This is to certify that the
thesis entitled

DEVELOPMENT OF LOW-COST AND EASY-TO-USE TOOLS FOR MULTIPLEX PATHOGEN DETECTION

presented by

Dieter Maurice Tourlousse

has been accepted towards fulfillment
of the requirements for the

Ph.D. degree in Environmental Engineering

D. Maurice Tourlousse

Major Professor's Signature

8/16/2010

Date

DEVELOPMENT OF LOW-COST AND EASY-TO-USE TOOLS FOR MULTIPLEX PATHOGEN DETECTION

By

Dieter Maurice Tourlousse

A DISSERTATION

Submitted to
Michigan State University
In partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Environmental Engineering

2010

ABSTRACT

DEVELOPMENT OF LOW-COST AND EASY-TO-USE TOOLS FOR MULTIPLEX PATHOGEN DETECTION

By

Dieter Maurice Tourlousse

In developing nations, deaths associated with infectious diseases are often due to the lack of appropriate tools for diagnosis rather than effective and economical prevention and treatment options. Compared to traditional techniques such as culturing and microscopy, nucleic acid amplification testing (NAAT) is faster and more accurate, but its adoptability is limited due the high cost and complexicity of the assay. Therefore, this work aimed at developing an affordable and easy-to-use NAAT device by leveraging advances in the fields of genomics, low-cost microfluidics, and isothermal techniques for NAAT.

The complexity of pathogen detection due to the extensive genetic variability that exists for many pathogens was evaluated to substantiate the need for multiplexed detection. In this context, a virulence and marker gene (VMG) biochip was developed for parallel detection of multiple waterborne pathogens, using several VMGs per pathogen and multiple probes per VMG. While the cost of the assay was, at the time of development, considered reasonably low on a per pathogen basis, recent advances in the arena of low-cost microfluidics drastically changed this notion. Therefore, development of a disposable microfluidic chip and accompanying device for pathogen detection was undertaken as the next step in this project.

Loop-mediated isothermal amplification (LAMP) was used instead of the more common polymerase chain reaction (PCR) since the former eliminates the need for

thermal cycling while achieving a level of accuracy and rapidity that is comparable or even better than that of PCR. In terms of assay format, a 64-channel polymer microfluidic chip was designed and fabricated for parallel analysis of four samples, with each sample being tested for up to 15 genetic markers. To improve usability of the chips in the hands of untrained personnel, the chip was designed to allow dispensing of the sample using a common pipettor in a single step, after which the chip is sealable using tape. For fabrication of the chips, a technique involving hot embossing of thin polymer films using inexpensive molds and a sacrificial thermoplastic counter tool was established. Furthermore, a novel application of xurography and film lamination was developed for facile integration of tens to hundreds of microvalves and air vents for flow control. Also, by exploiting the high amplification yield of LAMP in combination with an optimal fluorescent dye, the optics could be drastically simplified. An array of individually addressable LEDs was employed to read the reaction wells with a single photodiode, without the need for mechanically moving components. This optical module was demonstrated to provide identical performance characteristics, in terms of assay accuracy and speed, with a commercial real time PCR machine but was significantly cheaper.

Based on these two key components, a fully functional prototype of the device was designed and assembled, with the necessary electronics and software being developed by other researchers involved in this project. This device and the 64-channel chip were validated for multiplexed detection of six major diarrheal pathogens. Due to its robustness, low cost and compact size, it is expected that this device could play a key role in improving human health in developing nations by providing accurate and rapid diagnostics in an affordable and easy-to-use format.

Acknowledgements

The following dissertation, while an individual work, could not have been completed without the help and support from a lot of people.

First, I would like to express my deepest gratitude to my mentor, Dr. Syed Hashsham, for his excellent guidance and providing me with a stimulating atmosphere for performing research. I will always remember and cherish his ability to get the most out of his students, while still maintaining an environment that is caring and patient. I also thank my committee members Dr. James Tiedje, Dr. Alison Cupples and Dr,. Volodymyr Tarabara for their valuable suggestions and insights.

Also, many thanks to all members of our research group, including Robert and Tiffany Stedtfeld, Gregoire Seyrig, Deric Patton, Farhan Ahmad and Samuel Baushke. They created a working environment that was not only stimulating and productive, but also enjoyable. The members of Dr. Tiedje's group also provided much help during the initial tasks in this project, and I thank them for their willingness to share their knowledge and expertise with us.

My parents and brothers always supported me and showed genuine interest in my studies. Without them, and other family members, this journey would not have been possible. Also, I want to thank my friends for the many joyous times we spend together and the support they provided during more challenging times.

And last but not least, I want to thank my wife, Maki Kosaka, for her continued love, support and understanding. I am looking forward to many joyful chapters in our life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

Microbial pathogens continue to pose a significant threat to human health globally, despite tremendous progress in the field of biomedical sciences during the last century. Worldwide, infectious diseases are responsible for nearly 15 million deaths each year, the overwhelming majority of which occur in developing nations (Morens et al., 2004). In many cases, these deaths are due to the lack of appropriate diagnosis rather than availability of effective and economical prevention and treatment options (Yager et al., 2008). In response to the need for more widely accessible diagnostics, the Bill and Melinda Gates Foundation identified 'development of technologies that allow assessment of individuals for multiple conditions or pathogens at point-of-care' as one of the Grand Challenges in Global Health (Varmus et al., 2003). Such technologies are also needed in developed nations because the number of infectious disease-related deaths in many industrialized countries, including the United States, has steadily increased since the historic low nearly three decades ago (National Intelligence Council, 2000). Various factors are responsible for this trend, including the rapid increase in the number of drug resistant pathogens, the geographic locations affected by drug resistance, and the breadth of resistance (Levy and Marshall, 2004), as well as the increased prevalence of new and re-emerging infectious diseases (Cutler et al. 2010).

Need for more affordable and simple multiplexed diagnostics

Techniques for pathogen detection need to be rapid and accurate, and also be able to screen for multiple pathogens and/or genetic markers in a single assay, which is known as multiplexed detection. Conventional techniques, such as culturing, microscopy and lateral flow test are, however, inadequate with respect to at least one of these parameters. Compared to these techniques, nucleic acid amplification testing (NAAT) is much faster, provides superior sensitivity and specificity, and is also easier to multiplex. Because of these benefits, they are now increasingly being employed in clinical laboratories in developed nations. However, NAAT is often not available in poor countries due to the high cost and complexity of the assay and required instrumentation. Therefore, an urgent need exists for devices for NAAT that are compact, robust, inexpensive and easy-to-use (Kiechle and Holland, 2009; Weigl et al., 2009). Importantly, support for development of such devices has increased in recent years due to several factors, including the increased threat of antibiotic resistance, the high cost of effective drugs, and the risk of pandemics due to globalization (Yager et al., 2008).

Enabling technologies

Advances in different fields play a key role in enabling development of devices for NAAT that are more affordable, easy-to-use and compact, without sacrificing in terms of diagnostic accuracy: genomics, low-cost microfluidics, and isothermal amplification techniques.

Genomics. Identification of novel genetic markers for diagnosis of infectious diseases and detection of pathogens is facilitated due to the exponentially increasing amount of genomic data available for many pathogenic and commensal microorganisms. This information has also revealed that extensive variability exists in both the occurrence and allele sequences of virulence genes, which are common markers for differentiating between pathogenic and non-pathogenic strains. Due to this variability, simultaneous screening for different genetic markers may be necessary for reliable detection. Furthermore, the availability of databases dedicated to virulence (Chen et al., 2005) and antibiotic resistance genes (Liu and Pop, 2009), including several in-houses databases generated by our group and maintained at the Functional Gene Pipeline Repository (http://fungene.cme.msu.edu/index.spr), also aids in identification of novel genetic markers with improved diagnostic value.

Low-cost microfluidics. Using polymers as alternatives to silicon and glass for fabrication (Becker and Gartner, 2008), the cost of microfluidic chips has been considerably reduced. Aside from their low cost and ease-of-fabrication, other benefits of polymers include their superior biocompatibility, optical transparency, and the availability of many different methods for surface treatment and modification. Furthermore, novel techniques such as xurography (Bartholomeusz et al., 2005) and lamination (Paul et al., 2007) allow rapid and inexpensive prototyping of relatively complex chips with integrated components such as valves and pumps. This progress has not only expanded the versatility of microfluidics, but also allowed researchers without access to complex fabrication facilities to develop microfluidic assays and devices.

Importantly, microfluidic chips are also attractive for multiplexed detection. The two most common genetic techniques for multiplexed detection are multiplex PCR and microarrays, but neither of these is well suited for integration in low-cost and compact devices, mainly due to the need for sensitive and multi-color optical modules, which are expensive and bulky. In microfluidic chips, multiplexing is accomplished by aliquoting the sample in a multitude of reaction wells, each containing a distinct primer set. This technique takes advantage of the significant reduction in assay volume and also the ability to employ a microfluidic network for delivering the sample to the different reaction wells. In essence, it replaces complex and expensive robotics and arduous manual dispensing. While many chips for NAAT have been developed over the years, few are sufficiently simple and robust to be used by minimally trained personnel and without bulky peripheral instrumentation. In many cases, sample dispensing and chip sealing are the main challenges that remain to be addressed.

Isothermal amplification techniques. Isothermal techniques for DNA/RNA amplification are ideal for integration in low-cost and compact microfluidic devices since thermal cycling is not necessary as the reaction proceeds at constant temperature. A number of such techniques currently exist (Gill and Ghaemi, 2008) and some of them have also been implemented in microfluidic chips. These include including helicase-dependent amplification (Ramalingam et al., 2009), recombinase polymerase reaction (Lutz et al., 2010), and loop-mediated isothermal amplification (Fang et al., 2010). Among these, LAMP has been most widely used in tube-based formats for detection of a

myriad of pathogens and is arguably also the most promising technique for integration in low cost multiplexed diagnostic devices (Mori and Notomi, 2009). LAMP employs a single enzyme along with four to six specific primers and is performed at a relatively high temperature (65°C), both of which provide the exquisite specificity of LAMP. Another salient feature of LAMP is the amount of DNA generated compared to PCR, which is much larger. This allows direct detection of the amplification based on turbidity formed as a result of pyrophosphate precipitation (Mori et al., 2001). Moreover, LAMP is not inhibited by substances that are often inhibitory to PCR (Bakheit et al., 2008; Liang et al., 2009), which simplifies sample processing.

Objectives and tasks of this project

The overall goal of this project is to develop low-cost diagnostic tools for multiplexed pathogen detection. To accomplish this goal, the following tasks were identified as part of this project:

(1) Evaluate the genomic complexicity of microbial pathogens and its impact on the selection on genetic targets for diagnostics and the need for multiplexed detection,

(2) Design and validate a virulence and marker gene biochip for parallel detection of multiple waterborne pathogens,

(3) Establish a methodology for rapid prototyping of thin film polymer chips,

(4) Evaluate the feasibility of real time fluorogenic LAMP in microfluidic chips using a simplified optical sensor that is capable of reading a multitude of reaction wells,

(5) Develop a microfluidic chip for multiplex LAMP that can be easily loaded and sealed in the hands of minimally trained users,

(6) Validate rapid and multiplexed detection of six major diarrheal bacteria using the microfluidic chip and optical sensor integrated in a fully functional prototype of a compact and inexpensive LAMP device.

With respect to the proposed instrument for LAMP, the main improvement in comparison with existing LAMP devices lies in the employment of microfluidic chips instead of conventional reaction tubes (Table 1.1). As a result of the much smaller assay volume in microfluidic chips, this significantly reduces reagent cost on a per gene basis. Furthermore, multiplexed detection is also much simplified since manual dispensing in individual tubes is not needed. Please note that the commercial instruments shown in Table 1.1. became known during the course of the project and hence the purpose of presenting a comparative evaluation is to illustrate the capabilities of the NAAT device developed as part of this project to the instruments that are available now.

TABLE 1.1 Comparison of the developed device with the Genie II and LA-200 Turbidimeter, two other commercially available instruments for real time LAMP.

Parameter	Genie II	LA-200 Turbidimeter	Developed device
Assay format	Tubes	Tubes	Microfluidic chips
Throughput	16 tubes	32 tubes	4 × 16 channels
Assay volume	10-200 μL per tube	25 μL per tube	1 μL 30 μL per sample
Assay readout	Fluorescence	Turbidimetry	Fluorescence
User interface	Touch screen	Computer	iPod Touch
Dimensions	20 × 20 × 30 cm	25 ×28 × 19 cm	< 20 × 7 × 15 cm
Weight	2 kg	5 kg	< 1 kg
Cost	Not available	$17,000	<$2000

Organization of this dissertation

This dissertation is organized into seven chapters focusing on various aspects of the overall project (Figure 1.1). Chapter One introduces the problem of multiplexed diagnostics in the context of low cost and point of use scenarios. It also sets the specifications of the device being developed and its validation focus. Chapter Two (published) presents the complexity of identifying suitable genetic markers, which illustrates the superiority of multiplexed detection compared to single gene based diagnostics. This chapter is also published and was written in the context of evaluating the relationship between the genetic content, in terms of virulence genes, of a pathogen and its health threat. Chapter Three (published) focuses on validation of the genetic markers for 12 waterborne pathogens using a coupled format of multiplex PCR and microarray hybridization on an in situ synthesized platform. This work was part of the initial work that led to the development of the low cost device being presented in this dissertation, and was co-authored with Sarah Miller. Chapter Four presents the evaluation and establishment of the two key components of the proposed LAMP device: i) low-cost polymer chips and ii) a simple optical module that is capable of reading multiple reaction wells using a single photodiode without using mechanically moving components. The results presented in this chapter provided the basis for proceeding forward with the design and fabrication of a 64-channel microfluidic chip for LAMP (presented in Chapter Five), and validation of a fully functional prototype of the device in which it can be used (Chapter Six). Electronics and software related developments were part of the doctoral dissertation of my colleague Dr. Robert Stedtfeld, and will therefore not be detailed in

this dissertation. Finally, Chapter Seven presents conclusions and accomplishments specific to this dissertation.

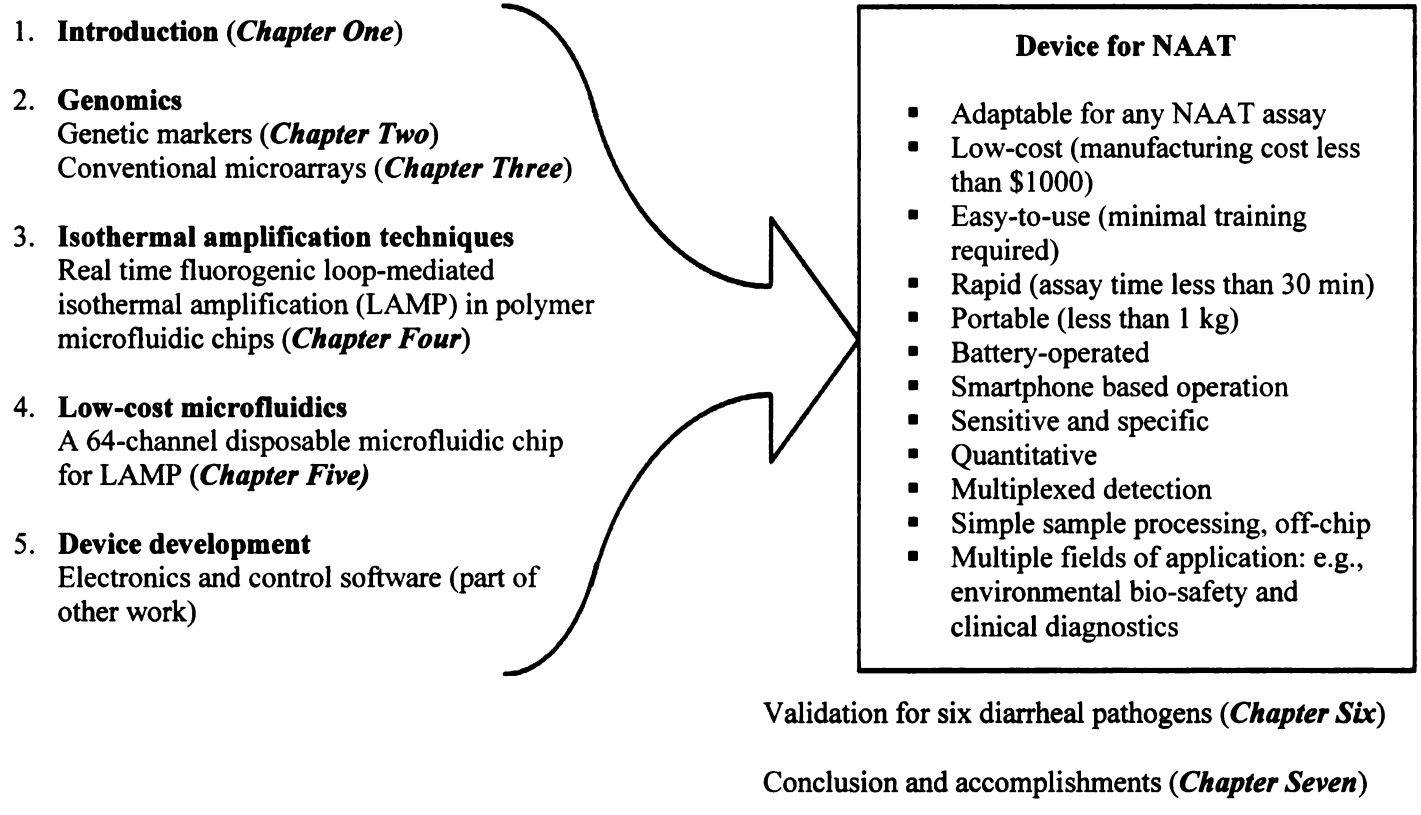


FIGURE 1.1 Content of the chapters in this dissertation and their organization.

REFERENCES

Bakheit, M.A., D. Torra, L.A. Palomino, O.M.M. Thekisoe, P.A. Mbati, J. Ongerth and P. Karanis, 2008. Sensitive and specific detection of *Cryptosporidium* species in PCR-negative samples by loop-mediated isothermal DNA amplification and confirmation of generated LAMP products by sequencing. Veterinary Parasitology, 158: 11-22.

Bartholomeusz, D.A., R.W. Boutte and J.D. Andrade, 2005. Xurography: rapid prototyping of microstructures using a cutting plotter. Journal of Microelectromechanical Systems, 14: 1364-1374.

Becker, H. and C. Gartner, 2008. Polymer microfabrication technologies for microfluidic systems. Analytical and Bioanalytical Chemistry, 390: 89-111.

Chen, L.H., J. Yang, J. Yu, Z.J. Ya, L.L. Sun, Y. Shen and Q. Jin, 2005. VFDB: a reference database for bacterial virulence factors. Nucleic Acids Research, 33: D325-D328.

Cutler, S.J., A.R. Fooks and W.H. van der Poel, 2010. Public health threat of new, reemerging, and neglected zoonoses in the industrialized world. Emerging Infectious Diseases 16: 1-7.

Fang, X.E., Y.Y. Liu, J.L. Kong and X.Y. Jiang, 2010. Loop-mediated isothermal amplification integrated on microfluidic chips for point-of-care quantitative detection of pathogens. Analytical Chemistry, 82: 3002-3006.

National Intelligence Council, 2000. The global infectious disease threat and its implications for the United States. Downloaded in July 2010 from http://www.dni.gov/nic/special_globalinfectious.html.

Gill, P. and A. Ghaemi, 2008. Nucleic acid isothermal amplification technologies - a review. Nucleosides Nucleotides & Nucleic Acids, 27: 224-243.

Kiechle, F.L. and C.A. Holland, 2009. Point-of-care testing and molecular diagnostics: miniaturization required. Clinics in Laboratory Medicine, 29: 555.

Levy, S.B. and B. Marshall, 2004. Antibacterial resistance worlwide: causes, challenges and responses. Nature Medicine 10: S122-S129.

Liang, S.Y., Y.H. Chan, K.T. Hsia, J.L. Lee, M.C. Kuo, K.Y. Hwa, C.W. Chan, T.Y. Chiang, J.S. Chen, F.T. Wu and D.D. Ji, 2009. Development of loop-mediated

isothermal amplification assay for detection of *Entamoeba histolytica*. Journal of Clinical Microbiology, 47: 1892-1895.

Liu, B. and M. Pop, 2009. ARDB - Antibiotic Resistance Genes Database. Nucleic Acids Research, 37: D443-D447.

Lutz, S., P. Weber, M. Focke, B. Faltin, J. Hoffmann, C. Muller, D. Mark, G. Roth, P. Munday, N. Armes, O. Piepenburg, R. Zengerle and F. von Stetten, 2010. Microfluidic lab-on-a-foil for nucleic acid analysis based on isothermal recombinase polymerase amplification (RPA). Lab on a Chip, 10: 887-893.

Morens, D.M., G.K. Folkers and A.S. Fauci, 2004. The challenge of emerging and re-emerging infectious diseases. Nature, 430: 242-249.

Mori, Y., K. Nagamine, N. Tomita and T. Notomi, 2001. Detection of loop-mediated isothermal amplification reaction by turbidity derived from magnesium pyrophosphate formation. Biochemical and Biophysical Research Communications, 289: 150-154.

Mori, Y. and T. Notomi, 2009. Loop-mediated isothermal amplification (LAMP): a rapid, accurate, and cost-effective diagnostic method for infectious diseases. Journal of Infection and Chemotherapy, 15: 62-69.

Paul, D., A. Pallandre, S. Miserere, J. Weber and J.L. Viovy, 2007. Lamination-based rapid prototyping of microfluidic devices using flexible thermoplastic substrates. Electrophoresis, 28: 1115-1122.

Ramalingam, N., T.C. San, T.J. Kai, M.Y.M. Mak and H.Q. Gong, 2009. Microfluidic devices harboring unsealed reactors for real-time isothermal helicase-dependent amplification. Microfluidics and Nanofluidics, 7: 325-336.

Varmus, H., R. Klausner, E. Zerhouni, T. Acharya, A.S. Daar and P.A. Singer, 2003. Grand challenges in global health. Science, 302: 398-399.

Weigl, B.H., D.S. Boyle, T. de los Santos, R.B. Peck and M.S. Steele, 2009. Simplicity of use: a critical feature for widespread adoption of diagnostic technologies in low-resource settings. Expert Review of Medical Devices, 6: 461-464.

Yager, P., G.J. Domingo and J. Gerdes, 2008. Point-of-care diagnostics for global health. Annual Review of Biomedical Engineering, 10: 107-144.

CHAPTER TWO

VIRULENCE FACTOR ACTIVITY RELATIONSHIPS (VFAR): CHALLENGES AND DEVELOPMENT APPROACHES

Tourlousse, Dieter M.; Stedtfeld, Robert D.; Baushke, Samuel W.; Wick, Lukas M.; Hashsham, Syed A. (2007) Water Environment Research 79(3):246-59.

INTRODUCTION

Virulence factor activity relationships (VFAR) is a predictive approach proposed by the National Research Council's (NRC's) Committee on Drinking Water Contaminants (Washington, D.C.) to classify and rank waterborne pathogens (NRC, 1999). Classification and prioritization of chemical and microbial contaminants is necessary to develop a contaminant candidate list (CCL), which serves as the basis for regulation (U.S. EPA, 2005). In essence, VFAR is loosely based on the concept of quantitative structure activity relationships, which have been used for many years to predict health effects of chemical substances and aid in their regulation (Cronin et al., 2003). As such, VFAR is expected to integrate the growing knowledgebase in the area of genomics (and other levels of cellular organization, depending on their availability) of pathogenic microorganisms with clinical information, and combine this with parameters related to the host and environment, to predict potential health risks associated with waterborne microorganisms. Because of the availability of information, it is anticipated that VFAR will first be developed at the genomic level. Later, it will be extended to include information from transcriptomics, proteomics, and phenomics.

To begin the process, the NRC Committee suggested that all pathogenic microorganisms may be described in terms of descriptor variables and response or outcome variables. Descriptor variables are those that may prove useful in predicting the response of the microorganisms. Examples of descriptor variables related to the microorganisms include toxins, adhesion and invasion factors, and protein secretion pathways. Response variables are virulence, potency, and persistence. Measures of response variables include minimum infectious dose (or other measures of dose-response relationships), mortality, morbidity, survival time in the environment, and disinfection kinetics. VFAR is then the mathematical model or algorithm that links descriptor variables to response variables in a quantitative and predictive manner. It was suggested that, for the purpose of developing VFAR, the definition of virulence might even be expanded to include clinical virulence (severity of the disease), pathogenicity (the ability to cause disease), and persistence (survival time) of the microorganism in the environment. The committee substantiated the feasibility of the VFAR concept by stating the following:

- VFAR is scientifically valid and robustly applicable;
- VFAR will be able to extend a known relationship between a virulence attribute and human disease to a situation in which the attribute is found in a new or unexpected circumstance or in microorganisms that have not been heretofore recognized as potentially pathogenic;
- VFAR will account for the likelihood that adverse human outcomes (i.e., disease) will continue to be discovered in association with the action or

presence of a microbial contaminant, microbial gene, or gene product in the clinical setting;

- VFAR is congruent with the direction that other government, private, and public agencies are taking; and
- the development of VFAR will satisfy the requirements that development of a CCL be systematic, scientifically sound, transparent, and involve broad public participation.

Although the above aspects of VFAR were highlighted to support its feasibility, they may also serve as guidelines during its development. A number of exploratory projects have been initiated by the U.S. Environmental Protection Agency (Washington, D.C.) to identify data availability and gaps for the development of VFAR. Evidently, VFAR needs to capture the widest possible range of molecular and clinical information related to waterborne pathogens, encompassing bacteria, viruses, and protozoa, and synthesize this information to enable a more systematic development of a microbial CCL. It will most likely require extensive new analysis, research, and development. The ultimate goal of VFAR is to develop predictive capabilities for waterborne pathogens by linking response to descriptor variables. Because of the inherent complexity and multifunctional nature of virulence factors, the American Academy of Microbiology (Washington, D.C.) report suggested that VFAR, when developed, must be used with considerable caution (Cangelosi et al., 2004).

Following the NRC report, the CCL Classification Work group of the National Drinking Water Advisory Council (NDWAC) (Washington, D.C.) suggested prioritizing waterborne pathogens by evaluating them for five attributes-potency, severity, prevalence, persistence or mobility, and magnitude (NDWAC, 2004). According to this report, potency, or the amount of contaminant required to cause illness, could be scored between I and 11, with the most potent pathogens being those that are known to cause a high rate of morbidity and water-related disease in healthy individuals (score 11), and the least potent being those for whom genetic sequences are available in searchable databases but with no documented virulence genes (score 1). Severity, or the seriousness of the health effects, could be scored with the summation of responses (yes = 1, no - 0) to a series of 12 carefully constructed questions based on morbidity and mortality, location and intensity of the infectious processes, extent of infection, time lost to illness, requirement of medical intervention, and associated chronic manifestation or disabilities. The total score (0 to 12) would then represent the severity of the adverse health effects caused by the microorganism. Prevalence, or the frequency of the microbial contaminant occurring in drinking water, could be scored with the most prevalent pathogens being those that are detected in drinking water (given a score of 7), and the least prevalent being those that are not detected in drinking water and have a narrow host range, limited primarily to humans (given a score of 1). Persistence/ mobility describe the potential for amplification under ambient conditions, sedimentation and absorption characteristics, and survival characteristics. It could be scored between 2 to 5, with the highest scoring pathogens being those that are stable for weeks or longer and are able to amplify or are protected from symbiotic relationships, and the lowest scoring being those that typically

die rapidly in water (i.e., in days). Magnitude is the contaminant concentration relative to the level that causes a perceived health effect. It could be scored between 0 and 6, with the highest scoring pathogens being those that have caused numerous recently documented waterborne disease outbreaks in the United States or other developed countries, and the lowest scoring being those that have never caused a waterborne disease outbreak anywhere or its biological properties diminish its ability to do so. The NDWAC report suggests selecting pathogens for the CCL using a "prototype classification" algorithm based on the above attributes. This classification scheme represents a significant first step in making the CCL process transparent and systematic and provides guidelines to extract usable and quantifiable data from an extremely complex and dynamic research area. Such classification of pathogenic attributes can also serve as the basis for systematic VFAR development. The VFAR models should enable the prediction and comparison of attributes, such as potency, severity, and persistence, based on comprehensive analysis of descriptor variables. Prevalence and magnitude can then be accessed through sampling and monitoring of water and outbreak samples.

The primary step in the VFAR development process is the collection of information related to descriptor and response variables. Unfortunately, this information is scattered in various searchable and non-searchable databases and journal articles, mostly in a form that necessitates further processing. It is also becoming evident that environment and host-related factors are as critical as pathogen-related descriptors in predicting response variables and therefore must be considered among the descriptor variables. A VFAR algorithm must also account for the quality and availability of data, and biological

variability. Specifying probability distributions for the input descriptor variables and carrying these uncertainties through to the predicted outcomes can accomplish this.

This paper summarizes some of the challenges that must be addressed during the collection, interpretation, and measurement of information related to VFAR, with a focus on genetic descriptors. However, similar care may need to be taken for descriptors at higher levels of cellular organization. This paper also proposes three critical elements of the VFAR development and validation process. The first is the construction of a comprehensive VFAR database housing all information related to descriptor and response variables. The second is the development of mathematical and statistical models to synthesize this information and relate response to descriptor variables. A Bayesian approach is discussed as an example of the type of modeling tools more suitable for VFAR development. The third element is a high-throughput monitoring strategy to identify the prevalence of genetic descriptors of virulence in drinking water supplies and sources. On-chip polymerase chain reaction (PCR) is presented as an economical means to monitor hundreds to thousands of genetic descriptors at a specificity and sensitivity level that is relevant to waterborne pathogens.

CHALLENGES IN DEVELOPING VFAR

The NRC report defined VFAR in very broad terms, perhaps to allow multiple interpretations and comprehensive development of the concept; this itself poses a challenge. Interpreted broadly, VFAR must synthesize virulence-associated information at all levels of cellular organization, for all classes of waterborne pathogens

(encompassing bacteria, protozoa, and viruses), and combine this with relevant host- and environment-related factors. It might even be proposed that such a comprehensive analysis would contribute to our understanding of the emergence of new waterborne pathogens. In narrower interpretations, VFAR should at least be able to prioritize selected waterborne pathogens through comparison, with an initial emphasis on comparative genomics and descriptors of human health. As evident from Figure 2.1, this paper employs the former interpretation. It illustrates that a broad set of descriptor variables (associated with the microorganism, host, and environment) determines the response (virulence, potency, and persistence). The predicted response has an associated uncertainty that is the result of biological variability and uncertainties associated with descriptor and response variables. Response variables then serve as the basis for ranking pathogens within the CCL and assigning new pathogens to the CCL for regulatory purposes.

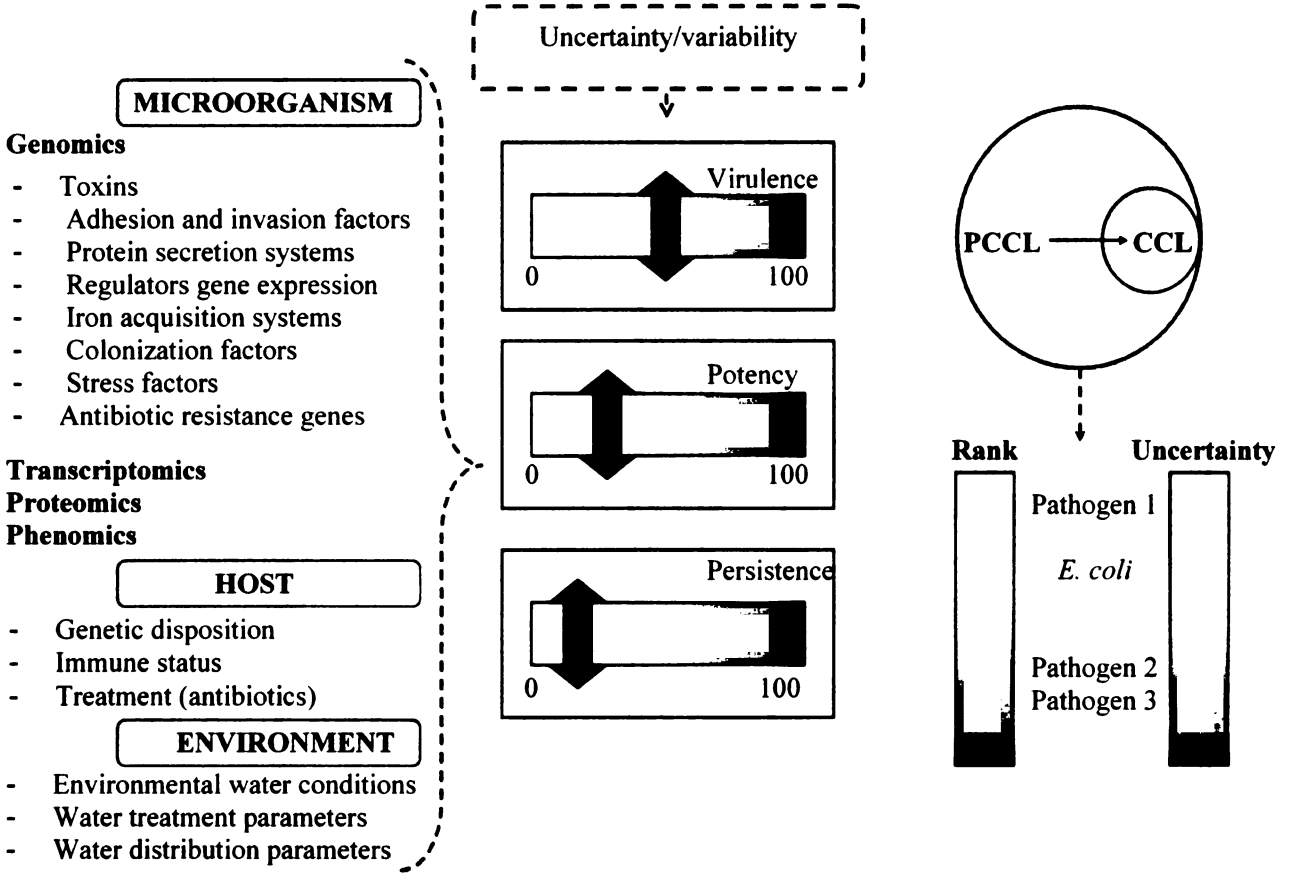


FIGURE 2.1 Conceptual depiction of the components of a VFAR model. It assimilates descriptor variables related to the microorganism, host, and environment to predict response variables (potency, persistence and virulence). Predictions have an associated error term because of uncertainty and biological variability in descriptor and response variables. Assigning new pathogens to the CCL and ranking of pathogens within the CCL are based on the predicted response variables. Images in this thesis/dissertation are presented in color.

Because virulence genes represent one of the major descriptors and arguably the first step in VFAR development, virulence gene based VFAR is used as an example to illustrate some of the challenges that must be addressed during collection and interpretation of the relevant information. First, the genetic variability that exists among pathogens and the challenges this poses for VFAR development are discussed. Subsequently, the limitations of incorporating genes involved in environmental persistence are pointed out. The growing understanding that many virulence factors are multi-functional and can also be found in nonpathogenic microorganisms is also discussed. The importance of incorporating host specific genes responsible for disease susceptibility or resistance as descriptors in VFAR is also illustrated. Finally, the challenges associated with the measurement of genetic descriptors, with an emphasis on allele discrimination using hybridization-based technologies, are presented. Without addressing these and similar issues in other descriptors, it may be difficult to develop a robust and scientifically sound predictive tool.

Information related to ecology and epidemiology of the microorganisms also needs to be integrated. Ecological principles include interactions between microorganisms and the biotic and abiotic environment. For example, the presence of the protozoon *Acanthamoeba astronyxis* significantly increased the resistance of *Burkholderia pseudomallei* to disinfection (Howard and Inglis, 2005). A similar protective bacterial-protozoan interaction was also observed for *Legionella pneumophila* and *Hartmannella vermiformis* (Donlan et al., 2005). Similarly, epidemiological studies provide insights to contamination sources, exposure risks, and susceptibility among populations and

individuals. For example, epidemiological studies recognized the consumption of untreated drinking water from private wells and water sources as a risk factor for the acquisition of *Campylobacter* infections (Said et al., 2003). Outbreaks of *L. pneumophila* infections have been linked to contaminated aerosol-producing devices, such as cooling towers and fountains (Hlady et al., 1993; Brown et al., 1999).

Genetic diversity

The main premise of VFAR is that the virulence and potency of a microorganism can be linked to its genetic and other descriptors. Although a vast amount of evidence is available to support this hypothesis, VFAR based on genomics alone still faces major challenges. In particular, the extensive genetic diversity that exists among pathogens has important implications for their development and interpretation. Although the studies discussed below demonstrate potential correlations between information on a particular virulence gene and the outcome of an infection, it should be recognized that virulence genes cannot be seen as independent entities and need to be evaluated within the context of the whole genome. This means that the sequence or occurrence of a particular virulence gene alone is not sufficient to predict the virulence and potency of a strain and will need to be complemented with information on other virulence genes and, if possible, the entire genome.

