THE EFFECT OF TARGET SECONDARY STRUCTURE ON MICROARRAY DATA QUALITY

by

Vladyslava Grygoriyivna Ratushna

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Information Technology

Charlotte

2010

Approved by:

Dr. Cynthia J. Gibas

Dr. Jennifer W. Weller

Dr. Anthony A. Fodor

Dr. Julie M. Goodliffe

Dr. James D. Oliver

ABSTRACT

VLADYSLAVA GRYGORIYIVNA RATUSHNA. The effect of target secondary structure on microarray data quality. (Under the direction of DR. CYNTHIA J. GIBAS)

DNA microarrays have become an invaluable high throughput biotechnology method, which allows a parallel investigation of thousands of cellular events in a single experiment. The principle behind the technology is very simple: fluorescently labeled single stranded target molecules bind to their specific probes deposited on the microarray surface. However, the microarray data rarely represent a yes or no answer to a biological community, but rather provide a direction for further investigation. There is a complicated quantitative relationship between a detected spot signal and the amount of target present in the unknown mixture. We hypothesize that physical characteristics of probe and target molecules complicate the binding reaction between target and probe. To test this hypothesis, we designed a controlled microarray experiment in which the amount and stability of the secondary structure present in the probe-binding regions of target as biophysical properties of nucleic acids varies in a known way. Based on computational simulations of hybridization, we hypothesize that secondary structure formation in the target can result in considerable interference with the process of probe-target binding. This interference will have the effect of lowering the spot signal intensity. We simulated hybridization between probe and target and analyzed the simulation data to predict how much the microarray signal is affected by folding of the target molecule, for the purpose of developing a new generation of microarray design and analysis software.

DEDICATION

Sviatoslav & Olga

You grew along with this dissertation!

ACKNOWLEDGMENTS

I would like to thank all the members of my graduate committee Dr. Cynthia J. Gibas, Dr. Jennifer W. Weller, Dr. Anthony A. Fodor, Dr. Julie M. Goodliffe and Dr. Dr. James D. Oliver for scientific mentoring, and support during the work on this project. Thank you for sharing with me your extraordinary knowledge of bioinformatics, biophysics and molecular biology. It was a great pleasure to work under your guidance.

I would like to give special thanks to Chaevia and Chaevien Clendinen for performing excellent large scale plasmid DNA extractions; Dr. Ra'ad Gharaibeh for sharing some of his BioPerl codes; Timothy Hamp for generously sharing his knowledge of molecular lab techniques. I would like to thank the NIH for financial support of our investigation, and the UNC Charlotte GASP awards.

I am especially grateful to my academic advisor Dr. Cynthia J. Gibas not only for the scientific guidance, but also for friendship, patience and care. You have created an atmosphere of warmth and kindness in our little bioinformatics lab, and I am truly happy to be a part of it.

I would like to thank Patricia Artis and Dr. Larry Mays for being the heart of our department.

I am also grateful to my friends, colleagues, and office mates at UNC Charlotte for support and all of the great time that we had together.

Finally, I would like to thank my family for the many of years of love and support especially my mother Anna Ratushna and my husband Dr. Vladimir Gantovnik.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

aRNA	Antisense RNA
array CGH	array comparative genomic hybridization
bp.	base pair
cDNA	complementary DNA
CDS	coding sequence
cmRNA	complementary mRNA
ΔG°	Gibbs free energy at standard state conditions
ΔH°	enthalpy at standard state conditions
ΔS°	entropy at standard state conditions
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
K_{eq}	equilibrium constant
(LATE)-PCR	linear-after-the-exponential-PCR
$[Mg^{2+}]$	magnesium ion concentration
mRNA	messenger RNA
NA	nucleic acid
$[Na^{+}]$	sodium ion concentration
NMR	nuclear magnetic resonance
OMP DE	Oligonucleotide Modeling Platform Developer Edition
ORF	open reading frame
PCR	polymerase chain reaction

RISC	RNA-induced silencing complex
RNA	ribonucleic acid
RNAi	RNA interference
RNA-Seq	whole transcriptome shotgun sequencing
rRNA	ribosomal RNA
RT-nPCR	reverse-transcription and nested polymerase chain reaction
RT-QPCR	real-time quantitative polymerase chain reaction
RT-PCR	reverse-transcription polymerase chain reaction
shRNA	short hairpin RNA
siRNA	small interfering RNA
SNP	single nucleotide polymorphism
TdT	terminal deoxynucleotydil transferase
T_{m}	melting temperature
Visual OMP	Visual Oligonucleotide Modeling Platform

CHAPTER 1: INTRODUCTION TO THE BINDING PROBLEM

1.1 Overview of Microarray Technology

One of the most challenging tasks of modern biotechnology is to observe how cells regulate their function; one way to approach this question is by determining gene expression levels and how they change during the course of processes such as cell differentiation and tissue morphogenesis, or how they respond to environmental stresses or disease conditions. Early attempts to address this issue were performed in a low-throughput fashion using Northern blot analysis [1]. This technology was later almost entirely superseded by RT-PCR (reverse transcription polymerase chain reaction) [2]. The introduction of microarray technology in 1995 [3] dramatically increased the pace at which gene expression analysis was performed.

The DNA microarray is a high throughput biotechnology method, which allows detection of thousands of unique nucleotide sequences in a single parallel experiment. Although microarrays are a sophisticated modern high throughput experimental platform, they rely on a very simple and fundamental fact of molecular biology, which is the sequence specific nature of hybridization between complementary nucleic acid strands. As the overall cost of microarrays decreases, the technique has become more and more popular, and despite the emergence of

RNA-Seq as a competing technology [4], more microarray experiments were performed this year than in any preceding year. A common application of microarray technology is detection of gene expression by hybridization of mRNA transcripts to the microarray, but microarrays can also be used to detect DNA or gene products characteristic of a species for diagnostic purposes, and for comparison of uncharacterized bacterial species or strains with characterized genomes via array CGH [5, 6].

In a microarray experiment, a collection of known *probes* is used to specifically separate *target* molecules from an unknown mixture. Probes are deposited in known locations on the microarray slide surface, one probe to each spot. The target mixture is then labeled with fluorescent tags and applied to the slide surface. The amount of fluorescent signal at each spot on the microarray is detected using a laser scanner or imager, and the signal intensity is taken to approximate the amount of each specific target in the unknown mixture.

All living organisms access and utilize the data stored in their genomic databases through transcription of their genes into RNA molecules, some of which are later translated into proteins, which in turn may undergo further modifications. The double stranded, antiparallel character of the DNA molecule, which makes up most of the genetic material was discovered over half a century ago [7] and was the first step towards the understanding of nucleic acid hybridization.

The invention of the PCR (polymerase chain reaction) process occurred only in 1983 [8-10]. It was a revolutionary technological breakthrough, which gave

scientists access to DNA sequences of interest in unlimited quantities. Subsequent major innovations in PCR technology included the invention of reverse transcriptase PCR and real time PCR techniques [2, 11, 12]. From that time on, the quantitative detection of single gene expression was possible. However, that was just the beginning of the challenge. At the same time, sequencing technology was also revolutionized. As the sequencing of genomes from dozens of different organisms was completed, and the existence of thousands of different genes and a large number of regulatory elements was revealed, the amount of available gene sequence information about which expression data could be collected overwhelmed the capabilities of single-gene PCR-based expression assays. Single-gene expression methods also could not reveal complex regulatory relationships. A new approach was necessary to address the gene expression issue on a genomic scale. This led to development of the microarray technology.

The first microarray experiment was reported in 1995 [3]. The key idea behind microarray experiments was very simple. It relied on the same property of nucleic acids that was used in the first step of the PCR reaction: two complementary nucleic acid strands sooner or later will anneal to each other. However, the rest of the technology was quite different from either PCR or RT-PCR. The reaction volume was miniature and the array had the parallel assay power of many thousands of RT-PCR reactions. Known probe sequences are chemically or biochemically synthesized and then the molecules are attached to a solid surface: in the first experiments PCR products and UV-crosslinking were used. Next a solution mixture of fluorescently

labeled transcripts, or ‘ targets’, was added into the hybridization chamber. It seemed like a perfect tool for the genome-wide transcript analysis. Unfortunately, the same property that causes nucleic acids to form a double helix with a complementary strand, which made microarrays such a powerful tool, was also one cause of potential weaknesses of this technology. Single stranded nucleic acid molecules readily hybridize not just to their precise complementary sequences, but also to some less perfectly complementary, ‘non-specific’, nucleic acid sequences; depending on the sequence they may also hybridize internally, to themselves. Cross-hybridization of microarray probes to unintended targets contributes to the total signal [13-17], and there are suggestions in the context of other hybridization-based technologies [18-24] that formation of unimolecular structure may interfere with the intended hybridization.

Nevertheless, microarray technology is still a powerful and popular diagnostic and research tool, and until recently no technology has been able to provide similar access to genome-wide expression information. Many approaches have been tried to improve the quality of data produced using DNA microarrays. Very long cDNA probes used in early experiments [3, 25, 26] had ill-defined properties [16, 25, 27], and have gradually been superseded by use of 24-70mer probes, which have sensitivity similar to that of cDNA probes [28] while being relatively uniform in their other properties. While sensitivity decreases with shorter probe length [16, 29-31], oligonucleotides smaller than 70 nucleotides are the most frequently used; in comparative studies the loss in sensitivity becomes most obvious

for oligos below 35 nucleotides in length [32]. The major advantage in using a shorter 60-70 nucleotide gene specific synthetic DNA sequences (such as those produced by Agilent, Operon, etc.), instead of a bulky 500-3,000 nucleotide long transcript is more reproducible responses. Given some care in the design process, the probes on the slide surface were less folded and more accessible for hybridization. Nonspecific UV-crosslinking, which was at first used to attach the probes to the surface of the slide, caused damage to the probe DNA and rendered some of them inaccessible for the target hybridization. Alternative techniques of probe attachment to the surface were later developed, such as attaching a reactive group to one end and using contact or inkjet printing methods to deliver the probe to a slide having complimentary chemistry. Additional improvements have included the addition of various linkers to the end of the probe to be attached (such as poly-Lysine linkers) to avoid known electrostatic surface effects.

From the perspective of experimental design, microarray experiments have a serious, innate flaw: there are a vast number of variables and a small number of samples. Statistical analysis of microarray data has therefore been a highly active area of research and many competing models and methods have been published [33-37], which address the image analysis and noise reduction issues for the microarrays. Some of these are highly platform-specific, while others are more flexible and can be applied to analyze data from virtually any microarray platform. While development of such algorithms is invaluable, even the best of them are often unable to explain all the discrepancies that are observed in the data from the real

microarray slides. Statistical analysis is a necessary step to prove that observed data are indeed valid and reliable rather than obtained by chance. However, such analysis is not a panacea, and should not be used as a band-aid to cover-up a poor experimental design.

As of today, it is still unclear why some of the spots on microarrays fail to produce a good signal (false negative spots). Many biology labs and medical centers routinely carry out microarray experiments, and a large number of research laboratories continuously work on improvement of microarray data analysis tools. However, since the cost of doing quality control experiments is higher than the cost of developing new statistical models, the number of researchers working on the techniques for improvement of the quality of the microarray assay design remains comparatively small. That is, more attempts are made to fit the existing experimental microarray data to statistical models, than are made to experimentally test proposed variables arising from the biological, chemical and physical processes proposed to cause microarray spot failures. For example, the Affymetrix chip originally allowed one half of the chip space to be occupied by single mismatch probes in their 25-nt probe designs. These probes were used to estimate the highest possible background level for each perfect match probe and were intended to help standardize the data. At first, this design seemed like a direct way to eliminate all ambiguous spot signals. Unfortunately, despite the sacrifice of the 50% of the microarray chip area, some mismatch probes were found to produce brighter spots than their perfect match counterparts, creating another level of data ambiguity. The

use of this design persisted until in 2003 the riddle of bright mismatches was partially resolved by Naef and Magnasco [38], who discovered that the difference between the perfect match and mismatch probe intensities strongly correlates with the base in the middle position of the 25-meric probe. This study showed that the thermodynamics of probe binding on the high-density oligonucleotide arrays is very different from that of solution experiments. The results of this study clearly indicated that fluorescent labels severely interfere with the probe-target binding, often causing the perfect match probes to produce a weaker signal than their mismatch counterparts. Another outcome of the study showed that the thermodynamics of probe binding on the high-density oligonucleotide arrays is very different from that of solution experiments. Finally, the study also clearly indicated that internal fluorescent labels sterically interfere with the probe-target binding. Another problem caused by keeping the mismatch probes on the arrays is the possibility that a single nucleotide polymorphism variant maps to a position in the probe. This can cause either low specific binding to the perfect match probe or a high specific binding to its mismatch counterpart.

Exploration of the biological, chemical and biophysical factors that can affect the extent of probe-target hybridization on a microarray chip is necessary, or we will not be able to accurately model the effect of these factors on the measurements. Purely statistical data cleansing methods do not permit mechanisms to be revealed, leading to the persistence of design strategies that result in data artifacts.

1.2 Comparison of Microarray Platforms

As was mentioned earlier, not all microarrays are produced in the same manner. There are several microarray platforms for measurement of gene expression, which differ considerably in the array format, probe nature, slide chemistry as well as the recommended amounts of targets and hybridization time and temperature [39]. These platforms can be roughly grouped into 9 categories, mainly based on their manufacturer: Motorola CodeLink [31], Affymetrix [40, 41], Agilent [30], NimbleGen [42], ABI [39, 43], Febit [44], and Illumina [45] as well as Core lab – manufactured mechanically spotted cDNA and oligo arrays [46].

Affymetrix GeneChips are the oldest and most abundant commercial platform for determining transcriptional gene profiles. These are ready-to-use, high-density, short-oligomer arrays, which often yield highly reproducible results. All GeneChips have short 25-mer oligos built on the slide surface via the *in situ* chemical synthesis method called photolithograpy, based on solid-phase DNA synthesis reagents [41]. The Affymetrix microarray platforms vary according to the purpose and number of design revisions, but generally places from 4-11 perfect match probes per feature on a target molecule in order to optimize the signal to noise ratio [40, 41]. One of the advantages of the Affymetrix arrays, as well as the other commercially supplied platforms, is the automation of probe handling and bookkeeping, which reduces the risk of human error during probe and array production. Automation virtually eliminates the danger of mixing up the gene

products while handling the large gene clone libraries and massive number of different PCR products.

While the most frequently used, because of the gene coverage, GeneChips have several characteristics that make analysis challenging including the relatively short and thus insensitive probes, and the presence of mismatch probes on some designs. Probe length is not a trivial matter: while Affymetrix microarrays have a sensitivity of 1:100,000, Agilent arrays (using 60-mers) show a 10 fold higher sensitivity value of 1:1,000,000 and PCR product arrays (using 500-800-mers typically) have 3 times higher sensitivity, at~ 1: 300,000 compared to Affymetrix [39]. However, there are applications, such as SNP detection or exon junction detection, for which longer oligo probes are completely unsuitable [47-49]. Looking at the sensitivity of the Agilent arrays compared to the short oligo platforms such as Affymetrix, and taking into account that the oligo probe specificity increases with the probe length, it becomes clear that a long oligo platform will be more suitable for the purpose of our investigation in terms of both specificity and sensitivity. However, on the long oligo array platforms, which involve the use of the non-sheared target molecules, sequence specificity and high sensitivity can be hindered by the presence of the interfering secondary structure in probe and target, which may lead to partial unintended cross-hybridization.

Another commercial supplier is Agilent, whose stock arrays carry 60-mer oligo probes (shorter probes can be requested on custom designs), which are deposited onto the slide surface via an inkjet printing technology [50]. These arrays

generally have only one, and sometimes two, specific probe per gene placed on the slide surface. Longer oligomer probes typically form more stable probe-target duplexes during hybridization step than do shorter ones, because the binding energy for shorter probe–target hybrid complexes is lower [16]. Thus, using longer oligo probes increases the array sensitivity. To maintain high specificity requires considering both self-hybridization (internal structures) and cross-hybridization, the potential for which increases with increasing length. The high hybridization efficiency and specificity of the Agilent arrays was shown to be due to both steric and non-steric effects [30], including the presence of polynucleotide linkers, which hold the probe above the glass surface of the slide. It is impossible to place linkers on the Affymetrix platform as the photolithographic process builds right on the surface of the slide. Thus, the electrostatic effects from the charged surface of the slide become unavoidable and cause the surface end of the probe to become inaccessible. The charge or electric potential of the dielectric slide surface was shown to interfere with the hybridization especially under the low salt conditions [51]. Another study [30] suggested that the first ten to fifteen bases at the surface end of the oligo probe may not be accessible during the hybridization reaction at all possibly due to the electrostatic interactions. The oligo probes are also known to be prone to forming duplexes with non-helical properties on a positively charged surface [52]. For the purpose of our study of the variable probe binding site accessibility on a target it is crucial to ensure the probe access to the entire probe binding region. This was one of the major reasons why we choose the Agilent arrays

as our microarray platform.

While less used in large biomedical and model system studies than formerly, pin-spotted glass-slide microarrays are ubiquitous in basic research, and represent our own platform of choice. These arrays may be made with wide range of probe lengths and chemistries (from PCR products to linker coupled short or long oligos). The probes are spotted onto a functionalized slide surface using solid or hollow pins. A disadvantage is of the less controlled spot shape and grid layout produced by spotting robots [53, 54]. This class of arrays is invaluable to researchers working in non-model organisms, for which commercial array platforms may not be available, for small-scale experiments, and for studies focused on improving microarray technology. They allow for relatively low cost testing of probes, different hybridization parameters, such as time, temperature, various additives, linkers, mismatches, and of preparation methods for target mixtures and labels.

1.3 Physico-Chemical Factors That Affect Probe-Target Hybridization

A fluid filled microarray chamber is a closed reaction vessel, which contains a non-catalyzed chemical reaction, called probe-target hybridization or duplex formation. The laws of thermodynamics govern all of the processes inside this chamber including all intended and unintended hybridization reactions, whose interpretation rests on achieving thermodynamic equilibrium. A process will proceed only if it is energetically favorable, and there must be sufficient time for the system to achieve equilibrium. While solution equilibrium studies of nucleic acid hybridization have been done for 50 years [55], microarrays add the diffusion

limited probes and the large surface-volume interface as additional constraints. Given these changes, what are the physico-chemical factors that affect the probe-target hybridization on a microarray? How much is currently known about them? What can be done to improve the microarray design in light of these factors?

With respect to the probes, factors that have been studied include sequence contributions to sequence sensitivity and specificity [29], probe length and probe density [16], duplex melting temperature [56], G/C percentage of the probe sequence [57], probe location with respect to the 3'-end of the transcript [58], probe secondary structure [24, 59] and presence of mismatches in the binding region [60-62]. Since the effects of many of these factors are not well understood, a lot of probe design software vaguely acknowledge their importance by incorporating as their probe design criteria the parameters that supposedly account for the effect of these factors but in fact are inadequate. Synthetic oligonucleotide probes are designed using extensive computational analysis. These include sequence comparison against the full range of possible targets to ensure probe specificity, GC content analysis to approximate binding affinity among all of the probes on the array, and self-complementarity analysis to reduce the potential that the probe will fold into a stable unimolecular structure [63-65]. These features are discussed in more detail in the following sections.

1.3.1 Sequence Specificity

An ideal probe is specific to a single genomic or transcript target. Meaning that, under the real experimental conditions with several thousands of different

fluorescently labeled targets floating around, a truly specific probe should hybridize only to its designated target. Probe specificity is a rather sensitive quality, which decreases for both the shorter oligos and the very long gene transcripts. On one hand, the specificity of very short oligo probes goes down with decreasing probe length, because smaller oligos can hybridize to their unintended targets simply by chance [66]. On the other hand, the specificity of very long probes also decreases with this time with the increasing probe length, due to an increased risk of a part of the probe annealing to an unintended target [66]. The issue of microarray probe specificity is escalated by the fact, that the sequences of the real gene transcripts are far from being random nucleotide collections, meaning that two different targets can share regions of high sequence similarity. Papers by Kane [29, 67] and Zhou [68, 69] established empirical rules for balancing specificity and sensitivity that are still commonly used for selection of specific microarray probes. According to these rules, under standard hybridization conditions, a 50-mer probe has the potential to hybridize with any target sequence with which it shares 75 % sequence similarity or shares at least 15 consecutive bases. These sequence specificity rules (often called Kane's criteria) were specifically established for 50-mer probes, and therefore are not directly applicable to the probes of any other length. In fact, these rules were established in experiments using a very limited number of probes, but ever since have been widely accepted as dogma. Although, there have been other studies conducted in our research group indicating that perhaps Kane's rules are not strict enough to ensure real probe specificity (Dr. Gharaibeh unpublished work on

minimum nucleation). Similar studies were performed for the short oligo microarrays, which included the effect of point mutations on the probe-target affinity [70]. Other sequence specificity restrictions need to be taken into consideration (such as the presence of poly-Nt stretches, the influence of target secondary structure and free energy of probe-target hybridization [65]), and the rules should be completely revised for the short oligo probes as well as for the 60 or 70-mers.

1.3.2 Probe Length

One of the major advantages of synthetic oligo arrays is the ability to control the length of the probe. Depending on the experimental system and question a number of experiments have demonstrated that the optimum oligo probe length is between 50 and 70 nucleotides, although some arrays are designed with probes 27-35 nucleotides long [16, 28, 29]. Very short probe length renders the probe specificity and leads to significant cross hybridization simply by chance. It has been shown that ideal probe length, which allows one to eliminate most of the probe stable secondary structure, but maintain sufficient probe specificity, is between 50 and 70 nucleotides [71-73]. In an expression microarray experiment, probes are designed to be specific to particular coding sequences in the genomes, while behaving in a thermodynamically uniform way. The uniform probe length is intended to assure that different hybridization reactions for the probes with similar biophysical parameters would simultaneously reach the equilibrium in the time

allotted for the hybridization under the same hybridization conditions across the microarray chip.

1.3.3 Duplex Melting Temperature

Although a uniform probe length does help to achieve uniform hybridization profiles, from a biophysical point of view a uniform melting temperature for the probe-target duplex formation and released Gibbs free energy are much more important parameters by which to characterize the uniformity of hybridization across the microarray. Probe G/C content can be used to approximate the duplex melting temperature, and is much less elaborate to calculate. Therefore, it is almost always included in the algorithms for the microarray probe design. G/C content outside the 50 ± 5 % range is not desirable, because it imposes more limits on the available probe sequence space, although this is modulated by sequence actually present in the target genome. In conjunction with other factors it may increase the chance of non-specific probe-target hybridization. For species with high or low G/C content it may become a limiting criterion for the probe selection.

1.3.4 Probe Secondary Structure

Presence or absence of competing secondary structure in a probe molecule is a crucial factor in success or failure of the entire microarray experiment. Although most schematic representations of microarrays show probe molecules as nice straight poles sticking out from the slide surface, real microarray probes may look rather different. Depending on attachment chemistry, density and length [74-76] they do not stick out upright but tend to bend and bind to almost any kind of

aromatic ring or positively charged ion that is sufficiently close. A common artifact comes from self- annealing, forming complex internal secondary structures. One of the major advantages of short oligo arrays is that it is easier to choose a probe sequence that will have a minimal likelihood of forming intramolecular hydrogen bonds, and therefore will be virtually free of any kind of stable secondary structure. In most microarray design algorithms, self-complementarity is used as a proxy for true modeling of secondary structure.

Despite incorporation of these factors into the process of microarray probe design, some probes still fail to produce signal in the presence of target. The only conclusion is that there must be more factors and properties that can affect probe-target hybridization on the chip. Such factors may currently be poorly understood and therefore are not currently evaluated during the probe design process.

1.3.5 Thermodynamic Equilibrium

The target molecules (most often fluorescently labeled cDNA molecules, although cRNAs are the target in Affymetrix expression arrays systems) hybridize dynamically, in a reversible reaction, to the probe oligomers to form relatively stable double helices. All kinds of hybridization reactions between the probe and its designated target, the probe and an unintended target, two different targets in solution, as well as other interactions, which occur on the microchip are trying to reach equilibrium [77]. Therefore, for known concentrations of reactants the concentrations of reaction products can be predicted if all side reactions are known. Meaning that for the known concentrations probe and target the final concentration

of a heteroduplex can be estimated. In order to obtain the most accurate signal from the microarray spot, probes and targets should be allowed to hybridize long enough for the entire system to reach equilibrium [78]. In any chemical reaction the K_{eq} reflects the proportion of reactants that do not form product. The reasons could be very different: imperfect probe density on a microchip, instability of hybridization complex or presence of additives, which influence the hybridization capacity and kinetics [79]. According to one mathematical model, developed for heterogeneous DNA-DNA hybridization [80], there are two different mechanisms by which targets can hybridize with their complementary probes: direct hybridization from the solution, and hybridization of molecules that were first adsorbed nonspecifically to the array surface, and subsequently diffused across the surface until coming into proximity with a probe. It was shown that nonspecific adsorption of single-stranded DNA on the surface followed by two-dimensional diffusion significantly enhances the overall hybridization rate [81]. Heterogeneous hybridization depends strongly on the rate constants for DNA adsorption/desorption in the non-probe-covered regions of the surface, the two-dimensional (2D) diffusion coefficient and the size of probes and targets. The diffusion of the single stranded target NA is constantly interrupted by repeated association and dissociation with the immobilized oligonucleotide molecules. Experimental studies show that the hybridization efficiencies of 5'-end support-bound oligonucleotides are 75-80% for single-stranded oligonucleotide targets and 40-50% for long double-stranded targets,

respectively [82]. Other current studies support the idea that DNA hybridization occurs via a competitive displacement [83].

1.3.6 Ionic strength of hybridization solution

Presence of various salts and other chemical compounds alters both probe-target hybridization and the amount of structure present on these molecules [84, 85]. For example, the addition of formamide reduces the hybridization temperature and unwinds some of the stable secondary structure on both probe and target molecules. The ionic strength of hybridization solution is in part responsible for the specificity of probe-target binding on a chip. There are empirical corrections to hybridization equations, which account for DNA thermodynamics with different concentrations of sodium, magnesium, urea, DMSO and formamide.

1.4 Secondary Structure in Nucleic Acids

The secondary structure of a single stranded RNA and DNA molecule represents a collection of all hydrogen bonds between the nucleotide bases that can be represented in a plane.

1.4.1 Biophysical Properties of Nucleic Acids

It is a common knowledge that nucleic acid molecules are polymer chains made of 4 types of bases each: adenine, guanine, thymine and cytosine (in case of the DNA, and may carry modifications like C-methylation), and adenine, guanine, uracil and cytosine (which often carry additional modifications in case of the RNA). The property of the nucleic acids to fold into condensed secondary and 3-dimensional structures is attributed to their ability to form additional hydrogen

bonds using the sugar, which can be broken and rejoined relatively easily by specific enzymes as well as mechanical or chemical forces or high temperature.

The problem of predicting the 3-dimensional structure of a RNA molecule or a single stranded DNA molecule lays in the fact that even at equilibrium these molecules can simultaneously form a whole ensemble of structures in a solution with the correct dielectric constant, particularly aqueous solutions, with the number of possible structures depending on the length and sequence of the nucleic acid. The predominant structure at equilibrium has the lowest ΔG energy and is considered the most stable, 'optimal' structure. Suboptimal structures have higher ΔG values. Suboptimal structures need not resemble the optimal structure. The library of structural motifs for DNA and RNA includes the Watson-Crick, and non-Watson-Crick interactions, internal and terminal mismatches, dangling ends energies, hairpins and bulges as well as internal loops and multibranched loops under several salt and temperature conditions. These were described in a number of publications by the Turner, Mathew and SantaLucia research groups at the University of Rochester and Wane State University [86-92] along with the nearest neighbor parameters for predicting stability of nucleic acid secondary structure. This data is summarized in the Nearest Neighbor Parameter Data Base [88].

1.4.2 Multi-State Hybridization Model

Often given very little consideration for the purpose of commercial speed and probe selection algorithm simplicity the thermodynamics of probe-target hybridization is the only force that drives the entire microarray experiment to

success [65, 84, 92, 93]. One obvious issue with applying these structure prediction methods to microarray hybridization simulation is due to the probes being restricted by one end on a solid support, causing the change in the overall thermodynamics of hybridization [94]. In addition to equilibrium state, the kinetics of a chemical reaction can affect the measurements. The rate of hybridization reactions depends on such thermodynamic parameters as the Gibbs free energy of duplex formation and the free energy of probe and target folding thus bring up the issue of the microarray hybridization kinetics.

The simplest kinetic model that describes the probe-target hybridization assumes the ideal world situation, in which specific targets find their designated specific probes with no side products or degradation:

$$\text{Probe} + \text{Target} \Leftrightarrow \text{Hybrid}$$

This model is called the two-state approximation model for hybridization, and the equilibrium constant K_{eq} for the two-state model can be calculated from the following equilibrium equation:

$$K_{eq} = \frac{[\text{Hybrid}]}{[\text{Probe}][\text{Target}]},$$

where [Probe], [Target] and [Hybrid] are the concentrations of probe, target and hybrid molecules at equilibrium correspondingly. The equilibrium constant K_{eq} characterizes the nature of a particular hybridization reaction and is independent on the total species concentration, however it changes with temperature, salt concentration and pH, and can be affected by the presence of different additives that affect the solvent, such as the DMSO, betaine, glycerol and formaldehyde. The

equilibrium constant can be predicted for every two-state hybridization reaction from the following equation:

$$K_{eq} = e^{-\frac{\Delta G_T^o}{RT}}.$$

The Gibb's free energy for a nucleic acid to form a duplex from a random coil under the standard state conditions can be calculated from the following formula, given that the ΔH° and ΔS° are accurately predicted by the nearest neighbor model:

$$\Delta G_T^0 = \Delta H^0 - T \times \Delta S^0.$$

Thus, knowing the K_{eq} we can solve the equilibrium equation and find the fraction of bound targets for the two-state hybridization model.

In reality there are many hybridization and folding reactions proceeding simultaneously on the chip all reaching their equilibria at different times, and competing with the intended probe-target hybridization reaction. These other reactions will decrease the concentration of specific hybrid formation. Therefore, a multi-state equilibrium should always be analyzed, when the non-independent multi-component systems are simulated. The reactions which occur on the chip during the microarray hybridization include, but are not limited to, the following processes:

- formation of the stable secondary structure on the target molecule (especially in the probe binding site or in its vicinity),
- probe secondary structure formation (although probes are almost always pre-screened against the self complementarity, there are rarely completely secondary structure free),

- non-specific target hybridization (will overall add the strength to the spot signal intensity, causing the false positive results),
- hybridization of the specific target to a non-specific probes (reduces the concentration of the free specific target),
- specific target homodimer formation (reduces the concentration of the free specific target as well), and
- heterodimer formation between a specific target and any other target species (reduces the concentration of the free specific target as well).

With short oligonucleotide probes and currently accepted probe densities, the concentration of probe homo- or heterodimer formation is negligible; the multi-state equilibrium that occurs during the microarray hybridization has the following reactions involved:

$$
\begin{array}{ccccc}
\mathrm{Probe}^{folded}_{specific} & & \mathrm{Target}^{folded}_{specific} & & \mathrm{Target}^{folded\,/\,unfolded}_{specific\,/\,non-specific} \Leftrightarrow \mathrm{Homo-/Heterodimer}^{NNN}_{target} \\
\Updownarrow & & \Updownarrow & + & \\
\mathbf{Probe}^{unfolded}_{specific} & + & \mathbf{Target}^{unfolded}_{specific} & \Leftrightarrow & \mathbf{Hybid}_{specific} \\
+ & & + & & \\
\mathrm{Target}^{unfolded}_{non-specific} & & \mathrm{Probe}^{unfolded}_{non-specific} & & \\
\Updownarrow & & \Updownarrow & & \\
\mathrm{Hybrid}^{N}_{non-specific} & & \mathrm{Hybrid}^{NN}_{non-specific} & &
\end{array}
$$

where *N*, *NN* and *NNN* refers to a number of possible hybrids resulting from the interactions with the different probe and target species. Every one of the reactions on the diagram above has its own equilibrium constant, which governs its particular reaction and contributes to the changes in free and unfolded probe and target concentration for the specific reaction.

1.5 Experimental Evidence for the Secondary Structure Effects in Other Nucleic Acid Based Platforms

The nucleic acid secondary structure has been of a concern in a number of different experimental platforms. However, there have not been a lot of experimental attempts to prove that the stable secondary structure on single stranded DNA or RNA is in fact affecting the quality of obtained results. A raise in studies addressing the issue of accessibility of folded single-stranded nucleic acids was due to a boost in the RNAi technology with the goal of targeting appropriate sites for siRNA. The impact of target RNA secondary structure in RNA interference experiments was described in a number of publications [95-97]. The results from these experiments suggest that the binding of siRNAs is affected by the stable stem-loop structures on target RNA. A unique and rapid method for determining the accessible sites on RNA molecules was described by Allawi et al. [98]. This method is independent of the target length and does not require any labeling. The accessible regions are determined via RNA hybridization to sequence-randomized libraries of DNA oligonucleotides, which are later extended using a reverse transcriptase and PCR amplified. This method although very fast and simple allows to map the regions of RNA accessible for hybridization, which most often, but not always, coincide with the regions of the stable secondary structure. Other technologies, such as DNA-templated organic synthesis showed that the ideal sequences for the this method should avoid having both heavy secondary structure or no secondary structure at all [99]. A surface plasmon resonance biosensing study, which specifically addressed

the influence of the secondary structure on DNA hybridization demonstrated the significant interference [100].

1.6 Secondary Structure Modeling

1.6.1 Secondary Structure Modeling Algorithms

RNA and DNA structure is very sensitive to a solution's temperature, ionic strength and the presence of other molecular species. A variety of computational algorithms were created to address this task, all of them have their advantages and disadvantages [101-106]. The most precise experimental ways to obtain a 3-dimensional image of a single stranded nucleic acid is by obtaining its structure via X-ray crystallography or NMR spectroscopy. The first requires that there be a regular enough structure to form a crystal and may not reflect the primary solution structure, so NMR studies are preferred. However, the amount of time and cost needed for such investigations does not scale to testing large scale RNA/DNA structure predictions. Therefore, a number of computational algorithms were developed in the recent years to model for modeling the structure of nucleic aid molecules from their base sequence.

It has been known for a while that the structure of the single stranded nucleic acid molecules is hierarchical, meaning that there are several levels of structural organization: the primary, secondary, tertiary and quaternary [107]. The sequence of nucleotides represents the primary level of structural organization, and it comes from sequence databases. The next level of organization is the secondary structure, which includes, but is not limited to, the canonical Watson-Crick base pairing. It is

the secondary structure that is primarily responsible for the stability of the entire conformation [108-112], while the tertiary structure represents the overall 3-dimensional conformation stabilized by the hydrogen bonds and other intramolecular interactions, and the quaternary level includes the interactions with the other molecular species (frequently electrostatic). It is often not necessary to know the tertiary structure in order to accurately predict the stable secondary structure formation.

There have been several publications reviewing the nucleic acids' secondary structure prediction algorithms, which show the overall trend in the secondary structure predictions from the energy minimization models [113] towards the kinetic models [114] and resolution of pseudoknots [115]. There are three classic approaches for prediction of nucleic acid structure from its sequence: the thermodynamic, comparative and hybrid methods. The comparative sequence analysis is the oldest method. It uses the sequence alignments to identify the compensatory base changes. The thermodynamic approaches involve the calculation of an energy model for an RNA secondary structure and search for the RNA structure with the lowest free energy, while the hybrid approaches combine both thermodynamic and comparative analysis information in order to predict the secondary structure of the RNA.

In a thermodynamic method of a secondary structure prediction, a free energy of folding of a particular structure is calculated using a nearest-neighbor model. The model assumes that the binding and stacking energy of a particular base

pair depends on the identity of a neighboring base pair [87, 92]. The dynamic programming algorithm is used as a computational method to find the lowest free energy conformation in the thermodynamic structure prediction, as the number of possible structural conformations grows exponentially with length [116]. The dynamic programming algorithm allows for checking the energy of formation for all possible structures without actually generating these structures. The calculation proceeds in two steps: first, the lowest free energy is calculated for all short and then long sequence fragments until the minimum free energy for the entire sequence is determined, and then the lowest energies are traced back to compute the exact structure with the lowest energy of formation. The accuracy of the energy minimization method has been studies extensively by and revealed the 72% accuracy, for the cases where pseudoknots were not taken into the account [117]. Today the free energy minimization approach for the secondary structure prediction took a step forward and can be used to predict some pseudoknot formations at the expense of computational time and general applicability [118-120].

1.6.2 Secondary Structure Modeling Software

One of the most respected software tools, that utilizes the dynamic programming algorithm and the nearest neighbor thermodynamic parameters, for prediction of nucleic acid folding and hybridization, taking the approach of free energy minimization, is Mfold by M. Zuker [117, 121]. This RNA and DNA structure prediction software accounts for canonical as well as non-canonical Watson-Crick

base pairs, and assigns energy penalties for various types of loops, bulges, mismatches and dangling ends. One of the reasons for the wide use of Mfold is the existence of the experimental data in support of the Mfold predictions. The major drawback of this software is that it weights each possible structure in the ensemble of optimal and suboptimal structures that a molecule can form equally. This means, that bonds that form only in rare conformations are considered equal to the bonds, which are present in the lowest-energy structure. Therefore, it is possible that, at some positions in the molecule, the secondary structure is over weighted. The Mfold software has been recently expanded to the form of a hybridization prediction server called HYBRID, which computes the partition functions for systems containing two molecules in solution that can fold as well as hybridize with each other. The calculations are performed over a wide range of temperatures, and the nucleotide accessibilities are measured as a fraction of all optimal and suboptimal structures, in which the nucleotides are found in a single-stranded conformation. The RNAstructure software also uses a dynamic programming algorithm with an energy model based on thermodynamic parameters for the nearest neighbor model [87, 117, 122].

The Vienna RNA package [123, 124] is another package that performs RNA secondary structure prediction through the energy minimization. This software uses three kinds of dynamic programming algorithms for structure prediction: the minimum free energy algorithm to yield a single optimal structure, the McCaskill partition function [105] algorithm to calculate base pair probabilities in the

thermodynamic ensemble, and the suboptimal folding algorithm [125] to generate all suboptimal structures within a given energy range of the optimal energy. The package also performs secondary structure comparisons, using either string alignment or tree-editing [126]. The drawback of this software is that it is not applicable to DNA folding calculations. The general advantage of the dynamic programming algorithms is that they can find the global minimum free energy of an RNA sequence relatively fast. The disadvantage of these algorithms is that not all energy rules (such as pseudoknots) can be incorporated using the dynamic programming paradigm. In addition, the kinetic properties of the reactions are completely ignored. Sfold [127, 128] is a non-commercial software package for prediction of probable RNA secondary structures and nucleotide accessibility based on the Ding and Lawrence algorithms [129] for RNA folding. The Sfold program assigns accessibility based on an ensemble-weighted average of secondary structures and gives the probability of a secondary structure at equilibrium. This algorithm uses the Turner free energy rules [130-132]. The Sfold and the RNAfold programs perform the secondary structure calculations a lot faster than the Mfold. The Pfold web server, by Knudsen and Hein [133, 134], predicts the RNA secondary structure using stochastic context-free grammars and computes a maximum likelihood secondary structure for the given alignment. It finds a consensus secondary structure given a sequence alignment. Another state of the art nucleic acid folding platform is Visual OMP [92, 135]. This software system utilizes the most recent nearest neighbor parameters from J. SantaLucia, which have never been

published and are considered proprietary. This program suite allows an experimentalist to take into account the hybridization environment (such as presence of formamide, DMSO and different concentrations of various ions). Unfortunately, the exact algorithm utilized in this software is also a proprietary secret and remains a mystery for the end users. The license for the Visual OMP is rather expensive. There are several structure modeling approaches that combine thermodynamic and comparative information: the Bayesfold method [136], which uses a Bayesian approach to compute base-pair probabilities given the mutual information, fraction of complementary base pairs, and the average RNAfold pairing probabilities; the Juan and Wilson method [137] that scores a potential base pair by using a linear combination of terms originating from the RNAfold thermodynamic structure predictions, a co-variation score in the alignment, and a correction term for loops of different lengths, and the ILM web server [138, 139] with the adjustable relative weight of a thermodynamic and comparative scores.

In our experience, the use of the information about the accessible sites present on the RNA surface obtained by Allawi et al. [98] using the extendable sites method for evaluation of the Mfold, Sfold and Vienna RNA structure prediction software showed several local correlations in the different regions of the structure, while occasionally disagreeing with some or all three of the predicted structures. It is important to keep in mind that some of the software can provide the information about base pairing probabilities, while all described experimental methods for

evaluation of the predicted RNA structures can only give an “accessible” or “not accessible” type of answer.

1.7 The Issue of Stable Target Secondary Structure

1.7.1 The Secondary Structure Of Target Molecules

Similar to the structure of probes interfering with target binding, structure in target will interfere with probe binding. Every target molecule carries a certain degree of a secondary structure, but not all of this structure has a potential to interfere with the probe-target binding. The major concern occurs with respect to the stable structure formations being present in the active probe binding sites. Since the probe binding sites are always parts of a much larger and complex molecule, their nucleotides are able to produce intramolecular hydrogen bonds with the nucleotides from the neighboring target regions. Some studies have shown that at least the 170 bp regions surrounding the probe binding sites on the target should be taken into consideration, when trying to evaluate the target accessibility [140]. Thus, despite having the complementary sequences, microarray probes and their binding sites on the target have the ability to form the secondary structure of considerably different abundance and stability. Our published computational analysis [141] has shown that these structures can be so stable that they neither unfold completely when the target mixture is heated to 65˚C, which is greater than the common microarray hybridization temperature, nor disappear completely when the target molecules are sheared down to 50 nucleotide long fragments, the shortest that completely matches the probe length in our experiments. 3-dimensional

structures such as hairpins and stacked regions have the potential to block regions of the target molecules from hybridizing to their intended probes. Our modeling result showed that such formations would not convert completely to a random coil with increasing hybridization temperature, more extensive shearing, or both [141]. These secondary structures are very likely to interfere with the annealing of the targets to their intended probes, resulting in a misinterpretation of the amount of target present in the sample.

1.7.2 Potential Effects of Target Secondary Structure on Microarray Hybridization

Microarray experiments are chemical reactions and any given probe-target pair represents a simple thermodynamic system. However the sample mixtures are very complex, and the array taken as a whole is a complex reagent. The potential for side reactions is very large. An intramolecular structure in either probe or target that occludes the interaction site effectively decreases the concentration of that species available to form the duplex product. The longer the target molecule, the higher chance that stable secondary structures can form. Because target length is highly variable, both intrinsically and from sample-handling differences, probe selection methods usually only model probe structure. This assumption is often justified by saying that the probe and the probe-binding site on the target have complementary sequences, and therefore, should have an equal amount of self-complementary sequence that would facilitate folding into stable structure. In reality, the probe-binding region of the target is a part of a much larger and like more stable target structure. Occasionally it could be buried deep inside this

structure. Also, the nucleotides in the probe-binding region of the target (unlike the nucleotides of the probe) have a choice to either bind to other nucleotides in the probe-binding site or bind to the nucleotides in the target sequence outside of the probe-binding site. Thus, freeing a probe-binding site from the secondary structure during the hybridization on the chip is very different from unwinding the probe secondary structure and requires more energy.

1.7.3 Testing the Effects Target Secondary Structure on Probe Hybridization

Based on the predicted physical properties of probe and target, my hypothesis is that the presence of the stable secondary structure on the target molecules affects the probe-target binding on the microarray enough to alter the hybridization signal from the spot. Secondary structure-free probes with uniform biophysical properties will have the probe-binding sites occupied by the secondary structure to a different extent, and therefore, will produce signals of a different intensity. We have showed that this situation is what standard models of single-stranded nucleic acid folding predict, for a microarray designed to detect gene transcripts from *Brucella suis* 1330 [141]. We will test this hypothesis by performing a computational simulation of the microarray experiment using multiple probes designed for the same target, placing probes in both predicted folded regions, and in secondary structure free regions. We predict that this effect will be more prominent when RNA targets are used rather than the DNA, and will decrease with the length of the target sequence. Substances that alter the dielectric constant of the solvent, like formamide and DMSO or use of a higher hybridization

temperature, will reduce secondary structure, but at the cost of sensitivity since this changes the position of the equilibrium towards the reactants. We will test the effect of each of these factors in a computational simulation.

CHAPTER 2: SECONDARY STRUCTURE IN THE TARGET AS A CONFOUNDING FACTOR IN SYNTHETIC OLIGOMER MICROARRAY DESIGN

A few years ago while working on a microbial genome comparison project, we got involved in design of a diagnostic microarray for the three pathogenic species of genus *Brucellae.* These bacteria are well known for their high G/C content: around 65%. The commercial microarrays for *Brucella suis* 1330, which were purchased by our collaborators, were designed using a software program called *pick70* and had 70-mer oligo probes attached to the slide surface. Our previous experience with the *Brucella suis* 1330 genome showed that it was practically impossible to create even a 60-mer array for this microbial organism which would contain at least one secondary structure-free probe per gene. The major reason was a high G/C content of the genome sequence of this particular species. However, creating a 50-mer array for *Brucella* was much less difficult. Thus, getting suspicious about the effectiveness of the newly purchased microarrays and being highly interested in the degree of secondary structure present on the target molecules, we decided to conduct a secondary structure modeling simulation using a genome wide set of *Brucella suis* 1330 open reading frames available through GenBank [142].

For this purpose we used a widely accepted nucleic acid structure modeling software called Mfold [117, 121, 143-145] to predict the presence of stable

secondary structures in target molecules from a complete ORFeome of *Brucella suis* 1330 (3262 CDSs). We modeled stable 3D structure formation in both cDNA and cRNA targets with temperatures and solutions related to the actual hybridization conditions (solvent characteristics) used for performing the microarray experiments, and examined a range of lengths (pseudo-sheared) of fragments, from 50 nt up to full length transcripts. The results of our modeling study were published in BMC genomics [141] and are attached as an appendix.

All modeling simulations were performed at four different temperatures: 37, 42, 52 and 65°C. The stabilities and extent of the target secondary structure were normalized to the global mean full target length of 851 nt. As expected, the average free energy change on global secondary structure formation in *Brucella suis* 1330 was a lot smaller for the full DNA transcripts rather than the full RNA transcripts and ranged from -130 to – 40 kcal/mole (for the DNA at 37 to 65°C) and from -270 to – 110 kcal/mole (for the RNA at 37 to 65°C). The extent of the target secondary structure was calculated for each nucleotide based on the presence of the stable secondary structure in >50% of the predicted conformations. The fraction of internally H-bonded and therefore inaccessible nucleotides in the full length *Brucella suis* 1330 transcripts ranged from 65 to – 55% (for the DNA at 37 to 65°C) and from about 55 to 30% (for the RNA at 37 to 65°C).

The accessibility of the original 70-mer probe binding regions in the targets was calculated at different fractional accessibility cut-offs (25, 50 and 75%). It was shown that the average number of inaccessible bases within the probe-binding site

greatly resembled the normalized data obtained for the full length targets. Since genus *Brucella* is known to have a higher than average GC content (of about 67%), small sets of transcripts were modeled for e organisms having more balanced GC content (such as *Escherichia coli*: GC% of 50%) and an AT-rich genome (*Lactococcus lactis:* GC% of 37%). Even there, our modeling results indicated the likely presence of abundant secondary structures.

An extensive shearing simulation was performed for the 627 bp long *ureG-1* gene transcript from *Brucella suis* 1330. The random shearing of the target was modeled for fragments of 200, 100 and 50 nt long. Sheared fragments were chosen starting at every 10th residue. When random shearing of the target was modeled, an overall destabilization of the secondary structure was observed. However, shearing did not eliminate the stable structures completely.

Overall, our computational study showed that the stable secondary structure is highly abundant on the full length target molecules, and that the actual probe-binding sites reflect the same amount of the secondary structure as the full length molecules. This property seems to persist across a range of temperatures for both DNA and RNA, with the stability getting lower as the temperature rises. This conclusion suggested that the actual probe-target hybridization process may be greatly affected due to the probe binding site being actively involved in the intramolecular hydrogen bond formation. However, the obtained data represented purely computational results prone to various algorithm and approach related errors. An experimental study is needed to obtain direct evidence to whether the

stable secondary structure on targets actually interferes with the probe binding. Knowing to what extent the structures in the probe binding site can affect the actual hybridization will help to develop this nucleic acid property into a criterion for the microarray probe design software, and for correct interpretation of the measurements.

2.1 Abstract

2.1.1 Background

Secondary structure in the target is a property not usually considered in software applications for design of optimal custom oligonucleotide probes. It is frequently assumed that eliminating self-complementarity, or screening for secondary structure in the probe, is sufficient to avoid interference with hybridization by stable secondary structures in the probe-binding site. Prediction and thermodynamic analysis of secondary structure formation in a genome-wide set of transcripts from *Brucella suis 1330* demonstrates that the properties of the target molecule have the potential to strongly influence the rate and extent of hybridization between transcript and tethered oligonucleotide probe in a microarray experiment.

2.1.2 Results

Despite the relatively high hybridization temperatures and 1M monovalent salt imposed in the modeling process to approximate hybridization conditions used

---

This chapter is adapted from Ratushna *et al*. [141. Ratushna, V.G., J.W. Weller, and C.J. Gibas, *Secondary structure in the target as a confounding factor in synthetic oligomer microarray design.* BMC Genomics, 2005. **6**(1): p. 31.]

in the laboratory, we find that parts of the target molecules are likely to be inaccessible to intermolecular hybridization due to the formation of stable intramolecular secondary structure. For example, at 65°C, 28 ± 7% of the average cDNA target sequence is predicted to be inaccessible to hybridization. We also analyzed the specific binding sites of a set of 70mer probes previously designed for *Brucella* using a freely available oligo design software package. 21 ± 13% of the nucleotides in each probe binding site are within a double-stranded structure in over half of the folds predicted for the cDNA target at 65°C. The intramolecular structures formed are more stable and extensive when an RNA target is modeled rather than cDNA. When random shearing of the target is modeled for fragments of 200, 100 and 50 nt, an overall destabilization of secondary structure is predicted, but shearing does not eliminate secondary structure.

2.1.3 Conclusion

Secondary structure in the target is pervasive, and a significant fraction of the target is found in double stranded conformations even at high temperature. Stable structure in the target has the potential to interfere with hybridization and should be a factor in interpretation of microarray results, as well as an explicit criterion in array design. Inclusion of this property in an oligonucleotide design procedure would change the definition of an optimal oligonucleotide significantly.

2.2 Background

Sequence-specific hybridization of a long, single-stranded, labeled DNA or RNA target molecule to shorter oligonucleotide probes is the basis of microarray

experiments. In such experiments, gene specific *probe* molecules are either synthesized in situ or are printed to the microarray surface, and are either non-specifically cross-linked to the surface or are attached specifically using a method such as amino or epoxide reactive groups to surface coatings of poly-Lysine or otherwise functionalized surfaces [146, 147]. *Target* molecules (most often fluorescently labeled cDNA molecules, although cRNA and aRNA are used in some protocols, and labeling methods vary) hybridize transiently to the probe oligomers until they form more stable double helices with their specific probes [80]. At some point, the rate of on and off reactions balance and the reactions reach equilibrium, and the concentration of the target in the sample solution can be calculated. Transcript abundance is assessed by estimation, based on the relative intensity of signal from each spot on the array. This interpretation of array data relies on the assumption that each hybridization reaction goes to completion within the timeframe of the experiment, that the behavior of all pairs of intended reaction partners in the experiment is somewhat uniform, and that labeling is equal for all species and samples have equal numbers and types of cells.

There are three major types of DNA microarrays, which differ in the approach used for probe design: Affymetrix type microarrays [148], which assay a single strand, usually in the same sense as the transcript, with a distributed set of 25-mer oligonucleotides, full length cDNA microarrays, in which long cDNA molecules of lengths up to several hundred bases are crosslinked to the slide surface to probe either sense or the antisense strand of a transcript copy and with very

complex binding interactions [3], and synthetic long-oligomer probe microarrays, which usually assay each transcript only once, in either the sense or antisense form. The last class of microarrays encompasses a variety of commercial and custom platforms, and the optimal probe length depends on the particular experimental design. Oligo lengths ranging from 35 to 70 nucleotides have been shown to perform well under different conditions [29-31, 149], though recent studies have shown that oligomers of up to 150 nucleotides may be desirable for assessing transcript abundance [16]. In general, the use of synthetic oligomers has been shown to result in improved data quality [71, 72] relative to cDNA arrays, and 70mers have been shown to detect target with a sensitivity similar to that of full length cDNA probes [28]. Short probes have been promoted because they facilitate finding unique sequence matches while forming fewer, and less stable, hairpin structures and because they display more uniform hybridization behavior overall, and they are sensitive to small sequence changes when such are needed, as with SNP arrays [150] and splice junction arrays [151]. However, the need for sensitivity and detection of rare transcripts drives the use of long-oligonucleotide arrays in gene expression experiments [16, 152]. In this study, we have modeled the accessibility of transcripts to hybridization with 70mer oligonucleotides.

A number of oligonucleotide design software packages have been published in recent years, each having design strengths in one of a number of criteria [64, 153-156]. Several factors are considered by almost all microarray design software packages: in particular, the sequence specificity of the probe-target interface and the

overall balance of GC content across the array. Unique regions of the target sequence are identified using sequence comparison methods; the unique regions become the search space for probe selection based on other criteria. The number of probes per sequence and location of the probe in the sequence also restrict sequence availability. A relatively uniform melting profile generally is achieved simply by selecting for probes with similar GC content and uniform or close-to-uniform length, although some design methods explicitly compute the duplex melting temperature for each candidate probe-target pair and filter unique probes to find those which match a specified range of melting temperatures. Another biophysical criterion that is sometimes applied is the elimination of probes having the ability to form stable intramolecular structures under the conditions of the experiment. This is usually done by eliminating regions of self-complementarity, although at least one design program [64] does explicitly compute the melting temperature of the most stable structure to form in the probe molecule and uses that information to filter out stable secondary structures in the probe.

Few of the available array design packages explicitly consider the possible structures of the transcript-derived molecules in the sample solution and their impact on whether the microarray will provide an effective assay, although the OligoDesign web server [156] does compute this information for use in design of locked nucleic acid probes. It has been shown that a hairpin of as little as six bases in an oligonucleotide can require a 600-fold excess of the complementary strand to displace the hairpin even partially [157]. Since the target molecules are generally

longer than the probe and may be of a different chemistry, it is not sufficient to conclude that their behavior will mirror that of the complementary probe. Prediction of secondary structure in a sample transcript using a standard nucleic acid secondary structure prediction algorithm (Mfold) demonstrates that while longer-range interactions are reduced at high temperatures, stable local structures persist in the transcript even at high salt concentration and high temperature (Figure 1). Because unimolecular reactions within the target can occur on a much shorter timescale than the diffusion-mediated, bimolecular, duplex hybridization reaction, competition for binding by intramolecular structures is expected to kinetically block the specific probe annealing sites on the target sequence in some cases and result in misinterpretation of the signal obtained from the assay if these effects are not taken into account.

In order to estimate the prevalence of stable secondary structure in long target molecules, and thus the impact such structures might have on the analysis of microarray data, we have modeled secondary structure formation in mRNA transcripts of the intracellular pathogen *Brucella suis*. We have assessed the stability of structures formed in the transcript and the accessibility of the binding sites of optimal probes generated using commonly applied design criteria. Because random shearing of the full-length target molecule is used in some protocols, we have also modeled the effects of shearing to an average length on the prevalence of secondary structure in selected targets.

2.3 Methods

Prediction and thermodynamic analysis of secondary structure was performed for all protein-coding gene transcripts predicted from 3264 CDSs in the *Brucella suis* 1330 genome. *Brucella suis* has a relatively high (57%) genomic GC content. *Brucella suis* was chosen for this experiment because our collaborators have previously acquired a custom synthetic oligomer microarray for this organism, developed using standard oligo array design software, and we have access to both target sequences and to a set of unique probe sequences that define the interaction sites for which expression results have been obtained by the laboratory.

In order to determine whether Brucella sequences form atypical structures we randomly picked and analyzed 50 gene coding sequences from compositionally balanced genome (*Escherichia coli*), and 50 from the GC-poor genome of the nonpathogenic AT-rich gram-positive bacterium *Lactococcus lactis* (35% genomic GC content). The *Brucella suis* genes ranged in length from 90 to 4,803 bp, with an average transcript length of 851 bp. The *E. coli* genes ranged in length from 140 to 2,660 bp, with an average transcript length of 792 bp. The range of GC content in the genes chosen was 37% to 57% with an average value of 50%, which is reasonably representative of the *E. coli* genome. The *L. lactis* genes chosen ranged in length from 140 to 2,730 bp., with an average transcript length of 765 bp., and ranged in GC content range from 30% to 42% with an average value of 35%.

2.3.1 Microarray Design

70-mer probes for each *Brucella suis* target were previously designed (Stephen Boyle, personal communication) using ArrayOligoSelector (pick70) [153]. ArrayOligoSelector uses sequence uniqueness, self-complementarity, and sequence complexity as criteria but does not explicitly evaluate ΔG of secondary structure formation for the probe. 72% of the probes designed using this method were found to contain secondary structures with melting temperatures greater than 65°C, and 10% contained secondary structures with melting temperatures greater than 80°C. The Brucella probes defined the interaction sites within the target transcripts for which structural accessibility was evaluated.

2.3.2 Secondary Structure Prediction

Probe and transcript secondary structure were predicted using the Mfold 3.1 software package [117, 121]. Mfold identifies the optimal folding of a nucleic acid sequence by energy minimization and can identify suboptimal foldings within a specified energy increment of the optimum as an approach to modeling the ensemble of possible structures that a single-stranded nucleotide molecule can assume. We modeled secondary structure in the single-stranded target, modeling the target both as DNA and as RNA, at a range of temperatures which is inclusive of hybridization temperatures commonly used in microarray protocols: 37°C, 42°C, 52°C and 65°C. The modeling conditions were chosen within the allowed settings of Mfold to approximate a microarray experiment: solution conditions of 1.0 M sodium concentration and no magnesium ion were used. The free energy increment for

computing suboptimal foldings, ΔΔG, was set to 5% of the computed minimum free energy. The default values of the window parameters, which control the number of structures automatically computed by Mfold 3.1, were chosen based on the sequence length. Free energy changes on formation of secondary structure were extracted from the Mfold output.

2.3.3 Accessibility Calculation

Accessibility in folded single-stranded DNA or RNA has recently begun to be addressed in a few experimental studies, mainly with the goal of targeting appropriate sites for RNAi. Because the structure of single-stranded nucleotide molecules is much more dynamic than that of proteins, with each molecule likely to exist in an ensemble of structures, and because the 3D structure of these molecules is rarely known, there is not yet a consensus representational standard of per-residue accessibility for single-stranded nucleic acids. Ding et al. [128, 158] implement probability of single-strandedness, when the weighted ensemble of likely structures is taken into account, as an accessibility criterion. However, use of their Sfold server, with batch jobs limited to 3500 bases, is not currently practical for a genome-scale survey of accessibility. Another approach to accessibility prediction is McCaskill's partition function approach [105] which can be used to compute base pair probabilities and summary pairing probability for any base. This approach is implemented in RNAFold [159], a component of the Vienna RNA package.

In this study, we chose to use the less physically rigorous approximation of probability of single strandedness as a simple fraction of predicted optimal and

suboptimal structures in which a residue is found to be part of a single stranded structure, as computed by Mfold. Accessibility scores derived from MFold predictions have been used in limited studies of RNA structure focused on hammerhead ribozymes [160], antisense and siRNA targeting [161, 162] and have been shown to be predictive in cases where some experimental measure of accessibility has been made[163]. While MFold-derived accessibility scores may not be completely optimal, they have been used with reasonable success to predict accessibility in the siRNA targeting context, and so we use MFold here.

2.3.4 Shearing Simulation

Random shearing of the target mixture is an approach that is often offered as a solution for the problem of target secondary structure. The actual content of a sheared mixture of DNA or RNA fragments is complex. Shearing breaks the molecule not in predictable locations, but in random locations that give rise to a distribution of fragments around an average fragment length. In order to simulate the effects of different degrees of shearing on structure formation and stability in a transcript, we picked fragments of 200, 100, or 50 bases in length, choosing the start position via a sliding window of 10 bases. Secondary structure prediction for all fragments derived from every transcript in the B. suis genome is computationally intensive and produces an extremely large amount of output. Since our initial goal was to determine how much the method would affect the number and type of secondary structures probes would be expected to bind the shearing simulation was performed for fragments derived from the 300 bp Ure-1A gene of *B. suis*. Secondary

structure and thermodynamics were computed for each of these fragments individually.

2.4 Results

2.4.1 Extent and Stability of Target Secondary Structure

Our modeling results obtained for the genome-wide set of intact single-stranded DNA or RNA targets demonstrate that stable secondary structures are widespread in target mixtures from *Brucella suis* (Figure 2) and in randomly chosen transcripts from the genomes of *E. coli* and *L. lactis*. Figure 2 shows the ΔG of formation for the most stable predicted secondary structure of the full-length transcript, as a function of reaction temperature. The major energy components of the MFold ΔG are hydrogen bond energy and base pair stacking energy. These can be assumed to have a roughly linear relationship with transcript length. In order to make energies from different-length transcripts comparable, energies were normalized by computing a per-residue folding ΔG for each transcript and then multiplying that value by the global mean target length, for all transcripts considered from all organisms, of 851 bp. Average ΔG of secondary structure formation decreases with increasing temperature, but even at 65°C, the average ΔG of secondary structure formation for a full-length transcript is -98.2 kcal/mol (-27.9 kcal/mol when modeled as cDNA), meaning that the transcript is quite stable in that structure and a considerable energy input will be required to displace or melt the remaining structure. The trend in ΔG of secondary structure formation from the high-GC genome of *B. suis* to the low-GC genome of *L. lactis* is a decrease in overall

stability. The average normalized ΔG of secondary structure formation for transcripts selected from the GC-balanced genome (*E. coli*) is near 70% of the average for *Brucella*, while the average ΔG for transcripts from the GC-poor genome (*L. lactis*) are even lower (30% at 52°C). However, even in the most GC-poor genome, stable target secondary structure in the single-stranded target is widespread.

Our results demonstrate that a significant fraction of nucleotide sites in the average target mixture, whether single stranded DNA or RNA, will be found in stable secondary structure under the hybridization conditions used in oligonucleotide microarray experiments, and will be relatively inaccessible for intermolecular interactions. Figure 3 shows the percentage of nucleotides that are in a double-helical state in at least 50% of the secondary structure conformations predicted by MFold, at various reaction temperatures. The measure of accessibility used is the fraction of structures in which a nucleotide is found in a single-stranded conformation, when all optimal and suboptimal structures predicted are considered.

Figure 4 is a plot of the average ΔG of structure formation when shearing of the target molecule is simulated by dividing the target into overlapping 200, 100, and 50mer fragments. Shearing the target into smaller fragments destabilizes secondary structure, especially at very short fragment lengths. However, shearing does not eliminate occlusion of nucleotides by secondary structure, even in the shortest fragments examined. When a DNA target is modeled at 52°C, for example, the double stranded fraction decreases by only about 30% – from 41% to 29% –

when the target is simulated as sheared into 50mer fragments. However, in hybridization experiments involving low copy number targets and longer oligos, creating extremely short target fragments may reduce or eliminate the signal on the chip, because the target can not be sheared specifically to present an unbroken hybridization site for the probe, and so some fragments will be created that match the probe only partially.

2.4.2 Interference of Secondary Structure with the Hybridization Site

Figure 5 shows the average percentage of nucleotides within a probe binding region in the target that are inaccessible, when different fractional accessibility cutoffs are used to classify the sites. Even when a relatively demanding criterion – double-strandedness in over 75% of optimal and suboptimal structures – is used to classify a nucleotide as inaccessible, an average of 21 ± 13% of nucleotides in the probe binding region are found in stable secondary structures at 65°C. Figure 6 shows a representative transcript and the challenge it presents to hybridization when modeled as full-length cDNA and fragments of various lengths.

2.5 Discussion

Lack of bioinformatics tools that incorporate experimentally validated biophysical properties of nucleic acids as a criterion for synthetic oligomer probe design is a major challenge for do-it-yourself microarray designers. One biophysical characteristic, which we predict will reduce the binding efficiency of microarray probes to their targets, is the propensity of long single-stranded DNA or RNA molecules to form stable secondary structure. 3-D structures such as hairpins and

stacked regions have the potential to pre-empt target nucleotides, thus blocking regions of the target molecules from hybridizing to their intended probes. Prediction and thermodynamic analysis of secondary structure at a range of temperatures in full length target sequences, as well as in subsequences formed by *in silico* shearing, revealed the likely presence of stable secondary structures in both full-length target and sheared target mixtures. These structures do not convert completely to random coil with either increasing hybridization temperature, more extensive shearing, or both. These secondary structures may therefore compete with the intended target for effective probe annealing in a microarray experiment, resulting in a misinterpretation of the amount of target present in the sample.

