binding for control set and weakest binding for target set). The program uses well-established thermodynamic models of hybridization in solution, as the more complex surface effects are still under active study and are also dependant on the microarray technology used. The program code can be easily extended to take account of more specific models, if needed. The tables for both DNA-DNA and DNA-RNA nearest-neighbor hybridization thermodynamics are included with the program. It is also possible to use a custom table of thermodynamic parameters, necessary if very specific experimental conditions are used. SLICSel is available from web interface at http://bioinfo.ut.ee/slicsel/

SLICSel validation

A series of hybridization experiments were conducted to validate the SLICSel program by testing the specificity of the SLICSel designed oligonucleotide probes and their suitability for the use in development of diagnostical technology. In total 97 oligonucleotide probes were designed complementary to the different regions of *S. pneumoniae*'s tmRNA (the main target molecule). Control tmRNA molecules were from five other bacteria: *S. pyogenes, S.agalactiae*, GrC/G streptococcus, *K.pneumoniae* and *M.catarrhalis*. All tmRNA sequences were synthesized i*n vitro* and then hybridized individually to the panel of *S.pneumoniae* tmRNA specific probes on microarray. Figure 1 shows the scatter plot of relative signal intensities of control tmRNA hybridizations onto microarray probes according to their binding energy difference $\Delta\Delta G$ between target and control RNA. From a total of 463 hybridization events only 20 (~4.3%) gave relative signal intensities higher than preset 10% false positive signal threshold condition. For the remaining 443 hybridizations (95.7%) the control signals remained under the threshold level. As shown in the Figure 1, designing probes with higher binding energy difference ($\Delta\Delta G$) decreases the possibility of a false positive signal. For example, choosing the probes with the minimum ΔG difference of 4 kcal/mol was sufficient to avoid all the false-positive bindings over the threshold while in the case of ΔG difference 2 kcal/mol 6 signals remained over the 10% signal threshold (~1.5% of hybridizations). The average hybridization signal intensities of target and control tmRNAs (all five together and individually) are shown on a bar chart and complementary table in Figure 2. Nearly fivefold increase of the probe specificity was achieved with $\Delta\Delta G$ condition 4 kcal/mol as the average false-positive control tmRNA signal intensity dropped from 2.46% to 0.55%. All of the average false-positive hybridization signals of individual tmRNAs were lower with higher minimum $\Delta\Delta G$ criteria. In general,

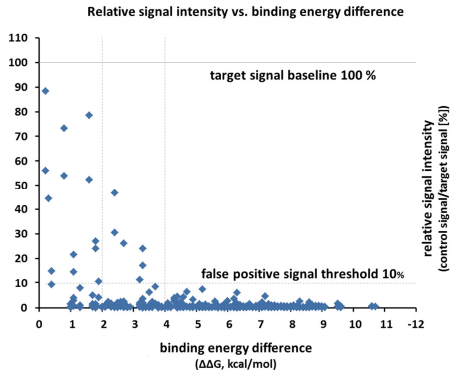


Figure 1 Non-specific microarray signal intensities from hybridization experiments with target and five control tmRNAs. Non-specific tmRNA signal intensities are divided by the corresponding probe-specific signal intensity. Target hybridization signal intensity is given on a y-axis as a 100% signal baseline. A maximum false-positive signal threshold is shown as horizontal 10% dotted line. Microarray signals are distributed along x-axis according to the calculated binding energy difference between the specific and non-specific binding (ΔG_{target} - $\Delta G_{control}$ ($\Delta\Delta G$) 0.2...10.7 kcal/mol). Dotted vertical lines separate the probes with binding energy difference smaller then 2 and 4 kcal/mol, respectively.

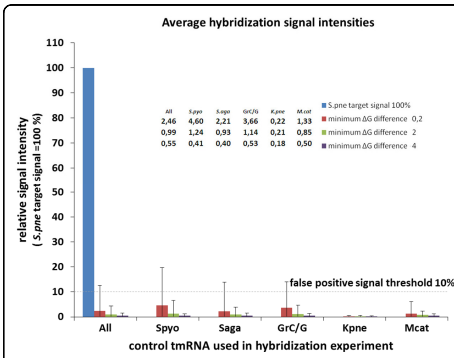


Figure 2 Average microarray signal intensities with target *Streptococcus pneumoniae* and five different control tmRNAs. The probe-specific target tmRNA hybridization signal average is shown as a 100% bar. Control signal averages of five different tmRNAs (*Streptococcus pyogenes (S.pyo), Streptococcus agalactiae (S.aga)*, GrC/G streptococcus (GrC/G), *Klebsiella pneumoniae (K.pne)* and *Moraxella catarrhalis (M.cat))* are given as a percentage of the target signal. Three different average bars for control tmRNAs represent the average hybridization signal intensities with probes' minimum ΔG differences 0.2; 2 and 4 kcal/mol compared to the hybridization with target molecule ($\Delta\Delta G = 0$). Error bars show SD of control signal averages. All of the average signal values of the control tmRNA hybridization reactions are shown on the table added onto the graph.

control tmRNAs from bacteria belonging to the Streptococcus genus showed stronger than/or near average false-positive hybridization signals while signals of more distant *K.pneumoniae* and *M.catarrhalis* remained under the overall average. *K.pneumoniae* tmRNA produced lowest average false-positive signals in all three different minimum ΔΔG conditions and had no signals over the 10% threshold. All of the false-positive signals greater than 10% were contributed by 10 single microarray probes. After removal of those problematic probes the average hybridization signal intensities were under 1% for all the different control tmRNAs.

NASBA-microarray technology

To test the SLICel designed probes for their potential use in microbial diagnostics; a new microarray was designed that consisted of the 25 best performing probes out of 97 according to their specificity and the sensitivity in the validation experiments. For control purposes oligonucleotide probes specific to *S.pyogenes, S.agalactiae*, *K.pneumoniae* and *M.catarrhalis* were also added to the microarray. tmRNA molecules of *S. pneumoniae* were amplified from three different total RNA dilutions (equaling to 0.1, 1 and 10 CFU, respectively) and labeled for microarray hybridization. Microarray signals were obtained with all three total RNA dilutions in all of the three parallel experiments including the 10 fg of total RNA sample equivalent to 0.1 CFU. According to the total RNA input into the NASBA reaction, microarray signals increased correspondingly with 0.1 CFU being the lowest and 10 CFU the highest in three replicate experiments (figure 3). Hybridization experiments with NASBA amplified negative control solution provided no significant signals over the background level on microarray. NASBA control experiments with excess amounts of total RNA mix from 4 control species (*S.pyogenes, S.agalactiae*, *K.pneumoniae* and *M.catarrhalis*) were performed to verify the specificity of the NASBA-microarray based detection method. 10 pg of total RNA from each of the control species were added, making the background RNA ratio to target RNA 4 × 10:1, 4 × 100:1 and 4 × 1000:1, respectively. Addition of control total RNA-s to NASBA reaction did not cause any changes to the microarray signal intensities; all of the *S.pneumoniae* target dilutions were amplified and detected on the microarray while the negative control remained blank. The capability of the described NASBA-microarray method to detect tmRNA from low amounts of bacteria was also confirmed experimentally when the total RNA was prepared from dilutions of *S.pneumoniae* cultures (0.1 to10

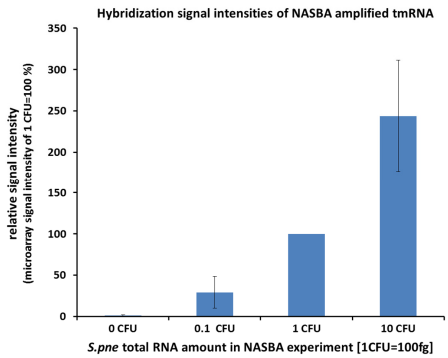


Figure 3 Microarray signal intensities of NASBA amplified tmRNA from *Streptococcus pneumoniae* total RNA dilutions. The microarray signal intensity of NASBA amplified 1 CFU total RNA was set as a 100% in all three parallel experiments. Rest of the RNA dilution hybridization signals from 0 CFU (equal volume of NASBA water as negative control), 0.1 and 10 CFU represent their relation to 1 CFU signal as a percentage. Error bars show ± 1 SD of signal averages over three parallel experiments. 1 CFU equivalent of total RNA stands for 100 fg of RNA from *S.pneumoniae*.

CFU) instead of using total RNA dilutions, making the experiment setup closer to real-world diagnostic situations where only small amounts of target bacteria may be present.

Discussion

We selected tmRNA as a marker molecule for technological tool development in bacterial diagnostics because they are present in all bacteria [1,2] in high copy number and they contain both conserved as well as highly divergent regions [3]. Presence of intact RNA molecules can additionally indicate the viability of the bacterial population in the analyte solution [26]. These characteristics make tmRNA a suitable marker molecule in microbial diagnostics. Although the aforementioned properties also apply to16S rRNA (and its corresponding gene), possibly the best known and most used marker in diagnostic and phylogeny studies, the need for investigation of novel alternative marker molecules like tmRNA remains as 16S rRNA often cannot be used to detect and distinguish closely related species [4,31]. For microarray-based detection technologies, the signal strength is determined by the number of target molecules hybridized to probes, i.e. by the equilibrium point of hybridization, and can thus be theoretically predicted using the nearest-neighbor thermodynamic model. The same model, incorporating mismatches, can also be used to predict the signal strength of nonspecific hybridizations - i.e. false-positive signals. In our approach the goal was not to design probes with maximum affinity, but instead maximize the difference of affinity between specific and nonspecific hybridization at annealing temperature. The microarray hybridization experiments conducted with tmRNA specific probes gave information about the concept of designing probes using NN thermodynamic modeling in SLICSel and whether the tested probes are suitable for further species detection and identification. In general the hybridization experiments with *in vitro* synthesized target and control tmRNA molecules proved that SLICSel designed probes are highly capable of specific bacterial identification. By implementing stringent binding energy difference criteria during probe design SLICSel can minimize the possibility of designing probes that would result in false-positive signals. In our validation experiment the hybridization binding energy difference $\Delta\Delta G$ 4 kcal/mol between control and target tmRNA was sufficient to eliminate all the false-positive control signals over the needed threshold level (Figure 1). We achieved an almost fivefold increase in average probe specificity by using stringent $\Delta\Delta G$ criteria 4 kcal/mol (Figure 2). Although, the efficiency of average SLICSel designed probe is high, there is no 100% guaranteed approach for the *in silico* oligonucleotide probe design for hybridization based experiments with surface-immobilized probes. Additional probe specificity evaluation *in vitro* and low quality probe removal still remain as necessary steps in any microarray experiment [20]. In our case the removal of 10 probes was needed to assure that hybridization signals with control tmRNAs remain safely under the determined 10% threshold level. We designed a new microarray incorporating only the optimum *S.pneumoniae* specific probe sequences for the detection of labeled tmRNA products amplified using NASBA. A key characteristic of the NASBA-microarray technology, especially in microbial diagnostics, is that the detection and the identification of the correct target can be optimized at two different points in the experimental protocol. The selection of oligonucleotide primer set determines the specificity of the NASBA amplification phase while a second level of specificity is provided by the SLICSel designed immobilized microarray probes. Specific amplification of a single RNA molecule or wider selection of various RNAs in case of multiplex-NASBA is possible. Certain rules have been described for the NASBA primer pair design [32], but as no convenient software has yet been developed it remains somewhat a trial-and-error approach. In our case the primer set was designed according to the aforementioned rules to amplify a near full length tmRNA molecule from *S.pneumoniae*. We included additional control probes specific to *S.pyogenes*, *S.agalactiae*, *K.pneumoniae* and *M.catarrhalist* in the microarray to determine the specificity of NASBA amplification step conducted in the presence of a non-*S.*

pneumoniae total RNA background. The composition of capture probes on the microarray depends on the overall goal of the experiment. In our case the objective was to specifically detect tmRNA molecules from *S.pneumoniae* total RNA and test the sensitivity of the method previously described by us [29]. Our intention was to investigate whether the method is capable of detecting 1 CFU by using tmRNA as a target molecule. Previous works have shown that detection of 1 CFU by using NASBA amplification of rRNA [33] or tmRNA [6] is possible. The addition of highly parallel microarray based detection to this amplification technology could represent a significant advance in microbial diagnostics; particularly in situations where high number of different bacterial species may be present (such as environmental samples) or in clinical settings where it is necessary to identify one particular infection causing species from a large panel of potential pathogens. We successfully detected and identified *S.pneumoniae* tmRNA molecules from all three different dilutions of total RNA used in experiments (Figure 3). Our experiments proved that 0.1 CFU equivalent total RNA was sufficient to produce strong reproducible hybridization signals on our microarray. Addition of background total RNAs to the NASBA reaction mix provided no signals on control probes on microarray, confirming the high specificity of NASBA-microarray technology and also its components: NASBA primers and microarray probes. In case of the specific tmRNA detection from 0.1 CFU equivalent of *S.pneumoniae* total RNA, the amount of non-specific RNA exceeded the target 4000 times. The described high level of achieved specificity and sensitivity demonstrates the potential and suitability of NASBA-microarray technology for the purpose of pathogen detection in microbial diagnostics or more complex analysis of microbial taxa in environment.