Diversity in Gene Content. Virulence genes are not always evenly distributed among various strains of a pathogenic species. For example, Thong et al. (2005) investigated the presence of four virulence genes (*setlA*, *setlB*, *ial*, and *ipaH*) among 110 clinical isolates

of *Shigella* spp. and observed that three of the genes were present in 45 of the tested isolates, while the other was detected in all of them. A similar variability in the distribution of virulence genes among clinical isolates has been described for *Staphylococcus aureus* (Becker et al., 2004); *Aeromonas* spp. (Chacon et al., 2004); *Campylobacter jejuni* (Rozynek et al., 2005); *Pseudomonas aeruginosa* (Feltman et al., 2001); and *Escherichia coli* (Blanco et al., 2004). Table 2.1 illustrates this point by listing the incidence of 12 selected virulence genes among 9 fully sequenced genomes of *E. coli* and *Shigella*. The presence or absence of specific virulence genes translates into different pathotypes and should be accounted for when developing VFAR. In the case of *E. coli*, the existence of specific combinations of virulence genes has been exploited to differentiate between pathotypes (Bekal et al., 2003). This study also revealed that some *E. coli* isolates may carry previously unknown combinations of virulence genes, which could lead to the emergence of novel pathotypes.

The main challenge of VFAR is to correlate information on the genetic make-up of a microorganism to its virulence and potency. Many cases have been described in which the pathogenic trait of a particular strain can be linked to its virulence gene content. In the case of *Helicobacter pylori*, for example, the most severe disease outcome is commonly associated with strains harboring a specific DNA region encoding the immunogenic protein cagA and a type IV secretion system (Censini et al., 1996; Nilsson et al., 2003). Similarly, the presence of the *stx2* and *eae* genes in *E. coli* strains has been reported to correlate with the severity of disease outcomes in humans (Boerlin et al., 1999; Beutin et al., 2004). Such examples demonstrate our growing knowledge about the basis of

interstrain variability in virulence and potency. However, the origin of this variability remains largely elusive in many other pathogenic species. For example, in *Listeria monocytogenes*, major virulence factors (internalin A and B, listeriolysin, phospholipase A and B, the actin-assembly inducing protein actA, and the gene expression regulator prfA) are conserved in both virulent and in less virulent strains (Doumith et al., 2004). Analysis of the presence or absence of these genes is therefore insufficient to estimate the pathogenic threat associated with a given strain.

In addition to the unequal distribution of known virulence genes, whole genome comparisons highlighted that the overall diversity in gene content among closely related strains may be significant. For example, Salama et al. (2000) compared the genomes of 15 clinical isolates of *H. pylori* and observed that up to 18% of genes were specific to the genome of a given strain. Similarly, Pearson et al. (2003) observed that approximately 16% of the genes present in the sequenced strain of *C. jejuni* (strain NCTC111168) were either absent or divergent among 18 strains from different sources. Underlining the main premise of VFAR, genome comparisons of more and less potent strains can also lead to the identification of unknown genes that might explain the difference in virulence (Whittam and Bumbaugh, 2002), as evidenced for *L. monocytogenes* (Doumith et al., 2004), *C. jejuni* (Poly et al., 2004), and *Leptospira interrogans* (Nascimento et al., 2004). Clearly, more research is needed to elucidate the contributions of those newly identified genes to virulence and VFAR should allow for this uncertainty due to incomplete knowledge.

TABLE 2.1 Incidence of selected virulence genes in several fully sequenced strains of *E. coli* and *Shigella*.

Strain	Type	*eae*	*stx1*	*stx2*	*set1A*	*senA*	*ehxA*	*hylA*	*espP*	*sat*	*iucC*	*shu/chu*	*ipaH*
E. coli K12 MG1655	-	-	-	-	-	-	-	-	-	-	-	-	-
E. coli W3110	-	-	-	-	-	-	-	-	-	-	-	-	-
E. coli O157:H7 EDL 933	EHEC	•	•	•	-	-	•	-	•	-	-	•	-
E. coli O157:H7 Sakai	EHEC	•	•	•	-	-	•	-	•	-	-	•	-
E. coli CFT073	UPEC	-	-	-	•	-	-	•	-	•	•	•	-
S. flexneri 2a str. 301	EIEC	-	-	-	•	•	-	-	-	-	•	-	•
S. sonnei Ss046	EIEC	-	-	-	-	•	-	-	-	-	•	-	•
S. dysenteriae Sd197	EIEC	-	•	-	-	•	-	-	-	-	-	•	•
S. boydii 4Sb227	EIEC	-	-	-	-	•	-	-	-	-	•	-	•

• present; - absent. EHEC: enterohemorrhagic *E. coli*; UPEC: uropathogenic *E. coli*; EIEC: enteroinvasive *E. coli*. Genome sequences were retrieved from GenBank and analyzed for the presence of 12 selected virulence genes using Blast search.

Allelic Diversity. Sequences of a virulence gene may vary considerably among different isolates. This is known as allelic variability and has been documented for many virulence genes, including the *vacA* gene of *H. pylori* (Atherton et al., 1995); the *trh* gene of *Vibrio parahaemolyticus* (Kishishita et al., 1992); the *eae* gene of *E. coli* (Adu-Bobie et al., 1998; Zhang et al., 2002); the *hly* and *actA* genes of *L. monocytogenes* (Jeffers et al., 2001); and the *prn* and *ptx* genes of *Bordetella pertussis* (Cassiday et al., 2000). Table 2.2 illustrates this point by summarizing the sequence variability among a number of selected virulence genes. Sequences were manually harvested from GenBank using the Basic Local Alignment Search Tool (Blast), aligned and analyzed using MEGA (Kumar et al., 2004). It can be seen that the maximum sequence diversity among the analyzed sequences ranges from 20% for the *eae* gene of *E. coli* to 1.6% for the *toxR* gene of *V.*

parahaemolyticus. The average nucleotide diversity of these two genes is 14 and 0.67%, respectively. These need to be interpreted with caution, however, because the allelic diversity might be underestimated when the number of available sequences is limited or has been collected in limited number of studies. Figure 2.2 displays the phylogenetic neighbor-joining trees (Saitou and Nei, 1987) of *toxR* of *V. parahaemolyticus*, *iap* of *L. monocytogenes*, and *eae* of *E. coli*, demonstrating low, medium, and high allelic diversity, respectively. Differences in gene sequence can result in significant differences in virulence, and potency, indicating that variability in virulence gene sequence should be assessed while developing VFAR.

In a number of studies, a correlation has been observed between the genotype of a virulence gene and the clinical outcome of infection. Atherton et al. (1995), for example, demonstrated a correlation between the presence of the *vacA s1* genotype and the development of peptic ulcer disease resulting from *H. pylori*. A similar correlation between the presence of specific *stx2* gene variants and the risk for the development of hemolytic-uremic syndrome has been documented for *E. coli* (Friedrich et al., 2002). However, a correlation between the genotype of a particular virulence gene and the severity of the disease is not always supported. In the case of the *vacA* gene, Tan et al. (2005) could not establish an association between the *vacA* genotype detected in patients and the clinical outcome of infections. Information about the frequency of presence or absence of such relationships is not yet available.

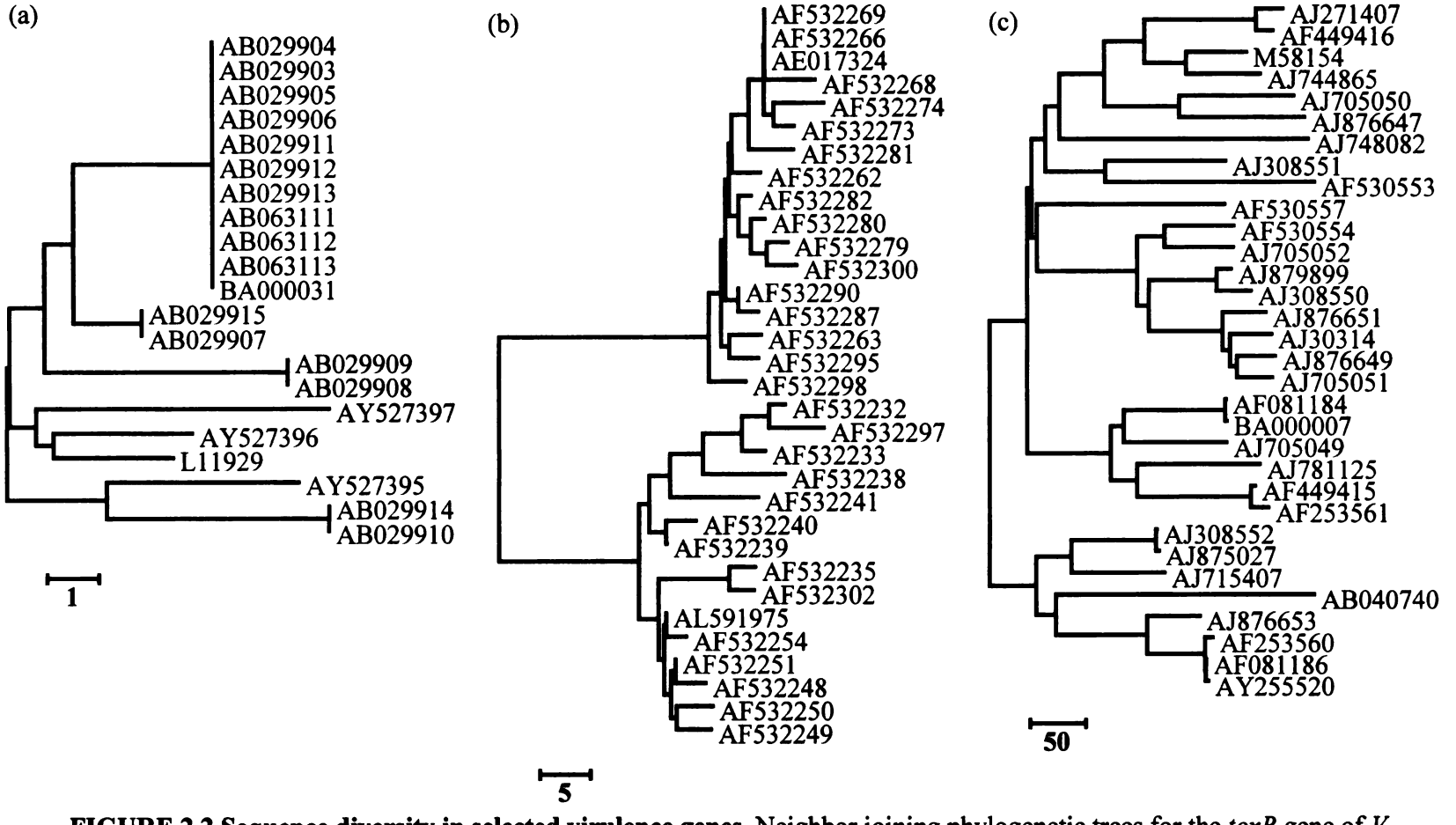


FIGURE 2.2 Sequence diversity in selected virulence genes. Neighbor joining phylogenetic trees for the *toxR* gene of *V. parahaemolyticus* (a), the *iap* gene of *L. monocytogenes* (b), and the *eae* gene of *E. coli* (c). The branch length represents the number nucleotide differences. Note the different scale among the trees.

TABLE 2.2 Sequence diversity statistics for selected virulence genes.

Pathogen	Gene	No. of sequences	No. of sites	Average diversity		Maximum diversity	
				No.	%	No.	%
E. coli	*eaeA*	32	2771	380	14	562	20
H. pylori	*vacA*	77	3756	306	8.1	602	16
H. pylori	*cagA*	89	3090	220	7.1	388	13
E. coli	*stx1A*	41	942	22.7	2.4	77	8.2
H. pylori	*ureA*	36	717	17.0	2.4	37	5.2
L. monocytogenes	*iap*	32	1395	31.3	2.2	64	4.6
L. monocytogenes	*plcB*	116	841	20.2	2.4	37	4.4
V. parahaemolyticus	*tdh*	20	570	10.8	1.9	20	3.5
C. perfringens	*plc*	18	1197	14.9	1.2	28	2.3
V. cholerae	*ctxA*	30	777	1.71	0.22	17	2.2
V. cholerae	*ctxB*	33	375	2.64	0.7	7	1.9
V. parahaemolyticus	*toxR*	21	879	5.93	0.67	14	1.6

Genes are sorted according to decreasing maximum sequence diversity. The number of sites indicates the number of sites for comparison after exclusion of gaps and missing data. Sequences were retrieved from GenBank using Blast search, aligned and analyzed using MEGA. Phylogenetic trees for the genes that are underlined are depicted in Figure 2.2.

A preliminary analysis to rank 15 waterborne microorganisms, including the bacterial members of the current CCL, was performed in this study using an in-house virulence and marker gene (VMG) database of close to 3000 sequences encompassing more than 90 pathogens collected to develop a pathogen biochip (Miller et al., 2008). Each genome was screened for the presence of virulence genes in the VMG database using Blast search. Bit scores (a measure of the length and quality of the alignment obtained using Blast) were recorded for each virulence gene hit. All hits were sorted according to the bit score for each genome and plotted in Figure 2.3. The x-axis is simply the hit number with

a different identity for each genome. It can be observed that (1) bit scores for pathogens start high and remain so for at least 10 to 20 sequences, (2) bit scores for non-pathogens start at low values and remain so, and (3) ranking solely based on genomic comparisons will be a function of bit score (i.e., it will depend on the length of sequence string being compared).

This analysis was undertaken to demonstrate that, when all the pertinent information about descriptor and response becomes available, a VFAR-based ranking should be feasible. It also illustrates that the power of the database to identify and rank a pathogen will depend on its comprehensiveness and accuracy--traits that are not necessarily compatible when it comes to the mode of collecting data (automated *versus* manual). An example of how a VFAR database may fail is illustrated by evaluating *Anabaena variabilis*, which is a CCL organism against the VMG database in the above analysis. The VMG database did not contain any virulence genes from this organism. Because the virulence genes of *A. variabilis* were missing from the database, the analysis ranked it with *Lactobacillus acidophilus*, a non-pathogen located closest to the x-axis (Figure 2.3). Thus, information related to descriptor and response variables must be comprehensive, but carefully, evaluated for inclusion in VFAR database to ensure accurate ranking.

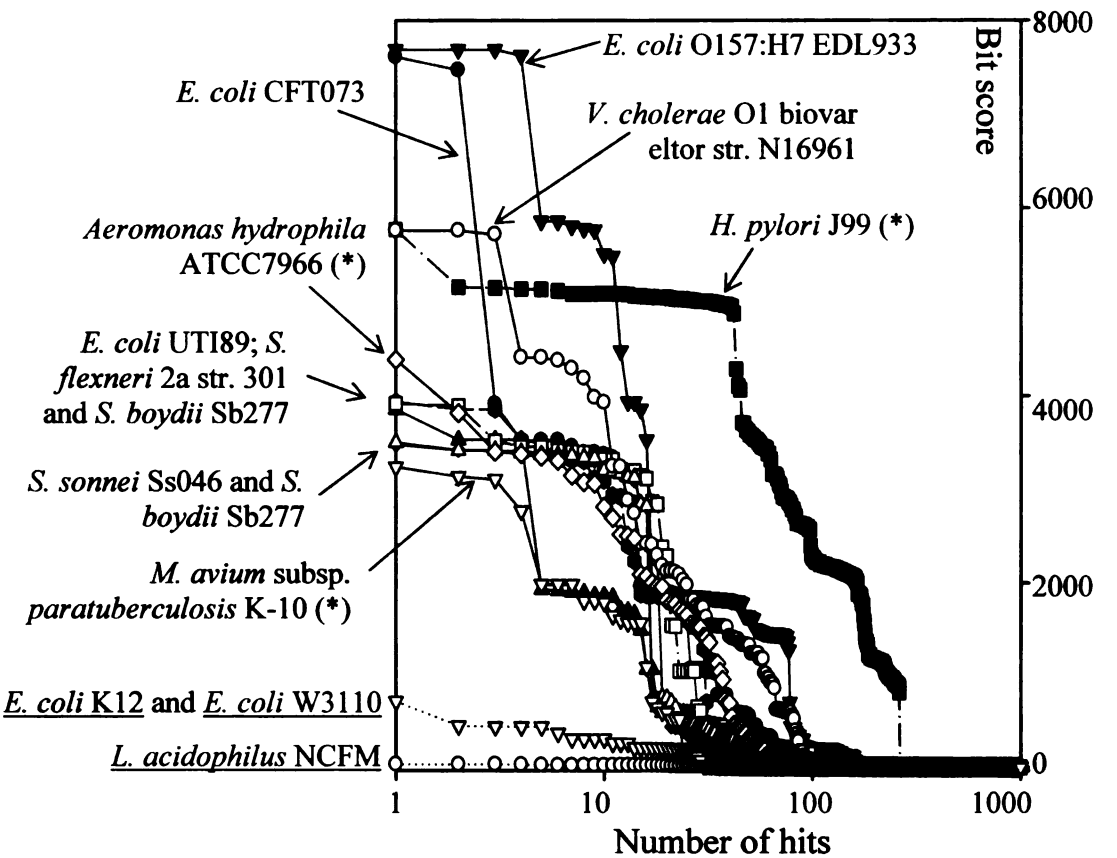


FIGURE 2.3 Ranking of 15 waterborne bacteria based on preliminary analysis of their genomic contents. Genomes of the analyzed bacteria were screened for the presence roughly 3000 virulence gene sequences encompassing more than 90 bacterial pathogens using Blast search, and bits scores were obtained for each gene in the database. The bit scores (*y-axis*) are plotted in descending order (*x-axis*) for each genome. Only the highest 1000 bit scores are shown. Non-pathogenic bacteria are underlined, and pathogens in the current CCL are marked (*).

Genetic descriptors of persistence

The survival or persistence of pathogens outside the host is an important response variable. Pathogens released in the environment are exposed to stressful conditions different from those encountered during host infection. Low temperatures; solar irradiation; and competitive, antagonistic, and predatory interactions with indigenous microbial populations are only few of these stresses. Pathogens and other microorganisms

cope with this stress by the induction of a number of survival mechanisms, such as entering the viable but non-cultivable state, spore and cyst formation, induction of the starvation-stress response, and formation of biofilms on abiotic and biotic surfaces. The molecular mechanisms and genes involved in stress and survival responses are only beginning to be understood. This currently limits the extent to which descriptors of persistence can be included in VFAR. More information about descriptors is required to predict the survival mechanisms of microorganisms in the environment.

Virulence genes in non-pathogenic microorganisms

Pathogenic and nonpathogenic microorganisms have developed similar molecular mechanisms to deal with adverse conditions, and genes encoding these stress-related factors can be found in both. Most studies on interactions between microorganisms and human hosts have initially focused on pathogens. Many genes identified in these studies were thought to play a critical role in pathogenicity and therefore called virulence genes. It is now becoming clear that a number of these genes are probably involved in more general interactions between microorganisms and hosts and can also be found in non-pathogens (Holden et al., 2004). Grozdanaov et al. (2004) analyzed the genome of the probiotic *E. coli* strain Nissle 1917 and observed that a number of genes (i.e., genes encoding adhesion factors, proteases, and iron acquisition systems), thought to be linked to the pathogenicity of *E. coli*, are also present in this nonpathogenic strain. More potent virulence factors, such as alpha-hemolysin and P-fimbrial adhesions, were not detected and explained the nonpathogenic nature of *E. coli* strain Nissle 1917. Zhang et al. (2003) documented that the majority of adhesion genes previously described in clinical isolates

of *S. epidermis* were also present in the non-infection-associated strain ATCC12228. More potent exotoxins and exoenzymes, such as the alpha-hemolysin, toxic shock syndrome toxin, staphylokinase, and hyaluronidase, were not found. Other factors, such as a number of proteases and nucleases and regulatory proteins, were detected in *S. epidermis* strain ATCC12228. The detection of genes previously linked to pathogenicity in nonpathogenic strains indicates that these genes contribute to general adaptability and fitness and are not solely associated with virulent traits. In developing VFAR, these factors not directly involved in pathogenicity should be treated differently than more potent virulence factors.

Another complexity is the detection of partial virulence gene arrays in nonpathogens. The Brazilian National Genome Project Consortium (2003) described the presence of genes encoding a type III secretion system in *Chromobacterium violaceum*, which is only slightly virulent to humans. Detailed analysis revealed that some key genes in the type III secretion system along with other pathogenicity related genes were missing. Similarly, Nilsson et al. (2003) observed that less virulent strains of *H. pylori* harbored a partially deleted *cag* pathogenicity island.

The above-described findings obviously have important implications for VFAR, and care must be taken when attempting to predict the virulence of a microorganism based on information on a limited number of virulence genes. Genes encoding the most toxic virulence factors should be given higher priority, because those are anticipated to be

absent in non-pathogens. Similarly, presence of most or all genes encoding a functional unit instead of a limited number of genes within the unit should be given a higher weight.

Effect of host genetics

The host plays a major role in determining the outcome of an infection. Variability in disease susceptibility and resistance has been observed at both the individual and population level. In the case of *H. pylori*, infections resulted in severe disease development in 10% of all infected individuals (Nguyen et al., 1999). A similar host-related variability in disease development has been documented for intracellular pathogens, including *Salmonella*, *Mycobacterium tuberculosis*, and *L. monocytogenes* (Ottenhoff et al., 2002; van de Vosse et al., 2004). A major part of the host's disease susceptibility and resistance is inheritable and has been linked to mutations in specific host genes. For example, increased susceptibility to infections with pathogens, such as *L. monocytogenes* and *S. enterica* serovar Typhimurium, has been associated with mutations in genes that encode major proteins in the type 1 cytokine cascade, such as interferon-gamma and its receptor (Ottenhoff et al., 2002). *P. aeruginosa* only causes disease in patients with cystic fibrosis, and this has been explained by mutations in the cystic fibrosis conductance regulator gene (Pier, 2002). Increased risk of developing gastric carcinoma resulting from *H. pylori* infection is linked to pro-inflammatory polymorphisms in the interleukin-1 gene cluster (Correa, 2005). Knowledge on the molecular mechanisms and genes involved in disease resistance and susceptibility is still limited. New information is rapidly becoming available, and incorporating these data to

the VFAR algorithm is expected to reduce the uncertainties associated with its predictive ability.

Measurement of genetic descriptors

Genetic descriptors of virulence can be measured using different molecular tools. In light of the above discussion, sequencing of all waterborne pathogens and related strains would be the best option. However, until the cost of sequencing and closing whole genomes is reduced to only a few thousand dollars per genome, it can only be used selectively. Sequencing by hybridization may become a more economical option, but is not yet a mature technology. Microarrays and high-throughput PCR are the two main approaches to evaluate the presence of thousands of genetic descriptors and their alleles. Ideally, a tool for the measurement of genetic descriptors for VFAR should be able to discriminate single nucleotide polymorphisms among different alleles. The specificity of microarrays and PCR, however, is lower than that of sequencing, because hybridization-based methods are prone to cross-hybridization. When using microarrays, finding specific probes that are able to discriminate between closely, related alleles is not trivial, as illustrated by the following discussion.

The ability to detect a single base-pair mismatch in a probe-target duplex on a microarray strongly depends on the position and type of the mismatch (Wick et al., 2006). Generally, mismatches in the center of a probe-target duplex are easier to detect than mismatches towards the ends of the duplex (Wick et al., 2006). Therefore, known single nucleotide polymorphisms should be placed in the center of a probe. This imposes

further restrictions on probe design, in addition to conditions that must be fulfilled, including the probe's guanine and cytosine (G+C) content, secondary structure, and target range, making probe design more cumbersome. Furthermore, microarray-based differentiation of nucleotide sequences is based on comparison of signal intensities. This makes it necessary to include probes for all possible or at least all known sequence polymorphisms on the microarray. New sequences not represented by probes on the microarray cannot be detected. The rapidly increasing number of allelic types for many virulence genes shows that a VFAR detection method must also be able to specifically identify and discriminate related allelic sequences.

KEY ELEMENTS OF VFAR

Databases of descriptor and response variables

The first step in VFAR development is active collection, manipulation, and storage of available data related to descriptor and response variables. However, this information is highly scattered and in formats that cannot be used without processing. The three categories of information related to descriptor variables that must be collected include (1) pathogen-related descriptors, with initial focus at the genomic level; (2) descriptor variables related to host; and (3) descriptors related to the environment. Information related to response variables includes virulence, potency, and persistence of the microorganism (i.e., minimum infectious dose, mortality, and survival time in the environment). Multiple datasets should be consulted to capture the variability in these descriptors. Importance of descriptor variables in predicting the outcome and uncertainties associated with their measurement should also be specified. All this

information should be housed in a flexible and interactive database, which allows integration and correlation of all relevant data. Information about outbreaks and sequence-based tracking approaches should also be included to verify the validity of the descriptors. An overview of virulence genes of selected pathogens in our database is presented in Table 2.3. Virulence genes can be categorized in three functional classes (Wassenaar and Gaastra, 2001). Class I consists of classical virulence genes (i.e., genes coding for toxins, adhesion, and invasion factors). These genes are directly involved in host-pathogen interactions and responsible for host damage and expected to be absent in nonpathogenic organisms. Class II consists of genes that encode for factors essential for the activity of the classical virulence factors (i.e., secretion systems and regulators of gene expression). Class III consists of genes related to the specific life style of a pathogen (i.e., genes coding for iron uptake mechanisms and factors implicated in host colonization and evasion of the host's immune response). More elaborate classification schemes, reflecting the probability of the microorganism to cause disease, may also be used (Wassenaar and Gaastra, 2001). Such refined categorization can serve as the basis for assigning weights to specific virulence genes in VFAR.

Genes responsible for antibiotic resistance also need to be incorporated to such databases. This is because antibiotic resistance influences the response to treatment and determines the persistence of a microorganism in the host and in the environment. Antibiotic-resistant microorganisms are appearing increasingly in surface waters, wastewater, and drinking water (Schwartz et al., 2003). As is the case with virulence

genes, exchange of antibiotic resistance genes among microorganisms can lead to the emergence of new microbial threats.

Databases of virulence and antibiotic resistance genes could serve as the starting point for genetic descriptors related to the microorganism. Online databases focusing on virulence factors have been developed in recent years, for example, by Chen et al. (2005; http://www.mgc.ac.cn/VFs/main.htm). This database currently houses information about 1979 virulence and related factors/ genes covering 18 genera and is updated on a regular basis. A virulence factor database housing information on hundreds of pathogens and virulence genes is currently being constructed by the Los Alamos National Laboratory (http://www.tvfac. lanl.gov/). A pilot-scale effort for storing genetic descriptors is described by Jenkins et al. (2004; http://flyingcloud.cme.msu.edu/vfarl). It must be noted that none of the above-described databases are currently suitable to explore the development of VFAR, because they do not contain information about response variables or host-and. environment-related descriptors. Future databases housing VFAR descriptor and response variables must be comprehensive and include tools to explore the relationships between descriptor and response variables. They may even integrate Bayesian modeling approaches (described below) to explore the linkages between descriptor and response variables.

Mathematical models for VFAR

Development of VFAR algorithms that relate descriptor variables to response variables is the second and arguably the most challenging element in the development

process; it depends on mathematical approaches that are able to integrate a variety of quantitative and qualitative data and also allow transparent expert intervention. Mathematical frameworks available are the following: (1) deterministic: by transforming the mechanistic understanding of the processes into equations; (2) empirical: by fitting equations to experimental observations; (3) neural-network-based: by training computational networks to learn from example data sets; and (4) probabilistic: by incorporating uncertainty to causation. Deterministic approaches require knowledge of the fundamental relationships between molecular data and observed phenotypes. Empirical and neural-network-based approaches require large amounts of data. Probabilistic approaches (also known as Bayesian, causal or belief networks, or knowledge maps) are developed using as much data as available, and their predictions improve as more data becomes available and are incorporated. Bayesian approaches are most useful to do the following: (1) compute the probability of any event, given any evidence; (2) evaluate the effect of intervention; (3) provide the most likely scenario that explains the evidence; and (4) make rational decisions.

Many diagnostic tools using probabilistic models with expert-level decision making capabilities exist today in medical diagnosis, genetic pedigree analysis, speech recognition, gene sequence/ expression analysis, and machine learning or artificial intelligence. Bayesian approaches have great potential for the development of VFAR because (1) VFAR needs to use molecular data, which may come from all cellular levels, from genomics to proteomics and phenomics; (2) data related to descriptor and response variables may be quantitative or qualitative; (3) uncertainties associated with descriptor

and response variables are larger than uncertainties associated with physicochemical data; (4) some of the data (i.e., occurrence) used for developing VFAR may be sparse, spatially and temporally; and (5) all microbial data suffers from the lack of knowledge about microbial populations that are yet to be discovered. Bayesian approach is perhaps the only mathematical tool that can address all the above issues and also incorporate uncertainty.

To illustrate the components of a VFAR algorithm, we again refer to Figure 2.1, which depicts linking a set of descriptor variables to response variables in a manner that incorporates the uncertainties associated to the response variables and allows expert intervention. The VFAR development makes use of data collected from various sources and organization levels, often using very different methods and techniques. It must still be transparent and systematic to satisfy the requirements of the CCL process and allow for the uncertainty in data quality and quantity. Major challenges in building a VFAR algorithm using the collected data comes from the lack of knowledge about the importance of data on different descriptor variables and the associated probability distributions.

TABLE 2.3 Examples of virulence factors for selected pathogens.

Pathogen	Class I	Class II	Class III
A. hydrophila	aerolysin, hemolysin, cytotonic enterotoxin, heat-stable enterotoxin, heat-labile enterotoxin, type IV pili	type III secretion system, type II secretion system	
C. jejuni	cytolethal distending toxin CDT, adhesion factor, invasion factor	translocation CDT, type IV secretion system, two-component regulator	
H. pylori	vacuolating cytotoxin, cytotoxin-associated gene, adhesion factor	type IV secretion system	
P. aeruginosa	exoenzyme, type IV pili	type III secretion system, type II secretion system	pyochelin biosynthesis,
Y. enterocolitica	heat-stable enterotoxin, adhesion factor, attachment invasion locus, invasin	type III secretion system, type III secretion system	yersiniabactin synthesis, yersiniabactin receptor
V. cholerae	cholera toxin, zona occludens toxin, accessory cholera toxin, RTX toxin, toxin-coregulated pilus, type IV pili	type II secretion system, two-component regulator, transport and secretion RTX toxin	
B. pertussis	pertussis toxin, dermonecrotic toxin, tracheal cytotoxin (tct), adenylate cyclase AC, filamentous haemagglutinin FA, pertactin, tracheal colonization factor	secretion and activation AC, secretion FA, type III secretion system, two-component regulator	

Table 2.3 Continued

Salmonella	type I fimbriae, curli, long polar fimbriae, secreted proteins	two-components regulator, ferric uptake regulator, type III secretion system, regulators	
S. aureus	hemolysins, exfoliative toxin, leukocidin, enterotoxins, toxic shock syndrome toxin, collagen binding protein, eslastin binding protein, autolysin	regulators, two-component regulator	iron uptake factors
C. perfringens	enterotoxin, alpha toxin, beta toxin, iota toxin, epsilon toxin, perfringolysin or theta-toxin, collagenase or kappa-toxin	two-component regulator	
L. monocytogenes	listeriolysin, autolysin, invasion-associated protein	regulator	
L. pneumophila	RTX toxin, type IV pili	two-component regulator, type IVb secretion system, type II secretion system, type I secretion system	iron uptake factors
E. coli EHEC	hemolysin, Shiga toxin, intimin, intimin receptor	type III secretion system, secretion and activation hemolysin	iron uptake factors

The genes presented in each class are examples, and the list is neither exhaustive nor exclusive. Genes are classified according to their function (Wassenaar and Gaastra, 2001). According to this classification scheme, virulence genes fall into three main categories: classical virulence genes (class I), genes that encode for factors essential for the activity of the classical virulence factors (class II), and genes related to the specific life style of a pathogen (class III). Such and more advanced sorting of virulence genes can potentially be adopted to rank the importance of the genes, with genes in class I given in the highest score.

High-throughput molecular monitoring

The CCL Classification Work group of NDWAC indicated that a proactive and robust approach to determine the occurrence of waterborne pathogens would permit an effective means to identify organisms responsible for waterborne disease outbreaks. Monitoring constitutes the third critical element for VFAR development and validation. Traditionally, the detection of waterborne pathogens is based on time-consuming, culture-based methods, which also depend on the ability to grow the microorganisms from complex environmental matrices. Cultivation-independent molecular detection methods allow to screen for thousands of genes selected from most organisms in the microbial PCCL. It is the most economical and direct approach to determine the cause of waterborne disease outbreaks and the organisms responsible for them. Molecular tools have been used to detect virulence and antibiotic-resistant genes in different types of water samples (Schwartz et al., 2003; Sen and Rodgers, 2004). A comprehensive sampling strategy and high-throughput monitoring provides extensive information about microorganisms that are rarely monitored in treated drinking water and its sources, including microbial agents of unknown etiology. The presumption is that the presence and abundance of specific genes can be correlated to exposure and health risks. Evidently, selection of candidate gene targets is critical in the monitoring step. A broad screen targeting virulence and antibiotic-resistant genes of all microorganisms in the preliminary CCL can be used initially. Designing a broad screen to identify unknown threats has already been used successfully for viruses in clinical samples (Wang et al., 2003). The screening may later be narrowed to a carefully selected and validated set of genes that proved most valuable in predicting pathogenic threats. For successful application in VFAR development,

molecular monitoring will need to be complemented with epidemiological surveillance (i.e., information on disease outbreaks).

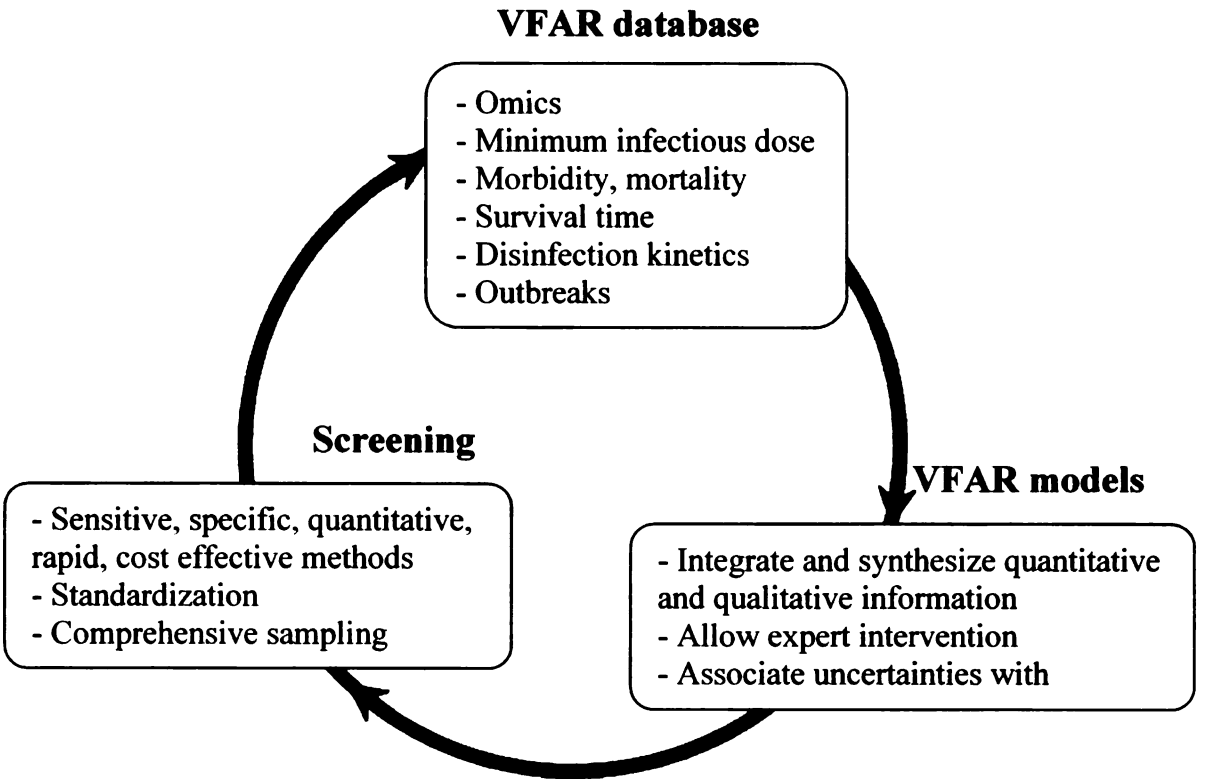


FIGURE 2.4 Components of the VFAR development and validation process. The cyclic nature of the process reflects the need to continuously update and improve information used to develop and validate VFAR.