2.5.1 Applying Target Secondary Structure as a Criterion in Array Design

Based on the results of this *in silico* experiment, secondary structure prediction in the target is being used to develop a new criterion for oligonucleotide probe design. Our results from this modeling experiment demonstrate that the implicit assumption used until now – that eliminating probe secondary structure by avoiding self-complementarity eliminates target secondary structure as well – is valid only when the target and probe are of the same length. Use of target secondary structure as an explicit criterion will allow for masking or preferentially avoiding the regions of the target sequence in which base pairs are directly involved in secondary structure formation, to eliminate these regions from the sequence for the purpose of the search for the optimal probe.

In this study we have assigned accessibility scores to sites in the target sequence based only on the fraction of predicted structures within 5% of the energy optimum, in which a residue is found in a single-stranded conformation. While this measure is not too computationally intensive to compute, and can be applied to genome-scale problems using readily available software (MFold), it is not the most physically rigorous definition of accessibility. By equally weighting each possible structure in the ensemble of optimal and suboptimal structures that a molecule can form, it is possible that secondary structure at some positions in the molecule is overcounted; bonds which form only in rare conformations are considered equal to bonds which are present in the lowest-energy structure. The program Sfold [127, 128, 158] assigns accessibility based on an ensemble-weighted average of secondary structure. The program RNAfold [159], part of the Vienna RNA package, implements McCaskill's partition function approach [105] to arrive at pairing probabilities for each pair of bases in the sequence, from which a summary per-base accessibility can be derived. These methods are more rigorous than MFold and we expected they might produce somewhat different results, although it has also been shown that predicted binding states from MFold optimal structures perform almost as well as SFold and RNAFold predictions when applied to molecules of known 3D structure [127].

When we compared MFold-based accessibility predictions for an individual transcript to those generated by SFold and RNAFold, we found that the difference in average predicted accessibility over an entire transcript is small. We computed

accessibility for the transcript of human ICAM-1, which has been mapped experimentally to determine its accessibility [98]. The average fractional accessibility derived from MFold results is about 3–4% greater than that predicted by RNAFold or SFold. Therefore use of this fractional accessibility measure will not impose an unnecessary constraint on the design process relative to other predictive approaches. The accessibility profiles calculated for ICAM-1 using each method are shown in Figure 7. In each section of the figure, antipeak locations (having lower pairing probability and therefore likely to be more accessible) can be compared to the extendable sites detected by Allawi et al [98], which are indicated by green dots at the bottom of the plot. In each prediction, there are a number of apparently correct predictions and obvious errors, and it is not clear which method is yielding the best results at the residue level. A systematic, competitive test of these predictions against solution accessibility data gathered on various experimental platforms is called for, although available data sets for validation are still rare. In the absence of such validation, the MFold accessibility predictions are sufficient to predict the scope of the secondary structure problem in a genome-based array design, even if some details of the prediction are not correct. An experimental approach will eventually be required to determine which approach best represents the conditions of the microarray experiment.

2.5.2 Loop Length and Other Considerations

In this study, we focused specifically on the DNA/RNA base pairs that are actively involved in hydrogen bond formation. We realize that other accessibility

considerations will have to be added to the scoring scheme in practice. The structure of a long single stranded DNA or RNA molecule can contain many nucleotides that, while not part of a double-helical stem, remain inaccessible to hybridization due to their location inside small loops within the target secondary structure. A loop is a somewhat constrained structure as well, and the length at which it presents accessible sequence that favors hybridization has been shown to be on the order of 10 nucleotides and longer [161], while nucleotides found in shorter loops may be classifiable as inaccessible. However, there is a need for quantitative hybridization experiments that would elucidate how loops and loop-like structures in tethered long-oligo probe and target molecules affect the performance of assays, and we have chosen not to formulate a system for scoring the accessibility of single-stranded loop structures or weighting this criterion relative to the double-strandedness criterion until we have carried out some of these experiments.

Development of a target secondary structure criterion for oligonucleotide array design is expected to impose restrictions on the probe selection beyond the sequence similarity and melting temperature criteria that are currently used, especially in cases where short probe length restricts the annealing temperature used in the hybridization protocol to 22–37°. In the *B. suis* example, use of a low annealing temperature, e.g. 42°C which is the temperature used in some published 70-mer array experiments [28], would result in only about 30% of the average transcript being accessible for intermolecular hybridization, not counting 'free'

bases found in short loops in secondary structures. There will be greater design latitude for experiments carried out at higher hybridization temperatures. Recommended hybridization temperatures for long synthetic oligomer arrays may prove to be closer to 65°C, when only 50% of a typical RNA transcript or 30% of the corresponding cDNA molecule remain inaccessible.

2.5.3 To Shear or Not to Shear

We have shown here that while shearing reduces overall ΔG of secondary structure formation for individual molecules in the target solution, shearing does not in itself eliminate formation of secondary structure in single-stranded DNA or RNA. The question of whether shearing should be used for long oligomer arrays is still an open one. While some signal may be gained by reducing the stability of secondary structure in the target molecule, random shearing by its nature creates a mixture of targets that may have substantially different affinities. For instance, in a 300 nt transcript that is targeted by a 70mer oligonucleotide, there is nearly a one in four chance that a random break in the sequence will occur within the target site for which the probe is designed. Short fragments may present a substantially different binding site, and therefore have a different binding affinity, than the full-length transcript that is considered when the probe is designed. Binding of a sheared 50mer fragment to a 70mer probe leaves a dangling end in the probe. A break very close to one end or the other of the target site may create a target that still binds to the probe, though with reduced affinity; a break closer to the middle of the target

site may produce fragments that bind partially to the probe, competing for binding with perfect matches.

2.5.4 The Utility of Experimentally Validated Biophysical Criteria

In other experimental contexts where hybridization is critical to success, the impact of secondary structure in single stranded polynucleotides on results has been recognized and is now being systematically studied [97-100]. Intramolecular folding of mRNAs is so extensive that only 5–10% of most transcripts is accessible to binding of complementary nucleic acids; however the modeling of long molecules has not proven to give very accurate binding predictions [59, 164, 165]. In fact, array-based screens have been utilized to empirically select oligonucleotides that bind effectively to transcripts for siRNA experiments [165, 166]. Several studies have demonstrated that, at 37°C and 0 mM Mg2+ oligonucleotides of length >20 yield good binding/RNAseH digestion at low concentrations relative to shorter oligonucleotides (30 nM vs 300 nM compared) and found that microarray binding was a good predictor of siRNA activity despite the 3' tethering and 1M NaCl used in array experiments vs siRNA experiments [166]. Systematic "scanning" of mRNA sequences with libraries of short oligos [167] has also been shown to be successful in locating sites for siRNA targeting; however, such methods are likely to become extremely expensive if applied to the large number of targets in a microarray design. We have begun to develop an experimental approach to this problem, in which structure predictions like those used in this study are experimentally evaluated to

determine whether the structures we can predict using existing modeling approaches will detectably affect signal in the microarray context.

2.6 Conclusion

The results of the current study suggest a significant role for target secondary structure in hybridization to oligonucleotide arrays, which will warrant further investigation. Oligonucleotide probe binding sites in a significant fraction of transcripts are found in double-stranded conformations even in cases where self-complementarity was avoided during the probe design process. We find that at 52°C, for example, approximately 57% of probes designed for *Brucella* had binding sites in the target which were predicted to contain a stretch of unpaired bases of at least 14 nt in length; at 65°C, that fraction increased to 93%. Based on these findings we would expect that at 52°C only 57% of our probes would encounter optimal conditions for hybridization and therefore would demonstrate the expected behavior in the experiment, where intensity is expected to scale with target concentration. We predict that the remaining probes, which have shorter, or no, accessible sequences, will exhibit modified binding behavior, and we plan to conduct experiments to characterize this behavior. We have shown conclusively that avoiding self-complementarity in the probe when designing an oligonucleotide array is insufficient to eliminate secondary structure from the binding site in the target. By combining the procedure for systematic computational assessment of transcript accessibility described in this study with selective experimental validation of the impact of predicted accessibility on hybridization, we will develop

a useful criterion for avoiding troublesome secondary structure when designing microarray targets.

CHAPTER 3: EQUILIBRIUM SIMULATION OF DNA HYBRIDIZATION ON THE TARGET SECONDARY STRUCTURE MICROARRAY

3.1 Abstract

In this chapter, we describe the computational steps used to design a DNA microarray optimized to test the impact of secondary structure interactions and predict the probe-target hybridization levels on this array.

We used a genome-wide set of open reading frames from a bacterial organism *Brucella melitensis* 16M to select a set of targets with specific predicted folding properties, and then design long oligo probes to test the target secondary structure hypothesis. Even probes designed to be secondary structure free do not necessarily fall into the secondary structure free binding regions on targets. Our hypothesis states that the abundance and stability of the secondary structure in the probe-binding regions of targets prohibits complete binding to probes, causing them to produce a reduced spot signal. We produced a set of test microarrays, each carrying five experimental and one control probe per target, aimed at binding in the regions of varying secondary structure abundance and stability. These sequences were submitted to Agilent, and microarrays were produced. In addition, a subset of five targets and probes from the main array was used to produce a run of miniarrays, for the purpose of determining how microarray hybridization conditions affect the stability of the predicted folded structures, including the

temperature, salt and target concentration various chemicals affecting the dielectric constant of the solvent.

This study presents the results of a number of computational simulations predicting the competitive behavior of a complex mixture of target DNA on the *Brucella melitensis* 16 M target secondary structure miniarray. The study was performed using the OMP DE software [92, 135], which utilizes advanced thermodynamic prediction methods for modeling nucleic acid hybridization. The goal of this investigation is, by using computational simulation, to predict the effects of the target secondary structure and hybridization conditions, on the microarray readout. These predictions can then be tested using the printed arrays.

Our hypothesis is that probes that avoid regions with the stable secondary structures on the target would significantly outperform those for which there are more competing structures.

3.2 Introduction

3.2.1 The Importance of the Target Secondary Structure

Molecular assays for sequencing and gene expression by necessity use large DNA and RNA molecules as their substrates. Secondary structure, especially short-range formation of stable hairpin structures [140, 168, 169] including the exceptionally stable RNA hairpin loops containing four unpaired nucleotides [170, 171] has been found to affect the outcome of many experimental protocols, which rely on sequence specific hybridization. For example, short secondary structure motifs, which occur in a 20-mer oligo probes can render the probe insensitive by

decreasing hybridization-based signal up to 50-fold [140]. A range of non-microarray based experimental hybridization platform studies showed that presence of the stable secondary structure interferes with outcomes dependent on specific nucleic acid binding.

In classical gel sequencing of such structured nucleic acids as tRNA and 5S RNA, so called 'compressions' were observed on the gel and were attributed to the formation of secondary structure, and nucleotide modifications were used to relax the GC rich regions of the cDNA [172]. Using pyrosequencing (a non-gel-based DNA sequencing technique) facilitates the analysis of DNA sequence compressions, which occur due to secondary structures in the DNA fragment during gel electrophoresis and can lead to misreading of the sequence [173].

It has been known for over a decade that presence of stem-loop structures on the PCR template, created by the presence of inverted repeat sequences from transposable elements, causes the polymerase to jump during PCR amplification, resulting in amplifications of shorter aberrant PCR products [174]. It was shown that at physiological temperatures, formation of DNA duplexes considerably slowed down the polymerization reaction, while the formation of triplexes arrested the reaction completely [175]. Conventional RT-nPCR amplifications of RNA transcripts containing an extensive secondary structure with a significant energy barrier predicted by the DNASIS software, produced amplicons missing the folded regions [18], due to the the intrastrand misalignment of repeats. In this study, adding DMSO (a known structure relaxing agent [176, 177]) in the reverse transcription step

allowed production of the full length amplicon after nested PCR. Competitive PCR studies, which allow the exact quantification of very small amounts of nucleic acids by co-amplification of a DNA template with known amounts of a competitor DNA, showed that GC-rich regions of stable secondary structures reduce the yield of the competitive PCR reaction by acting as pause or permanent termination sites [169]. The same study also showed that addition of betaine allows the polymerase to complete the amplification. Another way the presence of secondary structure alters the PCR amplification is by facilitating the formation of an artifact PCR band, containing a heteroduplex formed by two different single stranded gene-specific amplicons [178].

Another hybridization based platform affected by target secondary structure is RNA interference mediated by siRNAs and shRNAs. Even through the sequence characteristics of small RNAs are crucial to perform successful eukaryotic gene knockdowns, it was shown experimentally that secondary structure on siRNA and shRNA antisense strand also plays an important role [179]. This study concluded that RNA interference is more effective when there are several free bases at the 5 ‘ or 3’- end of the binding site. The computational analysis of RNAi libraries showed that there is a strong inverse correlation between the stability of antisense strand secondary structure and gene silencing efficiency [180]. Many other RNAi studies demonstrated that the secondary structure stability of the target RNA could be used to predict the efficiency of gene silencing [181-183]. An experimental study of shRNA target site accessibility from 100 endogenous human genes [184] revealed

that there is a strong correlation between siRNA binding site accessibility and the GC content of the site, however the target site accessibility is more critical for efficient gene knockdown than GC content. Investigations of RNA interference and microRNA pathways [185] found that functional miRNA binding sites are generally found in regions of high target accessibility. Large scale siRNA screening [186] revealed that targets with poor binding site accessibilities are difficult to silence beyond a 70% knock down rate. Unlike the microarray probe selection algorithms, there are programs for siRNA design that explicitly compute the accessibility of the target site interacting with the siRNA, such as RNAxs developed by Ivo Hofacker [183].

It is in the nature of single stranded nucleic acids to fold and form uniquely shaped arrays of structures, consisting of loops, stems and pseudo-knots [92, 107]. When occurring in nature, such structures help the RNA molecules to fulfill transport, structural and catalytic functions. The effectiveness of siRNA appears to be much less sensitive to the secondary structure present in its target mRNA in the living cells. The reason is the presence of helicases associated with the RNA-induced silencing complex, which unwind the target mRNAs and make them accessible to the incoming siRNA [187]. Nevertheless, it has also been shown [186] that the presence of secondary and tertiary structure on the mRNA transcripts, as well as the localization of the target sequence inside the cellular organelles, may inhibit or sometimes completely suppress access of activated RISC to the target sequences, resulting in failure of effective gene silencing in the RNAi experiments.

Artificial microarray mixtures do not incorporate unwinding enzymes. The abundance and stability of the DNA and especially RNA target secondary structure therefore raises a concern about the value of the microarray spot intensities [141]. There have been many statistical methods created specifically to address the problem of microarray data interpretation [35-37, 188, 189]. There are also many data cleansing protocols that rely on statistical tests to eliminate the most variable probes from a sample, in an attempt to classify probes responsive to factors other than the intended experimental factor [33, 35-37, 190]. Such clarification has value, but results in a loss of data. We believe that the best way to improve microarray technology is by improving platform design. In this chapter we describe the design of a microarray experiment aimed to test the effects of the target secondary structure on the data quality. The design procedures used to identify secondary structure free regions for this artificial microarray can easily be added as criteria in an array design pipeline if future experiments establish that this is a necessary design step.

3.2.2 The Model Organism and Its Genome

The study of the effect of target secondary structure on microarray data quality is of a fundamental character, and therefore can be performed on sequences sourced from any genome. We used *Brucella melitensis* 16M, due to commercial availability of a genomic library via Open Biosystems [191]. The library contains a total of 3,198 ORFs. The *Brucella* genomes generally have about 57% GC content, making them somewhat challenging to probe.

3.2.3 The Logic Behind the Target Secondary Structure Microarray

A decade after the introduction of microarray technology, the obvious biophysical criteria such as a uniformity of hybrid melting temperature as well as free energies of hybridization and folding [192], and probe [63, 64, 193, 194] and target secondary structure [195] are finally being introduced into microarray oligo design pipelines. Our investigation aims to further understanding of what makes a "good microarray probe", by turning the attention of scientists towards the biophysical nature of the microarray chip and the target. The target secondary structure was selected as being the most direct way to demonstrate the powerful effects of biophysical properties of nucleic acids on the quality and interpretation of microarray data. We designed a DNA microarray experiment to test differences in hybridization to structured and unstructured regions of the full-length unsheared target. We used a widely-adopted secondary structure modeling software to design sets of microarray oligomer probes of uniform thermodynamic properties and no secondary structure, which target binding sites having a variable amount of stable secondary structure.

3.2.4 Equilibrium Simulation of Microarray Interactions

A computational microarray simulation was performed to provide predictions as to the outcome of having probes whose target binding sites are involved in internal secondary structure, of varying levels of stability. Probe structure was minimized to limit the structural variable to one of the hybrid partners. Factors varied in the multi-state equilibrium simulation were: the probe

concentrations, temperature, formamide and DMSO concentration, and extent of the internal loop and bulge calculations. The uniqueness of this study lays in the fact that we make predictions based on the computational modeling prior to doing the microarray experiment, and therefore have a full control of our experimental platform. Investigations of microarray performance more often use a trace back approach in search of an explanation for unexpected experimental outcomes. This is partly due to limitations from the modern computational engines' ability to handle mixtures as complex as present in real microarray experiments. However, our target secondary structure miniarray experiment can be used to model multiple probe-target interaction at and beyond the selected hybridization temperature, and eventually replicated in the lab.

This study is designed to give computational predictions on whether the presence of stable secondary structure in the probe binding sites of a target has any effect on the extent of duplex formation between probe and target. These predictions will then be tested with the arrays that we have designed. Accurate responses are gauged by the response of probes designed to bind to target elements free of internal secondary structure. In case of a computational simulation we will be looking at specific target percent bound as a computational correlate to the spot signal intensity. There is sequence dependence to duplex stability, so we expect final target fractions bound to differ for different targets. This is one reason for having designed multiple probes at different sites per target, to allow statistical tests to be performed to establish the significance of observed response. For the purpose of our

investigation it is essential to establish whether there are any statistically significant differences between the target percent bound to the secondary structure free controls and experimental probe-binding sites, which did not occur by chance. The greater is the differences in intended target percent bound across the probe set, the stronger is the evidence in support of our hypothesis. Our simulation design allows investigation of the effects of several secondary structure destabilizing agents, such as DMSO and formamide at different concentrations.

Another important issue that we expect to clarify through these computational simulations is whether computational models of the most stable structure in an ensemble is sufficient to predict most observations. An accurate secondary structure prediction algorithm is the key to successful implementation of target secondary structure as a microarray probe design criterion. Since HYBRID 3.7 models only a two-state equilibrium, we will use another software called OMP DE [135, 143, 196], which is capable of modeling a multi-state equilibrium, to predict the target probe-binding site accessibilities under a range of the secondary structure destabilizing conditions.

We have also used the OMP DE software to obtain predictions of the fractions bound for the probe-target hybrids. We expect to obtain the highest percent bound from the control probes, designed to bind to the predicted secondary structure free probe sites on the target. However, it is possible that the obtained target fractions bound will be of a very similar or equal strength from all probes in the set (including the internal positive controls). Such a result would mean that the abundance of the

secondary structure at the probe-binding site has no effect on probe-target hybridization (at least at equilibrium). The OMP DE software allows experimental conditions to be modified, so we will perform simulations of a temperature series and try to establish whether there are any differences in equilibrium binding under more stringent conditions. Other structure relaxing factors, such as formamide and DMSO concentrations will also be adjusted. We will also perform a series of computational simulations to establish an optimal maximum allowable internal loop and bulge length.

Our goal is to establish a target secondary structure tolerance threshold for the microarray probe design.

3.3 Methods

We identified 96 candidate targets and six probes for each target. One probe per target was placed in a secondary structure free region and the rest in structured regions. This set of probes was printed on an 8-pack format Agilent array. Prior to carrying out the expensive large-scale assay, however, we chose five top probe-target sets and printed them in a "miniarray" format. The miniarrays are cheaper to produce and therefore we can optimize several factors including experimental temperature, salt concentration, experimental time-to-completion, and inclusion of chemical denaturants. The complex molecular interactions on the miniarray were modeled and the intended and unintended duplex concentrations at equilibrium predicted using the OMP DE software. The probe selection algorithm and the conditions of the miniarray simulation are described below.

3.3.1 Probe Selection Criteria for Target Secondary Structure Array

The flowchart describing the probe-target set selection algorithm for the target secondary structure microarray is shown in FIGURE 8. In the very first step of the probe set selection procedure we used an in-house pipeline, designed by Dr. Raad Gharaibeh (unpublished work) to generate a list of 709,860 50-mer probes from approximately 3,000,000 potential probes corresponding to the *Brucella melitensis* 16M ORFeome, which were pre-screened to satisfy to the probe selection criteria described by Kane et al. [29] for both forward and reverse strands. These criteria insure the probe specificity, by allowing no more than 15 consecutive bases to be shared between the oligo and any non-specific target, as well as limit the maximum probe percent identity to 75%. Candidate probes were initially selected using YODA software [197].

The following quality screens were then applied to insure uniform melting and binding affinity properties: the probe-target hybridization temperature was aimed to be around 60°C (with the probe T_m range of between 81.7°C and 87.9°C as calculated by MELTING [198] or 74.9°C and 80.2°C, when corrected for the chip). The probe GC percentage was set not to exceed 8 % and the maximum length of a homopolymer to be 4 or less. Sodium [Na^+] and magnesium ion concentrations [Mg^{2+}] were set to be 0.6 and 0.0 M respectively, as this conforms to the Agilent hybridization protocols.

After T_m filtering, the remaining probes underwent further screening using HYBRID 3.5 [199], which is a part of UNAFold nucleic acid structure prediction

package developed by M. Zuker, to ensure that they form no stable secondary structure. We selected only those probes for which $\Delta G_{\text{probe secondary structure formation}} \geq 0$. This reduces the set of candidate probes to 186,273 secondary structure free probes.

As is typical for high-GC genomes, secondary structure free probes could not be designed for all *Brucella melitensis* 16M coding sequences. Only 2,938 ORFs of a total of 3,198 had specific matches within our selected set of 186,273 candidate probes. Only 2,739 of these targets were found to have more than 5 secondary structure free probes. It was a crucial factor for our target selection criteria, as the experiment assumed placing a minimum of 2 secondary structure free probes per sequence, of which at least one must correspond to a secondary structure free probe binding site in the target.

3.3.2 Evaluation of Probe-Binding Sites' Accessibilities and Target Selection

Once the set of candidate probes was narrowed as described above, we again used the HYBRID 3.5 software [199] to make secondary structure predictions for 2,739 full length ORF transcripts as non-sheared DNA targets. Gateway vectors carrying two promoters located on the opposite strands outside of the cloned gene region and allowing producing both cmRNA and mRNA have been discontinued. Therefore, in view of producing a compatible RNA target mixture for the possible future target RNA studies using the same microarray we decided to place the targets on the sense strand.

We then attempted to select a target set for which 2 secondary structure free probes exist and for which completely secondary structure free probe-binding sites exist and are located on the target with no more than 15 base overlap. The requirement for two probes to be placed in secondary structure free regions resulted in elimination of over 80% of all open reading frames. We decided that forcing all sequences on the array to have 2 internal probes in secondary structure free regions would work as a pre-filter for targets with abnormally low secondary structure abundance, which was not the aim of this particular study. Therefore, it was decided to keep only one internal control probe per target.

We relaxed the selection criteria to include only one internal control probe per target. 1,141 targets had at least one secondary structure free region where an optimized candidate probe could be placed. Using a locally developed Perl script we selected five secondary structure rich probe-binding regions target. A restriction was set to prohibit any binding site overlap of more than 15 bp, to ensure that all 5 selected probes do not end up probing the exact same secondary structure region. As a result of the above calculations we created a list of a total of 628 targets, each with 6 unique probes of varying probe-binding site secondary structure abundance on the target.

This list of probe sequences and binding site accessibilities was sorted by the abundance of structure in the probe binding site. Every probe-binding region was assigned a secondary structure score, based on the number of nucleotide bases actively involved in the stable secondary structure formation, with a maximum

possible score per probe-binding region being 50 bp, and a maximum possible score per target being 50 x 5= 250 bp. The list of 96 sorted targets, their corresponding probe sequences, probe-binding site locations, secondary structure scores and the actual base accessibilities for the *Brucella melitensis* 16M array can be found in TABLE 2. HYBRID 3.5 software evaluates the propensity of every base to be single stranded in the optimal conformation. Every 0 in the probe binding site accessibility sequence represents a base, which is involved in intramolecular secondary structure formation, while 1 represents a free base in the minimum free energy target conformation. For example, probe BMEII0462_1 has 34 zeros in its probe binding site, its inaccessibility score is 34 (or 68 %). BMEII0462 leads the list of the most inaccessible targets in *Brucella melitensis* 16M, because its total inaccessibility score across the 5 experimental probes comes to 133 out of 250 (or an average of 53.2 % inaccessibility). The least inaccessible probe set for *Brucella melitensis* 16M array has probe-binding site accessibility score of 105 of 250 (or an average of 42.0 % inaccessibility).

3.3.3 Selection of positive and negative controls

The nature of this array allows for one internal positive control per every target: a secondary structure probe, which destined to bind to a secondary structure free target region. All names for the internal positive controls consist of the ORF name and number 0.

As a last step in *Brucella melitensis* 16M target secondary structure array design we had to pick a few negative controls, which serve to insure the accurate

background calculations. The preference was given to six 50-mer probes derived from *Arabidopsis thaliana*, which consequently underwent multiple nucleotide substitutions to insure that they do not compete for the to any of the 96 *Brucella melitensis* 16M targets. The specificity criteria were set to meet and exceed those described by Kane et al. [29]. We used the WATER package from the software suit EMBOSS [200] to perform the specificity calculations. WATER uses the Smith-Waterman algorithm [201] to calculate the local alignment of a sequence to other sequences and creates a standard EMBOSS alignment file. All six positive controls had a maximum target similarity of 44%, and a stretch of at most 11 of consecutive nucleotides occurring in at most 2 targets.

Our last negative control probe is the least prone to unintended target binding. It was derived from the linker sequences used by Agilent. The Agilent linkers are pre-selected to have low unintended binding qualities. We used the sequence for one such linker to perform what could be called a 'computational mutagenesis'. We introduced a series of mutations into the 50-mer sequence each followed by the similarity analysis using the WATER software as described above. The resulting probe had the maximum similarity score of at most 39% with any of the targets and a maximum consecutive stretch of 10 nucleotides.

The names and sequences for all 7 negative control probes as well as the probe-target similarity details are given in TABLE 3.

3.3.4 Miniarray Construction

To optimize the experimental conditions (such as the temperature, target concentration, presence of additives, etc.) and establish their most interesting ranges for studying behavior of the target molecules on microarrays we decided to design so- called miniarrays (microarrays with a small number of probe spots). For our miniarrays we selected five of the probe-target sets from the top of the *Brucella melitensis* 16M array list (six probes per target) and the 7 negative control probes. All 37 probes were synthesized by Operon and had amino-C6-linkers attached to their 5' end to allow the probe to float above the glass surface of the miniarray slide. It is generally acknowledged [24, 202] that when it comes to electrostatic interactions between the surface of the positively charged glass slide and the negatively charged nucleic acid probe, the C12 linkers pull the probes to a safer distance from the slide. However, the cost of having the C12 linkers synthesized and added to the oligo probes is considerable. Therefore, the decision was made in favor of adding the C6 linkers, which balance the cost and efficiency.

We have purchased a complete *B. melitensis* 16M ORFeome clone library from the Open Biosystems [191] for the purpose of producing full length transcripts first for the top 5 genes to be placed on the miniarray, and later for the rest of the genes in the full size target secondary structure microarray list. This clone library was created using the recombination based Gateway cloning system from Invitrogen Inc., which would allow a relatively easy produstion of the RNA target mixture at a later time. During construction of this library some of the cloning reactions were

performed on the undetectable amounts of amplicons [191], which resulted in some of the clones containing empty vectors. In addition to that, liquid cultures of transformant pools [191], rather than the inoculated and sequenced single cell colonies, were grown to generate glycerol cell stocks, resulting in PCR product mixtures being stored together in a single tube. Therefore, we re-inoculated and re-sequenced all ORF clones selected for use on the target secondary structure mini- and microarray. The large-scale plasmid DNA extractions were performed using the Qiagen Maxi Prep kits. New glycerol stocks were created for all 96 genes of interest, this time containing only the single cell inoculates, and the plasmid DNA was sent off for professional sequencing at the comfort read level to MWG The Genomic Company. The reads from regions deviating from the GenBank reference sequence were repeated until the ambiguity was resolved. Sequencing revealed multiple SNPs and other sequence aberrations, which most likely were introduced during the PCR amplification step. The sequencing result has proved to be of great importance. The SNPs were so abundant in the sequencing results, that they affected approximately every 1000th nucleotide in the target sequences, sometimes interfering with the probe-binding site on the target. Our five top targets were no exception; in fact a few of the point mutations fell inside our selected probe-binding sites. A number of genes had been truncated from 4 and up to 1092 nucleotides. Of the 96 sequenced gene clones there were several empty vectors with no gene insert at all as well as several vectors with the wrong gene inserts. Overall, the sequencing results required us to produce a new Gateway clone library, designed specifically to satisfy

the needs of our targets secondary structure microarray experiment, with all the glycerol stocks having the single cell origin, and the gene regions sequenced from the start to what would have been the stop codon in the genomic sequence (note, that all stop codons were removed during the cloning).

In order to maintain consistency between our computational predictions of probe-target binding interactions on the miniarray and their experimental validation, we incorporated the point mutation and deletion information revealed by sequencing into both the oligo probe and the target sequence. The probe-binding sites of the top five targets intended for the miniarray hybridization were carefully screened for point mutations or deletions. The modified target sequences were then refolded using HYBRID 3.7 and new probe-binding accessibilities recalculated. The mutated miniarray probe sequences and their binding site accessibilities are stored in TABLE 4, and the corresponding mutated target sequences are available in TABLE 5. The three modified probe names' were marked with the letter "_M" and their sequences were rescreened for cross-hybridization potential using WATER [200]. The small letter "m" in the probe and targets sequences indicates that a mutated full length target transcript was used to calculate nucleotide accessibilities. The base accessibilities in the probe binding site for pairing are represented as "1"s, and inaccessible nucleotides are shown as "0"s.

Designed as described above the miniarrays are aimed to serve a dual purpose: provide the essential information as to which of the biophysical characteristics are the most vital to explore more thoroughly in the full size

microarray format, and to be the base as well as provide the first experimental evidence to the computational simulation of the effects of the target secondary structure on the microarrays.

3.3.5 Modeling the Nucleic Acids' Interactions on *Brucella melitensis* 16M Miniarray with the OMP DE software

The OMP DE stands for Oligonucleotide Modeling Platform. This software is capable of calculating the thermodynamics of multi-state equilibrium for various types of nucleic acids, similar to the equilibrium that occurs in the microarray hybridization chambers. We used the OMP DE and its advanced thermodynamic parameters to simulate the behavior of target molecules and probe-target hybridizations at different temperatures and with and without the use of different secondary structure relaxing additives. All experiments simulated the hybridization reactions between five DNA targets and the corresponding virtually secondary structure free probes plus the seven negative controls on the *Brucella melitensis* 16M miniarray. The 6 probes in each set were intended to anneal to the regions of the varying secondary structure and therefore competed for the target. The simulations were performed for a six-state equilibrium, which included the probe and target folding as well as target-target homo- and heterodimer formation. The only prohibited interactions were between two probes, because the probes are attached to the surface and therefore assumed to be unlikely to interact. At temperatures below 55 °C, the OMP DE had reached its computational limitations in the simulation engine: too many intense calculations were required to simulate the amount of unimolecular folded structures that occur at low temperature, and the

numerical analysis (which calculates percent bound and concentration) broke down and the numbers no longer converged.

Hybridization simulation temperature range was set between 55°C and 65°C, sodium concentration was set as 0.6 M and magnesium concentration as 0.0 M. The initial calculations were performed assuming no structure-disturbing additives in the hybridization solution, and later series of 55°C assay simulations were performed including DMSO and formamide. SantaLucia's linear corrections to solution thermodynamics were used to allow for a better agreement between solution predictions and microarray experiments: surface slope ΔG = 0.85, surface intercept ΔG = 2.33, surface slope ΔH = 1 and surface intercept ΔH = 24.0 at 37 °C and 1M NaCl. Unfortunately, currently the OMP software does not allow accounting for the effect of the amino-C6 linkers, so this parameter had to be left out. The optimal energy thresholds for monomer, homodimer and heterodimer species were set to filter out all insignificant species with ΔG of greater than 0 kcal/mole. The maximum bulge and internal loop length for all structures was set at 35 bases. We performed a test run with the larger maximum bulge and internal loop length of 100 bp, but it was very computationally intense and had no effect on percent bound for the heterodimer species both optimal and suboptimal. We also allowed for some suboptimal structure calculations, with the maximum number of suboptimal structures for the monomer foldings set at 20, homodimer set at 10 and heterodimer at 12. For all monomer species the maximum energy of 10 kcal/mole that the ΔG of a suboptimal structure can be away from the optimal structure was

used. The values for the suboptimal energy window for the homodimer and heterodimer species were set at 1 kcal/mole and 5 kcal/mole respectively. Under these threshold conditions the optimal structures at equilibrium represent over 98% of all target conformations' ensemble. All effective target concentrations were assumed to be 200 pM, and all effective probes concentrations equal 20 μM. The TAIL_FOLDING function was set to TRUE for all participating species (monomers, homo- and heterodimers) forcing the secondary structure calculations to proceed after the specific binding has occurred.

Given that this study represents a computational simulation of the equilibrium processes on the miniarray, and not the actual experimental procedure or its multiplex kinetic model, we do not take into the consideration such factors as time, target degradation and label location and time-dependent loss of signal intensity, and assume all our oligos and targets to be of completely consistent length.

3.4 Results

3.4.1 *Brucella melitensis* 16M Target Secondary Structure Mini- and Microarray

We completed the probe design for a microarray aimed to specifically test the role of the target secondary structure in competitive probe-target hybridization based on *Brucella melitensis* 16M genome. This array was specifically designed to test the secondary structure effects on the DNA, and not the RNA targets, and contained 96 probe-target sets. There were six completely secondary structure free probes designed per each target sequence. One of these probes served as an internal

positive control and the other five were placed in the regions of variable abundance of the target secondary structure. Seven negative control probes were selected for the *Brucella melitensis* 16M array. The microarray was manufactured using the Agilent technology. The experimental probe sequences, corresponding binding site accessibilities and relevant annotations are given in TABLE 2, and the information about the negative controls can be found in TABLE 3.

A target secondary structure miniarray was created as a subset of the larger *Brucella melitensis* 16M microarray using the top 5 probe-target sets plus the negative controls. The probes were attached to the glass surface of the slides *via* the amino-C6 linkers. Both probe and targets sequences were adjusted to represent the actual sequences available in the *Brucella melitensis* 16M Gateway clone library from the Open Biosystems, rather than the GenBank reference sequences. The probe and binding site information is located in TABLE 4, and the adjusted target sequences can be found in TABLE 5.

3.4.2 Extent of the Target Secondary Structure in the Probe Binding Sites

The statistics for the analysis of abundance of the of the secondary structure in the probe-binding sites from the optimal folds of 96 *Brucella melitensis* 16M DNA targets is given in TABLE 6. It shows the maximum, minimum and average inaccessibility scores of the probe-binding sites due to the secondary structure formation as well as the maximum length of consecutive stretches of both accessible and inaccessible nucleotides. As it was described earlier, the miniarrays represent a subset of *Brucella melitensis* 16M microarray. They contain the top five probe-target

sets, which were adjusted to reflect the mutations discovered in the course of the clone sequencing. The corresponding statistical parameters for the *Brucella melitensis* 16M miniarray is given in TABLE 7.

3.4.3 Hybridization Simulations Using the OMP DE: Temperature Series

We performed a complete equilibrium simulation for the miniarray experiment with no secondary structure relaxing additives being added to the hybridization solution. The simulations took place at a range of temperatures from 55 °C to 65 °C with 1 °C as a window step. The graphical representations of all the folded target molecules and their binding sites were created using the Visual OMP and are attached at the end of this manuscript as FIGURE 9 for the BMEII0462m, FIGURE 10 for BMEII0874m, FIGURE 11 for BMEII0685m, FIGURE 12 for BMEI0267m and FIGURE 13 for BMEI0682m *Brucella* transcripts. FIGURE 14 is an example of a probe-target hybrid. In particular it shows the BMEI0267m BMEI0267m_5 heterodimer modeled at 60 °C using no structure relaxing additives.

Computational results of these simulations are summarized in terms of target percent bound in TABLE 8. The percent bound was calculated for both optimal and suboptimal target structures and is represented in TABLE 8 as both target percent bound in optimal conformation and as a total target percent bound for each of the target species. The table contains the simulation data only for the folded monomer target species (left at negligible concentrations after the reactions have reached the equilibrium) and the specific probe-targets heterodimers. The percentage bound for non-specific heterodimer species were present in minuscule

concentrations of less than 10^{-30} %, and therefore were excluded from any further considerations. The computational simulation results for the miniarray at the range of temperatures described above agreed with our predictions for three (BMEII0462, BMEII0685 and BMEI0267) out of the five selected targets. The total target percents' bound accumulated by probes with the secondary structure free binding sites from these sets at 55 °C were 98.94%, 99.64% and 90.45% correspondingly, and practically removes target from other probes in their set, supporting the idea that the stable secondary structure n the target indeed interferes with the probe target binding. Temperature increase caused some of the secondary structure in the probe binding regions to relax, which is reflected in the TABLE 8 by gradual decrease in the total target forming a heteroduplex with the internal control probes, and increase in total target percent bound from the other 5 heteroduplexes for these three probe-target sets. The total (optimal plus suboptimal) target percent bound for BMEII0462, BMEII0685 and BMEI0267 at 60°C hybridization temperature ranged between 75.42 % and 88.52 %, and 63.69% to 65.88 % of target 5°C above the typical 60°C hybridization temperature.

FIGURE 15, FIGURE 16 and FIGURE 17 are a graphical representation of the computational hybridization results obtained using the OMP DE software for the miniarray hybridizations at 55°C, 60°C and 65 °C respectively based on the total target percent bound (optimal plus suboptimal structures). Under our computational conditions, the optimal lowest energy target conformations represented over 98% of all possible target structures in the ensemble.

The other two DNA targets (BMEII0874 and BMEI0682) had the highest percentage bound with one of the experimental probes, which contained some degree of the secondary structure. These results were also persistent across the entire temperature range. Further detailed probe sequence analysis revealed several possible reasons as to why these probes gathered more target molecules than their corresponding controls. For example, the probe-binding site for BMEI0682m_1_M gave highest percent bound for the BMEI0682m, when modeled with the OMP DE. A careful look at the probe sequence of BMEI0682m_1_M reveals two high energy GC clamps at both 3' and 5' ends of this probe (see TABLE 4). Such clamps are known to affect PCR primers but are not typically screened from probe design software, despite reports that a 4-G homopolymer creates abnormally bright spots on some arrays [203].

3.4.4 Hybridization Simulations Using the OMP DE: Formamide Series

From early experiments in solution biophysics and with microarrays [204, 205] it is known that changing the dielectric constant of the solvent affects that stabilizing ability of H-bonds. A common additive is formamide [206], and the OMP DE software includes a modeling parameter which allows simulation of the presence of formamide. We used OMP DE software to perform a complete equilibrium simulation for the miniarray experiment at varying concentrations of formamide as a secondary structure relaxing agent included in the hybridization solution. The simulation was performed at 60 °C in the presence of 0%, 5%, 10% and 15% formamide, with $[Na^+]$ held at 0.6M, $[Mg^{++}]$ at 0.0 M, and effective probe and target

concentrations kept at 20 μM and 200 pM respectively. The computational results of these simulations are summarized in terms of target percent bound in TABLE 9 at the end of this manuscript. The table contains the simulation data only for the folded monomer target species (left at negligible concentrations after the reactions have reached the equilibrium) and the specific probe-targets heterodimers. The non-specific heterodimer species were present in minuscule concentrations at less than 10^{-30} % and therefore were excluded from any further considerations. The percent bound was calculated for both optimal and suboptimal target structures and is represented in TABLE 9 as both target percent bound in optimal conformation and as a total target percent bound for each of the target species. FIGURE 15, FIGURE 18, FIGURE 19 and FIGURE 20 are graphical representations of the computational results obtained using the OMP DE software for the miniarray hybridizations at 60 °C using 0%, 5%, 10% and 15% formamide respectively based on the total target percent bound (optimal plus suboptimal structures). FIGURE 21 illustrates the BMEII0462m target under the 10% formamide conditions at 60 °C.

The computational simulation results for the miniarray at 60 °C at varying formamide concentrations described above agree with our predictions and show that the addition of secondary structure altering agents such as a formamide relaxes the secondary structure in the probe binding site and causes the predominant association of target with the internal control probes for BMEII0462, BMEII0685 and BMEI0267 to be reduced as the targets now readily hybridize to the other probes in the set. Raising the formamide concentration from 0% to 15 % causes the

targets to bind more uniformly to all probes in the set. However, at 15% we begin to observe a large percentage (9.5%) of BMEII0462m existing as a monomer species. At 20% formamide OMP DE fail to detect any probe –target interactions, meaning that at least within the simulation the formamide diminishes the strength of H-bonds to the extent that stable duplexese do not form under these conditions (this is a well-recognized fact that is implicit in the protocols that generally decrease the hybridization temperature several degrees for each percent of formamide added to a reaction).

3.4.5 Hybridization Simulations Using the OMP DE: DMSO Series

We also modeled the effect on probe-target of a second dielectric-constant altering agent, dimethylsulfoxide (DMSO). We once again used the OMP DE software to perform a complete equilibrium simulation for the miniarray experiment at varying concentrations of DMSO included in the hybridization solution. The simulation was performed at 60 °C in a presence of 0%, 2%, 5% and 8% DMSO. The computational results of these simulations are summarized in terms of target percent bound in TABLE 10. The table contains the simulation data only for the folded monomer target species (left at negligible concentrations after the reactions have reached the equilibrium) and the specific probe-targets heterodimers. The non-specific heterodimer species were present in minuscule concentrations at less than 10^{-30} % and therefore were excluded from any further considerations. The percent bound was calculated for both optimal and suboptimal target structures and is represented in TABLE 10 as both target percent bound in optimal conformation

and as a total target percent bound for each of the target species. FIGURE 15, FIGURE 22, FIGURE 23 and FIGURE 24 are a graphical representation of the computational results obtained using the OMP DE software for the miniarray hybridizations at 60°C with 0%, 2%, 5% and 8% DMSO respectively based on the total target percent bound (optimal plus suboptimal structures). FIGURE 25 illustrates the relaxed secondary structure on BMEII0462m under 5% DMS0 at 60 °C.

The computational simulation results for the miniarray at 60 °C at varying DMSO concentrations described above closely resemble those obtained in the formamide series and agree with our predictions relaxing the secondary structure in the target's probe binding diminishes the amount of target associate with the internal control probes for BMEII0462, BMEII0685 and BMEI0267 to fade. Under the DMSO conditions the targets now readily hybridize to the other probes in the set. Raising the DMSO concentration from 0% to 8 % causes the targets to bind more uniformly to all probes in the set. However, when 9% DMSO was modeled OMP DE failed to detect most probe –target interactions, meaning that at least on a computational level the presence of this much DMSO means the duplexes themselves are no longer stable at this temperature.

3.4.6 Simulations of Competitive Hybridization Using the OMP DE

All of our equilibrium simulations of array hybridizations described to this point included multiple probe competition for the same targets. The results obtained from these simulations suggest, that when the probes are in excess (which

is a the common assumption in the microarray data analysis) and multiple different probes are placed on the slide for each of the target species (which is common for alternative splicing, tiling, Affymetrix and NimbleGen arrays) there is a strong competition among the probes for binding to an intended target. Under such circumstances, presence of the secondary structure in the probe-binding site of the target is an important factor, and as our predictions have shown it may considerably shift the competition in favor of those probes that have secondary structure-free binding sites.

We have also investigated the effect of the effective probe concentration on the output from the multistate equilibrium simulations. It is clear that when the stock oligo solution is pipetted out by the robotic device and spotted several times on the surface of the slide, dried and re-suspended in pre-hybridization and hybridization solution the effective probe spot concentration has no longer anything in common with the concentration of the original solution. However knowing the probe concentration in the spotting solution, as well as the size of the pin heads and the final spot diameter can be used to calculate the effective probe concentration on the chip [207]. Following the *Ricelli etal.* protocol for the 20μM probe spotting solution, we obtained the effective probe concentration in the hybridization volume in the order of 2mM. Using the equilibrium model described by *Gharaibeh et al*. [208] we estimated the effective probe concentration to be close to 2 nM. We were surprised by the 1,000,000 fold difference in the calculation estimates. Our effective probe concentration of 20μM used throughout the modeling simulations fell

approximately in between these two values. In order to confirm that the effects of the target secondary structure, which were observed in the computational microarray simulations under the competitive conditions described above, would hold at different probe concentrations we performed two additional simulations at the effective probe concentration levels of 2 nM and 2 mM. The DNA target percent bound at 60°C 0.6 M [Na^+] for the 2nM, 2μM and 2mM probe concentration and proved to be exactly the same, with a miniscule variations at 2nM. Meaning that when the system reaches the equilibrium the fact that the probe concentration greatly exceeds the target concentration overrides the effects of the exact probe concentration value, and since all probes are spotted at the same nominal concentration, the probe concentration does not influence competition for target.

3.4.7 Simulations of Noncompetitive Hybridization Using the OMP DE

We next investigated whether the target secondary structure would affect the probe-target binding under the noncompetitive conditions.

To address this question through the equilibrium simulations we split the original 5 targets 5x6=30 experimental probes and 7 negative controls miniarray into six separate subsets. Each subset contained all 5 targets but only 1 probe per each target e.g.: BMEII0462m_0, BMEII0874m_0, BMEII0685m_0, BMEI0267m_0 and BMEI0682m_0, plus the 7 negative controls. The simulations were performed at 60°C, sodium concentration of 0.6 M and no additives conditions and the maximum allowed internal loop and bulge size set at 35 bp. The results of these simulations are available in TABLE 11. They clearly show that upon reaching the equilibrium

when there is only a single probe available, targets having significant amounts of secondary structure are found to be bound to the same extent as those with unconstrained binding sites. In our computational simulation in the presence of potential non-specific target competitors the total (optimal and suboptimal) target percent bound ranged between the 99.99% and 100% for all intended probe-target pairs. This means that when no probe competitor with the secondary structure free binding site was present on the array upon reaching the equilibrium all of our probes were eventually hybridized to their intended binding sites on the target whether these sites were up to 68% folded or virtually structure free.

3.5 Discussion

3.5.1 Insights from the Microarray Design Process

One of the conclusions resulting from the array design process is that at common hybridization temperatures, for hybridization assays of an average size bacterial genome, there are very few if any completely secondary structure free probe binding sites per target.

We hypothesized that placing even perfect (sensitive, specific and secondary structure free probes) in the regions of stable target secondary structure may alter the obtained spot signal intensity, and therefore will require new rules for the microarray probe selection. However, as our array design has shown, finding 50 nucleotides long completely secondary structure free stretch on a full length DNA transcript can be challenging even for a bacterial organism. This is especially true for the genome-wide arrays, where greater specificity constrains limit the probe

selection. In addition, higher than the average GC content results in more stable target folding, which also limits the search for the secondary structure free probe-binding site. The third factor, which significantly reduces the chances of finding the ideal probe binding regions, is the RNA nature of the target molecules. And the last, but not the least factor, which can render the task completely impossible is the low microarray hybridization temperature. At this point it seems like it would be wise, to design microarrays for selected sets of DNA gene transcripts (or perhaps split the pangenomic assays into smaller subsets), and run the hybridizations at 55°C-60°C. However, the reality of the microarray technology is that thousands of genes are tested in parallel on a single slide, and the experimental hybridization temperatures vary wide from 60°C and down to the room temperature [209]. At the same time, with prokaryotic organisms grown in culture it is still a common practice to have the total mRNA extracts fluorescently labeled and used as the target, rather than have the complementary DNA synthesized [210]. The organism's GC content is not a variable parameter, and it can be higher than average as it is with Brucellae, for which the GC content is around 57%.

Our analysis of the *Brucella melitensis* 16M ORFeome showed that when the target secondary structure criterion was applied in its most strict form (the probe-binding site is not allowed to contain any stable secondary structure) and Kane's probe specificity criteria [29] were applied, the majority of the target molecules at 60°C contained only one completely secondary structure free probe binding site. It is fair to note here that we were taking into consideration probes with rather

relaxed uniformity parameters, such as a GC variability range of 8%, and allowed T_m variation of $\pm$2°C. It is also fair to mention that Kane's criteria are currently being re-assessed as too relaxed in term of the minimum nucleation (consecutive match) parameter (Gharaibeh, Gibas unpublished data). Our testing also showed that out of 3,198 open reading frames in *Brucella melitensis* 16M as many as 2,057 transcripts had no completely secondary structure free probe binding sites at all. A similar analysis performed using less stringent GC variability range of 12%, and allowed T_m variation of $\pm$5°C showed that out of 3,271 open reading frames in *Brucella suis* 1330 exactly 2,700 transcripts had none or one completely secondary structure free probe binding site. At the same time a number of targets had no secondary structure free probe-binding sites at all. Therefore, to create real genome wide microarrays a more tolerant target secondary structure criterion may have to be developed, which will be based on the actual experimentally established cut-off for the structural stability abundance.

3.5.2 Loops and Loop Sizes on the Probe Binding Site

In the computational study, we focused specifically on those target base pairs that are actively involved in hydrogen bond formation. However, there are a number of other accessibility considerations that may prevent particular bases from the hydrogen bond formation and need to be taken into account when developing a scoring scheme. The structure of a long single stranded DNA or RNA molecule can contain many nucleotides, which, while not being a part of a double-helical stem, remain inaccessible to hybridization due to their location inside small loops (less

than 10 bases) within the target secondary structure. We hope that our modeling data will at least partially clarify how loops and loop-like structures inside the probe-binding sites on target molecules affect the performance the microarray assays.

3.5.3 Hybridization Simulations Using the OMP DE: Temperature Series

When looking at the probe-target percents bound from the equilibrium simulations for our miniarrays, it is obvious that 3 out of 5 selected probe-target sets completely supported our hypotheses of the target secondary structure interference with the intended duplex formation. However, the two other sets (BMEI0682m_and BMEI0874m) seem to contradict to our expectations. The detailed sequence analysis of probes in these two sets suggests several possible reasons as to why these sets gave an unexpected distribution of target percent bound. For example, at equilibrium the heavily folded for BMEI0682m_1_M probe-binding site is found in a double stranded conformation with its intended probe. A careful look at the probe sequence of BMEI0682m_1_M reveals two high energy GC clamps at both 3' and 5' ends of this probe (see TABLE 4). Such clamps are known to affect PCR primers but are not typically screened from probe design software, despite reports that a 4-G homopolymer creates abnormally bright spots on some arrays [203]. We hypothesize that they work as a lock to ensure that one the probe finds its binding spot on the target it will settle down and will not be able to travel any further. Another reason why the BMEI0682m target would prefer this particular probe out of the pool of six, could be due to its unusually high GC content of 62%,

especially when compared to the GC content of 54% for the internal control probe and 54-56% content of the rest of the probes. This leads to a conclusion that there are other properties of the nucleic acids that can have stronger effects on probe-target hybridization than presence of the target secondary structure alone. High GC content is one of them. It causes the hybrid to have larger free energy of formation and therefore results in a stronger probe-target interaction.

Another example is the BMEII0874m_1_M probe, which has also accumulated a significant amount of target, and its intended binding site. When we look at the probe sequence of BMEII0874m_1_M (see TABLE 4), nothing out of the ordinary seems to catch an eye. It is only when we look at the low abundance of the secondary structure in its binding site, that a thought occurs, that most likely we are facing a lower threshold of the target secondary structure effect. Indeed there are only 7 bases in the probe binding site of BMEII0874m_1_M, which are directly involved in the secondary structure formation (see the binding site accessibility in TABLE 4). This is a rather small number, when compared to 23-26 bases in the other probes from the BMEII0874 set. In fact, BMEII0874m_1_M resembles and most likely serves as a second secondary structure free internal control for this probe set, which also has a higher GC content than the original internal control probe.

3.5.4 Hybridization Simulations Using the OMP DE: Formamide Series

FIGURE 21 illustrates the BMEII0462m target under the 10% formamide conditions at 60 °C. It is interesting to mention here that in a situation like the one

shown on this figure, when there is virtually no secondary structure present on any of the probe binding sites, other forces govern the probe-target binding and may result in target accumulation in one probe spot rather than the other as it is shown on FIGURE 19, when most of the signal accumulates at probe BMEI0462m_5.

3.5.5 Discussion of Modeling Results

The results of our study predict an important and overlooked role of target secondary structure in estimating the concentration of the target in a sample mixture when the assay is hybridization to oligonucleotide arrays. Oligonucleotide probe binding sites are found in double-stranded conformations in a significant fraction of transcripts even in cases where self-complementarity was avoided during the probe design process. In chapter 2 we showed that at 52°C approximately 57% of probes designed for Brucella had binding sites in the target predicted to contain a stretch of unpaired bases of at least 14 nt in length; at 65°C, that fraction increased to 93%. We have demonstrated that under the competitive conditions, when multiple probes are placed on the array the presence of the secondary structure in the probe binding site greatly affects the probe competition and in some cases may lead to complete loss of association of target with the probe. At the same time other biophysical properties of the both probe and target molecules such as the probe's high GC content in combination with the poly-GC blocks can override the effects of the target secondary structure.

Our simulations showed that under competitive binding conditions probes which had the secondary structure rich probe-binding sites on DNA targets exhibit

modified behavior. As expected their unusual behavior changes with the temperature raise but does disappear completely. A similar conclusion can be made for the studies of the secondary structure altering additives. Strong secondary structure relaxing chemicals, like DMSO and formamide, destabilize some of the target secondary structure at expense of weakening the specific probe-target interactions.

Our probe design and analysis was performed using two different oligo modeling platform. The HYBRID software, which was used for the secondary structure prediction and oligo probe selection, models the two-state equilibrium, while the OMP DE software, which was utilized to perform the hybridization simulation, performs the multi- (to be exact seven-) state equilibrium. The OMP DE calculations are based on the proprietary nearest neighbor binding energies, which therefore cannot be directly compared to the nearest neighbor parameters used by HYBRID. However, we can use the amount of the secondary structure located at the binding site of BMEII0462m_2 (see TABLE 4) and the shape of the same probe-binding site predicted by OMP DE (see FIGURE 9) to state that the energies used by these two modeling platforms are different. Another example could be the secondary structure prediction for the entire BMEI0682m. When modeled with the HYBRID 3.7 software binding sites not only for BMEI0682m_0, but also for BMEI0682m_2, BMEI0682m_3 and BMEI0682m_5 have up to 50% of bases actively involved in the secondary structure formation (see TABLE 5). However, the target folding performed using the OMP software shows the same binding sites virtually

free of the secondary structure (See FIGURE 13). It would be of great interest to compare the Gibbs free energies of formation of the intramolecular bonds in the probe-binding sites alone as they are modeled by the OMP, however such data are not available in any of the output files. Therefore, our speculations on the effects of the target secondary structure in the probe binding site have to be restricted to the secondary structure abundance rather than stability.

Another important issue, which has been left outside the scope of this computational investigation, is the role of the small internal loops and bulges inside the folded probe-binding sites. In this study, we focused specifically on the DNA base pairs that are actively involved in hydrogen bond formation. The structure of a long single stranded DNA molecule can contain many nucleotides that, while not part of a double-helical stem, remain inaccessible to hybridization due to their location inside small loops within the target secondary structure. The nucleotides trapped inside the small loops play a controversial double role in the hybridization kinetics: they render themselves inaccessible, while destabilizing the surrounding base pairs. It has been shown in the RNA experiments that loops of less than 10 nucleotides long are barely accessible [161], which was not included in out secondary structure accessibility calculation due to the lack of the experimental quantitative characterization of the bases hidden inside the internal loops and bulges on the single stranded DNA in particular.

Another type of structural elements that which can play a considerable role in the target secondary structure stabilization are pseudoknots. A pseudoknot

consists of at least two stem-loops in which half of one stem is intercalated between the two halves of another stem. Because of the overlapping nature of the pseudoknots (base pairs can overlap each other in the sequence position) these structural elements are not easy to predict using the dynamic programming algorithms due to computational complexity, and as a result time and memory demands [118, 211]. Therefore, they cannot be predicted neither by HYBRID nor by the OMP DE. For this reason the pseudoknot calculations have been omitted from our study. A grammatical context-sensitive method [212], which ignore RNA molecular energy, can be used to predict the pseudoknots, but they require verification with other computationally intense methods.

3.5.6 Widely Used – Poorly Understood

The results of our computational investigation support the first conundrum postulated by *Pozhitkov et al.* [213] as to the current probe design parameters being incorrect, but disagree with the authors conclusion on solving the problem of the spot signal predictability by placing the multiple probes per target. In fact our entire study shows that predictions as to the strength of the microarray spot signal can and should be made, and the final microarray results should always be normalized for the binding sequence complexity.

There have been very few experimental studies investigating the possible effects of the transcript secondary structure on the microarray spot intensity. In 2001 studies of the surface plasmon resonance imaging of the 16S rRNA [214] suggested that special considerations should be given to large target nucleic acids

in terms of their secondary structure abundance and stability as well as slower hybridization kinetics. In 2003 *Chandler et al.* [215] concluded that relaxing the secondary structure is the major concern for successful detection of 16S rRNA on planar oligonucleotide microarrays for achieving specific hybridization. In 2004 Lane et al. [23] made an attempt to directly investigate the reason short (15-20-mer) oligonucleotide probes would not hybridize to their complementary PCR amplicons. Their results pointed at the amplicons secondary structure as the culprit behind the poor hybridization.

3.5.7 The Limitations of Our Computational Simulations

Our computational study, just as any other in the field of science other than precise mathematics had its limitations, which included but were not limited to the fact that our conclusions were based on the equilibrium simulations, which represent the end-point percentages of the targets bound. While it is highly desirable that the microarray hybridizations should reach the equilibrium before the slides are read and analyzed, it is not always a feasible procedure for the gene expression arrays. For example, it was estimated that it would take 41 days for the whole-transcriptome chips to reach the hybridization equilibrium $t_{½}$ [79], when hybridizing with the tissue cDNA. This would mean that when the typical microarrays are developed in 24-48 hours, the observed spot hybridizations represent the early stages of the probe-target hybridization far from the actual equilibrium concentrations. Therefore, applying a multistate multiplex kinetic

model such as the one currently being developed by *Gantovnik et al.* (in preparation) would allow for a more accurate real-time prediction of duplex hybridization.

Our computational predictions for the target percent bound are made for the microarray systems, which have reached the complete equilibrium. The discrepancies in the computational predictions and the observed microarray signals may arise from the fact that the real microarray scans are made way before the system has reached the complete equilibrium or the fragile equilibrium has been distorted in the stringent washing steps. The washes, which follow at the end of all microarray protocols represent non-equilibrium processes, in which some of the targets are being washed away [213]. Their primary aim is to remove the targets, which formed the unintended duplexes, however they do so by disrupting the system equilibrium.

Our results of the effects of the target secondary structure are bases on the computational simulations for the top 5 probe-target sets derived from a high GC organism, which carry the highest inaccessibility scores, and therefore represent an extreme rather than a common situation. Their secondary structure inaccessibility scores refer to the however are lowered by the fact that the dynamic modeling algorithms utilized by both HYBRID and OMP DE are not capable to predict the mode complex pseudo-knot formations. Performing the wet-lab experiments using our target secondary structure mini- and especially microarray platforms would produce the direct experiment evidence in support of our computational predictions or otherwise will prove them wrong. Performing the simulations for the full size

target secondary structure microarrays will provide more accurate information on setting the accessibility and energy thresholds in using the target secondary structure as the factor affecting the oligo hybridization on the microarrays.

Most of our simulations were performed for a highly competitive probe set design, in which several probes present in access competed for binding to the same target, which limits the obtained results to those array platforms on which probe competition exists. However, such platforms are extremely abundant including all of the Affymetrix, tiling and alternative splicing arrays. The equilibrium simulations performed at one probe one targets scale showed no target secondary structure dependency. However once again, they represented the end point estimates. Taking into account the kinetics of the probe–folded binding site hybridization is likely to show considerably different hybridization times for these probes versus their secondary structure free controls. We can test to see if equilibrium is reached in the experimental context by varying hybridization time, and this set of experiments is one planned use of the miniarray platform we have described.

3.5.8 Why Do the Microarrays Work at All?

So, considering the interference with binding predicted by computational simulations, why do the microarrays work at all? We suggest two explanations. First of all, there are techniques, which help to reduce the secondary structure on the nucleic acids, and the other reason lays in the assumption that sometimes the microarrays just seem to work rather actually supply the reliable data due to the presence of omitted variable bias in statistical analysis of microarray data.

It has been noticed for a long time that and it is described in greater detail in the introduction to this chapter that the nucleic acid secondary structure is bad for intended hybridization on all hybridization-based platforms. Therefore, numerous approaches have been developed to counteract this effect: from raising the temperature and adding the structure relieving chemical agents to shearing the unsuitably large and heavily folded transcripts, which is routinely performed for all Affymetrix arrays. As our computational simulations in the presence of DMSO, formamide and elevated temperature have shown such agents indeed help to eliminate some of the target secondary structure effects. Our 2005 study of the abundance and stability of the target secondary structure indicated that shearing while shearing reduces overall ΔG of secondary structure formation for individual molecules in the target solution, shearing does not in itself eliminate formation of secondary structure in single-stranded DNA or RNA [141]. While some signal may be gained by shearing the target molecule, random shearing by will creates a mixture of targets that may have substantially different affinities. With using the random shearing there will always be a danger that a random break in the sequence will occur within the target site for which the probe is designed. The mechanistic approach for target shearing should be avoided, because the short fragments that are produced by this procedure may represent substantially different binding sites, and therefore have a different binding affinity, than the full-length transcript that is considered when the probe is designed. At the same time there are many microarray protocols, which exclude the searing step [216, 217], and some array

designs which cannot compile with performing the shearing step, such as the arrays specifically created for detection of the alternative splicing.

It has been shown experimentally [215] that although, relaxing the secondary structure is the major concern for successful detection of large RNAs on oligonucleotide microarrays for achieving specific hybridization, sequence of the oligo probe could be more important than the presence of the structure. This agrees with the simulation results we observed for BMEI0682m_1_M, where high abundance of the secondary structure in the probe-binding site was overridden by the probe sequence.

The last reason as to why we believe that the microarrays actually work relies on the presence of the independent variables, which are a true cause of the changes in the dependent variable used to determine the R^2 in the course of the statistical data analysis. The examples of such variables could be the target secondary structure and two-state surface hybridization resulting in presence of false positive signals due to unintended binding of unstable high energy targets. As little information is usually given about the magnitude of these microarray parameters they are almost always omitted from any statistical interpretations causing the microarray data interpretation to be biased. Looking at this problem from another angle we can say that even the data sets with the high R^2 do not guarantee that the most appropriate set of independent variables has been chosen. These could be the reasons why sometimes as much as 29% discrepancies are observed between the microarray results and their RT-QPCR validations [218].

3.5.9 Next Computational and Experimental Steps

The next step in revealing the effects of the target secondary structure include obtaining direct experimental evidence to support of our results from the computational miniarray simulations. We currently own a set of 37 amino-C6 modified oligomer probes for our miniarrays produced by Operon, which will be randomized and placed in three replicates on a couple hundreds glass slides at the concentration of 20 μM. The slides will contain 5 bright spots to mark the corners for later easy grid placement. The target mixtures are currently being produced using pairs of primers (one of which is the biotin-labeled) via linear-after-the-exponential (LATE)-PCR amplification [219] of the 5 transcripts from the *Brucella melitensis* 16M clone library. The biotinylated primers for the LATE-PCR were designed using the optimized design criteria for high yields of specific single stranded DNA described by Pierce et al. [220]. In particular to maximize the reaction efficiency and specificity T_m of the limiting primer was kept 5°C above the T_m level for the excess primer. The T_m of the excess primer was kept close to the melting temperature of the double stranded product. As a back-up for the possible failure of the (LATE)-PCR amplification we designed a set of regular PCR primers with the T_m close to 60°C and one of the oligos carrying the biotin label. The single-stranded PCR products are intended to be purified using either the magnetic streptavidin-covered Dynabeads produced by Invitrogen or more expensive and more reliable streptavidin columns. The single stranded target will further undergo the 3'end-labeling using fluorescent nucleotides and terminal deoxynucleotydil transferase

known as TdT [221]. Thus produced fluorescently labeled target mixtures will undergo concentration and hybridization time testing on the miniarray slides. Once all of the hybridization criteria are optimized the temperature and additives' miniarray experiment series will be performed. The slides will be hybridized at 60°C long enough to insure that the reactions are approaching the equilibrium. The fluorescent signal will be scanned and analyzed using R and Spot software. It would also be interesting to make a 30 slide set of the 4-spot miniarrays, containing 4 replicas of the same probe to test the kinetics of the noncompetitive hybridization with respect to the effect of the target secondary structure.