Conclusions

We have presented a novel technological procedure for bacterial diagnostics and microbial analysis. The nearest-neighbor thermodynamics based SLICSel tool is not exclusive for tmRNA and microarray probe design, but can be used for any other hybridization based technology where DNA or RNA oligonucleotide probe design is necessary. The combination of NASBA amplification technology with microarray based fluorescently labeled RNA detection enabled us to detect tmRNA molecules from as low as 0.1 to 10 CFU of *S.pneumoniae* total RNA. Using the described approach different patient samples, food products or any analyte solution can be tested and screened in a highly parallel approach for several live pathogens or contaminants. SLICSel and NASBA-microarray technology can be used separately for different areas of microbial diagnostics including environmental monitoring, bio threat detection, industrial process monitoring and clinical microbiology.

Additional material

Additional file 1: Methods supplementary file. Additional file describing thoroughly all of the necessary data and reagents needed for the methods section

Acknowledgements
This work was funded by the SLIC-513771 EU grant and by targeted financing from Estonian Government SF0180027s10. This work was also funded by grant SF0180026s09 from the Estonian Ministry of Education and Research and by the EU through the European Regional Development Fund through the Estonian Centre of Excellence in Genomics. Authors would like to thank Indrek Valvas and Asper Biotech for microarray manufacturing

Author details
[1]Dept. of Biotechnology, Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia. [2]Estonian Biocentre, Tartu, Estonia. [3]Dept. of Bioinformatics, Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia. [4]Molecular Diagnostics Research Group, National Centre for Biomedical Engineering Science, National University of Ireland, Galway, Ireland.

Authors' contributions
OS conducted NASBA and microarray experiments, performed microarray analysis and drafted the manuscript. LK designed SLICSel and microarray probes, helped with microarray data analysis and drafted the manuscript. BG carried out the microbiological experiments and RNA extraction and helped to draft and review the manuscript. PP helped designing SLICSel and the microarray probes, helped in data analysis and drafted the manuscript. SP participated in NASBA and microarray experiments and data analysis. KT participated in NASBA and microarray experiments and helped in manuscript review. MM and TB conceived of the study, participated in its design and coordinated microbiological experiments. MR conceived of the study and participated in its design, conducted SLICSel design, helped to draft and review the manuscript. AK conceived of the study, participated in its design, coordinated NASBA and microarray experiments and helped to draft and review the manuscript. All authors read and approved the final manuscript.

Received: 12 November 2010 Accepted: 28 February 2011
Published: 28 February 2011

References

1. Keiler KC, Waller PR, Sauer RT: **Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA.** *Science (New York, N.Y.)* 1996, **271**:990-3.
2. Keiler KC, Shapiro L, Williams KP: **tmRNAs that encode proteolysis-inducing tags are found in all known bacterial genomes: A two-piece tmRNA functions in Caulobacter.** *Proceedings of the National Academy of Sciences of the United States of America* 2000, **97**:7778-83.
3. Muto A, Ushida C, Himeno H: **A bacterial RNA that functions as both a tRNA and an mRNA.** *Trends in biochemical sciences* 1998, **23**:25-9.
4. Schönhuber W, Bourhis GL, Tremblay J, Amann R, Kulakauskas S: **Utilization of tmRNA sequences for bacterial identification.** *BMC Microbiology* 2001.
5. O'Grady J, Sedano-Balbás S, Maher M, Smith T, Barry T: **Rapid real-time PCR detection of Listeria monocytogenes in enriched food samples based on the ssrA gene, a novel diagnostic target.** *Food microbiology* 2008, **25**:75-84.
6. O'Grady J, Lacey K, Glynn B, *et al*: **tmRNA–a novel high-copy-number RNA diagnostic target–its application for Staphylococcus aureus detection using real-time NASBA.** *FEMS microbiology letters* 2009, **301**:218-23.
7. Glynn B, Lacey K, Grady JO, *et al*: **Reusable surface plasmon resonance assay for the specific detection of streptococcus pneumoniae tmRNA.** *JOURNAL OF RAPID METHODS AND AUTOMATION IN MICROBIOLOGY* 2008, **16**:210-221.

8. Bryant P, Venter D, Robins-Browne R, Curtis N: **Chips with everything: DNA microarrays in infectious diseases.** *The Lancet infectious diseases* 2004, **4**:100-11.
9. Loy A, Bodrossy L: **Highly parallel microbial diagnostics using oligonucleotide microarrays.** *Clinica chimica acta; international journal of clinical chemistry* 2006, **363**:106-19.
10. Miller MB, Tang Y-W: **Basic concepts of microarrays and potential applications in clinical microbiology.** *Clinical microbiology reviews* 2009, **22**:611-33.
11. Ye RW, Wang T, Bedzyk L, Croker KM: **Applications of DNA microarrays in microbial systems.** *Journal of microbiological methods* 2001, **47**:257-72.
12. Wilson KH, Wilson WJ, Radosevich JL, *et al*: **High-Density Microarray of Small-Subunit Ribosomal DNA Probes.** *Applied and environmental microbiology* 2002, **68**:2535-2541.
13. Sugimoto N, Nakano M, Nakano S: **Thermodynamics-structure relationship of single mismatches in RNA/DNA duplexes.** *Biochemistry* 2000, **39**:11270-81.
14. Nugen SR, Leonard B, Baeumner AJ: **Application of a unique server-based oligonucleotide probe selection tool toward a novel biosensor for the detection of Streptococcus pyogenes.** *Biosensors & bioelectronics* 2007, **22**:2442-8.
15. SantaLucia J, Hicks D: **The thermodynamics of DNA structural motifs.** *Annual review of biophysics and biomolecular structure* 2004, **33**:415-40.
16. Pozhitkov AE, Tautz D, Noble PA: **Oligonucleotide microarrays: widely applied–poorly understood.** *Briefings in functional genomics & proteomics* 2007, **6**:141-8.
17. Zhang L, Wu C, Carta R, Zhao H: **Free energy of DNA duplex formation on short oligonucleotide microarrays.** *Nucleic acids research* 2007, **35**:e18.
18. Naiser T, Kayser J, Mai T, Michel W, Ott A: **Position dependent mismatch discrimination on DNA microarrays - experiments and model.** *BMC bioinformatics* 2008, **9**:509.
19. Hooyberghs J, Van Hummelen P, Carlon E: **The effects of mismatches on hybridization in DNA microarrays: determination of nearest neighbor parameters.** *Nucleic acids research* 2009, **37**:e53.
20. Pozhitkov A, Noble Pa, Domazet-Loso T, *et al*: **Tests of rRNA hybridization to microarrays suggest that hybridization characteristics of oligonucleotide probes for species discrimination cannot be predicted.** *Nucleic acids research* 2006, **34**:e66.
21. Li S, Pozhitkov A, Brouwer M: **A competitive hybridization model predicts probe signal intensity on high density DNA microarrays.** *Nucleic acids research* 2008, **36**:6585-91.
22. Hegde P, Qi R, Abernathy K, *et al*: **A Concise Guide to cDNA Microarray Analysis.** *Biotechniques* 2000, **29**:548-562.
23. 't Hoen P, de Kort F, van Ommen G, den Dunnen J: **Fluorescent labelling of cRNA for microarray applications.** *Nucleic Acids Research* 2003, **31**:20e-20.
24. Wilson WJ, Strout CL, DeSantis TZ, *et al*: **Sequence-specific identification of 18 pathogenic microorganisms using microarray technology.** *Molecular and cellular probes* 2002, **16**:119-27.
25. Compton J: **Nucleic acid sequence-based amplification.** *Nature* 1991, **350**:91-2.
26. Keer JT, Birch L: **Molecular methods for the assessment of bacterial viability.** *Journal of Microbiological Methods* 2003, **53**:175-183.
27. Morisset D, Dobnik D, Hamels S, Zel J, Gruden K: **NAIMA: target amplification strategy allowing quantitative on-chip detection of GMOs.** *Nucleic acids research* 2008, **36**:e118.
28. Mader A, Riehle U, Brandstetter T, *et al*: **Microarray-based amplification and detection of RNA by nucleic acid sequence based amplification.** *Analytical and bioanalytical chemistry* 2010, **397**:3533-41.
29. Scheler O, Glynn B, Parkel S, *et al*: **Fluorescent labeling of NASBA amplified tmRNA molecules for microarray applications.** *BMC biotechnology* 2009, **9**:45.
30. Kaplinski L, Scheler O, Parkel S, *et al*: **Detection of tmRNA molecules on microarrays at low temperatures using helper oligonucleotides.** *BMC biotechnology* 2010, **10**:34.
31. Yamamoto S, Harayama S: **PCR Amplification and Direct Sequencing of gyrB Genes with Universal Primers and Their Application to the Detection and Taxonomic Analysis of Pseudomonas putida Strains.** *Applied and environmental microbiology* 1995, **61**:3768.
32. Deiman B, van Aarle P, Sillekens P: **Characteristics and applications of nucleic acid sequence-based amplification (NASBA).** *Molecular biotechnology* 2002, **20**:163-79.
33. Loens K, Beck T, Goossens H, *et al*: **Development of conventional and real-time NASBA for the detection of Legionella species in respiratory specimens.** *Journal of microbiological methods* 2006, **67**:408-15.

doi:10.1186/1472-6750-11-17
Cite this article as: Scheler *et al.*: Detection of NASBA amplified bacterial tmRNA molecules on SLICSel designed microarray probes. *BMC Biotechnology* 2011 **11**:17.

IV

Kaplinski L, Möls M, Möls T, Remm M.
Predicting failure rate of Multiplex PCR
(submitted)

BIOINFORMATICS

Vol. 00 no. 00 2005
Pages 1–7

Predicting failure rate of multiplex PCR

Lauris Kaplinski [1,2,*], Märt Möls [2], Tõnu Möls [1], Reidar Andreson [1,2] and Maido Remm [1,2]

[1]Department of Bioinformatics, Institute of Molecular and Cellular Biology, University of Tartu, 23 Riia str., Tartu 51010, Estonia
[2]Estonian Biocentre, 23 Riia str., Tartu 51010, Estonia

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Associate Editor: XXXXXXX

ABSTRACT

Motivation: The success rate of multiplex PCR is lower compared to single-plex PCR, but the causes of this phenomenon are not well understood.

Results: We used the experimental data from 8,000 4-plex PCR experiments to determine which primer-pair parameters influence the failure rate of multiplex PCR. We discovered that the main determinant of the multiplex PCR failure rate is the quality of each single primer pair. Interactions between primer pairs have only a minor effect on the 4-plex PCR failure rate. In other words, if primers work well in single-plex PCR, they are also likely to work in multiplex PCR. Low quality primer pairs not only worked poorly in multiplex PCR but also increased the failure rate of other primer pairs in the same multiplex group.

Conclusion: The quality of a single primer pair can be predicted accurately from the primer sequence. Our results indicate that it is possible to choose successful multiplex PCR primers at the design phase for each primer pair separately by using stricter primer selection criteria to avoid the contamination of groups with low-quality pairs. This will facilitate primer design for multiplex PCR in future.

Contact: lauris.kaplinski@ut.ee

Supplementary information: Supplementary Data are available at *Bioinformatics* online.

1 INTRODUCTION

Multiplexing is an important technique for optimizing the cost of PCR experiments. By performing many PCRs in a single tube one can save time, chemicals and the amount of source DNA needed, and provide internal control of the reaction. Multiplex PCR is successfully used for example in bridging gaps in shotgun sequencing (Sorokin *et al*., 1996; Tettelin *et al*., 1999), detecting mutations and indels (Edwards and Gibbs, 1994; Shen *et al*., 2010), detecting pathogens (Elnifro *et al*., 2000; Strommenger *et al*., 2003; Gardner *et al*., 2009; Shi *et al*., 2012), quantitative analysis of GM foods (Rudi *et al*., 2003; Jinxia *et al*., 2011) and genotyping (Edwards and Gibbs, 1994).

The higher the level of multiplexing (i.e. the number of different PCRs performed in a single tube), the lower the cost per PCR. With standard PCR equipment and protocols multiplexing levels from two to low tens are routinely used (Edwards and Gibbs, 1994; Nicodème and Steyaert, 1997; Tettelin *et al*., 1999; Strommenger *et al*., 2003; Syvänen, 2005). With specific protocols and methods, multiplexing levels of hundreds are achievable (Fredriksson *et al*., 2007).

It is well known that in a multiplex environment, when the multiplex level increases, the average success rate of PCR decreases (Sorokin *et al*., 1996; Rudi *et al*., 2003; Rachlin *et al*., 2005). Although the causes of this phenomenon are not known, several factors have been associated with the degradation of the PCR success rate.

Suboptimal GC contents of products and primers and the difference between the GC contents of primers lower the PCR success rate (Mallona *et al*., 2011). Different trinucleotide patterns at the 3' end of the primer influence the success of PCR (Onodera and Melcher, 2004).

The correct prediction of primer melting temperature is crucial for the success of PCR. Too low an annealing temperature can cause the formation of false products and too high an annealing temperature decreases the PCR yield. The optimal annealing temperature depends on the melting temperatures of both primers and products (Rychlik *et al*., 1990). The nearest-neighbor thermodynamic model is currently the most precise method for calculating oligonucleotide melting temperatures (Chavali *et al*., 2005).