Comprehensive and standardized sampling will be crucial to the success of a high-throughput molecular monitoring in the development and validation of VFAR. Samples need to be collected covering the widest possible spatial and temporal range and be integrated with epidemiological and outbreak data. Both cost and analysis time of the chosen detection method will determine the scope of sampling and the amount of information that can be gathered during monitoring. Sample collection and concentration, nucleic acid extraction, and the method of target detection may need to be standardized to allow exchange of data between different laboratories and databases. This may even

require a regulatory mechanism. Standardized methods should be easily implemented by different laboratories but still meet all requirements of a screening tool useful for VFAR.

Standardization of methods for sample preparation is warranted, because it can significantly influence the accuracy of molecular detection. For example, optimized methods to ensure maximum yield of nucleic acids from oocysts are required for accurate detection of *Cryptosporidium parvum* (Nichols and Smith, 2004). Because of the need to screen for different pathogenic groups (i.e., bacteria, viruses, and protozoa), development of sample preparation and detection methods applicable to all pathogens is challenging (Straub and Chandler, 2003). For VFAR development, a number of complementary methods may need to be used to ensure accurate detection and monitoring of all pathogens. The application of multiple and complementary methods may also be needed to detect microbial targets at different abundance levels. A VFAR screening tool should be able to detect all pathogens at its lowest concentration in the environment or at least at the lowest concentration known to pose a risk.

High specificity and sensitivity of the molecular detection methods are key requirements of the selected monitoring tools. High sensitivity is critical, because pathogenic microorganisms in environmental matrices are anticipated to constitute minor populations (less than 1%, by weight of DNA). This can be illustrated by considering that, in the genome of a given organism, a functional gene is generally present in 1 to 3 copies in a background of, on average, 4000 other genes. Assuming a relative abundance of 1%, this implies that the sensitivity of gene detection methods for environmental

samples must be better than 1 target gene in 400 000 nontarget genes. This is a challenging task, when multiple molecular targets must be analyzed together, in some cases for allelic variability, in matrices with varying characteristics.

Molecular tools to monitoring of thousands of target genes at a sensitivity and specificity level that is relevant to waterborne pathogens do exist at present. For example, real-time PCR (RT-PCR) enables the detection of a single gene copy and can be used to detect single nucleotide polymorphisms for allele discrimination. Furthermore, the RT-PCR platform allows moderate multiplexing and is characterized by analysis times below 1 hour (Belgrader et al., 1999). Microarray technology enables extraordinarily high levels of multiplexing, but lacks the required sensitivity for environmental applications. A number of technologies are currently being developed to integrate the benefits of both platforms. One such technology is on-chip PCR (Mitterer et al., 2004; Pemov et al., 2005), which combines high specificity, sensitivity, and quantization capabilities of RT-PCR with the high throughput of microarrays. On-chip RT-PCR may provide an economical means to analyze a high number of target genes and samples at the required level of sensitivity and specificity for waterborne pathogens and VFAR development. Its current cost estimate is in the range of $0.23 per target gene per sample and is expected to go down to $0.04 per target gene per sample in the near future. This constitutes a more than 10-fold reduction in cost compared with RT-PCR in conventional multiwell plates. It should be stressed that such low costs can only be achieved by analyzing a large number of targets and samples together. Technological advancements that enabled the development of on-chip RT-PCR include the capability to (1) load primers and other

reagents in nano-to pico-liter reaction chambers, (2) perform target amplification in such small reaction chambers, and (3) perform real-time imaging used modified equipment.

On-chip, PCR-based monitoring may form the cornerstone for the high-throughput molecular screening during the development and validation of VFAR (Figure 2.4). Ideally, such a molecular monitoring platform should be combined with a system automating and integrating the different steps in the detection process, preferably in portable devices for in-field application. The need for such devices is well-recognized (Straub et al., 2005), although efforts to develop them are insufficient compared with the importance of the problem. The performance of these automated devices may need to be rigorously validated using internal and external controls to ensure the accuracy of the readouts. Quality controls for the on-chip PCR amplification device should include controls for sample dispensing and internal controls for amplification. The latter is essential to eliminate false negative results resulting from inhibition of the reaction, which can result from malfunctioning of the thermal cycler, incorrect PCR mixtures, and the presence of amplification inhibitors in environmental extracts (Burggraf and Olgemoller, 2004).

In an ongoing project, our group is developing molecular diagnostic tools based on microarrays and on-chip RT-PCR, targeting a comprehensive set of pathogens, including a broad range of waterborne pathogens. A similar approach for all available virulence gene sequences for the microorganisms in the preliminary CCL may serve as the basis for the development of VFAR and the CCL. Such an approach should ultimately lead to an

economical and accurate monitoring scheme based on a selected set of genes proven useful in the prediction of health threats. After initial ranking within the microbial CCL, monitoring techniques may be as diverse as permitted by the identity of the microorganisms. DNA-based monitoring methods may also need to be complemented with other techniques, such as the use of RNA as targets for the molecular detection, to allow differentiation between live/dead and active/ inactive microbial targets. Culture-based methods may be warranted to validate the results of molecular methods. Culturing can also provide isolates that can be used to evaluate their health effects.

A high-throughput molecular screening could be a flagship program focusing on monitoring the prevalence of virulence and antibiotic-resistant genes in drinking water supplies and sources. Numerous examples exist to illustrate the benefits of focused flagship programs, for example, PulseNet, focusing on food, instituted by the Centers for Disease Control (http://www.cdc.gov/), and BioWatch, focusing on air, overseen by the Department of Homeland Security (http://www.dhs.gov/). Parallel programs for water will complement the above two programs and help develop VFAR. Preferably, such a program for VFAR should allow the integration of information from different scientific fields and provide a transparent and effective means of sharing data among multiple laboratories. Rigorous method standardization and evaluation of new entries to ensure the quality and uniformity will determine its applicability.

LIMITATIONS OF VFAR

When fully developed, VFAR will be able to predict response variables from a given set of descriptor variables related to existing or other microorganisms closely related to known pathogens. This information can then be used to rank and prioritize waterborne pathogens for regulatory purposes. However, VFAR will also have some limitations that need to be recognized during its development and interpretation. A number of the limitations of VFAR are listed below.

- Information related to higher levels of cellular organization (i.e., transcriptomics and proteomics) of pathogenic microorganisms under environmental and clinical conditions is limited. The VFAR models will need to incorporate this as uncertainties associated with the predictions and allow for its inclusion in the future.
- Information related to response variables (i.e., health effects) of strains harboring different combinations of virulence and antibiotic-resistant genes is critical for the development of VFAR. However, data related to severity of health threats of different strains is currently very sparse and will limit the accuracy of VFAR predictions.
- Strains of pathogenic species can possess different combinations of virulence genes, and monitoring of these genes within a mixed microbial community does not yield information about which strains (or combinations of strains) are present. However, linking information about an assortment of target genes detected in a mixed microbial community to the presence of specific pathogens and ultimately to health effects is critical for the application of high-throughput monitoring. Therefore, high-

throughput monitoring may need to be complemented with culture-based isolation techniques to determine strain identities.

- The application of molecular monitoring approaches for microbial water quality assessment is relatively new. This is because molecular methods for environmental detection need to fulfill stringent requirements in terms of sensitivity, specificity, analysis time, and cost effectiveness. Advancements that enable molecular tools to meet these requirements will play a critical role in developing and validating VFAR.

- The predictive capabilities of VFAR will be limited by the ability of mathematical and statistical models to integrate various types of quantitative and qualitative information to comprehensive predictive tools.

- Even in its most developed form, VFAR, as presented in this paper, will not be a model to predict evolution of pathogens. However, the information collected and learned in the process of developing VFAR may serve as a critical resource in developing such models.

REFERENCES

Adu-Bobie, J., G. Frankel, C. Bain, A.G. Goncalves, L.R. Trabulsi, G. Douce, S. Knutton and G. Dougan, 1998. Detection of intimins alpha, beta, gamma, and delta, four intimin derivatives expressed by attaching and effacing microbial pathogens. Journal of Clinical Microbiology, 36: 662-668.

Atherton, J.C., P. Cao, R.M. Peek, M.K.R. Tummuru, M.J. Blaser and T.L. Cover, 1995. Mosaicism in valcuolating cytotoxin alleles in *Helicobacter pylori*. Association of specific *vacA* types with cytotoxin production and peptic ulceration. Journal of Biological Chemistry, 270: 17771-17777.

Becker, K., A.W. Friedrich, G. Peters and C. von Eiff, 2004. Systematic survey on the prevalence of genes coding for staphylococcal enterotoxins SEIM, SEIO, and SEIN. Molecular Nutrition & Food Research, 48: 488-495.

Bekal, S., R. Brousseau, L. Masson, G. Prefontaine, J. Fairbrother and J. Harel, 2003. Rapid identification of *Escherichia coli* pathotypes by virulence gene detection with DNA microarrays. Journal of Clinical Microbiology, 41: 2113-2125.

Belgrader, P., W. Benett, D. Hadley, J. Richards, P. Stratton, R. Mariella and F. Milanovich, 1999. Infectious disease - PCR detection of bacteria in seven minutes. Science, 284: 449-450.

Beutin, L., G. Krause, S. Zimmermann, S. Kaulfuss and K. Gleier, 2004. Characterization of Shiga toxin-producing *Escherichia coli* strains isolated from human patients in Germany over a 3-year period. Journal of Clinical Microbiology, 42: 1099-1108.

Blanco, J.E., M. Blanco, M.P. Alonso, A. Mora, G. Dahbi, M.A. Coira and J. Blanco, 2004. Serotypes, virulence genes, and intimin types of Shiga toxin (verotoxin)-producing *Escherichia coli* isolates from human patients: Prevalence in Lugo, Spain, from 1992 through 1999. Journal of Clinical Microbiology, 42: 311-319.

Boerlin, P., S.A. McEwen, F. Boerlin-Petzold, J.B. Wilson, R.P. Johnson and C.L. Gyles, 1999. Associations between virulence factors of Shiga toxin-producing *Escherichia coli* and disease in humans. Journal of Clinical Microbiology, 37: 497-503.

Brown, C.M., P.J. Nuorti, R.F. Breiman, A.L. Hathcock, B.S. Fields, H.B. Lipman, G.C. Llewellyn, J. Hofmann and M. Cetron, 1999. A community outbreak of Legionnaires' disease linked to hospital cooling towers: an epidemiological

method to calculate dose of exposure. International Journal of Epidemiology, 28: 353-359.

Burggraf, S. and B. Olgemoller, 2004. Simple technique for internal control of real-time amplification assays. Clinical Chemistry, 50: 819-825.

Cangelosi, G. A., N.E. Freitag, M.R. Buckley, 2004. From Inside to Outside: Environmental Microorganisms as Human Pathogens; American Academy of Microbiology: Washington, D.C.

Cassiday, P., G. Sanden, K. Heuvelman, F. Mooi, K.M. Bisgard and T. Popovic, 2000. Polymorphism in *Bordetella pertussis* pertactin and pertussis toxin virulence factors in the United States, 1935-1999. Journal of Infectious Diseases, 182: 1402-1408.

Censini, S., C. Lange, Z.Y. Xiang, J.E. Crabtree, P. Ghiara, M. Borodovsky, R. Rappuoli and A. Covacci, 1996. cag, a pathogenicity island of *Helicobacter pylori*, encodes type I-specific and disease-associated virulence factors. Proceedings of the National Academy of Sciences of the United States of America, 93: 14648-14653.

Chacon, M.R., L. Soler, E.A. Groisman, J. Guarro and M.J. Figueras, 2004. Type III secretion system genes in clinical *Aeromonas* isolates. Journal of Clinical Microbiology, 42: 1285-1287.

Chen, L.H., J. Yang, J. Yu, Z.J. Ya, L.L. Sun, Y. Shen and Q. Jin, 2005. VFDB: a reference database for bacterial virulence factors. Nucleic Acids Research, 33: D325-D328.

Correa, P., 2005. New strategies for the prevention of gastric cancer: *Helicobacter pylori* and genetic susceptibility. Journal of Surgical Oncology, 90: 134-138.

Cronin, M.T.D., J.S. Jaworska, J.D. Walker, M.H.I. Comber, C.D. Watts and A.P. Worth, 2003. Use of QSARs in international decision-making frameworks to predict health effects of chemical substances. Environmental Health Perspectives, 111: 1391-1401.

Donlan, R.M., T. Forster, R. Murga, E. Brown, C. Lucas, J. Carpenter and B. Fields, 2005. *Legionella pneumophila* associated with the protozoan *Hartmannella vermiformis* in a model multi-species biofilm has reduced susceptibility to disinfectants. Biofouling, 21: 1-7.

Doumith, M., C. Cazalet, N. Simoes, L. Frangeul, C. Jacquet, F. Kunst, P. Martin, P. Cossart, P. Glaser and C. Buchrieser, 2004. New aspects regarding evolution and virulence of *Listeria monocytogenes* revealed by comparative genomics and DNA arrays. Infection and Immunity, 72: 1072-1083.

Feltman, H., G. Schulert, S. Khan, M. Jain, L. Peterson and A.R. Hauser, 2001. Prevalence of type III secretion genes in clinical and environmental isolates of *Pseudomonas aeruginosa*. Microbiology, 147: 2659-2669.

Friedrich, A.W., M. Bielaszewska, W.L. Zhang, M. Pulz, T. Kuczius, A. Ammon and H. Karch, 2002. *Escherichia coli* harboring shiga toxin 2 gene variants: frequency and association with clinical symptoms. Journal of Infectious Diseases, 185: 74-84.

Grozdanov, L., C. Raasch, E. Schulze, U. Sonnenborn, G. Gottschalk, J. Hacker and U. Dobrindt, 2004. Analysis of the genome structure of the nonpathogenic probiotic *Escherichia coli* strain Nissle 1917. Journal of Bacteriology, 186: 5432-5441.

Hlady, W.G., R.C. Mullen, C.S. Mintz, B.G. Shelton, R.S. Hopkins and G.L. Daikos, 1993. Outbreaks of legionnaires' disease linked to a decorative fountain by molecular epidemiology. American Journal of Epidemiology, 138: 555-562.

Holden, M., L. Crossman and A. Cerdeno-Tarraga, 2004. Pathogenomics of non-pathogens. Nature Reviews Microbiology, 2: 91-91.

Howard, K. and T.J.J. Inglis, 2005. Disinfection of *Burkholderia pseudomallei* in potable water. Water Research, 39: 1085-1092.

Jeffers, G.T., J.L. Bruce, P.L. McDonough, J. Scarlett, K.J. Boor and M. Wiedmann, 2001. Comparative genetic characterization of *Listeria monocytogenes* isolates from human and animal listeriosis cases. Microbiology-Sgm, 147: 1095-1104.

Jenkins, T.M., T.M. Scott, J.R. Cole, S.A. Hashsham and J.B. Rose, 2004. Assessment of virulence-factor activity relationships (VFARs) for waterborne diseases. Water Science and Technology, 50: 309-314.

Kishishita, M., N. Matsuoka, K. Kumagai, S. Yamasaki, Y. Takeda and M. Nishibuchi, 1992. Sequence variation in the thermostable direct hemolysin-related hemolysin (*trh*) gene of *Vibrio parahaemolyticus*. Applied and Environmental Microbiology, 58: 2449-2457.

Kumar, S., K. Tamura and M. Nei, 2004. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. Briefings in Bioinformatics, 5: 150-163.

Miller, S.M., D.M. Tourlousse, R.D. Stedtfeld, S.W. Baushke, A.B. Herzog, L.M. Wick, J.-M. Rouillard, E. Gulari, J.M. Tiedje, and S.A. Hashsham, 2008 *In situ*-synthesized virulence and marker gene biochip for detection of bacterial pathogens in water. Applied and Environmental Microbiology, 74: 2200-2209.

Mitterer, G., M. Huber, E. Leidinger, C. Kirisits, W. Lubitz, M.W. Mueller and W.M. Schmidt, 2004. Microarray-based identification of bacteria in clinical samples by solid-phase PCR amplification of 23S ribosomal DNA sequences. Journal of Clinical Microbiology, 42: 1048-1057.

Nascimento, A., A.I. Ko, E.A.L. Martins, C.B. Monteiro-Vitorello, P.L. Ho, D.A. Haake, S. Verjovski-Almeida, R.A. Hartskeerl, M.V. Marques, M.C. Oliveira, C.F.M. Menck, L.C.C. Leite, H. Carrer, L.L. Coutinho, W.M. Degrave, O.A. Dellagostin, H. El-Dorry, E.S. Ferro, M.I.T. Ferro, L.R. Furlan, M. Gamberini, E.A. Giglioti, A. Goes-Neto, G.H. Goldman, M.H.S. Goldman, R. Harakava, S.M.B. Jeronimo, I.L.M. Junqueira-de-Azevedo, E.T. Kimura, E.E. Kuramae, E.G.M. Lemos, M.V.F. Lemos, C.L. Marino, L.R. Nunes, R.C. de Oliveira, G.G. Pereira, M.S. Reis, A. Schriefer, W.J. Siqueira, P. Sommer, S.M. Tsai, A.J.G. Simpson, J.A. Ferro, L.E.A. Camargo, J.P. Kitajima, J.C. Setubal and M.A. Van Sluys, 2004. Comparative genomics of two *Leptospira interrogans* serovars reveals novel insights into physiology and pathogenesis. Journal of Bacteriology, 186: 2164-2172.

NDWAC, 2004 Downloaded in July 2006 from http://www.awwa.org/Advocacy/govtaff/documents/CCLRatificationDraft.pdf

Nguyen, T.N., A.N. Barkun and C.A. Fallone, 1999. Host determinants of *Helicobacter pylori* infection and its clinical outcome. Helicobacter, 4: 185-197.

Nichols, R.A.B. and H.V. Smith, 2004. Optimization of DNA extraction and molecular detection of Cryptosporidium oocysts in natural mineral water sources. Journal of Food Protection, 67: 524-532.

Nilsson, C., A. Sillen, L. Eriksson, M.L. Strand, H. Enroth, S. Normark, P. Falk and L. Engstrand, 2003. Correlation between cag pathogenicity island composition and *Helicobacter pylori*-associated gastroduodenal disease. Infection and Immunity, 71: 6573-6581.

NRC, 1999 - Setting Priorities for Drinking Water Contaminants; National Academies Press: Washington, D.C.

Ottenhoff, T.H.M., F.A.W. Verreck, E.G.R. Lichtenauer-Kaligis, M.A. Hoeve, O. Sanal and J.T. van Dissel, 2002. Genetics, cytokines and human infectious disease: lessons from weakly pathogenic mycobacteria and salmonellae. Nature Genetics, 32: 331-331.

Pearson, B.M., C. Pin, J. Wright, K. I'Anson, T. Humphrey and J.M. Wells, 2003. Comparative genome analysis of *Campylobacter jejuni* using whole genome DNA microarrays. FEBS Letters, 554: 224-230.

Pemov, A., H. Modi, D.P. Chandler and S. Bavykin, 2005. DNA analysis with multiplex microarray-enhanced PCR. Nucleic Acids Research, 33.

Pier, G.B., 2002. CFTR mutations and host susceptibility to *Pseudomonas aeruginosa* lung infection. Current Opinion in Microbiology, 5: 81-86.

Poly, F., D. Threadgill and A. Stintzi, 2004. Identification of *Campylobacter jejuni* ATCC 43431-specific genes by whole microbial genome comparisons. Journal of Bacteriology, 186: 4781-4795.

Rozynek, E., K. Dzierzanowska-Fangrat, P. Jozwiak, J. Popowski, D. Korsak and D. Dzierzanowska, 2005. Prevalence of potential virulence markers in Polish *Campylobacter jejuni* and *Campylobacter coli* isolates obtained from hospitalized children and from chicken carcasses. Journal of Medical Microbiology, 54: 615-619.

Said, B., F. Wright, G.L. Nichols, M. Reacher and M. Rutter, 2003. Outbreaks of infectious disease associated with private drinking water supplies in England and Wales 1970-2000. Epidemiology and Infection, 130: 469-479.

Saitou, N. and M. Nei, 1987. The neighbor-joining method: a new method for reconstructing phylogentic trees. Molecular Biology and Evolution, 4: 406-425.

Salama, N., K. Guillemin, T.K. McDaniel, G. Sherlock, L. Tompkins and S. Falkow, 2000. A whole-genome microarray reveals genetic diversity among *Helicobacter pylori* strains. Proceedings of the National Academy of Sciences of the United States of America, 97: 14668-14673.

Schwartz, T., W. Kohnen, B. Jansen and U. Obst, 2003. Detection of antibiotic-resistant bacteria and their resistance genes in wastewater, surface water, and drinking water biofilms. FEMS Microbiology Ecology, 43: 325-335.

Sen, K. and M. Rodgers, 2004. Distribution of six virulence factors in *Aeromonas* species isolated from US drinking water utilities: a PCR identification. Journal of Applied Microbiology, 97: 1077-1086.

Stedtfeld, R. D., S.W. Baushke, S.M. Miller, J.M. Tiedje, S.A. Hashsham, 2006 Microchamber Biochip for Quantitative Detection of Human Pathogens in Environmental Samples. Proceedings of the 106th General Meeting of the American Society for Microbiology, Orlando, Florida, May 21-25; American Society for Microbiology: Washington, D.C.

Straub, T.M. and D.P. Chandler, 2003. Towards a unified system for detecting waterborne pathogens. Journal of Microbiological Methods, 53: 185-197.

Straub, T.M., B.P. Dockendorff, M.D. Quinonez-Diaz, C.O. Valdez, J.I. Shutthanandan, B.J. Tarasevich, J.W. Grate and C.J. Bruckner-Lea, 2005. Automated methods for multiplexed pathogen detection. Journal of Microbiological Methods, 62: 303-316.

Tan, H.J., A.M. Rizal, M.Y. Rosmadi and K.L. Goh, 2005. Distribution of *Helicobacter pylori cagA*, *cagE* and *vacA* in different ethnic groups in Kuala Lumpur, Malaysia. Journal of Gastroenterology and Hepatology, 20: 589-594.

Thong, K.L., S.L. Hoe, S.D. Puthucheary, R.M. Yasin, 2005 Detection of Virulence Genes in Malaysian *Shigella* Species by Multiplex PCR Assay. BMC Infectious Diseases, 15: 8.

U.S. EPA, 2005 - Drinking Water Contaminant Candidate List 2. Federal Register 70 (6), 9071

van de Vosse, E., M.A. Hoeve and T.H.M. Ottenhoff, 2004. Human genetics of intracellular infectious diseases: molecular and cellular immunity against mycobacteria and salmonellae. Lancet Infectious Diseases, 4: 739-749.

Wang, D., A. Urisman, Y.T. Liu, M. Springer, T.G. Ksiazek, D.D. Erdman, E.R. Mardis, M. Hickenbotham, V. Magrini, J. Eldred, J.P. Latreille, R.K. Wilson, D. Ganem and J.L. DeRisi, 2003. Viral discovery and sequence recovery using DNA microarrays. PLOS Biology, 1: 257-260.

Wassenaar, T.M. and W. Gaastra, 2001. Bacterial virulence: can we draw the line? FEMS Microbiology Letters, 201: 1-7.

Whittam, T.S. and A.C. Bumbaugh, 2002. Inferences from whole-genome sequences of bacterial pathogens. Current Opinion in Genetics & Development, 12: 719-725.

Wick, L.M., J.M. Rouillard, T.S. Whittam, E. Gulari, J.M. Tiedje and S.A. Hashsham, 2006. On-chip non-equilibrium dissociation curves and dissociation rate constants as methods to assess specificity of oligonucleotide probes. Nucleic Acids Research, 34.

Zhang, W.L., B. Kohler, E. Oswald, L. Beutin, H. Karch, S. Morabito, A. Caprioli, S. Suerbaum and H. Schmidt, 2002. Genetic diversity of intimin genes of attaching and effacing *Escherichia coli* strains. Journal of Clinical Microbiology, 40: 4486-4492.

Zhang, Y.Q., S.X. Ren, H.L. Li, Y.X. Wang, G. Fu, J. Yang, Z.Q. Qin, Y.G. Miao, W.Y. Wang, R.S. Chen, Y. Shen, Z. Chen, Z.H. Yuan, G.P. Zhao, D. Qu, A. Danchin and Y.M. Wen, 2003. Genome-based analysis of virulence genes in a non-biofilm-forming *Staphylococcus epidermidis* strain (ATCC 12228). Molecular Microbiology, 49: 1577-1593.

CHAPTER THREE

AN *IN SITU* SYNTHESIZED VIRULENCE AND MARKER GENE (VMG) BIOCHIP FOR THE DETECTION OF WATERBORNE PATHOGENS

Miller, Sarah M.[#]; Tourlousse, Dieter M.[#]; Stedtfeld, Robert D., Baushke, Samuel W.; Herzog, Amanda B.; Wick, Lukas M.; Rouillard, Jean-Marie; Gulari, Erdogan; Tiedje, James M.; Hashsham, Syed A. (2008) Applied and Environmental Microbiology 74(7):2200-9. [#] Both authors contributed equally to this study.

INTRODUCTION

Pathogen detection and identification using virulence and marker genes (VMGs) as genetic targets are receiving considerable interest due to the rapidly increasing availability of gene and whole-genome sequence information for most major pathogens. VMGs are expected to be more specific than the 16S rRNA gene due to the limited ability of the latter to differentiate among closely related microorganisms. Knowledge about the presence or absence of selected VMGs may also aid in estimating the threat posed by the detected microorganisms. Screening for multiple VMGs for each pathogen is necessary to reduce the potential of erroneous calls due to differences in the occurrence of various VMGs among pathogenic strains belonging to the same species (Tourlousse et al., 2007). Use of a single VMG or a limited number of VMGs may lead to an increased rate of false-negative calls if the selected VMGs are unevenly distributed among different strains.

One of the first microarrays for the detection of microbial pathogens using multiple VMGs was described by Wilson et al. (2002b). This microarray was validated for 11

bacterial, five viral, and two eukaryotic pathogens and employed genomic DNA of pathogens spiked in total DNA extracted from filtered air. This study also highlighted the advantage of using highly redundant probe sets to infer presence/absence calls based on the positive fraction (PF). The latter was defined as the number of positive probes for a given target divided by the total number of probes used to detect that target. Setting a sufficiently high threshold for the PF to define a target as present substantially reduces the rate of false-positive calls due to cross-hybridization. When redundant probe sets are used, the reliability of the presence/absence calls is expected to be somewhere between the reliability achieved by PCR-based detection of the VMGs and the reliability achieved by sequencing of the corresponding VMG amplicons. Utilization of redundant probe sets may also result in more efficient validation because the success level of probe design is generally high and presence/absence calls based on redundant probe sets are not affected by failure of a small number probes within a set.

Among the challenges in the development of pathogen detection microarrays is the limited ability to detect low-abundance target sequences within complex mixtures. Microarrays without target gene amplification only reliably detect microorganisms at a relative abundance of approximately 1 to 5% (Wu et al., 2001; Denef et al., 2003; Bodrossy and Sessitsch, 2004; Rhee et al., 2004; Kostic et al., 2005). Obviously, pathogens may be present at levels well below this limit in various matrices. Hence, microarray-based pathogen detection generally relies upon target gene amplification strategies, typically using PCR (Call et al., 2003; Call, 2005; Loy and Bodrossy, 2006). For conserved genes (e.g., the 16S rRNA gene), amplification using universal or group-

specific primer sets in multi-template PCR has been extensively used to enrich target genes prior to hybridization (Loy et al., 2002; Wilson et al., 2002a; Luebke et al., 2003; Loy et al., 2004; Warsen et al., 2004; DeSantis et al., 2005; Franke-Whittle et al., 2005; Maynard et al., 2005; Palmer et al., 2006; Sanguin et al., 2006a; Sanguin et al., 2006b; Siripong et al., 2006; Eom et al., 2007; Becker and Gartner, 2008). (Call et al., 2001; Chizhikov et al., 2001; Volokhov et al., 2002; Wilson et al., 2002b; Keramas et al., 2003; Al-Khaldi et al., 2004; Gonzalez et al., 2004; Lin et al., 2004; Panicker et al., 2004; Sergeev et al., 2004; Vora et al., 2005; Antwerpen et al., 2007; Lin et al., 2007). Methods for simultaneous amplification of VMGs, however, are less well developed. Frequently, multiple independent PCR assays or multiplex PCR have been used to amplify multiple VMGs to enhance probe signals and to improve detection limits for genotyping of isolates and pathogen detection in various matrices (Call et al., 2001; Chizhikov et al., 2001; Volokhov et al., 2002; Wilson et al., 2002b; Al-Khaldi et al., 2004; Lin et al., 2004; Panicker et al., 2004; Sergeev et al., 2004; Vora et al., 2005; Antwerpen et al., 2007; Lin et al., 2007). However, development of robust and sensitive multiplex PCR assays for multiple VMGs still requires careful primer design and significant optimization of the reaction parameters (Markoulatos et al., 2002). If a large number of VMGs could be simultaneously amplified in a robust manner in complex matrices, VMG-based microarrays could provide the required specificity, sensitivity, and target throughput, all in the same assay, for diagnostic purposes.

The sensitivity of microarrays is also influenced by the strength of probe signals obtained after hybridization. Probes with high target binding affinities may be preferred

for detecting low-abundance targets. However, such probes are also more prone to cross-hybridization (Held et al., 2003) and may decrease specificity. This trade-off between specificity and sensitivity needs to be assessed experimentally to define optimal probe selection criteria.

The main objective of this study was to design and validate an in situ-synthesized VMG biochip to detect 12 bacterial pathogens. Many of the pathogens included (*Aeromonas hydrophila*, *Helicobacter pylori*, *Legionella pneumophila*, *Pseudomonas aeruginosa*, *Vibrio cholerae*, *Vibrio parahaemolyticus*, and *Yersinia enterocolitica*) were waterborne. Other pathogens (*Clostridium perfringens*, *Salmonella*, *Staphylococcus aureus*, *Campylobacter jejuni*, and *Listeria monocytogenes*) were relevant to clinical diagnostics and food safety. DNA extracted from tap water, river water, and tertiary effluent from a municipal wastewater treatment plant served as the matrix used to assess sensitivity and specificity after samples were spiked with pathogen DNA. Target gene amplification using a split multiplex PCR assay allowed detection of pathogens at a relative abundance between 0.1 and 0.01%, depending on the pathogen and the VMG. Up to six VMG amplicons per pathogen and up to 35 probes per VMG amplicon were used to eliminate false-positive calls. The effect of characteristics of probes on their hybridization behavior was also evaluated in order to derive probe design rules yielding the best trade-off between sensitivity and specificity. The described VMG biochip may have applications in diagnostic areas where parallel screening of multiple pathogens with high levels of specificity is critical.

MATERIALS AND METHODS

Probe design and biochip synthesis

Two sets of probes were included on the VMG biochip. The first set of probes (n = 791) was designed to detect 35 VMGs covering 12 pathogens. In addition, 47 PCR primer sets were designed for these genes with each amplicon being targeted by 7 to 35 probes (Table 3.1). The second set of probes (n = 2,034) targeted 67 VMGs for the 12 pathogens mentioned above, as well as VMGs for five additional pathogens. For these genes, no primers were designed, and the probes analyzed as nontarget probes to assess hybridization specificity. All VMGs selected in this study were previously used to detect the selected pathogens in PCR-based assays. The probes were designed based on 671 sequences retrieved from GenBank and satisfied the following criteria: (i) the probes were complementary to all retrieved sequences of a given gene with species-level specificity, and (ii) the probes contained at least two mismatches to all other nontarget sequences within the database. The Gibbs free energy of probe-target duplex formation assuming complementary targets (ΔG) ranged from –14.1 to –23.9 kcal/mol; the mean was –18.4 kcal/mol, and the variability (standard deviation) was 2.1 kcal/mol.

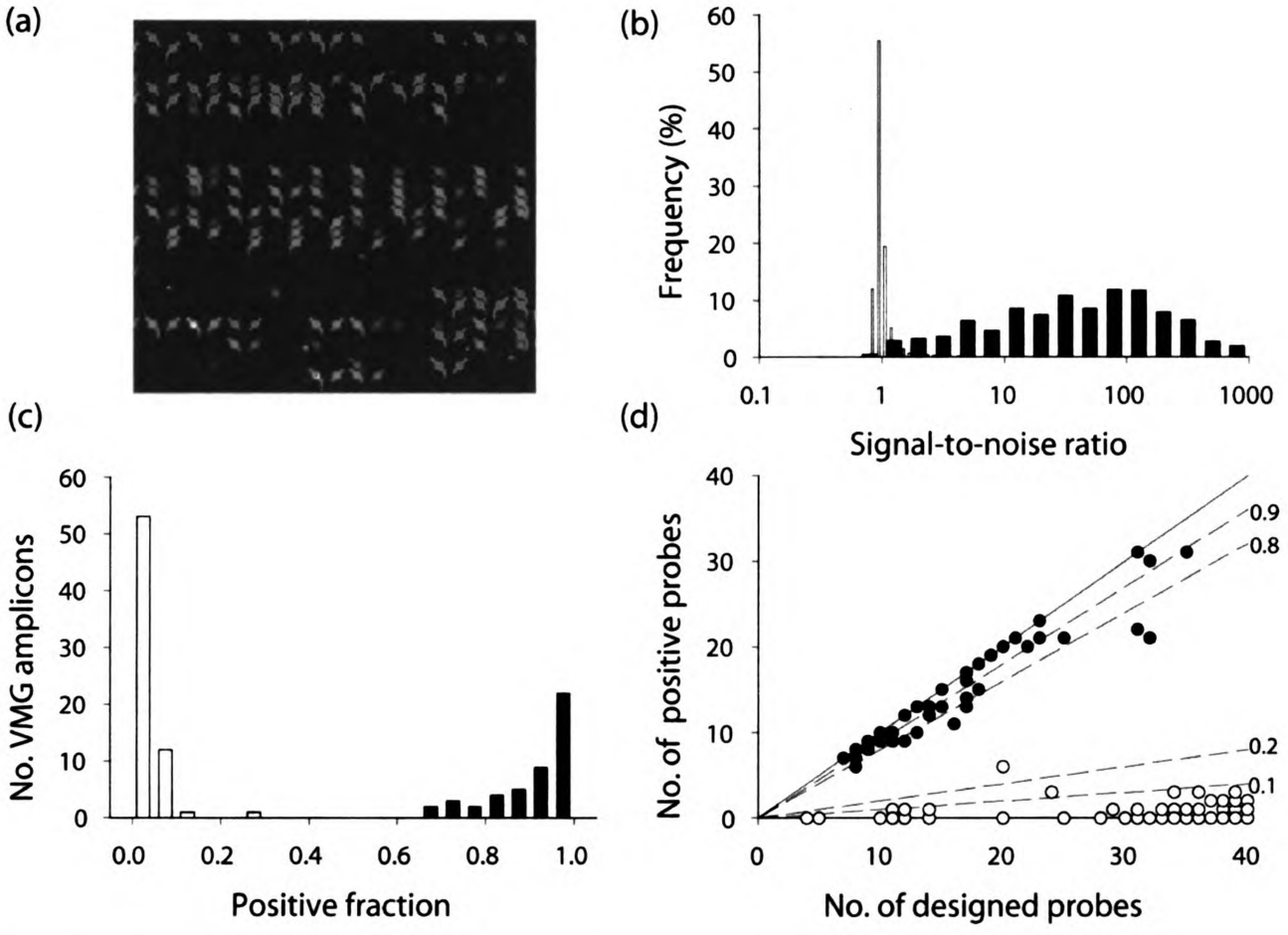


FIGURE 3.1 Evaluation of 791 target and 2,034 nontarget probes with a composite target mixture containing all 47 VMG amplicons. (a) Fluorescence image of a portion of the Xeotron chip after hybridization. (b) Distribution of the SNR for 791 target probes (filled bars) and 2,034 nontarget probes (open bars). (c) Distribution of the PF for 47 target probe sets (filled bars) and 67 nontarget sets (open bars). (d) Success of probe selection for target and nontarget probe sets (• and ○, respectively). The indicated slope of the dashed lines is numerically identical to the PF. Images in this thesis/dissertation are presented in color.

The probes were synthesized *in situ* on microfluidic biochips using a proprietary light-directed synthesis technology developed by the University of Michigan (Gao et al., 2001; Gao et al., 2004) and commercialized by Xeotron (Houston, TX; now part of Invitrogen, Carlsbad, CA). The microfluidic chip contained 8,000 micro-reactors with a diameter of 50 μm, which were interconnected with flow channels (Figure 3.1a). The *in situ* synthesis technology employs conventional phosphoramidite chemistry in conjunction with photogenerated acid-triggered deprotection of the 5'-hydroxyl groups of nucleotide phosphoramidite monomers with spatially resolved light patterns for light-directed acid deprotection are generated using a digital micromirror device. The probes were attached to the substrate via a spacer consisting of Ts and C_{18} with an effective length of 12 nucleotides. The biochip also contained 22 randomly spaced control spots containing solely linker chemistry to assess background signal intensity (Wick et al., 2006; Stedtfeld et al., 2007).