The next computational steps that would be important for this project are very diverse and range from the evaluation of accessibility probabilities for each base in the probe binding site to computational simulations of the known experimental alternative splicing or tiling arrays. The accessibility probabilities for the probe binding site bases for our miniarrays have already been calculated by the OMP DE as a part of the overall target gene folding, however they still need to be summarized and analyzed in terms of miniarray experiment. Similar calculations should be performed for the 96 gene microarray along with the actual equilibrium simulations. Once the experimental data for the target secondary structure mini- and microarray experiments become available the statistical correlations between the computational predictions and the experimental confirmations can be drawn. The experimental results from the 96 gene microarrays can also be used to set up some thresholds and guidelines as to designing the microarrays normalized for

sequence complexity. As the next step of validation of our hypothesis we can use the experimental data from preferably bacterial Agilent gene expression arrays that use two probes per gene sequence to see if the competition effect seems to occur in the real experiments. We can again make the secondary structure predictions and the form the equilibrium simulations for such an arrays using the OMP DE and test the correlations between the target secondary structure and the spot signal. It is likely that a strong correlation between the probe GC percent and the signal strength will be revealed, and the probe sets may have to be pre-screened for a more uniform GC content. Another option would be to use some of the alternative splicing or tiling arrays for which both the probe sequences and the spot intensities are available and see if our predictions of low signal abundance will correlate with the 'silent exons'. This is especially interesting given some recent observations that variable probe characteristics lead to overestimation of the alternative splicing events [151].

3.6 Conclusions

In this study we have performed a computational microarray simulation to evaluate whether the structures we can predict using existing nucleic acid modeling algorithms will have a significant affect on the microarray probe signal. Our results from this modeling experiment demonstrate that under the competitive probe-binding conditions, those probes, whose binding sites are occupied with the secondary structure loose the competition to the probes with completely secondary structure free sites, and in some cases may have their signal completely bleached

out. This observation brings a new meaning to analysis and understanding of the gene expression arrays involving the alternative splicing events. In fact it suggests that for the maximum accuracy of the results it is best to separate multiple oligo probes for the same target into several microarray chambers on the same slide. Such platforms are currently available from some microchip producing companies, such as Agilent. Our modeling study showed that while adding such structure relaxing agents as formamide and DMSO reduces the target secondary structure, it also weakens the specific probe-target interactions.

The future directions in investigation of the effects of the target secondary structure lays in obtaining the direct experimental evidence in support of the observed computational simulations as well as statistical computational analysis of publicly available results from the alternative splicing arrays in terms of the targets secondary structure abundance in the probe binding sites as well as the relative probe CG content.

REFERENCES

1. Alwine, J.C., D.J. Kemp, and G.R. Stark, *Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes.* Proc Natl Acad Sci U S A, 1977. **74**(12): p. 5350-4.

2. Mocharla, H., R. Mocharla, and M.E. Hodes, *Coupled reverse transcription-polymerase chain reaction (RT-PCR) as a sensitive and rapid method for isozyme genotyping.* Gene, 1990. **93**(2): p. 271-5.

3. Schena, M., et al., *Quantitative monitoring of gene expression patterns with a complementary DNA microarray.* Science, 1995. **270**(5235): p. 467-70.

4. Wang, Z., M. Gerstein, and M. Snyder, *RNA-Seq: a revolutionary tool for transcriptomics.* Nat Rev Genet, 2009. **10**(1): p. 57-63.

5. Shinawi, M. and S.W. Cheung, *The array CGH and its clinical applications.* Drug Discov Today, 2008. **13**(17-18): p. 760-70.

6. Oostlander, A.E., G.A. Meijer, and B. Ylstra, *Microarray-based comparative genomic hybridization and its applications in human genetics.* Clin Genet, 2004. **66**(6): p. 488-95.

7. Watson, J.D.a.C.F.H.C., *A Structure for Deoxyribose Nucleic Acid.* Nature, 1953. **171**: p. 737-738.

8. Mullis, K., et al., *Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction.* Cold Spring Harb Symp Quant Biol, 1986. **51 Pt 1**: p. 263-73.

9. Saiki, R.K., et al., *Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase.* Science, 1988. **239**(4839): p. 487-91.

10. Saiki, R.K., et al., *Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia.* Science, 1985. **230**(4732): p. 1350-4.

11. Gibson, U.E., C.A. Heid, and P.M. Williams, *A novel method for real time quantitative RT-PCR.* Genome Res, 1996. **6**(10): p. 995-1001.

12. Heid, C.A., et al., *Real time quantitative PCR.* Genome Res, 1996. **6**(10): p. 986-94.

13. Suzuki, S., et al., *Experimental optimization of probe length to increase the sequence specificity of high-density oligonucleotide microarrays.* BMC Genomics, 2007. **8**: p. 373.

14. Pozhitkov, A.E., et al., *Simultaneous quantification of multiple nucleic acid targets in complex rRNA mixtures using high density microarrays and nonspecific hybridization as a source of information.* J Microbiol Methods, 2008. **75**(1): p. 92-102.

15. Wu, C., R. Carta, and L. Zhang, *Sequence dependence of cross-hybridization on short oligo microarrays.* Nucleic Acids Res, 2005. **33**(9): p. e84.

16. Chou, C.C., et al., *Optimization of probe length and the number of probes per gene for optimal microarray analysis of gene expression.* Nucleic Acids Res, 2004. **32**(12): p. e99.

17. Evertsz, E.M., Au-Young,J., Ruvolo,M.V., Lim,A.C. and Reynolds,M.A, *Hybridization cross-reactivity within homologous gene families on glass cDNA microarrays.* Biotechniques, 2001. **31**: p. 1182, 1184, 1186.

18. Wei, S. and S.S. To, *Influence of RNA secondary structure on HEV gene amplification using reverse-transcription and nested polymerase chain reaction.* J Clin Virol, 2003. **27**(2): p. 152-61.

19. Hackermuller, J., et al., *The effect of RNA secondary structures on RNA-ligand binding and the modifier RNA mechanism: a quantitative model.* Gene, 2005. **345**(1): p. 3-12.

20. Wood, W.B. and P. Berg, *Influence of DNA Secondary Structure on DNA-Dependent Polypeptide Synthesis.* J Mol Biol, 1964. **9**: p. 452-71.

21. Eperon, L.P., et al., *Effects of RNA secondary structure on alternative splicing of pre-mRNA: is folding limited to a region behind the transcribing RNA polymerase?* Cell, 1988. **54**(3): p. 393-401.

22. Gamper, H.B., G.D. Cimino, and J.E. Hearst, *Solution hybridization of crosslinkable DNA oligonucleotides to bacteriophage M13 DNA. Effect of secondary structure on hybridization kinetics and equilibria.* J Mol Biol, 1987. **197**(2): p. 349-62.

23. Lane, S., et al., *Amplicon secondary structure prevents target hybridization to oligonucleotide microarrays.* Biosens Bioelectron, 2004. **20**(4): p. 728-35.

24. Southern, E., K. Mir, and M. Shchepinov, *Molecular interactions on microarrays.* Nat Genet, 1999. **21**(1 Suppl): p. 5-9.

25. Watson, A., et al., *Technology for microarray analysis of gene expression.* Curr Opin Biotechnol, 1998. **9**(6): p. 609-14.

26. Harrington CA, R.C., Retief J, *Monitoring gene expression using DNA microarrays.* Curr Opin Microbiol 2000. **3**: p. 285-291.

27. Kuo, W.P., et al., *Analysis of matched mRNA measurements from two different microarray technologies.* Bioinformatics, 2002. **18**(3): p. 405-12.

28. Wang, H.Y., et al., *Assessing unmodified 70-mer oligonucleotide probe performance on glass-slide microarrays.* Genome Biol, 2003. **4**(1): p. R5.

29. Kane, M.D., et al., *Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays.* Nucleic Acids Res, 2000. **28**(22): p. 4552-7.

30. Hughes, T.R., et al., *Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer.* Nat Biotechnol, 2001. **19**(4): p. 342-7.

31. Ramakrishnan, R., et al., *An assessment of Motorola CodeLink microarray performance for gene expression profiling applications.* Nucleic Acids Res, 2002. **30**(7): p. e30.

32. Relogio, A., et al., *Optimization of oligonucleotide-based DNA microarrays.* Nucleic Acids Res, 2002. **30**(11): p. e51.

33. Thompson, K.J., et al., *A white-box approach to microarray probe response characterization: the BaFL pipeline.* BMC Bioinformatics, 2009. **10**: p. 449.

34. Bickel, P.J., et al., *An overview of recent developments in genomics and associated statistical methods.* Philos Transact A Math Phys Eng Sci, 2009. **367**(1906): p. 4313-37.

35. Irizarry, R.A., et al., *Exploration, normalization, and summaries of high density oligonucleotide array probe level data.* Biostatistics, 2003. **4**(2): p. 249-64.

36. Li, C. and W.H. Wong, *Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection.* Proc Natl Acad Sci U S A, 2001. **98**(1): p. 31-6.

37. Gharaibeh, R.Z., A.A. Fodor, and C.J. Gibas, *Background correction using dinucleotide affinities improves the performance of GCRMA.* BMC Bioinformatics, 2008. **9**: p. 452.

38. Naef, F. and M.O. Magnasco, *Solving the riddle of the bright mismatches: labeling and effective binding in oligonucleotide arrays.* Phys Rev E Stat Nonlin Soft Matter Phys, 2003. **68**(1 Pt 1): p. 011906.

39. Hardiman, G., *Microarray platforms--comparisons and contrasts.* Pharmacogenomics, 2004. **5**(5): p. 487-502.

40. Lipshutz, R.J., et al., *High density synthetic oligonucleotide arrays.* Nat Genet, 1999. **21**(1 Suppl): p. 20-4.

41. Chee, M., et al., *Accessing genetic information with high-density DNA arrays.* Science, 1996. **274**(5287): p. 610-4.

42. Nuwaysir, E.F., et al., *Gene expression analysis using oligonucleotide arrays produced by maskless photolithography.* Genome Res, 2002. **12**(11): p. 1749-55.

43. Barbacioru, C.C., et al., *Effect of various normalization methods on Applied Biosystems expression array system data.* BMC Bioinformatics, 2006. **7**: p. 533.

44. Baum, M., et al., *Validation of a novel, fully integrated and flexible microarray benchtop facility for gene expression profiling.* Nucleic Acids Res, 2003. **31**(23): p. e151.

45. Kuhn, K., et al., *A novel, high-performance random array platform for quantitative gene expression profiling.* Genome Res, 2004. **14**(11): p. 2347-56.

46. Bowtell, D.D., *Options available--from start to finish--for obtaining expression data by microarray.* Nat Genet, 1999. **21**(1 Suppl): p. 25-32.

47. Tyagi, S., D.P. Bratu, and F.R. Kramer, *Multicolor molecular beacons for allele discrimination.* Nat Biotechnol, 1998. **16**(1): p. 49-53.

48. Kostrikis, L.G., et al., *Spectral genotyping of human alleles.* Science, 1998. **279**(5354): p. 1228-9.

49. Kwok, P.Y., *Methods for genotyping single nucleotide polymorphisms.* Annu Rev Genomics Hum Genet, 2001. **2**: p. 235-58.

50. Blanchard, K.a.H., *High-density oligonucleotide arrays.* Elsevier: Advanced Technology, 1996. **11**(6/7): p. 687-690.

51. Vainrub, A. and B.M. Pettitt, *Surface electrostatic effects in oligonucleotide microarrays: control and optimization of binding thermodynamics.* Biopolymers, 2003. **68**(2): p. 265-70.

52. Lemeshko, S.V., et al., *Oligonucleotides form a duplex with non-helical properties on a positively charged surface.* Nucleic Acids Res, 2001. **29**(14): p. 3051-8.

53. Mary-Huard, T., et al., *Spotting effect in microarray experiments.* BMC Bioinformatics, 2004. **5**: p. 63.

54. Dufva, M., *Fabrication of high quality microarrays.* Biomol Eng, 2005. **22**(5-6): p. 173-84.

55. Wetmur, J.G., *Hybridization and renaturation kinetics of nucleic acids.* Annu Rev Biophys Bioeng, 1976. **5**: p. 337-61.

56. Rangasamy Elumalai, J.K., Tom Watson , Jack Gardiner1, David Galbraith, David Henderson, Robin Buell, Shawn Kaeppler, Vicki Chandler *The Effects Of Temperature On Positive Hybridizing Spots And Variation Between Slides In Maize Microarray*, in *Plant & Animal Genomes XIII Conference*. 2005: Town & Country Convention Center, San Diego, CA.

57. Mulle, J.G., et al., *Empirical evaluation of oligonucleotide probe selection for DNA microarrays.* PLoS One. **5**(3): p. e9921.

58. Hulsman, M., et al., *Delineation of amplification, hybridization and location effects in microarray data yields better-quality normalization.* BMC Bioinformatics. **11**: p. 156.

59. Lima, W.F., et al., *Implication of RNA structure on antisense oligonucleotide hybridization kinetics.* Biochemistry, 1992. **31**(48): p. 12055-61.

60. Duan, F., et al., *Large scale analysis of positional effects of single-base mismatches on microarray gene expression data.* BioData Min. **3**(1): p. 2.

61. Rennie, C., et al., *Strong position-dependent effects of sequence mismatches on signal ratios measured using long oligonucleotide microarrays.* BMC Genomics, 2008. **9**: p. 317.

62. Seringhaus, M., et al., *Mismatch oligonucleotides in human and yeast: guidelines for probe design on tiling microarrays.* BMC Genomics, 2008. **9**: p. 635.

63. Rouillard, J.M., C.J. Herbert, and M. Zuker, *OligoArray: genome-scale oligonucleotide design for microarrays.* Bioinformatics, 2002. **18**(3): p. 486-7.

64. Rouillard, J.M., M. Zuker, and E. Gulari, *OligoArray 2.0: design of oligonucleotide probes for DNA microarrays using a thermodynamic approach.* Nucleic Acids Res, 2003. **31**(12): p. 3057-62.

65. Mueckstein, U., et al., *Hybridization thermodynamics of NimbleGen microarrays.* BMC Bioinformatics, 2010. **11**: p. 35.

66. Kreil, R., Russell, *Microarray Oligonucleotide Probes*, in *Methods in Enzymology*, O. Kimmel, Editor. 2006, Academic Press. p. 73-98.

67. Kane, M.D., *Aligning experimental design with bioinformatics analysis to meet discovery research objectives.* Cytometry, 2002. **47**(1): p. 50-1.

68. Liebich, J., et al., *Improvement of oligonucleotide probe design criteria for functional gene microarrays in environmental applications.* Appl Environ Microbiol, 2006. **72**(2): p. 1688-91.

69. He, Z., et al., *Empirical establishment of oligonucleotide probe design criteria.* Appl Environ Microbiol, 2005. **71**(7): p. 3753-60.

70. Naiser, T., et al., *Impact of point-mutations on the hybridization affinity of surface-bound DNA/DNA and RNA/DNA oligonucleotide-duplexes: comparison of single base mismatches and base bulges.* BMC Biotechnol, 2008. **8**: p. 48.

71. Yue, H., et al., *An evaluation of the performance of cDNA microarrays for detecting changes in global mRNA expression.* Nucleic Acids Res, 2001. **29**(8): p. E41-1.

72. Shi, S.J., et al., *DNA exhibits multi-stranded binding recognition on glass microarrays.* Nucleic Acids Res, 2001. **29**(20): p. 4251-6.

73. Letowski, J., R. Brousseau, and L. Masson, *Designing better probes: effect of probe size, mismatch position and number on hybridization in DNA oligonucleotide microarrays.* J Microbiol Methods, 2004. **57**(2): p. 269-78.

74. Halperin, A., A. Buhot, and E.B. Zhulina, *Brush effects on DNA chips: thermodynamics, kinetics, and design guidelines.* Biophys J, 2005. **89**(2): p. 796-811.

75. Peterson, A.W., R.J. Heaton, and R.M. Georgiadis, *The effect of surface probe density on DNA hybridization.* Nucleic Acids Res, 2001. **29**(24): p. 5163-8.

76. Jayaraman, A., C.K. Hall, and J. Genzer, *Computer simulation study of probe-target hybridization in model DNA microarrays: effect of probe surface density and target concentration.* J Chem Phys, 2007. **127**(14): p. 144912.

77. Binder, *Thermodynamics of competitive surface adsorption on DNA microarrays.* J. Phys.: Condens. Matter, 2006. **18**(18): p. S491-S523.

78. Sartor, M., et al., *Microarray results improve significantly as hybridization approaches equilibrium.* Biotechniques, 2004. **36**(5): p. 790-6.

79. Carletti, E., E. Guerra, and S. Alberti, *The forgotten variables of DNA array hybridization.* Trends Biotechnol, 2006. **24**(10): p. 443-8.

80. Hooyberghs, J., et al., *Breakdown of thermodynamic equilibrium for DNA hybridization in microarrays.* Phys Rev E Stat Nonlin Soft Matter Phys, 2010. **81**(1 Pt 1): p. 012901.

81. Chan, V., D.J. Graves, and S.E. McKenzie, *The biophysics of DNA hybridization with immobilized oligonucleotide probes.* Biophys J, 1995. **69**(6): p. 2243-55.

82. Gingeras, T.R., D.Y. Kwoh, and G.R. Davis, *Hybridization properties of immobilized nucleic acids.* Nucleic Acids Res, 1987. **15**(13): p. 5373-90.

83. Bishop, J., et al., *Competitive displacement of DNA during surface hybridization.* Biophys J, 2007. **92**(1): p. L10-2.

84. Lang, B.E. and F.P. Schwarz, *Thermodynamic dependence of DNA/DNA and DNA/RNA hybridization reactions on temperature and ionic strength.* Biophys Chem, 2007. **131**(1-3): p. 96-104.

85. Owczarzy, R., et al., *Effects of sodium ions on DNA duplex oligomers: improved predictions of melting temperatures.* Biochemistry, 2004. **43**(12): p. 3537-54.

86. SantaLucia, J., Jr. and D.H. Turner, *Measuring the thermodynamics of RNA secondary structure formation.* Biopolymers, 1997. **44**(3): p. 309-19.

87. Xia, T., et al., *Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs.* Biochemistry, 1998. **37**(42): p. 14719-35.

88. Turner, D.H. and D.H. Mathews, *NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure.* Nucleic Acids Res. **38**(Database issue): p. D280-2.

89. Kierzek, R., M.E. Burkard, and D.H. Turner, *Thermodynamics of single mismatches in RNA duplexes.* Biochemistry, 1999. **38**(43): p. 14214-23.

90. Diamond, J.M., D.H. Turner, and D.H. Mathews, *Thermodynamics of three-way multibranch loops in RNA.* Biochemistry, 2001. **40**(23): p. 6971-81.

91. Burkard, M.E., R. Kierzek, and D.H. Turner, *Thermodynamics of unpaired terminal nucleotides on short RNA helixes correlates with stacking at helix termini in larger RNAs.* J Mol Biol, 1999. **290**(5): p. 967-82.

92. SantaLucia, J., Jr. and D. Hicks, *The thermodynamics of DNA structural motifs.* Annu Rev Biophys Biomol Struct, 2004. **33**: p. 415-40.

93. Carlon, H., *Thermodynamics of RNA/DNA hybridization in high-density oligonucleotide microarrays.* Physica A: Statisticsl Mechanics and Its Applications, 2006. **362**(2): p. 433-449.

94. Oliviero, F., Colombi, Bergese, *On the difference of equilibrium constants of DNA hybridization in bulk solution and at the solid-solution interface.* Journal of Molecular Recognition, 2010.

95. Westerhout, E.M., et al., *HIV-1 can escape from RNA interference by evolving an alternative structure in its RNA genome.* Nucleic Acids Res, 2005. **33**(2): p. 796-804.

96. Ding, Y., C.Y. Chan, and C.E. Lawrence, *RNA secondary structure prediction by centroids in a Boltzmann weighted ensemble.* RNA, 2005. **11**(8): p. 1157-66.

97. Long, D., C.Y. Chan, and Y. Ding, *Analysis of microRNA-target interactions by a target structure based hybridization model.* Pac Symp Biocomput, 2008: p. 64-74.

98. Allawi, H.T., et al., *Mapping of RNA accessible sites by extension of random oligonucleotide libraries with reverse transcriptase.* RNA, 2001. **7**(2): p. 314-27.

99. Snyder, T.M., B.N. Tse, and D.R. Liu, *Effects of template sequence and secondary structure on DNA-templated reactivity.* J Am Chem Soc, 2008. **130**(4): p. 1392-401.

100. Fan-Ching Chien, J.-S.L., Huan-Ju Su, Li-An Kao, chung-Fan Chiou, Wen-yih Chen, Shean-Jen Chen, *An investigation into the influence of secondary structures on DNA hybridization using surface plasmon resonance biosensing.* Elsevier: Chemical Physics Letters, 2004. **397**(4-6): p. 429-434.

101. Waterman, S., *RNA secondary structure: a complete mathematical analysis.* Math. Biosci., 1978. **42**(3-4): p. 257-266.

102. Nussinov R., P., G., Griggs,J.R. and Kleitman,D.J., *Algorithm for loop matchings.* SIAM J. Appl. Math., 1978. **35**: p. 68-82.

103. Nussinov, R. and A.B. Jacobson, *Fast algorithm for predicting the secondary structure of single-stranded RNA.* Proc Natl Acad Sci U S A, 1980. **77**(11): p. 6309-13.

104. Zuker, M. and P. Stiegler, *Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information.* Nucleic Acids Res, 1981. **9**(1): p. 133-48.

105. McCaskill, J.S., *The equilibrium partition function and base pair binding probabilities for RNA secondary structure.* Biopolymers, 1990. **29**(6-7): p. 1105-19.

106. Ding, Y., *Statistical and Bayesian approaches to RNA secondary structure prediction.* RNA, 2006. **12**(3): p. 323-31.

107. Tinoco, I., Jr. and C. Bustamante, *How RNA folds.* J Mol Biol, 1999. **293**(2): p. 271-81.

108. Onoa, B., et al., *Identifying kinetic barriers to mechanical unfolding of the T. thermophila ribozyme.* Science, 2003. **299**(5614): p. 1892-5.

109. Banerjee, A.R., J.A. Jaeger, and D.H. Turner, *Thermal unfolding of a group I ribozyme: the low-temperature transition is primarily disruption of tertiary structure.* Biochemistry, 1993. **32**(1): p. 153-63.

110. Mathews, D.H., et al., *Secondary structure model of the RNA recognized by the reverse transcriptase from the R2 retrotransposable element.* RNA, 1997. **3**(1): p. 1-16.

111. Crothers, D.M., et al., *The molecular mechanism of thermal unfolding of Escherichia coli formylmethionine transfer RNA.* J Mol Biol, 1974. **87**(1): p. 63-88.

112. Woodson, S.A., *Recent insights on RNA folding mechanisms from catalytic RNA.* Cell Mol Life Sci, 2000. **57**(5): p. 796-808.

113. Mathews, D.H., *Revolutions in RNA secondary structure prediction.* J Mol Biol, 2006. **359**(3): p. 526-32.

114. Christoph Flamm, I.L.H., *Beyond energy minimization: approaches to the kinetic folding of RNA.* Monatshefte fur Chemie 2008. **139**: p. 447-457.

115. Anne Condon, H.J., *Computational prediction of nucleic acid secondary structure: Methods, applications and challenges.* Elsevier: Theoretical computer Science 2009. **410**: p. 294-301.

116. Eddy, S.R., *How do RNA folding algorithms work?* Nat Biotechnol, 2004. **22**(11): p. 1457-8.

117. Mathews, D.H., et al., *Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure.* J Mol Biol, 1999. **288**(5): p. 911-40.

118. Rivas, E. and S.R. Eddy, *A dynamic programming algorithm for RNA structure prediction including pseudoknots.* J Mol Biol, 1999. **285**(5): p. 2053-68.

119. Akutsu, T., *dynamic programming algorithms for RNa secondary structure prediction with pseudoknots.* Discrete Appl. Math., 2000. **104**(1-3): p. 45-62.

120. Dirks, R.M. and N.A. Pierce, *A partition function algorithm for nucleic acid secondary structure including pseudoknots.* J Comput Chem, 2003. **24**(13): p. 1664-77.

121. Zuker, M., *Mfold web server for nucleic acid folding and hybridization prediction.* Nucleic Acids Res, 2003. **31**(13): p. 3406-15.

122. Mathews, D.H., et al., *Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure.* Proc Natl Acad Sci U S A, 2004. **101**(19): p. 7287-92.

123. *Hofacker, F., Stadler, Bonhoeffer, Tacker, and Schuster, Fast Folding and Comparison of RNA Secondary Structures.* Monatsh.Chem, 1994. **125**: p. 167-188.

124. Hofacker, I.L., *Vienna RNA secondary structure server.* Nucleic Acids Res, 2003. **31**(13): p. 3429-31.

125. Wuchty, S., et al., *Complete suboptimal folding of RNA and the stability of secondary structures.* Biopolymers, 1999. **49**(2): p. 145-65.

126. Shapiro, B.A. and K.Z. Zhang, *Comparing multiple RNA secondary structures using tree comparisons.* Comput Appl Biosci, 1990. **6**(4): p. 309-18.

127. Ding, Y., C.Y. Chan, and C.E. Lawrence, *Sfold web server for statistical folding and rational design of nucleic acids.* Nucleic Acids Res, 2004. **32**(Web Server issue): p. W135-41.

128. Ding, Y. and C.E. Lawrence, *A statistical sampling algorithm for RNA secondary structure prediction.* Nucleic Acids Res, 2003. **31**(24): p. 7280-301.

129. Ding, Y. and C.E. Lawrence, *A bayesian statistical algorithm for RNA secondary structure prediction.* Comput Chem, 1999. **23**(3-4): p. 387-400.

130. Turner, D.H., Sugimoto, N., & Freier, S.M., *RNA structure prediction.* Annu. Rev. Biophys. Biophys. Chem., 1988. **17**: p. 167-192.

131. Jaeger, J.A., D.H. Turner, and M. Zuker, *Improved predictions of secondary structures for RNA.* Proc Natl Acad Sci U S A, 1989. **86**(20): p. 7706-10.

132. Serra, M.J. and D.H. Turner, *Predicting thermodynamic properties of RNA.* Methods Enzymol, 1995. **259**: p. 242-61.

133. Knudsen, B. and J. Hein, *RNA secondary structure prediction using stochastic context-free grammars and evolutionary history.* Bioinformatics, 1999. **15**(6): p. 446-54.

134. Knudsen, B.a.J.J.H., *Using stochastic context free grammars and molecular evolution to predict RNA secondary structure.* Bioinformatics, 1999. **15**(6): p. 446-454.

135. SantaLucia, J., Jr., *Physical principles and visual-OMP software for optimal PCR design.* Methods Mol Biol, 2007. **402**: p. 3-34.

136. Knight, R., A. Birmingham, and M. Yarus, *BayesFold: rational 2 degrees folds that combine thermodynamic, covariation, and chemical data for aligned RNA sequences.* RNA, 2004. **10**(9): p. 1323-36.

137. Juan, V. and C. Wilson, *RNA secondary structure prediction based on free energy and phylogenetic analysis.* J Mol Biol, 1999. **289**(4): p. 935-47.

138. Ruan, J., G.D. Stormo, and W. Zhang, *An iterated loop matching approach to the prediction of RNA secondary structures with pseudoknots.* Bioinformatics, 2004. **20**(1): p. 58-66.

139. Ruan, J., G.D. Stormo, and W. Zhang, *ILM: a web server for predicting RNA secondary structures with pseudoknots.* Nucleic Acids Res, 2004. **32**(Web Server issue): p. W146-9.

140. Koehler, R.T. and N. Peyret, *Effects of DNA secondary structure on oligonucleotide probe binding efficiency.* Comput Biol Chem, 2005. **29**(6): p. 393-7.

141. Ratushna, V.G., J.W. Weller, and C.J. Gibas, *Secondary structure in the target as a confounding factor in synthetic oligomer microarray design.* BMC Genomics, 2005. **6**(1): p. 31.

142. Paulsen, I.T., et al., *The Brucella suis genome reveals fundamental similarities between animal and plant pathogens and symbionts.* Proc Natl Acad Sci U S A, 2002. **99**(20): p. 13148-53.

143. SantaLucia, J., Jr., *A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics.* Proc Natl Acad Sci U S A, 1998. **95**(4): p. 1460-5.

144. Zuker, M., *Calculating nucleic acid secondary structure.* Curr Opin Struct Biol, 2000. **10**(3): p. 303-10.

145. Peyret, N., *Prediction of nucleic acid hybridization: parameters and algorithms*, in *Department of Chemistry*. 2000, Wayne State University: Detroit, MI.

146. Sawant, P.D., et al., *Hierarchy of DNA immobilization and hybridization on poly-L-lysine using an atomic force microscopy study.* J Nanosci Nanotechnol, 2005. **5**(6): p. 951-7.

147. Taylor, S., et al., *Impact of surface chemistry and blocking strategies on DNA microarrays.* Nucleic Acids Res, 2003. **31**(16): p. e87.

148. Lockhart, D.J., et al., *Expression monitoring by hybridization to high-density oligonucleotide arrays.* Nat Biotechnol, 1996. **14**(13): p. 1675-80.

149. Hughes, T.R. and D.D. Shoemaker, *DNA microarrays for expression profiling.* Curr Opin Chem Biol, 2001. **5**(1): p. 21-5.

150. Wong, G., et al., *Exploiting sequence similarity to validate the sensitivity of SNP arrays in detecting fine-scaled copy number variations.* Bioinformatics. **26**(8): p. 1007-14.

151. Gaidatzis, D., et al., *Overestimation of alternative splicing caused by variable probe characteristics in exon arrays.* Nucleic Acids Res, 2009. **37**(16): p. e107.

152. Walker, S.J., et al., *Long versus short oligonucleotide microarrays for the study of gene expression in nonhuman primates.* J Neurosci Methods, 2006. **152**(1-2): p. 179-89.

153. Bozdech, Z., et al., *Expression profiling of the schizont and trophozoite stages of Plasmodium falciparum with a long-oligonucleotide microarray.* Genome Biol, 2003. **4**(2): p. R9.

154. Chou, H.H., et al., *Picky: oligo microarray design for large genomes.* Bioinformatics, 2004. **20**(17): p. 2893-902.

155. Nielsen, H.B., R. Wernersson, and S. Knudsen, *Design of oligonucleotides for microarrays and perspectives for design of multi-transcriptome arrays.* Nucleic Acids Res, 2003. **31**(13): p. 3491-6.

156. Tolstrup, N., et al., *OligoDesign: Optimal design of LNA (locked nucleic acid) oligonucleotide capture probes for gene expression profiling.* Nucleic Acids Res, 2003. **31**(13): p. 3758-62.

157. Nguyen, H.K. and E.M. Southern, *Minimising the secondary structure of DNA targets by incorporation of a modified deoxynucleoside: implications for nucleic acid analysis by hybridisation.* Nucleic Acids Res, 2000. **28**(20): p. 3904-9.

158. Ding, Y. and C.E. Lawrence, *Statistical prediction of single-stranded regions in RNA secondary structure and application to predicting effective antisense target sites and beyond.* Nucleic Acids Res, 2001. **29**(5): p. 1034-46.

159. Hofacker, I., *Fast folding and comparission of RNA secondary structures.* Monatshefte fur Chemie, 1994.

160. Amarzguioui, M., et al., *Secondary structure prediction and in vitro accessibility of mRNA as tools in the selection of target sites for ribozymes.* Nucleic Acids Res, 2000. **28**(21): p. 4113-24.

161. Scherr, M., et al., *RNA accessibility prediction: a theoretical approach is consistent with experimental studies in cell extracts.* Nucleic Acids Res, 2000. **28**(13): p. 2455-61.

162. Amarzguioui, M. and H. Prydz, *An algorithm for selection of functional siRNA sequences.* Biochem Biophys Res Commun, 2004. **316**(4): p. 1050-8.

163. Kretschmer-Kazemi Far, R. and G. Sczakiel, *The activity of siRNA in mammalian cells is related to structural target accessibility: a comparison with antisense oligonucleotides.* Nucleic Acids Res, 2003. **31**(15): p. 4417-24.

164. Michel, F. and E. Westhof, *Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis.* J Mol Biol, 1990. **216**(3): p. 585-610.

165. Sohail, M., S. Akhtar, and E.M. Southern, *The folding of large RNAs studied by hybridization to arrays of complementary oligonucleotides.* RNA, 1999. **5**(5): p. 646-55.

166. Bohula, E.A., et al., *The efficacy of small interfering RNAs targeted to the type 1 insulin-like growth factor receptor (IGF1R) is influenced by secondary structure in the IGF1R transcript.* J Biol Chem, 2003. **278**(18): p. 15991-7.

167. Zhang, H.Y., et al., *mRNA accessible site tagging (MAST): a novel high throughput method for selecting effective antisense oligonucleotides.* Nucleic Acids Res, 2003. **31**(14): p. e72.

168. Anthony, R.M., et al., *Effect of secondary structure on single nucleotide polymorphism detection with a porous microarray matrix; implications for probe selection.* Biotechniques, 2003. **34**(5): p. 1082-6, 1088-9.

169. McDowell, D.G., N.A. Burns, and H.C. Parkes, *Localised sequence regions possessing high melting temperatures prevent the amplification of a DNA mimic in competitive PCR.* Nucleic Acids Res, 1998. **26**(14): p. 3340-7.

170. Woese, C.R., S. Winker, and R.R. Gutell, *Architecture of ribosomal RNA: constraints on the sequence of "tetra-loops".* Proc Natl Acad Sci U S A, 1990. **87**(21): p. 8467-71.

171. Varani, G., *Exceptionally stable nucleic acid hairpins.* Annu Rev Biophys Biomol Struct, 1995. **24**: p. 379-404.

172. Ambartsumyan, N.S. and A.M. Mazo, *Elimination of the secondary structure effect in gell sequencing of nucleic acids.* FEBS Lett, 1980. **114**(2): p. 265-8.

173. Ronaghi, M., et al., *Analyses of secondary structures in DNA by pyrosequencing.* Anal Biochem, 1999. **267**(1): p. 65-71.

174. Viswanathan, V.K., K. Krcmarik, and N.P. Cianciotto, *Template secondary structure promotes polymerase jumping during PCR amplification.* Biotechniques, 1999. **27**(3): p. 508-11.

175. Krasilnikov, A.S., et al., *Mechanisms of triplex-caused polymerization arrest.* Nucleic Acids Res, 1997. **25**(7): p. 1339-46.

176. Hohna, W.-H.S.a.B., *DMSO improves PCR amplification of DNA with complex secondary structure.* Trends in Genetics. **8**(7).

177. Kang, J., M.S. Lee, and D.G. Gorenstein, *The enhancement of PCR amplification of a random sequence DNA library by DMSO and betaine: application to in vitro combinatorial selection of aptamers.* J Biochem Biophys Methods, 2005. **64**(2): p. 147-51.

178. Separovic, E. and S.A. Nadin-Davis, *Secondary structure within PCR target sequences may facilitate heteroduplex production.* PCR Methods Appl, 1994. **3**(4): p. 248-51.

179. Patzel, V., et al., *Design of siRNAs producing unstructured guide-RNAs results in improved RNA interference efficiency.* Nat Biotechnol, 2005. **23**(11): p. 1440-4.

180. Matveeva, O.V., et al., *Optimization of duplex stability and terminal asymmetry for shRNA design.* PLoS One, 2010. **5**(4): p. e10180.

181. Shao, Y., et al., *Effect of target secondary structure on RNAi efficiency.* RNA, 2007. **13**(10): p. 1631-40.

182. Cullen, B.R., *Induction of stable RNA interference in mammalian cells.* Gene Ther, 2006. **13**(6): p. 503-8.

183. Tafer, H., et al., *The impact of target site accessibility on the design of effective siRNAs.* Nat Biotechnol, 2008. **26**(5): p. 578-83.

184. Chan, C.Y., et al., *A structural interpretation of the effect of GC-content on efficiency of RNA interference.* BMC Bioinformatics, 2009. **10 Suppl 1**: p. S33.

185. Obernosterer, G., H. Tafer, and J. Martinez, *Target site effects in the RNA interference and microRNA pathways.* Biochem Soc Trans, 2008. **36**(Pt 6): p. 1216-9.

186. Krueger, U., et al., *Insights into Effective RNAi Gained from Large-Scale siRNA Validation Screening.* OLIGONUCLEOTIDES, 2007. **17**(2): p. 237-50.

187. Welch, D.A.S.a.P.J. *An Unusual Path for RNAi Technology*. December 15, 2004; Available from: http://www.bio-itworld.com/archive/121504/rnai_path.html.

188. Allison, D.B., et al., *Microarray data analysis: from disarray to consolidation and consensus.* Nat Rev Genet, 2006. **7**(1): p. 55-65.

189. Nielsen, H.B., L. Gautier, and S. Knudsen, *Implementation of a gene expression index calculation method based on the PDNN model.* Bioinformatics, 2005. **21**(5): p. 687-8.

190. Irizarry, R.A., et al., *Summaries of Affymetrix GeneChip probe level data.* Nucleic Acids Res, 2003. **31**(4): p. e15.

191. Dricot, A., et al., *Generation of the Brucella melitensis ORFeome version 1.1.* Genome Res, 2004. **14**(10B): p. 2201-6.

192. Xia, X.Q., et al., *Evaluating oligonucleotide properties for DNA microarray probe design.* Nucleic Acids Res, 2010. **38**(11): p. e121.

193. Jourdren, L., et al., *Teolenn: an efficient and customizable workflow to design high-quality probes for microarray experiments.* Nucleic Acids Res, 2010. **38**(10): p. e117.

194. Dufour, Y.S., et al., *chipD: a web tool to design oligonucleotide probes for high-density tiling arrays.* Nucleic Acids Res, 2010. **38 Suppl**: p. W321-5.

195. Leparc, G.G., et al., *Model-based probe set optimization for high-performance microarrays.* Nucleic Acids Res, 2009. **37**(3): p. e18.

196. Bommarito, S., N. Peyret, and J. SantaLucia, Jr., *Thermodynamic parameters for DNA sequences with dangling ends.* Nucleic Acids Res, 2000. **28**(9): p. 1929-34.

197. Nordberg, E.K., *YODA: selecting signature oligonucleotides.* Bioinformatics, 2005. **21**(8): p. 1365-70.

198. Le Novere, N., *MELTING, computing the melting temperature of nucleic acid duplex.* Bioinformatics, 2001. **17**(12): p. 1226-7.

199. Markham, N.R. and M. Zuker, *DINAMelt web server for nucleic acid melting prediction.* Nucleic Acids Res, 2005. **33**(Web Server issue): p. W577-81.

200. Rice, P., I. Longden, and A. Bleasby, *EMBOSS: the European Molecular Biology Open Software Suite.* Trends Genet, 2000. **16**(6): p. 276-7.

201. Smith, T.F. and M.S. Waterman, *Identification of common molecular subsequences.* J Mol Biol, 1981. **147**(1): p. 195-7.

202. Sauer, P.a., *Quality control of chip manufacture and chip analysis using epoxy-chips as a mode.* Sensors and Actuators B: Chemical, 2003. **90**(1-3): p. 98-103.

203. Upton, G.J., W.B. Langdon, and A.P. Harrison, *G-spots cause incorrect expression measurement in Affymetrix microarrays.* BMC Genomics, 2008. **9**: p. 613.

204. Hung, T., K. Mak, and K. Fong, *A specificity enhancer for polymerase chain reaction.* Nucleic Acids Res, 1990. **18**(16): p. 4953.

205. Mamedov, T.G., et al., *A fundamental study of the PCR amplification of GC-rich DNA templates.* Comput Biol Chem, 2008. **32**(6): p. 452-7.

206. Sarkar, G., S. Kapelner, and S.S. Sommer, *Formamide can dramatically improve the specificity of PCR.* Nucleic Acids Res, 1990. **18**(24): p. 7465.

207. Riccelli, P.V., et al., *Hybridization of single-stranded DNA targets to immobilized complementary DNA probes: comparison of hairpin versus linear capture probes.* Nucleic Acids Res, 2001. **29**(4): p. 996-1004.

208. Gharaibeh, R.Z., et al., *Application of equilibrium models of solution hybridization to microarray design and analysis.* PLoS One, 2010. **5**(6): p. e11048.

209. Tao, S.C., et al., *Room-temperature hybridization of target DNA with microarrays in concentrated solutions of guanidine thiocyanate.* Biotechniques, 2003. **34**(6): p. 1260-2.

210. Gupta, V., et al., *Directly labeled mRNA produces highly precise and unbiased differential gene expression data.* Nucleic Acids Res, 2003. **31**(4): p. e13.

211. Akutsu, T., *Dynamic programming algorithms for RNA secondary structure prediction with pseudoknots.* Discrete Applied Mathematics, 2000. **104**(1-3): p. 45-62.

212. Cai, L., R.L. Malmberg, and Y. Wu, *Stochastic modeling of RNA pseudoknotted structures: a grammatical approach.* Bioinformatics, 2003. **19 Suppl 1**: p. i66-73.

213. Pozhitkov, A.E., D. Tautz, and P.A. Noble, *Oligonucleotide microarrays: widely applied--poorly understood.* Brief Funct Genomic Proteomic, 2007. **6**(2): p. 141-8.

214. Nelson, B.P., et al., *Surface plasmon resonance imaging measurements of DNA and RNA hybridization adsorption onto DNA microarrays.* Anal Chem, 2001. **73**(1): p. 1-7.

215. Chandler, D.P., et al., *Sequence versus structure for the direct detection of 16S rRNA on planar oligonucleotide microarrays.* Appl Environ Microbiol, 2003. **69**(5): p. 2950-8.

216. Francois, P., et al., *Comparison of amplification methods for transcriptomic analyses of low abundance prokaryotic RNA sources.* J Microbiol Methods, 2007. **68**(2): p. 385-91.

217. Anthony, R.M., et al., *Direct detection of Staphylococcus aureus mRNA using a flow through microarray.* J Microbiol Methods, 2005. **60**(1): p. 47-54.

218. Rajeevan, R., Vernon and Unger, *Use of Real-Time Quantitative PCR to Validate the Results of cDNA Array and Differential Display PCR Technologies.* Methods, 2001. **25**(4): p. 443-451.

219. Salk, J.J., et al., *Direct amplification of single-stranded DNA for pyrosequencing using linear-after-the-exponential (LATE)-PCR.* Anal Biochem, 2006. **353**(1): p. 124-32.

220. Pierce, K.E., et al., *Linear-After-The-Exponential (LATE)-PCR: primer design criteria for high yields of specific single-stranded DNA and improved real-time detection.* Proc Natl Acad Sci U S A, 2005. **102**(24): p. 8609-14.

221. Guerra, C.E., *Analysis of oligonucleotide microarrays by 3' end labeling using fluorescent nucleotides and terminal transferase.* Biotechniques, 2006. **41**(1): p. 53-6.

TABLES

TABLE 1: Gibbs Free Energy Upon the Secondary Structure Formation

Molecule	ΔG, kcal/mol			
	42°C		52°C	
	DNA	RNA	DNA	RNA
70-mer Probe	-6.8	N/A	-4.2	N/A
Full Length Target	-85.9	-188.4	-56.6	-- 140.2
200-mer Sheared Target	-25.5	-58.6	-15.9	-41.6
100-mer Sheared Target	-14.2	-25.7	-9.6	-18.0
50-mer Sheared Target	-6.1	-10.5	-4.2	-7.3

TABLE 2: Target Secondary Structure Array based on *Brucella melitensis* 16M ORFeome

INACCESSIBILITY SCORE FOR TARGET	TARGET NAME	PROBE INACCESSIBILITY SCORE	PROBE NAME & BINDING SITE LOCATION	PROBE SEQUENCE & PROBE-BINDING SITE ACCESSIBILITY
133	BMEII0462	24	BMEII0462_5	CTGATGCGCCAGCCTGCCCTCAAACGGATAGGCCACCATGTAGAAGCGGT
			1928 A 1977 G	00001000000111001000111111111000100111111111100001
		24	BMEII0462_4	GCTCATAGGTCTTGAGCGCATAGCCGCCGGTCGCCACAAAATCGAGAATG
			1386 C 1435 C	00000111111100000011100100100010010000111111111111
		24	BMEII0462_3	TCGAGAATGGTAATTTCCGGTTTTGCCCCGCCATCGACGGTGACAAGACC
			718 G 767 A	11111111111111110000010000100001111100001100001000
		27	BMEII0462_2	CGACCAACCCTTGGCCGCAAACCATTTGACGAAACGGTCCGGCAGAAGAA
			41 T 90 G	00000011111000000011000000011110000000111111111111
		34	BMEII0462_1	CCGGTGCGGGTCTCGATGGAAATATCAAGCCCCATTTCCTCCACCGGCAC
			355 G 404 G	00000000001000000111111000000100001111111000000001
		0	BMEII0462_0	GGCCCCAGAGACCCAGCGAATAATCAGTGGTGACAAAGCCCAGCGGATGC
			2022 G 2071 C	11111111111111111111111111111111111111111111111111
128	BMEII0874	23	BMEII0874_5	GTTATCAGCCACGTCACGAACATGCCCGCCATCAGGAATTCGCCATGAGC
			139 G 188 C	11111000100000111110000010001111111111000001111100
		24	BMEII0874_4	AACTTCCAACCGCTTCAAACCACCGATAGTCAGCCCTCGCGCTTCCGCCC
			2411 G 2460 T	11111111100001000111001000011111111110000011000000
		25	BMEII0874_3	CGACAATCGCGAATAGCCCGCCGAGCATCAGGCCGTTCAGAAGAATTTGC
			42 G 91 G	00111110001111100000100000000000000011111111111111
		26	BMEII0874_2	GACATGGTGTCGATCACGCCGATGATCAGCGCACCAACGAAGGCTCCAGG
			742 C 791 C	00001110000011100000011111000000111111111000001111
		30	BMEII0874_1	CTGGATCATGAGCCCCGTAGCACCGCCCGTAAGCCATCCAAAATGCAGCG
			1340 C 1389 G	00000000000100000001000001000011111111111111111000
		0	BMEII0874_0	GCGGCTCCCAGGCCAAAACAGAGAATGTAGATGCGCGAAACATTGACCCC
			574 G 623 C	11111111111111111111111111111111111111111111111111
128	BMEII0685	19	BMEII0685_5	CGGGCGCTCAACCGGCTTGTTGTTTTCGTCATAGATGCTGACGCCATTGC
			662 G 711 G	00011111111000001111111111111111111100000000011110
		20	BMEII0685_4	GGTTGCCCTTCGATACCAGCATATTGGTGTTCTGCGACTTCTGGACCGGG
			708 C 757 C	00111100100011000011111100001111111111111111000001
		24	BMEII0685_3	TCGCATCGGAGCCGAGGAACATTTCGCCTTCACCATAGCCAACCGCAAGC
			546 G 595 A	00000111100100011111111100010011111111100000000001111
		30	BMEII0685_2	CGCATAGACGAGCGGCGAGATGAATTCCGGCACCTCGGGCAGAACGATGG
			1694 C 1743 G	00011111111000000111111111110000010000000000000000
		35	BMEII0685_1	CGACAGTAACAGACTTGGCCAGATTGCGCGGCTGGTCCACGTCCGTTCCC
			1785 G 1834 G	00000000000000100000001111111110000100010000000111
		0	BMEII0685_0	AACACTGCATCCGATTCACGCGCGATGGTGGAGCCGGTCACGTTGAGCAC
			1123 G 1172 T	11111111111111111111111111111111111111111111111111
128	BMEI0267	16	BMEI0267_5	ATGCCGATGCCGCAGATCGCAAAGCCGATGATTGCAGTCGCCGAACTGTT
			907 A 956 T	00000000111111100001111110000111111111111111111111
		27	BMEI0267_4	CGACAAAGCCGATCAGTGATGGCGCGACCAACATGCCGGAATAGCCAAGC
			1041 G 1090 G	00111111111111111100000011111100000000000000000000111
		27	BMEI0267_3	GCATCATGGCCAGCGAACCGAGACCGCAGACCAGGATCATCAGGCCCATT
			168 A 217 C	00000100000100011111111110001000001000111111000111 1
		28	BMEI0267_2	GTCACCACATTCGGCACCAATGTGATGATGAGCAGCGCCGGAATGAAGGC
			280 G 329 C	00000000000111111100000000001111000001001111111111
		30	BMEI0267_1	CATGCGACAATGATCAGGATGGTGGCTGCGATCATCGTGACCAGCGCATG
			523 C 572 G	00000000000000011111111111111000000000111111000000
		0	BMEI0267_0	GTCAGCGTATCGGGCGAGGCTTGAGAAGGCGAGAACCACCAGCAGCAAGG
			1133 C 1182 C	11111111111111111111111111111111111111111111111111

TABLE 2: (Continued)

125 BMEI0682	17	BMEI0682_5	AACGCGCAATGCCGATATAGCCGGAAAGGGCCGTGAGAATAGGCAAGAGC
		1248 G 1297 T	11111000011100001111100001111111111111111111000111100
	22	BMEI0682_4	ACGGCAAGCGACGTGTCTTCCGTTTCGCCAAGGGTAGCGTTGATTTCTGC
		412 G 461 T	10000000000000011111111110000111111000011111111111
	26	BMEI0682_3	GGCAGCCAGAAACGATTCAAGCTTGAAGGTGACAACGAAACGCCACCACG
		611 C 660 C	11100001000111100010000111100000011111111000000111
	27	BMEI0682_2	GGAAGCCGGTCATCAGCTTGTAGAATGTGCTTTTCAGCTCCGCCCACTGA
		1539 T 1588 C	11000000111110000001111111100100000000111111100000
	33	BMEI0682_1	GCGCGCCCTGTCAGGAAGCCCATATACATGGTGACAAGGAAAGCGCCCGT
		1321 A 1370 C	01000000000001111110000111100000000011110000001001
	0	BMEI0682_0	TGATTTCTGCCTTTTCCGCTATCAGCCGTGCACGCTCTTCCCCAACGATG
		372 C 421 A	11111111111111111111111111111111111111111111111111
124 BMEI1324	17	BMEI1324_5	GTCGGCGCAAGTGACTGTTCCAGCTTGATCGTGAATGTCTTCGCGGCTGC
		1393 G 1442 C	11000011111110000111000011111111111111111000111100
	18	BMEI1324_4	CGAGGATTGCGTCTGGATCGACATTGTTACCCATTTCGCGCGCGATATCG
		1902 C 1951 G	11111111111111111111111111111100000010000000000001100
	21	BMEI1324_3	ACGATGAACCATTTGTCCATCACGAGCGCATCCTGGCCGAACGTTTGCTC
		2245 G 2294 T	11100000011111000000110000011111111111111111110000
	34	BMEI1324_2	GGGTCCAGCATTTCATCGAAGGCGACCTTTCCGGCCAGCGCGACGATCTT
		1813 A 1862 C	00011000000001111111100000010000000000000000011111
	34	BMEI1324_1	CAGAAGGGTCGCAGAGAAAATGCTGTTCATCGCCTGCCAGCGCGCAACGG
		1715 C 1764 G	00000100000011111000000001110000011111100000100000
	0	BMEI1324_0	CCAGTGTGAAGTCGAGTTTCGTTTCGGGTATGGCGTAAGGTGTCTGGCGA
		42 T 91 G	11111111111111111111111111111111111111111111111111
123 BMEII1020	20	BMEII1020_5	GCGCATAAAGCTGCTCCTTATCCTCATCGCAATCCGCACTGATCCAGACG
		1041 C 1090 C	00111111100000011111111000100011111000100011111111
	22	BMEII1020_4	AAGCGGCCCCTGCGGATAGCCAAGACCAAGCCTGACCGCCTGATCAATAT
		1445 A 1494 T	00000000000100000111111111111111111100100001111111
	25	BMEII1020_3	GGCCCCATGCGGAAACCTGCCGATTCGCGAAGTATCCGGTCGATATCGCC
		760 G 809 C	00000111111111111001001000001111100000001111110000
	26	BMEII1020_2	AGCATACGCGACAATATTTTCAGGATCCGAGAAGCCGCCCCATGCAAGCG
		1490 C 1539 T	11111110000011111111111100000000000000000001111100000
	30	BMEII1020_1	CCGTCTTCATAACGGTAGAAGCCCTGCCCCACCTTACGCCCGACATAACC
		934 G 983 G	00000111110000011111000011100000000000001000111111
	0	BMEII1020_0	GATGCGGGAGACGGCAGTTTGTGCATCTTCGGCGCTGATTTTGCCCTTTT
		290 A 339 C	11111111111111111111111111111111111111111111111111
123 BMEII0893	17	BMEII0893_5	GGCGAGAAGGCGGCATTCTCCACCTCGGTGAAATAGTTTTCCGGGTTGCG
		919 C 968 C	11111110000001111110000001111111111111110001111100
	18	BMEII0893_4	GGGAATTCGCCGTTCTCGATGGCCGAAAACAGATCTTCCTGCGTGGATTC
		736 G 785 C	11111111000011000001100001111111111111111000011110
	21	BMEII0893_3	TCGGCTTCCGCATTGGTCCAGTGCTTGTGGCCCTGCATGGTCTTGAAGTG
		664 C 713 A	00000111111111111110001001111110010001111111110000001
	29	BMEII0893_2	GCCTTCTTGCGTGTAGAATTTGAGGGCAAAGCCGCGCACGTCACGTTCGG
		335 C 384 C	11000010000001111111111000001100000010001110001001
	38	BMEII0893_1	ATCGTGACCTGATGCAGGCTTTCCGGCGACAGCGACCAGAAATCCCACAT
		505 A 554 T	00001000000000000010000000111111111000000010000000
	0	BMEII0893_0	GCAGATGGCTGTTCGACCGGACCATTGAACGAATTCGGCTCGTAATAGGC
		1171 G 1220 C	11111111111111111111111111111111111111111111111111
123 BMEII0564	22	BMEII0564_5	ACCTCCTTGATGCGCGCGCTGTCGCCTTCGATAAGTGCTCCGGCAAAAGG
		3433 C 3482 T	11000011111000000001111111111111111100100000011100
	22	BMEII0564_4	CCACGAAGCCGATGCCATCCCAGCCTGCCAGATCAGGGTCTTCGCAAAGG
		924 C 973 G	11111100001000100111100000000111111111111100000111
	24	BMEII0564_3	CAAGCGTCACCTCATTGGAAAGATCAAAGCCTGCGGAATTGCGGCGCGCG
		1584 C 1633 G	00000111111111111111111111111000000000011100000000
	26	BMEII0564_2	CACCTGCGCACGCTTGATCTCCGCATCCCAATAGGCGCCCTTCACCAGAC
		1058 G 1107 G	10000000000001111111100000110000111000011111111111
	29	BMEII0564_1	GAGTGCCTTCACCTTCGGCAGAAGTTCACCCATGACGCGCTCACTTTGCG
		785 C 834 C	11110001000001100001111100000001111000001001110010
	0	BMEII0564_0	AGGCTTTCCATGTCGTCGCGCTTGTAACGAACCACATGCAGGACCGGGCC
		2782 G 2831 T	11111111111111111111111111111111111111111111111111

TABLE 2: (Continued)

123 BMEI1492	18	BMEI1492_5	GCCACATGTAGGGCCACAGATTTCGCAATGTTTTCAGCGTCTCGCCGGAT
		30 A 79 C	00000001111111111111111100011111111110001110000111
	23	BMEI1492_4	ATAGAGCCAGACCGTCGCCGCCACGACCACGAGATAGGACAGGCCATAGG
		542 C 591 T	11111111111000000000000110000111111111110000000111
	24	BMEI1492_3	CAGGCTTTCCTGGGTCACGTCGCGCACATCCTGCCCATCGATGGTGATCG
		1256 C 1305 G	01111111111111111111100000100001000000000011110000
	24	BMEI1492_2	GAAAGCTCATGCATATGCACAAAGGTCTTGTAGGCAAGCTGGCGCACCGC
		340 G 389 C	00001100000001111111111111111111111100000000001000
	34	BMEI1492_1	TGAGGCCAGCCTGCACGATTTTCGCCGCATTATAGGCCAGCACCAGCATG
		249 C 298 A	11110000000000000010000000001111100000001000111111
	0	BMEI1492_0	TCCAGGTCTGGGTTGCAGCCTTTTCATATCGCGCCATAGCCCCATCGAAA
		744 T 793 A	11111111111111111111111111111111111111111111111111
123 BMEI0472	19	BMEI0472_5	AATATCGACCAGCCTGCCGGTAAAATAGAACAGCCCCGCCTCGACCAGCG
		161 C 210 T	11111110000011110000011111111111111000001111110000
	20	BMEI0472_4	CGCCAATCCAGCCGCCCGACAAGAAAGAGCGTGGTGATGCTGTAGATCAC
		532 G 581 G	11111111111110000000011111111110000000111111100000
	22	BMEI0472_3	GCACAAGACCAAGCGTGAACGAAACCTGCCCGACAGTGATCGCATTGCGG
		906 C 955 C	11111111111111111111111001000000111000000000000000
	28	BMEI0472_2	GGCGGATTACCAGAACCGTAATGCCCATGAAAAGAAGTTCCGGCCCCGCC
		267 G 316 C	00000011000000001111111111111111111100000000000000
	34	BMEI0472_1	GGCGCAACATGCCGTTCAGCTGCATCATCAACCGGCCAAACAGCATATTG
		963 C 1012 C	10000000000000000000111111111111110000000000000001
	0	BMEI0472_0	AAAGCGTGGACCAGCTGCGTTCGACATTATTCCTTCCCGCCTCGTCCAGA
		210 T 259 T	11111111111111111111111111111111111111111111111111
122 BMEII0473	19	BMEII0473_5	ACCGGAAGCTGGTTATAAGCTGCAATCCCGCCAATCAACCCGCCTATCAC
		52 G 101 T	00000000011111111111110000000000111111111111111111
	23	BMEII0473_4	ACGGGAGACTTTCCTGCGTGAAAACCAGCGATGATTTGCCAGCCGGTCGA
		3017 T 3066 T	01100011111000000011111000000111111111111000000111
	23	BMEII0473_3	CCAGCCAGCTATCCTGATTGACATTATCCACACTGTCCAGCACCGTTGCC
		777 G 826 G	11111111111111111111100000000011000000001111100001
	26	BMEII0473_2	CACACCCGAAATAGTGGACAGGGCTGGCGAGATGACATTCTGCGCAATGT
		467 A 516 G	00110001111111111111000100000110000000011111100000
	31	BMEII0473_1	GCACATCAGCCGCCGCCGCATCGATATTGCGCTTCAAGTCGAATTGCAGC
		276 G 325 C	00000100001000000010011111111000010001111110001000
	0	BMEII0473_0	GACTGTAAGGGCCTGTGACCTGTTTTGCAGCGTTCCAACCGGGGTCTGGT
		638 A 687 C	11111111111111111111111111111111111111111111111111
122 BMEII0130	21	BMEII0130_5	GCCTTCCTCATAGGCGACGTCATAAACCTTCTGCACCACGGTCTTGGCGG
		1187 C 1236 C	00011001111100100000000111111111111111111111000000
	22	BMEII0130_4	GCTTTTCCTTCAGTTGATTGAGCCCCGCCATAAGTTCGACACCGCGCGCC
		1056 G 1105 C	00001111111111111111110001000011111111110001000000
	23	BMEII0130_3	GACGTGAAGAAAGCACGCGTCAGCCCGTCTTCCTTGAAGAATTCGGCCAG
		298 C 347 C	01110000000000000001000000011111111111111111111111
	27	BMEII0130_2	AATAATAAGGCAGGCTGTCGAGCTGGTCGCGGATCGCCTTCTTCACCGGC
		198 G 247 T	11111111111111111100000001100001100000000000000000
	29	BMEII0130_1	ACCGCCCACCGCATCGAGAACCTTCTTGCCGTGAATGTCGATCACCTCGA
		116 T 165 T	10000001100000001111111111001001000100000111110000
	0	BMEII0130_0	CAGAGGCGAATTGGTCGTGCCGCGAAAGGTCGAATAATAAGGCAGGCTGT
		230 A 279 G	11111111111111111111111111111111111111111111111111
122 BMEI1691	18	BMEI1691_5	AACGTCATTGGCTTTCGTTCCTACCTGCCGATCAATGTCAGCGCGCGCGG
		1340 C 1389 T	00000111111111100000001111100001111111111111111000
	22	BMEI1691_4	CCGAAGGCCTCATCCTTACCTTTCTGGCTGGCGAGCGCGTTTTCATAGCG
		166 C 215 G	11111111110000000001001111111100000000111111111000
	22	BMEI1691_3	CCTTACCAAGCAAAGGCCCAACAACTGGCGCGCTCGTAATGAACGTTCGC
		39 G 88 G	00001111111100000111111111110000110001111100011000
	28	BMEI1691_2	CAGTCGCAAGCATGCCGTCAATTGCCTGTCGCGTTTGAAACAAGGTGCCC
		1020 G 1069 G	00110000110010011111111001001000011000000011111000
	32	BMEI1691_1	CCCTTGCGCTATTGGCTCGGTATTCACTGCTCGAGAATTGCGGAAAGGCT
		866 A 915 G	00000100010001100001111111111110000000100010000000
	0	BMEI1691_0	CGGCACCTTTGAAATTGGGCGAGATATCCATCAGCATTGCTTCCGGGCCA
		633 T 682 G	11111111111111111111111111111111111111111111111111

TABLE 2: (Continued)

122 BMEI0893	24	BMEI0893_5	GATGCCCGAAGTCGTGCCGCCGTCCATGCCAACGATCTGGAACCGAGTTT
		1769 A 1818 C	01100100111100100111111111111000010000011100000111
	24	BMEI0893_4	CAAGCCTTCCTCGATATGCCGGTGGATATTCTCCAGCACCACGATCGCGT
		1202 A 1251 G	11111111111000010010010000111110000100100110000111
	24	BMEI0893_3	CCTTGATAGGCATGGCCCCGAAAGCTTCCGGTGTCTGCAAGGTGGACTTC
		690 G 739 G	00011111111111000000000110000011000000011111111111
	25	BMEI0893_2	GCGCCCGAATCTTCTTCAGGTGCAAGCTCGCTGGACGTGTTCATGAACAG
		1618 C 1667 C	00100000011111000000100011100000000111111111111111
	25	BMEI0893_1	GCCGCCAGCGTGATCGTCATTGCGACAATCGAGCCGGTAATTTCCTTCAT
		1279 A 1328 C	00000111111000000011111000000011111111111100000001
	0	BMEI0893_0	AGCGAGCACCATGGCAAGCAGTGACAGGAGGTTGATCGAAAAGCCCAGCA
		1133 T 1182 T	11111111111111111111111111111111111111111111111111
121 BMEI2027	21	BMEI2027_5	GAGGCCGCGGCTGAAATCGCTTCGACCACCGATACGTCGCCAATCTCCAG
		1822 C 1871 C	01111000011111111111110000000000000011001111111111
	21	BMEI2027_4	GACCGCCTGTGCATAGGCCAGATTGACCCATTCGATGCGCCCATTGCCGT
		530 A 579 C	11110001100001110000000011111111111100001111111001
	24	BMEI2027_3	GCGCCCTTCACTGATAACCTGGCTCGCATTGGCCATATCGACCGCCTGTG
		569 C 618 C	00001111111111110000000010000110001111111111110001100
	27	BMEI2027_2	CGCACATTATGGCCGTCCCAGACCACAATGCGCTGGTCCTTCATGTTGAG
		127 C 176 G	00000000000000011111100000000000001111111111111111
	28	BMEI2027_1	GACATCGAGCAGCGAGCCATTGACCGTTTCGACCGTGAGATCAAAGGCGC
		317 G 366 C	11111000000000100010000001111111111111000000100010
	0	BMEI2027_0	CTGGTCCAGGGTTTCGGCATGGCTTTTGAGTGTGCGGCTCAGTTCGTCGC
		800 G 849 G	11111111111111111111111111111111111111111111111111
120 BMEII0217	21	BMEII0217_5	CTATCGAAGCAATAGTATCGGCCAGGAACGACGGGCTGCCCTTGACGTTT
		384 A 433 G	11110000000010001110010001111111111000100111111111
	21	BMEII0217_4	TGAACGTAAACTCGGTCTGGTCGCCATTGGCTGCATAGCTTTCGGCAAGC
		255 G 304 A	11110000110000011000011111111111111110000111110000
	23	BMEII0217_3	GCGTCCTTCACTTGGCGGAAGATGACCTCATCCACCTTCGGCGCTTCCCG
		697 C 746 C	11111111111111000000100111111001000000110000011111100
	24	BMEII0217_2	GATTCCGTCGCCATTAGCCTTTTCACGCCAGCTTGCAAACGGCACCGGCT
		1226 A 1275 C	11110001000000010011111111111111111001000000100011
	31	BMEII0217_1	CATTTTCCGCCTTAGCCTCGGTCGTGGCCACATCGCTGTTCACGATACCG
		516 C 565 G	11111110000000110000000000111000111100010000110000
	0	BMEII0217_0	GCGTGAAACCATTAGGATGACCGGCCTTGGCAAGCAGCTCCTTGGCTTTT
		1071 A 1120 C	11111111111111111111111111111111111111111111111111
120 BMEI1848	15	BMEI1848_5	CCCGCACTTTCGATTTCGCGCGCGACGAGCTGGCCGAGATCCTTCAGATG
		160 C 209 G	01000011111111110000100000011111111111111111111111
	18	BMEI1848_4	GCGGCGCGACAGACGGTCTATATCCGCCATGGTGAAATCGATTTCCGCTT
		878 A 927 C	00000001111111111111111100000000011111111111111100011
	19	BMEI1848_3	CTGCCGTCTTCACGATGCAGCCGTCTTCCGCCAGATTGCCGTAAAGCACG
		1275 C 1324 G	00001111111111111000010001111111111111000000111111000
	30	BMEI1848_2	CCATGGCGTCCACGAGGTCGAGCTTCTTGACAGAATCCTCCCACACGACC
		462 G 511 G	11110000100011000000001000011110000100111111100000
	38	BMEI1848_1	ATCGACGATCAGGTGGCCAGCTTCGACGAAGAGGCGCTTGCGGTCTGCAT
		680 A 729 T	11000000100000000001100000000000100000110001000111
	0	BMEI1848_0	GACTTCAGATAGCTGGTCGGGTAGAGCATTTCCTGCATTCCGGGGCCGCC
		1459 G 1508 C	11111111111111111111111111111111111111111111111111
119 BMEII1064	20	BMEII1064_5	ATGAAGATTGCCATCACCCGCATGGAACACATTAGCCACACGCAGGCCGT
		1256 A 1305 T	00001011111101000000001111111111111111111100011111000
	21	BMEII1064_4	CCGGTAATCAGCACCATTTCGATGCCGAGCACATTATTGGCGGTGAGGCC
		643 G 692 G	00111111000000011111000000011111111111110001111100
	21	BMEII1064_3	GCGGATTGCTTTTCGGGACATGCCCTCAAGCCATATCCGCCCTGAATTCC
		28 G 77 C	00001110000011111111111111100001001111001000011111
	22	BMEII1064_2	GACGCGCATCTGCTGGTCAAGATCGGTCTCGTTGAACATTTCCGGCATCA
		1445 T 1494 C	00000000000000011111111111111111111111111100000001
	35	BMEII1064_1	GTCACGACGCCAAGCAGACCTTCAGAGCCGGTCATGAGGCCAAGAAGATC
		739 G 788 C	00110000010000000000000110000111111110000110000000
	0	BMEII1064_0	AAGCAAGCCTCTGTTCTTCTGATTGCGACAGCGTGCAGGTGGTGGAACCG
		1074 C 1123 T	11111111111111111111111111111111111111111111111111