The presence of structural polymorphisms (SNPs, indels and CNVs) or non-unique regions in the genome influences the success rate of PCR (Piriyapongsa *et al*., 2009).

Nonspecific hybridization of primers to the template DNA can cause mispriming and the generation of false products even for unique primers if they contain a 3' non-unique region of sufficient length (Elnifro *et al*., 2000; Rudi *et al*., 2003; Miura *et al*., 2005; Kalendar *et al*., 2011). It is known that the number of nonspecific binding sites of primers affects the failure rate of the single-plex PCR reaction (Andreson *et al*., 2008). It is probably even more important for multiplex PCR because of the potential mispriming between the primers for different PCR primer pairs (Shen *et al*., 2010). It is also important when estimating the number of the nonspecific hybridization sites to consider not only perfect hybridization, but also hybridization with mismatches. While those bindings are weaker (having higher hybridization ΔG values), they are more abundant. Hybridization is a stochastic process and even weak bindings have a certain probability of starting polymerase reactions, wasting primers, enzyme and nucleotides and sometimes creating false products (Yamada *et al*., 2006). Because the (reverse

*to whom correspondence should be addressed.

complement of) primer sequence is copied one-to-one into the PCR product, false products, once created in the first cycles, will be amplified at a similar rate to the correct products (Kanagawa, 2003).

Even low complementarity between primers can cause the formation of primer dimers and the failure of PCR. This effect is stronger at higher multiplexing levels owing to the increasing number of possible combinations of primers (Brownie, 1997; Nicodème and Steyaert, 1997; Syvänen, 2005).

The factors that influence the rate of PCR, such as the differences between product lengths and primer melting temperatures, can degrade the success rate of multiplex PCR. As different primer pairs compete in the multiplex environment the reaction rates in the same group should be similar (Edwards and Gibbs, 1994; Syvänen, 2005). To achieve uniform PCR speed in different reactions and reduce the possibility of nonspecific hybridization, universal primers can be used. But even in that case the initial hybridization of target-specific primers is crucial for the success of PCR (Rudi *et al*., 2003; Shi *et al*., 2012).

It has been demonstrated that adjusting the reaction conditions such as the concentrations of primers, enzyme and Mg^{++} and annealing and extension times and temperatures reduces the failure rate of multiplex PCR (Edwards and Gibbs, 1994; Henegariu *et al*., 1997; Strommenger *et al*., 2003).

The most widely used primer design program, also routinely used to design multiplex PCR primers, is Primer3 (Untergasser *et al*., 2012). Although it allows the user to exert extensive control over the design parameters it lacks certain features, most notably the integrated check for nonspecific bindings. Several tools and models have been developed that complement primer design by estimating PCR efficiency under given experimental conditions (Mallona *et al*., 2011) and taking into account the potential nonspecific bindings to genomic DNA (Andreson *et al*., 2008; Qu *et al*., 2009; Andreson *et al*., 2006). A precise thermodynamic hybridization model can be used to predict the mispriming probabilities accurately on the basis of the longest 3' non-unique subsequences and thus eliminate primers with high failure probability (Miura *et al*., 2005).

Several programs have also been developed for the design of multiplex PCR either by applying more stringent primer design criteria (Kalendar *et al*., 2011; Qu *et al*., 2009) or by composing optimized groups of existing primers (Nicodème and Steyaert, 1997; Kaplinski *et al*., 2007) or integrating primer design with compatibility testing (Shen *et al*., 2010; Rachlin *et al*., 2005; Gardner *et al*., 2009). Finding the optimal multiplexing solution is an NP-complete problem (Rachlin *et al*., 2005), but heuristic algorithms can calculate near-optimal solutions (Rachlin *et al*., 2005; Nicodème and Steyaert, 1997). The achievable multiplexing level is critically dependent on the average probability of any two primer pairs working together (Rachlin *et al*., 2005). The algorithms for designing multiplex PCR try to minimize potential primer dimers, melting temperature differences and nonspecific hybridization with products (Kalendar *et al*., 2011; Qu *et al*., 2009; Yamada *et al*., 2006).

Thus in most cases, experimental validation of multiplex PCR sets is still necessary (Elnifro *et al*., 2000; Carter *et al*., 2010).

In the present paper we assess which primer-specific factors are related to the failure of multiplex PCR in three separate cases:

1. Which primer pair properties are related to the failure of a given PCR?
2. Which primer pair properties are related to the failure of other PCRs in the same tube?
3. Which primer pair combination properties are related to the failure of either one of these pairs?

We present experimentally verified factors for predicting the failure rate of multiplex PCR in four primer-pair groups. The first model (class A) and the second model (class B) can be used as additional selection criterion in the primer design process if multiplexing PCR is desired. We found no significant effect of primer pair combinations on the PCR failure rate.

2 MATERIALS AND METHODS

2.1 Selection of primers

A set of 300 PCR primer pairs was designed for randomly-chosen regions of the human genome using Primer3 software (Untergasser *et al*., 2012). These primer pairs were validated in a separate study (Andreson *et al*., 2008) using 10 independent single-plex PCR runs with each pair. We chose 209 primer pairs that had 100% success rates in the single-plex validation experiment. The primer and product sequences are listed in Supplementary Table S1.

2.2 Screening experiment

For initial evaluation we grouped these 209 primer pairs into 4000 groups: 2000 groups with two primer pairs each and 2000 groups with four each. Grouping was random, except that the minimal difference between product lengths in each group was 50 bp to ensure that the the presence of each individual product could be evaluated by gel electrophoresis.

The PCR conditions were as follows:

15 min pre-incubation at 95 °C, followed by seven touchdown cycles of 20 s at 95 °C, 30 s at 66 °C (decreasing 1 °C per cycle), 30 s at 72 °C; seventeen cycles of 20 s at 95 °C, 30 s at 58 °C, 30 s at 72 °C; sixteen cycles of 20 s at 95 °C, 30 s at 56 °C, 30 s at 72 °C, and final extension at 72 °C for 7 min. DNA was extracted from human blood cells by a modified salting-out method (Miller *et al*., 1988). The enzyme was Smart HotStart Taq (Naxo, Estonia). The reaction was performed in the buffer specified by the enzyme supplier, with 0.25 mM dNTP, 2.5 mM $MgCl_2$, 0.6 pM of primers and 15 ng of genomic DNA in a volume of 10 µl.

The PCR results were evaluated by examining the gel electropherograms by eye. The presence of a product with the correct length was recorded on three-grade scale (0 - no visible product band, 1 weak but distinguishable band, 2 strong product band). In addition, the presence of other product bands and smears (small nonspecific oligonucleotides) was recorded.

2.3 Main dataset

We created 6000 multiplex groups consisting of four primer pairs each, using the same methodology as for screening experiments. PCR experiments were carried out in Asper Biotech Ltd. and Quattromed Ltd., 3000 experiments in each laboratory, using different equipment and personnel. The PCR conditions were precisely the same as for the screening experiments and the results were evaluated with the same method. For the following analysis the results of 4-plex screening experiments (2000 PCR) were added to the main dataset, resulting in total of 8000 random multiplex groups.

2.4 Model types

The models for predicting the success/failure rate in multiplex PCR can be divided into three general classes:

Class A model (primer pair itself): The probability of the success of certain PCR depends on the properties of the same primer pair. I.e. if there are primer pairs A, B, C and D in a multiplex group, we predict the success of pair A depending on the variables describing primer pair A. The predicted

probability of success of primer pair A will be the same regardless of the other primer pairs present in the multiplex group:

$$\text{logit}(P(\text{success}|\text{primer } i; \text{experiment } j)) =$$
$$= \mu + \text{variables describing primer pair} i + \gamma_j + \tau_{labor(j)}$$

where $P(\text{success}|\text{primer } i; \text{experiment } j)$ is the probability of success of primer pair i in the j-th PCR experiment; γ_j is the random (batch) effect of the j-th PCR experiment ($E\gamma_j = 0$) and $\tau_{labor(j)}$ is a random effect of the laboratory in which the j-th experiment was performed ($E\tau_{labor} = 0$).

Class B model (other pairs in group): The probability of the success of certain PCR depends on the parameters of the same and other primer pairs present in the multiplex group. I.e. if there are primer pairs A, B, C and D in multiplex group, we estimate the probability of success of pair A using information available on the primer pairs A, B, C and D. But we always assume a primer in the multiplex group will have a similar effect on the outcome regardless of the other primers present. For example, if one primer in a multiplex group has a negative influence on the outcome of the other PCR experiments, then we assume its influence will always be negative, with the same effect size, regardless of the other primer pairs included.

$$\text{logit}(P(\text{success}|\text{primer } i; \text{experiment } j)) =$$
$$= \mu + \text{variables describing primer pair } i +$$
$$+ \text{variables describing other primer pairs in the } j\text{-th experiment} +$$
$$+ \gamma_j + \tau_{labor(j)}$$

Class C model (interaction between pairs): In this class of models the effect of a primer pair included in the PCR experiment can depend on the other primer-pairs included. For example, a primer-pair B included in the PCR experiment could increase the probability of PCR failure for some primers but decrease it for others. Models in this class can account for possible interaction effects among primer pairs.

$$\text{logit}(P(\text{success}|\text{primer } i; \text{experiment } j)) =$$
$$= \mu + \text{variables describing primer pair } i +$$
$$+ \text{variables describing other primer pairs in the } j\text{-th experiment} +$$
$$+ \text{variables describing pairs of primer pairs in the} j\text{-th experiment} +$$
$$+ \gamma_j + \tau_{labor(j)}$$

2.5 Parameters analyzed

For the subsequent analysis we calculated and used the following variable types:

1. The minimum, average and maximum of GC content og either primer in a pair, the length and GC content of the product of a given pair (Supplementary Table S1)
2. The minimum, average and maximum number of possible nonspecific binding sites for either PCR primer in the human genome (logarithmic, Supplementary Table S2)
3. The minimum, average and maximum thermodynamic stabilities of primer dimer and primer-product dimer between the primers and products of different PCRs

In addition to these variables we evaluated the upper limit of possible predictive power achievable for class A or B models using the most complete model from this particular class (maximum predictive power model). For example, to describe the maximum predictive power achievable by a class A model, we estimated the probability of success of a primer pair on the basis of the available data about the behavior of the same primer pair in PCR experiments, so each pair was ascribed an individual probability of success based on its own historical performance. We could not use the same approach for class C models because the number of possible pairwise interactions was too big.

Table 1. The failure rates of individual primer pairs and multiplex groups, concatenated screening and main experiments. Group failure was defined as the failure of any primer pairs in the given group.

	2-plex	4-plex
Experiments (groups)	2,000	8,000
PCR (primer pairs)	4,000	32,000
PCR failure rate	10.3%	22.6%
Group failure rate	17.9%	45.9%

The number of nonspecific primer binding sites in the human genome was calculated using a modified GenomeTester package (Andreson *et al*., 2006) using region lengths from 8 to 16 nucleotides, with 0-2 mismatches and requiring an exact match over 1-3 nucleotides at the 3' end of primer.

The strengths of primer dimers were calculated using the MultiPLX package (Kaplinski *et al*., 2007).

2.6 The model

The effect of variables describing primer pairs on the success rate of multiplex PCR was modeled by a generalized linear mixed model (mixed effects logistic regression). R package lme4 (R Core Team, 2013) was used for all calculations.

The success (1) or failure (0) of each primer pair in each multiplex group was treated as a single observation. Thus the result of each multiplex experiment gave us four observations, altogether 32, 000 possible dependent observations. To take into account the dependence among observations generated by experiment-specific effects, the experiment identifier was included in the model as a random factor.

The various models were tested using the following schema:

1. We randomly split the dataset (8,000 4-plex experiments, 32,000 observed primer pairs) into two equally-sized sets: training and test set. To evaluate the maximum predictive power all experiments were distributed randomly (so that each primer pair occurred in both sets). To evaluate parametric models, we distributed the observations into groups so that no primer-pair occurred in both sets as the observed pair.
2. All individual primer pairs from each group were used as separate observations.
3. Using the training set, we built a GLM model predicting the probability of success of PCR. Moving from the class A model (primer-specific factors) to class C (interactions), all variables of the previous model were also included in the new one.
4. We applied the model to the test set and calculated the relative improvement of the results if primer pairs or multiplex groups were eliminated on the basisi of a specific cutoff value of success rate prediction.

3 RESULTS

3.1 The results of the screening experiments

As the first step we determined the average failure rates of 2 and 4-plex PCRs in screening experiments. The results are presented in Table 1.

Table 2. Area Under Receiver Operating Characteristic (AUC) values of different models.

Model	AUC
Class A maximum predictive power	0.71
Class B maximum predictive power	0.77
Class A all variables	0.65
Class B all variables	0.73
Class C all variables	0.73

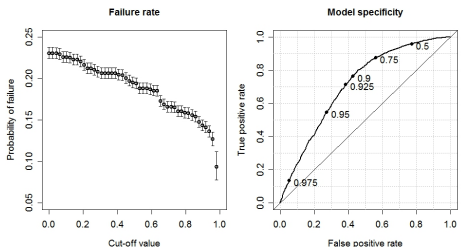


Fig. 1. The effect of applying a class A maximum predictive power model to the test dataset. Left - improvement of average PCR success rate after elimination of all primer pairs with predicted success rate below a certain cut-off value. X-axis applied cut-off value. Y axis estimated probability of the failure of a primer pair with 95% confidence intervals. Right ROC plot of model prediction. X-axis the proportion of remaining failing primer pairs (of all failing pairs) after eliminating those with predictions below the cut-off value. Y-axis the proportion of remaining working primer pairs (of all working pairs). Selected cut-off values from 0.5 to 0.975 are marked on the ROC curve.