Nucleic acid extraction

Cultures of *A. hydrophila* ATCC 7966, *C. perfringens* ATTC 12916, *Salmonella* sp. strain ATCC 13311, *L. monocytogenes* ATCC 15313, *P. aeruginosa* ATCC 10145, *V. parahaemolyticus* ATCC 43996, and *Y. enterocolitica* ATCC 55075 were obtained from the American Type Culture Collection (ATCC) (Manassas, VA). Cells were grown overnight under the conditions suggested by ATCC. For DNA extraction from 1 to 2 mL of culture the DNeasy tissue kit was used (Qiagen, Valencia, CA) following the manufacturer's instructions. For *V. cholerae* ATCC 39315, *C. jejuni* ATCC 700819, *H.*

pylori ATCC 700392, *S. aureus* ATCC 700699, and *L. pneumophila* ATCC 33152, purified DNA was obtained from the ATCC.

Tap water (40 liters), river water (10 liters; Red Cedar River, East Lansing, MI), and tertiary effluent (20 liters; wastewater treatment plant in East Lansing, MI) were filtered through 0.45-μm nitrocellulose filters (Millipore, Billerica, MA) and DNA extracted from the filters using the MegaPrep UltraClean soil DNA kit (Mo Bio Laboratories, Carlsbad, CA) following the manufacturer's instructions. The amount and quality of the extracted DNA were determined with the NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE).

Primer design and PCR conditions

A total of 47 gene-specific primer sets were designed for 35 VMGs flanking gene regions with high probe density. Primers were selected such that they had similar annealing temperatures (one set had an annealing temperature of 53°C, and another set had an annealing temperature of 58°C) and covered most alleles of a given VMG available in the GenBank database. The uniqueness of the primers was confirmed via BLAST search against the GenBank database. To ensure primer specificity, mismatches with related nontarget sequences were located near the 3' ends of the primers. For multiplex PCR, the primers were separated into five primer combinations, with each combination containing 9 or 10 primer pairs. Mixtures of primer sets were selected such that each pathogen (except *P. aeruginosa*) was targeted in at least two different multiplex PCR assays. Primers were synthesized by Integrated DNA Technologies (Coralville, IA).

PCR mixtures (25 µl) consisted of 1x PCR buffer, 2 mM MgCl2, 1.5 U (for monoplex PCR) or 3 U (for multiplex PCR) of AmpliTaq Gold (Roche Molecular Systems, Pleasanton, CA), each deoxynucleoside triphosphate at a concentration of 200 µM (Invitrogen), each primer at a concentration of 500 nM, 200 ng of bovine serum albumin (New England BioLabs, Beverly, MA), and 1 µl of DNA. After initial enzyme activation at 94°C for 10 min, 35 cycles of the following temperature regimen was used for amplification: denaturation at 94°C for 60 s, annealing at 53 or 58°C for 60 s, and elongation at 72°C for 60 s. This was followed by a final elongation step at 72°C for 7 min. PCR amplicons were purified using the QIAquick PCR purification kit (Qiagen) and quantified using the NanoDrop ND-1000 spectrophotometer.

Fluorescent labeling of PCR products

PCR products pooled from different PCR assays were labeled using an aminoallyl-labeled dUTP (aa-dUTP) incorporation and cyanine dye coupling protocol as described previously (Wick et al., 2006; Stedtfeld et al., 2007), with minor modifications. Briefly, PCR products were amplified, and aa-dUTP was incorporated into the amplification products with the Bioprime DNA labeling kit (Invitrogen, San Diego, CA) using a ratio of aa-dUTP to dTTP of 5:1 and an incubation time of 120 min. Purified products were then coupled with cyanine dye by incubation for 80 min in a 1:1 mixture of 0.1 M sodium carbonate buffer (pH 9.3) and *N*-hydroxysuccinimide ester cyanine dye (Amersham Biosciences).

Biochip hybridization and melting curve development

Hybridization and melting profile generation were performed using previously established protocols (Wick et al., 2006; Stedtfeld et al., 2007). The approach used for analysis of the melting profiles was based on methods used in previous studies utilizing this technique for microarrays using 16S rRNA targets (Liu et al., 2001; Urakawa et al., 2002; El Fantroussi et al., 2003; Li et al., 2004). Briefly, after priming of the biochips, target DNA (200 pmol of Cy dye) was hybridized overnight at 20°C using an M-2 hybridization station (Invitrogen [formerly Xeotron Corporation, Houston, TX]) with hybridization buffer containing 35% deionized formamide (Ambion), 6x SSPE (pH 6.6; Invitrogen), and 0.4% Triton X-100 (Sigma) (1x SSPE is 0.18 M NaCl, 10 mM NaH_2PO_4, and 1 mM EDTA [pH 7.7]). After the chips were washed, the initial point of the melting curve was obtained by washing each chip with high-stringency wash buffer (20 mM NaCl, 10 mM Na_2PO_4, 5 mM Na_2EDTA; pH adjusted to 6.6 with HCl) for 1.4 min at 25°C and imaging the chip. High-stringency wash buffer was degassed under a vacuum to prevent formation of air bubbles during the wash steps. Subsequent development of the melting curve was performed by using manual cycles of washing and scanning of the chip, repeated at 1°C intervals up to 60°C. At the end of this procedure, the chip was additionally stripped at 60, 45 and 30°C (2.5 min each) with nuclease-free water (Sigma). The tubing of the hybridization station was washed for 20 min before and after each experiment to prevent carry-over between experiments.

Experimental design

Assay specificity and sensitivity were evaluated with samples consisting of pathogen DNA spiked into DNA obtained from different water sources (tap water, river water, and tertiary effluent from a wastewater treatment plant), as has been done in numerous other validation studies (Wilson et al., 2002b; Maynard et al., 2005). A first set of samples was prepared by spiking 10 pg of DNA of each pathogen (equivalent to approximately 1,400 to 5,500 genome copies depending on the pathogen) into 10 ng of DNA from the different water samples. This yielded a relative abundance of 0.1% for each of the pathogens, calculated as a mass-based percentage of pathogen DNA in total DNA. Both spiked and raw water samples were subjected to the split multiplex PCR amplification step. The amplicons obtained from spiked water samples were then labeled with Cy3, and the amplicons obtained from the raw water samples labeled with Cy5. Equal amounts both labeled products (100 pmol of each dye or ~2.5 μg of DNA) were mixed and hybridized in duplicate. For the raw water samples, the 2,034 nontarget probes were evaluated to examine probe specificity. For the spiked samples, 673 probes were evaluated to examine probe sensitivity. A second set of samples contained pathogen DNA spiked at a relative abundance of 0.01% into DNA from river water and at a relative abundance of 0.001% in DNA from tertiary effluent.

Data acquisition and processing

Microarrays were scanned with a GenePix 4000B 16-bit laser scanner (Axon instruments, Union City, CA). Fluorescence signal intensities were extracted from the scanned images using GenePix5.0 (Axon Instruments, Union City, CA), which yielded

values between 0 and 65,535 arbitrary units. The median of all pixel intensities within a spot was used as raw spot intensity. Subsequent data analysis was done with Microsoft Excel (Microsoft, Redmond, WA), and plotting was done with SigmaPlot 9.0 (Systat Software, Point Richmond, CA). Raw spot intensities were divided by the mean signal of 22 empty spots to obtain signal-to-noise ratios (SNRs) (Stedtfeld et al., 2007). The SNR was computed for each wash step between 30 and 45°C, and the values subsequently averaged, which is equivalent to calculating an SNR based on the area under the melting curve within this temperature interval. The SNR was subsequently divided by the median SNR of the 2,034 nontarget probes and a probe signal considered positive if the SNR was greater than 3. The PF was computed by dividing the number of positive signal probes within a set by the number of probes within the set (Wilson et al., 2002b).

Gibbs free energy

The ΔG was calculated with the DINAMelt web server (SantaLucia, 1998; Dimitrov and Zuker, 2004; Bartholomeusz et al., 2005; Markham and Zuker, 2005). This parameter reflects the energy released during probe-target duplex formation and is sequence dependent. It provides a thermodynamic measure of the affinity between a probe and a target, with more negative ΔG being indicative of higher binding affinities. For all calculations, perfectly matching duplexes were assumed, and effects of target dangling ends were neglected. The temperature used for the calculations was 43°C instead of the actual hybridization temperature (20°C) to account for the presence of 35% formamide in the hybridization solution (Blake and Delcourt, 1996; Urakawa et al., 2002), and the Na+ concentration used was 1 M. A linear relationship between the G+C content of a probe

(%GC) and its ΔG was observed, which was used to convert the ΔG into a corresponding %GC.

RESULTS AND DISCUSSION

Probe behavior with amplicon mixtures generated by monoplex PCR

The performance of the VMG biochip was evaluated with a target mixture containing all 47 VMG amplicons, obtained in individual PCR amplifications. Of the 47 amplicons, 24 were labeled with Cy3 and 23 were labeled with Cy5. For the Cy3-labeled amplicons, 337 of 393 (85.5%) target probes displayed positive signals in all three replicates. Furthermore, only 10 of the 393 (2.5%) probes displayed positive signals in at least one of the replicates with the Cy5-labeled amplicons (i.e., nontarget amplicons), with a median SNR of 3.4. For the Cy5-labeled amplicons, 383 of 398 (96.2%) target probes yielded positive signals in all three replicates. Furthermore, only 18 of 398 (4.5%) probes displayed positive signals in at least one of the replicates with the Cy3-labeled amplicons (i.e., nontarget amplicons), with a median SNR of 3.5. Of the 2,034 nontarget probes, 27 (1.3%) and 34 (1.7%) displayed positive signals with the Cy3- and Cy5-labeled amplicons, respectively, in at least one of the replicates.

Overall, of the 791 target probes, 720 (91.0%) produced positive signals with SNRs up to 1,000. Of the 2,034 nontarget probes, 61 (3.0%) yielded positive signals with either the Cy3- or Cy5-labeled amplicons. The SNR for target and nontarget probes were well separated, except for a small fraction of the probes (Figure 3.1b), indicating the good discriminatory power of the probes. Altogether, 95% of the 5,650 individual probe-target

interactions (twice the total number of target and nontarget probes) were the expected interactions in terms of positive/negative probe signals.

The PF was calculated for each probe set (and corresponding amplicon) by dividing the number of positive probes by the number of probes within the set. A PF based on the number of initially designed probes is then equivalent to the success level of probe selection for individual probe sets. Of the 47 target probe sets, 22 displayed a PF of 1, implying that 100% of the designed probes yielded positive signals (Figure 3.1c). Of the remaining 25 sets, 18 displayed a PF between 0.8 to 1 and 7 displayed a PF between 0.6 to 0.8. Of the 67 nontarget probe sets, 34 displayed a PF of 0 and 31 displayed a PF between 0 and 0.1 (Figure 3.1c). The two remaining sets had a PF of 0.125 and 0.3. The success level of probe selection could also be illustrated by plotting the number of designed probes *versus* the number of positive probes (Figure 3.1d). As Figure 1 shows, the PF for both target and nontarget probes was not correlated with the number of designed probes.

Substantial variation in signal intensity was observed among probes for a given VMG amplicon (Figure 3.2b). Extensive unevenness in signal intensity among probes for a given target is well documented (Palmer et al., 2006; Pozhitkov et al., 2006; Bruun et al., 2007; Stedtfeld et al., 2007). This variability may be attributed to various factors, including probe and target secondary structure (Chandler et al., 2003; Lane et al., 2004; Ratushna et al., 2005; Becker and Gartner, 2008), target length (Antwerpen et al., 2007; Liu et al., 2007), ΔG (Held et al., 2003), and even the position of fluorescent labels

(Zhang et al., 2005). Also, sequence dissimilarities between the probes and hybridized targets may have contributed to this variability. When probe sets were sorted according to their median SNR, a trend toward increasing PF with a higher median SNR was apparent (Figure 3.2a). Interestingly, the %GC of amplicons displayed an analogous tendency (Figure 3.2b). This also partially explain the overall lower success rate for the probe sets hybridized with the Cy3-labeled amplicons (85.5%) than for the probe sets targeted by the Cy5-labeled amplicons (96.2%). The %GC for Cy3-labeled amplicons (36.5 ± 5.4%) was, on average, lower than the %GC for the Cy5-labeled amplicons (48.6 ± 7.9%). These trends imply that probe design for genes or genomes with a considerably lower %GC may be more challenging when continuous regions with a higher %GC are not present in the selected gene targets.

TABLE 3.1 Overview of the validation results.

Pathogen	Gene	No. of designed probes	Filter 1	Filter 2	Presence/absence calls (PF) in spiked water samples: Tap water (0.1% spike)	Tertiary effluent (0.1% spike)	River water (0.1% spike)	River water (0.01% spike)
A. hydrophila	*alt*	8	6	5	+ (0.60)	+ (1.00)	+ (1.00)	– (0.20)
	ast (1)	15	13	12	+ (0.71)	+ (0.79)	+ (0.71)	+ (0.58)
	ast (2)	10	9	7	+ (0.86)	+ (1.00)	+ (1.00)	– (0.07)
C. jejuni	*cadF*	25	21	21	+ (0.67)	+ (0.79)	+ (0.74)	– (0.26)
	cdtA	14	13	13	+ (0.92)	+ (0.85)	+ (0.92)	+ (0.77)
	cdtC	31	22	21	+ (0.74)	+ (0.88)	+ (0.88)	+ (0.52)
	hipO	14	12	12	+ (0.75)	+ (0.83)	+ (0.79)	– (0.38)
C. perfringens	*cpe* (1)	11	9	9	– (0.00)	– (0.06)	– (0.00)	– (0.11)
	cpe (2)	9	8	7	– (0.29)	– (0.07)	– (0.07)	– (0.00)
	plc (1)	17	14	14	– (0.04)	– (0.07)	– (0.07)	– (0.14)
	plc (2)	16	11	11	– (0.09)	– (0.27)	– (0.36)	– (0.18)
H. pylori	*flaA* (1)	18	18	18	+ (0.89)	+ (0.92)	+ (1.00)	+ (0.75)
	flaA (2)	8	8	7	+ (1.00)	+ (1.00)	+ (1.00)	– (0.21)
	ureB (1)	11	10	10	+ (0.70)	+ (0.85)	+ (0.75)	+ (0.50)

Table 3.1 Continued

	ureB (2)	23	23	23	+ (0.91)	+ (0.87)	+ (0.87)	– (0.28)
L. monocytogenes	*inlA*	19	19	19	+ (0.97)	+ (0.89)	+ (1.00)	– (0.45)
	inlB	10	9	9	+ (0.89)	+ (0.78)	+ (0.67)	– (0.33)
	lisO	17	13	11	+ (0.55)	+ (0.73)	+ (0.55)	– (0.27)
	plcA	32	21	16	+ (0.84)	+ (0.63)	+ (0.56)	– (0.13)
	plcB	18	15	15	+ (0.87)	+ (0.87)	+ (0.83)	– (0.43)
L. pneumophila	*dnaJ* (1)	17	16	16	+ (0.78)	+ (0.75)	+ (0.78)	– (0.28)
	dnaJ (2)	17	17	16	+ (0.81)	+ (0.75)	+ (0.75)	+ (0.53)
	mip	8	7	7	+ (0.71)	+ (0.57)	+ (0.57)	– (0.43)
P. aeruginosa	*oprL*	18	18	16	+ (1.00)	+ (0.94)	+ (1.00)	– (0.28)
Salmonella	*fimA*	21	21	20	+ (1.00)	+ (0.98)	+ (1.00)	+ (0.93)
	hilA (1)	12	12	12	+ (1.00)	+ (1.00)	+ (1.00)	+ (0.75)
	hilA (2)	23	21	21	+ (0.90)	+ (0.95)	+ (0.90)	– (0.45)
	invA	15	15	12	+ (0.92)	+ (0.92)	+ (0.58)	– (0.33)
	invE (1)	18	18	17	+ (0.97)	+ (0.91)	+ (0.65)	– (0.38)
	invE (2)	9	9	9	+ (0.56)	+ (0.72)	+ (0.61)	– (0.22)
S. aureus	*nuc* (1)	10	10	9	– (0.44)	+ (0.50)	+ (0.56)	– (0.0)
	nuc (2)	11	10	10	+ (0.60)	+ (0.60)	+ (0.65)	– (0.3)
	sec	12	9	8	– (0.38)	+ (0.63)	– (0.38)	– (0.19)

Table 3.1 Continued

	tsst (1)	17	17	17	– (0.35)	– (0.38)	– (0.29)	– (0.12)
	tsst (2)	13	10	10	+ (0.50)	+ (0.65)	+ (0.70)	– (0.15)
V. cholerae	*ctxB*	20	20	2	X	X	X	X
	hlyA	15	15	15	+ (0.94)	+ (0.97)	+ (0.94)	– (0.47)
	ompU	18	18	18	+ (0.79)	+ (0.82)	+ (0.53)	– (0.21)
	toxR	35	31	31	+ (0.92)	+ (0.90)	+ (0.85)	– (0.40)
	zot	31	31	31	+ (0.79)	+ (0.95)	+ (0.79)	– (0.45)
V. parahaemolyticus	*tdh*	20	20	19	+ (0.92)	+ (0.97)	+ (0.95)	+ (0.71)
	tlh (1)	19	19	19	+ (0.92)	+ (0.92)	+ (0.73)	– (0.38)
	tlh (2)	13	13	13	+ (0.92)	+ (0.92)	+ (0.95)	– (0.45)
	toxR (1)	20	20	19	+ (1.00)	+ (1.00)	+ (1.00)	+ (0.50)
	toxR (2)	12	12	12	+ (0.62)	+ (0.87)	+ (0.70)	– (0.30)
Y. enterocolitica	*ail*	32	30	30	+ (0.67)	+ (0.67)	+ (0.75)	– (0.33)
	ystA	7	7	6	+ (0.60)	+ (1.00)	+ (1.00)	– (0.20)
Total		**791**	**720**	**673**				

For each pathogen and VMG amplicon, the number of designed probes and the number of positive probes after hybridization of amplicons generated by monoplex (filter 1) and multiplex PCR (filter 2) are indicated. The presence (+) or absence (–) call for each amplicon was calculated based on the number of probes selected after application of filter 2.

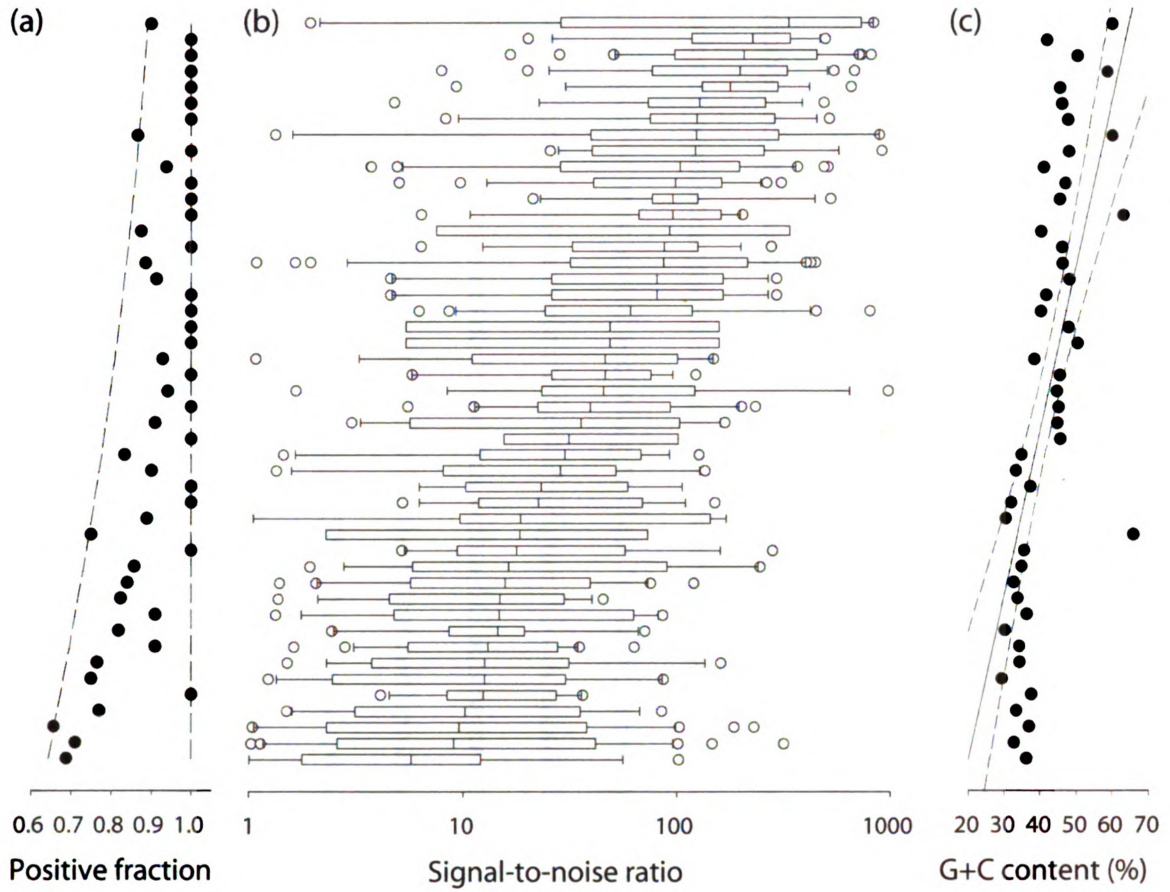


FIGURE 3.2 PF (a), SNR (b), and %GC (c) for all 47 target genes. Probe sets are sorted, from bottom to top, according to increasing median SNR. The boundaries of the boxes in panel b indicate the 25th and the 75th percentiles, and the whiskers indicate the 10th and the 90th percentiles. The median is given as a solid line, and outlying data points are shown as open symbols.

Success rate of probe design as a function of ΔG

In order to optimize the probe selection criteria, whether probes with positive or negative signals could be identified *in silico* based on evaluation of their theoretical (thermodynamic) properties was evaluated. Based on previous studies, it was anticipated that probe-target binding affinity (quantified as ΔG) would be the most valuable parameter for this screening. Only a weak linear relationship ($r^2 = 0.34$) was observed between ΔG and the natural logarithm of signal intensity (data not shown). This variability in signal intensity for probes with comparable ΔG precluded evaluation of probe design criteria in terms of signal intensity based on ΔG. To determine the relationship between ΔG and probe behavior, hybridization patterns were interpreted in terms of positive/negative signals, irrespective of their strength. For this analysis, all 791 target probes were sorted according to their ΔG and binned, and the percentage of positive probes in each bin calculated (Figure 3.3). A drastic decrease in the probe design success level was observed for probes with a ΔG less negative than –17 kcal/mol and a %GC less than 34.4 (as estimated using the linear relationship between %GC content and ΔG). A similar analysis with increasing SNR thresholds revealed analogous trends, but the results were shifted toward more negative ΔG (data not shown). The factors contributing to the unevenness in probe signal for a given target may also explain the lack of hybridization signal for probes with a highly negative ΔG. In accordance with our observations, Reyes-Lopez et al. (2003) previously reported that the proportion of predicted signals observed experimentally increased for 9-mer probes with increasingly negative ΔG.

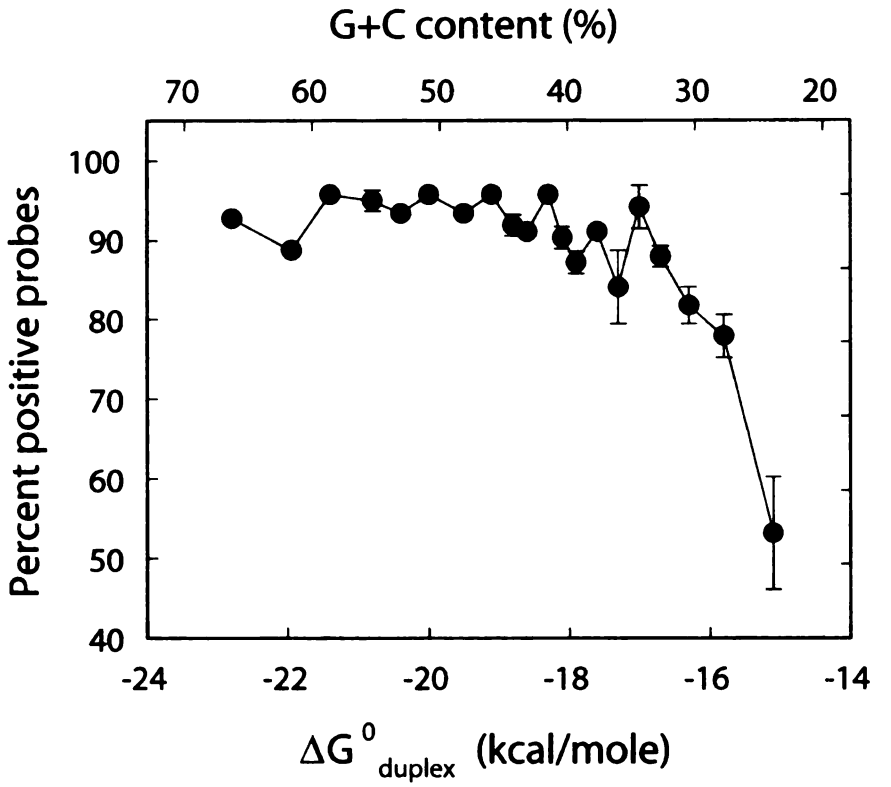


FIGURE 3.3 Success of probe design as a function of ΔG. Each symbol indicates the percentage of positive probes for bins of 40 probes, and the error bars indicate the standard deviations for three replicates. The percentage of positive probes is plotted at the median ΔG of each bin. The %GC was derived from the linear relationship between ΔG and %GC.

Based on this analysis, a ΔG threshold of –17 kcal/mol could be a valuable threshold for future oligonucleotide probe design. Although this criterion was demonstrated only for 18-mers in this study, a similar approach to assess the effect of ΔG on the success level of probe design could be adopted for longer probes. Also, this threshold is in accordance with the findings of Loy et al. (2005), who proposed –16 kcal/mol for 18-mers targeting the 16S rRNA gene. This suggestion was based on the observation that the majority of probes displaying positive signals (including cross-hybridization signals) were attributed to probe-target duplexes with a ΔG more negative than –16 kcal/mol. It should be noted that small deviations from these ΔG design thresholds may be observed

for hybridizations performed under different experimental conditions and/or for ΔG estimates obtained using other nearest-neighbor model parameters. Finally, the ΔG criterion was derived from hybridization patterns with high-abundance and low-complexity target mixtures. For target mixtures having very different compositions in terms of target sequence abundance and diversity, these rules should be applied with caution.

Multiplex amplification

A split multiplex PCR assay was developed to amplify all VMG amplicons. Of the 720 probes displaying positive signals after hybridization of monoplex PCR amplicons, 673 (93.5%) yielded positive signals after hybridization of multiplex PCR amplicons (Table 3.1). For 28 of the 47 probe sets, all probes yielded positive signals. In general, the probes that displayed negative signals with the multiplex PCR amplicons also yielded low signals with the monoplex PCR amplicons (data not shown). For the *ctxB* gene of *V. cholerae*, only 2 of 20 probes yielded positive hybridization signals after hybridization of amplicons generated by multiplex PCR. This was attributed to poor amplification in the multiplex PCR rather than to low probe hybridization efficiency. With monoplex PCR, an amplification product of the expected length was observed, verifying that the designed primers successfully amplified the *ctxB* gene of *V. cholerae* ATCC 39315. In addition, all probes targeting the *ctxB* gene yielded positive signals after hybridization of the *ctxB* gene amplicon generated by monoplex PCR. The 27 targeted probes (3.9%, excluding probes targeting the *ctxB* gene) displaying negative signals were distributed among multiple VMG amplicons, and hence redesign or further optimization of the multiplex

PCR assays was expected to be ineffective. In addition, the low levels of the signals of these probes with the monoplex PCR amplicons suggested that their sensitivity may be limited; therefore, these probes were masked during further analysis.

Performance of the VMG biochip with spiked water samples

The performance of the VMG biochip probe sets was further tested with samples containing pathogen DNA spiked into DNA extracted from three different water samples: tap water, tertiary effluent from a wastewater treatment plant, and river water. For pathogens spiked at a relative abundance level of 0.1%, the median PFs for 46 targeted VMG amplicons were 0.79, 0.87, and 0.75 for tap water, tertiary effluent, and river water, respectively (Figure 3.4). The median PFs for the 67 other probe sets were 0.05 for tap water, 0.09 for tertiary effluent, and 0.17 for river water (Figure 3.4). By applying a PF threshold of 0.5, the numbers of spiked VMG amplicons assigned as present were 39 for tap water (85%), 40 for tertiary effluent (87%), and 39 for river water (85%). For the nontarget probe sets, the PF was always less than 0.5, and consequently none of the corresponding genes were identified as present in the water samples. Probe sets targeting S. aureus VMG amplicons indicated the presence of this pathogen in all three spiked water samples, although the VMG amplicons assigned as present were in disagreement among samples. The VMG amplicons of *C. perfringens* could not be detected in any of the spiked water samples.

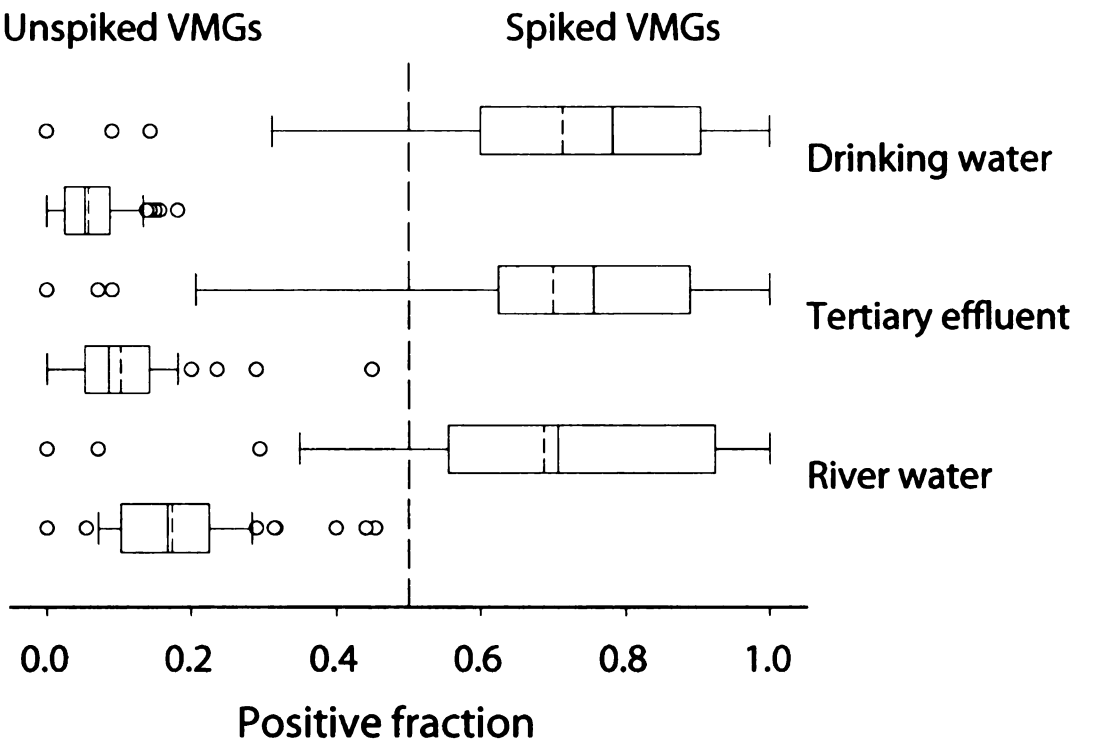


FIGURE 3.4 Performance of the VMG biochip with DNA from multiple water sources supplemented with pathogen DNA. DNA (10 pg) of each pathogen was spiked into 10 ng of DNA from different water sources, yielding a relative abundance for each pathogen of 0.1%. For spiked samples, 673 probes (46 VMG amplicons) were analyzed. For raw samples, 2,034 probes (67 VMG amplicons) were analyzed. The box plots indicate the distribution of the PF for target and non-target probe sets. The dashed line at a PF of 0.5 indicates the selected threshold for presence/absence calls. The boundaries of the boxes indicate the 25th and 75th percentiles, and the whiskers indicate the 10th and 90th percentiles. The median is given as a solid line, the mean is shown as a dashed line, and outlying data points are shown as open symbols.

When pathogen DNA was spiked at a relative abundance of 0.01% into DNA extracted from river water, the presence/absence calls for the VMG amplicons varied for a given pathogen (Table 3.1). A similar observation was made previously by Wilson et al. (2002b) and was explained by the variable sensitivity among probes. These researchers observed that for pathogens spiked at a fractional abundance of 0.025% (500 fg of pathogen DNA in 2 ng of DNA from air), only five out of eight probe sets for *Francisella tularensis* and three out of eight probe sets for *Y. pestis* were assigned as positive. For

pathogen DNA spiked into tertiary effluent DNA at a relative abundance level of 0.001% none of the target genes could be detected (data not shown). Hence, the detection limit of the VMG biochip was 0.1% to 0.01%, depending on the pathogen and VMG. This detection limit is within the range of sensitivity previously reported for microbial diagnostic arrays targeting either the 16S rRNA gene (Maynard et al., 2005; Palmer et al., 2006; Sanguin et al., 2006a) or multiple VMGs (Wilson et al., 2002b). However, advances that enhance the sensitivity are required to achieve reliable detection at abundance levels approaching the minimum infectious dose in environmental samples. These advances include improved experimental techniques for target gene enrichment and/or microarray signal enhancement and optimized data analysis algorithms.

The PF threshold applied in this study was less stringent than the cutoff values (at least 0.8) used by previous researchers (Wilson et al., 2002b; DeSantis et al., 2005). Both target gene sequence diversity and abundance affect the selection of an optimal threshold for the PF. Uncharacterized sequence diversity within the target gene may lead to reduced PFs due to the increased potential for mismatching probe-target duplexes. Low target gene abundance may also yield decreased PFs due to variability in probe sensitivity. In both cases, a lower threshold PF needs to be used to reduce false-negative calls. Although the concept of redundant probe sets for enhanced reliability in presence/absence calls has been exploited widely, the effect of target gene abundance on PFs is less well documented. As demonstrated in this study and by Wilson et al. (2002b), lower PFs are expected with decreasing target gene amounts, and the extent of the decrease may vary among probe sets.

The use of mismatch probes is a common strategy to identify cross-hybridization signals on microarrays and to enhance the reliability of presence/absence calls. In this study, excellent specificity was observed without inclusion of mismatch probes, even for environmental samples. A combination of various factors, including the use of redundant probe sets (with an average of 15 probes per VMG amplicon after experimental screening), the use of short probes with increased discriminatory power compared to long probes, reduction of sample complexity by target gene enrichment using multiplex PCR, and use of the melting curve approach, contributed to the elimination of false-positive calls.

Probe selectivity as a function of ΔG

A selective probe should yield high hybridization signals with target sequences and low signals with nontarget sequences. However, due to the biochemical nature of the probe-target hybridization process, these two criteria cannot be optimized independently. This translates into a trade-off between probe specificity and sensitivity that must be considered. Based on our analysis shown in Figure 3.3, ΔG was selected as the probe design parameter quantifying the trade-off between probe specificity and sensitivity. Probe sensitivity was derived from the hybridization patterns of the target probes, while probe specificity was quantified based on the hybridization patterns of the nontarget probes. The hybridization results with the raw tap water sample were omitted from this analysis due the minimal amount of nontarget probes with positive signals. The hybridization results for all three water samples spiked with pathogen DNA at a relative abundance of 0.1% were included in the analysis.

Both target and nontarget probes were sorted according to their ΔG and binned, and the percentage of positive probes computed for each bin. When the percentage of positive probes was plotted as a function of ΔG, two distinct regions were apparent (Figure 3.5b), and this effect was independent of the window size used for probe binning (data not shown). In region 1 (ΔG more negative than –19.3 kcal/mol), 93.8±2.8% of the target probes displayed positive signals, independent of the ΔG. The percentage of nontarget probes with positive signals was influenced more by ΔG and increased from 14.7% for probes with a ΔG of –19.9 kcal/mol to 42.7% for probes with a ΔG of –23.2 kcal/mol (average increase, 6.8% per ΔG). In region 2 (ΔG less negative than –19.3 kcal/mol), the percentage of positive target probes decreased rapidly, and only 31.3% of the probes yielded positive signals for probes with a ΔG of –15.6 kcal/mol (average decrease, –16.1% per ΔG). The percentage of nontarget probes with positive signal was significantly lower in this region (average, 9.2±3.9%).