TABLE 2: (Continued)

119 BMEII0676	17	BMEII0676_5	CCGAATCGCCCTCCACGATGAACAATTCCGCACCTGCTGCCGCATTCTGC
		1419 G 1468 G	11110000011111000001111111111111111111111110000111000
	21	BMEII0676_4	TGTGCTTCCACGGTCGCAAGCTTGTCTTTGGTCTGGCCGACGAATTCCGG
		1177 C 1226 A	11100001001111001000011111111111111000001111110000
	25	BMEII0676_3	ATGATGACTTCGAGCGTGGATTTGCCCGGCACCTGCGGATGCTCGTCCAC
		400 G 449 T	11111111110000001111111001000000010000011111100000
	26	BMEII0676_2	TGCCGCCATCGCCGGTCGGGATGGTATTGCAATAGGAATGGACAAAACCA
		984 T 1033 A	10001111100100011110000001000011111111111000010000
	30	BMEII0676_1	ATCTCGCCAAGCCCCTTGAAACGTCCGATTTCGATCTTGCCGCGTCCGGT
		1864 A 1913 T	11110000000000011100000110000111111110000000100011
	0	BMEII0676_0	TTCCAGCGAATTTCCACACCACCGAACAGATAGGCCTTCGAGCGCGCCAT
		736 A 785 A	11111111111111111111111111111111111111111111111111
119 BMEII0216	19	BMEII0216_5	GCTTTCGCGCTGGTAAAACTCGACATGCTCAATAACCGCGACGGGATCGT
		1688 A 1737 C	11111000000011110010001111111111111111110000111000
	20	BMEII0216_4	CAGCCTTGATGGAAGCCTGTACCATGACGGATGTGGCAAGAACGGTGCAG
		1809 C 1858 G	00000011111111100000100011111111111111111110001000
	22	BMEII0216_3	TCGAACCGGCGGATCAGCTTCATCTGGCTATACCATTTCAGCAGATCGGC
		85 G 134 A	11100100001000010001111111111111111111100010000100
	28	BMEII0216_2	CGCCGAAGAAGGCAACAGAAATTCCCTTCCGGTTGAGGATCTTGTCTGCG
		516 C 565 G	00010000000010001111000100000011111111111111001000
	30	BMEII0216_1	GTCGTCCACACTGGATGCCGACGGCCAAAGCGCATCGGGAATACGCAGAA
		1073 T 1122 C	00000000111000011100000011111000000111110000001111
	0	BMEII0216_0	ATCTGTAGCACTGCGCCTCGATGACGACCGGACCGCCCTCTTCCTCGATA
		801 T 850 T	11111111111111111111111111111111111111111111111111
119 BMEII0028	20	BMEII0028_5	CAGGAACATGCGGCTTTCCTCGCTCATACTGCGGATGATGTCGAACTCCT
		2240 A 2289 G	00000000011111111111000111111100011111111000001111
	22	BMEII0028_4	CTTGCGCGAGGCAGAAGTTCATCAGGACAGTTTTACCAGCGCCGGACTGG
		1371 C 1420 G	01111001110000000011111111111111110000001111100000
	24	BMEII0028_3	ATTCAAGCTGCTGTGAGCACAGTGCGCCGTGTTTGTCCTCATAGGTGCTC
		558 G 607 T	11111000110010000100011111111111000100001001100001
	26	BMEII0028_2	CGAGCTGGTCGATGGCAATGTCCATTTCCATGATCTGGGATGCGCTGCCG
		918 C 967 G	00000000011000010001111000100001111111111111111000
	27	BMEII0028_1	CAGCGAATAATGATATTCACCCATGACGAAATCGCCCCGGCCGAGCTGGT
		959 A 1008 G	01111111111111110001000100001111000010001000000000
	0	BMEII0028_0	GTTCATCGCATGTGATGTGTTGTGGCGCACCGTCTCCATGTCTTCGGCAA
		1016 T 1065 C	11111111111111111111111111111111111111111111111111
118 BMEII0340	11	BMEII0340_5	CCCGATGATACCTTTGCGGAACAGCAGGACACAGGCCATGAAGATAAGGC
		905 G 954 G	11111111111111111111111111111000000001111111111000
	16	BMEII0340_4	AAGGCTGCGAGCTTGTAGTGATTGACGGAAAACCCAAGCGATATGGCGCG
		631 C 680 T	00000111110000001111111111111111111111111111100000
	27	BMEII0340_3	GAATACCGAGCACGAGGCCCAGTTCCGGTGAGACGCCCCATTCCTTGACC
		270 G 319 C	00000000000111110000000111111111111111100001100000
	28	BMEII0340_2	GTTGCGAACTGGAAGACAAGCGCCTTCATGCCGCCCGCAAGACCTGCAAG
		697 C 746 C	11111111111111100000000000001100001000000000111100
	36	BMEII0340_1	CGATGCCGCCGAGCAACGTCATCAGGATGACCTCGCCCGACATTTGCCAA
		762 T 811 G	00000000000000001000001111000001000000000011111111
	0	BMEII0340_0	ATGTTGCAAGCCAGTTCTTGAGTGCGACGACGAAGGCCGCGCCGACAATG
		828 C 877 T	11111111111111111111111111111111111111111111111111
118 BMEI0754	17	BMEI0754_5	ACACCCGGAATATATTCCTTCGGAACCGAGCCGCCGACGATCTTCGATTC
		1582 G 1631 T	11111111111111000011100001111110010000000111111111
	22	BMEI0754_4	TTGGAGTGCATCTGCAACATACGGCCGATACGCTCACGCTTGCCCTTGAC
		1045 G 1094 A	00010011111001000001111000011111111100000011111111
	23	BMEI0754_3	CTTCCGGCGTCACAACCTCGACCTTCATGATCGGCTCGAGAAGCTGGGCG
		1800 C 1849 G	11100000011111111110000011111000001111111100000000111
	26	BMEI0754_2	CGCAAAGGATCGGATGGAACTTGACTTCGATCGTGCCCTTGCGGATCAGA
		741 T 790 G	00000000000011111111111111110000111100000000011110
	30	BMEI0754_1	CTTCAATCGTGAAGTCAACGTGGCCAGGCGTGTCGATGATGTTGAAGCGG
		240 C 289 G	00000111100000010000010001100000011111111110000011
	0	BMEI0754_0	CGGCCTTGGTCTTCGGCTCGATAGCGATTTCGATAACCGGATCAGGGAAT
		1218 A 1267 G	11111111111111111111111111111111111111111111111111

TABLE 2: (Continued)

118 BMEI0109	22	BMEI0109_5	TGCGCCCGATCTTGATGACCGTCGCGAATTCCCGCCCCTTTGCTTCAAAG
		534 C 583 A	00000000000111111111111111110001000000111111111100
	23	BMEI0109_4	AGATGCGTCCAGCCCCTTGGAAATCGCCACTTCCATGACGGTCTGCCCTG
		1316 C 1365 T	11111111111100000001111100000011110000001110000111
	24	BMEI0109_3	CCGACAGATGCATGATGTGCCTCCTTGTCGTGCGTTTCCGGCGGGAATGC
		1399 G 1448 G	10000000100011111000100110000011001111001111111111
	24	BMEI0109_2	ATCCCGGAAAAGGTGACGAACGCACCGGTATCCCAGCTCGCTTCCACGAG
		814 C 863 T	11100001111111111110000100000111100000100001111100
	25	BMEI0109_1	CGCCGCCCTGAAACACATCCACGACGAATTGCTCCATCAAAAGCCCGTCA
		270 T 319 G	00000000111111000001100000011111111111111111000000
	0	BMEI0109_0	TTCTTGACCATTGCAAGGGCGCGAACGAGGGCAGGAATATGGTTCAGCGC
		136 G 185 A	11111111111111111111111111111111111111111111111111
117 BMEI1645	17	BMEI1645_5	GGTGATATCGATGCCCAGATCGGAAGCACGCTCGCGGTTGATCTCCACCG
		2138 C 2187 C	01111110000001001111111111111111110000111111110000
	17	BMEI1645_4	CGCTCGAAGGAGGAGGTGGAGGAAATCGATTTCACGCCGGAAACACGCGC
		238 G 287 G	00010001001000011111111111111111111111111111000001
	26	BMEI1645_3	GGCAGATGCGCCGCTAAAATCCGTGCTGATCGTGATGACCGGCCGATCCA
		143 T 192 C	00001111000001111111111110000000010000010000111111
	28	BMEI1645_2	CATGGTGTCGGAGGTAACGGCAAGGCGCATCACCGGATCGGCGTTGGAAT
		419 A 468 G	00111001000111111111100000001000000011100000011111100
	29	BMEI1645_1	CAAGATCGGTCGGATCGTTGATCGGGTTGGTGGTCGAGACGCGCTTCACA
		2265 T 2314 G	00000000001111111111100001111000010001100000000111
	0	BMEI1645_0	TGCCCTGATTGCGGCGGCTCACCACGTTTTCCAGAACCACGATGGCATCA
		1212 T 1261 A	11111111111111111111111111111111111111111111111111
117 BMEI1629	20	BMEI1629_5	CGCCACTTCCTGCATGGCCTTGATCGCATCACCCGACGTATAGCCCGGAG
		2483 C 2532 G	00011111111110100001100000000111111111111110000111
	20	BMEI1629_4	GGTCACGGTCAATTGCATAGTGCCCGAACCAAGGCTCGAGGACTGCATAT
		230 A 279 C	11111111111111111100010000000111100000001000111111
	21	BMEI1629_3	CGATCTGGAAATAGACGTCATTGGTAAGCCCGCGCAGATCGGTCGCCAGA
		2730 T 2779 G	00000001111111111111111111111111000000000000000011
	21	BMEI1629_2	AACATAGAGTTCAAAGCCGCCCGTCGTGCTGAGACCCATGATCGGAGGCG
		1994 C 2043 T	00011111000000010000100111111111110000111111111110
	35	BMEI1629_1	CGAACTGCTTGTACATTTCGCCAACCAGACCGCCCATGAACGCAACCGGA
		1368 T 1417 G	00001000111100010000000000001100000111110001000000
	0	BMEI1629_0	TTCTGGCTCGATGAGCGCAGGATGATATTCTCGAAAGCGGCGGCGTCCGG
		724 C 773 A	11111111111111111111111111111111111111111111111111
117 BMEI1615	16	BMEI1615_5	GCGCCAAGGCCGCCAGCACAGGTATAATGCCGCCCGCTAAGATGGAAAAG
		511 C 560 C	00001110010010011111111111110010010011111111111111
	21	BMEI1615_4	GCCAGAAGCCCGCCCACAACCGCCACATAATCGCATTCATCAATTTCGCC
		259 G 308 C	00001100001001100001000111111111111111111111110000
	26	BMEI1615_3	CGAAAAGGAACGGTTCATGCTGCGCGTATCAGTGAAGCCGCTGGCAACCG
		914 C 963 G	00111111111111111111110000100001100001000000000000
	26	BMEI1615_2	TTGAAGCTCAAGGCCGGCTGTTCCAAGCGCACCAGCTCATAGGGGATGAG
		634 C 683 A	11111111000000111111111110000000001000000111110000
	28	BMEI1615_1	TGGTTTCGGCCAGAAGCTCCTTCGCCTTCTTGACCCGCAATTCCTCGCGC
		843 G 892 A	10000111000001111111110010000100001000011111100000
	0	BMEI1615_0	GACATCGATCATCGGCACCGGATCTTCCAGCGTTTCCTCGAAAATGCGGA
		716 T 765 C	11111111111111111111111111111111111111111111111111
117 BMEI0965	20	BMEI0965_5	CGCCGCGAATCGTTGCGCCCAGAATATGCATCGCCATCTTGAAGTCGTTG
		423 C 472 G	00011111111111100001000111110001000000011111111111
	21	BMEI0965_4	CGCTTCCGCCCACCTTCATGCCTTCGACCGAAGCCAGAATACCAATAATC
		1017 G 1066 G	00001000001111111100100001111000010011111111111111
	22	BMEI0965_3	CGACACCCATACGATCCATTTGCGTAACGATGGCGGCAATCGGAATTCCG
		498 C 547 G	00001000001111111111111111111111000000010001111000
	24	BMEI0965_2	GCGTTCGATCAGCCATGCGCCTGCCGTATCAAGGTTGGAAATGCCGGAAA
		188 T 237 C	00011111111111111000000011111110000100000000001111
	30	BMEI0965_1	AAGCCAAGCTTCGTGCACGCCCTTCAGACGCACATCCACATGCTTTGCGG
		254 C 303 T	10000000011000010000111111111110000100001000011100
	0	BMEI0965_0	GAGATCGGACAGGATGGTAAGCAGCGGAAGAATAACCACAAGCGCCACCA
		830 T 879 C	11111111111111111111111111111111111111111111111111

TABLE 2: (Continued)

117 BMEI0450	20	BMEI0450_5	CGACAAGGCCGACCAGAACGCCCGCAGAGGTGAACACATCCGTCATGATG
		453 C 502 G	11111111100010111110100011111111000000001111110000
	21	BMEI0450_4	AATGCGTGTCTTGAGATCGTGAACCTCGATAGCCCCACCCGCATTGGCCG
		665 C 714 T	00000010011111111111111111111111111111100100000000000
	22	BMEI0450_3	ACCGAGGAATTGATGACGTGCCAGCCCTGCCAGAGAATATTGATGGCCAC
		553 G 602 T	00000000011111111110010011110010011111111111111100000
	25	BMEI0450_2	GACCCGAATGGCCCACCACGCGCCGATTGCCGCAATCACATTGACAATCG
		131 C 180 C	00100111100100011111111000000001111111111100000000
	29	BMEI0450_1	GCGACCGTAATCAGAACGCCCTCCACCACAGCCGAAATATATTCCGCCTT
		223 A 272 C	01001000111111100010010000111001000001111000001001
	0	BMEI0450_0	CAGCGCATAGGCCACATATTTCAGCCCCATCACACCGAAAGCAACGGGAA
		44 T 93 G	11111111111111111111111111111111111111111111111111
116 BMEI0816	20	BMEI0816_5	TCGGCCAGCGGCTTCTTGATATGTTCCTGAATGACACGGGCAAGCGGACG
		2140 C 2189 A	00000100011111111111111111111111100001000011111110000
	21	BMEI0816_4	CCACACGGACGGTTCCACCATGCTTGAGCTTGCCGAAGAGCACCTCATCG
		2187 C 2236 G	11111111111111000000111111110001000001001111111000
	21	BMEI0816_3	GCCGTTATGGTCGGTCAGCGCCCCATGATCCATCACCTGCAACAAAATGT
		1760 A 1809 C	11111111001001111100100111111111111000000000111000
	26	BMEI0816_2	GCACAGAACCTGAATCGTGCGGTTTATCTCGCTGTCCCGGCCGATAAGCG
		596 C 645 C	01000000001111111111000011111111000010010000111000
	28	BMEI0816_1	AGCCGATAGCAGCTTTCGCCATATCCGACGCGCCCGCATTGGTCGTCATG
		1842 C 1891 T	00000111100000110000011110000000100111110010011111
	0	BMEI0816_0	GCCTCGATTTCCTTCACGCCGATGGTCTTCTTGCGCTTGTTTTCAGGCAG
		1270 C 1319 C	11111111111111111111111111111111111111111111111111
116 BMEI0600	21	BMEI0600_5	TGGCGCCGCTCCATTTAAGGTCGGTGGAAAGCAGTGCATTCTTATCGCTG
		2814 C 2863 A	00000100001111111100000001111111111111111111100000
	23	BMEI0600_4	TGCGTCAGGAATTGCACAAGCGGCTGCCGGTCGAAATTGACAGTGGGCTG
		3379 C 3428 A	00000100001111110000000011111111111111111111000000
	23	BMEI0600_3	CCTGCTCCATGCGGAACTCGGAAACGTCGAAACGCTCCCGATCGGCCTTG
		744 C 793 G	00011111111111111100000000001111000000000011111111
	24	BMEI0600_2	GCCGGTCAGTTCCGTCAATTGAAGCCCGTCCATGGTTTTCTGAAGCCGGG
		2720 C 2769 C	01000111111000100000001100001011111111111111010000
	25	BMEI0600_1	GCCACGCAGGAACGGCATCAGCTCCGCATCCATCGAAAACCTATCGACGG
		176 C 225 C	00011111111110001100000111111100000001111111100000000
	0	BMEI0600_0	CCCGCACGAAAGAACCATGCAGCGACGTCTTTCCAAGCCCTATTTCCAGA
		1332 T 1381 G	11111111111111111111111111111111111111111111111111
116 BMEI0053	15	BMEI0053_5	GGCAAGCATCGTCTGGCCTTTCTTCGTCTGCCACCATGACAGGCTTCCAG
		572 C 621 C	00011111111111111111111111110000001110000001111111
	20	BMEI0053_4	CCTTGCGCGCCATGGTTTCAATATCACCGCTGCCATCATGACGCACTGTC
		153 G 202 G	11100000000110000111111111111111111111000001111000
	21	BMEI0053_3	CTGCCAGAACGTCCGTCACCTTGGGCTTGCCTTCGGTAAGTGTGCCGGTT
		1533 A 1582 G	00000000011100010001111111111111111111111000100011
	24	BMEI0053_2	CGCTCCCATTTTCCGCAAGCACCGTCCTGTAGCCCAGCGCCGTAACTTTG
		498 C 547 G	00000100111111001000001001111111111000000001111111
	36	BMEI0053_1	CGGCATCGGCGGTTTCAAGTGCCACATCGGTTCCACCACCCATGGCAATG
		2163 C 2212 G	00000000000001111110000000100001110000110000000110
	0	BMEI0053_0	ATGGTCGTGGACTGCCGCTTGCCTTTCCTTGGCCAGTTGCTCCACACCAT
		218 A 267 T	11111111111111111111111111111111111111111111111111
115 BMEII0602	18	BMEII0602_5	ACGCATTCGTCCGCCAGTTCCTCGCCCAGTTCTTTCAGTGCCTTGGATGC
		1531 G 1580 T	00011110001100100111001001111111111100111001111111
	20	BMEII0602_4	CCGAAAAGGCCGATCATCTTCTTCAGCTCCGCCACGCCCTTTTCATTGGC
		2668 G 2717 G	00000000010011111111111111111111110010000001111110
	23	BMEII0602_3	CGATGGTCTGGTTCTTCTCGATGATGATCGCCGGGCCAGTGACAGTCTGG
		1950 C 1999 G	10000011111111111110000111100001110000001111111000
	25	BMEII0602_2	ATTCGTCATTGCCGAATGTCAGATTGTTCATGGTGCCCTGTGCCGCCGCC
		3144 G 3193 T	00000111111100000100001111111111111100001100010000
	29	BMEII0602_1	GCCTGCTACTTCCGTCTCGAAAGCGCGTTCATATTCGCCGTCGAAATGCG
		902 C 951 C	00111110000000111000000000000001111000111111111000
	0	BMEII0602_0	GAGAGAATGGCGAAGTCGAGCTTTTCCAGTGCACGGATCGTACGTCTGGT
		3391 A 3440 C	11111111111111111111111111111111111111111111111111

TABLE 2: (Continued)

115 BMEI1137	20	BMEI1137_5	GGCCCGACCGGAATACCATGAGCTTTGGCCCATTTGATAAAGTCGGCAAC
		1153 G 1202 C	00011100000111111000000111100000011111111111111111
	22	BMEI1137_4	TGGCGTCGGAAAGCTGCACGATACCCATCTTGTTGCCGGTGCGGGTCTTG
		3102 C 3151 A	10000111110001001011111010010001111100010001111111
	23	BMEI1137_3	CGCCCTGCGTTTCCTTGATCAGGTGATCGATCTTCGGGTGAATGGAGGCG
		2091 C 2140 G	00011100001100011111000111111111111110000011000001 1
	23	BMEI1137_2	GCTCGCTGCGTTTCCTGCTACCGTATCGTTCATTCGTCCACCCCGATCTT
		29 A 78 C	00111110001111000000000100000000011111111111111111
	27	BMEI1137_1	GGACCGCATCAGCCGAGCGCAGATAGAGTCGCATAGCCTTCTGCATACGG
		3312 C 3361 C	01000000011000011000000111111111111111100000011000
	0	BMEI1137_0	AGCCCGAAAGCACCTGCGCGATCTGCATCACCTGTTCCTGATAGACGATG
		2142 C 2191 T	11111111111111111111111111111111111111111111111111
115 BMEI0789	20	BMEI0789_5	GCGCGCATCGGCCAATTCACGTTCGAGCTTCTTGCGCTCATCGATAAGCG
		2231 C 2280 C	00001001000011111111111111111111100000011111111000 0
	20	BMEI0789_4	GATCAGCTCGGGATAGGCCTGTCCCATTTCGCGGATGAGAGCTGGCAACA
		977 T 1026 C	11100000000001111111100001111111111111100000011111
	23	BMEI0789_3	CCGAACCAGATCGTCTCGGTCGCAGCTTCGCCAGAACCAGTCCAATTGGC
		1291 G 1340 G	00000101111101000001111111110000000111001001111111
	23	BMEI0789_2	CGCCCATCGCCCAGAAATTATCGCTGGTCGGGATGCGGATGATACGGTCG
		444 C 493 G	00100010011111111111110010001000000000001111111111
	29	BMEI0789_1	ATGGTTTCGCCCTCCCGAACAGACGGCACTTCCACGCCATCACGTACGAG
		1414 C 1463 T	00000110000100011110001000011111000011110000011100
	0	BMEI0789_0	GGTCGGCAATCACACGATGGCTTGCACGGAAATCGCCTTCAGCCTTGACG
		786 C 835 C	11111111111111111111111111111111111111111111111111
114 BMEII0950	18	BMEII0950_5	CGGGGTGTAGGGCAGGCCGTCATCATAGCTCTTGGCGCAGTTTTCATCGC
		482 G 531 G	00111111111111000000111111111100000001111111111000
	23	BMEII0950_4	GGCATAACCGCCAATCATGTGCGTAGGCTTGGTGATGACGCGGGTGACGG
		2579 C 2628 C	00001111000010000111111110000011111110000001111111
	23	BMEII0950_3	GTGGTACATGAAGGTTGTTCCCTCCCTCATGCGCTGCGAAACTACGGCAC
		2462 G 2511 C	11111000000000011111111100000000011111111110000011
	23	BMEII0950_2	GTGATCCTGATAGAGCTGTTGACGGCCGGTCAGCGTGCGCCACGGGATCA
		2117 T 2166 C	01111000011111111111111111000110000000000000001111
	27	BMEII0950_1	ATGAAGTTCAGCACCAGATGCGGCTGTCCATCCGTCTTGGCGCGAATGGC
		2239 G 2288 T	11111111110000000000100011100000001111111100000001
	0	BMEII0950_0	GCCGCGATAGTTCATATCCATGTGGTACCAGTGATTGAGACCGGCGCCAA
		638 T 687 C	11111111111111111111111111111111111111111111111111
114 BMEII0478	14	BMEII0478_5	GCAACCGTTTCTTCAATGTAATGCGCCAGATCTGTGTCCGTGCATCCCGC
		637 G 686 C	11110000111111111111111111111111100010000000111111
	17	BMEII0478_4	AGCGATCATAGGCCGGTACTTTCACGCTGGAAGGCAAGCCCGACATGGTT
		48 A 97 T	00001111111111111111111110000011100000100001111111 1
	27	BMEII0478_3	GATCGAACCCCTCCAGCTTAGGCAAGGTTGGAGCGATCTCGATCTCCATC
		1020 G 1069 C	00000001010000011111111111110000010110000000011111
	28	BMEII0478_2	CGGTGGCCTGCCAGAAAACCCAGATAGGCAAGCGTCGAATGGCTGCCATT
		916 A 965 G	00001000000001111111111111000000001111000001100000
	28	BMEII0478_1	GGGGCGGCGTAGTGAGACACCAACAATTCCCCATTCATTTGACCCGCTCG
		182 C 231 C	00000000001111110000001111110000111111100001000011
	0	BMEII0478_0	GGGCCGAGCGTGGTGTCGCGGGATGAGCCAAATCATAAATAATATCGGGA
		465 T 514 C	11111111111111111111111111111111111111111111111111
114 BMEI1759	16	BMEI1759_5	TCCTCACCCTCTTTCAGCGAGATGGAATTGACGACTGCCTTGCCCTGAAC
		1390 G 1439 A	00000000111111111111111111111111111111111110010000001
	23	BMEI1759_4	ATGGAAACCCAGGCCCTGAATCTGCGTACCCATGGCGCCGTCGAGGATCA
		101 T 150 T	00000100100011111111100010010000001111111111111100
	24	BMEI1759_3	GCGATTGAGATCATAGACCATCGCCTCCATGCCATAATCGGCCTGCGCGA
		317 T 366 C	00011111111111111111110000011000000011110001100000 0
	25	BMEI1759_2	CGATATTCTTGCCGATATCGTGCACATCGCCCTTGACGGTCGCCATCAGC
		2319 G 2368 G	00000011111111000000001000100100111110010011111111
	26	BMEI1759_1	CAATGCCCGCCGTCACGACAAAACCGCCAACATAATCCTGTTTGCCGCTT
		3234 A 3283 G	00000000000111111111000100100011111111111000100100
	0	BMEI1759_0	TGTTCGGTCAGGATTTTATAGGCGCGGGTGCAGATTTCGATCTTGCGGGC
		1531 G 1580 A	11111111111111111111111111111111111111111111111111

TABLE 2: (Continued)

114 BMEI0943	20	BMEI0943_5	GACGCCAGATTGCAGGCCGTGTCGTCAAGGAACATATATTCCGAGCACGG
		1423 C 1472 C	00001001111111111111001000001111111111111100001111000
	21	BMEI0943_4	GCCGATCTTTTCCCACAGGTCGCGGGCTTTCAGCGTCTTCATGACACGGC
		1274 G 1323 C	00000011111000001111100000111111111111111111100000
	21	BMEI0943_3	ACCGAGCGCCAGAGCCACGAGGGAACGTCATTCTCCTCAACCTTCTTGAG
		205 C 254 T	11111110000111110000001111111111111000000111110000
	25	BMEI0943_2	GCCGCCAGCAAGCTTCTTGAACTTCACCAGTGCGAAGTCTGGCTCGATAC
		2063 G 2112 C	00000111111111111111100000001111000000011111100000
	27	BMEI0943_1	CGAGGCGATCGAACACCTGGCAAGCGTCCATTTCCGAACCGTAGCGTTCG
		291 C 340 G	11111111111111110000000000100000000000000011111100
	0	BMEI0943_0	GCAGGCCTTCATGATCGCGGTCAGATGCTGCTTGACGATCTTGGAGCCGG
		926 C 975 C	11111111111111111111111111111111111111111111111111
114 BMEI0501	19	BMEI0501_5	GCAGCCGCATCGCGCATCTGCTTTTCCAGATATTCGACATGGGAGGCCAG
		2221 C 2270 C	00001000111110001000011111111111111111000001111111
	20	BMEI0501_4	GCCGAGCAAAACCTGCGTACGGTCCTGATTTTCCAAACCTTCCACGAGGT
		419 A 468 C	10001000000000011111111111111111111111000001111110
	21	BMEI0501_3	GAAGCTTTTGAACTCGCCATATAGCTGCGCGGCCAGTGTCTTGTTCGGCG
		542 C 591 C	11111111111000100011111001110000010001111111100000
	25	BMEI0501_2	AACAAAGACTTGCCGCCTTCGTCATCCGTAACCACGCGATGCATCTTGCG
		2422 C 2471 T	11111111111110000010011111001000001001100001100000
	29	BMEI0501_1	CACCTGGGCCATGGTGAAGGTCTTGCCGGAACCTGTCACGCCGAGCAAAA
		458 T 507 G	00001111111100001000000000000111111111110001000000
	0	BMEI0501_0	CGGGAATTTTCTTCGGCTGTTTGGGGTTCTTTGCGCGTGGTTTTGCAGCC
		45 G 94 G	11111111111111111111111111111111111111111111111111
113 BMEII0881	18	BMEII0881_5	CTGGCCGACGGAAAGATAATTCGCCGCGCGCCAGTAACGGTTGATGAGTG
		47 C 96 G	00000110000111111111111000011000001111111111111111
	21	BMEII0881_4	GATTGCGCTTCTATCGCTCCCGGTTCGGTGACATTGACCGCATATTGGCG
		1141 C 1190 C	00000000111111111111111110000111111100001111100000
	21	BMEII0881_3	GCGGTTGAGATGCGCATAGATGAAATTGAGGCCGGGTGTCGTGCCCCAAT
		170 A 219 C	11111111111000001111111100000000111111111100000000
	24	BMEII0881_2	TGGCAACCGGCACCTGATGGGCGCGCCAATAGTTTTCAACAACCAGACCG
		894 C 943 A	00001111100000111110000000000011111111111111110000
	29	BMEII0881_1	GGCGAAGATGACGGGCCTGTTGGTGGTGAACATCCGGTCGAATTCCTCAT
		2024 A 2073 C	00011111111111000011110000000011010000000111000000
	0	BMEII0881_0	CGCTGGAAAATCGTGTCCAGTGTTGCGGCCATCTTCTGGTGCATGTCGGC
		736 G 785 G	11111111111111111111111111111111111111111111111111
113 BMEII0382	18	BMEII0382_5	CACCGTTGAGCGTCGTTGCGATATCCTTGGACGATATACCAAGCTTGCGC
		2148 G 2197 G	00001111000000000011111111111111111111111100111100
	18	BMEII0382_4	GTCGCGCGCCTCCACATCCCTGGCAACGACCACCATCTGCCCAAAATATG
		1841 C 1890 C	00000110001111111111110001000000011111111111111111
	24	BMEII0382_3	CCGTGACCTGATCCATCATCTTGTCAAGTTCGGCGCCAAATTCCAGCAGG
		888 C 937 G	00011000000111111111111000000100000111111100111100
	26	BMEII0382_2	GAATGACGCGGCCCGGCTCGTTCCAGTTATAATTGACCACCCCAAGCCGC
		2073 G 2122 C	11111110000000001100100111111111110010011000000000
	27	BMEII0382_1	CCATCATTTCCACCGCAATCATGGCATCATCGACCAGAAGGCCGAGCGCG
		1191 C 1240 G	11111111111010000010000000111100000111100000100001
	0	BMEII0382_0	TGATACGGAACCAGGCAACATAGAGCGCGGGCAGGAACAGGAGCGTGATT
		2982 A 3031 A	11111111111111111111111111111111111111111111111111
113 BMEI1387	14	BMEI1387_5	CAGGATCGTGCCCTTGAAGCCGATTTTGTGCTTGTATTCCACCACGAGGC
		638 G 687 G	11110000100001110000001111111111111111111111111111
	20	BMEI1387_4	CGGCACATTGTTCGGGAACTGATCGGTGTCCCAGCCGGATTGGTAATCGT
		896 A 945 G	00001111111100000111111111110000000000011111111111
	21	BMEI1387_3	GAAGCCTCTTGCCGAACTCGCCGTTCCAGCCCGCATAACGCTCATCAAGC
		1143 G 1192 C	11100000111111111000000110000100000111111111111110
	26	BMEI1387_2	ATCATACTGATGTTTGGTCGGCTCCTGCGGCTTGGGCTCGATCAGGATCG
		680 C 729 T	00001110000111111111111111100000000001000011110000
	32	BMEI1387_1	ATTCGTGCCGCCGGTCGTGAAGCCGCCTGCCAGCAGAACCTGATAATAGG
		959 C 1008 T	11111100000011111111000000000101000000000010000100
	0	BMEI1387_0	TGAGCAGGGTTTCATAGCCTTCGCGACCGCCCCAAAGCACGTAATTCTCT
		543 A 592 A	11111111111111111111111111111111111111111111111111

TABLE 2: (Continued)

113 BMEI1320	18	BMEI1320_5	CGGTTTGTGCTTCGACGGATCAACCTGCCAAAGCTCCTTCAGGCCGATGC
		647 G 696 G	11111111111111111110000000001111111111110000000011
	19	BMEI1320_4	CGCCGATCACCGCGCCCTCGTCATTATAGAGTACTTCGGTCGCAGCGAAG
		423 C 472 G	11111110000000100111111111111111111110010000000110
	22	BMEI1320_3	TATGCGCGCCGACTTCCCCACCCTTTTCCAGCACGACCACGGAAAGTTCG
		114 C 163 A	11100000100111100000110000111111111111111100001100
	26	BMEI1320_2	CCGAAGTCTTCTGCAATTCCATGTCACGCACTTGCAAATGCACCGGCTCG
		1413 C 1462 G	10001000000111111111111110000000000111111100000010
	28	BMEI1320_1	GAGTGAACCACGCGCGCCCTCGCCGATCAGCACATATTTACCGAGCAGCG
		539 C 588 C	00011111100000001111000001000011111111110000100000
	0	BMEI1320_0	AAAGCTGGTTGGTCCACATATCCAGGCCGCCCAGAGCAATGCCGATAGCG
		1200 C 1249 T	11111111111111111111111111111111111111111111111111
113 BMEI0961	17	BMEI0961_5	GCTATAGGCGCGCTCAAGCTGGTTGACCTCGGATTGGACGAGCGTTTCCA
		203 T 252 C	11111110000001000000001111111111111111111111111000
	21	BMEI0961_4	TCGTGGTGCGCTGCTCCAGAATCTCGCCATGATCCTGCAAGGTCTGGCTG
		1713 C 1762 A	00100011111111111111111111100100100000001111100001
	22	BMEI0961_3	CAGACGCCCATCGATATCGGCAAGGTTGCGCGTGGCGGCATCAATCAGCT
		2681 A 2730 G	00110001001111111111111111001000000000111111110000
	22	BMEI0961_2	ATCAGAACATTGCCGACGCTTTCGCTGATCGAGGTCAGATCGTCGCCGAT
		946 A 995 T	00000111111111111111111110000000011111000000011110000
	31	BMEI0961_1	GCGGTTCGTGTTTTCCACAAAGCTCGTCGTAGAGTCCACCAGACGCCCAT
		2720 A 2769 C	00110000001111100000000100100111111000000110001001
	0	BMEI0961_0	CGAGAAGGTTTCGGCGATTTCACGGGTGCGGGCAATGAGGTTTTCGTTGA
		848 T 897 G	11111111111111111111111111111111111111111111111111
113 BMEI0897	14	BMEI0897_5	CCCGCCTGCCCCATGAGGTGAAACAGTTCCGGGTCGGTGGTTTCTTCCAG
		136 C 185 G	00000110001111111111111111111000111000111111111111
	15	BMEI0897_4	ATCACACCCCACGAAATGCCATAGCGCGCGCGGTTGAAACAGCCAAACGG
		733 C 782 T	11111111111111111111111100000000111110010000111110
	24	BMEI0897_3	GCGACGATTGCACGCTCATCATGGAACGATATCCCGAATCAATGCGCTCC
		267 G 316 C	00011111111110000000000011100011110001111110000111
	28	BMEI0897_2	GTGATATTCAATCTGGATACCGTTGCCGCCATGCATATCACGGGCCTGGC
		1025 G 1074 C	11001111110000000100000110000001001111111110010000
	32	BMEI0897_1	TCGTGCGTGCCCTCATAGGTGTTGACGGTTTCCAGATTCTGCGCGTGGCG
		1081 C 1130 A	00011000000011111111111100000100000000000000111000
	0	BMEI0897_0	GACCGGAAACATTCGGCAGGATTGCGTCTTCGGAAACTTCAACCCCATCC
		675 G 724 C	11111111111111111111111111111111111111111111111111
113 BMEI0799	15	BMEI0799_5	GGCGATCAACAAGGGTCTGGTATTCCGGCTCGTCCTTATAGGCTGTGCCG
		1713 C 1762 C	00000111111111111111111001000011111111111111110000
	16	BMEI0799_4	GAAGGCCAGCTCCTGCACCAGTGTCGCGCCAGCCTCCTGCATATGATAGC
		743 G 792 C	11000000001111111111111111100000000111111111111111
	18	BMEI0799_3	CGATGGCGAAGAAGAACGACAGCCGGCCTGCAAAATCATCGACATTGAGG
		840 C 889 G	00001100000011111111100000010011111111111111111111
	31	BMEI0799_2	CGTATAAAGCGGCTTGACGTCGATGCCTTCCGGTGTGTGCCAGACGAGGC
		122 G 171 G	00111100000001001001111100000000001111110000010000
	33	BMEI0799_1	CGGCTTCTTCGGATCAAATTCCTTCATGATGCGTGACCAGAGCAGGCGCG
		929 C 978 G	00000000000100000000000000001111111111111111100000
	0	BMEI0799_0	GGTGTTGCGGACCATGAACTCCTTGAGAATGTCGTTCTGAATGGTGCCGG
		596 C 645 C	11111111111111111111111111111111111111111111111111
113 BMEI0449	15	BMEI0449_5	CCAGATCGGGATTGACGCGGTCGAATATCTCCTCTGCCACCGGTGAACGC
		1674 G 1723 G	00000100001111111111111111111111111111000000011111111
	19	BMEI0449_4	TCATATCGCCATTGAAATGCATCATGCCGATCGCGCCGCCATACCATGCG
		1329 C 1378 A	11110000100011111111111000100001111111111111001000
	23	BMEI0449_3	TCGGCTTCTTCCTCGTCAGGGTTGGAATCGAACAGAAGCGTCGCCCCTGC
		1435 G 1484 A	00000111111111111111111111111100100000000011100000000
	25	BMEI0449_2	AACGCGGCTCTTGTCGTTGCGGTCCACATCCGAACACATGGTCAGCTCGG
		1070 C 1119 T	11110001011110001000011111110000000011111111111000000
	31	BMEI0449_1	GATGATCCCGTCTTCCGTTTCGGCTGTGACGAGAAAATCATCCGGCAGGC
		1994 G 2043 C	00000000000011111111111111100000001110000000000010
	0	BMEI0449_0	TCGACGATGATACCGCCTGCCGTTTCGTGCTGGAATATCTCGCTATCCGC
		16 G 65 A	11111111111111111111111111111111111111111111111111

TABLE 2: (Continued)

112 BMEII0671	17	BMEII0671_5	AGCATCTGACATCGGCGCCGACATGAACCATTTATGGCCCGTAAGCCGCC
		644 G 693 T	11110000000100000011111111111111111111111111110000
	20	BMEII0671_4	ACGTCCAGCGCCATCACATTGCCGGAACCTTCCCATATCGCATTGACCGG
		1279 C 1328 T	11111100000001110010001111111111111111100010011000
	20	BMEII0671_3	CCCATGCCAAGCGTAACACCCTGTTTGCGGAAGGCAGGCTTTTGCGAAAA
		520 T 569 G	10000010000001111100000011000111111111111111111111
	27	BMEII0671_2	CGCAATCTGCATGAGAAGCGGGTCGCCCAGATAAGCGTTCGTGCCGGTGA
		65 T 114 G	00011000111110001000011000001111000001111100001111
	28	BMEII0671_1	CGCAGGCGCAATCTTGCAGACCCAATATTTCACCACCGGCGTCATGGAGC
		1148 G 1197 G	00010000100010001111110001000100000000011111111111
	0	BMEII0671_0	GGCGGCCACTCTCCAGCGGATTGTCTTCCCATATGGACGAATGCAGCCCC
		312 G 361 C	11111111111111111111111111111111111111111111111111
112 BMEI2002	11	BMEI2002_5	CTCGACCCAGGCATCGCCATTGTCGCCCTTGACGATCTTGTAAGGAACAA
		275 T 324 G	11111111111111111111111111111111111100000000111110000
	15	BMEI2002_4	CGCTTCTTGTCGGCCTCGGCATTGGCTTCGGCGTCCTTGACCATCTTTTC
		1534 G 1583 G	00111111111111111111111100000000001000111111111111
	22	BMEI2002_3	CGGCCGACTGGTTGTCCTCAGCCGTGGAGAAGGTCTGCGACTTCTTGGTC
		1254 G 1303 G	00000001111111111000000000011111000111000111111111
	31	BMEI2002_2	ACACGTCGAAAGTACCGCCGCCAAGGTCATAGACAGCGATGGTCTTGCCT
		567 A 616 T	00000011111110000000000111111111100000000001110000000
	33	BMEI2002_1	CTTCCTTGAGGCGCTGCAGGGCAAGCTTGTCGTTCTTCAGGTCGATGCCA
		741 T 790 G	01100111100100000000100001111111000100100000000000
	0	BMEI2002_0	GCGGATCTGGTGCTCCTTGCCGGTACCCTTATCCTTGGCCGAAACGTTGA
		1448 T 1497 C	11111111111111111111111111111111111111111111111111
112 BMEI1766	18	BMEI1766_5	GTGATCTGGTAAAGCTCCTCACCCTTCTTCTCCACGCCCAGAATGCCGAT
		1408 A 1457 C	00001111111111111110000100011111111111111111100000001
	21	BMEI1766_4	CACCAATGGCCGAGTTTTCGTCCGCCGAGCCGCCAAGCGTGATCTGGTAA
		1446 T 1495 G	00001110010001111000100111111111110001000011111111
	24	BMEI1766_3	CACATCGGCCACCAGTTCCATCTGTTCAGCACTGGCATCGCCCGGAATGC
		1037 G 1086 G	11111000100000001111111111111000000011101000110000
	24	BMEI1766_2	TTTCCACCTCTGCCAGCTCTGCCAAAATCTGCGGAACATCCTTCAGCGCA
		288 T 337 A	10000000110010011110010011110000011111111111110000
	25	BMEI1766_1	AGCTCGCCCGCCTCGTTCTTCTTCAGTTGCAGGCCGATATCGTGCACCTG
		541 C 590 T	00110001000011111111110000010000111111111110000 1000
	0	BMEI1766_0	GCGTTTGAAAGGCTCCATTCCCACCCGGCGATAAGCGGCAAGGAAAGTTT
		1593 A 1642 C	11111111111111111111111111111111111111111111111111
112 BMEI1529	18	BMEI1529_5	CGAGGGCATTATCCACGTCCTGCTCGCCATCCACATCGCCGGAATTTTCC
		1719 G 1768 G	00001111111111111111110000000111111111100001111000
	18	BMEI1529_4	AGATGGGCGACGTAGAAATCACCCTTCTTCGGATCGGAATGGACATGCGC
		298 G 347 T	00010000001111111100000010001111111111111111111111
	21	BMEI1529_3	GGGTGACGAAGCATTTCTGGTTGGCGCGGATGGTAAGGCGGATGACTTCC
		840 G 889 C	11111111000100000000111100011111111111000011111000
	27	BMEI1529_2	AGCAGATTGTCGGCTTCCGGCGTCAACACCGCATCAATAGCGTCGTAGCG
		1885 C 1934 T	00011111111110000000001111000100011111111000010001000
	28	BMEI1529_1	CGGACACTTCTTCCAGAAGCCCGTCGTCTTCCACCAGCTCAAGGCCGGAA
		726 T 775 G	00011000000111000000000111111111110000011111000000 11
	0	BMEI1529_0	TTCCGGCGCAGAAACAGATGGCCCCTTGATGTCCTCGTGCACGTCCTTGG
		197 C 246 A	11111111111111111111111111111111111111111111111111
112 BMEI1223	10	BMEI1223_5	CCATTCTTCGAGCGAAAGACCTGTCACTGCCTTCGATACAGCTTCGGCGA
		1571 T 1620 G	10000111111111111111111111111111111111111111111000000
	20	BMEI1223_4	CACTGCAGCTGTAATGGATGCCATGCACAAGAACGCCGTGCCTGCCAATT
		1416 A 1465 G	11111111111111111100100100000011111000000100100111
	23	BMEI1223_3	ATTCACGCGCACCCGAACCATCCGATTATGACCGGTTAACAGCGCTTGCT
		1472 A 1521 T	11111000000111111111110000011110000011110000011111000
	26	BMEI1223_2	GGCAGAAATCGCGTAACAGGCTTGCCCACTTTGTGATTCCCGGCCGGCAG
		259 C 308 C	00010001110001000000000000111111111111000001111111
	33	BMEI1223_1	CCGAGTGTACGGCAAGCCTTGTGTTTCGCTTCAGCCTCCAGGAAAGCTTC
		145 G 194 G	00000111110000000001110000100000010001111000000 0111
	0	BMEI1223_0	TCTCTCTTCAGGCCTCAGAGCATCAAGTGCACCTGATCGACCAAAGAGGG
		749 C 798 A	11111111111111111111111111111111111111111111111111

TABLE 2: (Continued)

112 BMEI0878	19	BMEI0878_5	CTTGCCCAAGGCTTCCAGCGTCTGGTTGTAATAGGGCGACGAGGAGCGCG
		995 C 1044 G	00011110000011000111110001111111111111111111100000
	20	BMEI0878_4	GCCAGCGTCAGGTAGTCGAGCCCCACATCGTTGAGGAATTGCAGGCGCTC
		1462 G 1511 C	00000000111111111111111111111111111100000000000011
	21	BMEI0878_3	GCTGGCCTTCGGCATAGATCGTATCGAAAGCAAGCGAGGATTTGCCCGAG
		111 C 160 C	00010000000111111111111100000001000011111111111111
	25	BMEI0878_2	AGTGCGATGACCCTATCGACCATCTGGCTCACCGTCTGGCTTTCGATCGG
		385 C 434 T	00110000001001111111111110010000001000011111111000
	27	BMEI0878_1	GACGCAGCCTTGAACAACGTCTCGATCAGGAAGGTGGACTTGCCTCCGCC
		1996 G 2045 C	00000100111110010000010000001111111111111001000010
	0	BMEI0878_0	AAGTGCATCTCGATCTTGATGACGCCGTCGCCCTGACATGCCTCGCAGCG
		2281 C 2330 T	11111111111111111111111111111111111111111111111111
112 BMEI0647	15	BMEI0647_5	GCAGAACCATAGGCTGCGAACATGGGGCGGTAATACATGGGCTGCGGGGT
		1417 A 1466 C	00001111111110000011111111111111111100000011111111
	20	BMEI0647_4	ATCTGCGAATTTGCCGGACTCTATCGAGCCGATTACATCGCCGACACCGG
		1262 C 1311 T	11111111111111111000011110000000011100001111110000
	23	BMEI0647_3	CACCTGCTTCCAGCTGCTCGATCAGAGGTGCAATGCCGGTGGAATTTCCC
		606 G 655 G	00000000111000111111111110000001110000001111111111
	26	BMEI0647_2	CGTTCGAGAGAGCGGCCTCCACCTGCTGCGGCGCGATATAGTGAACATGG
		417 C 466 G	00111111111100001100000100000000011111111111000000
	28	BMEI0647_1	CGCCAAGCCTTATCTTGTCCCCAATTGTCGGGCCATAAAGGTCGGCGTAT
		24 A 73 G	00000110000111100100011111111000100110000100000111
	0	BMEI0647_0	GGATTGGTCGAGCTTGGCAGGATGTTATTCTGGCCCACCACTTTCAGCAG
		868 C 917 C	11111111111111111111111111111111111111111111111111
111 BMEII0339	17	BMEII0339_5	CGCCTTGCCTTCACCGTGCGGATATTGCCATCCTGATCGAAAGTAACGCC
		691 G 740 G	00011000011111111110000011110000011111111111111111
	19	BMEII0339_4	CTTGGCGGTGTTCCAGCGGATACCGTGCTTCTGGTTCACCTTGAAATAGG
		548 C 597 G	11111100001100001111111111111100001100001111111000
	23	BMEII0339_3	ATTCCAGAATCTGCGGCGATCCCACAGCCCCGGCAGCGAGAATGACTTCC
		741 G 790 T	11111111111100000000111111111110000000000011110000
	25	BMEII0339_2	GGACGATATTTCTGCAAGGGCGCTTGTGAAACGATATGGCGGGTGAGGCG
		1273 C 1322 C	00100011100100000111000111111111111110000011100000
	27	BMEII0339_1	CGCCTGCCGCCTTTTCCAGATCTTCCTGCGTTGCATAATTCGGGCCGGGT
		1335 A 1384 G	00000100000111111111111111100000000100100011000011
	0	BMEII0339_0	CCCACAAGCGTACTTGCCTTCTCGTTGAGTGTCGCAATGCCGGTGACTTT
		904 A 953 G	11111111111111111111111111111111111111111111111111
111 BMEII0149	17	BMEII0149_5	CGTCGTATCGAAGGCGCCGAGATCGAGTGTGCCACCATTGACTGTCGCGG
		1526 C 1575 G	00100111111111001001111111000111000111111111100011
	18	BMEII0149_4	CTGGATTATTGCCGCGCCGCCGAAAGTGTTGTCGCCGCTCACCTTGATGG
		1061 C 1110 G	11111111110000000001000111111110001001111111111110
	22	BMEII0149_3	TGCCGCTGAAGGTATTGGCACCCGTCAACGACAGTATACCGTCTCCCGCC
		360 G 409 A	00000000010011111111100100000000011111111111111111
	25	BMEII0149_2	GCCGTGAAGTTGCGATCGCTGCTGTCACCGGCACCCAAATAGCTCAGTAT
		1210 A 1259 C	00100001000000110000111110000000001111111111111111
	29	BMEII0149_1	TGCACCGTCGACGCGTAAATTGCCTGTTCCAGAAACAGTGCCGGCAAGCG
		1370 C 1419 A	11100001000000111111111000111000111110000000000000
	0	BMEII0149_0	CCGCTAACGCCATCGCCGCTTGGACTCCAGAAGGCATTACTCGTATTCCA
		79 T 128 G	11111111111111111111111111111111111111111111111111
111 BMEI1076	20	BMEI1076_5	GCCAGCACCGAAATGCCGAGCGTTACGAACTGGAGGCGCATGAAGGGAAT
		1408 A 1457 C	00001100000000000010000011111111111111111111111111
	20	BMEI1076_4	CCGCATGATATTGGCAGTCGTCTCCAGATCGGTCTTGCCTGCATTCTGCA
		233 T 282 G	00111111111100000100000111111111111111100001110000
	22	BMEI1076_3	GGGCGCTATCGATATTCAGCTCCGCGAGCCGCTCGTTGATATCTTCCAGA
		1563 T 1612 C	00000111111111110000011111000000111111111000011100
	22	BMEI1076_2	AGAAGCTGCGCATCGTAAAGGCCCGGAACCTCCACGCGAATCTGATTGCG
		334 C 383 T	11110011100110000000001111100000000011111111111111
	27	BMEI1076_1	ACGGTCTGCTCCGCATCGTCGCCCACTTCCTGACTGCCGATAATGACAAG
		1645 C 1694 T	00000100011100011110000000100000100001111111111111
	0	BMEI1076_0	GATTGCGCCCCTCTCCCTCGACAACGGGATTGCCATAGCCAAGCTCTTCC
		291 G 340 C	11111111111111111111111111111111111111111111111111

TABLE 2: (Continued)

```
110 BMEII0504  16      BMEII0504_5     GCATTCCGTCCTTGCCGCGCCTGTAGCCCGCTTCATCCAAAAGTGCGTTC
                       1068 G 1117 C   00111000100100000000011111111111111111111111111111
               17      BMEII0504_4     GCACGATCCGGCAGCACACGGTAGATTACTTCATCCAGATAGGGCTCACC
                       676 G 725 C     11100000010000011111111111111111111111110000111100
               20      BMEII0504_3     TTCTGCACTTCTTGCGCTGCCCATGTAGCTTTTGGCCAGAATGCGAGCGC
                       58 G 107 A      11111111111111110001000111111111000100011000010000
               22      BMEII0504_2     GGCGACATCGGCTGAAGTAAACGGCTTGCCATCGTGCCAGTTGACGCCTT
                       317 A 366 C     00001001000011111111111111111111111000011001000000
               35      BMEII0504_1     CATTGGCATAGCCGCCCTGATTGGAGAAAGGAACACCCGCCGGAATGCCG
                       1344 C 1393 G   10000100100000011111111111000000000100000000000000
               0       BMEII0504_0     ACACGACACGGCCAAGGTTTTGCAGTGGCTTCCAGACTTCCATCGCGGAA
                       369 T 418 T     11111111111111111111111111111111111111111111111111

110 BMEI1281   17      BMEI1281_5      GAGATTGCCAAGCGCCTTGTAATCATCCGCCGACAGGGTGAGGTGATGCG
                       872 C 921 C     11111001000000001111111111000011111111111111111000
               19      BMEI1281_4      CGTCCGTTACGAATGCCAATGTCGCGCTGGCCGATACCGTCATGATTGAC
                       145 G 194 G     11111111111111000011111111100001111000000111110000
               22      BMEI1281_3      AGCCGGCTTTCGATCCTGCCTGTTCATGCGTGATCGTTTCGCGGCGTTTC
                       1245 G 1294 T   00000001110000011111111111111110000011100000111111
               22      BMEI1281_2      CCATACCGGATGGCTCGATGGATCACCCTCAACGCGCAATCCTTCACGTT
                       671 A 720 G     00001111100000100011111111111100000011111111100000
               30      BMEI1281_1      CACAACGCCGGGTAGGATATGCAGGCCAGTGCAGTCAATCACTTCACCGG
                       230 C 279 G     00001100000001111111111000000010001111100010000000
               0       BMEI1281_0      ATCGAATGTTGCGGCCATGGGTGTCTCCTGCTTTTCGGATGGGACTTGCG
                       71 C 120 T      11111111111111111111111111111111111111111111111111

110 BMEI1055   19      BMEI1055_5      GGGCCGCAAGATAGGCACTTTCCGCAATGGTCAGTTCGCCCACCGACTTG
                       567 C 616 C     00000011111111111111111111111111100001001100000000111
               19      BMEI1055_4      ATTCTTCAGATTCGTGACCACCGCACGCGCAAGGCCACCGAAATCGAGGC
                       287 G 336 T     11111111111100001111111000011111000011001111001000
               22      BMEI1055_3      TTCCGCCGATATTGACGATGCGCATTGCCCATTTCATATCGTCGGCAGCG
                       1161 C 1210 A   11000100000011111111111111111111111000000100010000
               23      BMEI1055_2      CTTGCGATCATAGGTCTGCTTGGAGCTCAGCAGGAAGTTCTTCGCCACCT
                       383 A 432 G     00000000111111100000111111110000011111111111000001
               27      BMEI1055_1      CGGCACACGGAAATCAACCTTCGGCGTTCGAGCCAAAGCCTGCTCCATGA
                       2195 T 2244 G   00111110000011000100000001111110000000100011111111
               0       BMEI1055_0      TTGCGCGCTTCAAGCTGCAATTTGGGATTGAGGGTCGTGCGAACCGACAA
                       871 T 920 A     11111111111111111111111111111111111111111111111111

110 BMEI1042   18      BMEI1042_5      ATAGGTGTTGTTCGACTTGCCAGCCGAAATGCCCTTGGAGATGATGCGGC
                       1145 G 1194 T   11111111110000000111100000001111111111111111110000
               18      BMEI1042_4      TTCTCGTTGATACGGAAATAGGTCGAAAGCTCCATCGGGCAGCGAACGCC
                       673 G 722 A     11110001111000111111111111111111000110000011111000
               18      BMEI1042_3      TGAGAAGTTCGGGATCCACCTCATCCAGCGACTTCGGGCCGGACTGGTTC
                       279 G 328 A     00000001111111111111111111111111110010000001000111
               22      BMEI1042_2      GGCAATCGCCGCGCTTGGTGACGAAATTATAGATACCGCCCTTGCCGTCC
                       909 G 958 C     00011110000111100000001111111111100100111100010011
               34      BMEI1042_1      CGCTTCCGAAATCGAGCAGAAGATGACGCCCGCCTTCGACAGTTCTTCCT
                       500 A 549 G     00000000111111100001000001111000010000000001000011
               0       BMEI1042_0      CGATTTTCGGATAATCAACGCGCGCCCAGGTCGGCTCTTCCATGGTGAGC
                       195 G 244 G     11111111111111111111111111111111111111111111111111

110 BMEI0414   14      BMEI0414_5      GCGATGAGAATGGTGCAGACCGCGTTGATTTCCGGCGTCACACCAAGGCG
                       676 C 725 C     00111111111100000000000011111111111111111111111111
               21      BMEI0414_4      ATTGCGGAACAATTCGCCATACCATTTGGTCGAGAACCCGGCCCAGACCG
                       116 C 165 T     00000001111111111111000111100011111111000011110000
               22      BMEI0414_3      CAGATCGTCGAGCGACAGGGTGAAAGCAAGCATCCAGCCGGAAACCACGG
                       560 C 609 G     01100001111100001100001111111111100100001111111000
               23      BMEI0414_2      AACCGCGGTCGAAATTCATCGCCACGAAGAGCAGAAGCAGCGACAGGCCG
                       342 C 391 T     11110000000111111000000111111111111111000000110000
               30      BMEI0414_1      CCGCTTCTTCCAGTGACCGGTCAAAGCTGACGAGGCGCGATTGCACCACG
                       444 C 493 G     00001111111111111000000000100000100000111100000100
               0       BMEI0414_0      GCAGGTAGAGAAAGGCAAAACCCAGCGCCACCGATACGATATTGAAGCGG
                       18 C 67 C       11111111111111111111111111111111111111111111111111
```

TABLE 2: (Continued)

109 BMEII1056	13	BMEII1056_5	CTGCATGTCGGCCAGTTCCTGTACGCCTTCCTCACCCTTGGCATTGCCTG
		791 C 840 G	11111111111100011100011111111111111111111000001111100
	18	BMEII1056_4	CGAGCTTGTCGATGGCGCGCAACACATTGAGACGCTTGGCTGCATTGCCT
		564 A 613 G	00011111111111000000000011111111111111110001111110001
	22	BMEII1056_3	GATAAGGCTTCGGCAGGGTCTCGAAATTTTCCGCCACATAGCGCGCCAGC
		273 G 322 C	11111100110000100001011111111000001111111000000111
	24	BMEII1056_2	CGGCGTTTCCACCGGTTCGAAGCCATAAAGGTCATAGACTTCGCGGATGG
		98 C 147 G	00000111100000111000000011111111111100000001111111
	32	BMEII1056_1	CAATTGCGCACCCTTGGTAAAGTCGCCACTCTCGTCCAGACGCCCCTTGC
		650 G 699 G	01100000000011111111110001001100001001000000001000
	0	BMEII1056_0	TTGCGTATCCTTGTCCATCACCAGCACGACCACTGGGCCAACCGTGTCGG
		1121 C 1170 A	11111111111111111111111111111111111111111111111111
109 BMEII0356	20	BMEII0356_5	GTGATGATAATCCTCCGGCCCATCGAACACGACCGCCTTGCCTTCAAAGG
		1304 C 1353 C	10000001111110000001111111111111111111000001111000
	20	BMEII0356_4	ATTTCTGCCAGTCATCCACCGTCAGTTCCACACCGATATGGCGCGCGAGC
		909 G 958 T	11110001000011111000010001111111111111110000111100
	21	BMEII0356_3	TCCTGATATTTGACCGCGTTTTCCAGCACGGCCCCGGTTTCCATCTGGCC
		1759 G 1808 A	11111111111110000011100000000000011111111111110000
	23	BMEII0356_2	TTCAGGTCTTCATGAACCATCTCGACAATACGCCGCCCCGTCAGATAGGC
		763 G 812 A	01111111111111000000110000111110000110000000011111
	25	BMEII0356_1	TTCCTGAATGGGATGAACCGGAAATTCGATGACGATGCCACCCGCCTCGC
		278 G 327 A	00001111110001100000011111111110000100111110010000
	0	BMEII0356_0	ATGGGACGGCCGGATTGCAGCTCTTCCTGGCTCAGCCCGTAATTCATGTA
		142 T 191 T	11111111111111111111111111111111111111111111111111
109 BMEII0239	16	BMEII0239_5	ATAAAGGCTCACCGCATCGCCCAGAATTTTCATCGCGCCGAACTGCCGCC
		314 G 363 T	11111111100100011111111111111111111000100111100000
	17	BMEII0239_4	TCAAAAGCTCGCCGGTCGCAAAACGCCAGTCTTCCGAGGCAATATGGAAG
		834 C 883 A	10111111110100010011111111111100000111111111100000
	21	BMEII0239_3	CCGTCACCAGCGCACTTATCAATTCGCGGTCCGGCAAAAGATAGGGCGTC
		1038 G 1087 G	00011000100010001111000100000011111111111111111111
	26	BMEII0239_2	GCATGGCGCGGTCCACCAGCTTCAGATTGTTGACACCCGGCACCACGATA
		1113 T 1162 C	00111000001100000000001111111111111110000010000111
	29	BMEII0239_1	CCCATCATCATGCGATGGTTCGTTTCGAGGTTCTTGTCCGGCCCTGACCC
		952 G 1001 G	00001001000100100000100010001001001111111111100100
	0	BMEII0239_0	CGCCATCGATGGTCATGAGCTTTGAGTGATTGAAGGCCCCGCCAGCACGC
		1206 G 1255 G	11111111111111111111111111111111111111111111111111
109 BMEI1720	16	BMEI1720_5	GGGCGGTGATATTTGAACGAGCGGCCCGTCAGATGTTCGCCATTGGCCAA
		121 T 170 C	01111111111111111001000001111111111110001110000011
	21	BMEI1720_4	GTCTCGCTCCAATAGGAAGACGTGACGTAGCCAAGCATGGTCATCGGCAC
		2800 G 2849 C	01110000111110000000001110000000111111111111111111
	23	BMEI1720_3	GGCACGTTGATCTCGAAGCCGAGTTCGCCGGTGAAGCTCATGCGGAAGAG
		2419 C 2468 C	00000010001000011110001100000011111111111111111111
	24	BMEI1720_2	GAAGCGAACCGCCCAGTTCCGCCTGTTCGTCGCAAAGCATAACCTTCGCG
		579 C 628 C	11111000000100011110000001100100100000111111111111
	25	BMEI1720_1	CGCGATTGACGAGGCTGCCCGCAACATGCTCACGCACATCGATGATGGCA
		1056 T 1105 G	00011100001100000100000001000111100011111111111111
	0	BMEI1720_0	GTGAAGCGAACGGACTTGGCCTTGTTGACGCGGCTTTTGTTCTGAATGCG
		16 C 65 C	11111111111111111111111111111111111111111111111111
109 BMEI1488	12	BMEI1488_5	TATCCGACGTTGCCTGGAAGATCGCGCCGGAAGCAATCAACTGTTTGCGC
		456 G 505 A	11100000011111111111111110000001111111111111111111
	16	BMEI1488_4	TTCAACCTGGCGGATCGCTTCCACGAAACGGTCGGTTGAGGAGGAGTGCC
		539 G 588 A	11111111000001100011110001111111111111111111100000
	24	BMEI1488_3	GTGTAGGCTCAATGAAGGTGCGGCCCACATAATGGTTGCGGATGATGCCA
		1041 T 1090 C	11111111111100000010000100111111111110000000001000
	25	BMEI1488_2	GTTCATGCCCATGTTCTTGCGCGCCACATAGACGCTGCGGCCACCCACTA
		884 T 933 C	11110000010000111111000001111100000110000001111111
	32	BMEI1488_1	GAATCATCCACCAGAACGACCCGCTTGCCTTCAATCATAGCGCGATTGGC
		1129 G 1178 C	01000000000111111001000011100000011111100000100000
	0	BMEI1488_0	ATTGCGCAGGATCGTCTCGCCGGTTGTGGAATAACGCACATGGCCGATGG
		326 C 375 T	11111111111111111111111111111111111111111111111111