Although 2-plex experiments would have simplified the subsequent analysis, we considered the 10% failure rate to be too low for an informative statistical failure model. Thus the following experiments were designed and performed using groups of four primer pairs.

3.2 Maximum predictive power models: The failure of a primer pair in multiplex PCR and its effect on other pairs is consistent among groups.

As a first step we constructed and tested a class A maximum predictive power model predicting the failure rate of individual primer pairs from their ID code. Although this approach cannot be used to predict the success or failure of new primer pairs, the correlation between previous and current failure rates shows how consistently the same primer pairs work across experiments. Lack of correlation would indicate no primer-specific (class A) factors at all and all failures would be caused by external factors.

The class A maximum predictive power model had clear predictive value on the test dataset as indicated by the AUC (Area Under Receiver Operating Characteristic) value of 0.713 (Table 2).

Applying the class A maximum predictive power model, i.e. excluding primers that had previously worked poorly, roughly halves the multiplex PCR failure rate in the test dataset (Fig. 1).

From these results we can infer that the failure of certain primer pairs in the multiplex environment is consistent across experiments. Thus, certain

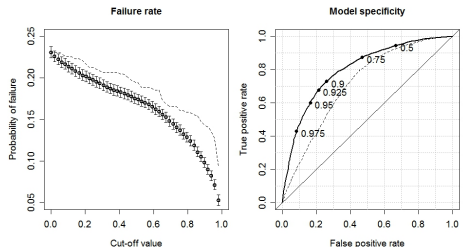


Fig. 2. The effect of applying a class B maximum predictive power model to the test dataset. Left - improvement of average PCR success rate after elimination of all primer pairs with predicted success rate below a certain cut-off value. X-axis applied cut-off value. Y axis estimated probability of the failure of a primer pair with 95% confidence intervals. Dashed line comparable failure rate of class A model. Right ROC plot of model prediction. X-axis the proportion of remaining failing primer pairs (of all failing pairs) after eliminating those with predictions below the cut-off value. Y-axis the proportion of remaining working primer pairs (of all working pairs). Dashed line comparable plot of class A model. Selected cut-off values from 0.5 to 0.975 are marked on the ROC curve.

primer-pair-specific properties can probably be used to predict the failure rate, and further analysis is justified.

Next, we constructed a class B maximum predictive power model predicting the failure rate of individual primer pairs depending on their ID code and the ID codes of the other primer pairs in the group. As with the class A maximum predictive power model, this approach, although not usable for actual prediction, shows whether there are systematic effect of certain primer pairs on the failure rates of other primer pairs in solution.

According to the likelihood ratio test, the class B maximum predictive power model was statistically significantly better than the class A model ($p < 2.2 * 10^{-16}$). The predictive value of the class B model was also clearly better than that of the class A model (Table 2, Fig. 2).

From these results we can see that the success or failure of a single PCR in the multiplex environment depends on both the given primer pair and the other primer pairs in the group. Certain primer pairs either increase or decrease the failure rates of others.

3.3 Class A model (the effect of the primer pair itself): The failure rate of a primer pair in multiplex PCR depends on the number of its nonspecific binding sites.

We constructed and tested a class A parametric model using all previously-calculated numerical properties of primer pairs (GC%, number of nonspecific binding sites in the human genome, etc). Applying this model to the test dataset also demonstrated a clear improvement of PCR quality, although slightly less than by applying the maximum predictive power model (Table 2, Fig. 3). For example, eliminating all primer pairs with predicted failure rate over 5% lowers the actual failure rate by 36% (from 22% to 14%).

To determine the most important variables correlated with the failure of a primer pair in multiplex PCR we used the stepwise AIC algorithm on the class A parametric model.

We tested class A models with the five and ten best variables separately on the test dataset and compared the results with the maximum predictive power model and the all-parameter model. The results are portrayed on Figure 4. The full model descriptions are given in Supplementary Data.

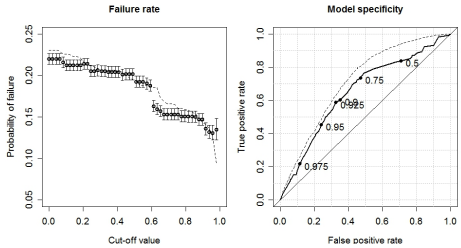


Fig. 3. The effect of applying a class A parametric model with all variables. Left - improvement of the average PCR success rate after elimination of all primer pairs with predicted success rates below a certain cut-off value. X-axis applied cut-off value. Y axis estimated probability of the failure of a primer pair with 95% confidence intervals. Dashed line the comparable failure rate of the class A maximum predictive power model. Right ROC plot of model prediction. X-axis the proportion of remaining failing primer pairs (of all failing pairs) after eliminating those with the prediction below the cut-off value. Y-axis the proportion of remaining working primer pairs (of all working pairs). Selected cut-off values from 0.5 to 0.975 are marked on the ROC curve. Dashed line the ROC curve of the class A maximum predictive power model.

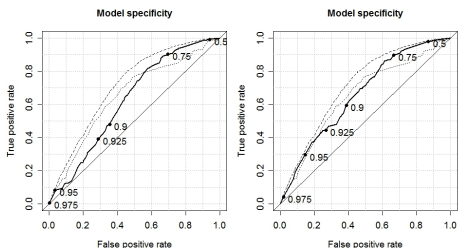


Fig. 4. The effect of applying class A parametric models with five and ten variables. Left - ROC plot of five-parameter model prediction. X-axis the proportion of remaining failing primer pairs (of all failing pairs) after eliminating those with predictions below the cut-off value. Y-axis the proportion of remaining working primer pairs (of all working pairs). Selected cut-off values from 0.5 to 0.975 are marked on the ROC curve. Dashed line the ROC curve of the maximum predictive power model. Dotted line ROC curve of the all parameter model. Right ROC plot of 10-parameter model prediction. Axes and labeling are the same as in the left graph.

3.4 Class B model (the effect of other primer pairs): The failure rate of a primer pair in multiplex PCR depends on the number of nonspecific binding sites of all primer pairs in the group

We constructed a class B model using the minimal, average and maximal values of all calculated primer pair parameters of all other primer pairs in the multiplex group. Applying this model to the test dataset demonstrated clear improvement of PCR quality over the comparable class A model. According to the likelihood ratio test, the class B parametric model with all variables

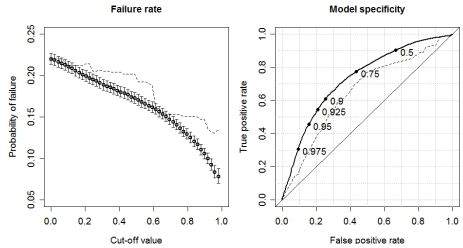


Fig. 5. The effect of applying the class B parametric model with all variables. Left - improvement of average PCR success rate after elimination of all primer pairs with predicted success rate below a certain cut-off value. X-axis applied cutoff value. Y axis estimated probability of the failure of a primer pair with 95% confidence intervals. Dashed line comparable failure rate of class A parametric model. Right ROC plot of model prediction. X-axis the proportion of remaining failing primer pairs (of all failing pairs) after eliminating those with the prediction below the cut-off value. Y-axis the proportion of remaining working primer pairs (of all working pairs). Dashed line comparable plot for application of the class A parametric model. Selected cut-off values from 0.5 to 0.975 are marked on the ROC curve.

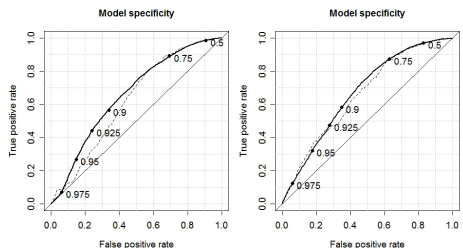


Fig. 6. The effect of applying class B parametric models with five and ten variables. Left - ROC plot of five-parameter model prediction. X-axis the proportion of remaining failing primer pairs (of all failing pairs) after eliminating those with predictions below the cut-off value. Y-axis the proportion of remaining working primer pairs (of all working pairs). Selected cut-off values from 0.5 to 0.975 are marked on the ROC curve. Dashed line comparable ROC curve of five-parameter class A model. Right ROC plot of ten-component model prediction. Axes and labeling are the same as in graph A. Dashed line comparable ROC curve of ten parameter class A model.

is statistically significantly better than the class A parametric model ($p < 2.2 * 10^{-16}$). The class B model was also clearly a better predictor than the class A model (Table 2, Fig. 5).

We evaluated the five and ten component class B models separately and compared those against the corresponding class A models. The results are portrayed in Figure 6. The full model description is given in Supplementary Data.

3.5 Class C model (the effect of interactions): We were unable to detect the effect of primer-primer interactions on the failure of multiplex PCR reliably.

Although the class C parametric model was statistically significantly better than the class B model ($p = 0.0009$), it offered no increased accuracy of prediction on the basis of the results from the test dataset.(Table 2). Thus we could not establish whether or to what extent primer dimers influence the quality of multiplex PCR.

4 DISCUSSION

In the present study we used primer pairs that had 10/10 success rates in independent single-plex PCR experiments. Although the PCR protocols used in the multiplex and single-plex environments were identical, roughly one reaction in four failed in the multiplex PCR.

It has been demonstrated previously that the number of nonspecific genomic binding sites in a eucaryotic genome is correlated with the failure rate of single-plex PCR (Andreson *et al*., 2008). Our first hypothesis was that the same factors that cause a single-plex PCR to fail affect the success of multiplex PCR even more strongly.

The second hypothesis was that too many nonspecific binding sites of certain primer pairs can affect not only the result of this PCR but also the results of other primer pairs in the same group.

The third hypothesis was, that the stability and nature of primer dimers influences the failure rate of multiplex PCR. All dimers compete for specific hybridization sites on the template DNA, lowering the effective primer concentration. Dimers that have 3' ends of both primers in the hybridized state can initiate an unwanted polymerase reaction, depleting primers and nucleotides and generating small false products. We could not demonstrate any primer-dimer effects in our experiments. As the number of possible interactions between primer pairs grows in the proportion to 2^n it is possible that such interactions will be more important in larger groups.

We were able to demonstrate that about half of the failures in 4-plex PCR can be attributed to primer-specific factors (Fig. 1). In other words, certain primer pairs have higher than average failure rates in multiplex PCR groups, no matter what other primer pairs are in those groups. Even if we do not know which factors are causing those to fail, we can use this information to optimize experiments by not multiplexing the primer pairs that have previously shown high failure probabilities.

According to our analysis the most important factors that increase the probability of PCR failure in multiplex PCR are the GC contents of both primers and of the PCR product, and the number of nonspecific binding sites of primers in the genome. We tested many different methods for counting the number of genomic binding sites and we found the best predictor was C16_1_MAX (the biggest number of binding sites of either primer in the pair, in the human genome, with 16 nucleotides from the 3' end hybridized with maximum of one mismatch). Nevertheless one has to bear in mind that these different methods of counting the number of nonspecific hybridization sites give highly correlated results. It is possible, for example, to replace the variable C_16_1_MAX in a class A model with the variable C_15_1_MAX with only minimal loss of predictive power, so the chosen list of factors has a certain arbitrariness. The all-variable model incorporates various different hybridization site counts, possibly simulating the non-linear and sequence-dependent effects of nonspecific hybridization.

It is known that the nearest neighbor thermodynamic model gives a much better estimate of the actual nonspecific hybridization strength than counting nucleotide pairs with mismatches, On the other hand it is much slower to calculate so it was not practical for our models.

Primer-pair-specific factors had the strongest effect on failure of multiplex PCR. Nevertheless we were able to demonstrate that in addition to the primer pair itself, other primer pairs in the multiplex group affect the PCR outcome. In our experiment, three other members of group had a much smaller effect on the outcome of certain PCR than the primer-pair itself. We can only speculate that if more reactions coincide in a single tube, the negative effects of other primer pairs add up and could become the dominant cause of multiplex PCR failure.

It is also worth noticing that other primer pairs in the multiplex group affect the outcome of the PCR reaction in similar way to the observed primer pair. I.e. the more nonspecific binding sites any primer-pair in group has, the lower the probability of success for all primer pairs in the group.

Even if the precise mechanisms of the failure of multiplex PCR remain unknown, we could demonstrate that by applying certain additional primer design criteria, or applying existing criteria more strictly, the success rate of PCR can be increased. This is especially helpful in the primer design phase, where in designing primers for multiplex PCR we can apply additional criteria to eliminate those candidate sequences that will fail with higher probability. If, on the other hand, the list of primers is already fixed, we can still decide how to distribute them into groups. If certain primers will fail with high probability it could be desirable to amplify those sequences in the single-plex environment.

We envision three possible applications of our results.

First, in designing primers for multiplex PCR, it is useful to mask repetitive words in the human genome with stricter criteria than for single-plex PCR so that primer pairs are guaranteed to have few nonspecific hybridization sites in the genome.