As evident from Figure 3.5b, an increase in the percentage of positive target probes was always associated with an increase in the percentage of positive nontarget probes. Probe selectivity, reflecting the ability of a probe to detect intended target sequences and exclude nontarget sequences, quantifies this trade-off and is the parameter that needs to be maximized. Probe selectivity, calculated by determining the difference in the percentages of positive probes for target and nontarget probes as a function of ΔG, was estimated based on the linear regression lines in Figure 3.5b. The highest selectivity (~80%) was observed for probes with a ΔG of –19.3 kcal/mol and a %GC of 47.2 (Figure 3.5a). Thus, probes with a ΔG of –19.3 kcal/mol provide the best trade-off between

sensitivity and specificity. Probes with a ΔG deviating from this optimum displayed lower selectivity. The selectivity was more than 70% for probes with a ΔG between –18.6 and –21.1 kcal/mol, which correspond to a %GC content between 42.3 and 56.1. Interestingly, the decrease in selectivity in region 2 was approximately two-fold greater than that in region 1 (–6.6% per ΔG for region 1 and –14.0% per ΔG for region 2). It should be noted that for DNA samples containing more complex nontarget sequences, the increase in the percentage of positive nontarget probes (i.e., cross-hybridization) for increasingly negative ΔG may be more pronounced, while for more abundant target sequences the decrease in the percentage of positive target probes for decreasingly negative ΔG may be suppressed.

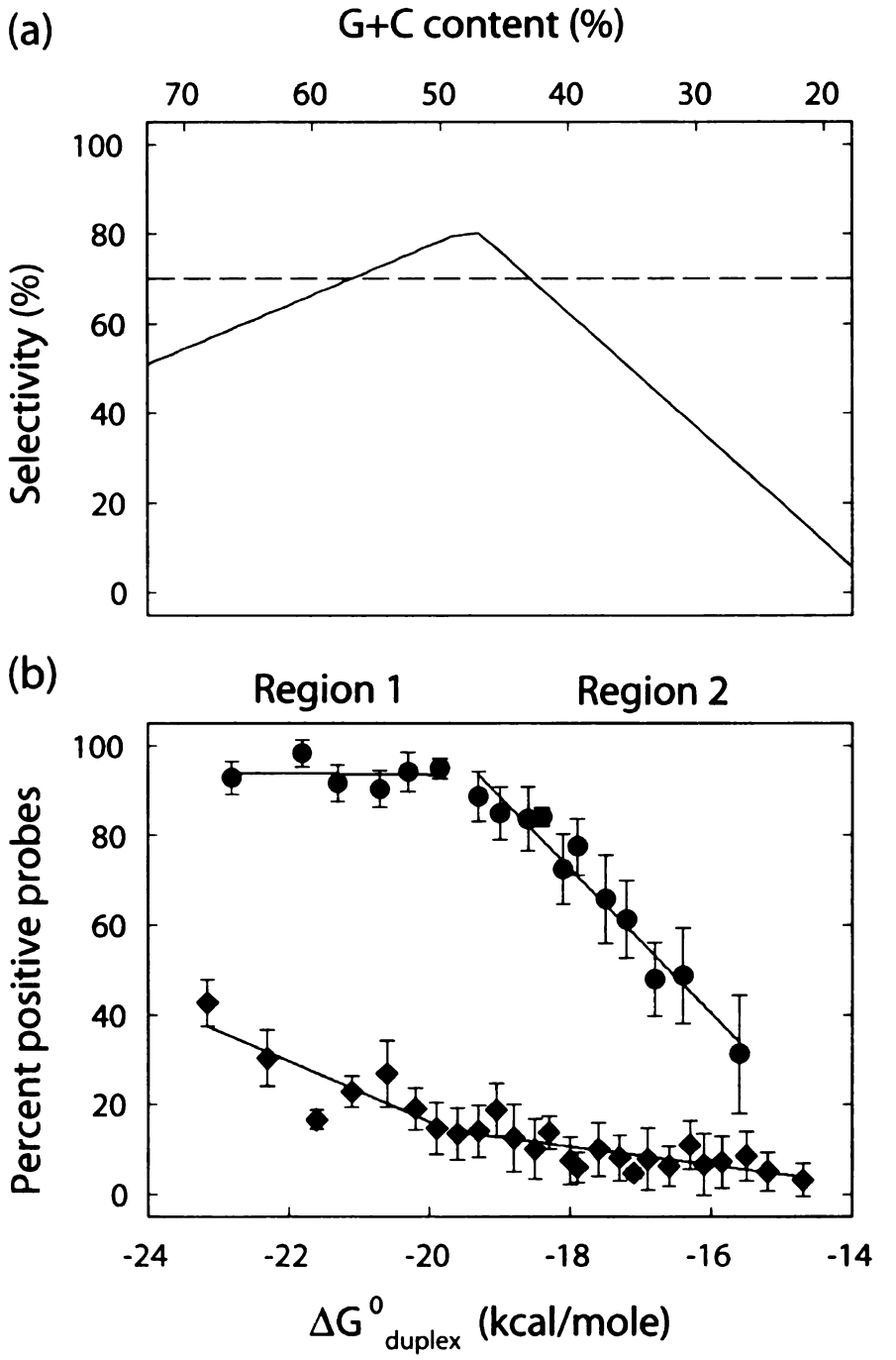


FIGURE 3.5 Derivation of optimal probe design criteria. (a) Selectivity, expressed as the difference in the percentage of positive probes for target and nontarget probes, as a function of ΔG. (b) Percentage of positive probes as a function of ΔG for target and nontarget probes (• and ♦, respectively). Each symbol indicates the percentage of positive probes for bins of 40 target and 80 nontarget probes, with the error bars representing the standard deviation for samples and replicates (n = 6 for target probes and n = 4 for nontarget probes). The %GC was derived from the linear relationship between ΔG and %GC.

A number of studies have demonstrated the effectiveness of ΔG as a theoretical probe selection parameter for both long and short probes and RNA or DNA targets (Luebke et al., 2003; Matveeva et al., 2003; Reyes-Lopez et al., 2003; Taroncher-Oldenburg et al., 2003; Rhee et al., 2004; Tiquia et al., 2004; He et al., 2005; Loy et al., 2005; Liebich et al., 2006; Antwerpen et al., 2007). In contrast, Pozhitkov et al. (2006) recently suggested that all theoretical (thermodynamics-based) screening of oligonucleotide probes should be omitted due to poor correlations between ΔG (and ΔG for intra- and intermolecular self-structures) and experimentally observed signal strengths for rRNA targets. Similarly, only a weak linear correlation was observed between probe signal intensity and ΔG (data not shown). This indicates that hybridization signals cannot be accurately predicted with current thermodynamics models derived from hybridizations in solution. However, thermodynamic parameters may still be indicative of probe behavior by reflecting the probability that a probe will yield signal intensity higher than a given threshold. Based on the effect of ΔG on probe selectivity, an optimal ΔG of –19.3 kcal/mol is suggested for 18-mer oligonucleotide probes under the described hybridization conditions. These probes should yield the highest confidence in presence/absence calls for a given number of probes per target gene. For probes with either more or less negative ΔG with lower selectivity, more probes are needed to attain high confidence in presence/absence calls. In general, probes with a more negative ΔG displayed better sensitivity but poorer specificity, which is interestingly analogous to the effect of length. In general, long probes (50- to 70-mers) are more sensitive than short probes (15- to 30-mers) but display lower specificity.

Cost and flexibility in microarray synthesis are often recognized as two of the major limiting factors for the adaptation of microarray technology for diagnostic purposes (Hashsham et al., 2004; Call, 2005). In this study, a maskless light-directed *in situ* microarray synthesis technology developed at the University of Michigan was employed. This technology employs a digital micromirror device to generate preprogrammed light patterns on the chip surface, triggering deprotection of the 5'-hydroxyl group in conventional phosphoramidite monomers. Synthesis of oligonucleotides using this chemistry provides high fidelity and a stepwise yield and also allows synthesis of probes that are up 100 nucleotides long. In addition, synthesis of new chips is low cost, rapid, and flexible and involves simply uploading a list of probe sequences in the optical unit. Other advantages of in situ probe synthesis include high spot uniformity and probe molecule density. In addition, continuous recycling of the hybridization solution in the microfluidic chips used in this study provides increased signal uniformity within a spot and increased reproducibility of the hybridization signals. The higher cost per chip compared to conventional glass slides is the major limitation of this platform. The ability to rapidly synthesize biochips with updated and reiterated probe sets is considered critical for diagnostic purposes as new gene sequence information is appearing almost daily and this information should be incorporated in the probe selection exercise and data analysis as soon as it becomes available.

CONCLUSIONS

In summary, we developed and evaluated a coupled format of multiplex PCR and DNA biochip for simultaneous detection of 12 bacterial pathogens in water. Because of

the use of redundant probe sets targeting multiple VMGs, false-positives eliminated. Pathogens could be detected at a relative abundance of 0.1 to 0.01%, depending on the pathogen. Analysis of the hybridization patterns also showed that probes with a ΔG of -19.3 kcal/mole provided the best trade-off between sensitivity and specificity. In future studies, the VMG biochip will be applied to additional environmental samples, and the presence/absence calls will be verified independently using real-time PCR.

REFERENCES

Al-Khaldi, S.F., K.M. Myers, A. Rasooly and V. Chizhikov, 2004. Genotyping of *Clostridium perfringens* toxins using multiple oligonucleotide microarray hybridization. Molecular and Cellular Probes, 18: 359-367.

Antwerpen, M.H., M. Schellhase, E. Ehrentreich-Forster, F. Bier, W. Witte and U. Nubel, 2007. DNA microarray for detection of antibiotic resistance determinants in Bacillus anthracis and closely related *Bacillus cereus*. Molecular and Cellular Probes, 21: 152-160.

Blake, R.D. and S.G. Delcourt, 1996. Thermodynamic effects of formamide on DNA stability. Nucleic Acids Research, 24: 2095-2103.

Bodrossy, L. and A. Sessitsch, 2004. Oligonucleotide microarrays in microbial diagnostics. Current Opinion in Microbiology, 7: 245-254.

Bruun, G.M., R. Wernersson, A.S. Juncker, H. Willenbrock and H.B. Nielsen, 2007. Improving comparability between microarray probe signals by thermodynamic intensity correction. Nucleic Acids Research, 35: 8.

Call, D.R., 2005. Challenges and opportunities for pathogen detection using DNA microarrays. Critical Reviews in Microbiology, 31: 91-99.

Call, D.R., M.K. Borucki and F.J. Loge, 2003. Detection of bacterial pathogens in environmental samples using DNA microarrays. Journal of Microbiological Methods, 53: 235-243.

Call, D.R., F.J. Brockman and D.P. Chandler, 2001. Detecting and genotyping *Escherichia coli* O157 : H7 using multiplexed PCR and nucleic acid microarrays. International Journal of Food Microbiology, 67: 71-80.

Chandler, D.P., G.J. Newton, J.A. Small and D.S. Daly, 2003. Sequence versus structure for the direct detection of 16S rRNA on planar oligonucleotide microarrays. Applied and Environmental Microbiology, 69: 2950-2958.

Chizhikov, V., A. Rasooly, K. Chumakov and D.D. Levy, 2001. Microarray analysis of microbial virulence factors. Applied and Environmental Microbiology, 67: 3258-3263.

Denef, V.J., J. Park, J.L.M. Rodrigues, T.V. Tsoi, S.A. Hashsham and J.M. Tiedje, 2003. Validation of a more sensitive method for using spotted oligonucleotide

DNA microarrays for functional genomics studies on bacterial communities. Environmental Microbiology, 5: 933-943.

DeSantis, T.Z., C.E. Stone, S.R. Murray, J.P. Moberg and G.L. Andersen, 2005. Rapid quantification and taxonomic classification of environmental DNA from both prokaryotic and eukaryotic origins using a microarray. FEMS Microbiology Letters, 245: 271-278.

Dimitrov, R.A. and M. Zuker, 2004. Prediction of hybridization and melting for double-stranded nucleic acids. Biophysical Journal, 87: 215-226.

El Fantroussi, S., H. Urakawa, A.E. Bernhard, J.J. Kelly, P.A. Noble, H. Smidt, G.M. Yershov and D.A. Stahl, 2003. Direct profiling of environmental microbial populations by thermal dissociation analysis of native rRNAs hybridized to oligonucleotide microarrays. Applied and Environmental Microbiology, 69: 2377-2382.

Eom, H.S., B.H. Hwang, D.H. Kim, I.B. Lee, Y.H. Kim and H.J. Cha, 2007. Multiple detection of food-borne pathogenic bacteria using a novel 16S rDNA-based oligonucleotide signature chip. Biosensors & Bioelectronics, 22: 845-853.

Franke-Whittle, I.H., S.H. Klammer and H. Insam, 2005. Design and application of an oligonucleotide microarray for the investigation of compost microbial communities. Journal of Microbiological Methods, 62: 37-56.

Gao, X.L., E. Gulari and X.C. Zhou, 2004. *In situ* synthesis of oligonucleotide microarrays. Biopolymers, 73: 579-596.

Gao, X.L., E. LeProust, H. Zhang, O. Srivannavit, E. Gulari, P.L. Yu, C. Nishiguchi, Q. Xiang and X.C. Zhou, 2001. A flexible light-directed DNA chip synthesis gated by deprotection using solution photogenerated acids. Nucleic Acids Research, 29: 4744-4750.

Gonzalez, S.F., M.J. Krug, M.E. Nielsen, Y. Santos and D.R. Call, 2004. Simultaneous detection of marine fish pathogens by using multiplex PCR and a DNA microarray. Journal of Clinical Microbiology, 42: 1414-1419.

Hashsham, S.A., L.M. Wick, J.M. Rouillard, E. Gulari and J.M. Tiedje, 2004. Potential of DNA microarrays for developing parallel detection tools (PDTs) for microorganisms relevant to biodefense and related research needs. Biosensors & Bioelectronics, 20: 668-683.

He, Z.L., L.Y. Wu, X.Y. Li, M.W. Fields and J.Z. Zhou, 2005. Empirical establishment of oligonucleotide probe design criteria. Applied and Environmental Microbiology, 71: 3753-3760.

Held, G.A., G. Grinstein and Y. Tu, 2003. Modeling of DNA microarray data by using physical properties of hybridization. Proceedings of the National Academy of Sciences of the United States of America, 100: 7575-7580.

Keramas, G., D.D. Bang, M. Lund, M. Madsen, S.E. Rasmussen, H. Bunkenborg, P. Telleman and C.B.V. Christensen, 2003. Development of a sensitive DNA microarray suitable for rapid detection of *Campylobacter* spp. Molecular and Cellular Probes, 17: 187-196.

Kostic, T., A. Weilharter, A. Sessitsch and L. Bodrossy, 2005. High-sensitivity, polymerase chain reaction-free detection of microorganisms and their functional genes using 70-mer oligonucleotide diagnostic microarray. Analytical Biochemistry, 346: 333-335.

Lane, S., J. Evermann, F. Loge and D.R. Call, 2004. Amplicon secondary structure prevents target hybridization to oligonucleotide microarrays. Biosensors & Bioelectronics, 20: 728-735.

Li, E.S.Y., J.K.K. Ng, J.H. Wu and W.T. Liu, 2004. Evaluating single-base-pair discriminating capability of planar oligonucleotide microchips using a non-equilibrium dissociation approach. Environmental Microbiology, 6: 1197-1202.

Liebich, J., C.W. Schadt, S.C. Chong, Z.L. He, S.K. Rhee and J.Z. Zhou, 2006. Improvement of oligonucleotide probe design criteria for functional gene microarrays in environmental applications. Applied and Environmental Microbiology, 72: 1688-1691.

Lin, B.C., K.M. Blaney, A.P. Malanoski, A.G. Ligler, J.M. Schnur, D. Metzgar, K.L. Russell and D.A. Stenger, 2007. Using a resequencing microarray as a multiple respiratory pathogen detection assay. Journal of Clinical Microbiology, 45: 443-452.

Lin, B.C., G.J. Vora, D. Thach, E. Walter, D. Metzgar, C. Tibbetts and D.A. Stenger, 2004. Use of oligonucleotide microarrays for rapid detection and serotyping of acute respiratory disease-associated adenoviruses. Journal of Clinical Microbiology, 42: 3232-3239.

Liu, W.T., H.L. Guo and J.H. Wu, 2007. Effects of target length on the hybridization efficiency and specificity of rRNA-based oligonucleotide microarrays. Applied and Environmental Microbiology, 73: 73-82.

Liu, W.T., A.D. Mirzabekov and D.A. Stahl, 2001. Optimization of an oligonucleotide microchip for microbial identification studies: a non-equilibrium dissociation approach. Environmental Microbiology, 3: 619-629.

Loy, A. and L. Bodrossy, 2006. Highly parallel microbial diagnostics using oligonucleotide microarrays. Clinica Chimica Acta, 363: 106-119.

Loy, A., K. Kusel, A. Lehner, H.L. Drake and M. Wagner, 2004. Microarray and functional gene analyses of sulfate-reducing prokaryotes in low-sulfate, acidic fens reveal cooccurrence of recognized genera and novel lineages. Applied and Environmental Microbiology, 70: 6998-7009.

Loy, A., A. Lehner, N. Lee, J. Adamczyk, H. Meier, J. Ernst, K.H. Schleifer and M. Wagner, 2002. Oligonucleotide microarray for 16S rRNA gene-based detection of all recognized lineages of sulfate-reducing prokaryotes in the environment. Applied and Environmental Microbiology, 68: 5064-5081.

Loy, A., C. Schulz, S. Lucker, A. Schopfer-Wendels, K. Stoecker, C. Baranyi, A. Lehner and M. Wagner, 2005. 16S rRNA gene-based oligonucleotide microarray for environmental monitoring of the betaproteobacterial order "*Rhodocyclales*". Applied and Environmental Microbiology, 71: 1373-1386.

Luebke, K.J., R.P. Balog and H.R. Garner, 2003. Prioritized selection of oligodeoxyribonucleotide probes for efficient hybridization to RNA transcripts. Nucleic Acids Research, 31: 750-758.

Markham, N.R. and M. Zuker, 2005. DINAMelt web server for nucleic acid melting prediction. Nucleic Acids Research, 33: W577-W581.

Markoulatos, P., N. Siafakas and M. Moncany, 2002. Multiplex polymerase chain reaction: A practical approach. Journal of Clinical Laboratory Analysis, 16: 47-51.

Matveeva, O.V., S.A. Shabalina, V.A. Nemtsov, A.D. Tsodikov, R.F. Gesteland and J.F. Atkins, 2003. Thermodynamic calculations and statistical correlations for oligo-probes design. Nucleic Acids Research, 31: 4211-4217.

Maynard, C., F. Berthiaume, K. Lemarchand, J. Harel, P. Payment, P. Bayardelle, L. Masson and R. Brousseau, 2005. Waterborne pathogen detection by use of oligonucleotide-based microarrays. Applied and Environmental Microbiology, 71: 8548-8557.

Palmer, C., E.M. Bik, M.B. Eisen, P.B. Eckburg, T.R. Sana, P.K. Wolber, D.A. Relman and P.O. Brown, 2006. Rapid quantitative profiling of complex microbial populations. Nucleic Acids Research, 34.

Panicker, G., D.R. Call, M.J. Krug and A.K. Bej, 2004. Detection of pathogenic *Vibrio* spp. in shellfish by using multiplex PCR and DNA microarrays. Applied and Environmental Microbiology, 70: 7436-7444.

Pozhitkov, A., P.A. Noble, T. Domazet-Loso, A.W. Nolte, R. Sonnenberg, P. Staehler, M. Beier and D. Tautz, 2006. Tests of rRNA hybridization to microarrays suggest that hybridization characteristics of oligonucleotide probes for species discrimination cannot be predicted. Nucleic Acids Research, 34.

Ratushna, V.G., J.W. Weller and C.J. Gibas, 2005. Secondary structure in the target as a confounding factor in synthetic oligomer microarray design. BMC Genomics, 6.

Reyes-Lopez, M., A. Mendez-Tenorio, R. Maldonaldo-Rodriguez, M.J. Doktycz, J.T. Fleming and K.L. Beattie, 2003. Fingerprinting of prokaryotic 16S rRNA genes using oligodeoxyribonucleotide microarrays and virtual hybridization. Nucleic Acids Research, 31: 779-789.

Rhee, S.K., X.D. Liu, L.Y. Wu, S.C. Chong, X.F. Wan and J.Z. Zhou, 2004. Detection of genes involved in biodegradation and biotransformation in microbial communities by using 50-mer oligonucleotide microarrays. Applied and Environmental Microbiology, 70: 4303-4317.

Sanguin, H., A. Herrera, C. Oger-Desfeux, A. Dechesne, P. Simonet, E. Navarro, T.M. Vogel, Y. Moenne-Loccoz, X. Nesme and G.L. Grundmann, 2006a. Development and validation of a prototype 16S rRNA-based taxonomic microarray for Alphaproteobacteria. Environmental Microbiology, 8: 289-307.

Sanguin, H., B. Remenant, A. Dechesne, J. Thioulouse, T.M. Vogel, X. Nesme, Y. Moenne-Loccoz and G.L. Grundmann, 2006b. Potential of a 16S rRNA-based taxonomic microarray for analyzing the rhizosphere effects of maize on *Agrobacterium* spp. and bacterial communities. Applied and Environmental Microbiology, 72: 4302-4312.

SantaLucia, J., 1998. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. Proceedings of the National Academy of Sciences of the United States of America, 95: 1460-1465.

Sergeev, N., M. Distler, S. Courtney, S.F. Al-Khaldi, D. Volokhov, V. Chizhikov and A. Rasooly, 2004. Multipathogen oligonucleotide microarray for environmental and biodefense applications. Biosensors & Bioelectronics, 20: 684-698.

Siripong, S., J.J. Kelly, D.A. Stahl and B.E. Rittmann, 2006. Impact of prehybridization PCR amplification on microarray detection of nitrifying bacteria in wastewater treatment plant samples. Environmental Microbiology, 8: 1564-1574.

Stedtfeld, R.D., L.M. Wick, S.W. Baushke, D.M. Tourlousse, A.B. Herzog, Y.M. Xia, J.M. Rouillard, J.A. Klappenbach, J.R. Cole, E. Gulari, J.M. Tiedje and S.A. Hashsham, 2007. Influence of dangling ends and surface-proximal tails of targets on probe-target duplex formation in 16S rRNA gene-based diagnostic arrays. Applied and Environmental Microbiology, 73: 380-389.

Taroncher-Oldenburg, G., E.M. Griner, C.A. Francis and B.B. Ward, 2003. Oligonucleotide microarray for the study of functional gene diversity in the nitrogen cycle in the environment. Applied and Environmental Microbiology, 69: 1159-1171.

Tiquia, S.M., L.Y. Wu, S.C. Chong, S. Passovets, D. Xu, Y. Xu and J.Z. Zhou, 2004. Evaluation of 50-mer oligonucleotide arrays for detecting microbial populations in environmental samples. Biotechniques, 36: 664-+.

Tourlousse, D.M., R.D. Stedtfeld, S.W. Baushke, L.M. Wick and S.A. Hashsham, 2007. Virulence factor activity relationships: Challenges and development approaches. Water Environment Research, 79: 246-259.

Urakawa, H., P.A. Noble, S. El Fantroussi, J.J. Kelly and D.A. Stahl, 2002. Single-base-pair discrimination of terminal mismatches by using oligonucleotide microarrays and neural network analyses. Applied and Environmental Microbiology, 68: 235-244.

Volokhov, D., A. Rasooly, K. Chumakov and V. Chizhikov, 2002. Identification of *Listeria* species by microarray-based assay. Journal of Clinical Microbiology, 40: 4720-4728.

Vora, G.J., C.E. Meador, M.M. Bird, C.A. Bopp, J.D. Andreadis and D.A. Stenger, 2005. Microarray-based detection of genetic heterogeneity, antimicrobial resistance, and the viable but nonculturable state in human pathogenic *Vibrio* spp. Proceedings of the National Academy of Sciences of the United States of America, 102: 19109-19114.

Warsen, A.E., M.J. Krug, S. LaFrentz, D.R. Stanek, F.J. Loge and D.R. Call, 2004. Simultaneous discrimination between 15 fish pathogens by using 16S ribosomal DNA PCR and DNA microarrays. Applied and Environmental Microbiology, 70: 4216-4221.

Wick, L.M., J.M. Rouillard, T.S. Whittam, E. Gulari, J.M. Tiedje and S.A. Hashsham, 2006. On-chip non-equilibrium dissociation curves and dissociation rate constants as methods to assess specificity of oligonucleotide probes. Nucleic Acids Research, 34.

Wilson, K.H., W.J. Wilson, J.L. Radosevich, T.Z. DeSantis, V.S. Viswanathan, T.A. Kuczmarski and G.L. Andersen, 2002a. High-density microarray of small-subunit ribosomal DNA probes. Applied and Environmental Microbiology, 68: 2535-2541.

Wilson, W.J., C.L. Strout, T.Z. DeSantis, J.L. Stilwell, A.V. Carrano and G.L. Andersen, 2002b. Sequence-specific identification of 18 pathogenic microorganisms using microarray technology. Molecular and Cellular Probes, 16: 119-127.

Wu, L.Y., D.K. Thompson, G.S. Li, R.A. Hurt, J.M. Tiedje and J.Z. Zhou, 2001. Development and evaluation of functional gene arrays for detection of selected genes in the environment. Applied and Environmental Microbiology, 67: 5780-5790.

Zhang, L., T. Hurek and B. Reinhold-Hurek, 2005. Position of the fluorescent label is a crucial factor determining signal intensity in microarray hybridizations. Nucleic Acids Research, 33.

CHAPTER FOUR

REAL TIME FLUOROGENIC LOOP-MEDIATED ISOTHERMAL AMPLIFICATION IN POLYMER CHIPS USING AN INEXPENSIVE AND COMPACT MULTIPLEXED OPTICAL SENSOR

INTRODUCTION

In developed nations, nucleic acid amplification testing (NAAT) is rapidly replacing conventional techniques for diagnosis of infectious diseases due to its superior speed, sensitivity and specificity. Presently, NAAT is however too costly and complex to be employed in developing countries, where the burden of infectious diseases is most severe (El Ekiaby et al., 2010). A pressing need therefore exists for more affordable NAAT devices that can be used in low-resource settings and by minimally trained personnel at point of care (Varmus et al., 2003; Weigl et al., 2009). Low cost microfluidics and lab-on-a-chip (Auroux et al., 2004; Zhang et al., 2006; Chen et al., 2007) along with compact and low-cost optical sensors for assay readout (Kuswandi et al., 2007) are the main technologies enabling development of such devices.

Devices for NAAT can furthermore be simplified by using isothermal methods for nucleic acid amplification rather than the common polymerase chain reaction (PCR), which requires thermal cycling. Several such techniques have been developed in recent years, including helicase-dependent amplification (Vincent et al., 2004), recombinase polymerase reaction (Piepenburg et al., 2006), and loop-mediated isothermal amplification (Notomi et al., 2000; LAMP). The latter technique in particular (i.e.,

LAMP) has attracted tremendous attention for simple and robust NAAT (Mori and Notomi, 2009). Aside from its isothermal character, appealing features of LAMP include: excellent specificity due to the use of four to six specific primers and relatively high amplification temperature, need for only a single enzyme, and superior tolerance to substances that typically inhibit PCR (Seyrig et al., 2010a). Another salient attribute of LAMP is the vast amount of DNA generated during amplification (Notomi et al., 2000). This permits direct visualization of positive amplification based on turbidity formed as a result of pyrophosphate precipitation, either visually at the end of the reaction or in real time using a temperature-controlled turbidity sensor (Mori et al., 2004). The high amplification yield also simplifies fluorescence detection due to increased signal strength in the presence of DNA binding dyes, which can be further enhanced using a non-inhibiting dye at high concentration (Seyrig et al., 2010b). The latter is crucial with the prospect of reducing the cost and size of NAAT devices since these are mostly determined by the requirements of the optical module.

Today, LAMP is most often performed in conventional reaction tubes. Integration of the technique in microfluidic devices is arguably the next logical step to further enhance its usability and versatility (Mori and Notomi, 2009). Compared to tubes, microfluidic chips employ a much smaller assay volume, which reduces reagent consumption and associated cost, and is also beneficial for multiplexed detection by assay parallelization. Several studies previously reported miniaturization of LAMP in poly(methyl acrylate) microchips (Hataoka et al., 2004) and polyacrylimade-based micro-chambers (Lam et al., 2008). Toward low-cost and compact instruments, Fang et al. (2010) described a

poly(dimethylsiloxane)-glass multiwell chip and integrated turbidity sensor. The latter consisted of high-intensity red light emitting diode (LED), optical fibers and a phototransistor, which is comparable to the sensor developed by Lee et al. (2008). Compared to turbidity, fluorescence sensing may be better suited for high throughput chips since the former typically requires a sufficiently long optical path and co-linear positioning of the light source and detector, which may complicate chip design. In this regard, in the studies by Fang et al. (2010) and Lee et al. (2008) monitoring of only a single reaction well was demonstrated. Furthermore, the time to positive detection may also be shorter for fluorogenic LAMP (Aoi et al., 2006).

Compact and robust devices for LAMP that allow multiplexed NAAT by parallel sensing of a multitude of reaction wells in a microfluidic format are currently lacking. In this work, we focus on the evaluation of two key components of such an instrument: disposable amplification chips and a temperature-controlled photodiode-based optical sensor. The optical module was drastically simplified by taking advantage of the high fluorescence signal generated using SYTO 81 dye at elevated concentration. To eliminate mechanical movement, which is common in multiplexed photodiode-based sensors (Dishinger and Kennedy, 2008), the optical module consists of an array individually LEDs (one per reaction well) and a single photodiode. To read the different wells, the LEDs are lit in a time-staggered fashion (i.e., lit one at a time) and the signal from the photodiode then assigned to the wells based on the illumination pattern (Ren et al., 2009). To prevent optical cross-talk, which may be problematic in polymer chips (Irawan et al., 2005), the chips were fabricated as thin film shell microstructures, *via* a modified hot

embossing technique using a moldable counter tool. Using a novel primer set for the diarrheal pathogen *Shigella*, the system was demonstrated to perform comparably to an expensive and bulky real time PCR instrument.

EXPERIMENTAL SECTION

Chip design and fabrication

The chips consisted of an array of seven reaction wells with a volume of approximately 2.5 μL each. The wells had a V-shaped layout, consisting of inlet channel for sample dispensing, an outlet channel for air venting, and sensing well. During incubation, the chips were placed vertically in order for air bubbles, when formed, to move away from the sensing region under the influence of gravity.

The chips were fabricated as shell microstructures out of 180 μm ZeonorFilm® (ZF14-180; Zeon Chemicals, Louisville, KY) by hot embossing using a thermoplastic counter tool (acrylonitrile butadiene styrene, ABS from K-mac Plastics). The embossing mold was fabricated *via* stereolithography (SLA) out of high resolution Somos® NanoTool™ (FineLine Prototyping; Raleigh, NC). The mold was thermally cured and coated with a thin layer of nickel (SLArmor^{TM}) to facilitate de-embossing. A landing shoulder was placed around the length of the mold to prevent excessive lateral flow of the ABS (Yao and Kuduva-Raman-Thanumoorthy, 2008), and the mold contained features for eight separate chips to expedite the fabrication process (Figure 4.1b).

To emboss the chips, a sheet of ZeonorFilm® was sandwiched between the ABS tool and embossing mold, and preheated to 150°C in a heated press (Carver press model 4386; Carver, Wabash, IN). A pressure of 1000 kg was then applied for 5 min (Figure 4.1a), after which the system was cooled to 105 °C while maintaining the embossing pressure. After further cooling at room temperature for 1 min, the embossing mold was removed, leaving the chip in the ABS tool, as illustrated schematically in Figure 4.1a. The chips were then removed from the ABS tool and cleaned with 1% Liqui-Nox (Alconox), rinsed with copies amounts of distilled water, soaked in isopropanol, and air-dried in a 65 °C oven. To obtain enclosed microchannels, the chips were laminated with high performance optical film (MicroAmp® Optical Adhesive Film; Applied Biosystems; Foster City, CA). Prior to lamination, 1 mm sample dispensing ports and cross-shaped air vents were patterned into the film using a commercial-grade knife plotter (Bartholomeusz et al., 2005; CraftROBO model CC33OL-20; Graphtec). The patterned film was visually aligned onto the chips, and briefly pressed at 1000 kg (at room temperature) using a piece of silicon rubber (70 A, McMaster-Carr) to obtain a uniform and hermetic seal. To prevent damaging its protruding features, the chip was placed back in the sacrificial ABS tool for bonding.

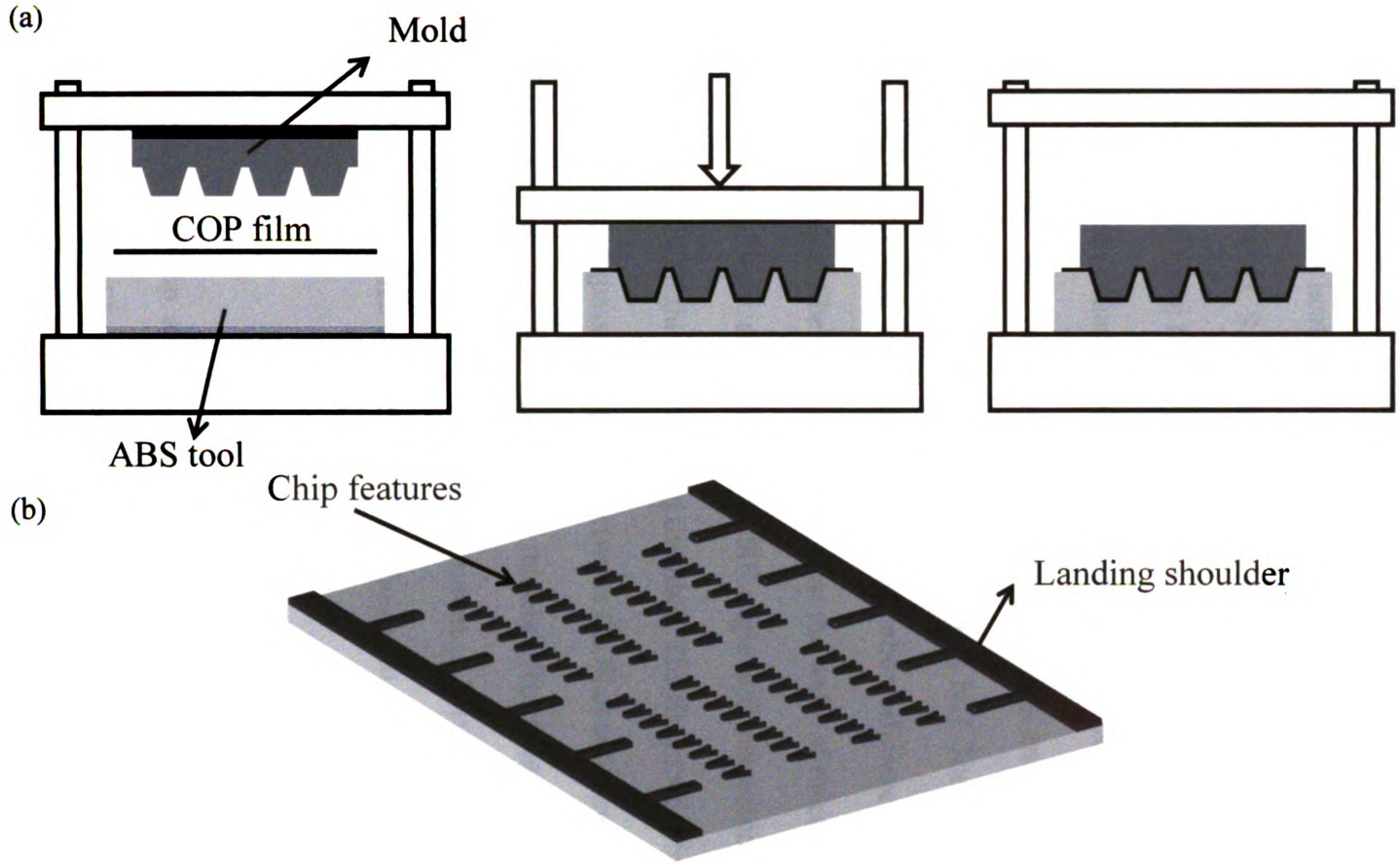


FIGURE 4.1 Method for hot embossing of the thin film polymer chips. (a) Schematic of the hot embossing process using a moldable counter tool. (b) Rendering of the embossing mold with landing shoulder around the length of the mold.