TABLE 2: (Continued)

109 BMEI1471	13	BMEI1471_5	CCACGATGGAGGCAAGCGTCACAAATTCGTTCTTGTCCCGAACCGGCAGG
		753 C 802 G	00011100000111111111111111111111111111111111111100000
	18	BMEI1471_4	ATGCCGACGCGCCAAGCAAAACCAGCACGACAAGCGACAGCATGAAGTTC
		213 G 262 T	00000010001111111111111111111111100010000001111111
	20	BMEI1471_3	TACATGGCCGCCCGTGCCGTCCGCAACAAAATAGAGGTCGTCGGTCTGGA
		1064 T 1113 A	11111111000110001001000001111111111111001000001111
	28	BMEI1471_2	TGGGCTTCTCGATATCCGACTTGAATATCGGGCGGTCTGACGGTTTGCCT
		924 A 973 A	00000001100000000111110000000000000111111111111111
	30	BMEI1471_1	GGAATGGCATATTCGCCCGCCTTCAGGTCGTTTTCATGCCCGAGCGTGCG
		418 C 467 C	11100000000110001000110000011111111111000000001000
	0	BMEI1471_0	CCCAAAGGTCGTCCACCAGTTTCTTCTGCGAGGCGATCATGCGGTCAATT
		687 A 736 G	11111111111111111111111111111111111111111111111111
109 BMEI1327	19	BMEI1327_5	GCCTCCAGCGCCCTGCCAACCATAAGTTCGGCAAGATCGGAAAAGGCCTG
		1861 C 1910 C	00010001111111111111111111111000000000111111111000011
	20	BMEI1327_4	GATCAGGTCCACATCCGAGCCCGCCGTCAGTTCGCGACTGCCAAGCTTAC
		1979 G 2028 C	11111111111111111000010010000111110000100110000111
	22	BMEI1327_3	CCGCATACTGCTTTTCCACCACCTTGAGCGCGGCGAGGAATATTTCGGAA
		1203 T 1252 G	00011100001111111111111111000000000001111111100001
	24	BMEI1327_2	GGGAAACGCGGGCAAACCCTTCATCATCTTCAGGCAGGATATGGGTCTGC
		1122 G 1171 C	00011100000000000001111111111111111111111111000000000
	24	BMEI1327_1	CCTTATGGGCGTGAATCTGGCGCTTGATCGAATGGACATCGGCAATTGCC
		816 G 865 G	00111100001111110000000111100001111111100001111000
	0	BMEI1327_0	AATAGTTGAGCTCGCGCGCGCCGAACTTGCCCATGCCAAGAACGATCCAG
		414 C 463 T	11111111111111111111111111111111111111111111111111
109 BMEI1089	15	BMEI1089_5	CGCGCCGAGCACATAGATCATCTCATTGAAGCCGTGGTCGTACTTGTCCT
		1025 A 1074 G	11111111111111111111111111110000000001111110010010000
	21	BMEI1089_4	TGGCCGATACATATTCGACATTGACCTTCTTGCCCGCGCCAAGCCTGGAG
		345 C 394 A	00010000111111111111111111111000100001000111100010
	23	BMEI1089_3	TCGGACTTCATCAGCGGACGGTCAATATCATCGCCCACTTCCGTAGAACG
		934 C 983 A	11111111111100000000111111111110000000010000111000
	24	BMEI1089_2	CGCTCTCGTCGGTCAGCTTTTCAAAATGCGACCCCATGGCCACGCGGTCA
		1431 T 1480 G	00010000011111111111111111110000010001111000100000
	26	BMEI1089_1	CGTGCGGTCCTTCACGATGGCCAGTGCTTCGGCTTCAGGCATTTCCAGAA
		665 T 714 G	00000001001111111110010001000000001111000011111111
	0	BMEI1089_0	GACGCTTGACATAACCGCCATGATCCGCGCCGAGCACATAGATCATCTCA
		1050 T 1099 C	11111111111111111111111111111111111111111111111111
109 BMEI0275	14	BMEI0275_5	CCTGTTCCGTCCGCTTGTTGCCCTGATAGACATGATCGCGCTTTGGCTGG
		2964 C 3013 G	00000011111111111111111111111111111111110000000011111
	18	BMEI0275_4	CGCTTGTCCTGCGGTCTACAAAGCGTTTCGTCAACCTCTCATGCAGCGCA
		1401 T 1450 G	00000111111111111111100000011111111111111110000111000
	21	BMEI0275_3	GAAACGGAATCCTTCCAGTTGCCCGACATTATGGCCCTCGACGATCACGT
		1511 A 1560 C	11110000011100000111111111111111100000000100111110
	25	BMEI0275_2	CGTGCGCCATTGCAGCACCTTCACCGGATCGAAATTATGCGCTTCCACCC
		893 G 942 G	00000001001110000000000010011111111111111111111100100
	31	BMEI0275_1	GCGAGACGGTAGAACATCGGATCGAAGGCCGGATCAACCACGACCGAGGT
		2176 A 2225 C	00011000001110000001110001000001111111110001000000
	0	BMEI0275_0	TGGTTGCCACGATGCTGGTTTCCACGATTCTGGCGCTGGCGATTCTGGTC
		2743 G 2792 A	11111111111111111111111111111111111111111111111111
108 BMEII0531	17	BMEII0531_5	CGGATGCGCCACGATATAAACCGCCGCCACCGACCAGTAAGGATTGGGCA
		107 T 156 G	00000000100111111111111111111111111111111110000000111
	19	BMEII0531_4	CACCCCATGTAGAGGCATCCCTGCTGTGGCGGTCGATTTCCTCGATCAGC
		888 G 937 G	11111111111111000111100011110000000011111100001111 0
	23	BMEII0531_3	CGCAGAGCACCATGCCCAGCAGAAATTGCGACGGGATAACCAGCAGCATA
		1386 T 1435 G	00010000011100000100011111111111111111110001000011111
	23	BMEII0531_2	CAATCGCCACGGTATAGCCAGCCAGCATGAAGAGATAGCTGCGCGGCGTG
		324 C 373 G	00000010001111111111111111111111111001000000000000
	26	BMEII0531_1	TACACGGCCCAGCGCATAATCGAAGGTGGTTTCAGGAGCCTGCACCACTG
		383 C 432 A	11111100001100001111111110000011000111100010000000
	0	BMEII0531_0	CAAGGCGAATTCAAACACGAAAGCGGCGGCAATCACGACCATAAGCAGCC
		1271 G 1320 G	11111111111111111111111111111111111111111111111111

TABLE 2: (Continued)

108 BMEII0392	20	BMEII0392_5	GTCGGCATCCACTTCAATGTCGAGCGAGAGAAAATCCTCATTGCGCCCGC
		491 G 540 C	00011111111111111110000000011111111111111111000000000
	21	BMEII0392_4	ATTTCGGTTCAATCTTGACCAGTTCGGCCAGTGGGTCTTGCAGGCGGCGC
		1146 G 1195 T	11100000111111100000111110000011111111111111000000
	22	BMEII0392_3	CATCGCTTCCAGCAATGCAGCACCAAAAGCGCCGGTTGAAGGCTGTTCCT
		2051 A 2100 G	00000001110000000100111111111111111111111100011111000
	22	BMEII0392_2	CTCCGTCACACGAACCTTCACCACGTCGCCCGCCTTCACCACTTCATGCG
		1877 C 1926 G	11111111111111111111110000110000011110000000000000
	23	BMEII0392_1	CGGCAACATTCGTCACGGTGCCTTCCAGCATCATGCCGGGTTTCAGATCC
		1743 G 1792 G	00011111110010000100001111000010000100111111111111
	0	BMEII0392_0	TTCACATCTGGCACTTCCGGCGTGATGTAGGATGCGGCAAGCTCGGCAGG
		226 C 275 A	11111111111111111111111111111111111111111111111111
108 BMEI1504	12	BMEI1504_5	GCGCCGCAATCGCAATATTGTTGAAAACGCAAAAGCCCATGGCCCTGCTG
		471 C 520 C	00001000111111111111111111111111111111111111000110011
	18	BMEI1504_4	GCCACGAAAACATTGTCCGCAGCACCCGACATGACATCATCGACCGCGGC
		391 G 440 C	00111111111111111100000000001111111111111111000000
	22	BMEI1504_3	CAAGGCGGTAGAAATCCGGCCCCTCCAGCTCGCTCATCAACGCGCGAATG
		141 C 190 G	11111111111111111100000100010000010011110010000011
	26	BMEI1504_2	ATCCTGCGTGCCATTGCCGTGGTGTACATCCCAATCGACAATTGCGATGC
		551 G 600 T	00001000011111111111111110000010000111110000100000
	30	BMEI1504_1	ATATTGCCCACGCCCGTTTCGTGCTTGTCACCACTTCCCGGATAGAGCGG
		649 C 698 T	11111000000001100000000111110000010000001111111000
	0	BMEI1504_0	CTCCATGAGCTTGCCCGTCGCCCAATCGAAATCCGACTCATCGAGATTGA
		851 T 900 G	11111111111111111111111111111111111111111111111111
108 BMEI1155	15	BMEI1155_5	GGCCTGCGTGGTCACGTTGAGAAGGTGGTTCAGGATACGGCCGATTTCCG
		311 C 360 C	00100100100110000111111111111111111111111111111111000
	16	BMEI1155_4	CACGTGGTAGCCTTCCGTGTAAAGCTTGAAGTGATGGATGAGCGCCTCCA
		923 T 972 G	11100001000011000111111111111111111111111110000011111
	19	BMEI1155_3	GATAGGTCTTGGCTTCCATCAGCTTTTCGGTGCCGCGATGCAGCAGACCG
		126 C 175 C	11111111111111111111111111111100000000001110000001000
	28	BMEI1155_2	CAGCATGTGGCCCCGGCAAAGGAAATCCATGGCTTGAAGATGGGCAAAGC
		1082 G 1131 G	00111110000000000001111110001000001111111000011100
	30	BMEI1155_1	AGAAGCCCCATGCCCATGCGTCTTCCAGCTTCACAACGCCGATATCGACG
		603 C 652 T	11100010000010111111110001000110000000000111100000
	0	BMEI1155_0	TCCAGAAGACGTTCGACGGCAAGCGCATAGGCATGTTCCTGATTCATCGG
		217 C 266 A	11111111111111111111111111111111111111111111111111
108 BMEI0884	17	BMEI0884_5	ACCTTGTCGCCCTCTGCCAGATTGATCCCGCGCACACCGATAGAATTGCG
		2086 C 2135 T	11111111111111111111111111000001110000000011110000
	17	BMEI0884_4	GGGCAATCAGGTCTTCGTCGTCCATTTCCGCGCCGCCAAAACCGATTTCG
		1482 C 1531 C	00111111111111111111111111110000000001111000011110
	20	BMEI0884_3	GCAGACGCCACAGCTTTTCCTTGTAAACGATACCGCGCGACGAGAAGAAC
		1698 G 1747 C	00011000000011111111111111111110000000000111111111
	22	BMEI0884_2	GCACGGTTTCCTCACGAGCCTTCAACTCCTGATACCGATCTTCCGACAGC
		2286 G 2335 C	11000001111000010000111110000100011111111111111000
	32	BMEI0884_1	TCGGGACCGGGAATCAATTCCATCAGTTCCGACAGTTCGATGGCCGGATT
		550 A 599 A	00001000000001111000011000010000111110001001000111
	0	BMEI0884_0	CGGCACATGTCAGCAGCACATCGTCGAATTCCGTACAGGTATCGACCGAG
		1986 C 2035 G	11111111111111111111111111111111111111111111111111
108 BMEI0876	17	BMEI0876_5	TTCAGCGGACATCGGGCTCATGCTGCGCAGTTTCTTCAGCTCGGCCAGAG
		878 C 927 A	11110001111111111111110001110001111111100000011111000
	18	BMEI0876_4	ACCGACTTCTTCTTCGTCTTCAGGATTTCCGTCACGGCCTTGCGCGCAAG
		1717 C 1766 T	00000011111111111111111111111111111100000011000000
	18	BMEI0876_3	GGAATGGAGAGCAACCAGTCGAGATAATTGCGAACCACCGTCGCTTCAGC
		922 G 971 C	11110000110000110000011111111111111111111000111100
	22	BMEI0876_2	GGATGCCGCGGCCGAAATCGATTCCTTCATCACGTCACGGAGATTACCCG
		1943 C 1992 C	11111100001100011000000111111111111100000111110000
	33	BMEI0876_1	GGGCCTGCACCGCCAGATATTCGACGATGCGTTCCTTGACCTTGCCAAGG
		1044 C 1093 C	00000100000000011111111000000000111000000011111000000
	0	BMEI0876_0	CCTCACTGTCGCCAAGCTCCTTCTGGATCGCCTTCATCTGCTCATTGAGA
		753 T 802 G	11111111111111111111111111111111111111111111111111

TABLE 2: (Continued)

108 BMEI0195	17	BMEI0195_5	GGCCGTCCGTCAGGCGTCCGTCGTCCAGCACCTGCAACAGAACATTGAAG
		2229 C 2278 C	00100110011111111110010010000100011111111111111111
	17	BMEI0195_4	CGGGAGGAGCACCGATAAGGCGGCTCACGGAATGCTTTTCCATGAATTCC
		2076 G 2125 G	00011111111111111111111111000000011111100000000111111
	21	BMEI0195_3	GACGAGGGAACCGTCGAGAATATCACCAAGCAGGATACGTTCGGCCAGCG
		2654 C 2703 C	00000111110000011111111111111100000001001111111100
	25	BMEI0195_2	ACCAATGTGTGCATTTCGTCGATGAAGAGGATGATCTGCCCGGCAGCCGT
		976 A 1025 T	01110011100000010000011111111111111111000001100000
	28	BMEI0195_1	CGCAATGCAATTCACCGCGCGCCAGAGCAGGCTTCAAAAGGTTCGACGCA
		1056 T 1105 G	00100000000000100001001000111111111111111111000000
	0	BMEI0195_0	CGAAGGTCTGGGCAGACTGGATGAAGCCGCGGACCCGTTCTGTATATTTT
		183 A 232 G	11111111111111111111111111111111111111111111111111
107 BMEII1043	14	BMEII1043_5	CGGCAGATCGGCATAAGCGACGTTTTCACCTTCATCATGGGCGAGTGTGC
		2117 G 2166 G	00001000111111111111111111111111111111100000011111100
	16	BMEII1043_4	CGCTTCCAGTTGGCGGGCATTATCCATCCATGCTTCAAAGTCTTCACGGC
		1067 G 1116 G	01111111111100000000011111111111111111111111000000
	21	BMEII1043_3	TTGTTGGCGTCACCCTTCTTGCCGTTATCGTCAATGACACGGGCTGGCCC
		1207 G 1256 A	11111111111111111100010011110010001111100000000000
	28	BMEII1043_2	CCATTCGGCAGGCGTCGGGCGGAACTGTTCAGCATGAATGGAGACCGACT
		2462 A 2511 G	00111100011000100000000011111100000011111100000111
	28	BMEII1043_1	TGGCGCAACATTCTTCCACCAGCTTATCGGCGAAGACCAGCCTTTCACCG
		837 C 886 A	00000100111110010000011111100000001111111110000000
	0	BMEII1043_0	TTGTGGCGTATTGCGGAAAATAACCGGCTTCTTGGAGCGCCAGGAATGCG
		1325 C 1374 A	11111111111111111111111111111111111111111111111111
107 BMEII0751	17	BMEII0751_5	TTTCAGCATGAAAGCCTGTTCGCGCTCGCAAAGCTCCTGATGGCCCGGCT
		773 A 822 A	00000111000000011111111111111111111111100000111111
	19	BMEII0751_4	CGCGGTTTTCGTATGCGCATCCACATCATCAGACTTAGTGACGGCCTCGT
		644 A 693 G	11111111111111000000000111111111111111110000000000
	21	BMEII0751_3	CTCATACCAGCCAACCGTGCCGTCATCGAACAGAACCTGGAGATGGCCGT
		509 A 558 G	00011111111111100111001111111111000100001001000001
	23	BMEII0751_2	GCCAGATCGGGCTTCAGCTCTTTCAGCGCCTCATGGAAGGACGGATGAAT
		28 A 77 C	00000000000001000010011100100111111111111111111111
	27	BMEII0751_1	CGATAACGAGTTTACGCCCGTTGGCCCTTGCACAGTCGATTACCCGGCGT
		186 A 235 G	11110000001111111000000111111100000011000111110001000
	0	BMEII0751_0	GGTTCTTGTGCGTTTCCCAGGTTGCACCGCTCGACTGCTGGTTCAGGTTC
		318 G 367 C	11111111111111111111111111111111111111111111111111
107 BMEII0487	13	BMEII0487_5	TCGAGAACTTGACGTCCTCCCGAAGCTTGAATGTGTAGGATCGGCCATCC
		246 G 295 A	10000111111111111111111111111111111111100000111110000
	14	BMEII0487_4	CCTTGCCGTCTCATCAGTCGAGACAAGAACCTGACCGATCTCCGCATCGA
		1364 T 1413 G	11111100000001111100000001111111111111111111111111
	19	BMEII0487_3	GTTCTGGGCAAACATCTGGTTCGGAGAATAAAGGTGCGGATTGAGCGGCC
		116 G 165 C	00001111100000110000011111111111111111111111100000
	25	BMEII0487_2	ACCGACCTTACGCAGATCCGCCTGCACGATCTCGGCCATCGATTTGCTGA
		1127 T 1176 T	11111111110000001111000000111111000010000010000111
	36	BMEII0487_1	ACAACTCGATCGCCAGAGGCTGCCCATCCTTCTCCCGGATATCGCCGCTT
		1056 A 1105 T	00001000001000011000000000110000011100000110000111
	0	BMEII0487_0	TGGGACCCCAGTAGCTATCATTGCGGGTAAATACGTCGTGCTCGCCAAGT
		585 A 634 A	11111111111111111111111111111111111111111111111111
106 BMEII0979	18	BMEII0979_5	GGCAGCAAACGTTTCCGGCTTGTGTTCTTTCAGCCAGAGCAGTTTCGGAG
		437 C 486 C	00000011111111100000011111111111000000111111111111
	19	BMEII0979_4	GCCATATATCGTCGCTCGACTGCTCGACAATACCGCCCGCCTCGCGCCAT
		120 A 169 C	00011110100000101111111111111111111000001111000011
	20	BMEII0979_3	GCTGCGGGATCGACGCCAGCAACCCTGACCACTTCGCGAACACTTTCACA
		178 T 227 C	01100000010000010000001001111111111111111111111111
	24	BMEII0979_2	CGTCACAAGATCGGGATAAGCACCGGACGCAACCGCACCTAACATGGCCG
		1439 C 1488 G	11111111111111111111100000000010000010011000010000
	25	BMEII0979_1	ATGATCCATCCAGACGATGATGTTGCGCGCCCGATCGTTGGAAGGACCAA
		299 T 348 T	11111000001111100000111110000011111000000111110000
	0	BMEII0979_0	CTGTATCAAGGTCCAGCCCCGCAATAACCGCACGCGCATCAGGATCGGCG
		1170 C 1219 G	11111111111111111111111111111111111111111111111111

TABLE 2: (Continued)

106 BMEII0852	17	BMEII0852_5	CCGGTATGGCAATGCCGTCCTGCTCGATCAACTGCTTGTAGCGCGTCAGA
		1191 T 1240 G	00001000000011111111111111111111111111111111000000111
	18	BMEII0852_4	CTTCCGCAACCCGTTCCGCCGTTTCCAGATCATTGCGCAAGGTGACGAGC
		627 G 676 G	00010000001111111111111111111110000001000111111111
	19	BMEII0852_3	GCGCGCGATAGGAAATGAAGATCGTGAAGGAGCGCCCGTCATTACTCACC
		462 G 511 C	00000000011111111111110000111100000011111111111111
	26	BMEII0852_2	GCAAACAAGCCGCCAAGCCCAAGCCCGACCAGAAGCCAGTTAATCAGCCT
		1291 A 1340 C	00111111000000000000100100001111111111111110000100
	26	BMEII0852_1	GGCGCTCACGTACTGCCGGATCGAAATCGCGCGCATTCTTCGACAGGGTG
		381 C 430 C	00000000001111111111111110000000001111110000000111
	0	BMEII0852_0	CGTCTGGGCTGGCCTGATCTTGTCCAAAGCCTCCCGAAGCATCGTGCCTG
		1346 C 1395 G	11111111111111111111111111111111111111111111111111
106 BMEII0451	16	BMEII0451_5	GTGTTGAGCGTGGCCACCATGTGCTGAACGAAGGCTAGATCGCCCATGTT
		991 A 1040 C	00001111111111111000010011111111111111111001000011
	17	BMEII0451_4	AACAGATCGGCTTTCAGCATCGACTCGCGAATACGGCCGTCGCCTGAGCC
		1087 G 1136 T	11111000000111111111111111000011111000111100011110
	18	BMEII0451_3	TGGCAAGTTCGGCCTGCACATCGATCGGCTTTTCGGGCTCGAGCGTATCG
		1425 C 1474 A	11111111111111111000000001100011111111111000110000
	27	BMEII0451_2	CGCACCCCAGCTATCCAGCGAGAAGGGCGGGTTGGCGATTACTCGATCAT
		875 A 924 G	00010000100011110001100001000011111000111100011111
	28	BMEII0451_1	TTTTGATCGACGACCAGCGCGCGCGTTCAACGAGAAAGAACTCATGCTCG
		234 C 283 A	11111000000000000001000011110000011111111111110000
	0	BMEII0451_0	GCAGCCCCTGACGTATGGACTCCTTGGCCTCTTCCTCGAAACGATCAGAA
		156 T 205 C	11111111111111111111111111111111111111111111111111
106 BMEI1602	7	BMEI1602_5	AATCGCCTGCCATGCCTTGCGGTGCGACAGGAAACAGGCAATCTCGTTGC
		197 G 246 T	11100000001111111111111111111111111111111111111111
	18	BMEI1602_4	GCGGAGTGTGCAGGCGGCGTTTATAAATGCGCGCAATCTCGTCTTTGTTG
		126 C 175 C	11111111100000001111111111100000100011111111111000
	25	BMEI1602_3	CGTCTTCCACGATGAAGCCCGCATCAAGCTTCCGGTCGATAATCGCCTGC
		237 G 286 G	00011111000001100000111110000011111111111000000000
	26	BMEI1602_2	CCGCGGCCCGGTATGCACCCATGAAGTTTTCCGTCAGTGCAATATCGTCT
		282 A 331 G	00000111111111110100000000000000011111111110000011
	30	BMEI1602_1	GGATGATGATCCGCACCGCTTCGTCCCGGAAAACCTCACGGCCATGCTCG
		384 C 433 C	01111111110001000111000100000000000000000101111100
	0	BMEI1602_0	CTTGCGGTGCGACAGGAAACAGGCAATCTCGTTGCGGCTGAGGGCAAAGG
		182 C 231 G	11111111111111111111111111111111111111111111111111
106 BMEI1569	12	BMEI1569_5	ACATGCCCATTCCTGCTTGCCGAGGCGATAGATGGTGCCGTCCAGTTCGG
		1535 C 1584 T	11110000011111111111111111111111111100001111111000
	15	BMEI1569_4	CTTGTCCTGCGCCATGTCGAAAATATGTGCTTCATAGCCGAGCCTGCGCA
		272 T 321 G	11111111000001111111111111111111111100001000110001
	19	BMEI1569_3	GCAGAACGCCGCTGTGATAATCGTTGACGCCAAGATAGTCGGCAAGCTTC
		1767 G 1816 C	11111100000111111111111110000100001100100001111111
	29	BMEI1569_2	GACGCGCACATAGGCAACACCGGAAATCTTTCCCAGTGCATCCTCAATGG
		152 C 201 C	11000001111000110000000001111100000000011000111111
	31	BMEI1569_1	TTCTGTTTGCCTGCCCGAAGCCGCAACGCGTTTGTCACCACCACGATGGA
		2152 T 2201 A	11111111001100100100110000000011001000000000001001
	0	BMEI1569_0	CGCATCATGTAGTCGAGTGTGCGGCCGATCAACAGGAAGAAGAGAAGCGT
		640 A 689 G	11111111111111111111111111111111111111111111111111
106 BMEI1230	11	BMEI1230_5	CATACCCGTCGACATGGTTCGCCAACTCTCGCTCGACATGACGGACGACG
		51 C 100 G	11111111111111111111111111111111111111111100000000000
	21	BMEI1230_4	CGCGCACTAATGCTCGCACTACATCCTCGTCCGTCTGGCTGTCAAACCTG
		108 C 157 G	00100011110001001000111111100000000111111111111111
	22	BMEI1230_3	CGCTGTTCCGACGATCACGACGAGCTTTAACCAGCGGAAGGAGAAGACCA
		510 T 559 G	00000100001001111001000111111010000011111111111111
	24	BMEI1230_2	AGTGAATGGGAAGGTCTTGCTCACGCGCAGACCAGGCACCTCAACTGTGC
		248 G 297 T	11111111111100000100111100110000000000011111110000
	28	BMEI1230_1	CGATCTGTTCCAGCGTCGAGTTAATGTGGTTAGCCGCGGCTTCACCTTCG
		405 C 454 G	00010001110001100000111111111111111000000000001000
	0	BMEI1230_0	CCTGTGCTGATCATACCCGTCGACATGGTTCGCCAACTCTCGCTCGACAT
		62 A 111 G	11111111111111111111111111111111111111111111111111

TABLE 2: (Continued)

106 BMEI0471	14	BMEI0471_5	GGTGAAGCAGCGAGGTGTCCTGCGTGACCATGCCTATATTGGCTCTCAAT
		1320 A 1369 C	11111110000000011111111111111111000111111000111111
	19	BMEI0471_4	GGTCGAGCGTGCATCGGCCTGTTCTTCCGAAACGACCCGCAAACGTGGAA
		632 T 681 C	00000111111111111111111111111111110000000000110000011
	21	BMEI0471_3	AGCATTTCCTCGCTGGCATCCGGGCGTCCATACATGATGTTCTCACGCAC
		1375 G 1424 T	11111110000000011100001000011111111111111111111000010
	22	BMEI0471_2	TTCGACAATGCGCCCCTTGTCCATCACCACGAGACAATCAAGCGCCGCAA
		1715 T 1764 A	11111111100000010000011111111111000001100000011111
	30	BMEI0471_1	CCCCTGGCCACCTTCAGCGCCGGAGCATTAGGAACGTCAACGACCTCATG
		1051 C 1100 G	00111000100001111110000100001000110011110010001000
	0	BMEI0471_0	GCGGTCTGCATCAGCTTGGTCGATATCCGTCCAGCAAACTCGTCCTGAAA
		421 T 470 C	11111111111111111111111111111111111111111111111111
106 BMEI0455	17	BMEI0455_5	ATAGTGCCCCTTGATATGTTCCATATTGACCGTTTCGGCCACGCCCGGCG
		836 C 885 T	11111111111110000011100000111111111111110000111000
	17	BMEI0455_4	GCAAGGCCGCCGGGTAGAAATCGCGCGTGGCATCGCCGAATTCATTGAAA
		474 T 523 C	00111111111111111111110000111111100000001000011111
	22	BMEI0455_3	CCCAAGGGCAGGCATAGGAAACATAGAGATGATAGCGGCCAGGCTCTGCC
		147 G 196 G	00100010010011111111111111111111110000000000111000
	24	BMEI0455_2	GTGCTGCCTGTGGTGCCGTAGAAAGTCCAGCCTTCCTCCGCCATCAGGAA
		265 T 314 C	00001111111110010010000001111110000001001001111111
	26	BMEI0455_1	GAAACGATGGTTTCGCGCTGCTTGTCCCAGAGCACCGGCACCGTTACACG
		385 C 434 C	11100001111111100100001111111100001000000001100100
	0	BMEI0455_0	GGTCGCGGGTATAGTTCCACAGGTTTGGATAATCTGCGAGCCTGCGAAGA
		777 T 826 C	11111111111111111111111111111111111111111111111111
106 BMEI0007	17	BMEI0007_5	CGCGCGGATCGATAAAGTCATATTCAATGGCATAGCCCGGCTGGAGCAGC
		1074 G 1123 G	10000111111111111111111111111111111111000000000000011
	19	BMEI0007_4	GCATCGATTTCACGAACAAGATGGCCTTTTCCCAGCCCGCCAATCGCCGG
		217 C 266 C	11111111111111111110000000011100000011111111100000
	22	BMEI0007_3	GCGATAGGGTTCGCTCACGCCGCGAGACGTCAGATCATCCACCATCACAC
		1301 G 1350 C	00001111110000000000000000100111111111111111111111
	22	BMEI0007_2	TTCGTTCGCCATCGCAGACAAGATCAGCAGCACCACCCTCAACAACGGTG
		426 C 475 A	11001000011111111111111100010010000001011111010001
	26	BMEI0007_1	AAGCGAAACTGGATGCCTGCTCTATCAGCCACGCGCCCCATCAGGCCATC
		271 G 320 T	00001111111111000000001000000011111111111111000000000
	0	BMEI0007_0	CAAGAGCTTCGGCCGAACTCATGGGTTCATCCTGACGCAATTCGAATCCG
		39 C 88 G	11111111111111111111111111111111111111111111111111
105 BMEI1818	12	BMEI1818_5	CCTCATCAAGCTGAGGCAGAATTCGTGTCACGGGAAGATCGGGCAGGCTC
		282 G 331 G	00000111111000000011111111111111111111111111111111
	19	BMEI1818_4	CAGGAAGGCCAGAATACTGCCCGTTTCATCCACCAGTGCATCGCGGATTG
		896 C 945 G	00001100100011110001001111111111111111111111000001111
	19	BMEI1818_3	GGGAGAGTCCATGCCACATTCGCCAGAAGACAAGAATGCGCGCCCTGACG
		111 C 160 C	00000000111111111111111111111111110000010010000111
	24	BMEI1818_2	CGCCATTCAGGTCCGGATCATCCAGGATCATGCGCGCAAAGACGCCCTCG
		588 C 637 G	00001111111111111111111110000100000000001111100100000
	31	BMEI1818_1	TCCAGCGTTGCCGACATGACGAGGATTTTGAGATCGTCGCGCAGTGCCGC
		718 G 767 A	00000000000001000000001111111111111111100000001000
	0	BMEI1818_0	GTGTCACGGGAAGATCGGGCAGGCTCGTCAAAGGCTTTCCATCTTTTGCA
		258 T 307 C	11111111111111111111111111111111111111111111111111
105 BMEI1575	18	BMEI1575_5	GCCAGCGAGACCTGCCCTTCTAGATAGAAATGGTCCTGACCGCCCAGATA
		547 T 596 C	11100000000111111111111111111111111111111000001000010
	19	BMEI1575_4	TCGCCGGTCAGCGTATCAACACTAACCTCGGAACAGGCCGCGCCATAAGC
		1867 G 1916 A	10000111100001111111111111111111111000000000001111
	20	BMEI1575_3	CATCGTTTTCGCCCGGCACGTCCTTGCAGGTGAGGATATCGACCACGCCG
		222 C 271 G	11111111100000000010001111111111111111111110001100000
	23	BMEI1575_2	CAGCCCATGCCCTGCACGAAGCCGCCTTCTATCTGGCCGATATCGATGGC
		1972 G 2021 G	00001000001110000011000010011111111111111111111000
	25	BMEI1575_1	TCATGTGCAGGCCCTGCCCCATTTCCGTGCCACCATGGTTCATATGCACC
		1434 G 1483 A	11110000010001000010010001000011111111111111100001
	0	BMEI1575_0	GCGCATAGAGATAATCCACGCCCAGAATAGTGCCGTCATCGTCGAAGCCC
		876 G 925 C	11111111111111111111111111111111111111111111111111

TABLE 2: (Continued)

105 BMEI1341	16	BMEI1341_5	GATCATCGGATCGGGGCTCTGGGTCGCCGTACCGCAATAGACACGCATAT
		326 A 375 C	00001111000001000111111111100111001111111111111111
	16	BMEI1341_4	GTGCCATAAATGAAGGGAAGTGCGCCACCTTCTTCCGCTGTCGCCACGCG
		76 C 125 C	11111111111111111111110001001111111111001000000000
	19	BMEI1341_3	GGTCGCCCGTTCCATCCGACGCCTGATAACCGAAATATTCCGACCAGTCC
		1335 G 1384 C	00100000011000010000000111111111111111111111111111
	22	BMEI1341_2	CTGCCCGGCCCGCTCATAGCCATGCACTCCAAACACATTCAGAAGCGGCT
		2222 A 2271 G	10001000001111111000001000101111111111111111000001
	32	BMEI1341_1	CAGCCACCTCCACACCATCGATGCGGTTATAATGCCTGTTGCTCCACCAG
		1671 C 1720 G	00000001100000011111000000111110001000100111001000
	0	BMEI1341_0	CTCCACCTTCCATATCGTGCTCGGCTTGCGGACAGAGAGGCTCGTAACAC
		764 G 813 G	11111111111111111111111111111111111111111111111111
105 BMEI0835	6	BMEI0835_5	CACCTCAAACGGCAGGACAGTCAGCACGTGGTCGATAGAGGCGCGCATGG
		416 C 465 G	11111111111111111111111111111111111111111110000001111
	20	BMEI0835_4	AACAACCTCCAGCTTCTCGATACGCGCCGCCAGTTGCTCGACAGCCTTGC
		659 G 708 T	11000000111111111111111111000001111111000000011110000
	23	BMEI0835_3	GCGCCCTTAACGCATCAATCAGATCAGCGCCCAGCAAGTCGCCGGATTCT
		48 A 97 C	00001111100000111111111100000100111100100000111111
	25	BMEI0835_2	TTAAACCGCTGCGCTTCCAGAACCAGTTGCGACGCACGTGCCTGCAGGAT
		535 A 584 A	11111110001000010000000001111000000000111111111111
	31	BMEI0835_1	CTGTCGATCAACAGCACACAATCGGGCTTTTCGACCACGATCTGCCGCGC
		253 G 302 G	00000111100000011111110000001100011110001100000000
	0	BMEI0835_0	CCATGACTTCCACCTCAAACGGCAGGACAGTCAGCACGTGGTCGATAGAG
		426 C 475 G	11111111111111111111111111111111111111111111111111

TABLE 3: Negative Controls for Target Secondary Structure Array based on *Brucella melitensis* 16M ORFeome

PROBE NAME AND SEQUENCE	NUMBER OF TARGETS WHITHIN A CERTAIN SIMILARITY RANGE	THE LONGEST CONSECUTIVE STRETCH OCCURING IN A CERTAIN NUMBER OF TARGETS
CONTROL_AT_3057M GAAGAGGTTCTTGCGAGGGAGATCCGAGAGTTCAGGAGCCCAACAGACTT	0-30% 54 31-35% 21 36-40% 12 41-45% 9 46-50% 0 51-55% 0 56-60% 0 61-65% 0 66-70% 0 71-75% 0 75-100% 0	1: 0 2: 0 3: 5 4: 19 5: 23 6: 25 7: 9 8: 9 9: 2 10: 2 11: 2 12: 0 13: 0 14: 0 15: 0 16+: 0
CONTROL_AT_6369M GCGACCGTCAGGGAATACTTCTTTGACTCCGTCACAGTAACAGTCACCTT	0-30% 67 31-35% 11 36-40% 14 41-45% 4 46-50% 0 51-55% 0 56-60% 0 61-65% 0 66-70% 0 71-75% 0 75-100% 0	1: 0 2: 0 3: 9 4: 13 5: 25 6: 17 7: 15 8: 5 9: 7 10: 5 11: 0 12: 0 13: 0 14: 0 15: 0 16+: 0
CONTROL_AT_7455M TTTGAAGTGAAAGGCGAGGAGTACTCATGTCCGTCGTTGCCAGAACTCCC	0-30% 72 31-35% 13 36-40% 10 41-45% 1 46-50% 0 51-55% 0 56-60% 0 61-65% 0 66-70% 0 71-75% 0 75-100% 0	1: 0 2: 0 3: 2 4: 15 5: 26 6: 23 7: 17 8: 3 9: 9 10: 1 11: 0 12: 0 13: 0 14: 0 15: 0 16+: 0

TABLE 3: (Continued)

CONTROL_AT_15884M	0-30%	68	1:	0
TAGACTCGATGGAAGCTCTGGTGAAGGTACAAGTGGGACTGAGCAGATTT	31-35%	12	2:	0
	36-40%	13	3:	6
	41-45%	3	4:	11
	46-50%	0	5:	22
	51-55%	0	6:	19
	56-60%	0	7:	15
	61-65%	0	8:	10
	66-70%	0	9:	8
	71-75%	0	10:	5
	75-100%	0	11:	0
			12:	0
			13:	0
			14:	0
			15:	0
			16+:	0
CONTROL_AT_17104M	0-30%	73	1:	0
GATACTTCTCCCAGAGAGTCCCTGAGCCACCAAGTGTGTCAACGCATGTT	31-35%	7	2:	0
	36-40%	10	3:	4
	41-45%	6	4:	22
	46-50%	0	5:	29
	51-55%	0	6:	21
	56-60%	0	7:	6
	61-65%	0	8:	11
	66-70%	0	9:	3
	71-75%	0	10:	0
	75-100%	0	11:	0
			12:	0
			13:	0
			14:	0
			15:	0
			16+:	0
CONTROL_AT_22534M	0-30%	70	1:	0
ATGGGAGGGTGTGAGAAGTTTGGGGATGTTGAGATGGCTGAGTGGGTGTT	31-35%	19	2:	0
	36-40%	3	3:	6
	41-45%	4	4:	19
	46-50%	0	5:	25
	51-55%	0	6:	14
	56-60%	0	7:	8
	61-65%	0	8:	14
	66-70%	0	9:	9
	71-75%	0	10:	0
	75-100%	0	11:	1
			12:	0
			13:	0
			14:	0
			15:	0
			16+:	0

TABLE 3: (Continued)

>CONTROL_AGILENT_1_50RC	0-30%	74	1:	0
GGTCTATCCCGGCCACATACGGAACGCTATGTGATACGTATAGTAGGATA	31-35%	14	2:	0
	36-40%	8	3:	2
	41-45%	0	4:	20
	46-50%	0	5:	27
	51-55%	0	6:	13
	56-60%	0	7:	12
	61-65%	0	8:	11
	66-70%	0	9:	9
	71-75%	0	10:	2
	75-100%	0	11:	0
			12:	0
			13:	0
			14:	0
			15:	0
			16+:	0

TABLE 4: Target Secondary Structure Miniarray based on *Brucella melitensis* 16M ORFeome

INACCESSIBILITY SCORE FOR TARGET	TARGET NAME	PROBE INACCESSIBILITY SCORE	PROBE NAME & BINDING SITE LOCATION	PROBE SEQUENCE & PROBE-BINDING SITE ACCESSIBILITY
133	BMEII0462m	24	BMEII0462m_5	CTGATGCGCCAGCCTGCCCTCAAACGGATAGGCCACCATGTAGAAGCGGT
			1928 A 1977 G	00001000000111001000111111111000100111111111100001
		24	BMEII0462m_4	GCTCATAGGTCTTGAGCGCATAGCCGCCGGTCGCCACAAAATCGAGAATG
			1386 C 1435 C	00000111111100000011100100100010010000111111111111
		24	BMEII0462m_3	TCGAGAATGGTAATTTCCGGTTTTGCCCCGCCATCGACGGTGACAAGACC
			718 G 767 A	11111111111111110000010000100001111100001100001000
		27	BMEII0462m_2	CGACCAACCCTTGGCCGCAAACCATTTGACGAAACGGTCCGGCAGAAGAA
			41 T 90 G	00000011111000000011000000011110000000111111111111
		34	BMEII0462m_1	CCGGTGCGGGTCTCGATGGAAATATCAAGCCCCATTTCCTCCACCGGCAC
			355 G 404 G	00000000001000000111111000000100001111111000000001
		0	BMEII0462m_0	GGCCCCAGAGACCCAGCGAATAATCAGTGGTGACAAAGCCCAGCGGATGC
			2022 G 2071 C	11111111111111111111111111111111111111111111111111
128	BMEII0685m	19	BMEII0685m_5	CGGGCGCTCAACCGGCTTGTTGTTTTCGTCATAGATGCTGACGCCATTGC
			647 G 696 G	00011111111000001111111111111111111111000000000111100
		20	BMEII0685m_4	GGTTGCCCTTCGATACCAGCATATTGGTGTTCTGCGACTTCTGGACCGGG
			693 C 742 C	00111100100011000011111100001111111111111111000001
		24	BMEII0685m_3	TCGCATCGGAGCCGAGGAACATTTCGCCTTCACCATAGCCAACCGCAAGC
			531 G 580 A	00000111100100011111111000100111111110000000001111
		30	BMEII0685m_2	CGCATAGACGAGCGGCGAGATGAATTCCGGCACCTCGGGCAGAACGATGG
			1679 C 1728 G	00011111111000000111111111100000100000000000000000
		35	BMEII0685m_1	CGACAGTAACAGACTTGGCCAGATTGCGCGGCTGGTCCACGTCCGTTCCC
			1770 G 1819 G	00000000000000100000001111111110000100010000000111
		0	BMEII0685m_0	AACACTGCATCCGATTCACGCGCGATGGTGGAGCCGGTCACGTTGAGCAC
			1108 G 1157 T	11111111111111111111111111111111111111111111111111
128	BMEI0267m	16	BMEI0267m_5	ATGCCGATGCCGCAGATCGCAAAGCCGATGATTGCAGTCGCCGAACTGTT
			907 A 956 T	00000000111111100001111110000111111111111111111111
		27	BMEI0267m_4	CGACAAAGCCGATCAGTGATGGCGCGACCAACATGCCGGAATAGCCAAGC
			1041 G 1090 G	00111111111111111000000111110000000000000000000111
		27	BMEI0267m_3	GCATCATGGCCAGCGAACCGAGACCGCAGACCAGGATCATCAGGCCCATT
			168 A 217 C	00000100000100011111111100010000010001111110001111
		28	BMEI0267m_2	GTCACCACATTCGGCACCAATGTGATGATGAGCAGCGCCGGAATGAAGGC
			280 G 329 C	00000000000111111100000000001111000001001111111111
		30	BMEI0267m_1	CATGCGACAATGATCAGGATGGTGGCTGCGATCATCGTGACCAGCGCATG
			523 C 572 G	00000000000000001111111111111100000000011111100000
		0	BMEI0267m_0	GTCAGCGTATCGGGCGAGGCTTGAGAAGGCGAGAACCACCAGCAGCAAGG
			1133 C 1182 C	11111111111111111111111111111111111111111111111111
117	BMEI0682m	17	BMEI0682m_5	AACGCGCAATGCCGATATAGCCGGAAAGGGCCGTGAGAATAGGCAAGAGC
			1248 G 1297 T	11110000111000011111000011111111111111111000111100
		22	BMEI0682m_4	ACGGCAAGCGACGTGTCTTCCGTTTCGCCAAGGGTAGCGTTGATTTCTGC
			412 G 461 T	10000000000000001111111111000011111100001111111111
		26	BMEI0682m_3	GGCAGCCAGAAACGATTCAAGCTTGAAGGTGACAACGAAACGCCACCACG
			611 C 660 C	11100001000111100010000111100000011111111000000111
		27	BMEI0682m_2	GGAAGCCGGTCATCAGCTTGTAGAATGTGCTTTTCAGCTCCGCCCACTGA
			1539 T 1588 C	11000000111110000001111111100100000000111111100000
		25	BMEI0682m_1_M	**GCGCGCCCTG**TCAGGAAGCCCgTATACATGGTGACAAGGAAA**GCGCCCG**T
			1321 A 1370 C	01000000000001111111111111111110000011110000001001
		0	BMEI0682m_0	TGATTTCTGCCTTTTCCGCTATCAGCCGTGCACGCTCTTCCCCAACGATG
			372 C 421 A	11111111111111111111111111111111111111111111111111

TABLE 4: (Continued)

```
105 BMEII0874m 23      BMEII0874m_5   GTTATCAGCCACGTCACGAACATGCCCGCCATCAGGAATTCGCCATGAGC
                       112 G 161 C    11111000100000111110000010001111111111000001111100
               24      BMEII0874m_4_M AACTTCCAACCGCcTCAAACCACCGATAGTCAGCCCTCGCGCTTCCGCCC
                       2384 G 2433 T  11111111100001000111001000011111111110000011000000
               25      BMEII0874m_3   CGACAATCGCGAATAGCCCGCCGAGCATCAGGCCGTTCAGAAGAATTTGC
                       15 G 64 G      00111110001111100000100000000000000011111111111111
               26      BMEII0874m_2   GACATGGTGTCGATCACGCCGATGATCAGCGCACCAACGAAGGCTCCAGG
                       715 C 764 C    00001110000011100000011111000000111111111000001111
               7       BMEII0874m_1_M tTGGATCATGAGCCCCGTAGCACCGCCCGTAAGCCATCCAAAATGCAGCG
                       1313 C 1362 A  11111111110010000011111111111111111111111111111111
               0       BMEII0874m_0   GCGGCTCCCAGGCCAAAACAGAGAATGTAGATGCGCGAAACATTGACCCC
                       547 G 596 C    11111111111111111111111111111111111111111111111111
```

TABLE 5: Target Sequences For *Brucella melitensis* 16M Miniarray

TARGET NAME	TARGET SEQUENCE
BMEII0462m	aTGACACTGACCCGCAAGCCCATCGCAGCATCTGCCACCTTTCTTCTGCCGGACCGTTTCGTCAAATGGTTTG CGGCCAAGGGTTGGTCGCCGCGTGCGCACCAGCTGGAGCTTCTGGCGCGCGCCGAACAGGGACAGTCCACGCT TCTGATCGCACCCACCGGCGCGGGCAAGACGCTGGCCGGGTTTCTGCCCGCGCTGGTTGATCTGGAAAAGCGG CGAAGCAAAGCAATTTCAGGAAAAGTGGGAACCGGTTTTCCGTCCGGAATTGCGCCAAAATCAAAAAGTGGCG TTCACACGCTCTATATCTCGCCGCTGAAGGCGCTTGCTGTCGATATTCGCCGCAATCTCACTGTGCCGGTGGA GGAAATGGGGCTTGATATTTCCATCGAGACCCGCACCGGCGACACACCTGCCCACAAGCGTCAGCGCCAGAAA CTGGCGCCGCCGGATATTCTGCTGACCACGCCGGAGCAGCTTGCACTTTTGATCGCGTCCAGGGTGGCGGAGC AATTTTTCAAAGATTTGCGCTATGTGGTGCTGGATGAACTGCACTCGCTGGTGATTTCCAAGCGCGGGCATTT GCTGGCGCTGGGGCTGGCACGGCTGCGCCGCCTGCAACCACAGTTGCAGACCATCGGTCTTTCCGCCACCGTG GCGGAGCCGGATGAGTTGCGCCGCTGGCTGGTGGAGCAGGATGGGACCGGGTCGATGGCCGGTCTTGTCACCG TCGATGGCGGGGCAAAACCGGAAATTACCATTCTCGATTCTAAGGAGCGGGTGCCATGGGCGGGGCACTCCTC ACGCTATGCCATTCCCGATATTTACGCGGCGATCCGCCAGCACCGAACGACGCTTTTGTTCGTTAATACGCGC AGTCAGGCGGAGATGCTTTTTCAGGAGCTCTGGCGGGTCAATGAGGAGACGTTGCCGATTGCGCTGCATCACG GCTCGCTCGATGCCGGCCAGCGCCGCAAAGTGGAACAGGCCATGGCCGCCAATACATTGCGGGCGGTGGTCGC GACCTCGACGCTTGATCTCGGCATCGATTGGGGCGATGTCGATCTGGTCATCCATGTCGGCGCGCCGAAGGGC GCAAGCCGCCTTGCGCAGCGCATCGGGCGCGCCAATCACCGCATGGACGAGCCAAGCCGCGCCATTCTGGTGC CCGCCAATCGCTTCGAGGTGATGGAGTGTCGTGCCGCGCTTGATGCCAATTATCTGGGCGCGCAGGATACGCC GCCCTTGATCGACGGGGCGCTGGATGTGCTGGCGCAGCATGTGCTGGGCATGGCCTGTGCCGAGCCGTTCAAC GCCGATCAGCTTTATCGTGAGGTGCAAAGTGCGGCCCCTTATGCAAACCTTCCGCGAGACACCTTCGACCGCA TTCTCGATTTTGTGGCGACCGGCGGCTATGCGCTCAAGACCTATGAGCGGTTTGCGAAAATCCGCAAGACGGT GGACGGCACGTGGCGGGTGTCCAATCCGCGCATCGCGCAGCAATATCGCCTCAATATCGGCACCATTGTCGAA GCGCCGGAACTGAATGTGCGGCTCACGCGCGGCGGCAAGGGCGCGAATGCGCGGGGCGGGCGCGTGCTGGGCC GCATCGAGGAATATTTTCTGGAAACACTAACGGCTGGCGATACCTTCCTTTTCGCCGGAAAGGTGCTGCGCTT TGAGGGTATTCGGGAAAATGAATGCATCGCCTCCAATGCTGCCGGGCAGGACGCAAAAATTCCCGTCTATGCG GGCGGCAAATTTCCGCTTTCCACCTATCTCGCCGCGCAGGTGCGCGCCATGCTTGCCGACCGCGCGCGCTGGC AATTCCTGCCGCAGCAGGTGCGCGAATGGCTGGAGGTGCAGCAATGGAAATCCGTGCTGCCCGGCACGGACGA ACTTTTGATCGAAACCTTCCCGCGCGGCAACCGCTTCTACATGGTGGCCTATCCGTTTGAGGGCAGGCTGGCG CATCAGACGCTCGGCATGTTGCTGACGCGCCGTCTGGAGCGCATGGGCGCGCATCCGCTGGGCTTTGTCACCA CTGATTATTCGCTGGGTCTCTGGGGCCTCAAGGATATGGCGTCCATGATCCGCATGGGCAGGCTTAACCTTTC GCGCCTGTTCGATGAAGACATGTTGGGCGACGATCTGGAAGCGTGGCTGGATGAAAGCTATCTTCTCAAGCGC ACGTTCCGTAATTGCGCCGTGATTTCCGGGCTTATCGAGCGCCGTCATCCGGGGCAGGAAAAGACCGGGCGTC AGGTGACGGTTTCGACGGATCTTATCTACGACGTTCTGCGCAGCCACGAGCCGGACCACATTCTCTTGCAGGC CACACGCGCGGATGCGGCAACGGGGCTTTTGGACATCAAGCGGCTTGGCGACATGCTGGCGCGTGTGAAAGGG CATATCCTGCACAAGCCGCTTGACCAGATTTCGCCGCTGGCGCTGCCGGTGATGCTCGAAATCGGGCGCGAGC GCGTGGCGGGCGAGGGCGATGAAATGCTGCTTGAAGAAGCAGCCGACGATCTTGTCAGGGAAGCAATGCAATG c

TABLE 5: (Continued)

BMEII0874m

ATGACTACCTTTCTGCAAATTCTTCTGAACGGCCTGATGCTCGGCGGGCTATTCGCGATTGTCGCCGTCGGCC
TGACGCTCATCTTCGGCATCGTCAAGGTCGTGAATTTCGCTCATGGCGAATTCCTGATGGCGGGCATGTTCGT
GACGTGGCTGATAACCACAAAGCTGGGGCTTCATCCTTATGCGGCGGTTATCATTGTTCTGCCTGCGATGTTC
ATTCTGGGTGCGCTGACCCAGCGCCTTCTCATTCAGCCGCTAATGGCTTCTGATGACGGTCATGCACAGATTT
TCGCAACCGTTGGCCTGTCCACGGCGATGATTAATCTGGCCCTGCTGATCTTCGGTGCGGATATAGCCAATAC
ACCGAATTTCGGACTGCGGACCCCAATCGAAATCGGGCCGCTTCGCGTACTTACCGGGCAGGTCTTTATTTTT
CTGGGCGCCATTGTGCTTGTCGTTGCGCTTCAACTGTTCCTGAAGAACAGCCAGACTGGCCACGCAATCCGCG
CCGTTGCTCAGCATCGCAGTGCGGCGGAATTGATGGGGGTCAATGTTTCGCGCATCTACATTCTCTGTTTTGG
CCTGGGAGCCGCCTGTGTCGGACTGGCTGCGGTTCTGATTGCACCGCTCTATCCGACTTCTTCAAATATCGGC
ACCTATTTCGTGCTGACGGCCTTCGTGGTTGTGGTGCTCGGTGGCCTTGGCTCGATCCCTGGAGCCTTCGTTG
GTGCGCTGATCATCGGCGTGATCGACACCATGTCGGGCTACTACATCGGATCAGACCTGCGGGAAGCCGTCGT
ATTCGGCATTTTCCTCCTGATCCTCATCCTCAAGCCTTCTGGCCTCTTTGGCAAGCAGCTTAATCTTTCGCAT
TTGTCCTCAGGAATATTCATGAGCATCAGTGACACCATAGACCGCGCCGGCATGAATAAGTCCAGATTGGGCA
AAGGGCAGATCGCCTTCTGTGTTTTGCTGGCTTCGCTTCTGCTTCTGCCGCTGGCAATCAACAATGCCTTCGT
CTCGCATATCTTTATAACCATTTGCCTTTTCGCCGCCCTGTCCACGGCTTGGAACATCGTGGGTGGCTTTGCA
GGCCAGATGTCGCTCGGCCATGCCGTTTTCTACGGTATCGGCGGTTACACAGGCGTGATCCTGTTCAATATGG
GGATCAGCCCATGGTTCAGCATGTTCATCGGCGCGTTCATCGCCGCGCTGGTAGGCATGGTCATATCCTATCC
CTGCTTTCGTCTGAAAGGCCCGTTCTATTCGCTGGCGTCCATTGCGTTTCTGGAAGTGTTTCGCGTGCTGGCG
CTGCATTTTGGATGGCTTACGGGCGGTGCTACGGGGCTCATGATCCAaCTCAAGCTCGGCTGGGTCTGGATGG
TTTTCCGTGAACGCTGGCCGTCGTTGCTGATCGTATTCGGCATGTTGCTGGTGACGCTTGCAATCACCTGGGC
GGTTCGCCGCTCGCGTCTGGGCTTTTATCTGGTTGCCACGCGTGAGCGTGAATCGGCCGCACGCGCCGCCGGC
GTTCGCACCGTTCGCGTGCGTTTGATCGCGGTCGCCATATCGTCGGCACTCTGCGCGATGCTGGGCACATTCC
ATGCGATGTATCTGACATTCATTGAACCTGCTGCGATGTTCTCGCTCGCCTTCTCGATCCAGATCGCAATGTT
CGCCCTGATTGGTGGTCTTGGTACCGTGGCTGGCCCCCTATTGGGTGCGGTGCTTCTCGTTCCTATCACGGAA
TGGGCGCGTGCTTCACTCGGTGCTTCGGCCCTCGGCCTGCATGGCTTCGTCTACGGCCTTGTCCTGATCCTCG
TCGTACTTTTCATGCCGAACGGCATCATGGGGGCGATCAACCGCTTTGTCCGCAAGCCGCAAGATAGTGAAGA
AACGGCAACGGCACGAACGGAGCCAATTGCGGCTGTGCCGGCCAGGGCCATTAAAGCGCCGTCGCCGGACCGT
GCGGGGATCGGGCAGGATATTCTGCGCGTGCAGAACCTGAACAAGCATTTCGGTGGCTTGCATGTGACGCGCA
ATGTCAGCTTTACCCTGCGCGAAGGTGAAGTACTCGGTTTGATCGGCCCCAATGGTGCGGGCAAGACCACATT
GTTCAACATGATTTCTGGTTTTCTTGCCCCGGATGAGGGTACGGTCAACCTGTGTGGGGCGGACGGCCAATTC
CATGCTCCGAAAAACCCGGCGGATTTTGCGGCGCTGGGACTTGGCCGCACCTTCCAGATCGTGCAGCCGTTTG
CGGCCATGACGGTCGAGGAAAACATTATGGTGGGGGCTTTCTATCGCCACCACCATGAAAAGGATGCCCGTGA
AGCGGCACGGGAAACCGCCTGGCGCATGGGGCTTGGCCCCTTGCTCGGGGCGGAAGCGCGAGGGCTGACTATC
GGTGGTTTGAgGCGGTTGGAAGTTGCCCGCGTCATGGCGATGGAACCGCGCATTCTGCTGCTTGATGAAGTGA
TGGCCGGTATCgACCAGACTGATGTTCGGCGCGCTATCGACCTGATGCTGTCCATCCGCGACAGCGGTGTTTC
GATCATCGCCATTGAGCACGTCATGCAGGCCGTCATGTCGCTCTCGGACCGCGTTATCGTCCTGGCGTCGGGC
GAGGTGATAGCCCAGGGGCGGCCGCAGGATGTGGTGCGCGATCCTCAGGTTGTCGAAGCCTATCTGaGCAAGG
AGTTTGCACATGCTCACGCTTGc

TABLE 5: (Continued)

BMEII0685m

ATGTGCGGAATCATCGGCATTATCGGAAATGACGAGGTCGCTCCGCGTCTCGTGGACGCATTGAAGCGCCTTG
AATATCGCGGCTACGATTCCGCCGGCATTGCCACATTGCAGAATGGCAGGCTCGACCGCCGCCGCGCCGAAGG
CAAACTCGTCAATCTGGAAAAGCGTCTTGCGGGCGAGCCGCTTCCGGGCGTGATCGGCATCGGCCATACCCGT
TGGGCAACCCATGGCAGGCCGGTGGAGCACAATGCGCATCCGCATATCACCACACGTCTTGCCGTGGTTCACA
ATGGAATCATCGAAAATTTCGCCGAATTGCGCGCCATGCTGGAAGCCGAAGGCCGCAAATTTGAAACGGAAAC
CGACACGGAAGCCGTCGCCCATCTGGTGACGCGCGAACTGGAAAAGGGCAAGTCGCCGGTGGAAGCCGTGCGC
GATTGCCTGCCGCATCTCAAAGGCGCTTTTGCACTCGCCTTCCTGTTTGAGGGCGATGAAGAACTGCTGATCG
GCGCACGCCAGGGGCCGCCGCTTGCGGTTGGCTATGGTGAAGGCGAAATGTTCCTCGGCTCCGATGCGATTGC
GCTCGCACCTTTCACCGATACCATCTCCTATCTGGAAGATGGCGACTGGGCTGTGCTGACCCGCAATGGCGTC
AGCATCTATGACGAAAACAACAAGCCGGTTGAGCGCCCGGTCCAGAAGTCGCAGAACACCAATATGCTGGTAT
CGAAGGGCAACCATCGCCACTTCATGCAGAAGGAAATGTTCGAGCAGCCGGAAGTCATTTCCCACACGCTTGC
CAATTATCTCGACTTCACGACGGGCAAGGTGCGCAAGGAAGCGATCGGTATCGATTTCAGCAAGGTCGATCGC
CTGACGATCACCGCTTGCGGCACGGCCTATTATGCCGCAACGGTTGCGAAATACTGGTTTGAACAGATTGCGC
GCCTGCCGGTCGATAGCGATATCGCGtCGGAATTCCGCTACCGCGAAATGCCGCTCTCGAAGGATTCGCTGGC
CATGTTCGTTTCGCAGTCGGGCGAAACGGCGGATACACTTGCTTCGCTGCGCTATTGCAAGGCGCAGGGCCTG
AAAATCGCCTCGGTGCTCAACGTGACCGGCTCCACCATCGCGCGTGAATCGGATGCAGTGTTCCCGACGCTCG
CAGGCCCTGAAATCGGCGTTGCTTCCACCAAGGCCTTCACCTGCCAGCTTTCGGCCATGGCCTCACTCGCTAT
TGCGGCGGCGCGTGCGCGTGGTGCAATCGACGAGGTTCGCGAGCAGGAACTGGTGCACCAGCTTTCCGAAGCG
CCGCGTTTCATCAATCAGGTTTTGAAGCTTGAAGACCAGATTGCTGTCGTCTGCCATGACCTGTCGAAGGTCA
ATCATGTGCTATATCTCGGTCGCGGCACGTCCTTCCCGCTCGCCATGGAAGGCGCGCTGAAGCTCAAGGAAAT
CTCCTATATCCACGCCGAAGGCTATGCGGCAGGTGAGTTGAAGCATGGGCCGATTGCGCTCATCGATGAAACC
ATGCCGGTGATCGTCATCGCACCATCTGATCGTCTCTATGAGAAGACCGTGTCGAACATGCAGGAAGTGGCTG
CGCGCGGCGGGCGCATCATCCTCATCACCGACAAGAAGGGGGCAGAAAGCGCCAGCATCGACACGATGGCCAC
CATCGTTCTGCCCGAGGTGCCGGAATTCATCTCGCCGCTCGTCTATGCGCTGCCGATCCAGATGCTCGCCTAT
CACACGGCAGTCCTTATGGGAACGGACGTGGACCAGCCGCGCAATCTGGCCAAGTCTGTTACTGTCGAGTAc

BMEI0267m

ATGGTACAGGTTTCCGGACGCGCAGATATGGCGCCTGCGCGTGCGCCGATTGTGACCAGAGAGCGGATTGCTG
TTGCGCTCCTATTTCTGATGAACGGCTATATTTTCGGTGGCTGGGCCCCAAAAATCCCGGAATTTGCAGAACG
TCTCGGGCTTGATAGCGCCGGAATGGGCCTGATGATCCTGGTCTGCGGTCTCGGTTCGCTGGCCATGATGCCG
GTCGCAGGCGCTCTTTCCGCGCATCGCGGTTCGGGCATTGTGGTGCGCATCTTTGCCCTTGCCTTCATTCCGG
CGCTGCTCATCATCACATTGGTGCCGAATGTGGTGACGGCGGTGATCGTTATGCTCTATTTCGGCGGAACCAT
GGCTGCGATGGATGTGTCCATGAACGCCAATGCGGTCGCAGTCGAGAAAAAGATGCGCCGGGCAATCATGTCG
TCCTGTCACGCCTTCTGGAGCCTTGGTGGCCTCATCGGCGCGGCAACGGGCGGTTTTCTCATCGCGCAATTCG
GCTCGATCGTTCATGCGCTGGTCACGATGATCGCAGCCACCATCCTGATCATTGTCGCATGGCCCTCGATCAT
CCCGGATACGGAGCATCATCATCCAGACGGCGAGAAGCAGAAGCTTGCGCTGCCGCGCAATCCCCTGCCGTGG
CTTGTCGGCGTCATGGCGCTGTTTTCCATGGTGCCGGAAGGGGCGATCCTTGATTGGGGCGCCTATCATATGC
GCCAGGATCTCGGCGCTTCCGTCACGGTGGCAGGTTTCGGCTTCGCGGCATTTTCGGGTTCCATGGCCGTTAT
GCGCTTTGCGGGTGATCTGGTGCGCGACCGATTCGGTGCCGTGAAAACCCTGCGTGCCTGCACGGCCATCGCC
ATCATAGGCATGTTGATCGTCGGTTTTGGAAACAGTTCGGCGACTGCAATCATCGGCTTTGCGATCTGCGGCA
TCGGCATTTCCAACATGGTGCCGATAGCTTTCTCGATGGCTGGCAACATGCCGGGCGTCAATCCAAGTGTCGG
CCTGTCGATAGCCACCACGCTTGGCTATTCCGGCATGTTGGTCGCGCCATCACTGATCGGCTTTGTCGCCAGG
CATAGCGGCTTCGGTGTGGTGTTTCTGGCGCTTCCGGCCTTGCTGCTGGTGGTTCTCGCCTTCTCAAGCCTCG
CCCGATACGCTGACCATAAGCATTAc

TABLE 5: (Continued)