Second, after creating a list of candidate primer pairs, the number of nonspecific genomic hybridization sites should be included among the final selection criteria, preferably in the form of a model that estimates the quality of PCR. Also, according to our results, slightly lower primer GC content is preferable for multiplex PCR.

Third, if the primer pairs are already chosen or if only one primer pair is available for a certain region, we can use our failure model to predict both the probability of PCR failure in the multiplex environment and the probability of it disturbing other primer pairs. If either of these is too high, we can opt to amplify this primer pair in a single-plex environment.

ACKNOWLEDGEMENT

The authors would like to thank Eneli Oitmaa and Kai Jõers for supervising the experimental part of this study.

Funding: This work was supported by Enterprise Estonia (Grant EU19730); Estonian Ministry of Education and Research (grant SF0180026s09); EU (European Regional Development Fund through the Estonian Centre of Excellence in Genomics).

REFERENCES

Andreson,R., Reppo,E., Kaplinski,L. and Remm,M. (2006) GENOMEMASKER package for designing unique genomic PCR primers, *BMC bioinformatics*, **7**, 172.

Andreson,R., Möls,T. and Remm,M. (2008) Predicting failure rate of PCR in large genomes, *Nucleic acids research*, **36**, e66.

Brownie,J. (1997) The elimination of primer-dimer accumulation in PCR, *Nucleic Acids Research*, **25**, 3235-3241.

Carter,I.W.J. *et al*. (2010) PCR Methodology. In Schuller,M., Sloots,T.P., James,G.S., Halliday,C.L., Carter,I.W.J. (eds.), *PCR for Clinical Microbiology* Springer, Netherlands, Dordrecht, pp. 11–47.

Chavali,S., Mahajan,A., Tabassum,R., Maiti,R. and Bharadwaj,D. (2005) Oligonucleotide properties determination and primer designing: a critical examination of predictions, *Bioinformatics*, **21**, 3918-3925.

Edwards,M. and Gibbs,R. (1994) Multiplex PCR: advantages, development, and applications, *Genome Research*, **3**, S65-S75.

Elnifro,E.M., Ashshi,A.M., Cooper,R.J. and Klapper,P.E. (2000) Multiplex PCR: optimization and application in diagnostic virology, *Clinical Microbiology Reviews*, **13**, 559-570.

Fredriksson,S., Banèr,J., Dahl,F., Chu,A., Ji,H., Welch,K. and Davis,R.W. (2007) Multiplex amplification of all coding sequences within 10 cancer genes by Gene-Collector, *Nucleic acids research*, **35**, e47.

Gardner,S.N., Hiddessen,A.L., Williams,P.L., Hara,C., Wagner,M.C. and Colston,B.W. (2009) Multiplex primer prediction software for divergent targets, *Nucleic acids research*, **37**, 6291-304.

Henegariu,O., Heerema,N.A., Dlouhy,S.R., Vance,G.H. and Vogt,P.H. (1997) Multiplex PCR: critical parameters and step-by-step protocol, *Biotechniques*, **23**, 504-11.

Jinxia,A., Qingzhang,L., Xuejun,G., Yanbo,Y., Lu,L. and Minghui,Z. (2011) A multiplex nested PCR assay for the simultaneous detection of genetically modified soybean, maize and rice in highly processed products, *Food Control*, **22**, 1617-1623.

Kalendar,R., Lee,D. and Schulman,A.H. (2011) Java web tools for PCR, in silico PCR, and oligonucleotide assembly and analysis, *Genomics*, **98**, 137-44.

Kanagawa,T. (2003) Bias and artifacts in multitemplate polymerase chain reactions (PCR), *Journal of bioscience and bioengineering*, **96**, 317-23.

Kaplinski,L., Andreson,R., Puurand,T. and Remm,M. (2007) MultiPLX: automatic grouping and evaluation of PCR primers, *Bioinformatics*, **21**, 1701-1702.

Mallona,I., Weiss,J. and Egea-Cortines,M. (2011) pcrEfficiency: a Web tool for PCR amplification efficiency prediction, *BMC bioinformatics*, **12**, 404.

Miller,S.A., Dykes,D.D. and Polesky,H.F. (1988) A simple salting out procedure for extracting DNA from human nucleated cells, *Nucleic acids research*, **16**, 1215.

Miura,F., Uematsu,C., Sakaki,Y. and Ito,T. (2005) A novel strategy to design highly specific PCR primers based on the stability and uniqueness of 3-end subsequences, *Bioinformatics (Oxford, England)*, **21**, 4363-70.

Nicodème,P. and Steyaert,J. (1997) Selecting optimal oligonucleotide primers for multiplex PCR, *Proc Int Conf Intell Syst Mol Biol*, **5**, 210-213.

Onodera,K. and Melcher,U. (2004) Selection for 3 end triplets for polymerase chain reaction primers, *Molecular and Cellular Probes*, **18**, 369-372.

Piriyapongsa,J., Ngamphiw,C., Assawamakin,A., Wangkumhang,P., Suwannasri,P., Ruangrit,U., Agavatpanitch,G. and Tongsima,S. (2009) RExPrimer: an integrated primer designing tool increases PCR effectiveness by avoiding 3 SNP-in-primer and mis-priming from structural variation, *BMC genomics*, **10** (Suppl. 3), S4.

Qu,W., Shen,Z., Zhao,D., Yang,Y. and Zhang,C. (2009) MFEprimer: multiple factor evaluation of the specificity of PCR primers, *Bioinformatics (Oxford, England)*, **25**, 276-8.

R Core Team (2013) R: A Language and Environment for Statistical Computing R Foundation for Statistical Computing, Vienna.

Rachlin,J., Ding,C., Cantor,C. and Kasif,S. (2005) Computational tradeoffs in multiplex PCR assay design for SNP genotyping, *BMC genomics*, **6**, 102.

Rachlin,J., Ding,C., Cantor,C. and Kasif,S. (2005) MuPlex: multi-objective multiplex PCR assay design, *Nucleic acids research*, **33**, W544-W547.

Rudi,K., Rud,I. and Holck,A. (2003) A novel multiplex quantitative DNA array based PCR (MQDA-PCR) for quantification of transgenic maize in food and feed, *Nucleic Acids Research*, **31**, e62.

Rychlik,W., Spencer,W. and Rhoads,R. (1990) Optimization of the annealing temperature for DNA amplification in vitro, *Nucleic Acids Research*, **18**, 6409-6412.

Shen,Z., Qu,W., Wang,W., Lu,Y. and Wu,Y. (2010) MPprimer: a program for reliable multiplex PCR primer design, *BMC Bioinformatics*, **11**, 143.

Shi,H., Trinh,Q., Xu,W., Zhai,B., Luo,Y. and Huang,K. (2012) A universal primer multiplex PCR method for typing of toxinogenic Pseudomonas aeruginosa, *Applied Microbiology and Biotechnology*, **95**, 1579-1587.

Sorokin,A., Lapidus,A., Capuano,V., Galleron,N., Pujic,P. and Ehrlich,S.D. (1996) A new approach using multiplex long accurate PCR and yeast artificial chromosomes for bacterial chromosome mapping and sequencing, *Genome Research*, **6**, 448-453.

Strommenger,B., Kettlitz,C., Werner,G. and Witte,W. (2003) Multiplex PCR Assay for Simultaneous Detection of Nine Clinically Relevant Antibiotic Resistance Genes in Staphylococcus aureus, *Journal of Clinical Microbiology*, **41**, 4089-4094.

Syvnen,A. (2005) Toward genome-wide SNP genotyping, *Nature Genetics*, **37**, S5-S10.

Tettelin,H., Radune,D., Kasif,S., Khouri,H. and Salzberg,S. (1999) Optimized multiplex PCR: efficiently closing a whole-genome shotgun sequencing project, *Genomics*, **62**, 500-507.

Untergasser,A., Cutcutache,I., Koressaar,T., Ye,J., Faircloth,B.C., Remm,M. and Rozen,S.G. (2012) Primer3–new capabilities and interfaces, *Nucleic acids research*, **40**, e115.

Yamada,T., Soma,H. and Morishita,S. (2006) PrimerStation: a highly specific multiplex genomic PCR primer design server for the human genome, *Nucleic acids research*, **34**, W665-W669.

CURRICULUM VITAE

Name: Lauris Kaplinski
Date of birth: April 8, 1971, Tartu
Citizenship: Estonian
Address: Department of Bioinformatics, Institute of Molecular and Cell Biology, University of Tartu
Riia str. 23B, 51010, Tartu, Estonia
Phone: +372 737 5001
E-mail: lauris.kaplinski@ut.ee

Education and professional employment:

1977–1989	Tartu I Keskkool
1989–1994	University of Tartu, Faculty of Biology and Geography, Department of Biochemistry
1994–1996	University of Tartu, Department of Philosophy
1998	BSc University of Tartu, Institute of Molecular and Cell Biology
2002–2004	Biodata Ltd. IT specialist
2004–	Estonian Biocentre programmer
2007–	PhD student, Department of Bioinformatics, University of Tartu, Institute of Molecular and Cell Biology

Scientific work:
My main research projects have focused on applying oligonucleotide hybridization model on PCR primer design and microarray hybridization. I have implemented an algorithm for automatic calculating of inter-primer interaction and optimized grouping for multiplex PCR. For microarray hybridization I developed a probe selection algorithm and program, and studied the possibility of breaking the secondary structure of target RNA molecules. I have also analyzed the primer-specific factors related to the failure of multiplex PCR.

List of publications:

1. Koua D, Laht S, Kaplinski L, Stöcklin R, Remm M, Favreau P, Lisacek F. (2013) Position-specific scoring matrix and hidden Markov model complement each other for the prediction of conopeptide superfamilies. Biochim Biophys Acta. 2013;1834(4):717–24
2. Koua D, Brauer A, Laht S, Kaplinski L, Favreau P, Remm M, Lisacek F, Stöcklin R. (2012) ConoDictor: a tool for prediction of conopeptide superfamilies. Nucleic Acids Res. 2012;40(Web Server issue):W238–41
3. Scheler O, Kindt JT, Qavi AJ, Kaplinski L, Glynn B, Barry T, Kurg A, Bailey RC. (2012) Label-free, multiplexed detection of bacterial tmRNA using silicon photonic microring resonators. Biosens Bioelectron. 2012;36(1): 56–61

4. Laht S, Koua D, Kaplinski L, Lisacek F, Stöcklin R, Remm M. (2011) Identification and classification of conopeptides using profile Hidden Markov Models. Biochim Biophys Acta. 2012;1824(3):488–92
5. Scheler O, Kaplinski L, Glynn B, Palta P, Parkel S, Toome K, Maher M, Barry T, Remm M, Kurg A. (2011) Detection of NASBA amplified bacterial tmRNA molecules on SLICSel designed microarray probes. BMC Biotechnol. 2011;11:17
6. Kaplinski L, Scheler O, Parkel S, Palta P, Toome K, Kurg A, Remm M. (2010) Detection of tmRNA molecules on microarrays at low temperatures using helper oligonucleotides. BMC Biotechnol. 2010;10:34
7. Scheler O, Glynn B, Parkel S, Palta P, Toome K, Kaplinski L, Remm M, Maher M, Kurg A. (2009) Fluorescent labeling of NASBA amplified tmRNA molecules for microarray applications. BMC Biotechnol. 2009; 9:45
8. Kaplinski L, Remm M. (2007) MultiPLX: automatic grouping and evaluation of PCR primers. Methods Mol Biol. 2007;402:287–304. Review
9. Andreson R, Kaplinski L, Remm M. (2007) Fast masking of repeated primer binding sites in eukaryotic genomes. Methods Mol Biol. 2007; 402:201–18. Review
10. Mägi R, Kaplinski L, Remm M. (2006) The whole genome tagSNP selection and transferability among HapMap populations. Pac Symp Biocomput. 2006:535–43
11. Andreson R, Reppo E, Kaplinski L, Remm M. (2006) GENOMEMASKER package for designing unique genomic PCR primers. BMC Bioinformatics. 2006;7:172
12. Kaplinski L, Andreson R, Puurand T, Remm M. (2005) MultiPLX: automatic grouping and evaluation of PCR primers. Bioinformatics. 2005; 21(8):1701–2

ELULOOKIRJELDUS

Nimi: Lauris Kaplinski
Sünniaeg: 8. aprill 1971, Tartu
Kodakondsus: Eesti
Aaddress: Bioinformaatika õppetool,
Molekulaar ja Rakubioloogia Instituut, Tartu Ülikool
Riia mnt. 23B, 51010, Tartu, Eesti
Telefon: 737 5001
E-mail: lauris.kaplinski@ut.ee

Haridus ja erialane teenistuskäik:

1977–1989	Tartu I Keskkool
1989–1994	Tartu Ülikooli Bioloogia-geograafiateaduskond, biokeemia eriala
1994–1996	Tartu Ülikooli filosoofiateaduskond, filosoofia eriala
1998	Tartu Ülikooli Bioloogia-geograafiateaduskond, biokeemia eriala BSc
2002–2004	Biodata OÜ (IT spetsialist)
2004–	Eesti Biokeskus (programmeerija)
2007–	Tartu Ülikooli Loodus- ja tehnoloogiateaduskond, bioinformaatika eriala doktorant

Teadustegevus:
Minu teadustöö on põhiliselt keskendunud oligonukleotiidide hübridisatsioonimudeli rakendamisele PCR praimerite disainis ja hübridisatsioonikiipidel. Ma olen kirjutanud algoritmi ja programmi praimeritevaheliste interaktsioonide arvutamiseks ning nende optimiseeritud grupeerimiseks multipleks PCR jaoks. Detektsioonikiipide jaoks olen kirjutanud proovidisaini algoritmi ja programmi ning uurinud võimalusi RNA sekundaarstruktuuri lõhkumiseks. Viimasel ajal olen tegelnud mudelitega, mis võimaldaksid ennustada multipleks PCR ebaõnnestumise tõenäosust praimerite omaduste põhjal.