Temperature-controlled optical module

The temperature-controlled optical module is shown schematically in Figure 4.2, and consists of the three main following components: (1) a non-transparent chip holder, (2) an optical sensor, and (3) a heater. The optics is comprised of seven individually addressable green LEDs (RL3-G4518; Super Bright LEDs, St. Louis, MO), seven 1 mm core polymer optical fibers (IF C U1000; Industrial Fiber Optics, Tempe, AZ – razor cut and diamond polished), a blue/green enhanced photodiode with built-in 500 MΩ photovoltaic amplifier (ODA-6WB-500M; OptoDiode Corp., Newbury Park, CA), and a colored glass 590 nm longpass filter (NT54-658; Edmund Optics, Barrington, NJ). The LEDs were placed directly below the reaction wells to improve illumination efficiency in the absence of lenses. Also, the optical fibers were butted against the walls of reaction wells to collect sufficient amounts of light, and oriented at a 90° angle with respect to the LEDs to prevent excessive LED light from entering the fiber. The fibers were butt-coupled to the photodiode and the longpass filter placed between the fiber ends and the photodiode to block the majority of LED light coupled into the fiber. The heater consisted of a Kapton heater (Omega) attached to a thin sheet of aluminum. A T-type thermocouple (Omega) was placed in an additional channel in the chip holder to monitor the temperature of the chip.

To assemble the setup, an ABS chip holder was fabricated using an additional SLA mold containing optical fiber alignment channels, following the fabrication technique outline above. Subsequently, 1 mm through holes were drilled through the chip holder below the center of the reaction wells to act as pinholes and prevent LED light from

reaching adjacent reaction wells. Additionally, 3 mm blind holes were drilled through the bottom of the chip holder in which the LEDs were inserted. The optical fibers were then inserted in the chip holder and fix in place using a small amount of glue. Because the LEDs and optical fibers were integrated as part of the chip holder, the system was highly robust and less prone to misalignment. To couple the fibers to the photodiode, a fixture was fabricated out of black Huntsman RenShape 7820 resin via SLA (Figure 4.2). After placing the chip into the holder, the aluminum top cover with heater was fastened on top using clamps and the entire system covered from ambient light.

Electronic controls

LabView was used for system control and data acquisition (National Instruments). To this end, all hardware was connected to a personal computer through a Multifunction USB Data Acquisition card (NI USB-6009). To control the temperature of the chip, a feedback mechanism involving (PID) and pulse width modulation (PWM) was implemented. Signals from the thermocouple were amplified and conditioned using a LabView 9211 thermocouple input module. The analog output generated from the PID controller was used to adjust the duty cycle of the PWM driver (DRV102T, Texas Instruments). For all experiments, the temperature of the chip was set to 63°C.

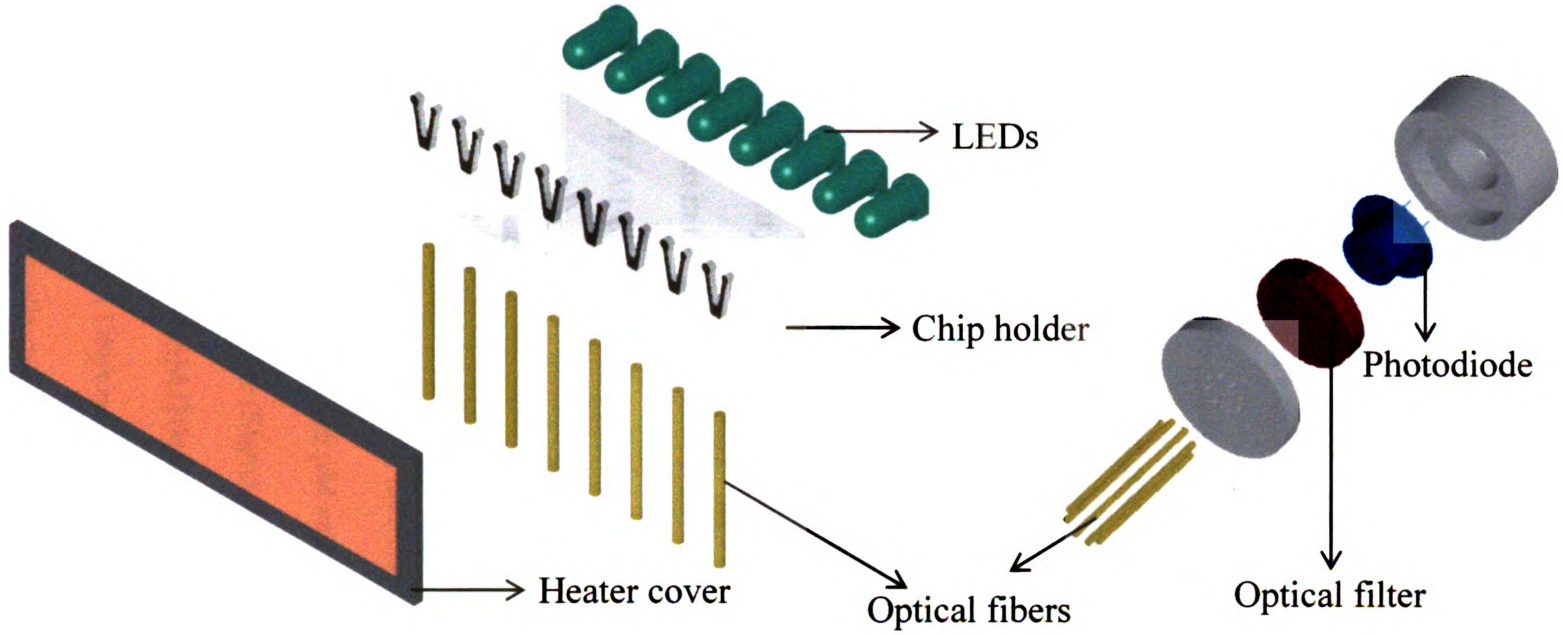


FIGURE 4.2 Schematic of the different components in the experimental setup. Individual reaction wells are read by time-staggered triggering of the LEDs and recording the fluorescence signal using the photodiode, with the optical fibers sending light from the wells to the photodiode. Images in this thesis/dissertation are presented in color.

To control the LEDs, a high-speed CMOS logic 4-to-16 line decoder/demultiplexer (CD74HCT4514; Texas Instruments) was used. High logic output from the decoder signaled LED drivers (STLA01; STMicroelectronics) to provide a constant current of 130 mA to the LEDs. For each measurement, the average of 100 readings collected during a 200 ms time period was used. An analog switch was used to power down the photodiode between measurements as this was observed to improve signal stability. For all experiments, recording of the signals was started when the chip reached 60°C, which took approximately 5 min from room temperature.

Loop-mediated isothermal amplification

The system was evaluated using a novel primer set for the *ipaH* gene of the diarrheal pathogen *Shigella*. The primers were designed using Primer Explorer V4, and Blast search used to verify the specificity of the primers. The sequences of the primers were as follows (written in the 5'-3' direction): F3: TGGCTGGAAAAACTCAGTGC; B3: TCTGACTTTATCCCGGGCAA; FIP; CATGTGAGCGCGACACGGTCGCAGTCTTTCGCTGTTGCT; BIP: AGGCATCAGAAGGCCTTTTCGACCAGAATTTCGAGGCGGAAC; LF: TCACAGCTCTCAGTGGCATCAG; and LB: CTCTCCCTGGGCAGGGAAAT.

Reaction mixtures contained 1.6 μM each of FIP and BIP primer, 200 nM each of F3 and B3 primer, 800 mM betaine (Sigma Aldrich), 1.4 mM of each dNTP (Invitrogen), 20 mM Tris-HCl (pH 8.8), 10 mM $(NH_4)_2SO_4$, 10 mM KCl, 8 mM $MgSO_4$, 8 mM Triton X-100, 0.64 units/μL of Bst DNA polymerase (New England Biolabs) and 20 μM of SYTO 81 (Invitrogen), unless otherwise stated. Genomic DNA from *Shigella flexneri* 2a str.

2457T was obtained from the American Type Culture Collection (Manassas, VA) and used for all experiments. Selected experiments were also run in parallel in a commercial real time PCR instrument from Bio-Rad Laboratories for comparison.

To calculate the time-to-positivity (TTP), raw signal intensities were baseline-corrected by subtracting the average signal obtained in the first five minutes. The TTP was then defined as the time at which the baseline-corrected signal reached an arbitrary cut-off of 100 mV.

RESULTS AND DISCUSSION

The research described herein is part of a project aimed at developing a low-cost, easy-to-use and compact device for LAMP in disposable microfluidic chips. Such devices have widely recognized potential to improve human health globally, by enabling rapid and accurate diagnosis of infectious diseases to be performed in low-resource settings and by minimally trained personnel (Yager et al., 2006). For the purpose of this study, a simple multiwell chip was used to circumvent challenges associated with distribution of a sample across multiple reaction wells, with the goal of evaluating the thin film chips in conjunction with the LED-photodiode based multiplexed optical sensor.

Microfluidic chips

The amplification chips were fabricated as thin film shell microstructures to prevent optical crosstalk between neighboring reaction wells (discussed below). To fabricate the chips, a moldable counter tool was used rather than a matching hard counter mold to

eliminate alignment issues and also reduces mold cost. Also, by placing a landing shoulder around the chip features in the mold, pressure build up was improved by preventing excessive lateral flow of the softened counter tool (Yao and Kuduva-Raman-Thanumoorthy, 2008). This design was crucial as the glass transition temperature of ABS is substantially lower than that of COP, as a result of which initial optimization of the embossing conditions using molds lacking a landing shoulder proved difficult. Using the modified mold, fabrication was highly robust and not drastically influenced by embossing pressure and/or temperature dwelling time. The here presented technique is comparable to rubber-assisted microforming, a recently developed method in which thermostable rubber serves as the counter tool (Nagarajan and Yao, 2009). A benefit of the latter technique may be that polymers other than COP can also be readily used while for the method used in this study more optimization may be needed to find a suitable moldable counter tool material to ensure physical and chemical compatibility. However, using a moldable counter tool provides the benefit that sharp features can be fabricated that may not be attainable using rubber-assisted microforming.

This study also demonstrates the utility of molds fabricated *via* SLA for hot embossing and rapid prototyping of polymer chips. Compared to other mold fabrication techniques, SLA is much cheaper and also provides the ability to fabricate complex features with few design constraints. For certain applications, the relatively high surface roughness of SLA molds may be problematic, e.g., for assays requiring carefully controlled fluid flow or when assays are performed at elevated temperature due to the propensity of air bubble formation. However, neither was critical in this study and we

have extensively utilized SLA molds for rapid and inexpensive prototyping of more complex chips.

Optical module

The main premise of this study was that simple optics would suffice for real time monitoring of LAMP due to its high amplification yield. Furthermore, our group previously reported that SYTO 81 is only marginally inhibiting at concentrations up to 20 μM (Seyrig et al., 2010b), which drastically enhanced signal strength. As expected, this effect was also observed in the here described chips (Figure 4.3). While a dye concentration of 2 μM was sufficient for detection, a concentration of 20 μM was used for all subsequent experiments, recognizing the small shift in TTP.

Importantly, the optical module described here is much cheaper and simpler than that typically used for fluorescence sensing in microfluidic chips (Kuswandi et al., 2007). More specifically, a simple colored glass emission filter with an emission spectrum shifted to the right of the maximal emission peak of SYTO 81, along with the orthogonal positioning of the LEDs with respect to the fibers, was sufficient to drop the optical background signal well below the baseline signal prior to amplification. Also, no lenses were necessary since the optical fibers were butt-coupled to the wells and photodiode, which reduces cost and also provides flexibility in terms of positioning of the chip with respect to the photodiode.

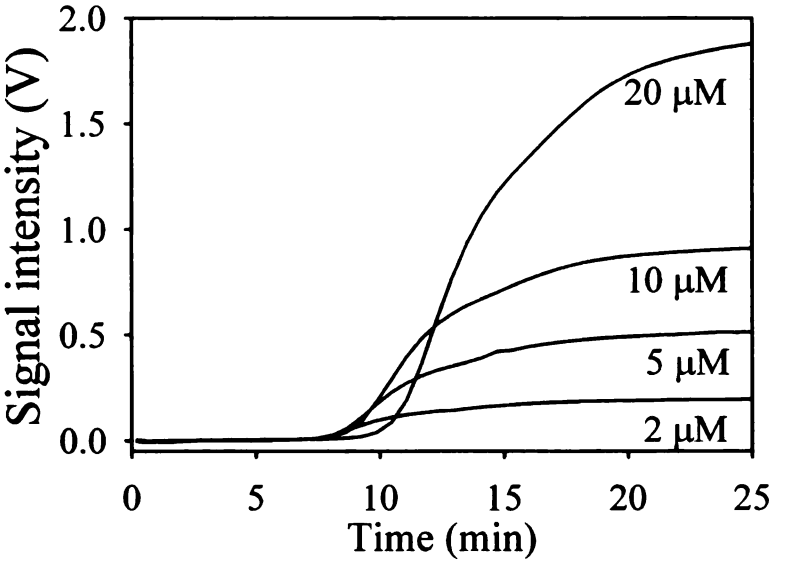


FIGURE 4.3 Enhancement of the signal intensity at elevated SYTO 81 concentrations.

Compared to previous work in which a similar optical module was used for fluorogenic LAMP with calcein dye (Tomita et al., 2008), the use of SYTO 81 presents a significant improvement as the former typically results in delayed amplification compared to SYTO 81 (Seyrig et al., 2010b).

Optical cross-talk and reproducibility

Optical cross-talk between closely spaced reaction wells can occur in polymer chips due to LED light or emitted fluorescence light propagating through the chip substrate as a result of its waveguiding properties (Irawan et al., 2005). To evaluate whether the shell-structured chip design effectively prevented this, the following experiment was performed. The central well of the chip was filled with water and the four surrounding wells with amplification reagents and DNA. As shown in Figure 4.4, no increase in signal

was measurable in the central well, indicating that optical crosstalk did not affect the accuracy of the measurements.

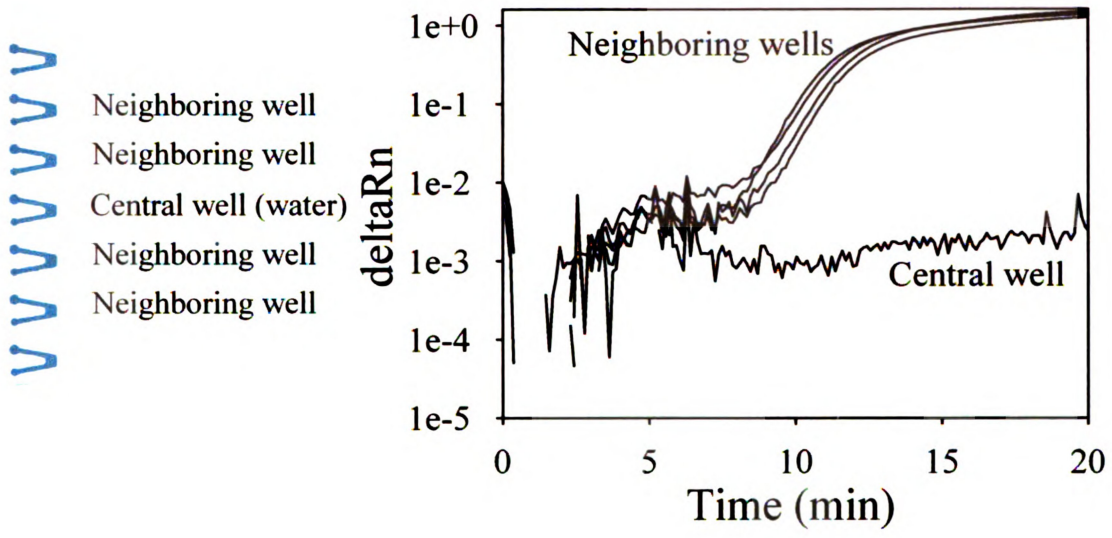


FIGURE 4.4 Lack of optical crosstalk between neighboring reaction wells. The left panel illustrates the experimental design and the right panel shows the amplification plots. Data are from a single representative experiment. Images in this thesis/dissertation are presented in color.

Evaluation of the intra- and inter-chip variability revealed that amplification and detection was comparable across wells and chips. While the raw signal intensities varied several-fold between different wells within a single chip, probably due to alignment differences in this manually assembled system, normalization of the amplification plots showed that all amplification curves displayed comparable shapes and positions (Figure 4.5). In terms of TTP, the coefficient of variation (CV) for the seven wells in a single chip was roughly 4%, with the average TTP not being significantly different for the three chips. The CV was however, as expected, higher than the CV observed using the commercial real time PCR instrument, which was less than 1%. A smaller CV was also

reported using a commercial real time turbidity meter (Mori et al., 2004), but comparison with the data here reported is difficult since the latter measures absorbance and was demonstrated using a different primer and high target gene concentration.

The data presented in Figure 4.5 also shows that TTP obtained using the chips and simple sensor may be slightly smaller than that in the commercial real time PCR machine. This observation is interesting as the latter contains a more expensive optical module with better sensitivity, but by optimization of the fluorescent dye comparable results can be generated using much simplified optics.

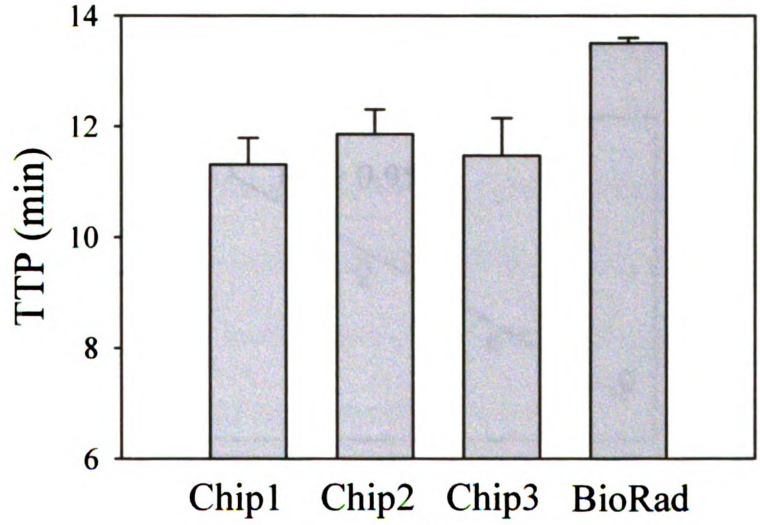


FIGURE 4.5 Intra-chip and inter-chip variability. Data represent the mean and standard deviation of the TTP for each chip and that of a commercial real time PCR machine.

Quantification

To further evaluate the system, a traditional standard curve was generated for genome copy numbers ranging from ~10 to 10^6 copies per channel. Good linearity was observed between the starting copy number and TTP with a correlation coefficient of 0.95 (Figure

4.6), indicating that the developed system is well suited for target gene quantification. The limit of detection was ~100 copies per reaction wells, which is similar to the sensitivity achieved in conventional tubes using a commercial real time PCR instrument.

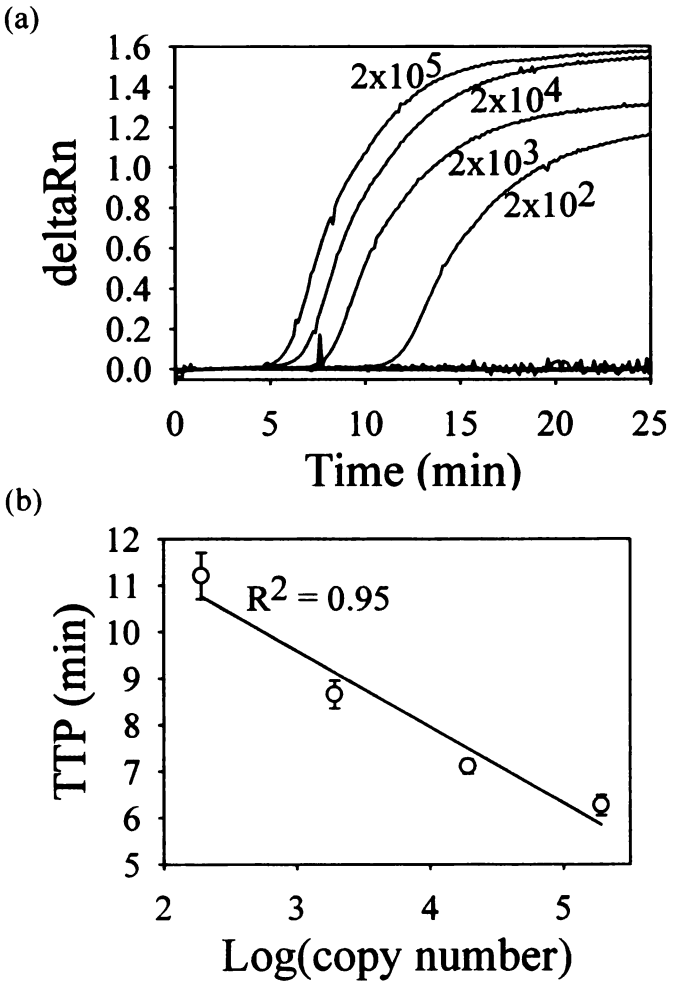


FIGURE 4.6 Target gene quantification. (a) Amplification curves for different target copy numbers and (b) standard curve. The amplification plots are from a single representative experiment and the symbols in the standard curve represent the average and standard error of three replicates.

These results also demonstrate that native COP is compatible with LAMP, although a minor level of inhibition may occur at low target copy numbers, as is apparent from the slight non-linearity of the calibration plot at the lowest copy number yielding positive

amplification. However, this should not impact quantification potential since other curve fits (e.g., a quadratic) can readily be used to account for this effect (data not shown). This observation is important since surface treatment, as was described in other studies using COP and the related polymer cyclo olefin copolymer chips for isothermal nucleic acid amplification (Furuberg et al., 2008; Lutz et al., 2010), adds an additional level of complexicity and cost to the chip fabrication process. Furthermore, initial evaluation of two different treatment techniques (UV-ozone and photografting) did not improve amplification efficiency at low target copy numbers (data not shown).

CONCLUSIONS

In summary, two key components of a low-cost and compact device for microfluidic NAAT using LAMP were developed: disposable amplification chips and a multichannel fluorescence sensor. The optics was drastically simplified by taking advantage of the high amplification yield of LAMP and a non-inhibiting DNA-binding (SYTO 81) at elevated concentration. Optical crosstalk was eliminated due to the use of shell-structured chips, which were fabricated using inexpensive molds *via* hot embossing using a moldable counter tool. Combined with film lamination and xurography, this technique provides an attractive means for rapid and inexpensive prototyping of microfluidic chips.

The developed system is well suited for integration in low-cost and compact devices. The total component cost for the seven-channel sensor was less than $100, with the majority of the cost stemming from the pre-amplified photodiode. Furthermore, due to the low-cost of the LEDs and optical fibers, moderate further multiplexing of the sensor

would not increase the cost by more than a few dollars. The sensor can also be build very compact due to the use of polymer optical fibers rather than lenses, which provides flexibility in positioning of the photodiode with respect to the chip. In its current setup, the size is mainly determined by the size of the LEDs but using surface Mount LEDs could reduce this.

Toward further development, the main step is to modify the chips so that a single pipetting step is sufficient for loading of a multitude of reaction channels. In addition, the chip and sensor throughput will be increased to 64, to enable analysis of multiple samples for several genetic markers in a single instrument run.

REFERENCES

Aoi, Y., M. Hosogai and S. Tsuneda, 2006. Real-time quantitative LAMP (loop-mediated isothermal amplification of DNA) as a simple method for monitoring ammonia-oxidizing bacteria. Journal of Biotechnology, 125: 484-491.

Auroux, P.A., Y. Koc, A. deMello, A. Manz and P.J.R. Day, 2004. Miniaturised nucleic acid analysis. Lab on a Chip, 4: 534-546.

Bartholomeusz, D.A., R.W. Boutte and J.D. Andrade, 2005. Xurography: Rapid prototyping of microstructures using a cutting plotter. Journal of Microelectromechanical Systems, 14: 1364-1374.

Chen, L., A. Manz and P.J.R. Day, 2007. Total nucleic acid analysis integrated on microfluidic devices. Lab on a Chip, 7: 1413-1423.

Dishinger, J.F. and R.T. Kennedy, 2008. Multiplexed detection and applications for separations on parallel microchips. Electrophoresis, 29: 3296-3305.

El Ekiaby, M., N. Lelie and J.P. Allain, 2010. Nucleic acid testing (NAT) in high prevalence-low resource settings. Biologicals, 38: 59-64.

Fang, X.E., Y.Y. Liu, J.L. Kong and X.Y. Jiang, 2010. Loop-Mediated Isothermal Amplification Integrated on Microfluidic Chips for Point-of-Care Quantitative Detection of Pathogens. Analytical Chemistry, 82: 3002-3006.

Furuberg, L., M. Mielnik, A. Gulliksen, L. Solli, I.R. Johansen, J. Voitel, T. Baier, L. Riegger and F. Karlsen, 2008. RNA amplification chip with parallel microchannels and droplet positioning using capillary valves. Microsystem Technologies-Micro-and Nanosystems-Information Storage and Processing Systems, 14: 673-681.

Hataoka, Y., L.H. Zhang, Y. Mori, N. Tomita, T. Notomi and Y. Baba, 2004. Analysis of specific gene by integration of isothermal amplification and electrophoresis on poly(methyl methacrylate) microchips. Analytical Chemistry, 76: 3689-3693.

Irawan, R., S. Tjin, P. Yager and D. Zhang, 2005. Cross-talk problem on a fluorescence multi-channel microfluidic chip system. Biomedical Microdevices, 7: 205-211.

Kuswandi, B., Nuriman, J. Huskens and W. Verboom, 2007. Optical sensing systems for microfluidic devices: A review. Analytica Chimica Acta, 601: 141-155.

Lam, L., S. Sakakihara, K. Ishizuka, S. Takeuchi, H.F. Arata, H. Fujita and H. Noji, 2008. Loop-mediated isothermal amplification of a single DNA molecule in polyacrylamide gel-based microchamber. Biomedical Microdevices, 10: 539-546.

Lee, S.Y., J.G. Huang, T.L. Chuang, J.C. Sheu, Y.K. Chuang, M. Holl, D.R. Meldrum, C.N. Lee and C.W. Lin, 2008. Compact optical diagnostic device for isothermal nucleic acids amplification. Sensors and Actuators B-Chemical, 133: 493-501.

Lutz, S., P. Weber, M. Focke, B. Faltin, J. Hoffmann, C. Muller, D. Mark, G. Roth, P. Munday, N. Armes, O. Piepenburg, R. Zengerle and F. von Stetten, 2010. Microfluidic lab-on-a-foil for nucleic acid analysis based on isothermal recombinase polymerase amplification (RPA). Lab on a Chip, 10: 887-893.

Mori, Y., M. Kitao, N. Tomita and T. Notomi, 2004. Real-time turbidimetry of LAMP reaction for quantifying template DNA. Journal of Biochemical and Biophysical Methods, 59: 145-157.

Mori, Y. and T. Notomi, 2009. Loop-mediated isothermal amplification (LAMP): a rapid, accurate, and cost-effective diagnostic method for infectious diseases. Journal of Infection and Chemotherapy, 15: 62-69.

Nagarajan, P. and D.G. Yao, 2009. Rubber-assisted micro forming of polymer thin films. Microsystem Technologies-Micro-and Nanosystems-Information Storage and Processing Systems, 15: 251-257.

Notomi, T., H. Okayama, H. Masubuchi, T. Yonekawa, K. Watanabe, N. Amino and T. Hase, 2000. Loop-mediated isothermal amplification of DNA. Nucleic Acids Research, 28.

Piepenburg, O., C.H. Williams, D.L. Stemple and N.A. Armes, 2006. DNA detection using recombination proteins. PLOS Biology, 4: 1115-1121.

Ren, K.N., Q.L. Liang, X. Mu, G.A. Luo and Y.M. Wang, 2009. Miniaturized high throughput detection system for capillary array electrophoresis on chip with integrated light emitting diode array as addressed ring-shaped light source. Lab on a Chip, 9: 733-736.

Seyrig, G., F. Ahmad, R.D. Stedtfeld, D.M. Tourlousse and S.A. Hashsham, 2010a. Simple, powerful, and smart: Using LAMP for low cost screening of multiple waterborne pathogens. In Environmental Microbiology: Current Technology and Water Applications. Caister Academic Press.

Seyrig, G., K. Towerey, R.D. Stedtfeld, T.M. Stedtfeld, D. M. Tourlousse, A.M. Cupples, J.M. Tiedje and S. A. Hashsham, 2010b. Evaluation of alternative fluorescent DNA dyes for faster real-time loop-mediated isothermal amplification. *In preparation.*

Tomita, N., Y. Mori, H. Kanda and T. Notomi, 2008. Loop-mediated isothermal amplification (LAMP) of gene sequences and simple visual detection of products. Nature Protocols, 3: 877-882.

Varmus, H., R. Klausner, E. Zerhouni, T. Acharya, A.S. Daar and P.A. Singer, 2003. Grand challenges in global health. Science, 302: 398-399.

Vincent, M., Y. Xu and H.M. Kong, 2004. Helicase-dependent isothermal DNA amplification. EMBO Reports, 5: 795-800.

Weigl, B.H., D.S. Boyle, T. de los Santos, R.B. Peck and M.S. Steele, 2009. Simplicity of use: a critical feature for widespread adoption of diagnostic technologies in low-resource settings. Expert Review of Medical Devices, 6: 461-464.

Yager, P., T. Edwards, E. Fu, K. Helton, K. Nelson, M.R. Tam and B.H. Weigl, 2006. Microfluidic diagnostic technologies for global public health. Nature, 442: 412-418.

Yao, D.G. and R. Kuduva-Raman-Thanumoorthy, 2008. An enlarged process window for hot embossing. Journal of Micromechanics and Microengineering, 18.

Zhang, C.S., J.L. Xu, W.L. Ma and W.L. Zheng, 2006. PCR microfluidic devices for DNA amplification. Biotechnology Advances, 24: 243-284.

CHAPTER FIVE

A SIMPLE 64-CHANNEL MICROFLUIDIC CHIP FOR LOOP-MEDIATED ISOTHERMAL AMPLIFICATION

INTRODUCTION

Devices for nucleic acid amplification tests (NAAT) that are capable of screening for multiple pathogens or genetic markers simultaneously are urgently needed to address the global burden of infectious diseases (Yager et al., 2008; Weigl et al., 2009). Multiplexed detection is necessary to accurately identify the etiologic agent(s) responsible for diseases that can be caused by multiple pathogens, such as infectious diarrhea (Antikainen et al., 2009), respiratory infections (Kerdsin et al., 2010), and for antibiotic resistance typing emanating from a number of different protein-coding genes (Martineau et al., 2000). Multiplex PCR and microarrays are two main platforms for multiplexed detection; while both are powerful in conventional laboratory settings they are not easily integrated in low-cost and compact devices. This is mainly due to the need for sensitive optics and complex reagent handling equipment.

A more promising method for multiplexed detection involves parallelization of the assays in microfluidic chips. A number of assays using polymerase chain reaction (PCR) have been developed by miniaturization of common multiwell plates (Matsubara et al., 2005; Morrison et al., 2006) or using a microfluidic network for sample distribution (Liu et al., 2003). More recently, isothermal techniques nucleic acid amplification have also been implemented in microfluidic chips (Hataoka et al., 2004; Furuberg et al., 2008; Lam

et al., 2008; Ramalingam et al., 2009b; Fang et al., 2010; Lutz et al., 2010). In many cases, however, sample dispensing and chip sealing represent the two main challenges that were not adequately addressed. Most chips required either many manual steps or bulky peripheral equipment for loading of the sample and sealing of the chips. Therefore, a need still exists for more user-friendly and robust microfluidic chips for multiplexed NAAT.

In this work, a polymer microfluidic chip for multiplexed NAAT using loop-mediated isothermal amplification (LAMP) is presented. The chip consists of four distinct arrays of 15+1 reaction channels containing dried primers, allowing parallel analysis of four samples for a multitude of genetic targets. The chip is relatively simple to use as it requires only a single pipetting step to dispense the sample in a distribution channel that routes the sample to the different reaction wells. After dispensing, the sample inlet port of the chip is sealed with tape. Basic development for such array-type chips and sample propagation involving capillary action (Furuberg et al., 2008; Ramalingam et al., 2009a) or centrifugation (Lutz et al., 2010) are not novel. In our design, pressure generated by depression of the plunger of the pipette is used to induce fluid flow. Another distinguishing feature of our chips is the use of robust pinch-type valves for sealing of the reaction wells. To fabricate these valves, a novel technique involving film lamination and xurography (Bartholomeusz et al., 2005) was developed. Finally, to demonstrate the usability of the chip, three major pathogens responsible for infectious diarrhea (*Campylobacter jejuni*, *Salmonella* and *Vibrio cholerae*) were detected using endpoint fluorogenic LAMP.

EXPERIMENTAL SECTION

Chip design and operation

The chip is roughly the size of credit card and contains four arrays of 15+1 reaction wells (Figure 5.1). For each array, the 15 reaction wells are connected to a sample distribution channel (250 μm × 250 μm) that delivers the sample to different reaction wells from a single inlet port. The other four channels are connected to an additional sample distribution channel and inlet port, with the purpose of accommodating negative assay controls. The volume of sample loaded in the chip is ~30 μL per array, of which roughly half remains in the distribution channel.

The chip is loaded using a conventional pipettor, requiring only a single pipetting step per sample. Since capillary action is not sufficient to fill the chips due to the hydrophobic nature of untreated COP and the presence of various features that may act as capillary valves, pressure is provided by depressing the plunger of the pipettor to push the sample through the sample distribution channel. Hence, the inlet port is designed to fit snugly around the pointed end of a 200 μL pipette tip (Figure 5.2b).

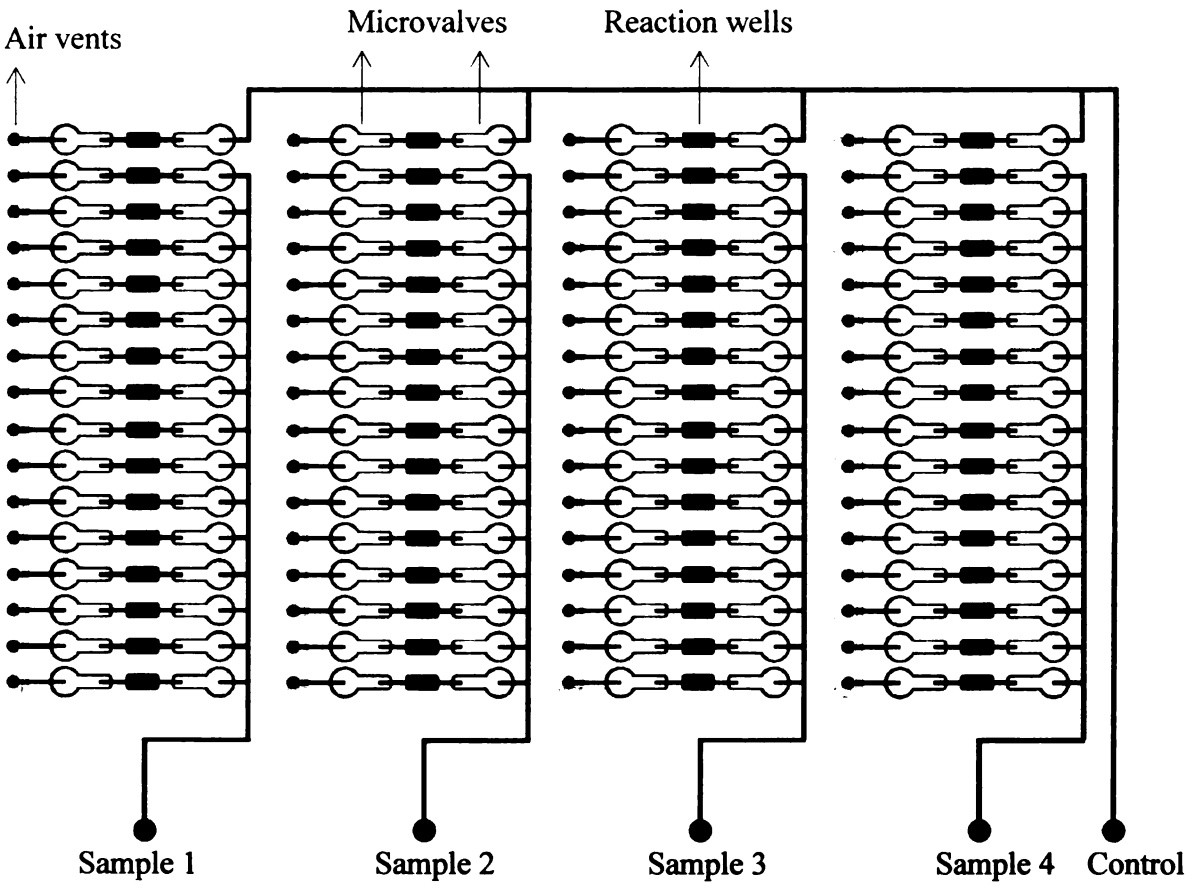


FIGURE 5.1 Schematic of the 64-channel microfluidic chip. The chip enables parallel analysis of four samples for up to 15 different gene targets per sample. Images in this thesis/dissertation are presented in color.