BMEI0682m

GTGGCGCAGAATGACAAGCAGCAAAACCCGCCTGCCTCGCAGTCCGCGCCGGCGGAAAAGCCGAATGGGCAGG
CAGCTTCCGCTGCCGATCAACGGCCCCCCGAACCCGTATCGAGCCCCGTATCAAGCATTGTACAGCAGCAGCG
GCCCACGTTCGATGAGCTGAAGCAACTGACGAAGAAGATCAACGAGCAGCTTTCCAAGTCCGCGTCGAGCGAT
GAGGCGCTTGCCAATCTCAAGCTTCAGCTTGACGCCCTTTCCAAGAAGCTTCTGGAAACGGGCGTTGCCTTCC
GCCCGCGCCTTACCGAGATCAACACGCGGCTGGAGCAGCTTGGCCCCGCGCCTTCGAGCGATCAGCCGGCGGA
GCCGGCCATCGTTGGGGAAGAGCGTGCACGGCTGATAGCGGAAAAGGCAGAAATCAACGCTACCCTTGGCGAA
ACGGAAGACACGTCGCTTGCCGTGAACCGCATGTCTGCGTTGATCGGCGATATGCGGCGTGATCTCTTCACCA
AGACGCTTTCGCAGCGCGTCAATCTGGATTCGACGCTTGGCAGCGAAGTGGTTTCGGCTGCAAGCGACCAGAT
GATATCGTTGTGGCGTATCGTGCGATCGTGGTGGCGTTTCGTTGTCACCTTCAAGCTTGAATCGTTTCTGGCT
GCCGCTTTCTTTGCCCTTGCCGCCGCATTGGTGCTTCAGTTCGGCGCGCAGCGGTTTCTGGGCGCCTTTTACC
GGCGCGACCCGTCCGTCGAATCCCCATCCTATCTGAGCCGTTTATCGGTTGCGTTCTGGTCAACGGTCATTCC
GTCAGCAGCCGTTGGCGTCTTTCTGGCGACGACATATTTCCTCCTCAATTACTTCAATGTGTTGAGGACGGAT
ATTGCATCGCTTTTCCAATCGCTTTTCATCGTATTGGGGCTGGTGTTCTTCATTCACCGGCTGGCCGTGGCCT
GCATCAGTTCCGATATGCCGCAATGGCGGCTGGTGCAGGTGGCCCCGCGCCCGGGGCATCTTCTGGCCTGGCT
GGTGACAGCCACCGCACTTACCAGCGGACTTGATTCCTTCTTTGGAACAGTCAATCGCATACTTTCCTCGCCA
CTCTCGCTGACTATGGCAAAAAGCCTGATCGCGACGGTGATTATCGGGGTGCTCATTCTGGCCATCGCCTTTG
TCAAGCCTGTCGAGCGAGAGAAGGATGGCGCTGTCCGTGCGTGGCCACGCGCTTTCAGGATATTCCTGATTTT
GATGGGGCTCTTGCCTATTCTCACGGCCCTTTCCGGCTATATCGGCATTGCGCGTTTCATCTCGCAGCAGATC
GTCGTGACGGGCGCTTTCCTTGTCACCATGTATAcGGGCTTCCTGACAGGGCGCGCGATTTCCGAGGAGCAGG
CCTTCGCCTCAAGCCGGATCGGCAAGGCAATGCGCGAGCGTTTTCATTTCGATGAAGCAACGCTCGACCAGCT
TGGCCTTCTGGCTGGTATTCTCATCAATCTGGTCGTCGCGCTGATCGGCATTCCGCTGGTTTTGATGCAGCTT
GGTTTTCAGTGGGCGGAGCTGAAAAGCACATTCTACAAGCTGATGACCGGCTTCCAGATCGGAAATTTCTCCA
TCTCGCTCATGGGGCTCCTGTCGGGTGTGCTGCTGTTTCTCATCGGCTATGTCCTGACGCGCTGGTTCCAGAA
CTGGCTGGATAACAGCGTCATGGCGCGTGGCCGGGTGGATTCGGGTGTGCGCAATTCCATCCGCACTGTTGTC
GGTTATGTCGGGCTTTGTCTCGCGGCGCTGATGGGCATTTCGGCGGCCGGGTTCAACCTTGCCAATCTCGCAC
TGATTGCTGGCGGCCTCTCTCTCGGTATCGGTTTCGGCCTCCAGAATATCGTCCAGAACTTTGTTTCCGGCCT
GATCCTGCTGGCAGAACGCCCCTTCAAGGTGGGCGACTGGGTGGAGGCGGGTACGGTCAGCGGCATTGTGAAG
AAGATCAGCGTGCGCGCGACGGAAGTGGAAACATTCCAGAAACAGTCGATTATCGTGCCGAATTCGACGCTCA
TCAACGGCAATGTCGGCAACTGGACGCACCGCAACAAGCTTGGCCGCATCGACATCAATGTGCAGGCTTCCTT
TACAGAAGACCCGCGCCGCGTCCACGCGCTTTTGCTGGAGATCGTGCGTGGCCATCCATCCATTCTGAAGAAC
CCGGAACCGTTCGTTTCCTTTCAGAGCATGACCGGTTCGCTGCTCGTTTTCGATGTTTATGCCCATGTGGCCG
ACATTACGTCGACCGGCAGTATCAAGAACGAATTGCGATTCCAGATTGTCGAGCGTTTCCATGAACAGGGGTT
GAGCCTGTCATCTTCCTCGACAGACCTTATATTGAAGGCCCCCGATGTGGAGAAACTTTCGGAACTGATGCAG
GAGGAAAAAGGACTTTCAGCGGCAGCCAGCAGGGAGAAGACGGGCGAAAAGAAGCCTGAGGAGGGTGACAAGG
ACGATCGTGCGTAc

TABLE 6: Extent of the Target Secondary Structure in the Probe Binding Sites for *Brucella melitensis* 16M Microarray

	EXTENT OF BINDING SITE INACCESSIBILITY						MAXIMUM CONSECUTIVE STRETCH FOR FOLDED BINDING SITE	
	SET OF 6 BINDING SITES			SINGLE FOLDED BINDING SITE				
	MIN	MAX	AVERAGE	MIN	MAX	AVERAGE	ACCESSIBLE BASES	INACCESSIBLE BASES
SCORE	105	133	125.4	6	38	25.08	41	19
%	42	53.2	50.2	12	76	50.16	82	38

TABLE 7: Extent of the Target Secondary Structure in the Probe Binding Sites for *Brucella melitensis* 16M Miniarray

	EXTENT OF BINDING SITE INACCESSIBILITY						MAXIMUM CONSECUTIVE STRETCH FOR FOLDED BINDING SITE	
	SET OF 6 BINDING SITES			SINGLE FOLDED BINDING SITE				
	MIN	MAX	AVERAGE	MIN	MAX	AVERAGE	ACCESSIBLE BASES	INACCESSIBLE BASES
SCORE	105	133	122.2	7	35	24.44	32	19
%	42	53.2	48.9	14	70	50.16	64	38

TABLE 8: Target Percent Bound on a Miniarray: Temperature Series

STRUCTURE	TEMPERATURE			
	55 °C		56 °C	
	OPTIMAL % BOUND	TOTAL % BOUND	OPTIMAL % BOUND	TOTAL % BOUND
FOLDED TARGET				
BMEII0462m	1.89E-23	2.51E-22	6.92E-23	1.04E-21
BMEII0874m	1.79E-23	3.11E-22	5.65E-23	9.85E-22
BMEII0685m	2.45E-25	3.15E-24	1.26E-24	1.70E-23
BMEI0267m	3.20E-24	4.72E-23	3.13E-24	4.21E-23
BMEI0682m	2.01E-23	2.88E-22	5.98E-23	8.95E-22
PROBE_TARGET HYBRID				
BMEII0462m+BMEII0462m_5	0.718459	0.72368135	1.16689	1.17548944
BMEII0462m+BMEII0462m_4	0.000564092	0.000567073	0.000858083	0.000863538
BMEII0462m+BMEII0462m_3	0.0130042	0.013169194	0.0128695	0.013038985
BMEII0462m+BMEII0462m_2	0.308334	0.310478573	0.476668	0.480076489
BMEII0462m+BMEII0462m_1	0.00589107	0.005929008	0.0600037	0.060396087
BMEII0462m+BMEII0462m_0	98.3626	98.9461362	97.679	98.270181
BMEII0874m+BMEII0874m_5	0.73788	0.74688005	0.679083	0.68734294
BMEII0874m+BMEII0874m_4_M	3.03369	3.05448555	2.14757	2.16276876
BMEII0874m+BMEII0874m_3	24.4596	24.46956641	24.2262	24.2380141
BMEII0874m+BMEII0874m_2	0.458484	0.465436301	0.577623	0.586656034
BMEII0874m+BMEII0874m_1_M	61.1944	61.557931	64.2534	64.7370943
BMEII0874m+BMEII0874m_0	9.60636	9.7057732	7.51032	7.5881123
BMEII0685m+BMEII0685m_5	0.00110907	0.001116827	0.00222037	0.002236464
BMEII0685m+BMEII0685m_4	0.00631068	0.006328709	0.00890947	0.00893578
BMEII0685m+BMEII0685m_3	0.0201172	0.020252392	0.0308316	0.031090525
BMEII0685m+BMEII0685m_2	0.0746631	0.075239522	0.662095	0.667665112
BMEII0685m+BMEII0685m_1	0.25517	0.25517	0.496499	0.496499
BMEII0685m+BMEII0685m_0	99.0456	99.641937	98.1979	98.793621
BMEI0267m+BMEI0267m_5	9.36657	9.5045677	2.82619	2.86830842
BMEI0267m+BMEI0267m_4	0.000152649	0.000152875	7.44E-05	7.45E-05
BMEI0267m+BMEI0267m_3	0.0291645	0.029433041	0.0117862	0.011897345
BMEI0267m+BMEI0267m_2	5.11E-06	5.14E-06	2.26E-06	2.27E-06
BMEI0267m+BMEI0267m_1	0.0133111	0.013421539	0.00609502	0.006145954
BMEI0267m+BMEI0267m_0	89.5638	90.45247	96.1566	97.113572
BMEI0682m+BMEI0682m_5	5.64818	5.72159434	5.25396	5.32262444
BMEI0682m+BMEI0682m_4	2.87034	2.8823719	3.51548	3.530236
BMEI0682m+BMEI0682m_3	12.5596	12.7749758	8.36716	8.5163976
BMEI0682m+BMEI0682m_2	0.161517	0.162661824	0.156648	0.157775386
BMEI0682m+BMEI0682m_1_M	64.6647	64.884062	70.9639	71.328958
BMEI0682m+BMEI0682m_0	13.4569	13.57436644	11.0515	11.1440677

TABLE 8: (Continued)

STRUCTURE	TEMPERATURE			
	57 °C		58 °C	
	OPTIMAL % BOUND	TOTAL % BOUND	OPTIMAL % BOUND	TOTAL % BOUND
FOLDED TARGET				
BMEII0462m	2.53E-22	4.78E-21	9.82E-22	1.80E-20
BMEII0874m	1.83E-22	3.62E-21	6.01E-22	1.13E-20
BMEII0685m	6.57E-24	8.72E-23	3.28E-23	4.31E-22
BMEI0267m	1.54E-23	2.13E-22	7.70E-23	1.08E-21
BMEI0682m	1.85E-22	2.71E-21	1.31E-22	1.98E-21
PROBE_TARGET HYBRID				
BMEII0462m+BMEII0462m_5	1.8312	1.84494374	10.1324	10.2095151
BMEII0462m+BMEII0462m_4	0.00123673	0.001246816	0.00888741	0.008964235
BMEII0462m+BMEII0462m_3	0.0150442	0.015249812	0.0162226	0.016452653
BMEII0462m+BMEII0462m_2	3.31671	3.3407328	3.2662	3.29018034
BMEII0462m+BMEII0462m_1	0.0833201	0.083881823	0.106775	0.107516993
BMEII0462m+BMEII0462m_0	94.1247	94.7139335	85.8251	86.3673192
BMEII0874m+BMEII0874m_5	0.649426	0.65733274	0.631031	0.63884088
BMEII0874m+BMEII0874m_4_M	1.9301	1.94419711	1.25496	1.26440477
BMEII0874m+BMEII0874m_3	19.3514	19.3627163	12.8847	12.8936979
BMEII0874m+BMEII0874m_2	0.739302	0.75112941	0.944952	0.96145725
BMEII0874m+BMEII0874m_1_M	70.6928	71.236019	78.7194	79.335488
BMEII0874m+BMEII0874m_0	5.98699	6.0485317	4.8562	4.9061896
BMEII0685m+BMEII0685m_5	0.00454542	0.004578761	0.00890663	0.008973565
BMEII0685m+BMEII0685m_4	0.0126082	0.012646677	0.0173493	0.017433304
BMEII0685m+BMEII0685m_3	0.0488058	0.049239287	0.0736046	0.074314144
BMEII0685m+BMEII0685m_2	1.01029	1.019604941	1.32681	1.34029918
BMEII0685m+BMEII0685m_1	0.966202	0.966202	1.75695	1.75695
BMEII0685m+BMEII0685m_0	97.3501	97.947685	96.2133	96.801975
BMEI0267m+BMEI0267m_5	4.29318	4.35802721	5.85707	5.9462224
BMEI0267m+BMEI0267m_4	0.000169702	0.000169967	0.000449259	0.000449985
BMEI0267m+BMEI0267m_3	0.0781897	0.078948331	0.143455	0.144889835
BMEI0267m+BMEI0267m_2	4.53E-06	4.56E-06	1.09E-05	1.09E-05
BMEI0267m+BMEI0267m_1	0.0156282	0.015756571	0.0408626	0.041193504
BMEI0267m+BMEI0267m_0	94.4563	95.547124	92.7867	93.867188
BMEI0682m+BMEI0682m_5	4.484	4.5435998	5.97182	6.06239105
BMEI0682m+BMEI0682m_4	3.10756	3.1206213	0.626028	0.62866273
BMEI0682m+BMEI0682m_3	5.87429	5.98391617	0.983409	1.00228591
BMEI0682m+BMEI0682m_2	0.159742	0.160909502	0.0382294	0.038511223
BMEI0682m+BMEI0682m_1_M	79.1092	79.541253	90.8013	91.319686
BMEI0682m+BMEI0682m_0	6.59576	6.649654131	0.941013	0.948456191

TABLE 8: (Continued)

STRUCTURE	TEMPERATURE			
	59 °C		60 °C	
	OPTIMAL % BOUND	TOTAL % BOUND	OPTIMAL % BOUND	TOTAL % BOUND
FOLDED TARGET				
BMEII0462m	4.62E-21	8.09E-20	2.07E-20	3.80E-19
BMEII0874m	2.26E-21	4.03E-20	9.12E-21	1.64E-19
BMEII0685m	1.65E-22	2.28E-21	6.98E-22	9.06E-21
BMEI0267m	3.85E-22	5.92E-21	1.89E-21	2.93E-20
BMEI0682m	4.39E-22	6.39E-21	1.45E-21	2.22E-20
PROBE_TARGET HYBRID				
BMEII0462m+BMEII0462m_5	14.0321	14.1417621	19.564	19.7190802
BMEII0462m+BMEII0462m_4	0.013196	0.013318009	0.018987	0.019176295
BMEII0462m+BMEII0462m_3	0.0213315	0.021678989	0.0274487	0.027909505
BMEII0462m+BMEII0462m_2	3.87119	3.90043864	4.52002	4.55484612
BMEII0462m+BMEII0462m_1	0.16643	0.16760398	0.251693	0.25352243
BMEII0462m+BMEII0462m_0	81.2289	81.7552006	74.9272	75.4254789
BMEII0874m+BMEII0874m_5	0.711422	0.72020754	0.883334	0.89564002
BMEII0874m+BMEII0874m_4_M	0.926313	0.93345154	2.24325	2.26107444
BMEII0874m+BMEII0874m_3	9.9312	9.93936137	8.34825	8.35645052
BMEII0874m+BMEII0874m_2	8.28026	8.4304667	11.3954	11.6065101
BMEII0874m+BMEII0874m_1_M	74.8303	75.426071	71.844	72.421896
BMEII0874m+BMEII0874m_0	4.50441	4.55044561	4.40572	4.45840168
BMEII0685m+BMEII0685m_5	0.0169104	0.017041313	0.122565	0.123541075
BMEII0685m+BMEII0685m_4	0.024505	0.024627709	0.0292297	0.029382595
BMEII0685m+BMEII0685m_3	0.112506	0.11367077	0.146479	0.148096156
BMEII0685m+BMEII0685m_2	1.63118	1.64992866	1.51338	1.53300094
BMEII0685m+BMEII0685m_1	3.29604	3.29604	18.2404	18.2404
BMEII0685m+BMEII0685m_0	94.3083	94.898658	79.4281	79.925586
BMEI0267m+BMEI0267m_5	7.60779	7.7250357	9.64143	9.7935098
BMEI0267m+BMEI0267m_4	0.00113474	0.001136609	0.0214752	0.021511263
BMEI0267m+BMEI0267m_3	0.245635	0.248153459	1.41801	1.43297265
BMEI0267m+BMEI0267m_2	2.70E-05	2.72E-05	0.000317531	0.000320116
BMEI0267m+BMEI0267m_1	0.0980403	0.098859503	0.224658	0.226567806
BMEI0267m+BMEI0267m_0	90.857	91.926757	87.4775	88.525081
BMEI0682m+BMEI0682m_5	4.96895	5.04576726	4.23936	4.30591375
BMEI0682m+BMEI0682m_4	0.579763	0.58220624	0.5429	0.54519089
BMEI0682m+BMEI0682m_3	0.756941	0.77214351	0.550866	0.562307317
BMEI0682m+BMEI0682m_2	0.0420385	0.042355496	0.0455668	0.045915884
BMEI0682m+BMEI0682m_1_M	92.2218	92.927293	93.3522	94.113117
BMEI0682m+BMEI0682m_0	0.630190428	0.630190428	0.424188	0.427594882

TABLE 8: (Continued)

STRUCTURE	TEMPERATURE			
	61 °C		62 °C	
	OPTIMAL % BOUND	TOTAL % BOUND	OPTIMAL % BOUND	TOTAL % BOUND
FOLDED TARGET				
BMEII0462m	8.93E-20	1.57E-18	3.59E-19	6.65E-18
BMEII0874m	3.72E-20	6.57E-19	1.40E-19	2.39E-18
BMEII0685m	3.33E-21	4.66E-20	1.46E-20	2.34E-19
BMEI0267m	9.15E-21	1.31E-19	4.33E-20	6.47E-19
BMEI0682m	5.33E-21	8.21E-20	1.93E-20	3.04E-19
PROBE_TARGET HYBRID				
BMEII0462m+BMEII0462m_5	28.6801	28.9109357	39.2494	39.5698549
BMEII0462m+BMEII0462m_4	0.0261987	0.026474284	0.0341156	0.034501176
BMEII0462m+BMEII0462m_3	0.0330371	0.033612139	0.0378397	0.03852171
BMEII0462m+BMEII0462m_2	4.91708	4.9559962	5.16262	5.20374641
BMEII0462m+BMEII0462m_1	0.359457	0.36214873	0.489497	0.49321644
BMEII0462m+BMEII0462m_0	65.2655	65.7108265	54.2855	54.6601019
BMEII0874m+BMEII0874m_5	1.09946	1.115003535	1.25842	1.27664523
BMEII0874m+BMEII0874m_4_M	1.90216	1.91774408	2.1933	2.21194218
BMEII0874m+BMEII0874m_3	7.04051	7.0487555	5.48113	5.48877529
BMEII0874m+BMEII0874m_2	15.7347	16.0354744	19.8176	20.2096226
BMEII0874m+BMEII0874m_1_M	69.2511	69.822484	61.4784	61.990966
BMEII0874m+BMEII0874m_0	4.01237	4.06053322	8.71694	8.8220267
BMEII0685m+BMEII0685m_5	0.191668	0.19321831	0.301733	0.304223321
BMEII0685m+BMEII0685m_4	0.0393694	0.039585698	0.0535085	0.053817343
BMEII0685m+BMEII0685m_3	0.217515	0.220106883	0.303096	0.307022832
BMEII0685m+BMEII0685m_2	1.57359	1.59736545	2.16675	2.20119136
BMEII0685m+BMEII0685m_1	20.3894	20.3894	22.6133	22.6133
BMEII0685m+BMEII0685m_0	77.0755	77.560275	74.0473	74.520451
BMEI0267m+BMEI0267m_5	12.3446	12.541296	15.6173	15.8703706
BMEI0267m+BMEI0267m_4	0.0592817	0.059384713	0.159512	0.160064532
BMEI0267m+BMEI0267m_3	0.418659	0.423168367	0.571566	0.577916631
BMEI0267m+BMEI0267m_2	0.000687226	0.000693199	0.00142157	0.001434644
BMEI0267m+BMEI0267m_1	1.21779	1.22844519	2.06347	2.08197837
BMEI0267m+BMEI0267m_0	84.7201	85.746967	80.2373	81.3082312
BMEI0682m+BMEI0682m_5	3.62903	3.68677699	3.18115	3.23316985
BMEI0682m+BMEI0682m_4	0.566251	0.56860779	0.593103	0.59561226
BMEI0682m+BMEI0682m_3	0.399603	0.408286627	0.302531	0.309355323
BMEI0682m+BMEI0682m_2	0.0474426	0.047811906	0.0464481	0.046815534
BMEI0682m+BMEI0682m_1_M	94.0062	94.97519	94.5448	95.571784
BMEI0682m+BMEI0682m_0	0.310746	0.313326136	0.24116	0.243229786

TABLE 8: (Continued)

STRUCTURE	TEMPERATURE			
	63 °C		64 °C	
	OPTIMAL % BOUND	TOTAL % BOUND	OPTIMAL % BOUND	TOTAL % BOUND
FOLDED TARGET				
BMEII0462m	1.36E-18	2.48E-17	5.15E-18	9.24E-17
BMEII0874m	5.76E-19	1.04E-17	2.35E-18	4.05E-17
BMEII0685m	6.40E-20	1.08E-18	2.69E-19	5.06E-18
BMEI0267m	2.02E-19	3.27E-18	9.17E-19	1.27E-17
BMEI0682m	7.56E-20	1.41E-18	3.51E-19	6.18E-18
PROBE_TARGET HYBRID				
BMEII0462m+BMEII0462m_5	50.6755	51.095239	58.5291	59.023229
BMEII0462m+BMEII0462m_4	0.0411164	0.041743512	0.0512555	0.052096843
BMEII0462m+BMEII0462m_3	0.0369152	0.037595605	0.0368541	0.037559137
BMEII0462m+BMEII0462m_2	5.0685	5.11012298	5.16863	5.20922737
BMEII0462m+BMEII0462m_1	0.620478	0.62533379	0.810848	0.81738235
BMEII0462m+BMEII0462m_0	42.7891	43.0898915	34.61	34.8604994
BMEII0874m+BMEII0874m_5	1.5707	1.59408572	1.68631	1.71181785
BMEII0874m+BMEII0874m_4_M	2.77011	2.79459399	3.68089	3.71420282
BMEII0874m+BMEII0874m_3	4.70597	4.71379137	4.18505	4.1931823
BMEII0874m+BMEII0874m_2	22.2253	22.6661337	22.9096	23.375458
BMEII0874m+BMEII0874m_1_M	60.3243	60.834308	60.0845	60.599709
BMEII0874m+BMEII0874m_0	7.30789	7.3970882	6.32785	6.405629
BMEII0685m+BMEII0685m_5	0.437379	0.441103962	0.607683	0.61293498
BMEII0685m+BMEII0685m_4	0.0714799	0.072009439	0.0943418	0.095070676
BMEII0685m+BMEII0685m_3	0.366557	0.37219922	0.403104	0.410413156
BMEII0685m+BMEII0685m_2	3.10382	3.1549237	4.41069	4.4852458
BMEII0685m+BMEII0685m_1	25.1272	25.1272	27.2475	27.2475
BMEII0685m+BMEII0685m_0	70.3793	70.832525	66.7202	67.148792
BMEI0267m+BMEI0267m_5	19.4418	19.7574813	23.5416	23.9352392
BMEI0267m+BMEI0267m_4	0.40805	0.409522936	0.698219	0.70082564
BMEI0267m+BMEI0267m_3	0.767495	0.77635713	1.03234	1.04461679
BMEI0267m+BMEI0267m_2	0.00289075	0.002919054	0.00572955	0.005808912
BMEI0267m+BMEI0267m_1	3.31831	3.34848799	3.10142	3.12798812
BMEI0267m+BMEI0267m_0	74.6816	75.7052169	70.1986	71.1855458
BMEI0682m+BMEI0682m_5	3.29258	3.34665816	3.20394	3.25792091
BMEI0682m+BMEI0682m_4	0.810174	0.81360604	4.33135	4.3494497
BMEI0682m+BMEI0682m_3	0.295668	0.302605767	0.28548	0.292384472
BMEI0682m+BMEI0682m_2	0.0523887	0.052901664	0.0575492	0.058123787
BMEI0682m+BMEI0682m_1_M	94.1576	95.244134	90.7161	91.811435
BMEI0682m+BMEI0682m_0	0.237976	0.240096692	0.228555	0.230681615

TABLE 8: (Continued)

STRUCTURE	TEMPERATURE	
	65 °C	
	OPTIMAL % BOUND	TOTAL % BOUND
FOLDED TARGET		
BMEII0462m	1.89E-17	3.41E-16
BMEII0874m	1.01E-17	1.75E-16
BMEII0685m	1.22E-18	2.23E-17
BMEI0267m	4.08E-18	5.72E-17
BMEI0682m	1.67E-18	3.00E-17
PROBE_TARGET HYBRID		
BMEII0462m+BMEII0462m_5	65.3128	65.877025
BMEII0462m+BMEII0462m_4	0.516699	0.525899915
BMEII0462m+BMEII0462m_3	0.0358446	0.036553445
BMEII0462m+BMEII0462m_2	5.04337	5.09340673
BMEII0462m+BMEII0462m_1	1.03067	1.03909509
BMEII0462m+BMEII0462m_0	27.2256	27.4280207
BMEII0874m+BMEII0874m_5	1.84775	1.87631721
BMEII0874m+BMEII0874m_4_M	5.22883	5.27744138
BMEII0874m+BMEII0874m_3	3.89438	3.90337881
BMEII0874m+BMEII0874m_2	24.36	24.8691014
BMEII0874m+BMEII0874m_1_M	57.747	58.251508
BMEII0874m+BMEII0874m_0	5.75133	5.8222555
BMEII0685m+BMEII0685m_5	0.726228	0.73273306
BMEII0685m+BMEII0685m_4	0.107295	0.108168488
BMEII0685m+BMEII0685m_3	0.433964	0.442552415
BMEII0685m+BMEII0685m_2	5.1783	5.2691447
BMEII0685m+BMEII0685m_1	29.7568	29.7568
BMEII0685m+BMEII0685m_0	63.2789	63.6905403
BMEI0267m+BMEI0267m_5	28.2689	28.7424633
BMEI0267m+BMEI0267m_4	1.09922	1.10346288
BMEI0267m+BMEI0267m_3	1.36193	1.37851208
BMEI0267m+BMEI0267m_2	0.0110889	0.011249859
BMEI0267m+BMEI0267m_1	2.85341	2.88347262
BMEI0267m+BMEI0267m_0	64.9515	65.8807661
BMEI0682m+BMEI0682m_5	3.31681	3.37360294
BMEI0682m+BMEI0682m_4	4.60158	4.621123
BMEI0682m+BMEI0682m_3	0.284661	0.292319557
BMEI0682m+BMEI0682m_2	0.0671076	0.067780792
BMEI0682m+BMEI0682m_1_M	90.2611	91.409039
BMEI0682m+BMEI0682m_0	0.233893	0.236141049

TABLE 9: Target Percent Bound on a Miniarray: Formamide Series

STRUCTURE	FORMAMIDE CONCENTRATION			
	0 %		5%	
	OPTIMAL % BOUND	TOTAL % BOUND	OPTIMAL % BOUND	TOTAL % BOUND
FOLDED TARGET				
BMEII0462m	2.07E-20	3.80E-19	2.65E-14	3.54E-13
BMEII0874m	9.12E-21	1.64E-19	2.92E-14	4.47E-13
BMEII0685m	6.98E-22	9.06E-21	7.29E-15	7.90E-14
BMEI0267m	1.89E-21	2.93E-20	1.24E-14	1.25E-13
BMEI0682m	1.45E-21	2.22E-20	1.08E-14	1.52E-13
PROBE_TARGET HYBRID				
BMEII0462m+BMEII0462m_5	19.564	19.7190802	80.9196	82.107225
BMEII0462m+BMEII0462m_4	0.018987	0.019176295	0.438108	0.448465512
BMEII0462m+BMEII0462m_3	0.0274487	0.027909505	0.0865046	0.088693192
BMEII0462m+BMEII0462m_2	4.52002	4.55484612	4.31936	4.37091557
BMEII0462m+BMEII0462m_1	0.251693	0.25352243	10.5307	10.637414
BMEII0462m+BMEII0462m_0	74.9272	75.4254789	2.32169	2.34729285
BMEII0874m+BMEII0874m_5	0.883334	0.89564002	1.96128	2.00549347
BMEII0874m+BMEII0874m_4_M	2.24325	2.26107444	53.9258	54.5661306
BMEII0874m+BMEII0874m_3	8.34825	8.35645052	4.56982	4.60907174
BMEII0874m+BMEII0874m_2	11.3954	11.6065101	13.3554	13.7763356
BMEII0874m+BMEII0874m_1_M	71.844	72.421896	22.3201	22.6603927
BMEII0874m+BMEII0874m_0	4.40572	4.45840168	2.35104	2.38254873
BMEII0685m+BMEII0685m_5	0.122565	0.123541075	1.02539	1.03530363
BMEII0685m+BMEII0685m_4	0.0292297	0.029382595	0.188874	0.191632659
BMEII0685m+BMEII0685m_3	0.146479	0.148096156	1.02539	1.05348417
BMEII0685m+BMEII0685m_2	1.51338	1.53300094	17.0234	17.3710061
BMEII0685m+BMEII0685m_1	18.2404	18.2404	37.9075	37.9075
BMEII0685m+BMEII0685m_0	79.4281	79.925586	42.1352	42.4409784
BMEI0267m+BMEI0267m_5	9.64143	9.7935098	38.693	39.5592966
BMEI0267m+BMEI0267m_4	0.0214752	0.021511263	2.55946	2.57954001
BMEI0267m+BMEI0267m_3	1.41801	1.43297265	23.1523	23.6287124
BMEI0267m+BMEI0267m_2	0.000317531	0.000320116	0.927508	0.94475352
BMEI0267m+BMEI0267m_1	0.224658	0.226567806	1.35306	1.37224212
BMEI0267m+BMEI0267m_0	87.4775	88.525081	31.3179	31.9154367
BMEI0682m+BMEI0682m_5	4.23936	4.30591375	2.87285	2.93663852
BMEI0682m+BMEI0682m_4	0.5429	0.54519089	4.8743	4.8948683
BMEI0682m+BMEI0682m_3	0.550866	0.562307317	0.284862	0.294423657
BMEI0682m+BMEI0682m_2	0.0455668	0.045915884	0.15805	0.160099186
BMEI0682m+BMEI0682m_1_M	93.3522	94.113117	89.9469	91.447021
BMEI0682m+BMEI0682m_0	0.424188	0.427594882	0.262946	0.266987082

TABLE 9: (Continued)

STRUCTURE	FORMAMIDE CONCENTRATION			
	10 %		15 %	
	OPTIMAL % BOUND	TOTAL % BOUND	OPTIMAL % BOUND	TOTAL % BOUND
FOLDED TARGET				
BMEII0462m	3.23E-07	3.92E-06	3.07396	9.507193
BMEII0874m	1.51E-07	1.18E-06	1.58E-08	1.99E-08
BMEII0685m	8.72E-08	6.12E-07	2.83E-10	1.04E-09
BMEI0267m	1.93E-07	1.03E-06	3.31E-10	1.45E-09
BMEI0682m	3.63E-24	7.87E-24	1.68E-10	1.12E-09
PROBE_TARGET HYBRID				
BMEII0462m+BMEII0462m_5	75.375	76.822731	66.423	68.6606877
BMEII0462m+BMEII0462m_4	0.849014	0.880632503	0.0375947	0.529835844
BMEII0462m+BMEII0462m_3	0.161182	0.16818895	0.0393375	0.494775423
BMEII0462m+BMEII0462m_2	4.40507	4.54665392	1.59222	9.682862
BMEII0462m+BMEII0462m_1	9.80917	9.9919698	8.51457	8.80826642
BMEII0462m+BMEII0462m_0	7.36195	7.5898005	6.39033	6.70516841
BMEII0874m+BMEII0874m_5	0.573005	0.599785222	0.55177	2.258239483
BMEII0874m+BMEII0874m_4_M	85.0179	86.7529122	77.0674	79.7492828
BMEII0874m+BMEII0874m_3	1.3151	1.33512347	0.121831	1.92550248
BMEII0874m+BMEII0874m_2	3.84341	4.01348069	1.61256	5.3228221
BMEII0874m+BMEII0874m_1_M	6.42327	6.60803964	0.568692	7.61093749
BMEII0874m+BMEII0874m_0	0.676579	0.69066758	0.622642	0.622667741
BMEII0685m+BMEII0685m_5	0.936011	0.95058786	0.903898	0.92430353
BMEII0685m+BMEII0685m_4	0.175034	0.179228532	0.166495	0.168018119
BMEII0685m+BMEII0685m_3	1.47259	1.53541042	1.20437	2.68897708
BMEII0685m+BMEII0685m_2	23.3645	24.0627306	22.2247	23.2798805
BMEII0685m+BMEII0685m_1	34.0844	34.2090793	32.9151	33.09415805
BMEII0685m+BMEII0685m_0	38.4624	39.0629481	37.1428	38.3270804
BMEI0267m+BMEI0267m_5	36.7597	38.092431	35.2511	36.7178413
BMEI0267m+BMEI0267m_4	5.82177	5.91289148	5.58285	5.70653367
BMEI0267m+BMEI0267m_3	22.3302	23.0154984	21.0928	21.9867278
BMEI0267m+BMEI0267m_2	0.881164	0.913974858	0.845002	0.859994276
BMEI0267m+BMEI0267m_1	1.30502	1.34197696	1.2327	1.27992201
BMEI0267m+BMEI0267m_0	29.7531	30.7232187	28.5321	29.7746347
BMEI0682m+BMEI0682m_5	2.84933	2.94063389	2.51422	2.580546256
BMEI0682m+BMEI0682m_4	4.83439	4.87396354	4.26582	4.29377893
BMEI0682m+BMEI0682m_3	0.28253	0.29608558	0.249302	0.253116157
BMEI0682m+BMEI0682m_2	0.232159	0.23810503	0.204855	0.204875025
BMEI0682m+BMEI0682m_1_M	89.2103	91.320828	78.7185	81.7350295
BMEI0682m+BMEI0682m_0	0.322695	0.330341237	0.165308	1.134186158

TABLE 10: Target Percent Bound on a Miniarray: DMSO Series

STRUCTURE	DMSO CONCENTRATION			
	0 %		2%	
	OPTIMAL % BOUND	TOTAL % BOUND	OPTIMAL % BOUND	TOTAL % BOUND
FOLDED TARGET				
BMEII0462m	2.07E-20	3.80E-19	7.80E-19	1.32E-17
BMEII0874m	9.12E-21	1.64E-19	3.33E-19	5.68E-18
BMEII0685m	6.98E-22	9.06E-21	3.69E-20	5.67E-19
BMEI0267m	1.89E-21	2.93E-20	1.18E-19	1.94E-18
BMEI0682m	1.45E-21	2.22E-20	4.30E-20	7.80E-19
PROBE_TARGET HYBRID				
BMEII0462m+BMEII0462m_5	19.564	19.7190802	54.3581	54.809203
BMEII0462m+BMEII0462m_4	0.018987	0.019176295	0.0619157	0.062793806
BMEII0462m+BMEII0462m_3	0.0274487	0.027909505	0.0468211	0.047668789
BMEII0462m+BMEII0462m_2	4.52002	4.55484612	4.79095	4.82959774
BMEII0462m+BMEII0462m_1	0.251693	0.25352243	0.765664	0.77175716
BMEII0462m+BMEII0462m_0	74.9272	75.4254789	39.2257	39.4789778
BMEII0874m+BMEII0874m_5	0.883334	0.89564002	1.53457	1.55714138
BMEII0874m+BMEII0874m_4_M	2.24325	2.26107444	3.93253	3.96713374
BMEII0874m+BMEII0874m_3	8.34825	8.35645052	5.06087	5.06972243
BMEII0874m+BMEII0874m_2	11.3954	11.6065101	22.6451	23.0946424
BMEII0874m+BMEII0874m_1_M	71.844	72.421896	59.0032	59.515925
BMEII0874m+BMEII0874m_0	4.40572	4.45840168	6.71269	6.7953932
BMEII0685m+BMEII0685m_5	0.122565	0.123541075	0.517288	0.521550713
BMEII0685m+BMEII0685m_4	0.0292297	0.029382595	0.0903762	0.090893786
BMEII0685m+BMEII0685m_3	0.146479	0.148096156	0.31187	0.316310357
BMEII0685m+BMEII0685m_2	1.51338	1.53300094	3.18345	3.23461324
BMEII0685m+BMEII0685m_1	18.2404	18.2404	27.8975	27.8975
BMEII0685m+BMEII0685m_0	79.4281	79.925586	67.5035	67.939115
BMEI0267m+BMEI0267m_5	9.64143	9.7935098	18.7403	19.0493848
BMEI0267m+BMEI0267m_4	0.0214752	0.021511263	0.465805	0.467455553
BMEI0267m+BMEI0267m_3	1.41801	1.43297265	0.967627	0.97825651
BMEI0267m+BMEI0267m_2	0.000317531	0.000320116	0.0197631	0.020012497
BMEI0267m+BMEI0267m_1	0.224658	0.226567806	3.29898	3.32823514
BMEI0267m+BMEI0267m_0	87.4775	88.525081	75.2131	76.15664
BMEI0682m+BMEI0682m_5	4.23936	4.30591375	2.88246	2.92997592
BMEI0682m+BMEI0682m_4	0.5429	0.54519089	3.89319	3.9096182
BMEI0682m+BMEI0682m_3	0.550866	0.562307317	0.285815	0.292299244
BMEI0682m+BMEI0682m_2	0.0455668	0.045915884	0.0491123	0.049590452
BMEI0682m+BMEI0682m_1_M	93.3522	94.113117	91.6212	92.585173
BMEI0682m+BMEI0682m_0	0.424188	0.427594882	0.231337	0.233293318

TABLE 10: (Continued)

STRUCTURE	DMSO CONCENTRATION			
	5 %		8 %	
	OPTIMAL % BOUND	TOTAL % BOUND	OPTIMAL % BOUND	TOTAL % BOUND
FOLDED TARGET				
BMEII0462m	8.10E-14	1.03E-12	1.46946	4.7777749
BMEII0874m	7.74E-14	1.05E-12	4.32E-09	5.45E-09
BMEII0685m	2.18E-14	2.11E-13	7.11E-11	2.28E-10
BMEI0267m	3.70E-14	3.51E-13	8.37E-11	3.75E-10
BMEI0682m	3.24E-14	4.16E-13	4.18E-11	2.80E-10
PROBE_TARGET HYBRID				
BMEII0462m+BMEII0462m_5	80.3901	81.594941	68.5056	70.7698463
BMEII0462m+BMEII0462m_4	0.522524	0.535183108	0.029543	0.417313633
BMEII0462m+BMEII0462m_3	0.111603	0.114490195	0.0318607	0.470084998
BMEII0462m+BMEII0462m_2	4.62774	4.68665132	1.49986	9.7007302
BMEII0462m+BMEII0462m_1	10.4618	10.56943	8.9152	9.21251415
BMEII0462m+BMEII0462m_0	2.46873	2.49929989	6.691	7.0227259
BMEII0874m+BMEII0874m_5	1.67479	1.71388726	0.525392	2.21686764
BMEII0874m+BMEII0874m_4_M	60.6189	61.3582095	79.1399	81.6880624
BMEII0874m+BMEII0874m_3	3.90228	3.9367202	0.0997434	1.840620359
BMEII0874m+BMEII0874m_2	11.2336	11.6006276	1.55884	5.3538532
BMEII0874m+BMEII0874m_1_M	19.0597	19.3558611	1.00592	7.1038903
BMEII0874m+BMEII0874m_0	2.00761	2.03467906	0.629801	0.638899187
BMEII0685m+BMEII0685m_5	0.997477	1.00725418	0.910426	0.937368902
BMEII0685m+BMEII0685m_4	0.183732	0.18646514	0.167697	0.171897163
BMEII0685m+BMEII0685m_3	1.16715	1.19984613	1.19488	2.68840138
BMEII0685m+BMEII0685m_2	19.5828	19.9966246	22.3852	23.4319315
BMEII0685m+BMEII0685m_1	36.3227	36.3227	33.1528	33.43737961
BMEII0685m+BMEII0685m_0	40.9881	41.2871527	37.411	38.6361379
BMEI0267m+BMEI0267m_5	38.1673	39.0813277	35.7743	37.4288861
BMEI0267m+BMEI0267m_4	3.45142	3.47899178	5.6657	5.80241668
BMEI0267m+BMEI0267m_3	23.1853	23.6635021	21.7316	22.6611606
BMEI0267m+BMEI0267m_2	0.914906	0.9324882	0.857543	0.893952815
BMEI0267m+BMEI0267m_1	1.33468	1.35405039	1.27003	1.33023683
BMEI0267m+BMEI0267m_0	30.8924	31.4896585	28.9555	30.1832473
BMEI0682m+BMEI0682m_5	2.87147	2.93611449	2.6522	2.758170432
BMEI0682m+BMEI0682m_4	4.87196	4.8925184	4.49994	4.554182789
BMEI0682m+BMEI0682m_3	0.280457	0.290159494	0.262984	0.271360764
BMEI0682m+BMEI0682m_2	0.175063	0.177377952	0.216097	0.21610755
BMEI0682m+BMEI0682m_1_M	89.9036	91.433202	83.0386	86.1681157
BMEI0682m+BMEI0682m_0	0.266417	0.270633321	0.155468	1.128968908

TABLE 11: OMP DE Simulation for the Noncompetitive Hybridization

STRUCTURE	OPTIMAL % BOUND	TOTAL % BOUND
PROBE SET #0		
BMEII0462m+BMEII0462m_0	99.3394	100
BMEII0874m+BMEII0874m_0	98.8184	100
BMEII0685m+BMEII0685m_0	99.3776	100
BMEI0267m+BMEI0267m_0	98.8166	99.999972
BMEI0682m+BMEI0682m_0	99.2032	99.9999539706
PROBE SET #1		
BMEII0462m+BMEII0462m_1	99.2784	100
BMEII0874m+BMEII0874m_1_M	99.202	99.999957
BMEII0685m+BMEII0685m_1	100	100
BMEI0267m+BMEI0267m_1	99.1571	100
BMEI0682m+BMEI0682m_1_M	99.1915	100
PROBE SET #2		
BMEII0462m+BMEII0462m_2	99.2354	99.9999959
BMEII0874m+BMEII0874m_2	98.1811	99.999993
BMEII0685m+BMEII0685m_2	98.7201	100
BMEI0267m+BMEI0267m_2	99.1924	99.999974
BMEI0682m+BMEI0682m_2	99.2397	99.99997
PROBE SET #3		
BMEII0462m+BMEII0462m_3	98.3489	99.999967
BMEII0874m+BMEII0874m_3	99.9019	100
BMEII0685m+BMEII0685m_3	98.908	99.9999644
BMEI0267m+BMEI0267m_3	98.9558	99.999964
BMEI0682m+BMEI0682m_3	97.9653	100
PROBE SET #4		
BMEII0462m+BMEII0462m_4	99.0129	100
BMEII0874m+BMEII0874m_4_M	99.2117	100
BMEII0685m+BMEII0685m_4	99.4796	99.999958
BMEI0267m+BMEI0267m_4	99.8324	100
BMEI0682m+BMEI0682m_4	99.5798	100
PROBE SET #5		
BMEII0462m+BMEII0462m_5	99.2136	100
BMEII0874m+BMEII0874m_5	98.626	99.9999915
BMEII0685m+BMEII0685m_5	99.2099	99.999982
BMEI0267m+BMEI0267m_5	98.4471	99.999963
BMEI0682m+BMEI0682m_5	98.4544	100

FIGURES

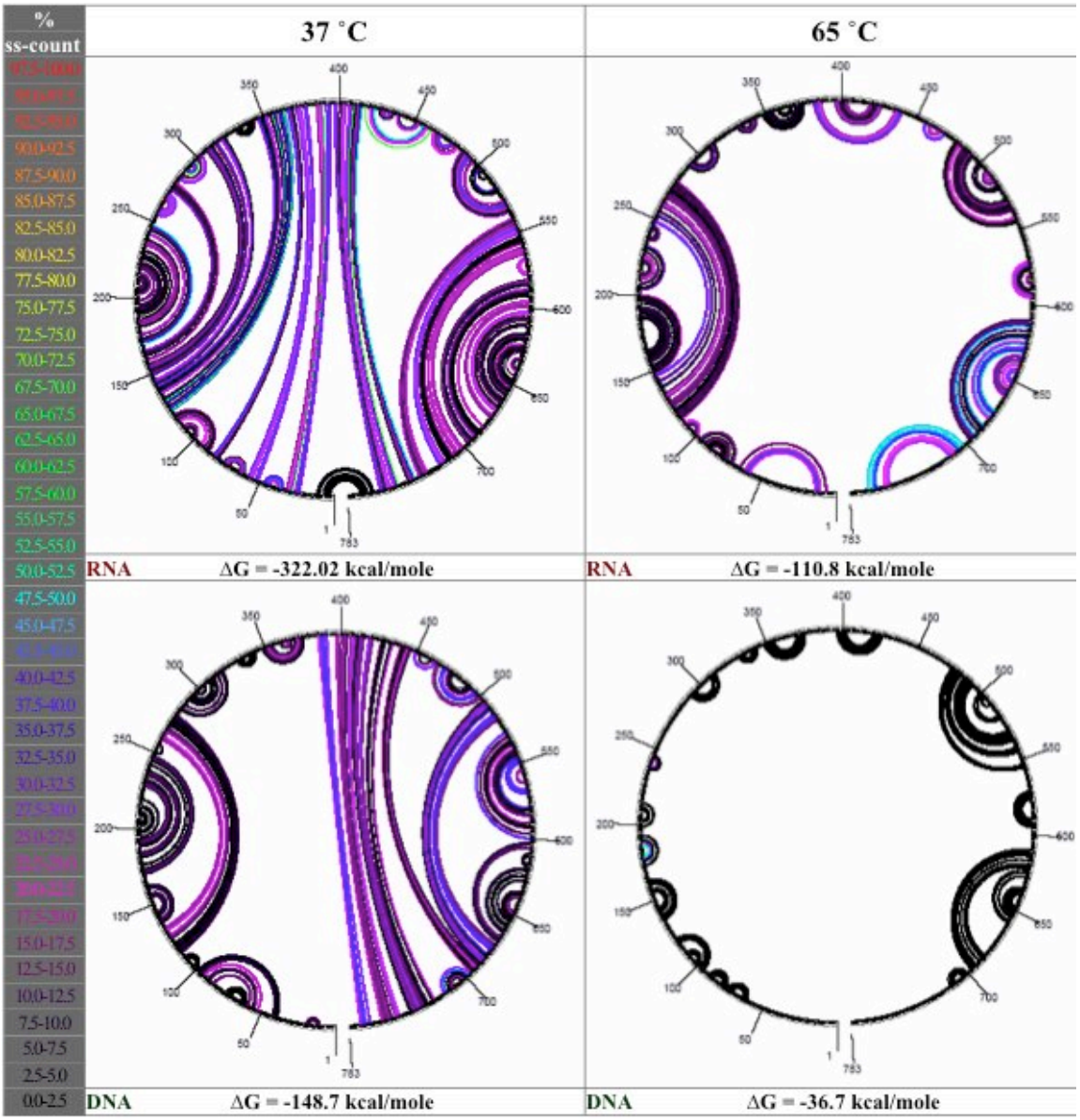

FIGURE 1: Secondary structure in a sample transcript

Circular diagrams of structure in a sample transcript (moeB homolog designated BR0004) from *Brucella suis* 1330. Circular diagrams show hydrogen bonds between individual nucleotides, color-coded according to single-strandedness – the fraction of structures in which that bond is not present. Black bonds indicate 0% single-strandedness; red bonds indicate 100% single-strandedness.

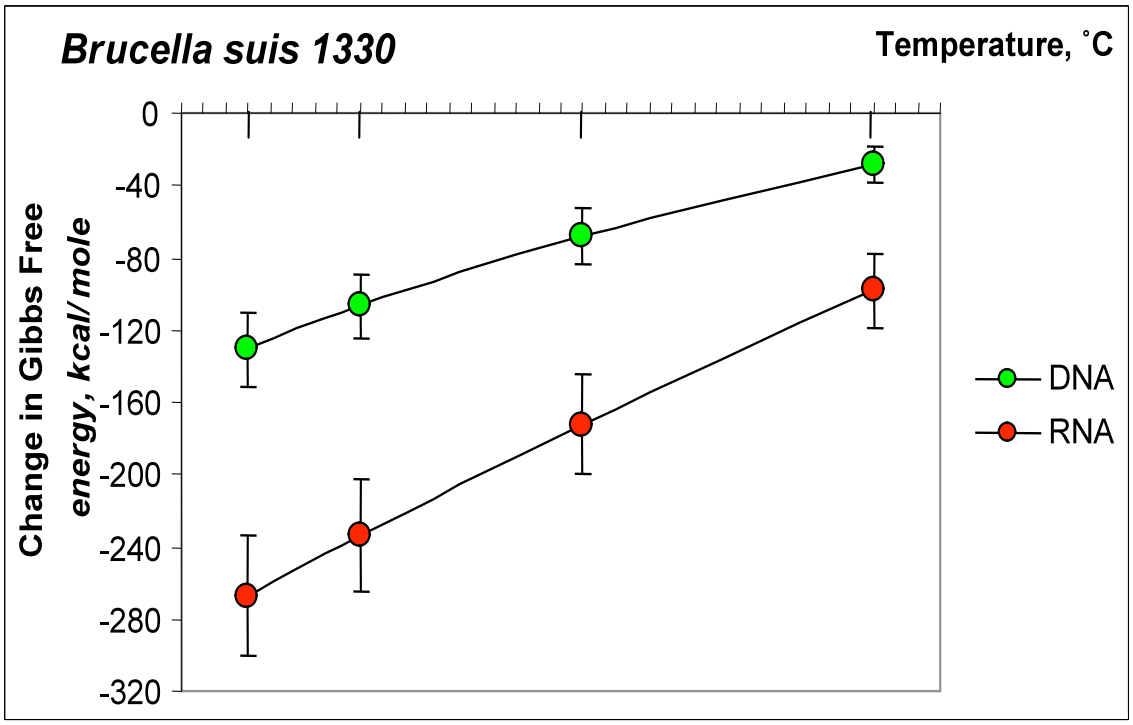

FIGURE 2: Stability of transcript secondary structure in *Brucella suis* 1330

Average free energy change on global secondary structure formation for *Brucella suis* 1330 targets, modeled as DNA or RNA. *ΔG* values are normalized to global mean target length.

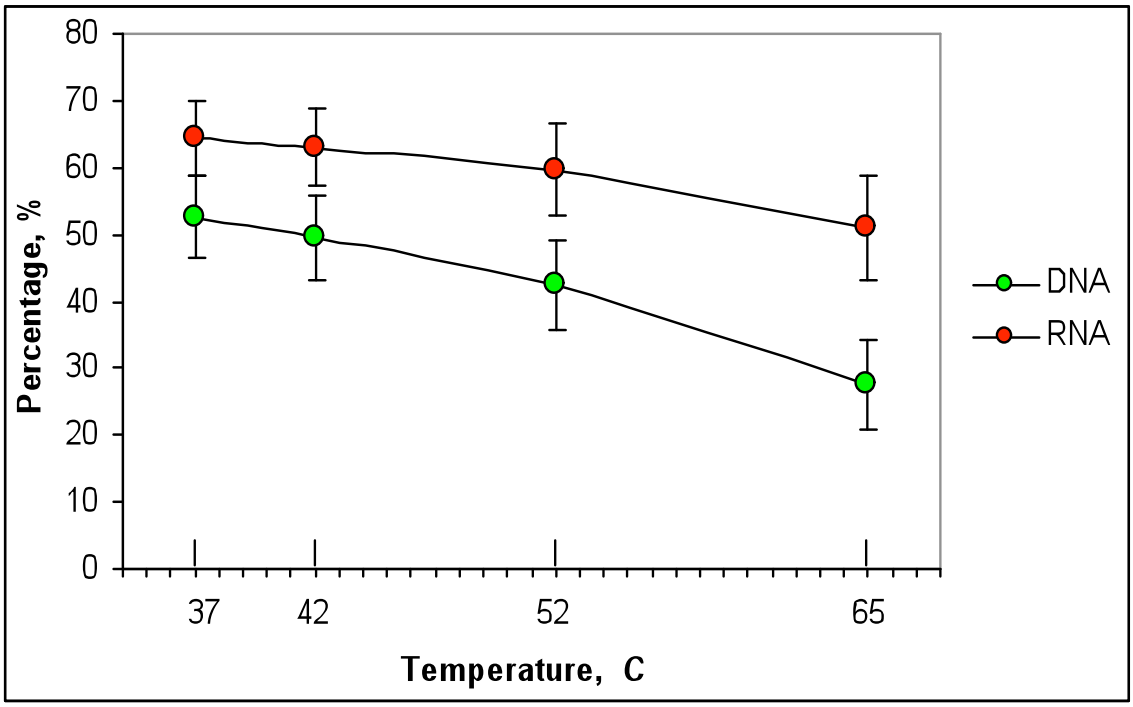

FIGURE 3: Fractional accessibility of nucleotides in the target

Fraction of the complete transcript classified as inaccessible due to the presence of stable structure in >50% of predicted conformations. Data shown are for 37, 42, 52 and 65°C simulations in *Brucella suis* 1330.

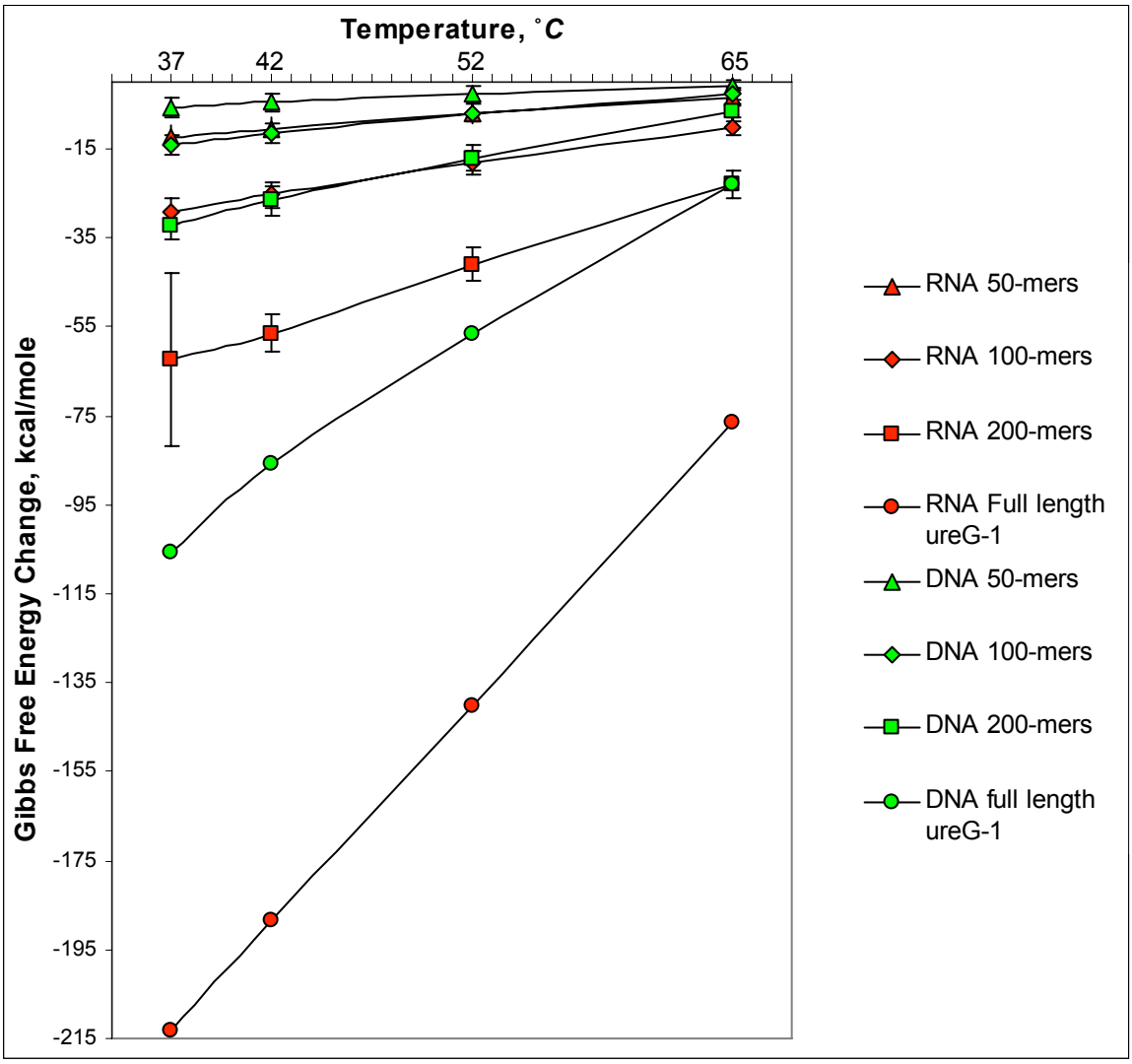

FIGURE 4: Stability of secondary structure in sheared fragments

Free energy change on secondary structure formation for the ureG-1 RNA transcript from *Brucella suis*. The transcript is modeled as sheared into fragments of length 200 nt, 100 nt or 50 nt; fragments are chosen starting at every 10th residue.

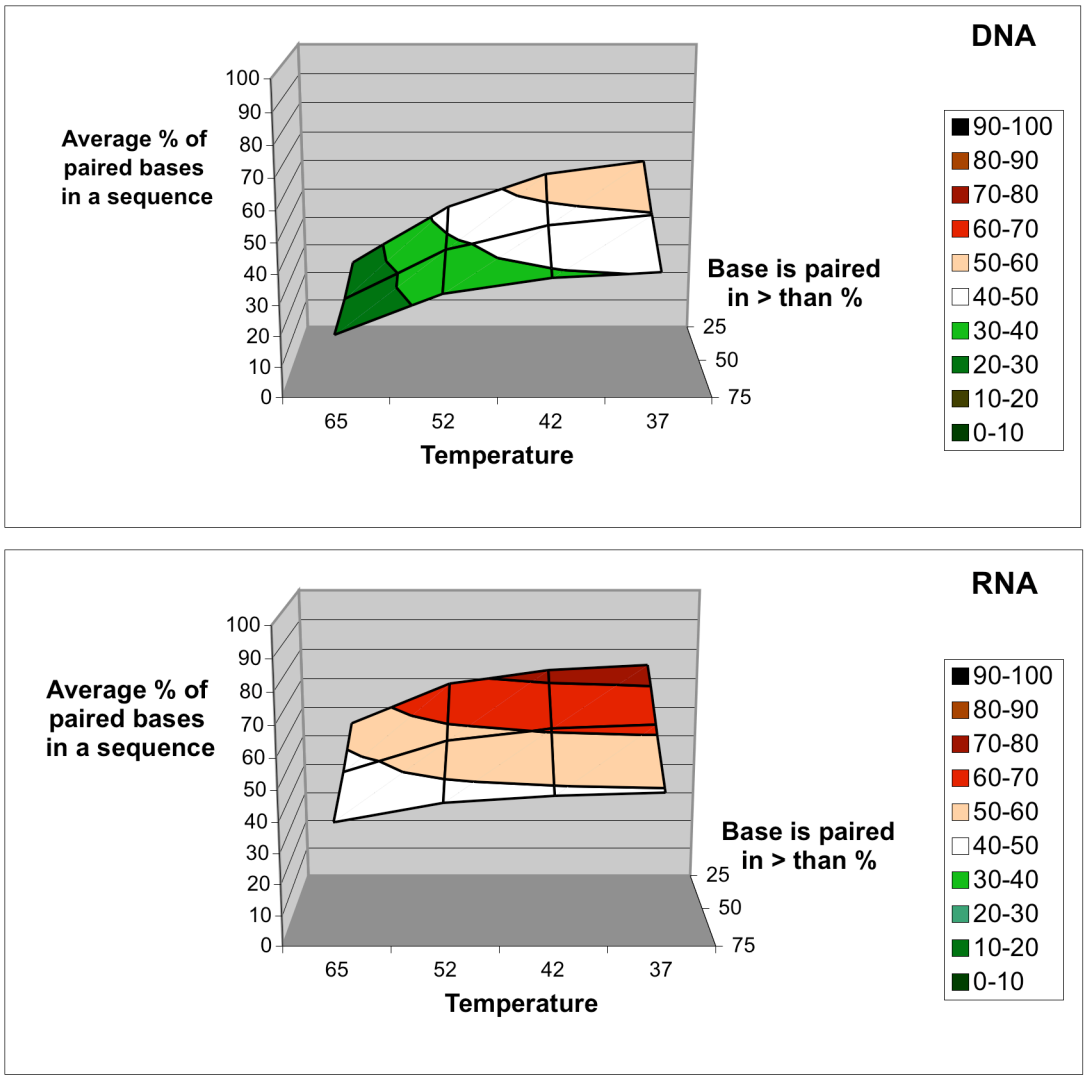

FIGURE 5: Accessibility of the probe-binding site

Fraction of the average probe binding site in the *Brucella* genomic array that is found to be inaccessible at 37°, 42°, 52° and 65°C, for DNA or RNA target. Inaccessible sites are defined here using three different cutoffs for the fraction of structures in which the site is base-paired: 25%, 50%, and 75%.

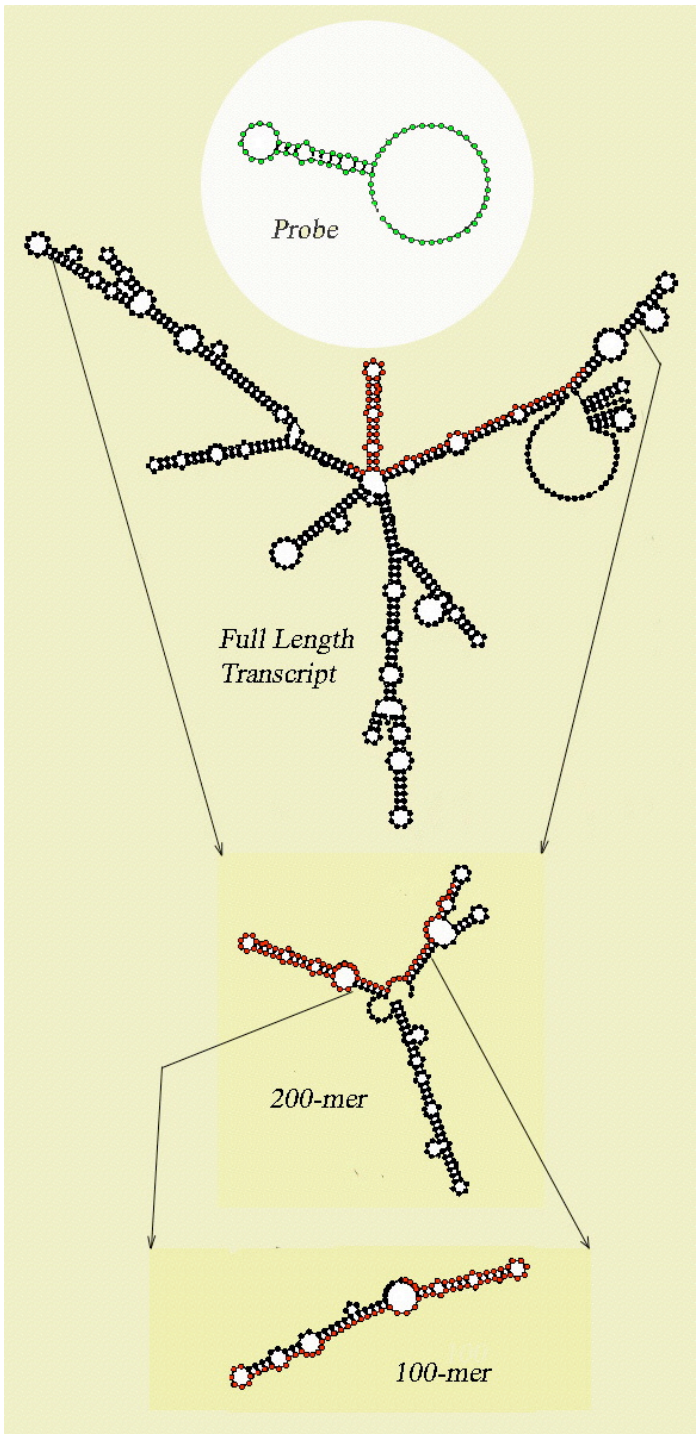

FIGURE 6: Structure in a binding site – full length target and sheared fragments

The position of a 70mer oligonucleotide probe (green) binding site (red dots) within a full-length optimal transcript structure, as well as examples of stable structure in 200mer & 100mer fragments which overlap the probe binding site. Corresponding ΔG values for these fragments modeled at 42° and 52°C are shown in TABLE 1.