Publikatsioonid:

1. Koua D, Laht S, Kaplinski L, Stöcklin R, Remm M, Favreau P, Lisacek F. (2013) Position-specific scoring matrix and hidden Markov model complement each other for the prediction of conopeptide superfamilies. Biochim Biophys Acta. 2013;1834(4):717–24
2. Koua D, Brauer A, Laht S, Kaplinski L, Favreau P, Remm M, Lisacek F, Stöcklin R. (2012) ConoDictor: a tool for prediction of conopeptide superfamilies. Nucleic Acids Res. 2012;40(Web Server issue):W238–41
3. Scheler O, Kindt JT, Qavi AJ, Kaplinski L, Glynn B, Barry T, Kurg A, Bailey RC. (2012) Label-free, multiplexed detection of bacterial tmRNA

using silicon photonic microring resonators. Biosens Bioelectron. 2012; 36(1):56–61
4. Laht S, Koua D, Kaplinski L, Lisacek F, Stöcklin R, Remm M. (2011) Identification and classification of conopeptides using profile Hidden Markov Models. Biochim Biophys Acta. 2012;1824(3):488–92
5. Scheler O, Kaplinski L, Glynn B, Palta P, Parkel S, Toome K, Maher M, Barry T, Remm M, Kurg A. (2011) Detection of NASBA amplified bacterial tmRNA molecules on SLICSel designed microarray probes. BMC Biotechnol. 2011;11:17
6. Kaplinski L, Scheler O, Parkel S, Palta P, Toome K, Kurg A, Remm M. (2010) Detection of tmRNA molecules on microarrays at low temperatures using helper oligonucleotides. BMC Biotechnol. 2010;10:34
7. Scheler O, Glynn B, Parkel S, Palta P, Toome K, Kaplinski L, Remm M, Maher M, Kurg A. (2009) Fluorescent labeling of NASBA amplified tmRNA molecules for microarray applications. BMC Biotechnol. 2009; 9:45
8. Kaplinski L, Remm M. (2007) MultiPLX: automatic grouping and evaluation of PCR primers. Methods Mol Biol. 2007;402:287–304. Review
9. Andreson R, Kaplinski L, Remm M. (2007) Fast masking of repeated primer binding sites in eukaryotic genomes. Methods Mol Biol. 2007; 402: 201–18. Review
10. Mägi R, Kaplinski L, Remm M. (2006) The whole genome tagSNP selection and transferability among HapMap populations. Pac Symp Biocomput. 2006:535–43
11. Andreson R, Reppo E, Kaplinski L, Remm M. (2006) GENOMEMASKER package for designing unique genomic PCR primers. BMC Bioinformatics. 2006;7:172
12. Kaplinski L, Andreson R, Puurand T, Remm M. (2005) MultiPLX: automatic grouping and evaluation of PCR primers. Bioinformatics. 2005; 21(8):1701–2

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS

1. **Toivo Maimets**. Studies of human oncoprotein p53. Tartu, 1991, 96 p.
2. **Enn K. Seppet**. Thyroid state control over energy metabolism, ion transport and contractile functions in rat heart. Tartu, 1991, 135 p.
3. **Kristjan Zobel**. Epifüütsete makrosamblike väärtus õhu saastuse indikaatoritena Hamar-Dobani boreaalsetes mägimetsades. Tartu, 1992, 131 lk.
4. **Andres Mäe**. Conjugal mobilization of catabolic plasmids by transposable elements in helper plasmids. Tartu, 1992, 91 p.
5. **Maia Kivisaar**. Studies on phenol degradation genes of *Pseudomonas* sp. strain EST 1001. Tartu, 1992, 61 p.
6. **Allan Nurk**. Nucleotide sequences of phenol degradative genes from *Pseudomonas sp.* strain EST 1001 and their transcriptional activation in *Pseudomonas putida.* Tartu, 1992, 72 p.
7. **Ülo Tamm**. The genus *Populus* L. in Estonia: variation of the species biology and introduction. Tartu, 1993, 91 p.
8. **Jaanus Remme**. Studies on the peptidyltransferase centre of the *E.coli* ribosome. Tartu, 1993, 68 p.
9. **Ülo Langel**. Galanin and galanin antagonists. Tartu, 1993, 97 p.
10. **Arvo Käärd**. The development of an automatic online dynamic fluorescense-based pH-dependent fiber optic penicillin flowthrought biosensor for the control of the benzylpenicillin hydrolysis. Tartu, 1993, 117 p.
11. **Lilian Järvekülg**. Antigenic analysis and development of sensitive immunoassay for potato viruses. Tartu, 1993, 147 p.
12. **Jaak Palumets**. Analysis of phytomass partition in Norway spruce. Tartu, 1993, 47 p.
13. **Arne Sellin**. Variation in hydraulic architecture of *Picea abies* (L.) Karst. trees grown under different enviromental conditions. Tartu, 1994, 119 p.
13. **Mati Reeben**. Regulation of light neurofilament gene expression. Tartu, 1994, 108 p.
14. **Urmas Tartes**. Respiration rhytms in insects. Tartu, 1995, 109 p.
15. **Ülo Puurand.** The complete nucleotide sequence and infections *in vitro* transcripts from cloned cDNA of a potato A potyvirus. Tartu, 1995, 96 p.
16. **Peeter Hõrak**. Pathways of selection in avian reproduction: a functional framework and its application in the population study of the great tit (*Parus major*). Tartu, 1995, 118 p.
17. **Erkki Truve**. Studies on specific and broad spectrum virus resistance in transgenic plants. Tartu, 1996, 158 p.
18. **Illar Pata**. Cloning and characterization of human and mouse ribosomal protein S6-encoding genes. Tartu, 1996, 60 p.
19. **Ülo Niinemets**. Importance of structural features of leaves and canopy in determining species shade-tolerance in temperature deciduous woody taxa. Tartu, 1996, 150 p.

20. **Ants Kurg**. Bovine leukemia virus: molecular studies on the packaging region and DNA diagnostics in cattle. Tartu, 1996, 104 p.
21. **Ene Ustav**. E2 as the modulator of the BPV1 DNA replication. Tartu, 1996, 100 p.
22. **Aksel Soosaar**. Role of helix-loop-helix and nuclear hormone receptor transcription factors in neurogenesis. Tartu, 1996, 109 p.
23. **Maido Remm**. Human papillomavirus type 18: replication, transformation and gene expression. Tartu, 1997, 117 p.
24. **Tiiu Kull**. Population dynamics in *Cypripedium calceolus* L. Tartu, 1997, 124 p.
25. **Kalle Olli**. Evolutionary life-strategies of autotrophic planktonic microorganisms in the Baltic Sea. Tartu, 1997, 180 p.
26. **Meelis Pärtel**. Species diversity and community dynamics in calcareous grassland communities in Western Estonia. Tartu, 1997, 124 p.
27. **Malle Leht**. The Genus *Potentilla* L. in Estonia, Latvia and Lithuania: distribution, morphology and taxonomy. Tartu, 1997, 186 p.
28. **Tanel Tenson**. Ribosomes, peptides and antibiotic resistance. Tartu, 1997, 80 p.
29. **Arvo Tuvikene**. Assessment of inland water pollution using biomarker responses in fish *in vivo* and *in vitro*. Tartu, 1997, 160 p.
30. **Urmas Saarma**. Tuning ribosomal elongation cycle by mutagenesis of 23S rRNA. Tartu, 1997, 134 p.
31. **Henn Ojaveer**. Composition and dynamics of fish stocks in the gulf of Riga ecosystem. Tartu, 1997, 138 p.
32. **Lembi Lõugas**. Post-glacial development of vertebrate fauna in Estonian water bodies. Tartu, 1997, 138 p.
33. **Margus Pooga**. Cell penetrating peptide, transportan, and its predecessors, galanin-based chimeric peptides. Tartu, 1998, 110 p.
34. **Andres Saag**. Evolutionary relationships in some cetrarioid genera (Lichenized Ascomycota). Tartu, 1998, 196 p.
35. **Aivar Liiv**. Ribosomal large subunit assembly *in vivo*. Tartu, 1998, 158 p.
36. **Tatjana Oja**. Isoenzyme diversity and phylogenetic affinities among the eurasian annual bromes (*Bromus* L., Poaceae). Tartu, 1998, 92 p.
37. **Mari Moora**. The influence of arbuscular mycorrhizal (AM) symbiosis on the competition and coexistence of calcareous grassland plant species. Tartu, 1998, 78 p.
38. **Olavi Kurina**. Fungus gnats in Estonia (*Diptera: Bolitophilidae, Keroplatidae, Macroceridae, Ditomyiidae, Diadocidiidae, Mycetophilidae*). Tartu, 1998, 200 p.
39. **Andrus Tasa**. Biological leaching of shales: black shale and oil shale. Tartu, 1998, 98 p.
40. **Arnold Kristjuhan.** Studies on transcriptional activator properties of tumor suppressor protein p53. Tartu, 1998, 86 p.

41. **Sulev Ingerpuu.** Characterization of some human myeloid cell surface and nuclear differentiation antigens. Tartu, 1998, 163 p.
42. **Veljo Kisand.** Responses of planktonic bacteria to the abiotic and biotic factors in the shallow lake Võrtsjärv. Tartu, 1998, 118 p.
43. **Kadri Põldmaa.** Studies in the systematics of hypomyces and allied genera (Hypocreales, Ascomycota). Tartu, 1998, 178 p.
44. **Markus Vetemaa.** Reproduction parameters of fish as indicators in environmental monitoring. Tartu, 1998, 117 p.
45. **Heli Talvik.** Prepatent periods and species composition of different *Oesophagostomum* spp. populations in Estonia and Denmark. Tartu, 1998, 104 p.
46. **Katrin Heinsoo.** Cuticular and stomatal antechamber conductance to water vapour diffusion in *Picea abies* (L.) karst. Tartu, 1999, 133 p.
47. **Tarmo Annilo.** Studies on mammalian ribosomal protein S7. Tartu, 1998, 77 p.
48. **Indrek Ots.** Health state indicies of reproducing great tits (*Parus major*): sources of variation and connections with life-history traits. Tartu, 1999, 117 p.
49. **Juan Jose Cantero.** Plant community diversity and habitat relationships in central Argentina grasslands. Tartu, 1999, 161 p.
50. **Rein Kalamees.** Seed bank, seed rain and community regeneration in Estonian calcareous grasslands. Tartu, 1999, 107 p.
51. **Sulev Kõks.** Cholecystokinin (CCK) — induced anxiety in rats: influence of environmental stimuli and involvement of endopioid mechanisms and erotonin. Tartu, 1999, 123 p.
52. **Ebe Sild.** Impact of increasing concentrations of O_3 and CO_2 on wheat, clover and pasture. Tartu, 1999, 123 p.
53. **Ljudmilla Timofejeva.** Electron microscopical analysis of the synaptonemal complex formation in cereals. Tartu, 1999, 99 p.
54. **Andres Valkna.** Interactions of galanin receptor with ligands and G-proteins: studies with synthetic peptides. Tartu, 1999, 103 p.
55. **Taavi Virro.** Life cycles of planktonic rotifers in lake Peipsi. Tartu, 1999, 101 p.
56. **Ana Rebane.** Mammalian ribosomal protein S3a genes and intron-encoded small nucleolar RNAs U73 and U82. Tartu, 1999, 85 p.
57. **Tiina Tamm.** Cocksfoot mottle virus: the genome organisation and translational strategies. Tartu, 2000, 101 p.
58. **Reet Kurg.** Structure-function relationship of the bovine papilloma virus E2 protein. Tartu, 2000, 89 p.
59. **Toomas Kivisild.** The origins of Southern and Western Eurasian populations: an mtDNA study. Tartu, 2000, 121 p.
60. **Niilo Kaldalu.** Studies of the TOL plasmid transcription factor XylS. Tartu 2000. 88 p.