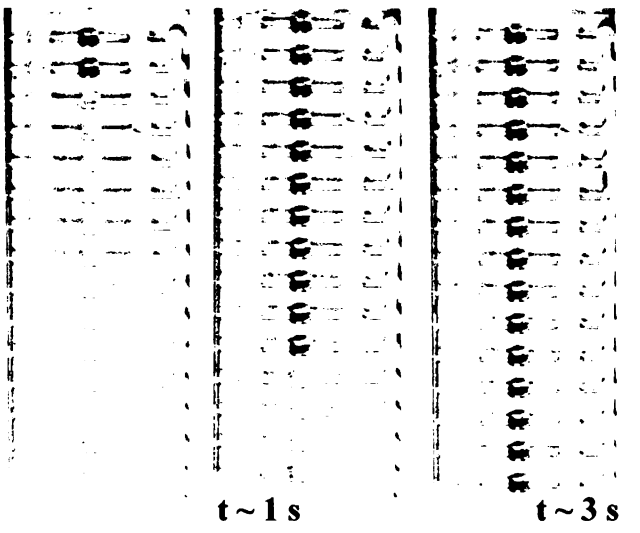
As the sample moves through the distribution channel, sequential filling of the reaction wells occur within a matter of seconds (Figure 5.2a). In the process, air inside the chip is purged out through air vents placed upstream of each reaction well. Gas-permeable hydrophobic membrane (GVHP 09050, Millipore) covering the air vents prevents liquid from exiting the chip due to the pressure generated with the pipettor.

After dispensing of the sample, the inlet ports are easily sealed with tape. The chip is then placed in a dedicated cartridge in which microvalves placed up- and downstream of each reaction well are closed under the action of an array of plungers, as described below. Because closing of the valves leads to fluid movement due to the non-zero dead volume of the valves, it was necessary to allow for backflow of a small portion of the sample without wetting the surface needed for sealing. Therefore, the inlets ports were placed in a cavity lower than the level of the surface for lamination (Figure 5.2b).

Chip fabrication

For rapid and inexpensive prototyping of the chips, lamination (Paul et al., 2007) and xurography (Bartholomeusz et al., 2005) were used. The latter is a simple prototyping technique that employs computer-controlled cutting plotter to generate structures in thin polymer films. High performance optical film with pressure sensitive adhesive (MicroAmp® Optical Adhesive Film; Applied Biosystems) was used for all layers to prevent inhibition of LAMP. A low-cost commercial grade cutting plotter (CraftROBO Pro model CC330L-20; Graphtec) equipped with a 60°, 0.9 mm blade and set to a cutting quality of 3 was used for xurography.

(a)



(b)

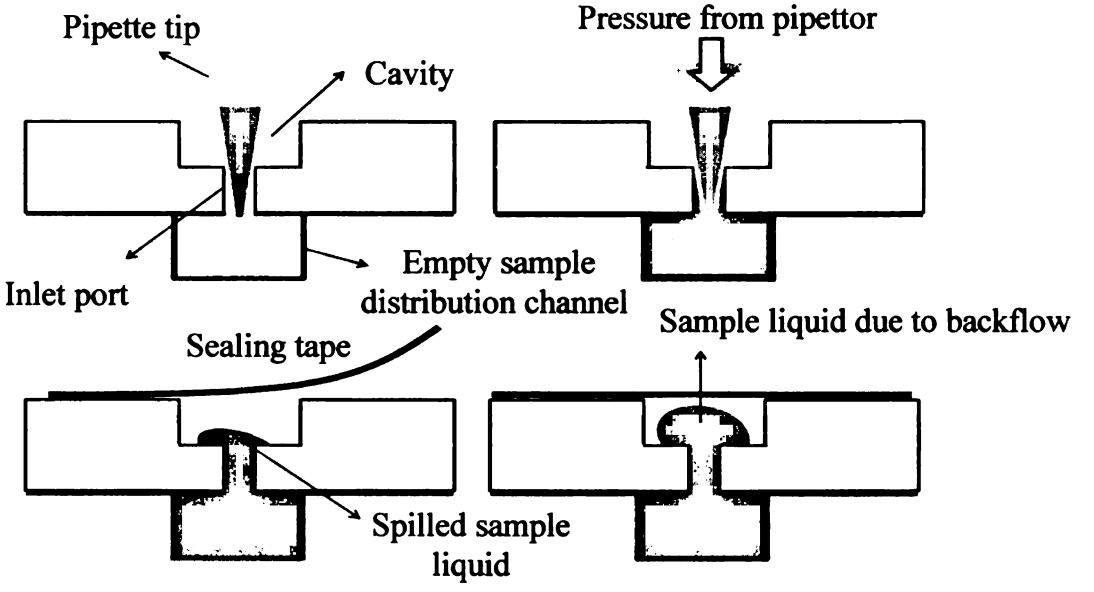


FIGURE 5.2 Microfluidic operation of the chip. (a) Time-lapsed images showing filling of the reaction wells as the sample passes through the distribution channel. (b) Schematic of the inlet port design. Images in this thesis/dissertation are presented in color.

The chip contains three layers, as illustrated in Figure 5.3: Layer 1 – a cyclo olefin polymer sheet (COP; ZeonorFilm®, Zeon Chemicals) with embossed channel and well features, Layer 2 – a first layer of adhesive optical film with 256 through-holes for routing of the sample through the valve seats, 64 through-holes for air venting and 5

through-holes for sample dispensing, and Layer 3 – a second layer of adhesive optical film with 69 through-holes for air venting, 5 through-holes for sample dispensing, and 128 valve features.

Layer 1 was fabricated as detailed in Chapter 3 *via* a modified hot embossing technique using a mold with landing shoulder fabricated by stereolithography and a moldable counter tool. To fabricate Layer 2, an adhesive film is placed on backing paper with the transparent film face-up, and Pattern 1 (see Appendix for the different patterns and process schemes) cut using a blade with its position in the holder adjusted so that the release liner remains intact. After cutting, the unnecessary material is then easily weeded out using transfer tape, leaving through-holes in the transparent film. To fabricate Layer 3, an adhesive film is placed on backing paper with the release liner face-up, and cutting of the necessary patterns performed in two steps. First, Pattern 2a, which generates the through-holes for air venting, is cut using a blade with its position adjusted so that all layers are pierced. Second, Pattern 2b, which generates both the valve seats and alignment guides for the hydrophobic membranes, is cut using a blade with its position adjusted so that it cuts only through the release liner. The unwanted material of the valve features is then removed using transfer tape, taking care not to detach the release liner from the adhesive layer.

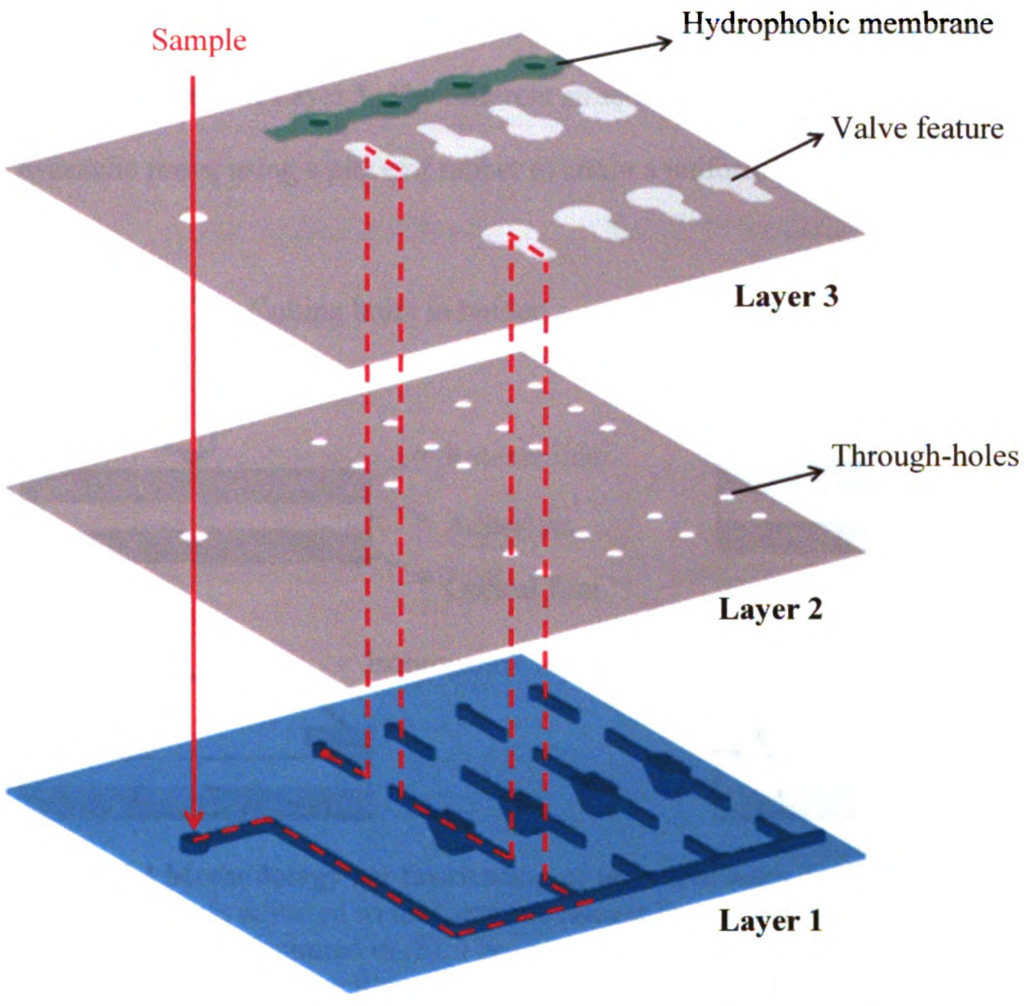


FIGURE 5.3 Rendering of the multilayer structure of the microfluidic chip. The dashed red line indicates the flow path of the sample liquid through the different layers during filling of the chip. For clarity, only a portion the chip and the sample flow for a single well are shown. Images in this thesis/dissertation are presented in color.

In the next step, the valve seats are formed by submerging the entire adhesive film in isopropanol for 10 to 15 min, which dissolves the adhesive not protected by release liner (Figure 5.4). The film is then cleaned using paper to remove any remaining adhesive; this step is crucial since any left-over adhesive may clog the through-holes in Layer 1. Subsequently, the features cut for alignment of the membranes are removed and the membranes put in place.

After releasing the liner from the patterned films, the chip is assembled by visually aligning them on top of Layer 1. The different structures are then firmly bonded together in a hydraulic press, using a piece of rubber to attain a uniform bond.

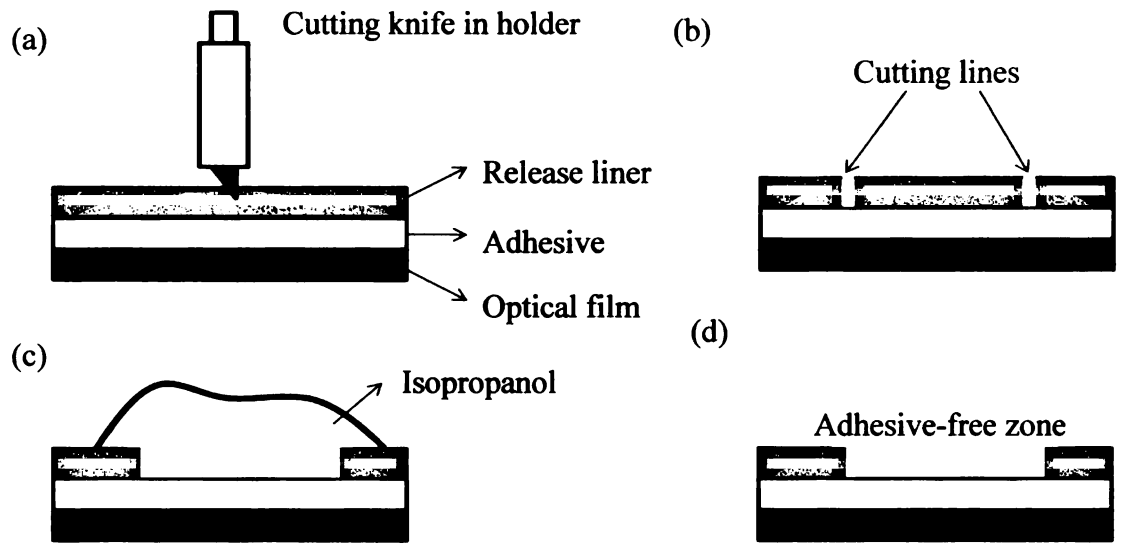


FIGURE 5.4 Methodology for fabrication of the valve seats. (a) The position of the blade in its holder is adjusted so that only the release liner is cut when placed face-up on backing paper, as illustrated in (b). The unnecessary material is then removed using transfer tape and the adhesive film placed in isopropanol (c), which dissolves the exposed portion of the adhesive layer (d).

Microfluidic valves

A schematic of the structure and working principle of the valves is shown in Figure 5.5a. The valves consist of a discontinued microchannel that routes the sample through the valve seats formed in Layer 2. When a plunger depresses Layer 2, the through-hole in Layer 1 is covered and the valve closed. This is demonstrated in Figure 5.5b, which shows a photograph of three reaction wells with closed valves. Since the plunger is fabricated out of clear plastic, the underlying channel with closed through-hole is visible.

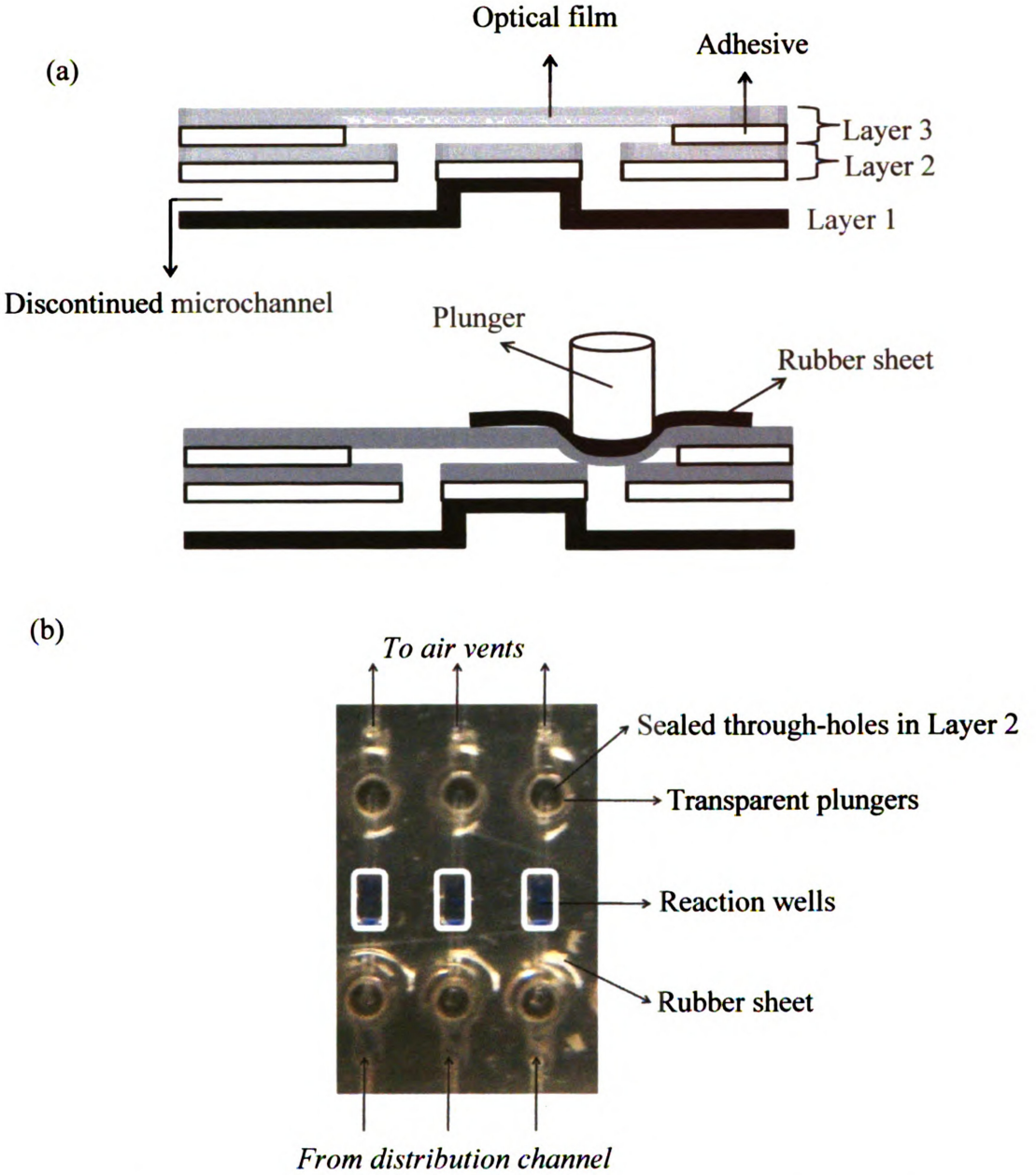


FIGURE 5.5 Illustration of the pinch-type microvalves. (a). Schematic of the structure and working principle of the microvalves. (b). Close-up photograph of three reaction wells with closed microvalves. Images in this thesis/dissertation are presented in color. The position of the reaction wells has been enhanced for clarity.

Experimental setup

The experimental setup for evaluation of the chips is shown schematically in Figure 5.6. It consists of a heated chip holder, a fixture for closing the microvalves, and

imaging system for endpoint detection of fluorogenic LAMP. As chip holder, one of the formed pieces of ABS that served as sacrificial counter tool for hot embossing was used. To heat the chip, a strip heater was placed below the chip holder and temperature controlled using proportional–integral–derivative (PID) control and pulse width modulation (PWM), implemented using LabView (National Instruments) as discussed in more detail in Chapter 4. To close the microvalves, a transparent lid with pre-aligned plungers was pressed down onto the chip using two toggle clamps. The imaging system consisted of a portable 50 mW green laser (Laserglow Technologies; Toronto, Canada), an engineered diffuser (Thorlabs, Newton, NJ), an 555 nm bandpass filter (Chroma Technologies, Bellows Falls, VT) and a monochrome CCD camera (Deep Sky Imager Pro; Meade Instruments Corporation, Irvine, CA).

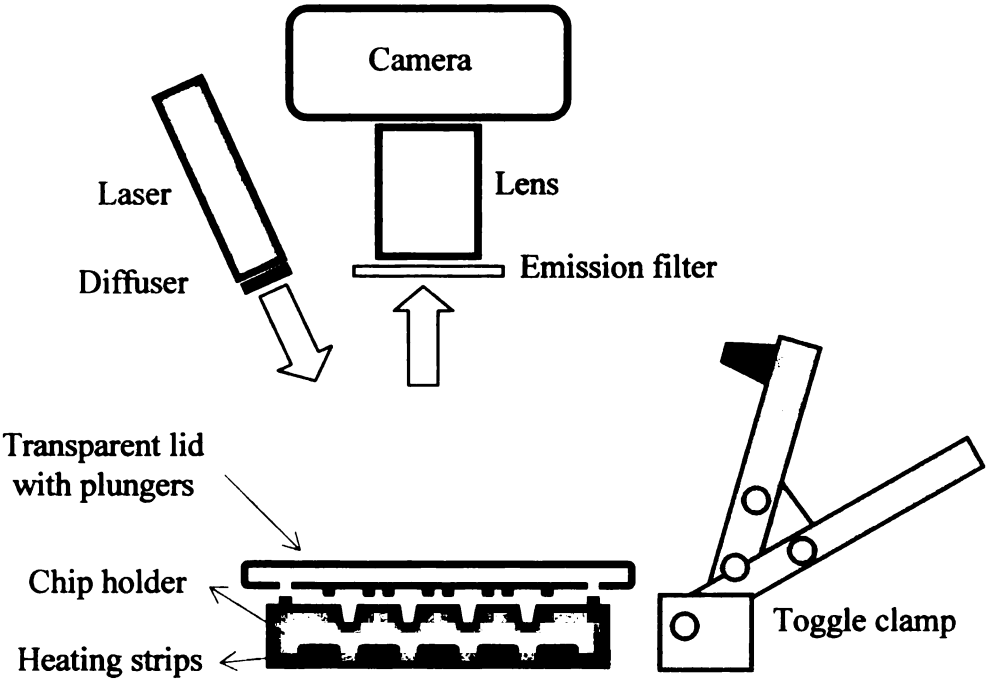


FIGURE 5.6 Schematic of the experimental setup for evaluation of the chip by endpoint fluorogenic LAMP.

Loop-mediated isothermal amplification

The chip was evaluated using six LAMP primers sets for three major diarrheal pathogens, *C. jejuni*, *Salmonella* and *V. cholerae* (Table 5.1). The primers were designed using Primer Explorer V4 and Blast search used to verify the specificity of the primers. The amount of primers dried in each reaction well was that needed to yield a final concentration of 1.6 μM each of FIP and BIP primer, 800 nM each of LB and FB primer, and 200 nM each of F3 and B3 primer after dissolution.

TABLE 5.1 Primers used in this study.

Pathogen Gene	**Primer sequences (5'-3')**
C. jejuni	FP: GCAAGACAATATTATTGATCGC
cj0414[a]	RP: GCAAGACAATATTATTGATCGC
	FIP: ACAGCACCGCCACCTATAGTAGAAGCTTTTTTAAACTAGGGC
	BIP: AGGCAGCAGAACTTACGCATTGAGTTTGAAAAAACATTCTACCTCT
	LF: CTAGCTGCTACTACAGAACCAC
	LB: CATCAAGCTTCACAAGGAAA
mapA	FP: CTAAAAATTCTCAATGCAGTTCT
	RP: ACCGCATTAAAATTCACATC
	FIP: ACAACATTGAATTCCAACATCGCTATGTGAAAGTCCTGGTGGT
	BIP: CACTTTAGACACTGGTATTGCTTTGACAAATAACTTTTTCCCTTTAGC
	LF: TGTATAAAAGCCCTTTAATCTTTGCTTCA
	LB: ATGATTTAAAAGAAGRGCAAAAGGTT
Salmonella	FP: CGGCCCGATTTTCTCTGG
invA	RP: CGGCAATAGCGTCACCTT
	FIP: GCGCAGCATCCGCATCAATAATATGGTATGCCCGGTAAACAG
	BIP: GAACGGCGAAGCGTACTGGACATCGCACCGTCAAAGGAA
	LF: CCTTCAAATCGGCATCRATACTCAT
	LB: AAGGGAAAGCCAGCTTTACG

Table 5.1 continued

hilA	FP: GGATCAGGTTCAATCCGAG
	RP: TGTACAATATTATCATTMCCATCGG
	FIP: TTCGTAATGGTCACCGGCAGAGTCTGCATTACTCTATCGTG
	BIP: AACTGCCGCAGTGTTAAGGATATCATCTGCCCGGAGAT
	LF: GCGCATACTGCGATAATCCCTT
	LB: CTTGAGCTCATGGATCAATTACGCC
V. cholerae	FP: CGAGTGGAAACGGTTGAAGA
toxR	RP: AGGGGAAGTAAGACCGCTAT
	FIP: GCACACTGCTTGAYTCTGCGTACGAAAGCGAAGCTGCTCAT
	BIP: AGCCACTGTAGTGAACACACCGTCGATTCCCCAAGTTTGGAG
	LF: ACAGATTCTGGCTGAGAGATGTC
	LB: CAGCCAGCCAATGTTGTGAC
rtxA	FP: ACCGTACTTGGTCTAACCGT
	RP: TTGCGTGTAGACATCTTCGG
	FIP: CCTGTCTCAACGCGTGAACTGATTGGCGATGTGACCTTTGATG
	BIP: TTAGTCCGTAAAGGCAAAGTGGGCGATGTGTCCGCTCAATGCG
	LF: TGTTCGCGGCACCAGCA
	LB: GATATTACTCTGCAAGGTGCTGG

[a] From Yamazaki et al. (2008)

The reaction mixture loaded in the chip contained 800 mM betaine (Sigma Aldrich), 1.4 mM of each dNTP (Invitrogen), 20 mM Tris-HCl (pH 8.8), 10 mM $(NH_4)_2SO_4$, 10 mM KCl, 8 mM $MgSO_4$, 8 mM Triton X-100, 0.64 units/μL of Bst DNA polymerase (New England Biolabs) and 20 μM of SYTO 81 (Invitrogen). Genomic DNA from *C. jejuni* strain NCTC 11168, *Salmonella enterica* strain LT2, and *V. cholerae* strain O1 biovar eltor str. N16961 was obtained from the American Type Culture Collection (Manassas, VA).

RESULTS AND DISCUSSION

Chip design and fabrication

The microfluidic chip is capable of analyzing four samples for up to 15 different genetic markers per sample. The chip is well suited for adoptation in low-resource settings as only a simple pipettor is needed for sample dispensing and sealing of the chips merely requires covering the sample inlet port with tape. Evidently, it is recognized that sample processing also needs to be addressed to provide a system with sample-in answer-out capabilities.

Propagation of the sample through the distribution channel relies only upon the pressure generated by the pipettor. Compared to capillary-driven flow, which is a common technique for low-cost microfluidics, this provides the benefit that design of the chip is simplified. This is because capillary-driven flow requires careful analysis of the flow behavior in expanding features that may be present in the chip (Zhang et al., 2005). Others have relied upon centrifugation for sample distribution in microfluidic chips for NAAT (Lutz et al., 2010), which requires additional instrumentation. The valves implemented in the chip are "normally-open" valves, which is important from the perspective of the user. It allows the user to directly load the sample, after which closing of the valves is accomplished in a dedicated chip cartridge. Furthermore, due to the shell-structure of the chips, alignment is much simplified, which is important considering the need for correct positioning of the plungers with respect to the chip.

In terms of fabrication, the developed method for locally removing adhesive is, to our knowledge, novel. This technique is highly versatile since a cutting plotter is used to define the features that need to be removed. A drawback of the use of optical film is its limited flexibility, as a result of which substantial pressure is necessary to activate the valves. Using a polydimethylsiloxane layer may be more appealing in this respect, but bonding of the latter to other polymeric substrates is challenging (Tang and Lee, 2010), which would complicate the fabrication process.

Optical isolation of assay chambers in the microfluidic chip

To evaluate the efficacy of the valves, it was evaluated whether fluidic cross-talk occurred between adjacent reaction wells. To this end, two different experiments were performed. In the first experiment, primers were dried in alternating reaction channels and the chip loaded with LAMP reaction mixture containing target DNA (Figure 5.7). In the second experiment, primers along with target DNA was dried down in alternating reaction channels, with the other channels only containing primer (Figure 5.7) and the chip loaded with solely reaction mixture.

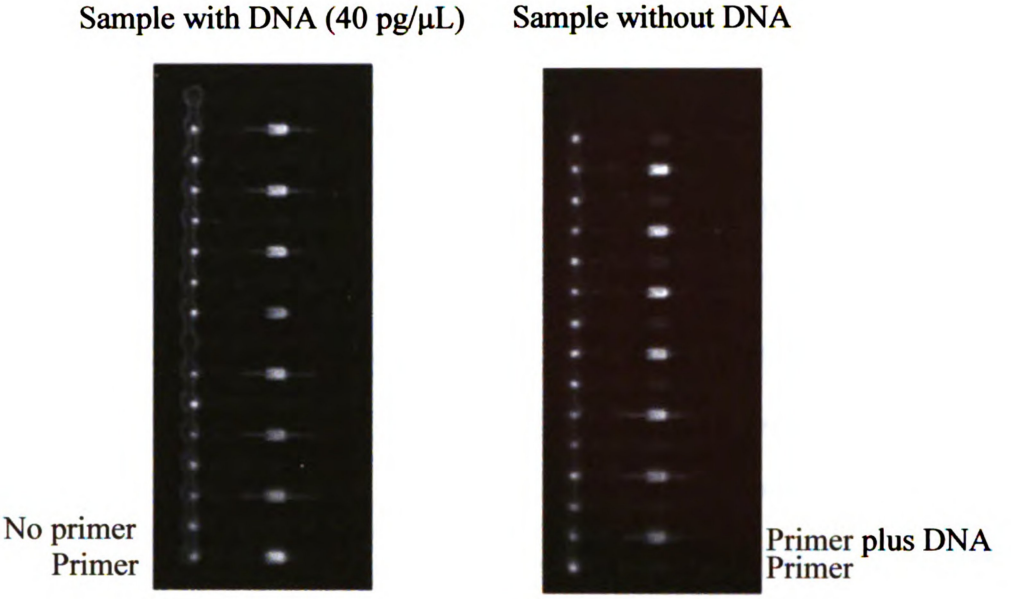


FIGURE 5.7 Fluidic isolation of the reaction wells. The content dried in the reaction wells is shown for the lower two wells, but this pattern applies to the entire chip. Images were taken after 40 min of amplification in the presence of 20 μM SYTO 81 dye.

Both experiments clearly demonstrated that no DNA (primers or amplicons) migrated between adjacent reaction channels, as no fluorescence was visible in the wells lacking primer in the first experiment and the wells lacking DNA in the second experiment. This suggests that the valves effectively isolated the reaction wells and each reaction well could be considered as an independent assay not influenced by the neighboring wells.

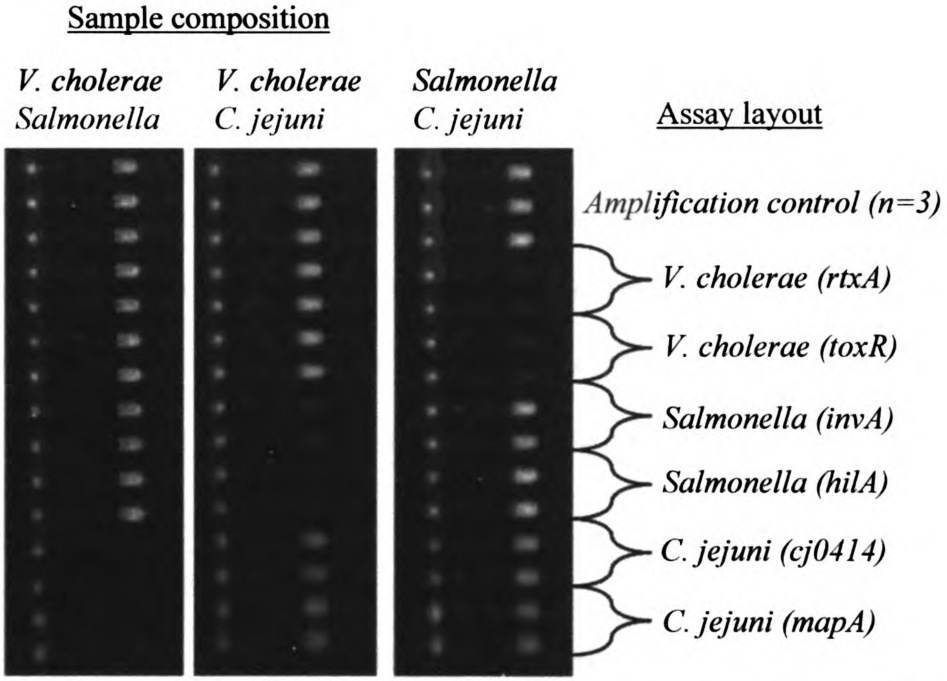


FIGURE 5.8 Multiplexed detection using the microfluidic chips. The samples contained 2 pg/μL of DNA from the indicated pathogens, which amounts to roughly 500 genomic copies per reaction well. The position of the different primer sets is indicated. Images were taken after 40 min of amplification in the presence of 20 μM SYTO 81 dye.

Furthermore, inspection of the chips after amplification verified that amplification predominantly occurred in the region enclosed by the valve pairs, demonstrating that the valves are effective in isolating the reaction wells. Sealing of the reaction wells also prevented formation of air bubbles, which always occurred in unsealed reaction wells due to the surface roughness of the molds.

Demonstration of parallel detection of multiple pathogens

To illustrate the utility of the chips for multiplexed pathogen detection, six different LAMP primer sets for three major diarrheal pathogens (*C. jejuni*, *Salmonella* and *V. cholerae*) were dispensed in duplicate in the chips along with three positive amplification controls (Figure 5.8). Three binary genomic DNA mixtures were then loaded in the chip

along with reaction mixture and subjected to LAMP for 40 min, after which the fluorescence was imaged.

For all pathogen mixtures, the pattern of positive/negative wells corresponded to those expected, which indicates that the chips performed as expected (Figure 5.8). It also suggests that the newly designed primers are specific for the genomic DNA tested, but additional testing is necessary with other bacteria and stool samples.

CONCLUSIONS

In summary, a robust and user-friendly polymer microfluidic chip for multiplexed LAMP was developed. The chip contains four arrays of 15+1 interconnected reaction channels, allowing analysis of multiple samples for several genetic targets. To load the sample and seal the chip, only a common pipettor and tape is required. This makes the chip very attractive for adoptation in resource-limited settings. Effective fluidic isolation of the reaction wells was accomplished using pinch-type microvalve pairs, which offer the possibility to further increase the density of the reaction wells since fluidic cross-talk between closely spaced wells is not a concern. In terms of fabrication, a novel method for monolithic integration of hundreds of pinch-type valves in a few simple step using a cutting plotter and lamination was presented.

Toward further development, the chip will be combined with a modified version of the optical module described in Chapter 4 to provide a low-cost and compact device for LAMP in disposable microfluidic chips with pre-dispensed and stabilized reagents. After

validation of the primer sets with field samples, this device could be used for detection of bacteria responsible for the majority of infectious agents responsible for diarrhea.

APPENDIX

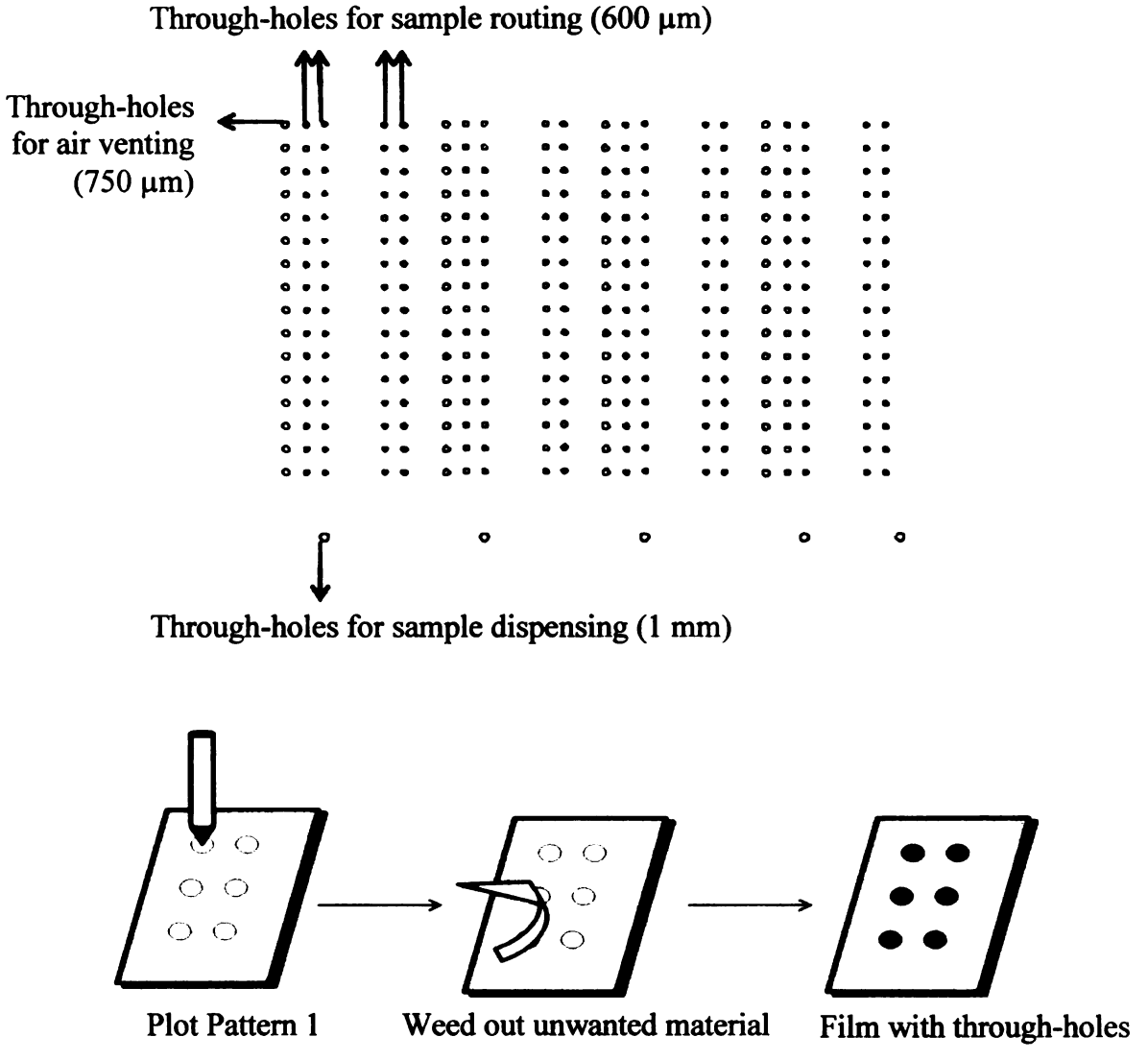


FIGURE A1. Schematic of Pattern 1 and process scheme.