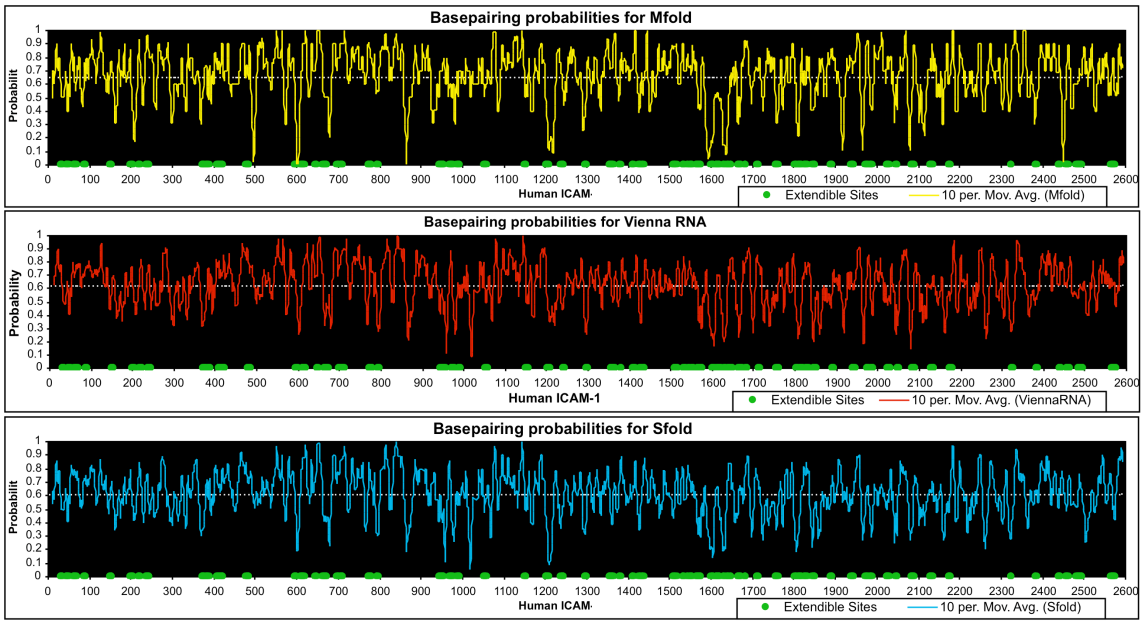

FIGURE 7: Accessibility prediction using three common methods

Pairing probabilities computed using RNAFold (top), MFold (middle) and SFold (bottom) for the human ICAM-1 transcript. Extendable sites detected by Allawi et al [21].

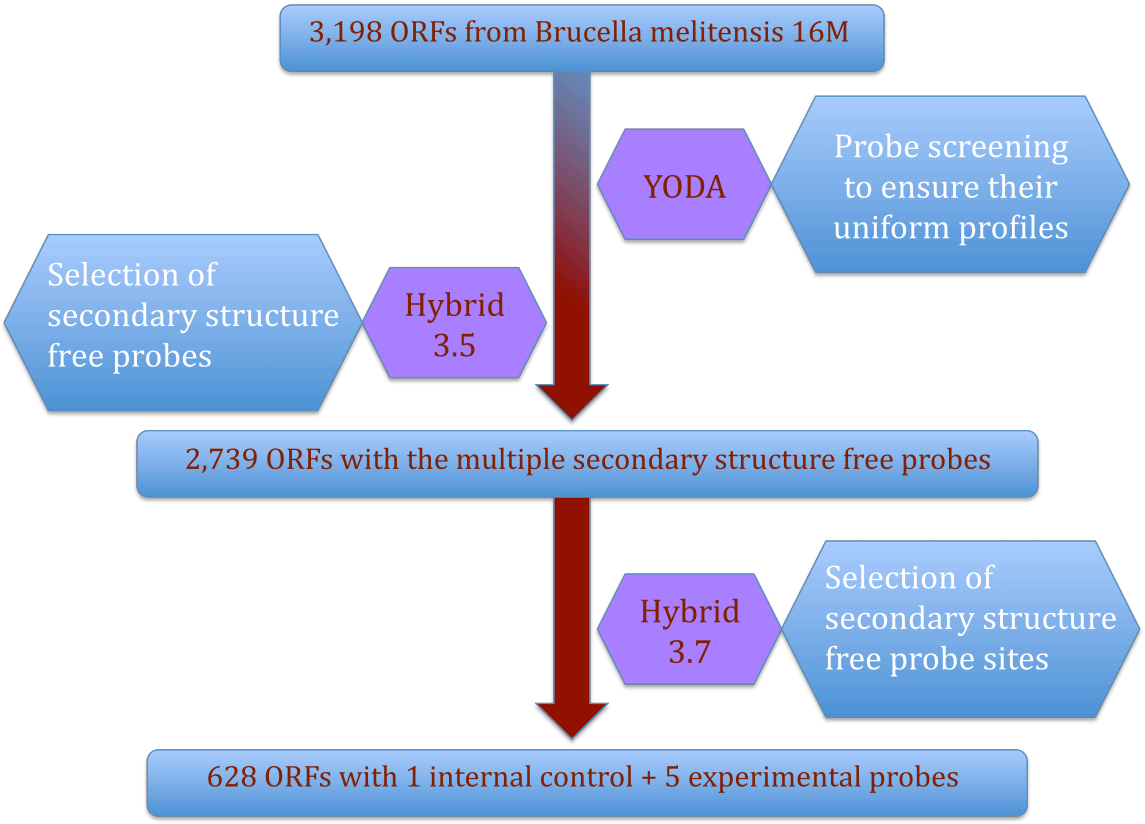

FIGURE 8: Flowchart of the target secondary structure microarray design

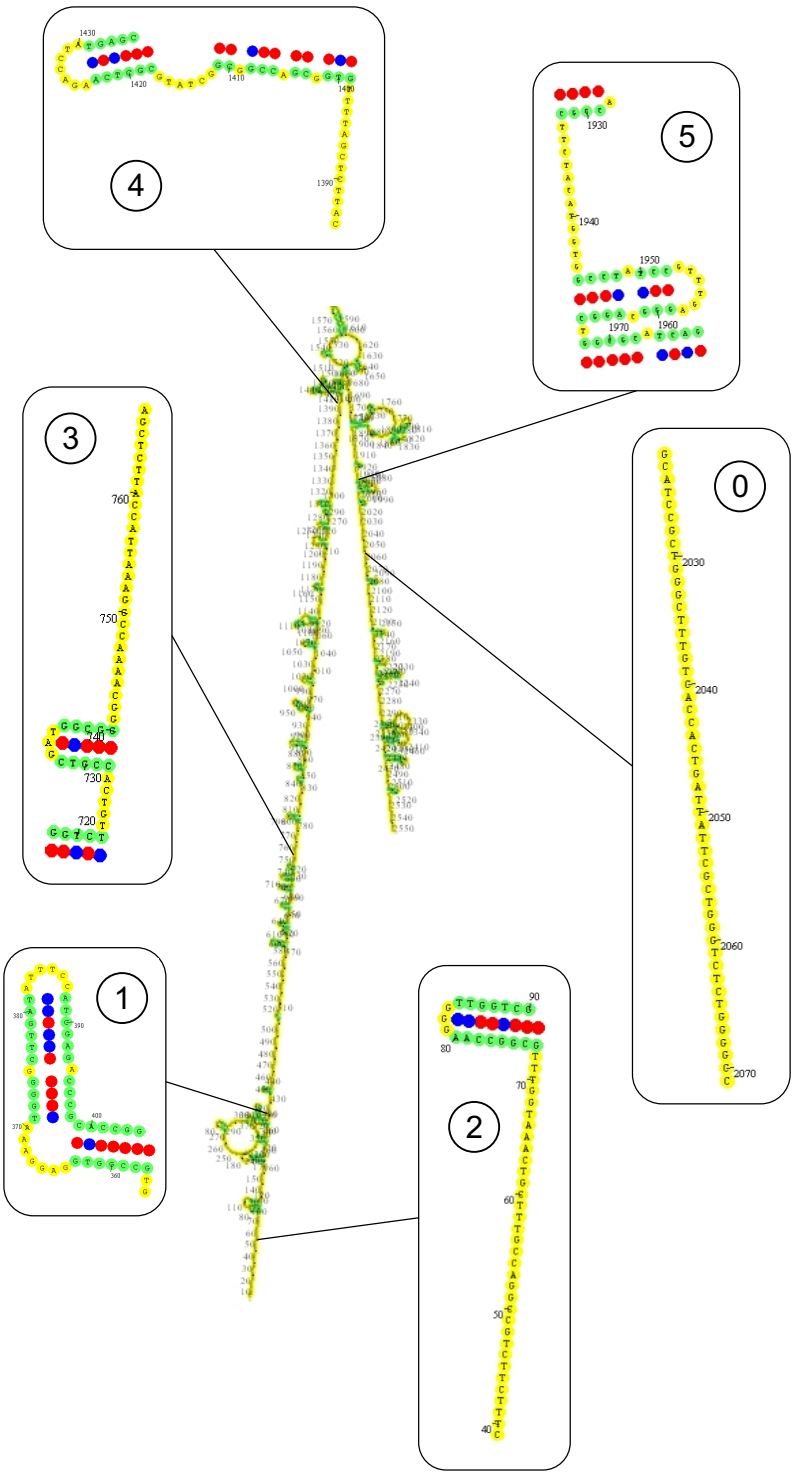

FIGURE 9: Secondary Structure on BMEII0462m and its binding sites at 60 °C

Bound nucleotides are drawn in bright green color, while all other bases are shown in yellow. The CG bonds are represented by red filled red circles, and the AT bonds and some non-Watson-Crick interactions are shown in blue.

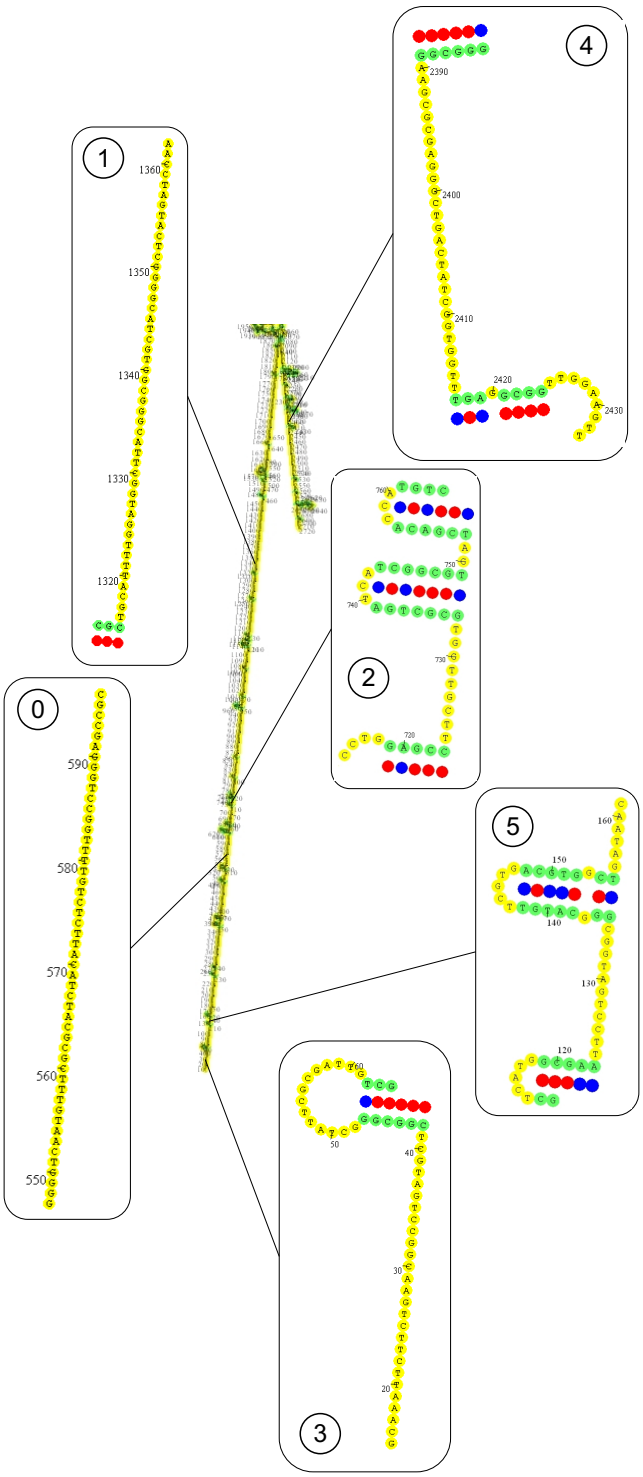

FIGURE 10: Secondary Structure on BMEII0874m and its binding sites at 60 °C

Bound nucleotides are drawn in bright green color, while all other bases are shown in yellow. The CG bonds are represented by red filled red circles, and the AT bonds and some non-Watson-Crick interactions are shown in blue.

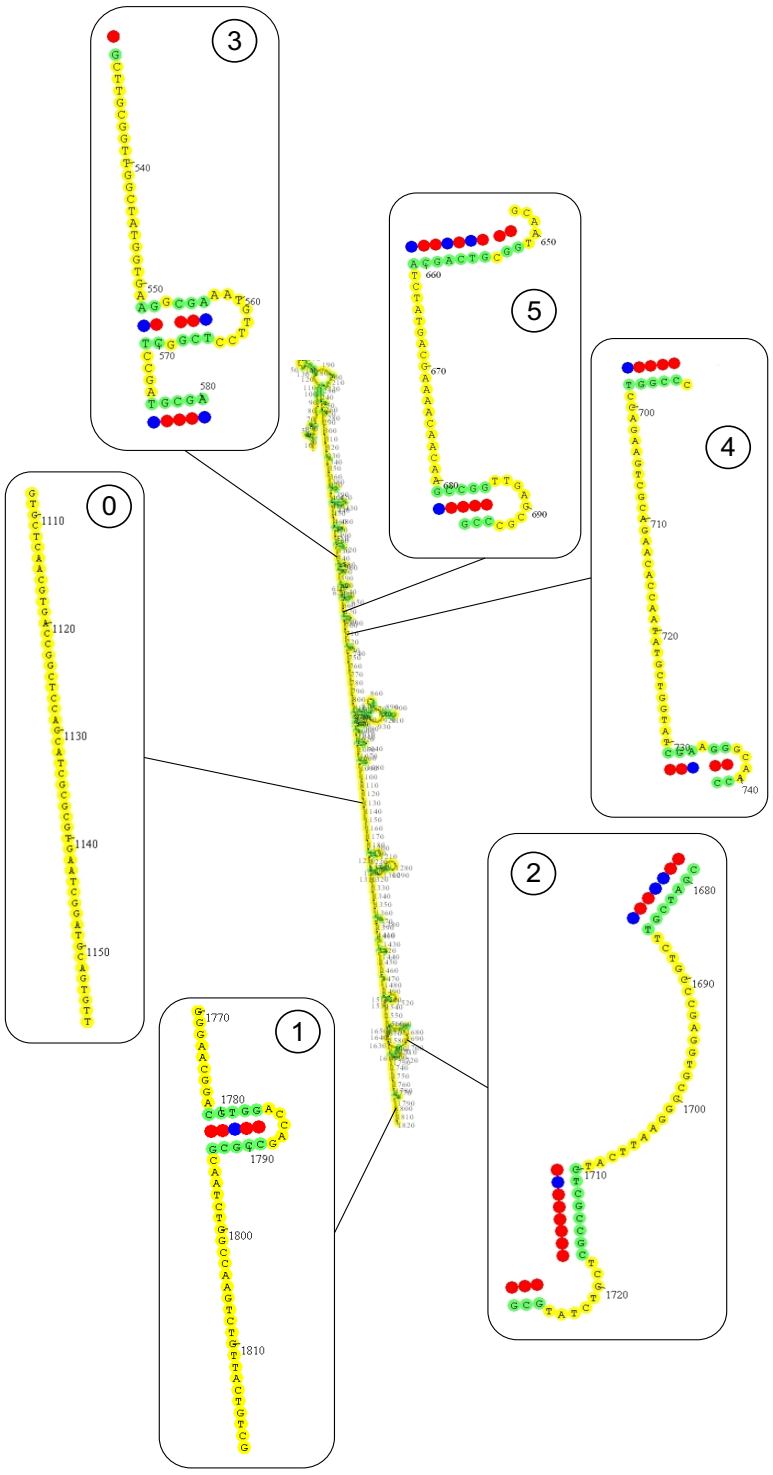

FIGURE 11: Secondary Structure on BMEII0685m and its binding sites at 60 °C

Bound nucleotides are drawn in bright green color, while all other bases are shown in yellow. The CG bonds are represented by red filled red circles, and the AT bonds and some non-Watson-Crick interactions are shown in blue.

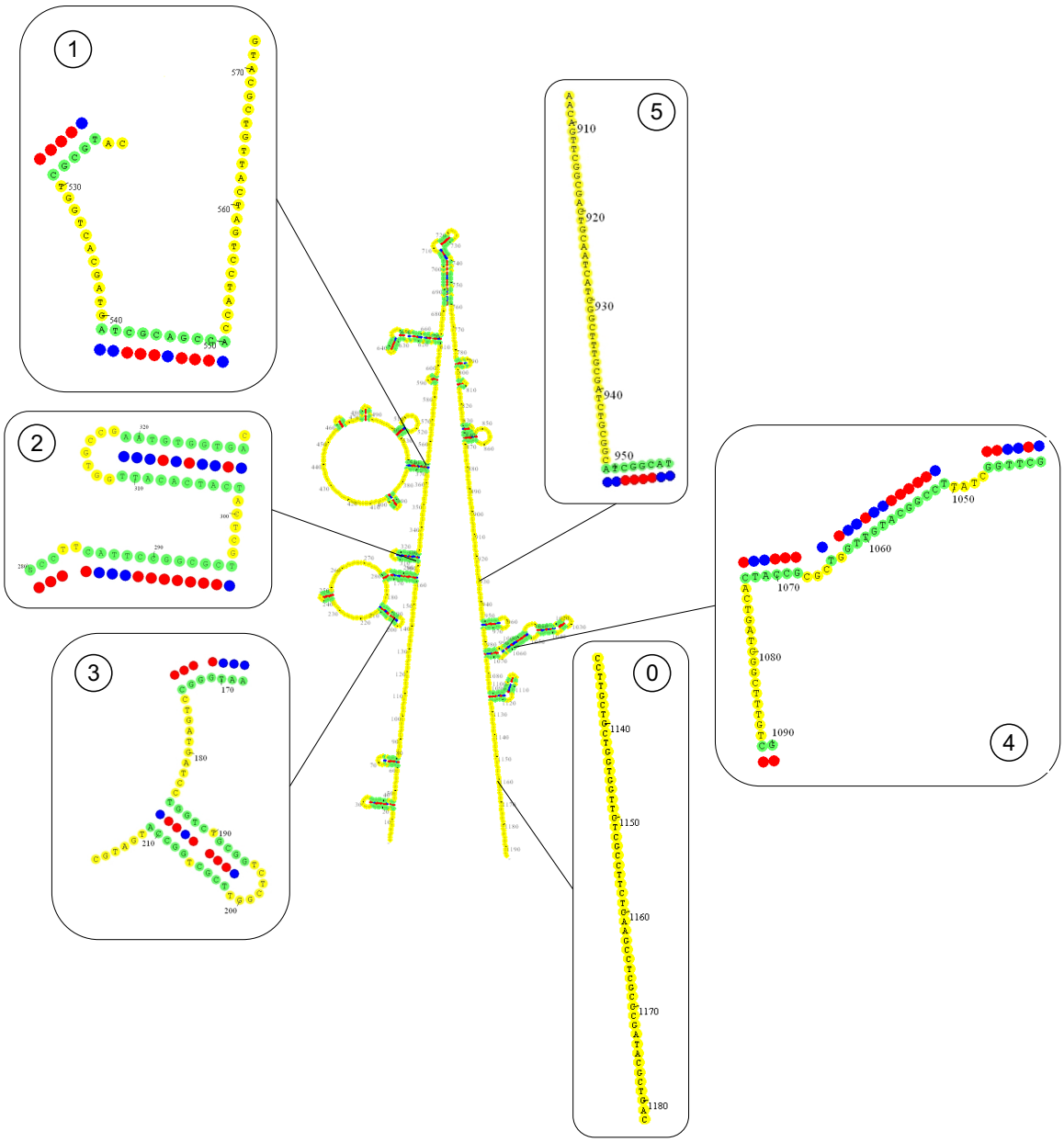

FIGURE 12: Secondary Structure on BMEI0267m and its binding sites at 60 °C

Bound nucleotides are drawn in bright green color, while all other bases are shown in yellow. The CG bonds are represented by red filled red circles, and the AT bonds and some non-Watson-Crick interactions are shown in blue.

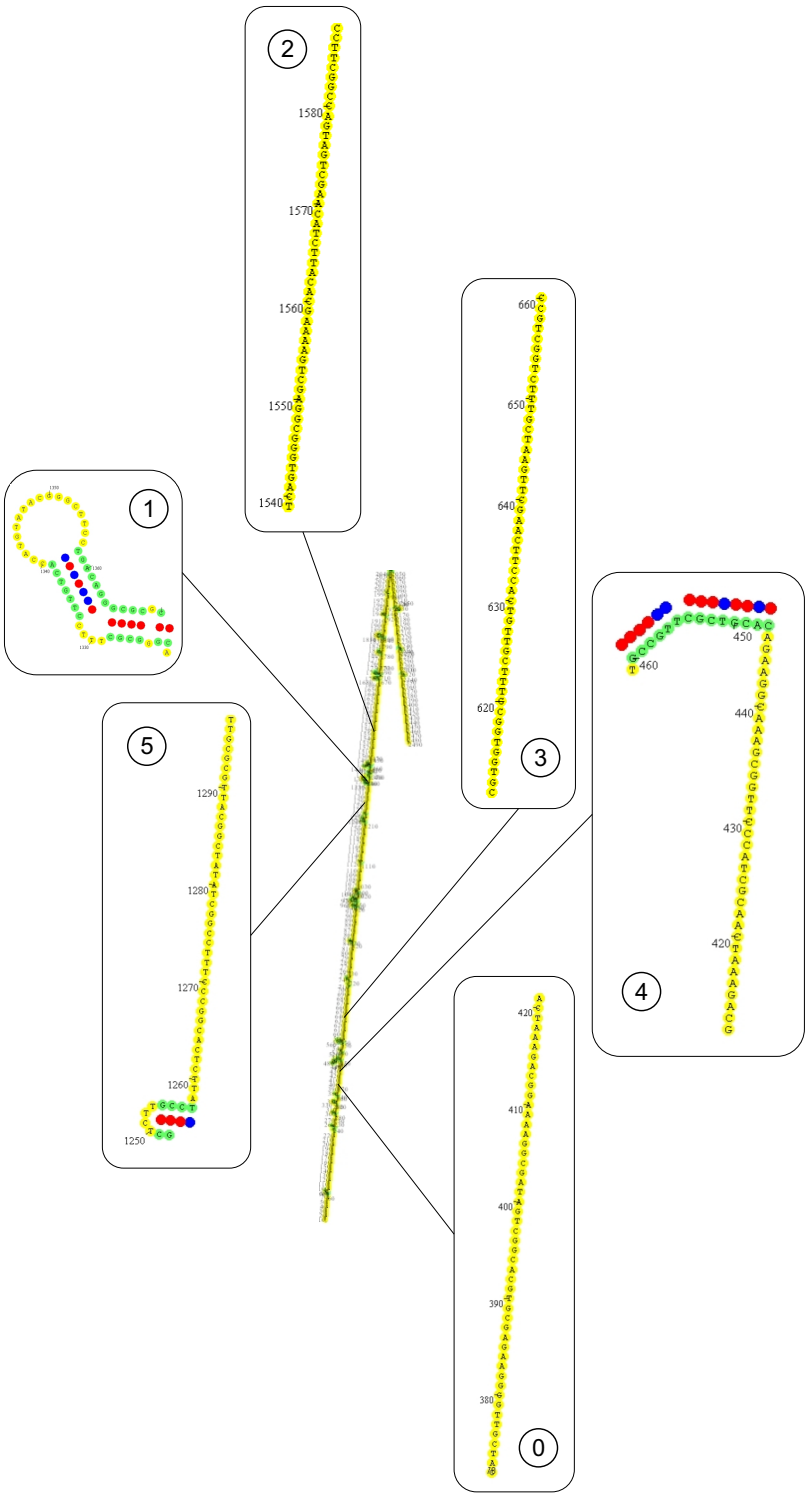

FIGURE 13: Secondary Structure on BMEI0682m and its binding sites at 60 °C

Bound nucleotides are drawn in bright green color, while all other bases are shown in yellow. The CG bonds are represented by red filled red circles, and the AT bonds and some non-Watson-Crick interactions are shown in blue.

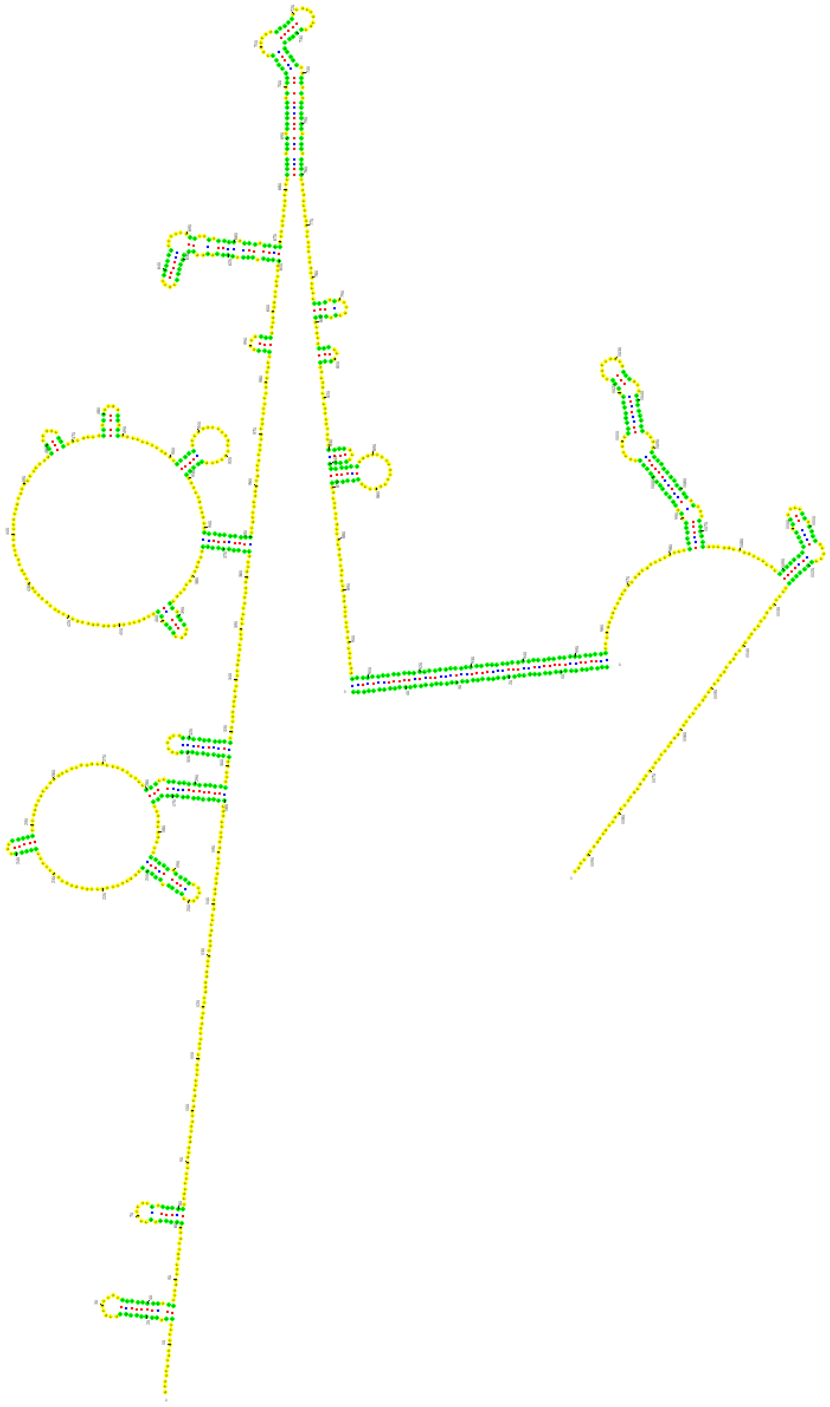

FIGURE 14: Secondary Structure on BMEI0267m – BMEO267m_5 Heterodimer at 60 °C and No Structure Destabilizing Additives Conditions

Bound nucleotides are drawn in bright green color, while all other bases are shown in yellow. The CG bonds are represented by red filled red circles, and the AT bonds and some non-Watson-Crick interactions are shown in blue.

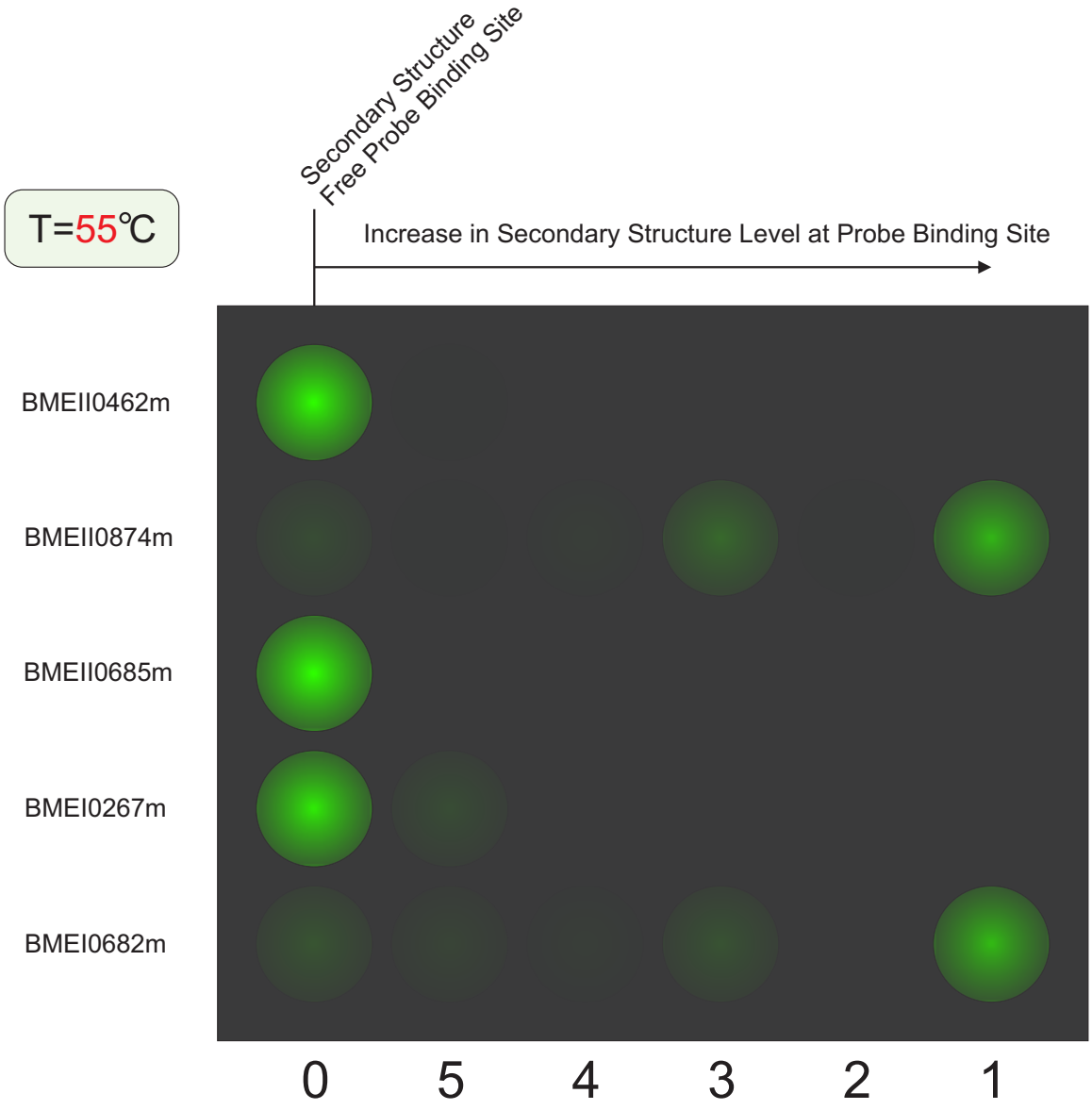

FIGURE 15: Graphical representation of computational simulation for the miniarray hybridization at 55 °C, no additives

The numbers 0, 5, 4, 3, 2, and 1 represent the probe numbers on the miniarray and are placed in the order of increasing secondary structure abundance in the probe-binding site.

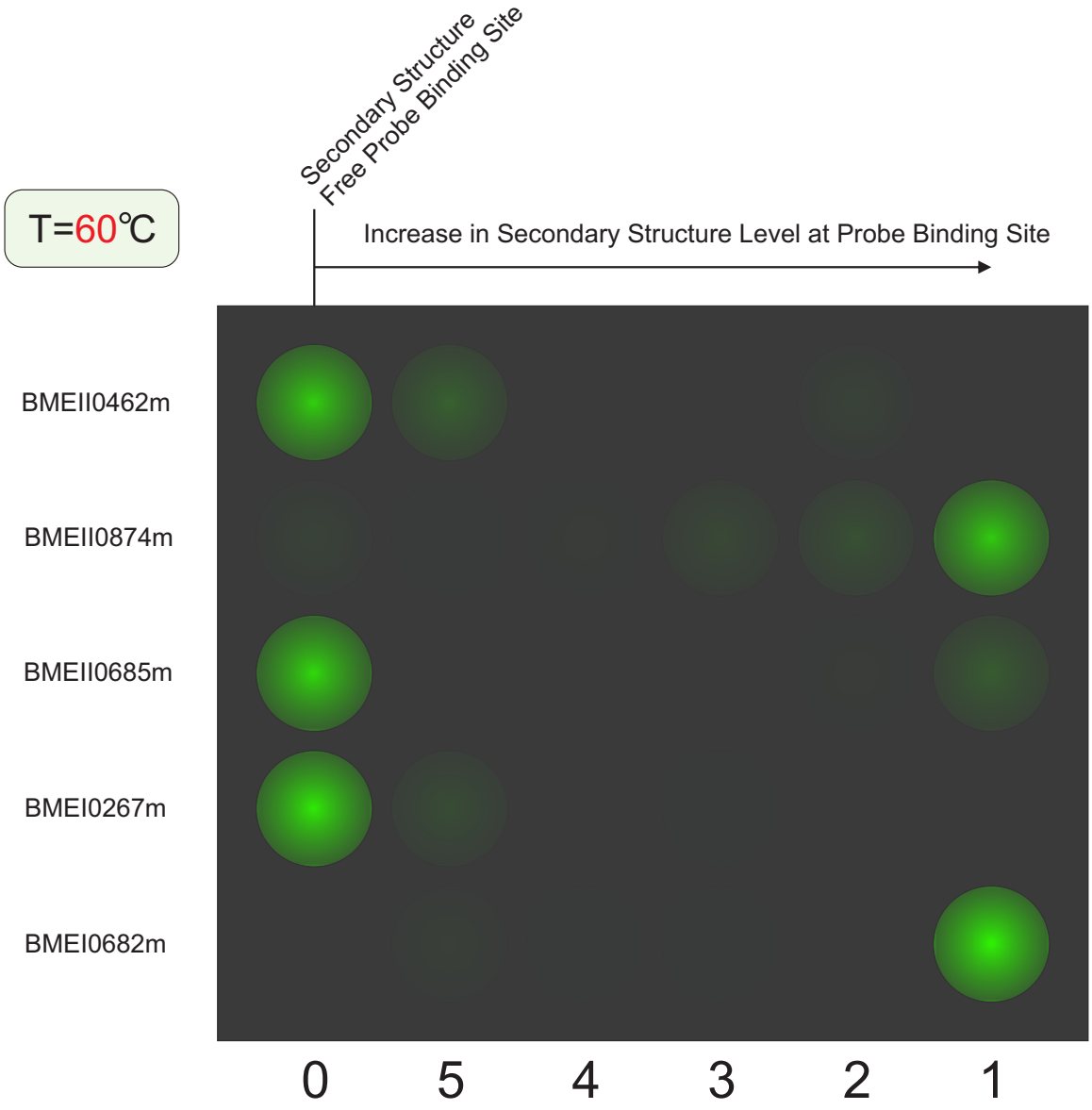

FIGURE 16: Graphical representation of computational simulation for the miniarray hybridization at 60 °C, no additives

The numbers 0, 5, 4, 3, 2, and 1 represent the probe numbers on the miniarray and are placed in the order of increasing secondary structure abundance in the probe-binding site.

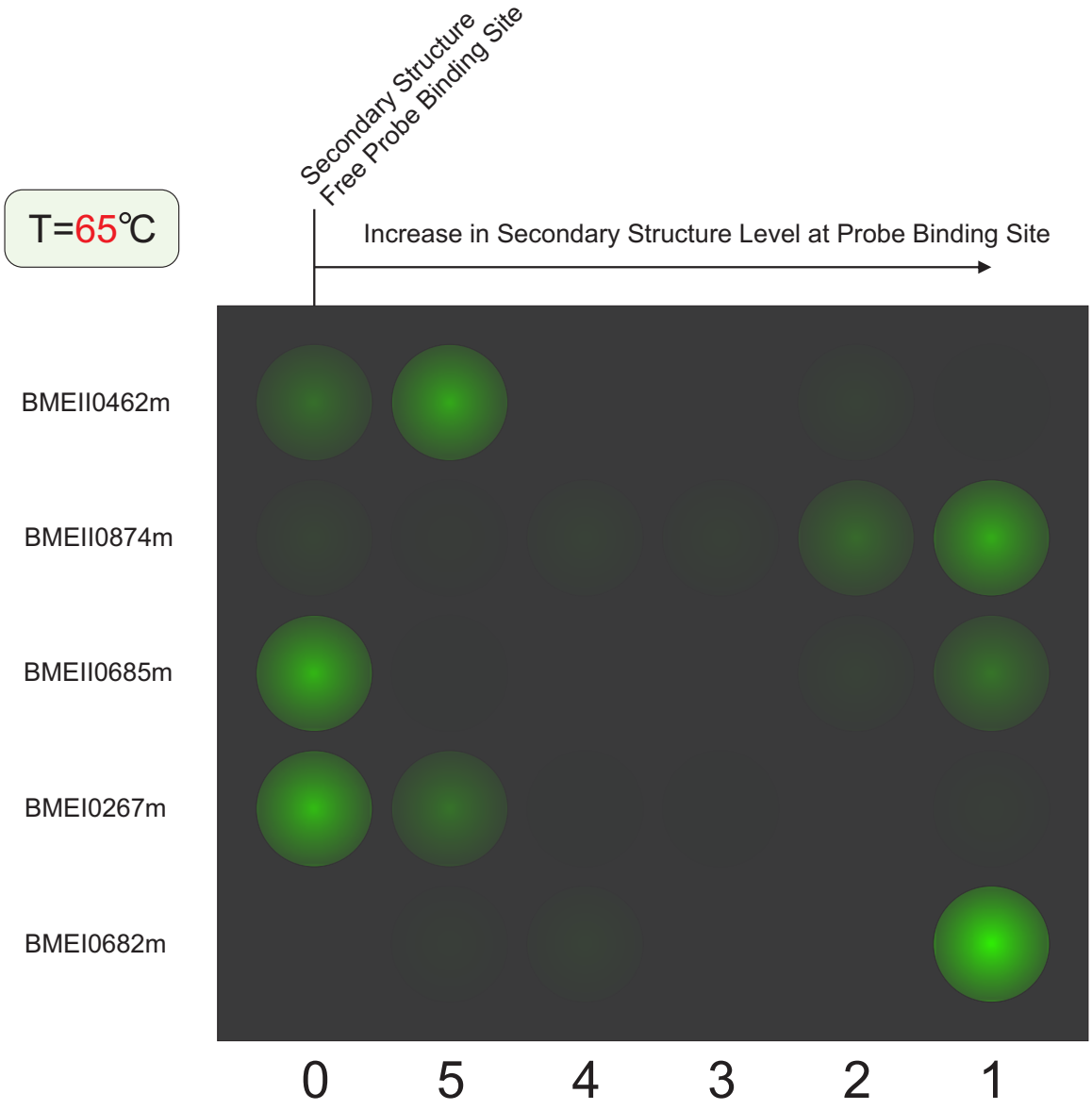

FIGURE 17: Graphical representation of computational simulation for the miniarray hybridization at 65 °C, no additives

The numbers 0, 5, 4, 3, 2, and 1 represent the probe numbers on the miniarray and are placed in the order of increasing secondary structure abundance in the probe-binding site.

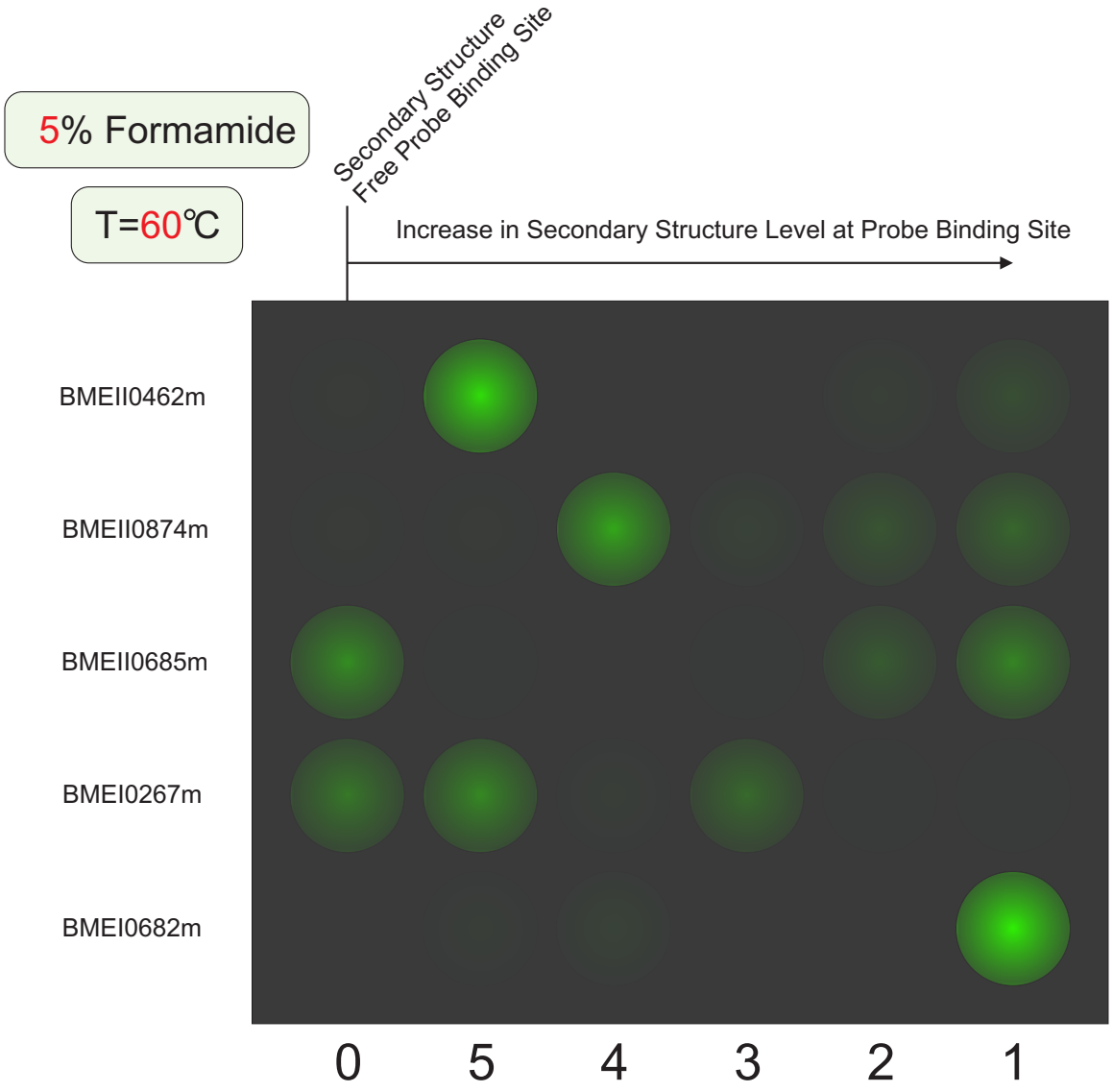

FIGURE 18: Graphical representation of computational simulation for the miniarray hybridization at 60 °C and 5% formamide

The numbers 0, 5, 4, 3, 2, and 1 represent the probe numbers on the miniarray and are placed in the order of increasing secondary structure abundance in the probe-binding site.

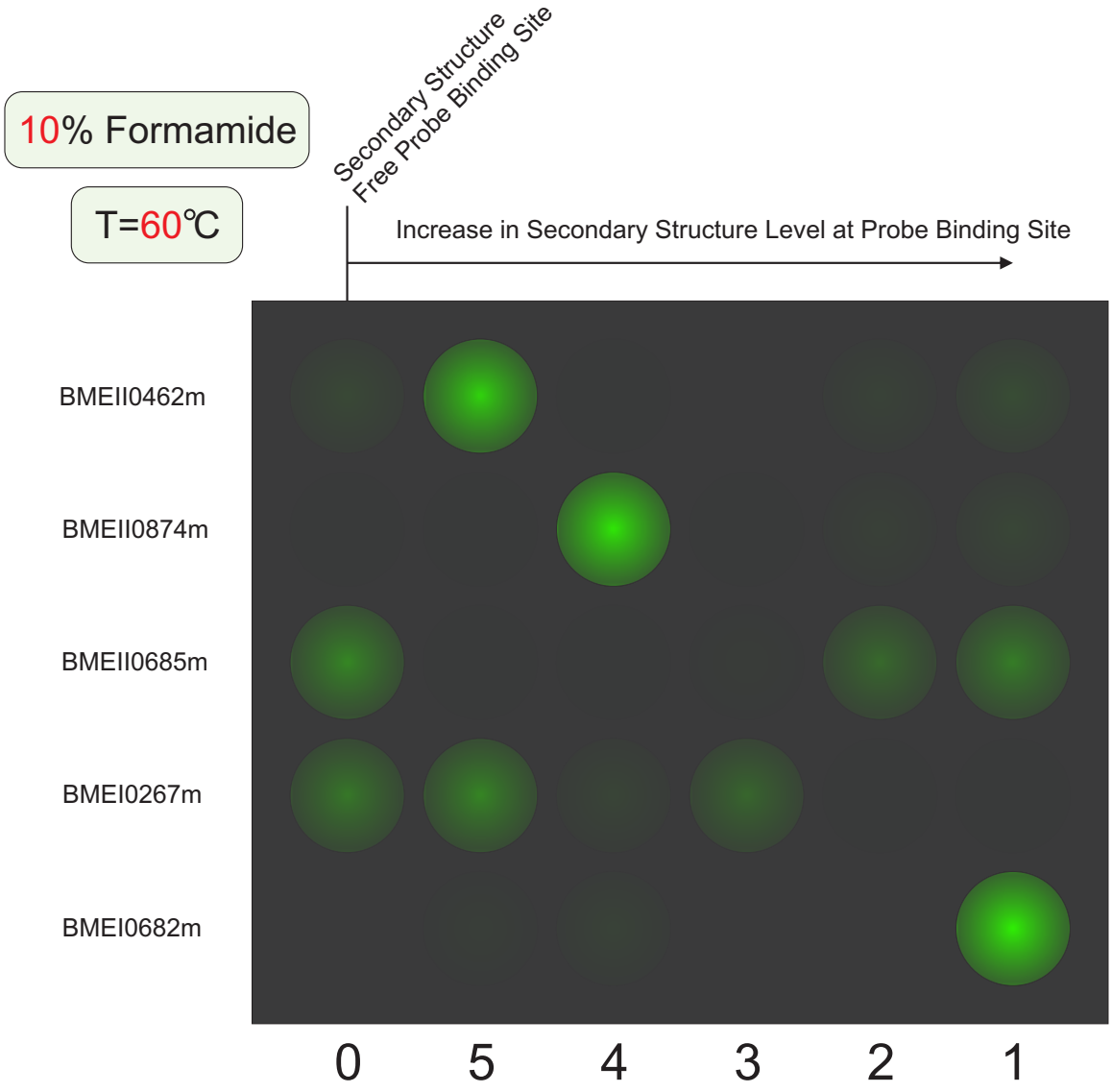

FIGURE 19: Graphical representation of computational simulation for the miniarray hybridization at 60 °C and 10% formamide

The numbers 0, 5, 4, 3, 2, and 1 represent the probe numbers on the miniarray and are placed in the order of increasing secondary structure abundance in the probe-binding site.

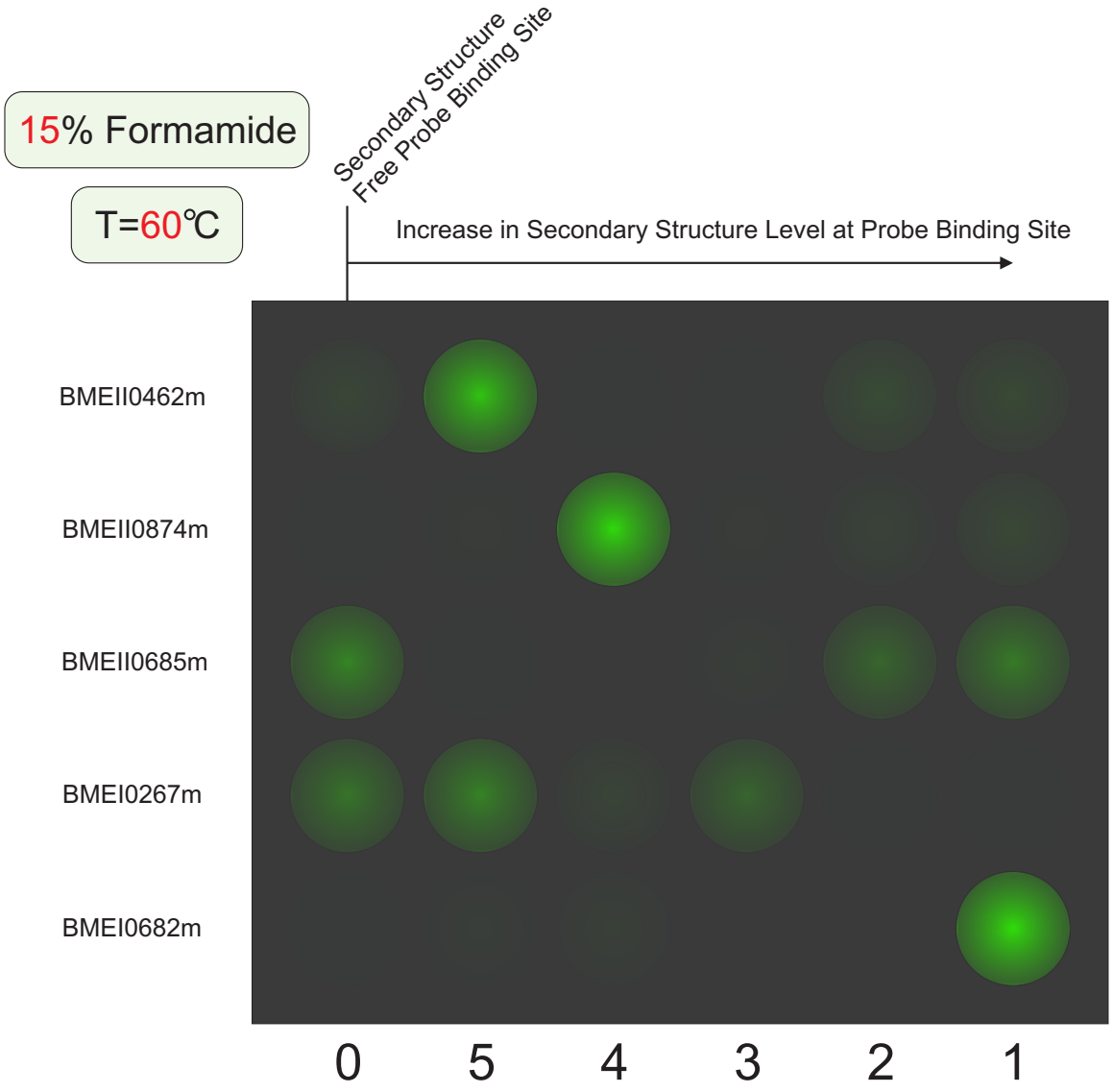

FIGURE 20: Graphical representation of computational simulation for the miniarray hybridization at 60 °C and 15% formamide

The numbers 0, 5, 4, 3, 2, and 1 represent the probe numbers on the miniarray and are placed in the order of increasing secondary structure abundance in the probe-binding site.

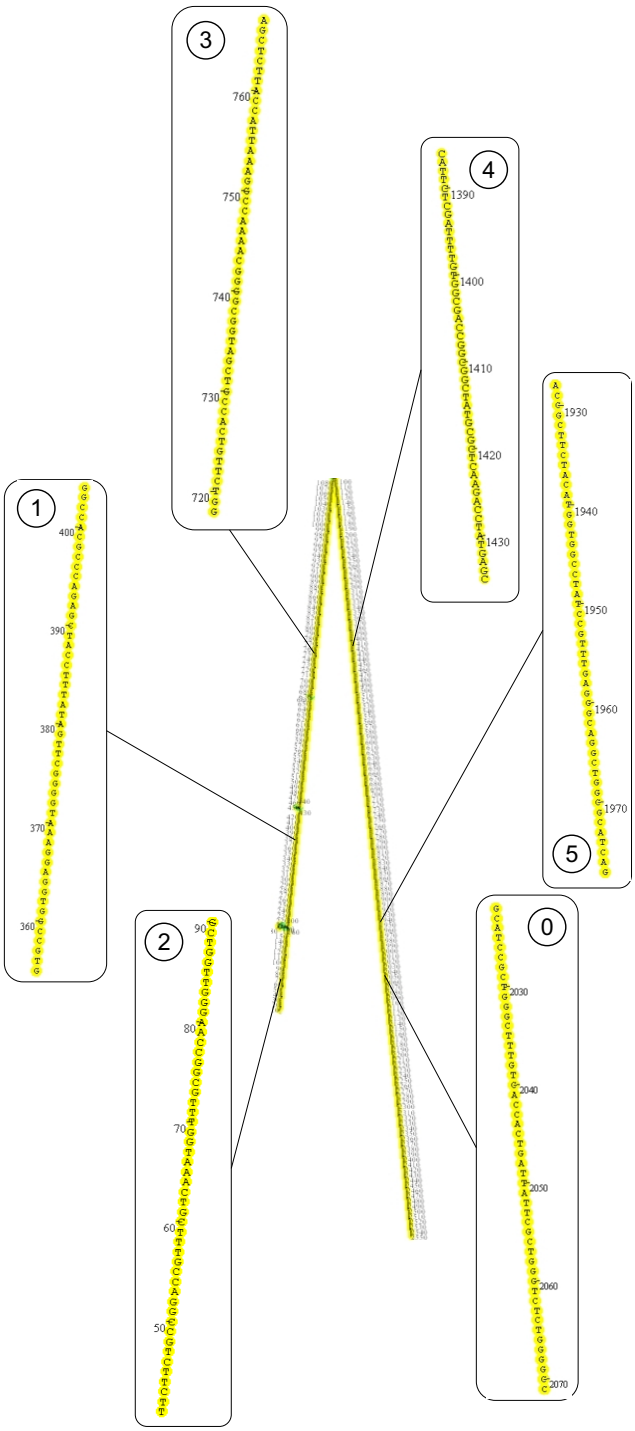

FIGURE 21: Relaxed Secondary Structure on BMEII0462m and Its Binding Sites at 10% Formamide at 60 °C

Bound nucleotides are drawn in bright green color, while all other bases are shown in yellow. The CG bonds are represented by red filled red circles, and the AT bonds and some non-Watson-Crick interactions are shown in blue.

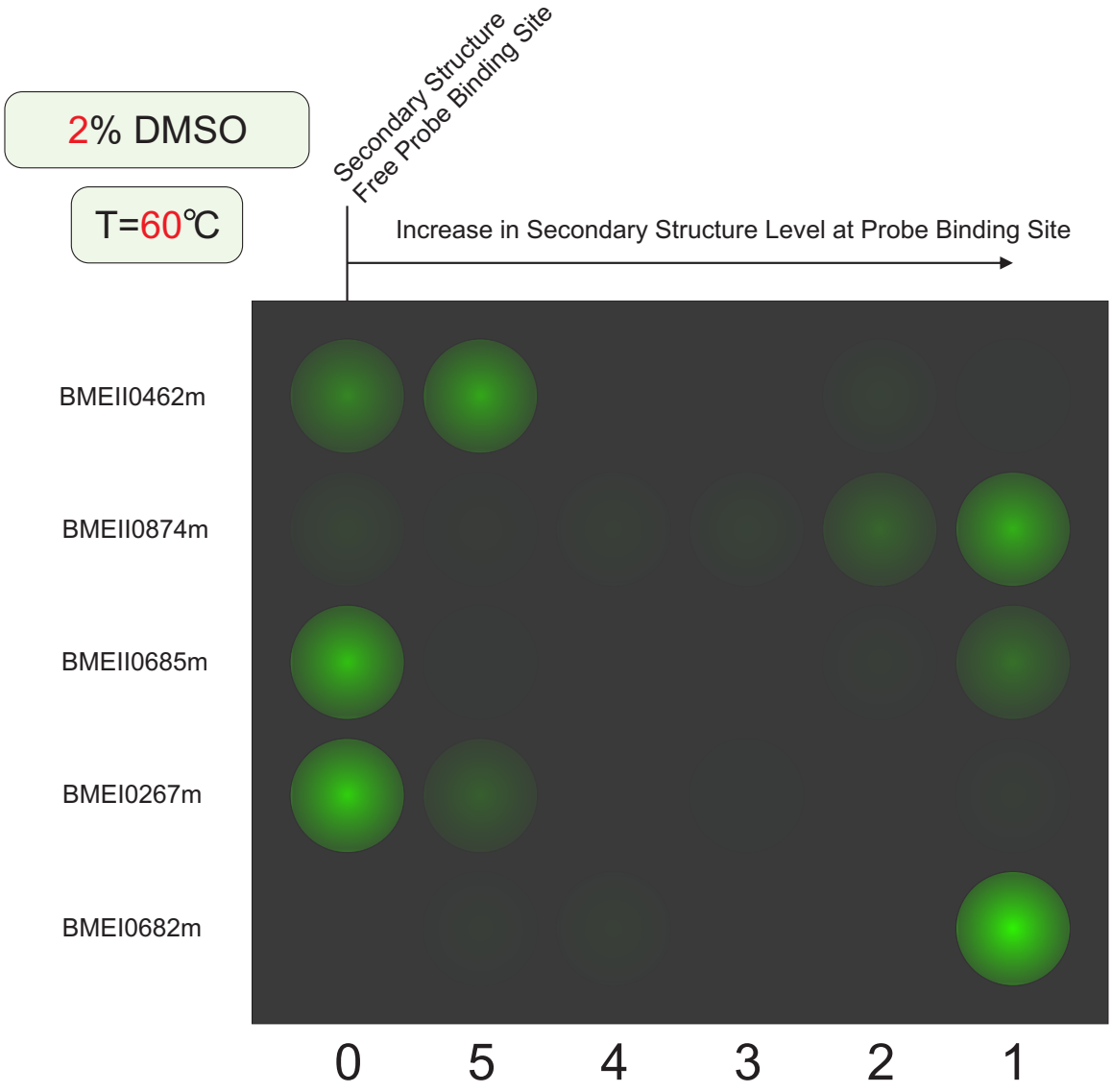

FIGURE 22: Graphical representation of computational simulation for the miniarray hybridization at 60 °C and 2% DMSO

The numbers 0, 5, 4, 3, 2, and 1 represent the probe numbers on the miniarray and are placed in the order of increasing secondary structure abundance in the probe-binding site.

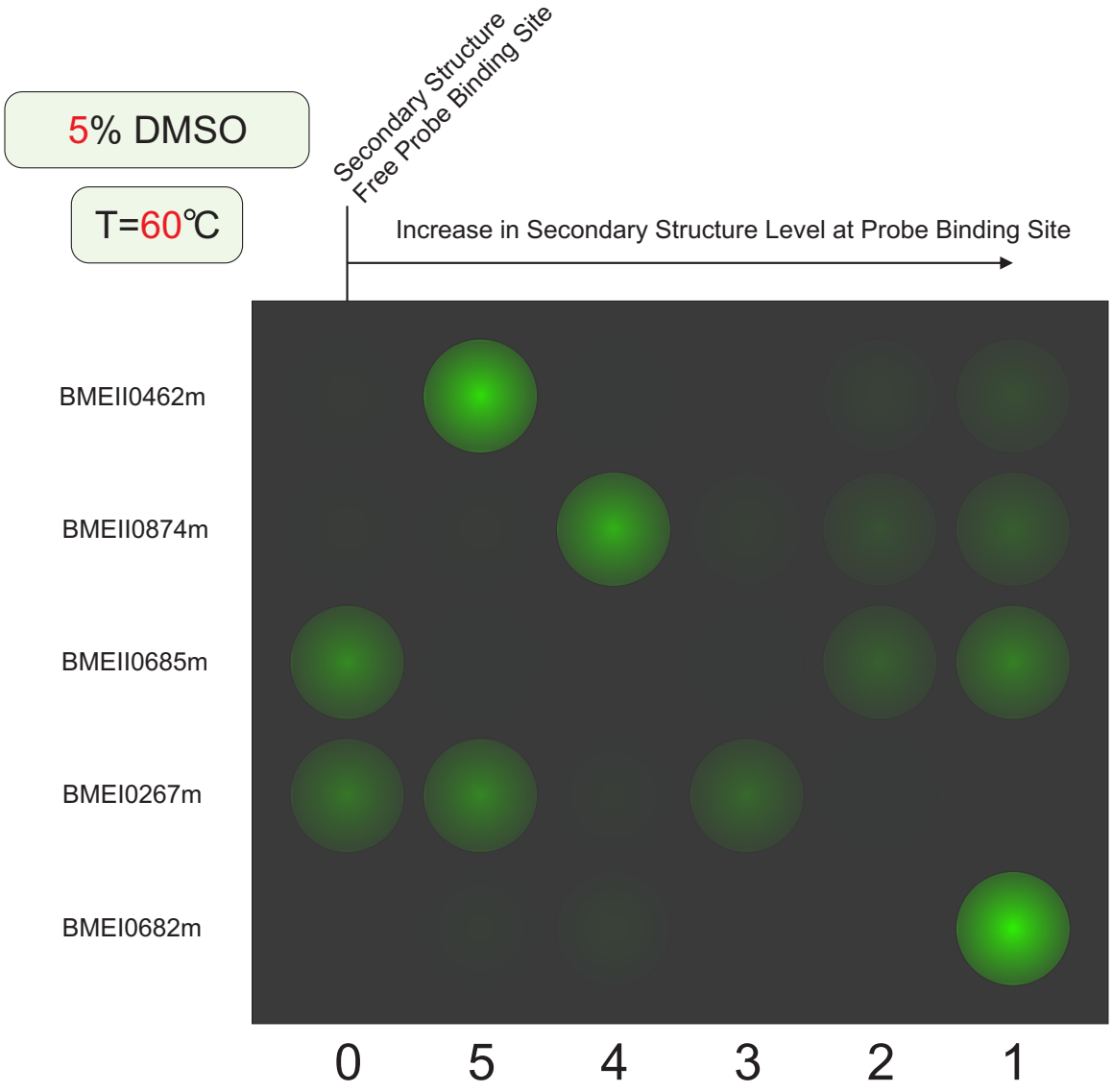

FIGURE 23: Graphical representation of computational simulation for the miniarray hybridization at 60 °C and 5% DMSO

The numbers 0, 5, 4, 3, 2, and 1 represent the probe numbers on the miniarray and are placed in the order of increasing secondary structure abundance in the probe-binding site.