61. **Dina Lepik.** Modulation of viral DNA replication by tumor suppressor protein p53. Tartu 2000. 106 p.
62. **Kai Vellak.** Influence of different factors on the diversity of the bryophyte vegetation in forest and wooded meadow communities. Tartu 2000. 122 p.
63. **Jonne Kotta.** Impact of eutrophication and biological invasionas on the structure and functions of benthic macrofauna. Tartu 2000. 160 p.
64. **Georg Martin.** Phytobenthic communities of the Gulf of Riga and the inner sea the West-Estonian archipelago. Tartu, 2000. 139 p.
65. **Silvia Sepp.** Morphological and genetical variation of *Alchemilla L.* in Estonia. Tartu, 2000. 124 p.
66. **Jaan Liira.** On the determinants of structure and diversity in herbaceous plant communities. Tartu, 2000. 96 p.
67. **Priit Zingel.** The role of planktonic ciliates in lake ecosystems. Tartu 2001. 111 p.
68. **Tiit Teder.** Direct and indirect effects in Host-parasitoid interactions: ecological and evolutionary consequences. Tartu 2001. 122 p.
69. **Hannes Kollist.** Leaf apoplastic ascorbate as ozone scavenger and its transport across the plasma membrane. Tartu 2001. 80 p.
70. **Reet Marits.** Role of two-component regulator system PehR-PehS and extracellular protease PrtW in virulence of *Erwinia Carotovora* subsp. *Carotovora*. Tartu 2001. 112 p.
71. **Vallo Tilgar.** Effect of calcium supplementation on reproductive performance of the pied flycatcher *Ficedula hypoleuca* and the great tit *Parus major,* breeding in Nothern temperate forests. Tartu, 2002. 126 p.
72. **Rita Hõrak.** Regulation of transposition of transposon Tn*4652* in *Pseudomonas putida*. Tartu, 2002. 108 p.
73. **Liina Eek-Piirsoo.** The effect of fertilization, mowing and additional illumination on the structure of a species-rich grassland community. Tartu, 2002. 74 p.
74. **Krõõt Aasamaa.** Shoot hydraulic conductance and stomatal conductance of six temperate deciduous tree species. Tartu, 2002. 110 p.
75. **Nele Ingerpuu.** Bryophyte diversity and vascular plants. Tartu, 2002. 112 p.
76. **Neeme Tõnisson.** Mutation detection by primer extension on oligonucleotide microarrays. Tartu, 2002. 124 p.
77. **Margus Pensa.** Variation in needle retention of Scots pine in relation to leaf morphology, nitrogen conservation and tree age. Tartu, 2003. 110 p.
78. **Asko Lõhmus.** Habitat preferences and quality for birds of prey: from principles to applications. Tartu, 2003. 168 p.
79. **Viljar Jaks.** p53 — a switch in cellular circuit. Tartu, 2003. 160 p.
80. **Jaana Männik.** Characterization and genetic studies of four ATP-binding cassette (ABC) transporters. Tartu, 2003. 140 p.
81. **Marek Sammul.** Competition and coexistence of clonal plants in relation to productivity. Tartu, 2003. 159 p

82. **Ivar Ilves.** Virus-cell interactions in the replication cycle of bovine papillomavirus type 1. Tartu, 2003. 89 p.
83. **Andres Männik.** Design and characterization of a novel vector system based on the stable replicator of bovine papillomavirus type 1. Tartu, 2003. 109 p.
84. **Ivika Ostonen.** Fine root structure, dynamics and proportion in net primary production of Norway spruce forest ecosystem in relation to site conditions. Tartu, 2003. 158 p.
85. **Gudrun Veldre.** Somatic status of 12–15-year-old Tartu schoolchildren. Tartu, 2003. 199 p.
86. **Ülo Väli.** The greater spotted eagle *Aquila clanga* and the lesser spotted eagle *A. pomarina*: taxonomy, phylogeography and ecology. Tartu, 2004. 159 p.
87. **Aare Abroi.** The determinants for the native activities of the bovine papillomavirus type 1 E2 protein are separable. Tartu, 2004. 135 p.
88. **Tiina Kahre.** Cystic fibrosis in Estonia. Tartu, 2004. 116 p.
89. **Helen Orav-Kotta.** Habitat choice and feeding activity of benthic suspension feeders and mesograzers in the northern Baltic Sea. Tartu, 2004. 117 p.
90. **Maarja Öpik.** Diversity of arbuscular mycorrhizal fungi in the roots of perennial plants and their effect on plant performance. Tartu, 2004. 175 p.
91. **Kadri Tali.** Species structure of *Neotinea ustulata*. Tartu, 2004. 109 p.
92. **Kristiina Tambets.** Towards the understanding of post-glacial spread of human mitochondrial DNA haplogroups in Europe and beyond: a phylogeographic approach. Tartu, 2004. 163 p.
93. **Arvi Jõers.** Regulation of p53-dependent transcription. Tartu, 2004. 103 p.
94. **Lilian Kadaja.** Studies on modulation of the activity of tumor suppressor protein p53. Tartu, 2004. 103 p.
95. **Jaak Truu.** Oil shale industry wastewater: impact on river microbial community and possibilities for bioremediation. Tartu, 2004. 128 p.
96. **Maire Peters.** Natural horizontal transfer of the *pheBA* operon. Tartu, 2004. 105 p.
97. **Ülo Maiväli.** Studies on the structure-function relationship of the bacterial ribosome. Tartu, 2004. 130 p.
98. **Merit Otsus.** Plant community regeneration and species diversity in dry calcareous grasslands. Tartu, 2004. 103 p.
99. **Mikk Heidemaa.** Systematic studies on sawflies of the genera *Dolerus, Empria,* and *Caliroa* (Hymenoptera: Tenthredinidae). Tartu, 2004. 167 p.
100. **Ilmar Tõnno.** The impact of nitrogen and phosphorus concentration and N/P ratio on cyanobacterial dominance and N_2 fixation in some Estonian lakes. Tartu, 2004. 111 p.
101. **Lauri Saks.** Immune function, parasites, and carotenoid-based ornaments in greenfinches. Tartu, 2004. 144 p.
102. **Siiri Rootsi.** Human Y-chromosomal variation in European populations. Tartu, 2004. 142 p.

103. **Eve Vedler.** Structure of the 2,4-dichloro-phenoxyacetic acid-degradative plasmid pEST4011. Tartu, 2005. 106 p.
104. **Andres Tover.** Regulation of transcription of the phenol degradation *pheBA* operon in *Pseudomonas putida*. Tartu, 2005. 126 p.
105. **Helen Udras.** Hexose kinases and glucose transport in the yeast *Hansenula polymorpha*. Tartu, 2005. 100 p.
106. **Ave Suija.** Lichens and lichenicolous fungi in Estonia: diversity, distribution patterns, taxonomy. Tartu, 2005. 162 p.
107. **Piret Lõhmus.** Forest lichens and their substrata in Estonia. Tartu, 2005. 162 p.
108. **Inga Lips.** Abiotic factors controlling the cyanobacterial bloom occurrence in the Gulf of Finland. Tartu, 2005. 156 p.
109. **Kaasik, Krista.** Circadian clock genes in mammalian clockwork, metabolism and behaviour. Tartu, 2005. 121 p.
110. **Juhan Javoiš.** The effects of experience on host acceptance in ovipositing moths. Tartu, 2005. 112 p.
111. **Tiina Sedman.** Characterization of the yeast *Saccharomyces cerevisiae* mitochondrial DNA helicase Hmi1. Tartu, 2005. 103 p.
112. **Ruth Aguraiuja.** Hawaiian endemic fern lineage *Diellia* (Aspleniaceae): distribution, population structure and ecology. Tartu, 2005. 112 p.
113. **Riho Teras.** Regulation of transcription from the fusion promoters generated by transposition of Tn*4652* into the upstream region of *pheBA* operon in *Pseudomonas putida*. Tartu, 2005. 106 p.
114. **Mait Metspalu.** Through the course of prehistory in india: tracing the mtDNA trail. Tartu, 2005. 138 p.
115. **Elin Lõhmussaar.** The comparative patterns of linkage disequilibrium in European populations and its implication for genetic association studies. Tartu, 2006. 124 p.
116. **Priit Kupper.** Hydraulic and environmental limitations to leaf water relations in trees with respect to canopy position. Tartu, 2006. 126 p.
117. **Heili Ilves.** Stress-induced transposition of Tn*4652* in *Pseudomonas Putida.* Tartu, 2006. 120 p.
118. **Silja Kuusk.** Biochemical properties of Hmi1p, a DNA helicase from *Saccharomyces cerevisiae* mitochondria. Tartu, 2006. 126 p.
119. **Kersti Püssa.** Forest edges on medium resolution landsat thematic mapper satellite images. Tartu, 2006. 90 p.
120. **Lea Tummeleht.** Physiological condition and immune function in great tits (*Parus major* l.): Sources of variation and trade-offs in relation to growth. Tartu, 2006. 94 p.
121. **Toomas Esperk.** Larval instar as a key element of insect growth schedules. Tartu, 2006. 186 p.
122. **Harri Valdmann.** Lynx (*Lynx lynx*) and wolf (*Canis lupus*) in the Baltic region: Diets, helminth parasites and genetic variation. Tartu, 2006. 102 p.

123. **Priit Jõers.** Studies of the mitochondrial helicase Hmi1p in *Candida albicans* and *Saccharomyces cerevisia*. Tartu, 2006. 113 p.
124. **Kersti Lilleväli.** Gata3 and Gata2 in inner ear development. Tartu, 2007. 123 p.
125. **Kai Rünk.** Comparative ecology of three fern species: *Dryopteris carthusiana* (Vill.) H.P. Fuchs, *D. expansa* (C. Presl) Fraser-Jenkins & Jermy and *D. dilatata* (Hoffm.) A. Gray (Dryopteridaceae). Tartu, 2007. 143 p.
126. **Aveliina Helm.** Formation and persistence of dry grassland diversity: role of human history and landscape structure. Tartu, 2007. 89 p.
127. **Leho Tedersoo.** Ectomycorrhizal fungi: diversity and community structure in Estonia, Seychelles and Australia. Tartu, 2007. 233 p.
128. **Marko Mägi.** The habitat-related variation of reproductive performance of great tits in a deciduous-coniferous forest mosaic: looking for causes and consequences. Tartu, 2007. 135 p.
129. **Valeria Lulla.** Replication strategies and applications of Semliki Forest virus. Tartu, 2007. 109 p.
130. **Ülle Reier**. Estonian threatened vascular plant species: causes of rarity and conservation. Tartu, 2007. 79 p.
131. **Inga Jüriado**. Diversity of lichen species in Estonia: influence of regional and local factors. Tartu, 2007. 171 p.
132. **Tatjana Krama.** Mobbing behaviour in birds: costs and reciprocity based cooperation. Tartu, 2007. 112 p.
133. **Signe Saumaa.** The role of DNA mismatch repair and oxidative DNA damage defense systems in avoidance of stationary phase mutations in *Pseudomonas putida.* Tartu, 2007. 172 p.
134. **Reedik Mägi**. The linkage disequilibrium and the selection of genetic markers for association studies in european populations. Tartu, 2007. 96 p.
135. **Priit Kilgas.** Blood parameters as indicators of physiological condition and skeletal development in great tits (*Parus major*): natural variation and application in the reproductive ecology of birds. Tartu, 2007. 129 p.
136. **Anu Albert**. The role of water salinity in structuring eastern Baltic coastal fish communities. Tartu, 2007. 95 p.
137. **Kärt Padari.** Protein transduction mechanisms of transportans. Tartu, 2008. 128 p.
138. **Siiri-Lii Sandre.** Selective forces on larval colouration in a moth. Tartu, 2008. 125 p.
139. **Ülle Jõgar.** Conservation and restoration of semi-natural floodplain meadows and their rare plant species. Tartu, 2008. 99 p.
140. **Lauri Laanisto.** Macroecological approach in vegetation science: generality of ecological relationships at the global scale. Tartu, 2008. 133 p.
141. **Reidar Andreson**. Methods and software for predicting PCR failure rate in large genomes. Tartu, 2008. 105 p.
142. **Birgot Paavel.** Bio-optical properties of turbid lakes. Tartu, 2008. 175 p.

25

143. **Kaire Torn.** Distribution and ecology of charophytes in the Baltic Sea. Tartu, 2008, 98 p.
144. **Vladimir Vimberg.** Peptide mediated macrolide resistance. Tartu, 2008, 190 p.
145. **Daima Örd.** Studies on the stress-inducible pseudokinase TRB3, a novel inhibitor of transcription factor ATF4. Tartu, 2008, 108 p.
146. **Lauri Saag.** Taxonomic and ecologic problems in the genus *Lepraria* (*Stereocaulaceae*, lichenised *Ascomycota*). Tartu, 2008, 175 p.
147. **Ulvi Karu.** Antioxidant protection, carotenoids and coccidians in greenfinches – assessment of the costs of immune activation and mechanisms of parasite resistance in a passerine with carotenoid-based ornaments. Tartu, 2008, 124 p.
148. **Jaanus Remm.** Tree-cavities in forests: density, characteristics and occupancy by animals. Tartu, 2008, 128 p.
149. **Epp Moks.** Tapeworm parasites *Echinococcus multilocularis* and *E. granulosus* in Estonia: phylogenetic relationships and occurrence in wild carnivores and ungulates. Tartu, 2008, 82 p.
150. **Eve Eensalu.** Acclimation of stomatal structure and function in tree canopy: effect of light and CO_2 concentration. Tartu, 2008, 108 p.
151. **Janne Pullat**. Design, functionlization and application of an *in situ* synthesized oligonucleotide microarray. Tartu, 2008, 108 p.
152. **Marta Putrinš.** Responses of *Pseudomonas putida* to phenol-induced metabolic and stress signals. Tartu, 2008, 142 p.
153. **Marina Semtšenko.** Plant root behaviour: responses to neighbours and physical obstructions. Tartu, 2008, 106 p.
154. **Marge Starast.** Influence of cultivation techniques on productivity and fruit quality of some *Vaccinium* and *Rubus* taxa. Tartu, 2008, 154 p.
155. **Age Tats.** Sequence motifs influencing the efficiency of translation. Tartu, 2009, 104 p.
156. **Radi Tegova.** The role of specialized DNA polymerases in mutagenesis in *Pseudomonas putida.* Tartu, 2009, 124 p.
157. **Tsipe Aavik.** Plant species richness, composition and functional trait pattern in agricultural landscapes – the role of land use intensity and landscape structure. Tartu, 2009, 112 p.
158. **Kaja Kiiver.** Semliki forest virus based vectors and cell lines for studying the replication and interactions of alphaviruses and hepaciviruses. Tartu, 2009, 104 p.
159. **Meelis Kadaja.** Papillomavirus Replication Machinery Induces Genomic Instability in its Host Cell. Tartu, 2009, 126 p.
160. **Pille Hallast.** Human and chimpanzee Luteinizing hormone/Chorionic Gonadotropin beta (*LHB/CGB*) gene clusters: diversity and divergence of young duplicated genes. Tartu, 2009, 168 p.
161. **Ain Vellak.** Spatial and temporal aspects of plant species conservation. Tartu, 2009, 86 p.