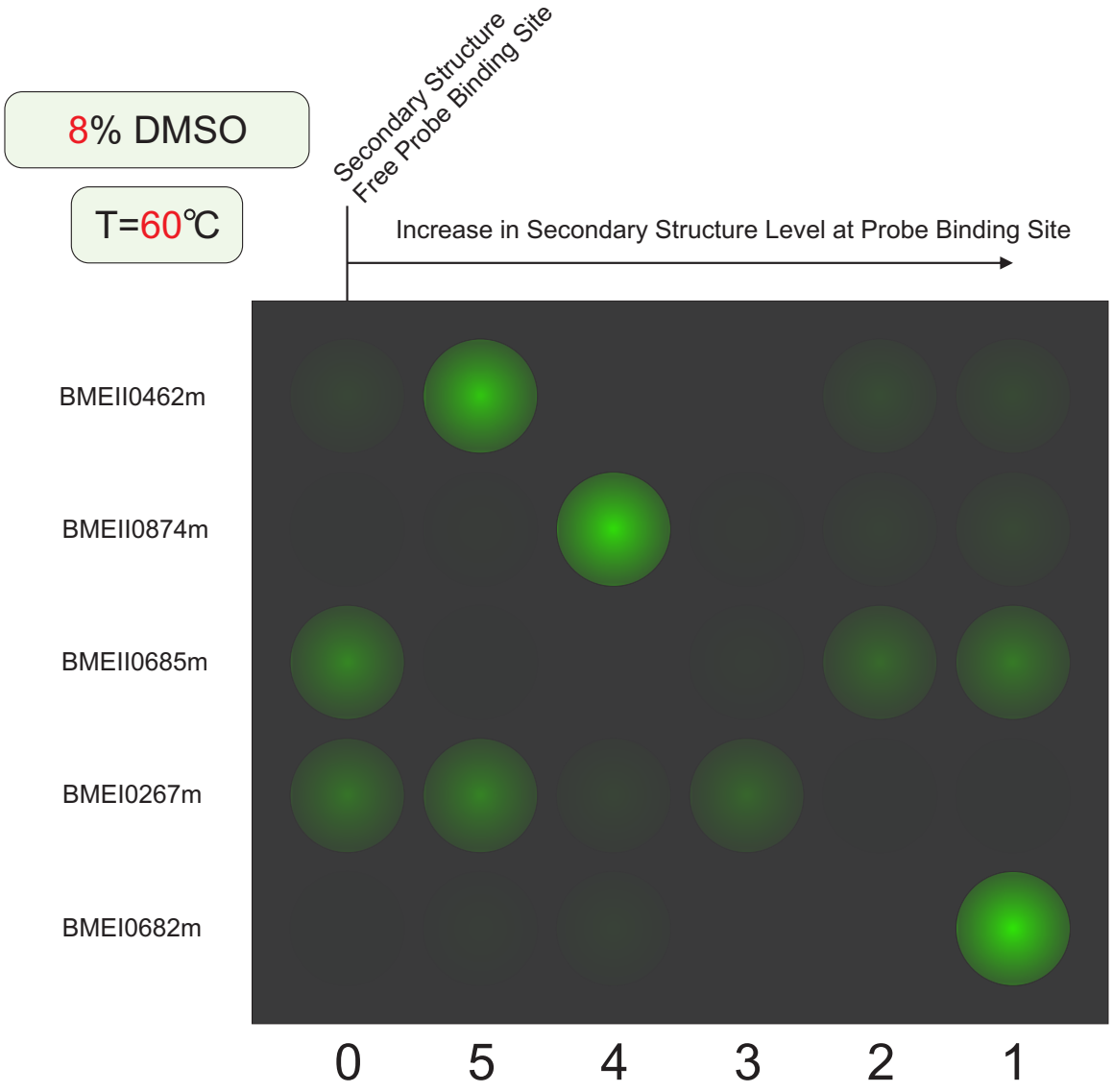

FIGURE 24: Graphical representation of computational simulation for the miniarray hybridization at 60 °C and 8% DMSO

The numbers 0, 5, 4, 3, 2, and 1 represent the probe numbers on the miniarray and are placed in the order of increasing secondary structure abundance in the probe-binding site.

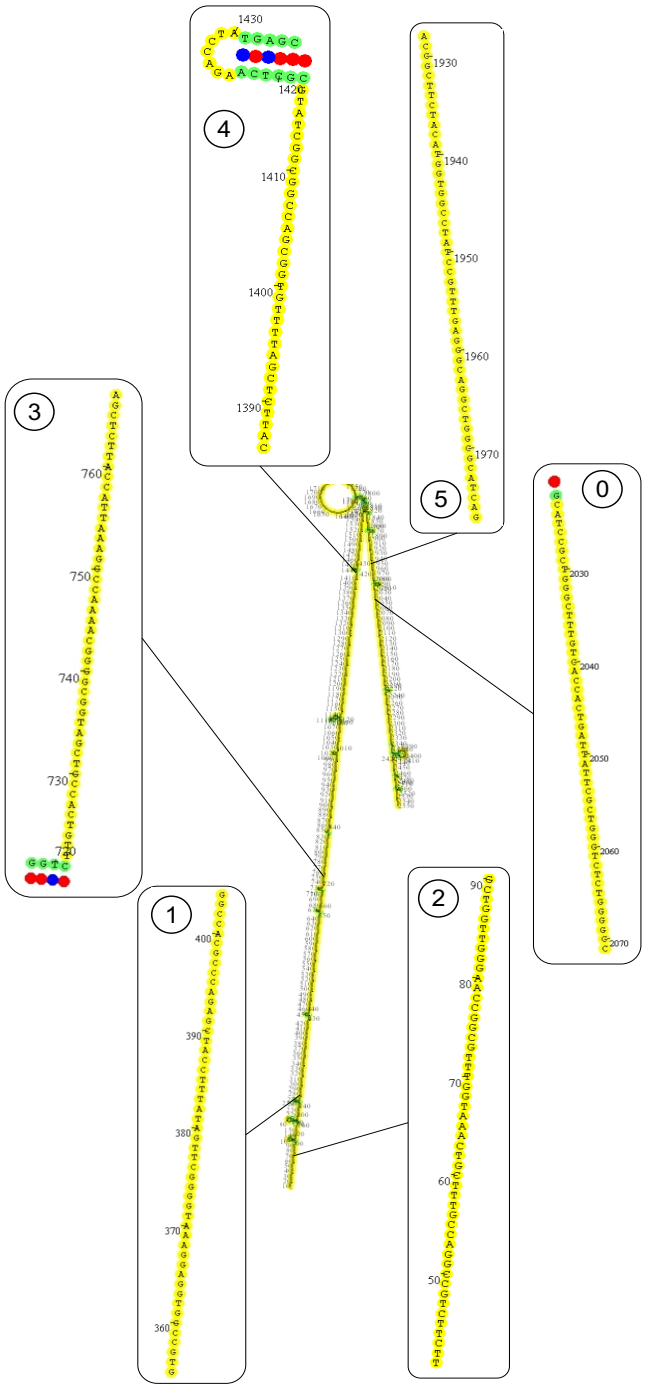

FIGURE 25: Relaxed Secondary Structure on BMEII0462m and Its Binding Sites at 5% DMSO at 60 °C

Bound nucleotides are drawn in bright green color, while all other bases are shown in yellow. The CG bonds are represented by red filled red circles, and the AT bonds and some non-Watson-Crick interactions are shown in blue.

APPENDIX A: SECONDARY STRUCTURE IN THE TARGET AS A CONFOUNDING FACTOR IN SYNTHETIC OLIGOMER MICROARRAY DESIGN

BMC Genomics

BioMed Central

Research article

Secondary structure in the target as a confounding factor in synthetic oligomer microarray design

Vladyslava G Ratushna[1], Jennifer W Weller[2] and Cynthia J Gibas*[1]

Address: [1]Department of Biology, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, 24061, USA and [2]School of Computational Science, Prince William Campus of George Mason University, Manassas, Virginia, 20110, USA

Email: Vladyslava G Ratushna - vratushn@vt.edu; Jennifer W Weller - jweller@gmu.edu; Cynthia J Gibas* - cgibas@vt.edu

* Corresponding author

Published: 08 March 2005

BMC Genomics 2005, **6**:31 doi:10.1186/1471-2164-6-31

Received: 10 September 2004

Accepted: 08 March 2005

This article is available from: http://www.biomedcentral.com/1471-2164/6/31

Abstract

Background: Secondary structure in the target is a property not usually considered in software applications for design of optimal custom oligonucleotide probes. It is frequently assumed that eliminating self-complementarity, or screening for secondary structure in the probe, is sufficient to avoid interference with hybridization by stable secondary structures in the probe binding site. Prediction and thermodynamic analysis of secondary structure formation in a genome-wide set of transcripts from *Brucella suis 1330* demonstrates that the properties of the target molecule have the potential to strongly influence the rate and extent of hybridization between transcript and tethered oligonucleotide probe in a microarray experiment.

Results: Despite the relatively high hybridization temperatures and 1M monovalent salt imposed in the modeling process to approximate hybridization conditions used in the laboratory, we find that parts of the target molecules are likely to be inaccessible to intermolecular hybridization due to the formation of stable intramolecular secondary structure. For example, at 65°C, 28 ± 7% of the average cDNA target sequence is predicted to be inaccessible to hybridization. We also analyzed the specific binding sites of a set of 70mer probes previously designed for *Brucella* using a freely available oligo design software package. 21 ± 13% of the nucleotides in each probe binding site are within a double-stranded structure in over half of the folds predicted for the cDNA target at 65°C. The intramolecular structures formed are more stable and extensive when an RNA target is modeled rather than cDNA. When random shearing of the target is modeled for fragments of 200, 100 and 50 nt, an overall destabilization of secondary structure is predicted, but shearing does not eliminate secondary structure.

Conclusion: Secondary structure in the target is pervasive, and a significant fraction of the target is found in double stranded conformations even at high temperature. Stable structure in the target has the potential to interfere with hybridization and should be a factor in interpretation of microarray results, as well as an explicit criterion in array design. Inclusion of this property in an oligonucleotide design procedure would change the definition of an optimal oligonucleotide significantly.

Background

Sequence-specific hybridization of a long single-stranded labeled DNA or RNA target molecule to shorter oligonucleotide probes is the basis of the gene expression microarray experiment. In this type of microarray experiment, gene specific *probe* molecules are either synthesized in situ or are printed to the microarray slide, and are either non-specifically cross-linked to the surface or are attached specifically using a method such as poly-Lysine linkers. *Target* molecules (most often fluorescently labeled cDNA molecules, although cRNA and aRNA are used in some protocols) hybridize transiently to the probe oligomers until they form stable double helices with their specific probes. At some point, the rate of on and off reactions reach equilibrium, and the concentration of the target in the sample solution can be calculated. Transcript abundance is assessed by the relative intensity of signal from each spot on the array. This interpretation of array data relies on the assumption that each hybridization reaction goes to completion within the timeframe of the experiment and that the behavior of all pairs of intended reaction partners in the experiment is somewhat uniform.

There are three major types of DNA microarrays, which differ in the approach used for probe design: Affymetrix type microarrays [1], which assay each transcript with a distributed set of 25-mer oligonucleotides, full length cDNA microarrays, in which long cDNA molecules of lengths up to several hundred bases are crosslinked to the slide surface to probe their complement [2], and synthetic long-oligomer probe microarrays, which usually assay each transcript only once. The latter class of microarrays encompasses a variety of commercial and custom platforms, and there has yet to emerge a consensus on an optimal probe length for particular experimental designs. Oligo lengths ranging from 35 to 70 nucleotides have been shown to perform well under different conditions [3-5], though recent studies have shown that oligomers of up to 150 nucleotides may be desirable for assessing transcript abundance [6]. In general, the use of synthetic oligomers has been shown to result in improved data quality [7,8] relative to cDNA arrays, and 70mers have been shown to detect target with a sensitivity similar to that of full length cDNA probes [9]. Short probes have been promoted because they facilitate finding unique sequence matches while forming fewer, and less stable, hairpin structures and because they display more uniform hybridization behavior overall. However, the need for sensitivity and detection of transcripts in low copy number drives the use of long-oligonucleotide arrays. In this study, we have modeled the accessibility of transcripts to hybridization with 70mer oligonucleotides.

A number of oligonucleotide design software packages have been published in recent years, each having design strengths in one of a number of criteria [10-14]. Several factors are considered by almost all microarray design software packages: in particular, the sequence specificity of the probe-target interface and the overall balance of GC content across the array. Unique regions of the target sequence are identified using sequence comparison methods; the unique regions become the search space for probe selection based on other criteria. The number of probes per sequence and location of the probe in the sequence also restrict sequence availability. A relatively uniform melting profile generally is achieved simply by selecting for probes with similar GC content and uniform or close-to-uniform length, although some design methods explicitly compute the duplex melting temperature for each candidate probe-target pair and filter unique probes to find those which match a specified range of melting temperatures. Another biophysical criterion that is sometimes applied is the elimination of probes having the ability to form stable intramolecular structures under the conditions of the experiment. This is usually done by eliminating regions of self-complementarity, although at least one design program [13] does explicitly compute the melting temperature of the most stable structure to form in the probe molecule and uses that information to filter out stable secondary structures in the probe.

Few of the available array design packages explicitly consider the possible structures of the transcript-derived molecules in the sample solution and their impact on whether the microarray will provide an effective assay, although the OligoDesign web server [14] does compute this information for use in design of locked nucleic acid probes. It has been shown that a hairpin of as little as six bases in an oligonucleotide can require a 600-fold excess of the complementary strand to displace the hairpin even partially [15]. Since the target molecules are generally longer than the probe and may be of a different chemistry, it is not sufficient to conclude that their behavior will mirror that of the complementary probe. Prediction of secondary structure in a sample transcript using a standard nucleic acid secondary structure prediction algorithm (Mfold) demonstrates that while longer-range interactions are reduced at high temperatures, stable local structures persist in the transcript even at high salt concentration and high temperature (Figure 1). Because unimolecular reactions within the target can occur on a much shorter timescale than the diffusion-mediated, bimolecular, duplex hybridization reaction, competition for binding by intramolecular structures is expected to block the specific probe annealing sites on the target sequence in some cases and result in misinterpretation of the signal obtained from the assay if these effects are not taken into account.

In order to estimate the prevalence of stable secondary structure in long target molecules, and thus the impact

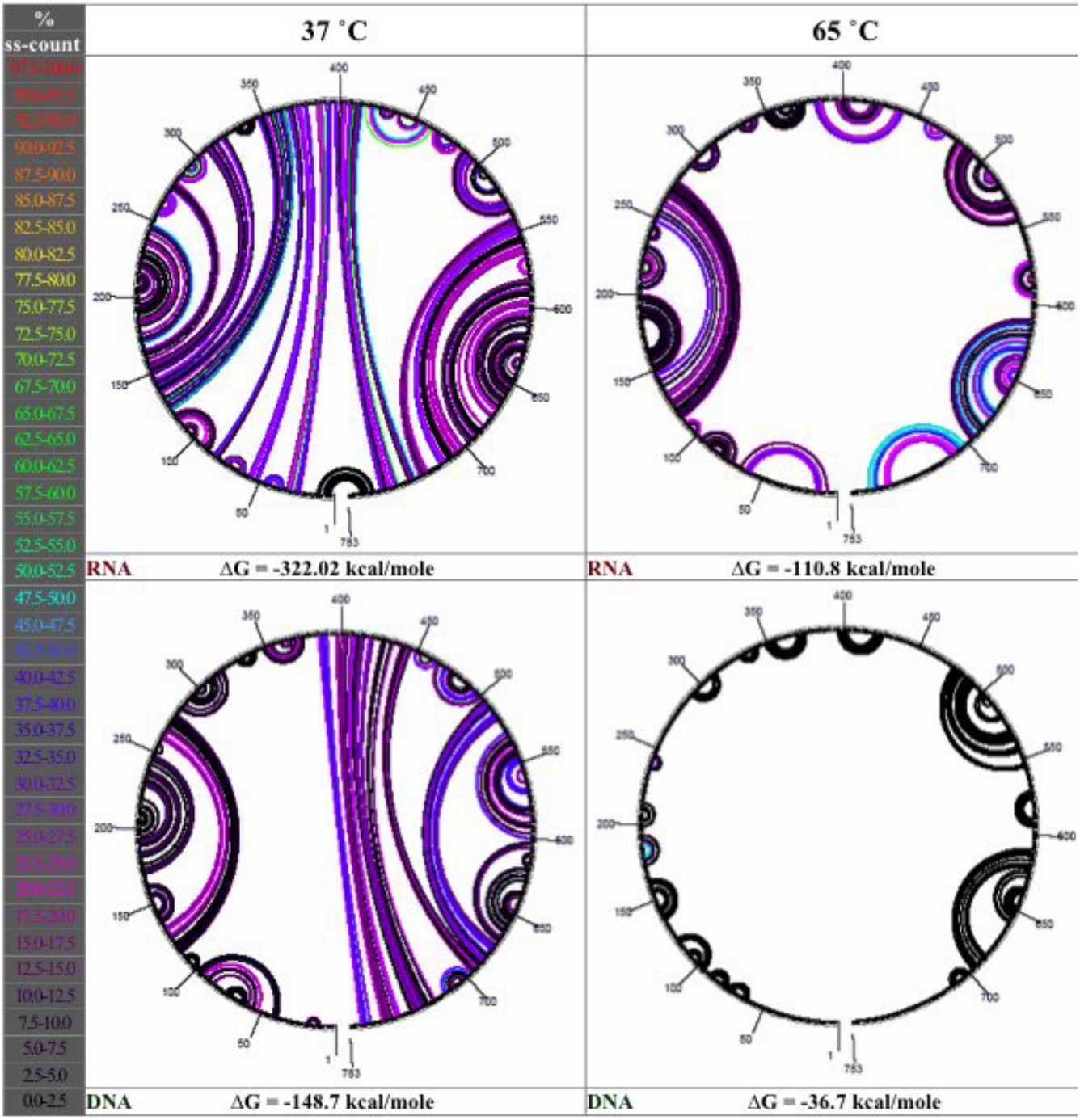

Figure 1
Secondary structure in a sample transcript. Circular diagrams of structure in a sample transcript (moeB homolog designated BR0004) from *Brucella suis*. Circular diagrams show hydrogen bonds between individual nucleotides, color-coded according to single-strandedness – the fraction of structures in which that bond is not present. Black bonds indicate 0% single-strandedness; red bonds indicate 100% single-strandedness.

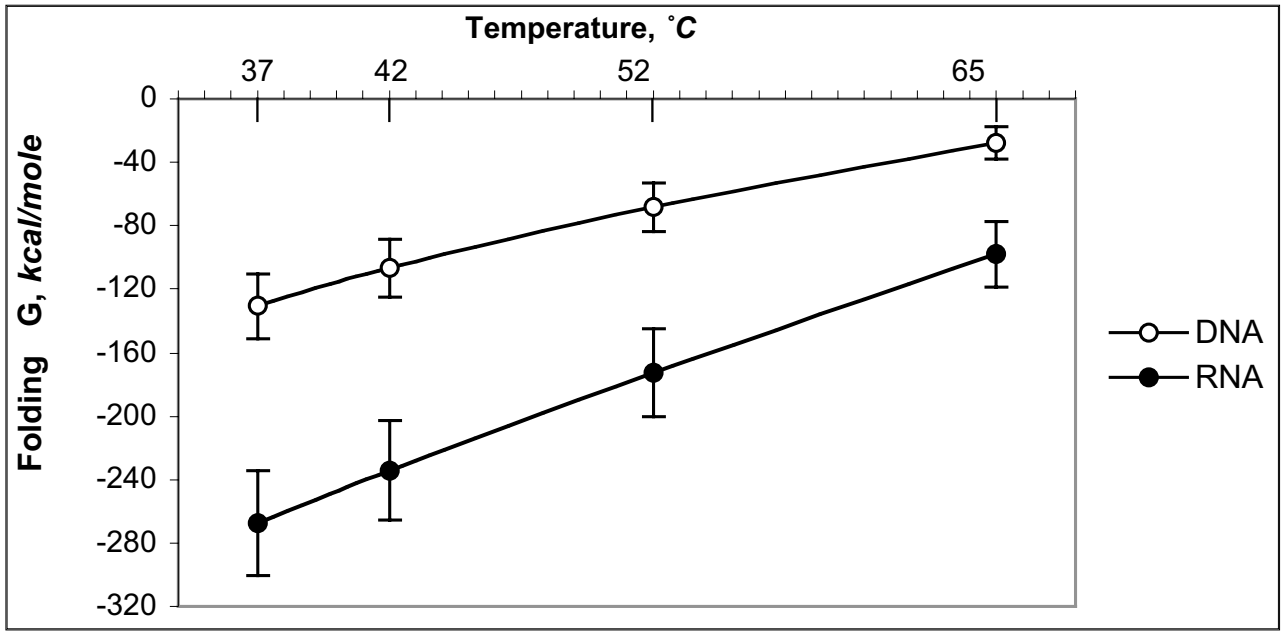

Figure 2
Stability of transcript secondary structure in *Brucella suis*. Average free energy change on global secondary structure formation for *Brucella suis* targets, modeled as DNA or RNA. ΔG values are normalized to global mean target length.

such structures might have on the analysis of microarray data, we have modeled secondary structure formation in mRNA transcripts of the intracellular pathogen *Brucella suis*. We have assessed the stability of structures formed in the transcript and the accessibility of the binding sites of optimal probes generated using commonly applied design criteria. Because random shearing of the full-length target molecule is used in some protocols, we have also modeled the effects of shearing to an average length on the prevalence of secondary structure in selected targets.

Results

Extent and stability of target secondary structure

Our modeling results obtained for the genome-wide set of intact single-stranded DNA or RNA targets demonstrate that stable secondary structures are widespread in target mixtures from *Brucella suis* (Figure 2) and in randomly chosen transcripts from the genomes of *E. coli* and *L. lactis*. Figure 2 shows the ΔG of formation for the most stable predicted secondary structure of the full-length transcript, as a function of reaction temperature. The major energy components of the Mfold ΔG are hydrogen bond energy and base pair stacking energy. These can be assumed to have a roughly linear relationship with transcript length. In order to make energies from different-length transcripts comparable, energies were normalized by computing a per-residue folding ΔG for each transcript and then multiplying that value by the global mean target length, for all transcripts considered from all organisms, of 851 bp. Average ΔG of secondary structure formation decreases with increasing temperature, but even at 65°C, the average ΔG of secondary structure formation for a full-length transcript is -98.2 kcal/mol (-27.9 kcal/mol when modeled as cDNA), meaning that the transcript is quite stable in that structure and a considerable energy input will be required to displace or melt the remaining structure. The trend in ΔG of secondary structure formation from the high-GC genome of *B. suis* to the low-GC genome of *L. lactis* is a decrease in overall stability. The average normalized ΔG of secondary structure formation for transcripts selected from the GC-balanced genome (*E. coli*) is near 70% of the average for *Brucella*, while the average ΔG for transcripts from the GC-poor genome (*L. lactis*) are even lower (30% at 52°C). However, even in the most GC-poor genome, stable target secondary structure in the single-stranded target is widespread.

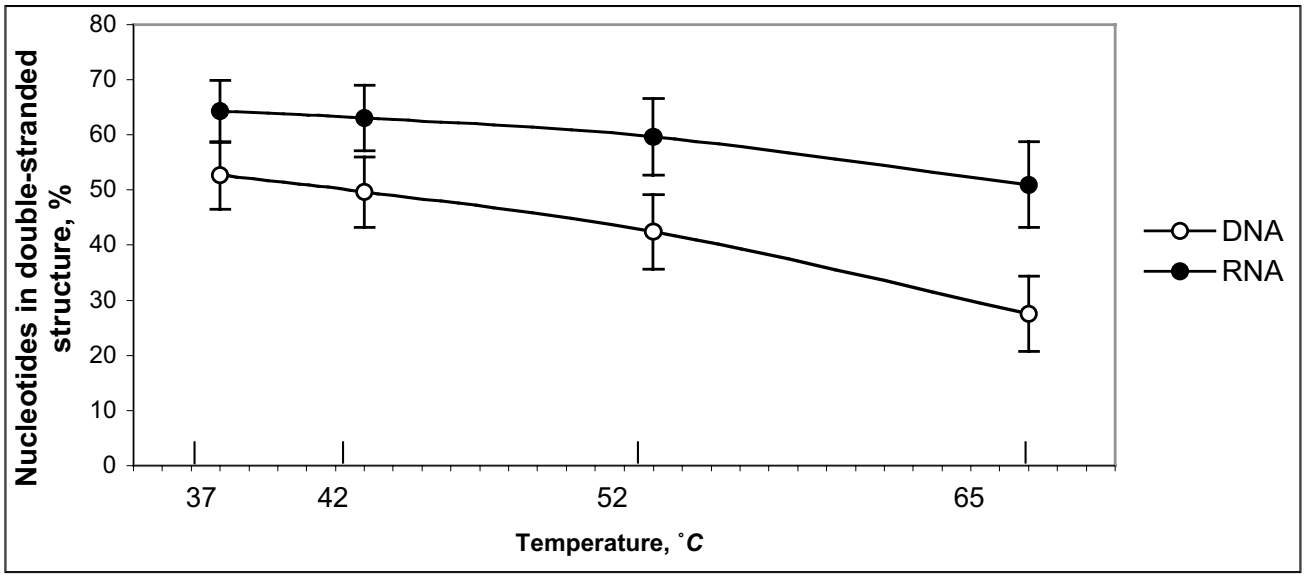

Figure 3
Fractional accessibility of nucleotides in the target. Fraction of the complete transcript classified as inaccessible due to the presence of stable structure in >50% of predicted conformations. Data shown are for 37, 42, 52 and 65°C simulations in *Brucella suis*.

Our results demonstrate that a significant fraction of nucleotide sites in the average target mixture, whether single stranded DNA or RNA, will be found in stable secondary structure under the hybridization conditions used in oligonucleotide microarray experiments, and will be relatively inaccessible for intermolecular interactions. Figure 3 shows the percentage of nucleotides that are in a double-helical state in at least 50% of the secondary structure conformations predicted by Mfold, at various reaction temperatures. The measure of accessibility used is the fraction of structures in which a nucleotide is found in a single-stranded conformation, when all optimal and suboptimal structures predicted are considered.

Extent and stability of target secondary structure

Figure 4 is a plot of the average ΔG of structure formation when shearing of the target molecule is simulated by dividing the target into overlapping 200, 100, and 50mer fragments. Shearing the target into smaller fragments destabilizes secondary structure, especially at very short fragment lengths. However, shearing does not eliminate occlusion of nucleotides by secondary structure, even in the shortest fragments examined. When a DNA target is modeled at 52°C, for example, the double stranded fraction decreases by only about 30% – from 41% to 29% – when the target is simulated as sheared into 50mer fragments. However, in hybridization experiments involving low copy number targets and longer oligos, creating extremely short target fragments may reduce or eliminate the signal on the chip, because the target can not be sheared specifically to present an unbroken hybridization site for the probe, and so some fragments will be created that match the probe only partially.

Interference of secondary structure with the hybridization site

Figure 5 shows the average percentage of nucleotides within a probe binding region in the target that are inaccessible, when different fractional accessibility cutoffs are used to classify the sites. Even when a relatively demanding criterion – double-strandedness in over 75% of optimal and suboptimal structures – is used to classify a nucleotide as inaccessible, an average of 21 ± 13% of nucleotides in the probe binding region are found in stable secondary structures at 65°C. Figure 6 shows a representative transcript and the challenge it presents to hybridization when modeled as full-length cDNA and fragments of various lengths.

Discussion

Lack of bioinformatics tools that incorporate experimentally validated biophysical properties of nucleic acids as a

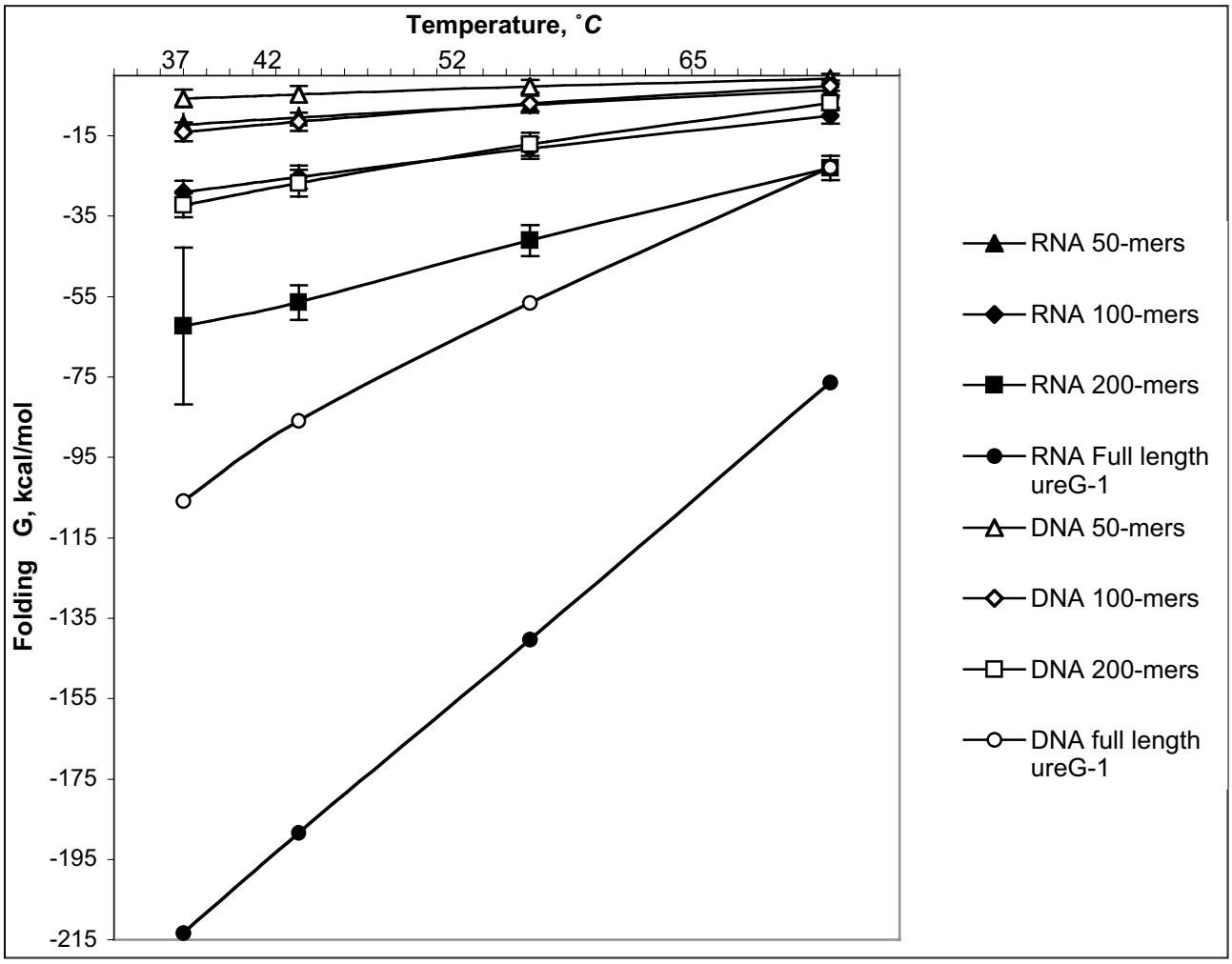

Figure 4
Stability of secondary structure in sheared fragments. Free energy change on secondary structure formation for the ureG-1 RNA transcript from *Brucella suis*. The transcript is modeled as sheared into fragments of length 200 nt, 100 nt or 50 nt; fragments are chosen starting at every 10th residue.

criterion for synthetic oligomer probe design is a major challenge for do-it-yourself microarray designers. One biophysical characteristic, which we predict will reduce the binding efficiency of microarray probes to their targets, is the propensity of long single-stranded DNA or RNA molecules to form stable secondary structure. 3-D structures such as hairpins and stacked regions have the potential to pre-empt target nucleotides, thus blocking regions of the target molecules from hybridizing to their intended probes. Prediction and thermodynamic analysis of secondary structure at a range of temperatures in full length target sequences, as well as in subsequences formed by *in silico* shearing, revealed the likely presence of stable secondary structures in both full-length target and sheared target mixtures. These structures do not convert completely to random coil with either increasing hybridization temperature, more extensive shearing, or both. These secondary structures may therefore compete with the intended target for effective probe annealing in a microarray experiment, resulting in a misinterpretation of the amount of target present in the sample.

Applying target secondary structure as a criterion in array design

Based on the results of this *in silico* experiment, secondary structure prediction in the target is being used to develop

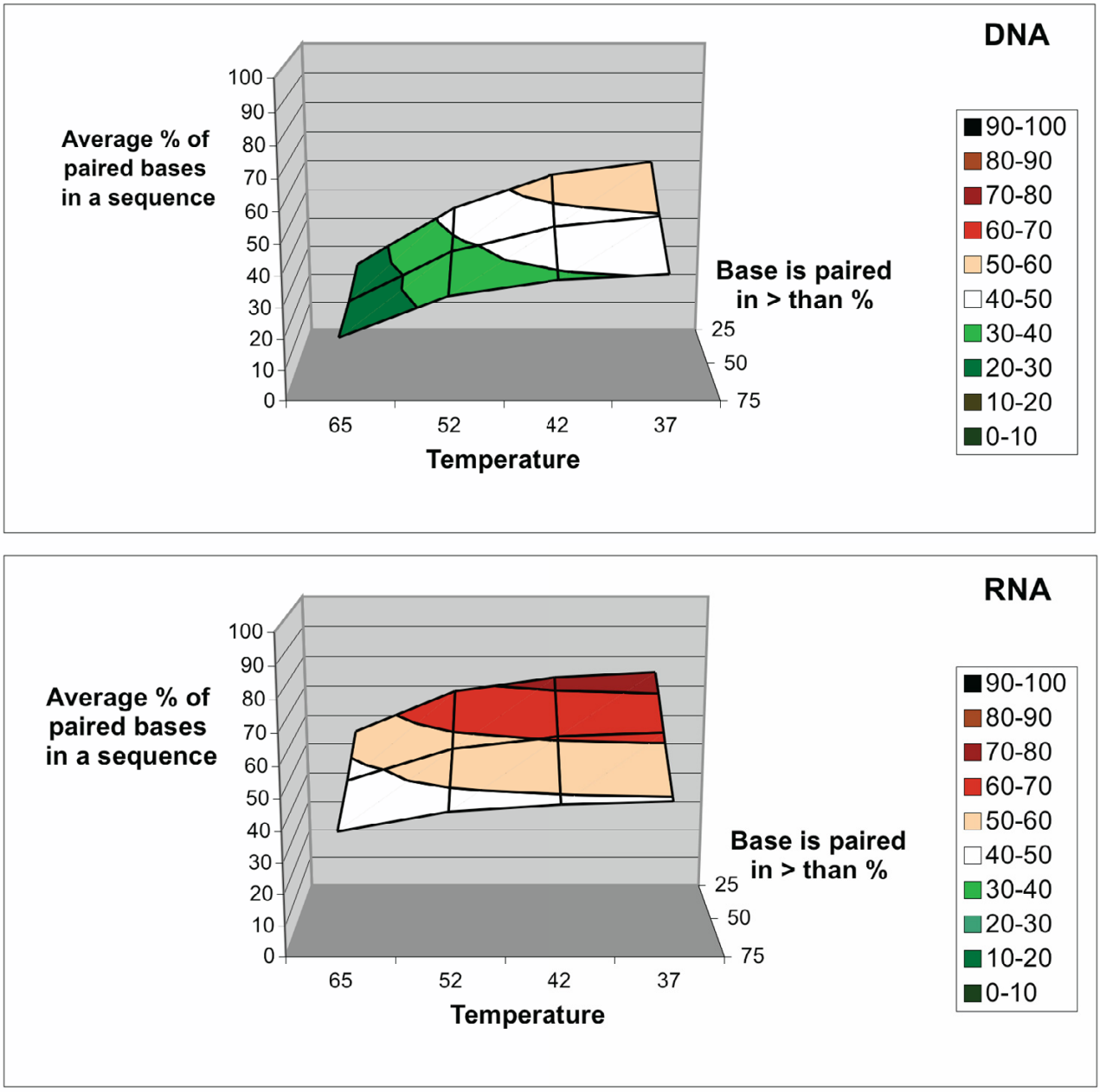

Figure 5
Accessibility of the probe binding site. Fraction of the average probe binding site in the *Brucella* genomic array that is found to be inaccessible at 37°, 42°, 52° and 65°C, for DNA or RNA target. Inaccessible sites are defined here using three different cutoffs for the fraction of structures in which the site is base-paired: 25%, 50%, and 75%.

a new criterion for oligonucleotide probe design. Our results from this modeling experiment demonstrate that the implicit assumption used until now – that eliminating probe secondary structure by avoiding self-complementarity eliminates target secondary structure as well – is valid only when the target and probe are of the same length. Use of target secondary structure as an explicit criterion will allow for masking or preferentially avoiding the

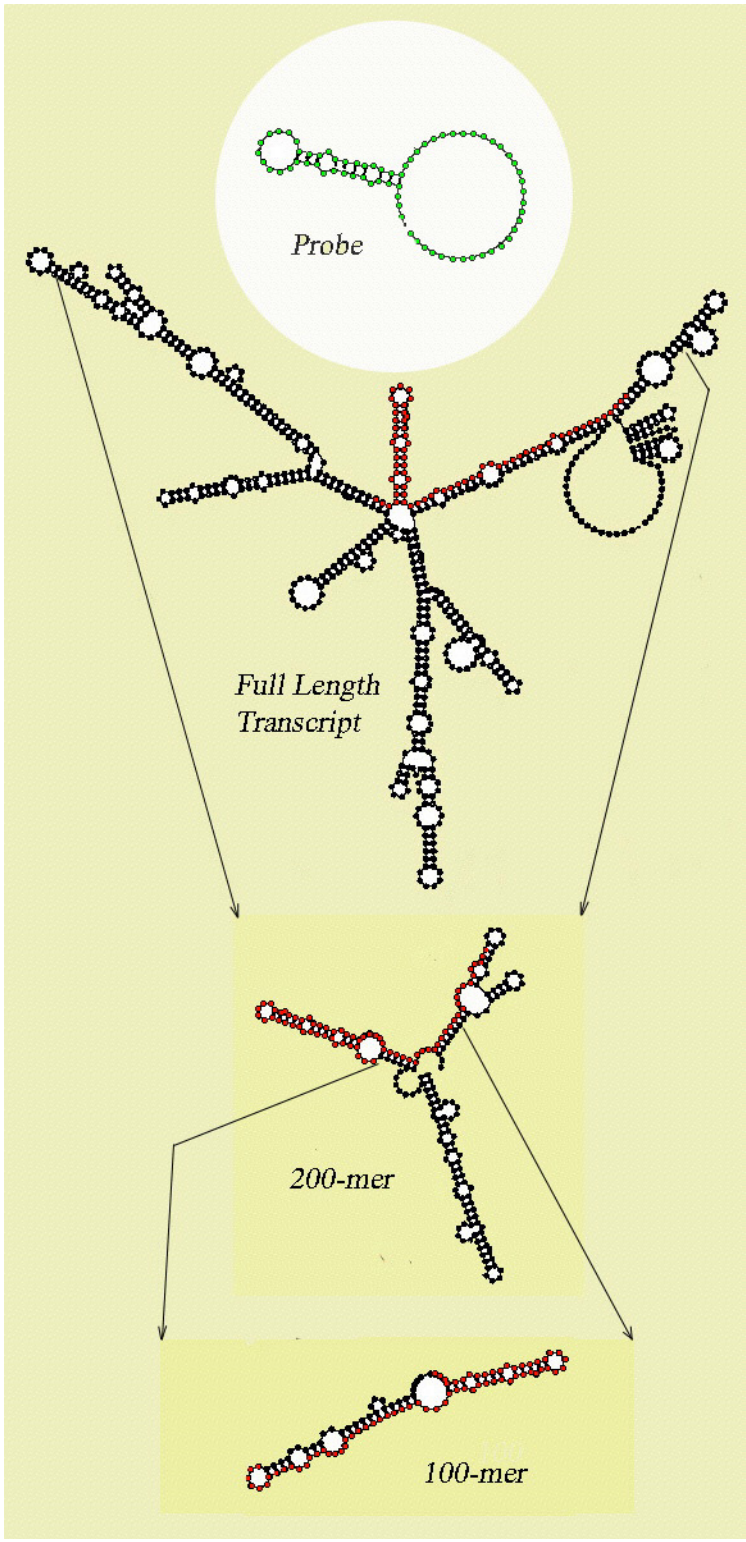

Figure 6
Structure in a binding site – full length target and sheared fragments. The position of a 70mer oligonucleotide probe (green) binding site (red dots) within a full-length optimal transcript structure, as well as examples of stable structure in 200mer and 100mer fragments which overlap the probe binding site. Corresponding ΔG values for these fragments modeled at 42° and 52°C are shown in Table 1.

regions of the target sequence in which base pairs are directly involved in secondary structure formation, to eliminate these regions from the sequence for the purpose of the search for the optimal probe.

In this study we have assigned accessibility scores to sites in the target sequence based only on the fraction of predicted structures within 5% of the energy optimum, in which a residue is found in a single-stranded conformation. While this measure is not too computationally intensive to compute, and can be applied to genome-scale problems using readily available software (Mfold), it is not the most physically rigorous definition of accessibility. By equally weighting each possible structure in the ensemble of optimal and suboptimal structures that a molecule can form, it is possible that secondary structure at some positions in the molecule is overcounted; bonds which form only in rare conformations are considered equal to bonds which are present in the lowest-energy structure. The program Sfold [16-18] assigns accessibility based on an ensemble-weighted average of secondary structure. The program RNAfold[19], part of the Vienna RNA package, implements McCaskill's partition function approach[20] to arrive at pairing probabilities for each pair of bases in the sequence, from which a summary per-base accessibility can be derived. These methods are more rigorous than MFold and we expected they might produce somewhat different results, although it has also been shown that predicted binding states from MFold optimal structures perform almost as well as SFold and RNAFold predictions when applied to molecules of known 3D structure [16].

When we compared MFold-based accessibility predictions for an individual transcript to those generated by SFold and RNAFold, we found that the difference in average predicted accessibility over an entire transcript is small. We computed accessibility for the transcript of human 1CAM-1, which has been mapped experimentally to determine its accessibility [21]. The average fractional accessibility derived from MFold results is about 3–4% greater than that predicted by RNAFold or SFold. Therefore use of this fractional accessibility measure will not impose an unnecessary constraint on the design process relative to other predictive approaches. The accessibility profiles calculated for ICAM-1 using each method are shown in Fig. 7. In each section of the figure, antipeak locations (having lower pairing probability and therefore likely to be more accessible) can be compared to the extendable sites detected by Allawi et al [21], which are indicated by green dots at the bottom of the plot. In each prediction, there are a number of apparently correct predictions and obvious errors, and it is not clear which method is yielding the best results at the residue level. A systematic, competitive test of these predictions against solution accessibility data gathered on various experimental platforms is called for, although available data sets for validation are still rare. In the absence of such validation, the MFold accessibility predictions are sufficient to predict the scope of the secondary structure problem in a genome-based array design, even if some details of the prediction are not correct. An experimental approach will eventually be required to determine which approach best represents the conditions of the microarray experiment.

Loop length and other considerations

In this study, we focused specifically on the DNA/RNA base pairs that are actively involved in hydrogen bond formation. We realize that other accessibility considerations will have to be added to the scoring scheme in practice. The structure of a long single stranded DNA or RNA molecule can contain many nucleotides that, while not part of a double-helical stem, remain inaccessible to hybridization due to their location inside small loops within the target secondary structure. A loop is a somewhat constrained structure as well, and the length at which it presents accessible sequence that favors hybridization has been shown to be on the order of 10 nucleotides and longer [22], while nucleotides found in shorter loops may be classifiable as inaccessible. However, there is a need for quantitative hybridization experiments that would elucidate how loops and loop-like structures in tethered long-oligo probe and target molecules affect the performance of assays, and we have chosen not to formulate a system for scoring the accessibility of single-stranded loop structures or weighting this criterion relative to the double-strandedness criterion until we have carried out some of these experiments.

Development of a target secondary structure criterion for oligonucleotide array design is expected to impose restrictions on the probe selection beyond the sequence similarity and melting temperature criteria that are currently used, especially in cases where short probe length restricts the annealing temperature used in the hybridization protocol to 22–37°. In the *B. suis* example, use of a low annealing temperature, e.g. 42°C which is the temperature used in some published 70-mer array experiments [9], would result in only about 30% of the average transcript being accessible for intermolecular hybridization, not counting 'free' bases found in short loops in secondary structures. There will be greater design latitude for experiments carried out at higher hybridization temperatures. Recommended hybridization temperatures for long synthetic oligomer arrays may prove to be closer to 65°C, when only 50% of a typical RNA transcript or 30% of the corresponding cDNA molecule remain inaccessible.

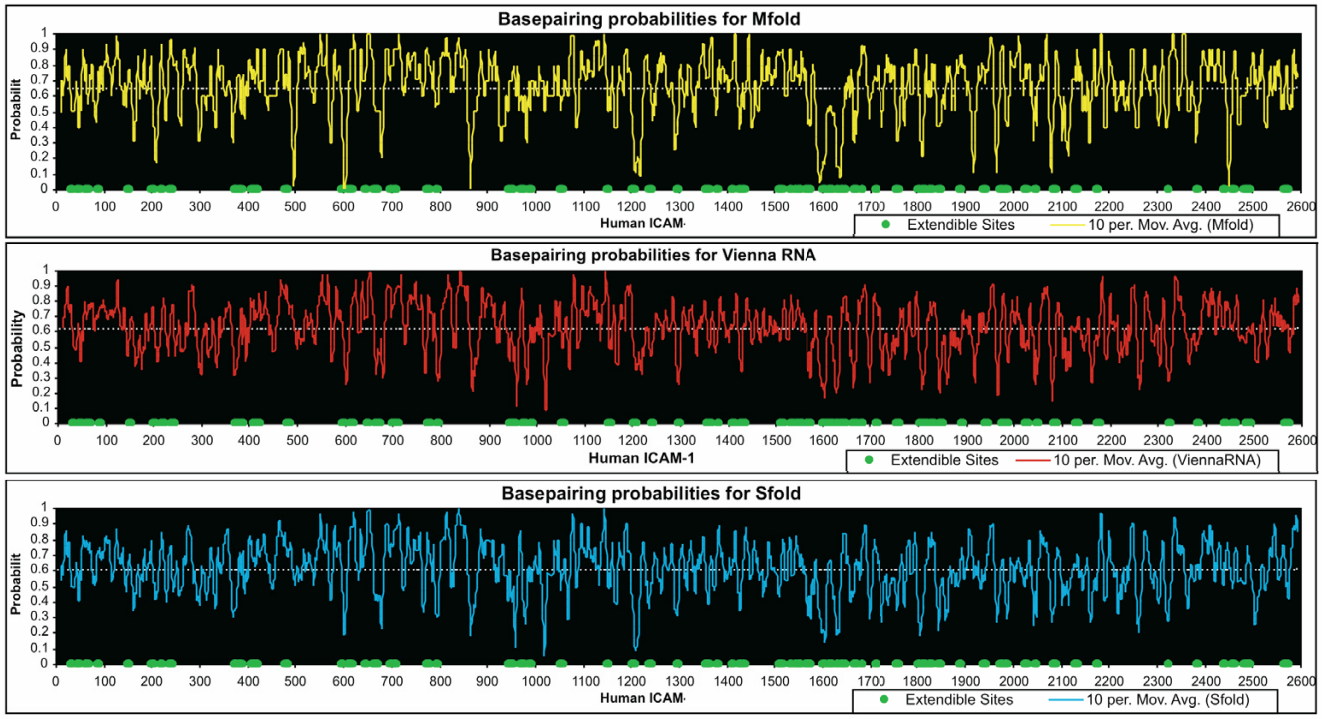

Figure 7
Accessibility prediction using three common methods. Pairing probabilities computed using RNAFold (top), MFold (middle) and SFold (bottom) for the human ICAM-1 transcript. Extendable sites detected by Allawi et al [21]

To shear or not to shear

We have shown here that while shearing reduces overall ΔG of secondary structure formation for individual molecules in the target solution, shearing does not in itself eliminate formation of secondary structure in single-stranded DNA or RNA. The question of whether shearing should be used for long oligomer arrays is still an open one. While some signal may be gained by reducing the stability of secondary structure in the target molecule, random shearing by its nature creates a mixture of targets that may have substantially different affinities. For instance, in a 300 nt transcript that is targeted by a 70mer oligonucleotide, there is nearly a one in four chance that a random break in the sequence will occur within the target site for which the probe is designed. Short fragments may present a substantially different binding site, and therefore have a different binding affinity, than the full-length transcript that is considered when the probe is designed. This is illustrated in Figure 8d, where binding of a 50mer sheared fragment to a 70mer probe leaves a dangling end in the probe. A break very close to one end or the other of the target site may create a target that still binds to the probe, though with reduced affinity; a break closer to the middle of the target site may produce fragments that bind partially to the probe, competing for binding with perfect matches.

The utility of experimentally validated biophysical criteria

In other experimental contexts where hybridization is critical to success, the impact of secondary structure in single stranded polynucleotides on results has been recognized and is now being systematically studied (18–21). Intramolecular folding of mRNAs is so extensive that only 5–10% of most transcripts is accessible to binding of complementary nucleic acids; however the modeling of long molecules has not proven to give very accurate binding predictions [23-25]. In fact, array-based screens have been utilized to empirically select oligonucleotides that bind effectively to transcripts for siRNA experiments [23,26]. Several studies have demonstrated that, at 37°C and 0 mM Mg2+ oligonucleotides of length >20 yield good binding/RNAseH digestion at low concentrations relative to shorter oligonucleotides (30 nM vs 300 nM compared) and found that microarray binding was a good predictor of siRNA activity despite the 3' tethering and 1M NaCl used in array experiments vs siRNA experiments [26]. Systematic "scanning" of mRNA sequences with libraries of short oligos [27] has also been shown to be successful in

locating sites for siRNA targeting; however, such methods are likely to become extremely expensive if applied to the large number of targets in a microarray design. We have begun to develop an experimental approach to this problem, in which structure predictions like those used in this study are experimentally evaluated to determine whether the structures we can predict using existing modeling approaches will detectably affect signal in the microarray context.

Conclusion

The results of the current study suggest a significant role for target secondary structure in hybridization to oligonucleotide arrays, which will warrant further investigation. Oligonucleotide probe binding sites in a significant fraction of transcripts are found in double-stranded conformations even in cases where self-complementarity was avoided during the probe design process. We find that at 52°C, for example, approximately 57% of probes designed for *Brucella* had binding sites in the target which were predicted to contain a stretch of unpaired bases of at least 14 nt in length; at 65°C, that fraction increased to 93%. Based on these findings we would expect that at 52°C only 57% of our probes would encounter optimal conditions for hybridization and therefore would demonstrate the expected behavior in the experiment, where intensity is expected to scale with target concentration. We predict that the remaining probes, which have shorter, or no, accessible sequences, will exhibit modified binding behavior, and we plan to conduct experiments to characterize this behavior. We have shown conclusively that avoiding self-complementarity in the probe when designing an oligonucleotide array is insufficient to eliminate secondary structure from the binding site in the target. By combining the procedure for systematic computational assessment of transcript accessibility described in this study with selective experimental validation of the impact of predicted accessibility on hybridization, we will develop a useful criterion for avoiding troublesome secondary structure when designing microarray targets.

Methods

Prediction and thermodynamic analysis of secondary structure was performed for all protein-coding gene transcripts predicted from 3264 CDSs in the *Brucella suis* 1330 genome. *Brucella suis* has a relatively high (57%) genomic GC content. *Brucella suis* was chosen for this experiment because our collaborators have previously acquired a custom synthetic oligomer microarray for this organism, developed using standard oligo array design software, and we have access to both target sequences and to a set of unique probe sequences that define the interaction sites for which expression results have been obtained by the laboratory.

In order to determine whether Brucella sequences form atypical structures we randomly picked and analyzed 50 gene coding sequences from a compositionally balanced genome (*Escherichia coli*), and 50 from the GC-poor genome of the nonpathogenic AT-rich gram-positive bacterium *Lactococcus lactis* (35% genomic GC content). The *Brucella suis* genes ranged in length from 90 to 4,803 bp, with an average transcript length of 851 bp. The *E. coli* genes ranged in length from 140 to 2,660 bp, with an average transcript length of 792 bp. The range of GC content in the genes chosen was 37% to 57% with an average value of 50%, which is reasonably representative of the *E. coli* genome. The *L. lactis* genes chosen ranged in length from 140 to 2,730 bp, with an average transcript length of 765 bp., and ranged in GC content range from 30% to 42% with an average value of 35%.

Microarray design

70-mer probes for each *Brucella suis* target were previously designed (Stephen Boyle, personal communication) using ArrayOligoSelector (pick70) [10]. ArrayOligoSelector uses sequence uniqueness, self-complementarity, and sequence complexity as criteria but does not explicitly evaluate ΔG of secondary structure formation for the probe. 72% of the probes designed using this method were found to contain secondary structures with melting temperatures greater than 65°C, and 10% contained secondary structures with melting temperatures greater than 80°C. The Brucella probes defined the interaction sites within the target transcripts for which structural accessibility was evaluated.

Secondary structure prediction

Probe and transcript secondary structure were predicted using the Mfold 3.1 software package [28,29]. Mfold identifies the optimal folding of a nucleic acid sequence by energy minimization and can identify suboptimal foldings within a specified energy increment of the optimum as an approach to modeling the ensemble of possible structures that a single-stranded nucleotide molecule can assume. We modeled secondary structure in the single-stranded target, modeling the target both as DNA and as RNA, at a range of temperatures which is inclusive of hybridization temperatures commonly used in microarray protocols: 37°C, 42°C, 52°C and 65°C. The modeling conditions were chosen within the allowed settings of Mfold to approximate a microarray experiment: solution conditions of 1.0 M sodium concentration and no magnesium ion were used. The free energy increment for computing suboptimal foldings, ΔΔG, was set to 5% of the computed minimum free energy. The default values of the window parameters, which control the number of structures automatically computed by Mfold 3.1, were chosen based on the sequence length. Free energy changes on

Table 1: Stability of a sample transcript – full length target and sheared fragments Folding ΔG of target transcript and fragment molecules shown in Figure 8, at hybridization temperatures commonly used for long oligomer arrays.

Molecule	^{2}G, kcal/mole			
	42°C		52°C	
	DNA	RNA	DNA	RNA
70-mer Probe	- 6.8	N/A	- 4.2	N/A
Full Length Target	- 85.9	- 188.4	-56.6	- 140.2
200-mer sheared Target	- 25.5	- 58.6	-15.9	- 41.6
100-mer sheared Target	-14.2	- 25.7	-9.6	-18.0
50-mer sheared Target (not shown)	- 6.1	-10.5	- 4.2	-7.3

formation of secondary structure were extracted from the Mfold output.

Accessibility calculation

Accessibility in folded single-stranded DNA or RNA has recently begun to be addressed in a few experimental studies, mainly with the goal of targeting appropriate sites for RNAi. Because the structure of single-stranded nucleotide molecules is much more dynamic than that of proteins, with each molecule likely to exist in an ensemble of structures, and because the 3D structure of these molecules is rarely known, there is not yet a consensus representational standard of per-residue accessibility for single-stranded nucleic acids. Ding et al. [17,18] implement probability of single-strandedness, when the weighted ensemble of likely structures is taken into account, as an accessibility criterion. However, use of their Sfold server, with batch jobs limited to 3500 bases, is not currently practical for a genome-scale survey of accessibility. Another approach to accessibility prediction is McCaskill's partition function approach [20] which can be used to compute base pair probabilities and summary pairing probability for any base. This approach is implemented in RNAFold [19], a component of the Vienna RNA package.

In this study, we chose to use the less physically rigorous approximation of probability of single strandedness as a simple fraction of predicted optimal and suboptimal structures in which a residue is found to be part of a single stranded structure, as computed by Mfold. Accessibility scores derived from MFold predictions have been used in limited studies of RNA structure focused on hammerhead ribozymes[30], antisense and siRNA targeting [22,31] and have been shown to be predictive in cases where some experimental measure of accessibility has been made[32]. While MFold-derived accessibility scores may not be completely optimal, they have been used with reasonable success to predict accessibility in the siRNA targeting context, and so we use MFold here.

Shearing simulation

Random shearing of the target mixture is an approach that is often offered as a solution for the problem of target secondary structure. The actual content of a sheared mixture of DNA or RNA fragments is complex. Shearing breaks the molecule not in predictable locations, but in random locations that give rise to a distribution of fragments around an average fragment length. In order to simulate the effects of different degrees of shearing on structure formation and stability in a transcript, we picked fragments of 200, 100, or 50 bases in length, choosing the start position via a sliding window of 10 bases. Secondary structure prediction for all fragments derived from every transcript in the B. suis genome is computationally intensive and produces an extremely large amount of output. Since our initial goal was to determine how much the method would affect the number and type of secondary structures probes would be expected to bind the shearing simulation was performed for fragments derived from the 300 bp Ure-1A gene of *B. suis*. Secondary structure and thermodynamics were computed for each of these fragments individually.

Authors' contributions

VGR participated in the design of the study, carried out the simulations and analysis, and drafted the manuscript. JWW participated in the design of the study and helped to draft the manuscript. CJG conceived of the study, participated in its design, coordinated the research and analysis, and drafted the manuscript.

Acknowledgements

The authors would like to acknowledge the Fralin Biotechnology Center for research startup support (CJG) and the Department of Biology for graduate assistantship support (VGR).

References

1. Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, Mittmann M, Wang C, Kobayashi M, Horton H, Brown EL: **Expres-**

sion monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol* 1996, **14(13):**1675-1680.
2. Schena M, Shalon D, Davis RW, Brown PO: **Quantitative monitoring of gene expression patterns with a complementary DNA microarray.** *Science* 1995, **270(5235):**467-470.
3. Kane MD, Jatkoe TA, Stumpf CR, Lu J, Thomas JD, Madore SJ: **Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays.** *Nucleic Acids Res* 2000, **28(22):**4552-4557.
4. Hughes TR, Mao M, Jones AR, Burchard J, Marton MJ, Shannon KW, Lefkowitz SM, Ziman M, Schelter JM, Meyer MR, Kobayashi S, Davis C, Dai H, He YD, Stephaniants SB, Cavet G, Walker WL, West A, Coffey E, Shoemaker DD, Stoughton R, Blanchard AP, Friend SH, Linsley PS: **Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer.** *Nat Biotechnol* 2001, **19(4):**342-347.
5. Ramakrishnan R, Dorris D, Lublinsky A, Nguyen A, Domanus M, Prokhorova A, Gieser L, Touma E, Lockner R, Tata M, Zhu X, Patterson M, Shippy R, Sendera TJ, Mazumder A: **An assessment of Motorola CodeLink microarray performance for gene expression profiling applications.** *Nucleic Acids Res* 2002, **30(7):**e30.
6. Chou CC, Chen CH, Lee TT, Peck K: **Optimization of probe length and the number of probes per gene for optimal microarray analysis of gene expression.** *Nucleic Acids Res* 2004, **32(12):**e99.
7. Shi SJ, Scheffer A, Bjeldanes E, Reynolds MA, Arnold LJ: **DNA exhibits multi-stranded binding recognition on glass microarrays.** *Nucleic Acids Res* 2001, **29(20):**4251-4256.
8. Yue H, Eastman PS, Wang BB, Minor J, Doctolero MH, Nuttall RL, Stack R, Becker JW, Montgomery JR, Vainer M, Johnston R: **An evaluation of the performance of cDNA microarrays for detecting changes in global mRNA expression.** *Nucleic Acids Res* 2001, **29(8):**E41-1.
9. Wang HY, Malek RL, Kwitek AE, Greene AS, Luu TV, Behbahani B, Frank B, Quackenbush J, Lee NH: **Assessing unmodified 70-mer oligonucleotide probe performance on glass-slide microarrays.** *Genome Biol* 2003, **4(1):**R5.
10. Bozdech Z, Zhu J, Joachimiak MP, Cohen FE, Pulliam B, DeRisi JL: **Expression profiling of the schizont and trophozoite stages of Plasmodium falciparum with a long-oligonucleotide microarray.** *Genome Biol* 2003, **4(2):**R9.
11. Chou HH, Hsia AP, Mooney DL, Schnable PS: **PICKY: oligo microarray design for large genomes.** *Bioinformatics* 2004.
12. Nielsen HB, Wernersson R, Knudsen S: **Design of oligonucleotides for microarrays and perspectives for design of multi-transcriptome arrays.** *Nucleic Acids Res* 2003, **31(13):**3491-3496.
13. Rouillard JM, Zuker M, Gulari E: **OligoArray 2.0: design of oligonucleotide probes for DNA microarrays using a thermodynamic approach.** *Nucleic Acids Res* 2003, **31(12):**3057-3062.
14. Tolstrup N, Nielsen PS, Kolberg JG, Frankel AM, Vissing H, Kauppinen S: **OligoDesign: optimal design of LNA (locked nucleic acid) oligonucleotide capture probes for gene expression profiling.** *Nucl Acids Res* 2003, **31(13):**3758-3762.
15. Nguyen HK, Southern EM: **Minimising the secondary structure of DNA targets by incorporation of a modified deoxynucleoside: implications for nucleic acid analysis by hybridization.** *Nucl Acids Res* 2000, **28(20):**3904-3909.
16. Ding Y, Chan CY, Lawrence CE: **Sfold web server for statistical folding and rational design of nucleic acids.** *Nucleic Acids Res* 2004, **32(Web Server issue):**W135-41.
17. Ding Y, Lawrence CE: **Statistical prediction of single-stranded regions in RNA secondary structure and application to predicting effective antisense target sites and beyond.** *Nucleic Acids Res* 2001, **29(5):**1034-1046.
18. Ding Y, Lawrence CE: **A statistical sampling algorithm for RNA secondary structure prediction.** *Nucleic Acids Res* 2003, **31(24):**7280-7301.
19. Hofacker ILFWSPFBSTMSP: **Fast folding and comparison of RNA secondary structures.** *Monatshefte f Chemie* 1994.
20. McCaskill JS: **The equilibrium partition function and base pair binding probabilities for RNA secondary structure.** *Biopolymers* 1990, **29(6-7):**1105-1119.
21. Allawi HT, Dong F, Ip HS, Neri BP, Lyamichev VI: **Mapping of RNA accessible sites by extension of random oligonucleotide libraries with reverse transcriptase.** *Rna* 2001, **7(2):**314-327.
22. Scherr M, Rossi JJ, Sczakiel G, Patzel V: **RNA accessibility prediction: a theoretical approach is consistent with experimental studies in cell extracts.** *Nucleic Acids Res* 2000, **28(13):**2455-2461.
23. Sohail M, Akhtar S, Southern EM: **The folding of large RNAs studied by hybridization to arrays of complementary oligonucleotides.** *Rna* 1999, **5(5):**646-655.
24. Lima WF, Monia BP, Ecker DJ, Freier SM: **Implication of RNA structure on antisense oligonucleotide hybridization kinetics.** *Biochemistry* 1992, **31(48):**12055-12061.
25. Michel F, Westhof E: **Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis.** *J Mol Biol* 1990, **216(3):**585-610.
26. Bohula EA, Salisbury AJ, Sohail M, Playford MP, Riedemann J, Southern EM, Macaulay VM: **The efficacy of small interfering RNAs targeted to the type 1 insulin-like growth factor receptor (IGF1R) is influenced by secondary structure in the IGF1R transcript.** *J Biol Chem* 2003, **278(18):**15991-15997.
27. Zhang HY, Mao J, Zhou D, Xu Y, Thonberg H, Liang Z, Wahlestedt C: **mRNA accessible site tagging (MAST): a novel high throughput method for selecting effective antisense oligonucleotides.** *Nucleic Acids Res* 2003, **31(14):**e72.
28. Zuker M: **Mfold web server for nucleic acid folding and hybridization prediction.** *Nucleic Acids Res* 2003, **31(13):**3406-3415.
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(5):**911-940.
30. Amarzguioui M, Brede G, Babaie E, Grotli M, Sproat B, Prydz H: **Secondary structure prediction and in vitro accessibility of mRNA as tools in the selection of target sites for ribozymes.** *Nucleic Acids Res* 2000, **28(21):**4113-4124.
31. Amarzguioui M, Prydz H: **An algorithm for selection of functional siRNA sequences.** *Biochem Biophys Res Commun* 2004, **316(4):**1050-1058.
32. Kretschmer-Kazemi Far R, Sczakiel G: **The activity of siRNA in mammalian cells is related to structural target accessibility: a comparison with antisense oligonucleotides.** *Nucleic Acids Res* 2003, **31(15):**4417-4424.

VITA

Vladyslava Ratushna was born on April 26, 1976 in Kyiv, Ukraine. In 1993 she finished the local high school with the Golden Medal Award, and was admitted to the National University of "Kiev-Mohyla Academy"(NaUKMA). As a sophomore, she participated in a one year exchange program with the University of Texas, Austin (UT Austin). Her undergraduate research interests concerned the effect of the radionuclide incorporation on a structure of rat bone tissue. In 1998 she graduated from NaUKMA with a bachelor's degree in Natural Sciences with a major specialization in Biology and a minor in Ecology. After that she entered a graduate program at the Department of Biophysics, Biochemistry and Molecular Biology at Iowa State University (ISU), where she worked in a *Zea mays* molecular genetics lab. Vladyslava transferred from Iowa to Virginia Tech and conducted bioinformatics research in biophysical aspects of microarray probe design at the Department of Biological Sciences. She continued her study of the effects of the target secondary structure on the quality of the microarray experiments in Dr.Gibas's lab at the University of North Carolina, Charlotte.