162. **Triinu Remmel.** Body size evolution in insects with different colouration strategies: the role of predation risk. Tartu, 2009, 168 p.
163. **Jaana Salujõe.** Zooplankton as the indicator of ecological quality and fish predation in lake ecosystems. Tartu, 2009, 129 p.
164. **Ele Vahtmäe.** Mapping benthic habitat with remote sensing in optically complex coastal environments. Tartu, 2009, 109 p.
165. **Liisa Metsamaa.** Model-based assessment to improve the use of remote sensing in recognition and quantitative mapping of cyanobacteria. Tartu, 2009, 114 p.
166. **Pille Säälik.** The role of endocytosis in the protein transduction by cell-penetrating peptides. Tartu, 2009, 155 p.
167. **Lauri Peil.** Ribosome assembly factors in *Escherichia coli.* Tartu, 2009, 147 p.
168. **Lea Hallik.** Generality and specificity in light harvesting, carbon gain capacity and shade tolerance among plant functional groups. Tartu, 2009, 99 p.
169. **Mariliis Tark.** Mutagenic potential of DNA damage repair and tolerance mechanisms under starvation stress. Tartu, 2009, 191 p.
170. **Riinu Rannap.** Impacts of habitat loss and restoration on amphibian populations. Tartu, 2009, 117 p.
171. **Maarja Adojaan.** Molecular variation of HIV-1 and the use of this knowledge in vaccine development. Tartu, 2009, 95 p.
172. **Signe Altmäe.** Genomics and transcriptomics of human induced ovarian folliculogenesis. Tartu, 2010, 179 p.
173. **Triin Suvi.** Mycorrhizal fungi of native and introduced trees in the Seychelles Islands. Tartu, 2010, 107 p.
174. **Velda Lauringson.** Role of suspension feeding in a brackish-water coastal sea. Tartu, 2010, 123 p.
175. **Eero Talts.** Photosynthetic cyclic electron transport – measurement and variably proton-coupled mechanism. Tartu, 2010, 121 p.
176. **Mari Nelis.** Genetic structure of the Estonian population and genetic distance from other populations of European descent. Tartu, 2010, 97 p.
177. **Kaarel Krjutškov.** Arrayed Primer Extension-2 as a multiplex PCR-based method for nucleic acid variation analysis: method and applications. Tartu, 2010, 129 p.
178. **Egle Köster.** Morphological and genetical variation within species complexes: *Anthyllis vulneraria* s. l. and *Alchemilla vulgaris* (coll.). Tartu, 2010, 101 p.
179. **Erki Õunap.** Systematic studies on the subfamily Sterrhinae (Lepidoptera: Geometridae). Tartu, 2010, 111 p.
180. **Merike Jõesaar.** Diversity of key catabolic genes at degradation of phenol and *p*-cresol in pseudomonads. Tartu, 2010, 125 p.
181. **Kristjan Herkül.** Effects of physical disturbance and habitat-modifying species on sediment properties and benthic communities in the northern Baltic Sea. Tartu, 2010, 123 p.

182. **Arto Pulk.** Studies on bacterial ribosomes by chemical modification approaches. Tartu, 2010, 161 p.
183. **Maria Põllupüü.** Ecological relations of cladocerans in a brackish-water ecosystem. Tartu, 2010, 126 p.
184. **Toomas Silla.** Study of the segregation mechanism of the Bovine Papillomavirus Type 1. Tartu, 2010, 188 p.
185. **Gyaneshwer Chaubey.** The demographic history of India: A perspective based on genetic evidence. Tartu, 2010, 184 p.
186. **Katrin Kepp.** Genes involved in cardiovascular traits: detection of genetic variation in Estonian and Czech populations. Tartu, 2010, 164 p.
187. **Virve Sõber.** The role of biotic interactions in plant reproductive performance. Tartu, 2010, 92 p.
188. **Kersti Kangro.** The response of phytoplankton community to the changes in nutrient loading. Tartu, 2010, 144 p.
189. **Joachim M. Gerhold.** Replication and Recombination of mitochondrial DNA in Yeast. Tartu, 2010, 120 p.
190. **Helen Tammert.** Ecological role of physiological and phylogenetic diversity in aquatic bacterial communities. Tartu, 2010, 140 p.
191. **Elle Rajandu.** Factors determining plant and lichen species diversity and composition in Estonian *Calamagrostis* and *Hepatica* site type forests. Tartu, 2010, 123 p.
192. **Paula Ann Kivistik.** ColR-ColS signalling system and transposition of Tn*4652* in the adaptation of *Pseudomonas putida.* Tartu, 2010, 118 p.
193. **Siim Sõber.** Blood pressure genetics: from candidate genes to genome-wide association studies. Tartu, 2011, 120 p.
194. **Kalle Kipper.** Studies on the role of helix 69 of 23S rRNA in the factor-dependent stages of translation initiation, elongation, and termination. Tartu, 2011, 178 p.
195. **Triinu Siibak.** Effect of antibiotics on ribosome assembly is indirect. Tartu, 2011, 134 p.
196. **Tambet Tõnissoo.** Identification and molecular analysis of the role of guanine nucleotide exchange factor RIC-8 in mouse development and neural function. Tartu, 2011, 110 p.
197. **Helin Räägel.** Multiple faces of cell-penetrating peptides – their intra-cellular trafficking, stability and endosomal escape during protein trans-duction. Tartu, 2011, 161 p.
198. **Andres Jaanus.** Phytoplankton in Estonian coastal waters – variability, trends and response to environmental pressures. Tartu, 2011, 157 p.
199. **Tiit Nikopensius.** Genetic predisposition to nonsyndromic orofacial clefts. Tartu, 2011, 152 p.
200. **Signe Värv.** Studies on the mechanisms of RNA polymerase II-dependent transcription elongation. Tartu, 2011, 108 p.
201. **Kristjan Välk.** Gene expression profiling and genome-wide association studies of non-small cell lung cancer. Tartu, 2011, 98 p.

202. **Arno Põllumäe.** Spatio-temporal patterns of native and invasive zooplankton species under changing climate and eutrophication conditions. Tartu, 2011, 153 p.
203. **Egle Tammeleht.** Brown bear (*Ursus arctos*) population structure, demographic processes and variations in diet in northern Eurasia. Tartu, 2011, 143 p.
205. **Teele Jairus.** Species composition and host preference among ectomycorrhizal fungi in Australian and African ecosystems. Tartu, 2011, 106 p.
206. **Kessy Abarenkov.** PlutoF – cloud database and computing services supporting biological research. Tartu, 2011, 125 p.
207. **Marina Grigorova.** Fine-scale genetic variation of follicle-stimulating hormone beta-subunit coding gene (*FSHB*) and its association with reproductive health. Tartu, 2011, 184 p.
208. **Anu Tiitsaar.** The effects of predation risk and habitat history on butterfly communities. Tartu, 2011, 97 p.
209. **Elin Sild.** Oxidative defences in immunoecological context: validation and application of assays for nitric oxide production and oxidative burst in a wild passerine. Tartu, 2011, 105 p.
210. **Irja Saar**. The taxonomy and phylogeny of the genera *Cystoderma* and *Cystodermella* (Agaricales, Fungi). Tartu, 2012, 167 p.
211. **Pauli Saag.** Natural variation in plumage bacterial assemblages in two wild breeding passerines. Tartu, 2012, 113 p.
212. **Aleksei Lulla.** Alphaviral nonstructural protease and its polyprotein substrate: arrangements for the perfect marriage. Tartu, 2012, 143 p.
213. **Mari Järve.** Different genetic perspectives on human history in Europe and the Caucasus: the stories told by uniparental and autosomal markers. Tartu, 2012, 119 p.
214. **Ott Scheler**. The application of tmRNA as a marker molecule in bacterial diagnostics using microarray and biosensor technology. Tartu, 2012, 93 p.
215. **Anna Balikova**. Studies on the functions of tumor-associated mucin-like leukosialin (CD43) in human cancer cells. Tartu, 2012, 129 p.
216. **Triinu Kõressaar.** Improvement of PCR primer design for detection of prokaryotic species. Tartu, 2012, 83 p.
217. **Tuul Sepp.** Hematological health state indices of greenfinches: sources of individual variation and responses to immune system manipulation. Tartu, 2012, 117 p.
218. **Rya Ero.** Modifier view of the bacterial ribosome. Tartu, 2012, 146 p.
219. **Mohammad Bahram.** Biogeography of ectomycorrhizal fungi across different spatial scales. Tartu, 2012, 165 p.
220. **Annely Lorents.** Overcoming the plasma membrane barrier: uptake of amphipathic cell-penetrating peptides induces influx of calcium ions and downstream responses. Tartu, 2012, 113 p.

26

221. **Katrin Männik.** Exploring the genomics of cognitive impairment: whole-genome SNP genotyping experience in Estonian patients and general population. Tartu, 2012, 171 p.
222. **Marko Prous.** Taxonomy and phylogeny of the sawfly genus *Empria* (Hymenoptera, Tenthredinidae). Tartu, 2012, 192 p.
223. **Triinu Visnapuu.** Levansucrases encoded in the genome of *Pseudomonas syringae* pv. tomato DC3000: heterologous expression, biochemical characterization, mutational analysis and spectrum of polymerization products. Tartu, 2012, 160 p.
224. **Nele Tamberg.** Studies on Semliki Forest virus replication and pathogenesis. Tartu, 2012, 109 p.
225. **Tõnu Esko.** Novel applications of SNP array data in the analysis of the genetic structure of Europeans and in genetic association studies. Tartu, 2012, 149 p.
226. **Timo Arula.** Ecology of early life-history stages of herring *Clupea harengus membras* in the northeastern Baltic Sea. Tartu, 2012, 143 p.
227. **Inga Hiiesalu.** Belowground plant diversity and coexistence patterns in grassland ecosystems. Tartu, 2012, 130 p.
228. **Kadri Koorem.** The influence of abiotic and biotic factors on small-scale plant community patterns and regeneration in boreonemoral forest. Tartu, 2012, 114 p.
229. **Liis Andresen.** Regulation of virulence in plant-pathogenic pectobacteria. Tartu, 2012, 122 p.
230. **Kaupo Kohv.** The direct and indirect effects of management on boreal forest structure and field layer vegetation. Tartu, 2012, 124 p.
231. **Mart Jüssi.** Living on an edge: landlocked seals in changing climate. Tartu, 2012, 114 p.
232. **Riina Klais.** Phytoplankton trends in the Baltic Sea. Tartu, 2012, 136 p.
233. **Rauno Veeroja.** Effects of winter weather, population density and timing of reproduction on life-history traits and population dynamics of moose (*Alces alces*) in Estonia. Tartu, 2012, 92 p.
234. **Marju Keis.** Brown bear (*Ursus arctos*) phylogeography in northern Eurasia. Tartu, 2013, 142 p.
235. **Sergei Põlme.** Biogeography and ecology of *alnus*- associated ectomycorrhizal fungi – from regional to global scale. Tartu, 2013, 90 p.
236. **Liis Uusküla.**. Tartu, 2013, 173 p.
237. **Marko Lõoke.** Studies on DNA replication initiation in *Saccharomyces cerevisiae.* Tartu, 2013, 112 p.
238. **Anne Aan.** Light- and nitrogen-use and biomass allocation along productivity gradients in multilayer plant communities. Tartu, 2013, 127 p.
239. **Heidi Tamm.** Comprehending phylogenetic diversity – case studies in three groups of ascomycetes. Tartu, 2013, 136 p.

240. **Liina Kangur.** High-Pressure Spectroscopy Study of Chromophore-Binding Hydrogen Bonds in Light-Harvesting Complexes of Photosynthetic Bacteria. Tartu, 2013, 150 p.
241. **Margus Leppik.** Substrate specificity of the multisite specific pseudouridine synthase RluD. Tartu, 2013, 111 p.