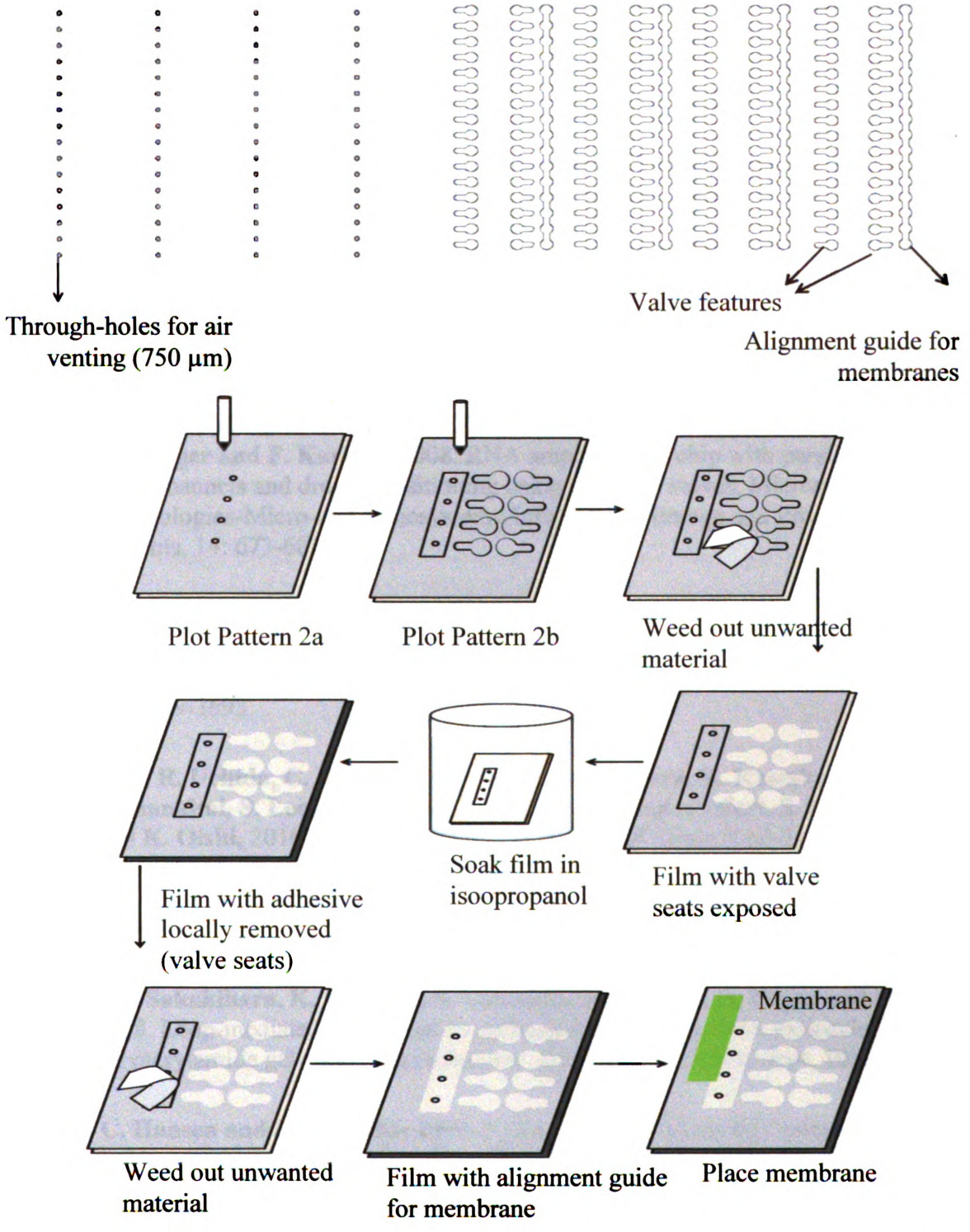


FIGURE A2. Schematics of Pattern 2a and 2b and process scheme. Images in this thesis/dissertation are presented in color.

REFERENCES

Antikainen, J., E. Tarkka, K. Haukka, A. Siitonen, M. Vaara and J. Kirveskari, 2009. New 16-plex PCR method for rapid detection of diarrheagenic Escherichia coli directly from stool samples. European Journal of Clinical Microbiology & Infectious Diseases, 28: 899-908.

Bartholomeusz, D.A., R.W. Boutte and J.D. Andrade, 2005. Xurography: Rapid prototyping of microstructures using a cutting plotter. Journal of Microelectromechanical Systems, 14: 1364-1374.

Fang, X.E., Y.Y. Liu, J.L. Kong and X.Y. Jiang, 2010. Loop-Mediated Isothermal Amplification Integrated on Microfluidic Chips for Point-of-Care Quantitative Detection of Pathogens. Analytical Chemistry, 82: 3002-3006.

Furuberg, L., M. Mielnik, A. Gulliksen, L. Solli, I.R. Johansen, J. Voitel, T. Baier, L. Riegger and F. Karlsen, 2008. RNA amplification chip with parallel microchannels and droplet positioning using capillary valves. Microsystem Technologies-Micro-and Nanosystems-Information Storage and Processing Systems, 14: 673-681.

Hataoka, Y., L.H. Zhang, Y. Mori, N. Tomita, T. Notomi and Y. Baba, 2004. Analysis of specific gene by integration of isothermal amplification and electrophoresis on poly(methyl methacrylate) microchips. Analytical Chemistry, 76: 3689-3693.

Kerdsin, A., R. Uchida, C. Verathamjamrus, P. Puangpatra, K. Kawakami, P. Puntanakul, S. Lochindarat, T. Bunnag, P. Sawanpanyalert, S. Dejsirilert and K. Oishi, 2010. Development of Triplex SYBR Green Real-Time PCR for Detecting Mycoplasma pneumoniae, Chlamydophila pneumoniae, and Legionella spp. without Extraction of DNA. Japanese Journal of Infectious Diseases, 63: 173-180.

Lam, L., S. Sakakihara, K. Ishizuka, S. Takeuchi, H.F. Arata, H. Fujita and H. Noji, 2008. Loop-mediated isothermal amplification of a single DNA molecule in polyacrylamide gel-based microchamber. Biomedical Microdevices, 10: 539-546.

Liu, J., C. Hansen and S.R. Quake, 2003. Solving the "world-to-chip" interface problem with a microfluidic matrix. Analytical Chemistry, 75: 4718-4723.

Lutz, S., P. Weber, M. Focke, B. Faltin, J. Hoffmann, C. Muller, D. Mark, G. Roth, P. Munday, N. Armes, O. Piepenburg, R. Zengerle and F. von Stetten, 2010. Microfluidic lab-on-a-foil for nucleic acid analysis based on isothermal recombinase polymerase amplification (RPA). Lab on a Chip, 10: 887-893.

Martineau, F., F.J. Picard, L. Grenier, P.H. Roy, M. Ouellette, M.G. Bergeron and E. Trial, 2000. Multiplex PCR assays for the detection of clinically relevant antibiotic resistance genes in staphylococci isolated from patients infected after cardiac surgery. Journal of Antimicrobial Chemotherapy, 46: 527-533.

Matsubara, Y., K. Kerman, M. Kobayashi, S. Yamamura, Y. Morita and E. Tamiya, 2005. Microchamber array based DNA quantification and specific sequence detection from a single copy via PCR in nanoliter volumes. Biosensors & Bioelectronics, 20: 1482-1490.

Morrison, T., J. Hurley, J. Garcia, K. Yoder, A. Katz, D. Roberts, J. Cho, T. Kanigan, S.E. Ilyin, D. Horowitz, J.M. Dixon and C.J.H. Brenan, 2006. Nanoliter high throughput quantitative PCR. Nucleic Acids Research, 34.

Paul, D., A. Pallandre, S. Miserere, J. Weber and J.L. Viovy, 2007. Lamination-based rapid prototyping of microfluidic devices using flexible thermoplastic substrates. Electrophoresis, 28: 1115-1122.

Ramalingam, N., H.B. Liu, C.C. Dai, Y. Jiang, H. Wang, Q.H. Wang, K.M. Hui and H.Q. Gong, 2009a. Real-time PCR array chip with capillary-driven sample loading and reactor sealing for point-of-care applications. Biomedical Microdevices, 11: 1007-1020.

Ramalingam, N., T.C. San, T.J. Kai, M.Y.M. Mak and H.Q. Gong, 2009b. Microfluidic devices harboring unsealed reactors for real-time isothermal helicase-dependent amplification. Microfluidics and Nanofluidics, 7: 325-336.

Tang, L. and N.Y. Lee, 2010. A facile route for irreversible bonding of plastic-PDMS hybrid microdevices at room temperature. Lab on a Chip, 10: 1274-1280.

Weigl, B.H., D.S. Boyle, T. de los Santos, R.B. Peck and M.S. Steele, 2009. Simplicity of use: a critical feature for widespread adoption of diagnostic technologies in low-resource settings. Expert Review of Medical Devices, 6: 461-464.

Yager, P., G.J. Domingo and J. Gerdes, 2008. Point-of-care diagnostics for global health. Annual Review of Biomedical Engineering, 10: 107-144.

Yamazaki, W., M. Taguchi, M. Ishibashi, M. Kitazato, M. Nukina, N. Misawa and K. Inoue, 2008. Development and evaluation of a loop-mediated isothermal amplification assay for rapid and simple detection of *Campylobacter jejuni* and *Campylobacter coli*. Journal of Medical Microbiology, 57: 444-451.

Zhang, L., T. Hurek and B. Reinhold-Hurek, 2005. Position of the fluorescent label is a crucial factor determining signal intensity in microarray hybridizations. Nucleic Acids Research, 33.

CHAPTER SIX

MULTIPLEXED DETECTION OF DIARRHEAL BACTERIA USING LOOP-MEDIATED ISOTHERMAL AMPLIFICATION IN A LOW-COST AND PORTABLE MICROFLUIDIC DEVICE

INTRODUCTION

Infectious diarrheal disease is the second leading cause of mortality and morbidity worldwide (Kosek et al., 2003; Black et al., 2010). More than 1.3 million children die each year due to diarrhea, the vast majority in developing nations (Black et al., 2010). Rapid and accurate identification of the responsible agents is crucial for effective management of the disease and also to prevent its spread in the community. However, the only available method in impoverished countries involves culturing, which takes 3-5 days and also requires considerable expertise.

Because infectious diarrhea can be caused by several pathogens, including protozoa, viruses and bacteria, multiplex detection techniques are necessary for accurate diagnosis. Several such methods based on nucleic acid amplification tests (NAAT) using multiplex PCR (Wang, 2008; Antikainen et al., 2009; O'Leary et al., 2009) and microarrays (You et al., 2008; Suo et al., 2010) have been developed, and some of these are commercially available, such as the EntericBio multiplex PCR system from Serosep Ltd. (Limerick, Ireland). Because these assays require bulky and expensive instrumentation they are, however, not typically available in developing nations. Therefore, a need exists for more affordable and easy-to-use devices for NAAT for diagnosis of diarrheal pathogens to help

health care providers prescribe more effective treatments and also provide valuable surveillance and epidemiological data.

The goal of this study was to apply a low-cost and compact device for loop-mediated isothermal amplification (LAMP) for rapid and accurate detection of six major diarrheal bacteria: *Campylobacter jejuni*, *Salmonella*, enterotogenic *Escherichia coli* (ETEC), *Vibrio cholerae*, and *Yersinia enterocolitica*. The instrument integrates the two key components of this device developed in previous work integrated in a fully packaged device: a 64-channel polymer microfluidic chip (described in Chapter 5) and a LED/photodiode-based optical sensor (described in Chapter 4). The pathogens targeted in this study are from the high priority list prepared by Bill and Melinda Gates Foundation. One of the aims of the Foundation is 'to develop and deliver innovative tools and approaches to help prevent, diagnose, and treat enteric and diarrheal diseases around the world.' The approach presented here addresses this need by providing a low-cost, battery-operated and easy-to-use device for multiplexed detection of diarrheal pathogens using NAAT.

DEVICE INTEGRATION

Toward integration of the device in a fully packaged and functional prototype, several tasks had to be completed. First, the electronics for control and data acquisition needed to be transferred to a custom designed printed circuit board (PCB), with all software being contained in a single microcontroller. Second, a housing needed to be designed to contain the chip holder with embedded heater, optical components (LEDs, optical fibers and

photodiode) and PCB. Third, a Wi-Fi data communication protocol needed to be established between the microcontroller and iPod Touch. Due to the breath of these tasks, a concerted effort from several students and collaborators was necessary, and the prototype presented here is based on this combined endeavor.

Embedded control and data acquisition

Embedding of the system involved design of a custom PCB and programming of a microcontroller. As described in the Ph.D. dissertation of my colleague Dr. Robert Stedtfeld, at the core of the instrument is a 32-bit ARM 7 microcontroller (LPC2378, NXP Semiconductors) for temperature control, triggering of the LEDs, acquiring timed signal readings from the photodiode and sensing the data to a Wi-Fi module (Mini-Microprocessor Wi-Fi Core Module Rabbit Module), which then sends the data to the iPod Touch (Figure 6.1).

To control the chip temperature, the microcontroller receives temperature readings from a thermocouple placed on the chip holder and adjusts the voltage applied to the heater using proportional–integral–derivative (PID) control and pulse width modulation (PWM). The microcontroller also trigger the LEDs in a predetermined pattern using four de-multiplexers connected to 64 surface mount 50 mA LED drivers. Finally, it also records the photodiode readings and sends it to a Wi-Fi module through a serial communication port. The Wi-Fi module then automatically routes the data in real time to the iPod Touch for plotting and storage. To start the instrument, the microcontroller receives a string from the iPod Touch that contains information about the requested assay time and temperature.

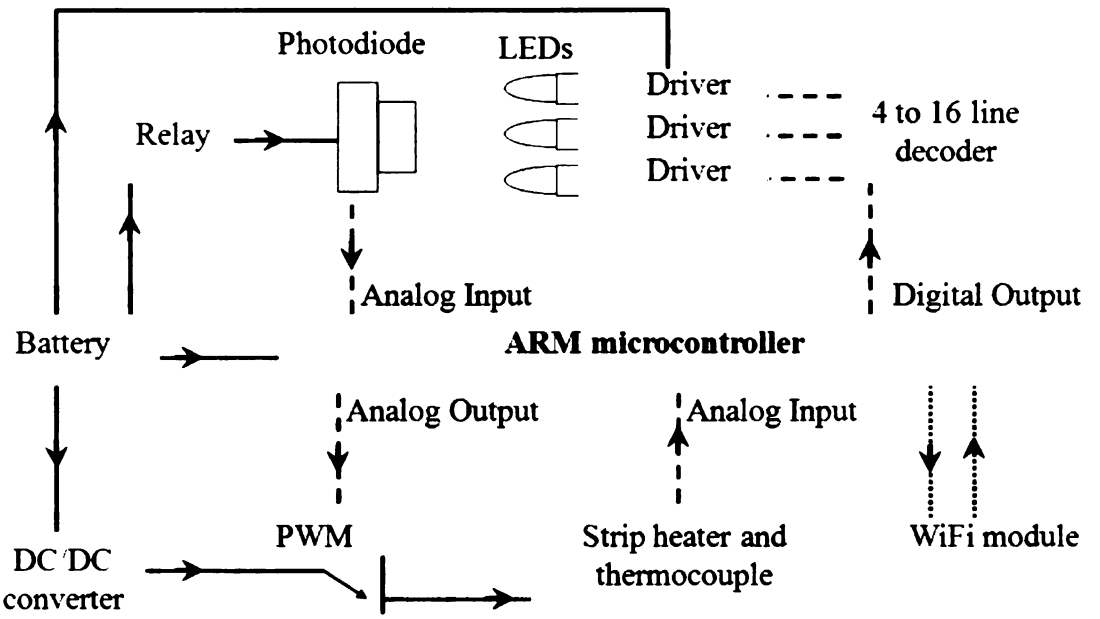


FIGURE 6.1 Functional block diagram of the modules of the embedded instrument. Kindly provided by Robert D. Stedtfeld. Images in this thesis/dissertation are presented in color.

GeneZ software

A specific application (GeneZ) was developed for the iPod Touch by Scott Price, another team members involved in this project. The software allows the user to set the reaction temperature and time (Figure 6.2), and start the assay. During the instrument run, data readings from the photodiode are continuously sent from the instrument to the iPod Touch, where the data is sorted and plotted for each reaction well. At the end of the assay, the software automatically calculates the time-to-positivity (TTP) for each reaction well, calculates averages for replicated assays and estimates the target gene copy number using a predefined calibration plot.

FIGURE 6.2 Screen shots GeneZ application for the iPod Touch. Images in this thesis/dissertation are presented in color.

Housing design and assembly

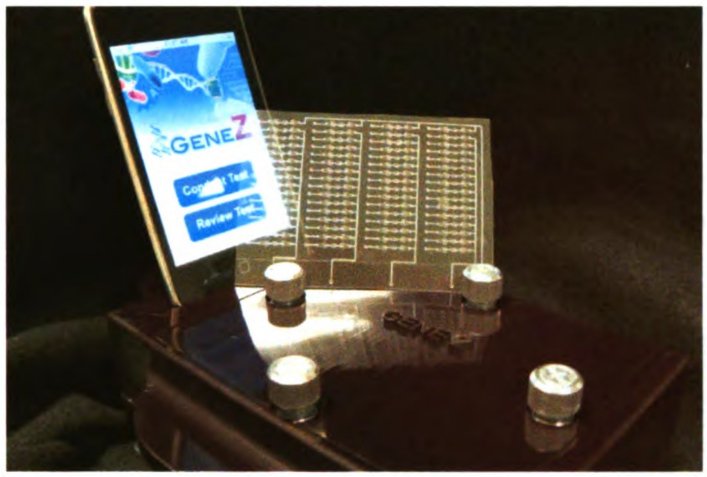
A first prototype of the instrument was completed as part of this study (Figure 6.3a). The housing measures 15 × 10 × 5 cm and is fabricated *via* stereolithography out of black Huntsman RenShape 7820 resin (FineLine Prototyping; Raleigh, NC). The housing consists of two parts that can be separated to facilitate assembly of the system.

The key functional component of the instrument is the chip holder with integrated heater, LEDs and optical fibers. Compared to the system evaluated in Chapter 4, the layout of the optics needed to be modified to permit sensing of a higher density chip (Figure 6.4). More specifically, rather than coupling the fibers to the wells at a 90° angle, the fibers needed to transect the chip holder and be oriented at a 135° angle. To improve

the coupling efficiency between the fibers to the reaction wells, one of the sides of the wells has a draft angle of 45°. Also, due to space constraints, a 750 μm fiber was used rather than the 1 mm fiber used in Chapter 4.

To assemble the system, a chip holder was fabricated out of black acrylonitrile butadiene styrene (ABS) by hot embossing with the mold used for fabrication of the chips. Flexible electrical resistance heating tape strip heater (TEMPCO Electric Heater Corporation; Wood Dale, IL) was embedded in the holder during embossing to achieve effective heating of the chip holder (Figure 6.4). Using this type of heating element rather than a more conventional thin film heater was necessary for the LEDs and optical fibers to be inserted in the holder, and proved highly effective and versatile. After fabrication of the holder with embedded heaters, holes were manually drilled for the fibers and LEDs. The latter were mounted on a custom-designed PCB board, as shown in Figure 6.3b, with spacers to raise the LEDs from the board. Prior to assembly of the housing, the fibers were glued in place and their ends bundled and coupled to the photodiode using a simple fixture. The chip holder was then placed on the PCB board with LEDs, with the fibers being routed in the 'wings' in the housing. This was arguably the most challenging step during assembly of the prototype due the limited space between the bottom of the chip holder and PCB.

(a)



(b)

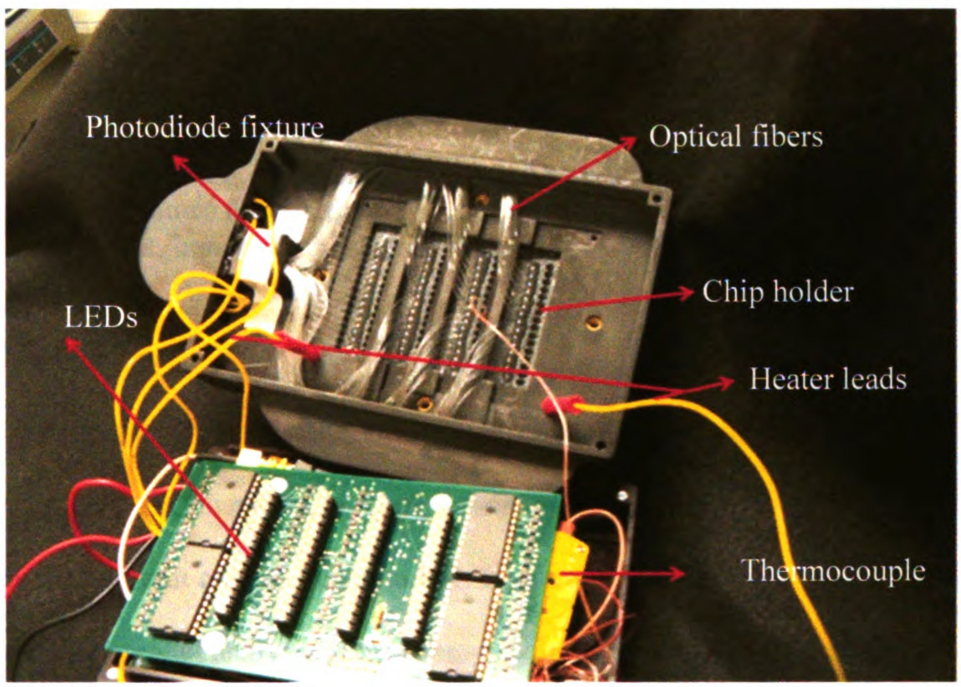


FIGURE 6.3 Photographs of the developed instrument. (a) The device along with the microfluidic chip and iPod Touch. (b) The disassembled prototype showing the various components in the device. Images in this thesis/dissertation are presented in color.

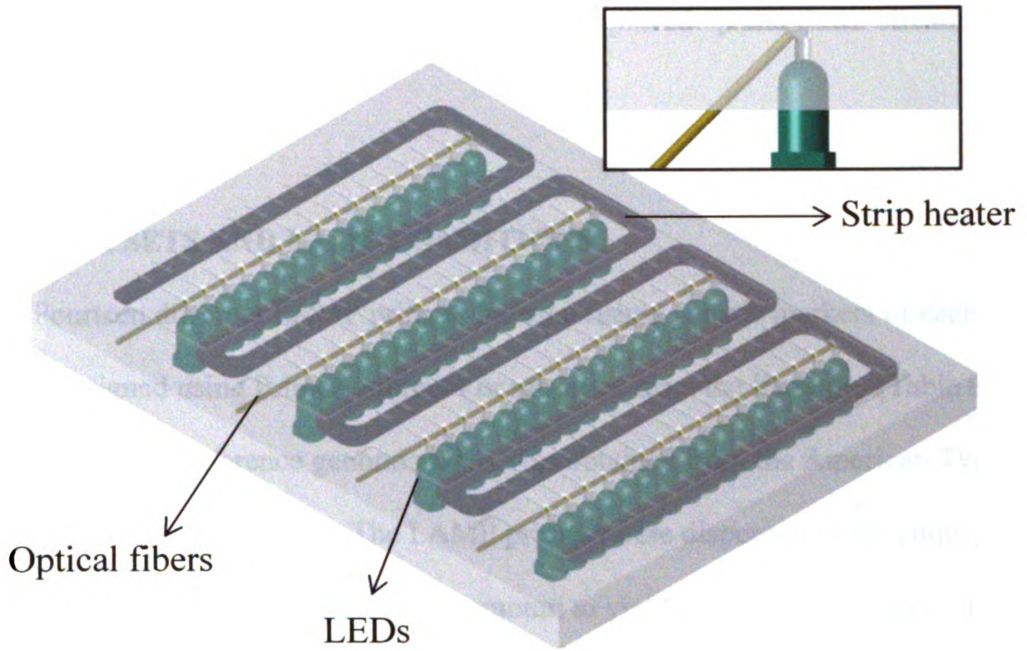


FIGURE 6.4 Rendering of the chip holder with embedded heater, LEDs and optical fibers. The inset shows the positioning of the LEDs and optical fibers with respect to the wells. Images in this thesis/dissertation are presented in color.

Microfluidic chips

The microfluidics chip layout, operating principle and the method for fabrication are described in detail in Chapter 4, and only a brief summary is provided below. The chips are the size of a credit card and contain four arrays of 15+1 reaction wells (1 μL per well), enabling parallel analysis of four samples for up to 15 pathogens or genetic markers. Each array of 15 reaction wells is connected to a sample inlet port in which the user dispenses a sample using a common pipettor. The distribution channel routes the sample to the different reaction wells in which primers were pre-dispensed and dried during chip fabrication. After dispensing, the chip is sealed and placed in the instrument. An additional layer of PCR tape is then placed on the chip to prevent evaporation through the hydrophobic membranes that serve as air vents. Upon closing of the lid, an array of

pre-aligned plungers closes microvalves that are placed up and down-stream of each reaction well.

PRIMER SETS AND LAMP CONDITIONS

Fourteen different LAMP primers for established genetic markers of each pathogen were designed using PrimerExplorer4 or retrieved from the literature (Table 6.1). For each pathogen, reference genomic DNA was obtained from the American Type Culture Collection (Manassas, VA). The LAMP primers were dispensed in the chips prior to assembly of the different layers in an amount to yield a final concentration of 1.6 μM each of FIP and BIP primer, 800 nM each of LF and LB primer, 200 nM each of F3 and B3 primer after dissolution in the reaction well. The LAMP mixture loaded in the chips consisted of 800 mM betaine (Sigma Aldrich), 1.4 mM of each dNTP (Invitrogen), 20 mM Tris-HCl (pH 8.8), 10 mM $(NH_4)_2SO_4$, 10 mM KCl, 8 mM $MgSO_4$, 8 mM Triton X-100, 0.64 units/μL of *Bst* polymerase (New England Biolabs), 20 μM of SYTO 81 dye (Invitrogen) and target DNA (1 μL per 25 μL LAMP mixture).

TABLE 6.1 Primers used in this study.

Pathogen/gene Reference		**Sequence (5'-3')**
C. jejuni		
cj0414	F3	GCAAGACAATATTATTGATCGC
(Yamazaki et al., 2008)	B3	CTTTCACAGGCTGCACTT
	FIP	ACAGCACCGCCACCTATAGTAGAAGCTTTTTTAAACTAGGGC
	BIP	AGGCAGCAGAACTTACGCATTGAGTTTGAAAAAACATTCTACCTCT
	LF	CTAGCTGCTACTACAGAACCAC
	LB	CATCAAGCTTCACAAGGAAA
mapA	F3	ACAAGATACTTTTGCTCAAGTT
This study	B3	GCATTAAAATTCACATCRACAA
	FIP	AACATCGCTAATGTATAAAAGCCCTCTCAATGCAGTTCTTGTGAAAG
	BIP	TTCAATGTTGTGCCAATAAACGCTTTAAAACCTTTTGCYCTTCT
	LF	TTTGCTTCAAAACCACCAGGA
	LB	CTGGTATTGCTTTGAAAAAGGTTTATTTAC
ETEC		
eltA	F3	GCCATTATATGCAAATGGCG
(Yano et al., 2007)	B3	CCTGCTAAGTGAGCACTTCT
	FIP	CTCATTATGCCCTCTGGGCAACTCTAGACCCCCAGATGA
	BIP	ATGATCACGCGAGAGGAACACAAAGTGGAAACATATCCGTCA
	LF	AAGACCTCCGGAACGTTTTA
	LB	ACCGGCTTTGTCAGATATGA
eltB	F3	ATGGCAGGCAAAAGAGAA
This study	B3	CCATACTGATTGCCGCAA
	FIP	GAGTCTATATGTTGACTGCCCGGATGGTTATCATTACATTTAAGAGCG
	BIP	TGAAAGGATGAAGGACACATTAAGAGGGGTTTTATTATTCCATACACA
	LF	ACTTCGACCTGAAATGTTGCG
	LB	ATCRCATATCTGACCGAGACCA
estIA	F3	CTCAGGATGCTAAACCAGT

Table 6.1 continued

(Yano et al., 2007)	B3	CAGAACAAATATAAAGGGAACTGTT
	FIP	TCATGCTTTCAGGACCACTTTTATTGAGTCTTCAAAAGAAAAAATCACACT
	BIP	AGTAGCAATTACTGCTGTGAATTGTCCCTTTATATTATTAATAGCACCCG
	LF	GTTGTAATCCTGCTTGT
Salmonella		
phoP	F3	GCCATTCCACATCGAAGAGGT
(Li et al., 2009)	B3	ATGAGAACATCAATGGTATGGC
	FIP	GGCGTGAGAGATCCACCTGGAATGCGCCGTAATAGCGGTC
	BIP	CACCATTATGGAAACGCTTATCCGCCGGATACAGCTGAAGCATC
	LF	CAGGTGATCAACATCCCGCC
	LB	CGGTAAAGTGGTCAGCAAAGAT
invA	F3	CGGCCCGATTTTCTCTGG
This study	B3	CGGCAATAGCGTCACCTT
	FIP	GCGCAGCATCCGCATCAATAATATGGTATGCCCGGTAAACAG
	BIP	GAACGGCGAAGCGTACTGGACATCGCACCGTCAAAGGAA
	LF	CCTTCAAATCGGCATCRATACTCAT
	LB	AAGGGAAAGCCAGCTTTACG
hilA	F3	GGATCAGGTTCAATCCGAG
This study	B3	TGTACAATATTATCATTMCCATCGG
	FIP	TTCGTAATGGTCACCGGCAGAGTCTGCATTACTCTATCGTG
	BIP	AACTGCCGCAGTGTTAAGGATATCATCTGCCCGGAGAT
	LF	GCGCATACTGCGATAATCCCTT
	LB	CTTGAGCTCATGGATCAATTACGCC
Shigella		
ipaH	F3	GCATGCCAACACCTTTTCC
This study	B3	TGATGGACCAGGAGGGTT
	FIP	CGACCTGTTCACGGAATCCGG-CTTGACCGCCTTTCCGATAC
	BIP	AGCAGTCTTTCGCTGTTGCTGC-TCCGGAGATTGTTCCATGTG
	LF	GGTATTGCGTGCAGAGACG

Table 6.1 continued

	LB	TGATGCCACTGAGAGCTGTG
V. cholerae		
ctxA	F3	TCGGGCAGATTCTAGACC
(Okada et al., 2010)	B3	GTGGGCACTTCTCAAACT
	FIP	TTGAGTACCTCGGTCAAAGTACTTCCTGATGAAATAAAGCAGTCA
	BIP	TCAACCTTTATGATCATGCAAGAGGGGAAACATATCCATCATCGTG
	LF	CCTCTTGGCATAAGACCACC
	LB	AACTCAGACGGGATTTGTTAGG
rtxA	F3	ACCGTACTTGGTCTAACCGT
This study	B3	TTGCGTGTAGACATCTTCGG
	FIP	CCTGTCTCAACGCGTGAACTGATTGGCGATGTGACCTTTGATG
	BIP	TTAGTCCGTAAAGGCAAAGTGGGCGATGTGTCCGCTCAATGCG
	LF	TGTTCGCGGCACCAGCA
	LB	GATATTACTCTGCAAGGTGCTGG
toxR	F3	CGAGTGGAAACGGTTGAAGA
This study	B3	AGGGGAAGTAAGACCGCTAT
	FIP	GCACACTGCTTGAYTCTGCGTACGAAAGCGAAGCTGCTCAT
	BIP	AGCCACTGTAGTGAACACACCGTCGATTCCCCAAGTTTGGAG
	LF	ACAGATTCTGGCTGAGAGATGTC
	LB	CAGCCAGCCAATGTTGTGAC
Y. enterocolitica		
phoP	F3	GCCATTCCACATCGAAGAGGT
(Li et al., 2010)	B3	ATGAGAACATCAATGGTATGGC
	FIP	GGCGTGAGAGATCCACCTGGAATGCGCCGTAATAGCGGTC
	BIP	CACCATTATGGAAACGCTTATCCGCCGGATACAGCTGAAGCATC
	LF	CAGGTGATCAACATCCCGCC
	LB	CGGTAAAGTGGTCAGCAAAGAT

RESULTS AND DISCUSSION

To demonstrate the performance of the developed instrument, four ternary mixtures of pathogen genomic DNA were loaded in the chip and subjected to LAMP. The amount of DNA loaded in the chip was around 40 pg per array, which yields roughly 1.3 pg of DNA (or ~500 genome copies) per reaction well. A fluorescence image was then recorded after 40 min of amplification using a simple imaging system, akin the setup described in Chapter 5.

As illustrated in Table 6.2, excellent concordance was observed between the presence/absence calls based on visual inspection of the endpoint signal intensity and the composition of the mixtures, for all but two genes. This shows that developed primer sets performed well in the microfluidic chips and integrated device. For one gene (*eltA* of ETEC), no amplification was observed in two separate mixtures. However, previous evaluation of this primer set using a commercial real time PCR instrument indicated that the primer successfully amplified its intended target at high copy number, indicating that the lack of detection was most likely due to poor sensitivity. The other gene (*cj0414* of *C. jejuni*) yielded positive amplification even the absence of its target in two mixtures, suggesting that this primer set is non-specific or cross-over contamination occurred with previous experiments.

In another experiment with the same sample mixtures, the reaction was monitored in real time using the 64-channel optical sensor to demonstrate the utility of the prototype for real time fluorogenic LAMP. The amplification plots for all primer sets (with the

exception of *eltA*) are shown in Figure 6.5. For all genes, an increase in signal intensity occurred within 30 min, indicating that the high density optical unit is capable of monitoring the reactions. However, the signal intensity was substantially lower than that observed using the seven well sensor described in Chapter 4. This can only in part be explained by the use of smaller fibers at this would also yield a decrease in signal of about ~1.8-fold based on the reduced area available for light collection. In addition, the baseline fluctuated more than previously, which was attributed to the formation of air bubbles when the valves were not effectively closed. Inspection of the wells after amplification revealed a substantial amount of air bubbles, which will impact signal intensity due to light scattering and reduced sample volume in the wells. For mixture 4, few air bubbles were observed and the corresponding amplification plots displayed fewer fluctuations. Furthermore, air bubble formation also leads to ejection of a portion of the amplification mixture out of the reaction wells, which can be expected to negatively impact assay reproducibility.

It should be noted that the time-to-detection observed in this experiment were significantly higher than these observed using a commercial real time PCR machine. However, this is not due to the modification in optics and accompanying lower signal intensity as previous experimentation with high brightness LEDs showed that the amount of signal does not impact the TTP (data not shown). The delayed amplification was most likely attributed to the fact that the chips used were prepared several days prior to the experiment and stored at room temperature. While dried primers can typically be stored without refrigeration, this indicates that evaluation of the storage conditions will be

needed toward product commercialization.

TABLE 6.2 Results from the validation experiment with ternary mixtures.

Primer	Mixture 1	Mixture 2	Mixture 3	Mixture 4
CJ_0414	+/+[a]	+/+	-/+	-/+
CJ_mapA	+/+	+/+	-/-	-/-
EC_eltA	+/-	-/-	+/-	-/-
ETEC_eltB	+/+	-/-	+/+	-/-
ETEC_estIA	+/+	-/-	+/+	-/-
Salm_invA	+/+	+/+	-/-	-/-
Salm_hilA	+/+	+/+	-/-	-/-
Salm_phoP	+/+	+/+	-/-	-/-
Shig_ipaH	-/-	+/+	+/+	+/+
VC_ctxA	-/-	-/-	+/+	+/+
VC_rtxA	-/-	-/-	+/+	+/+
VC_toxR	-/-	-/-	+/+	+/+
YE_ail	-/-	-/-	-/-	+/+
YE_phoP	-/-	-/-	-/-	+/+

[a] Expected/observed.

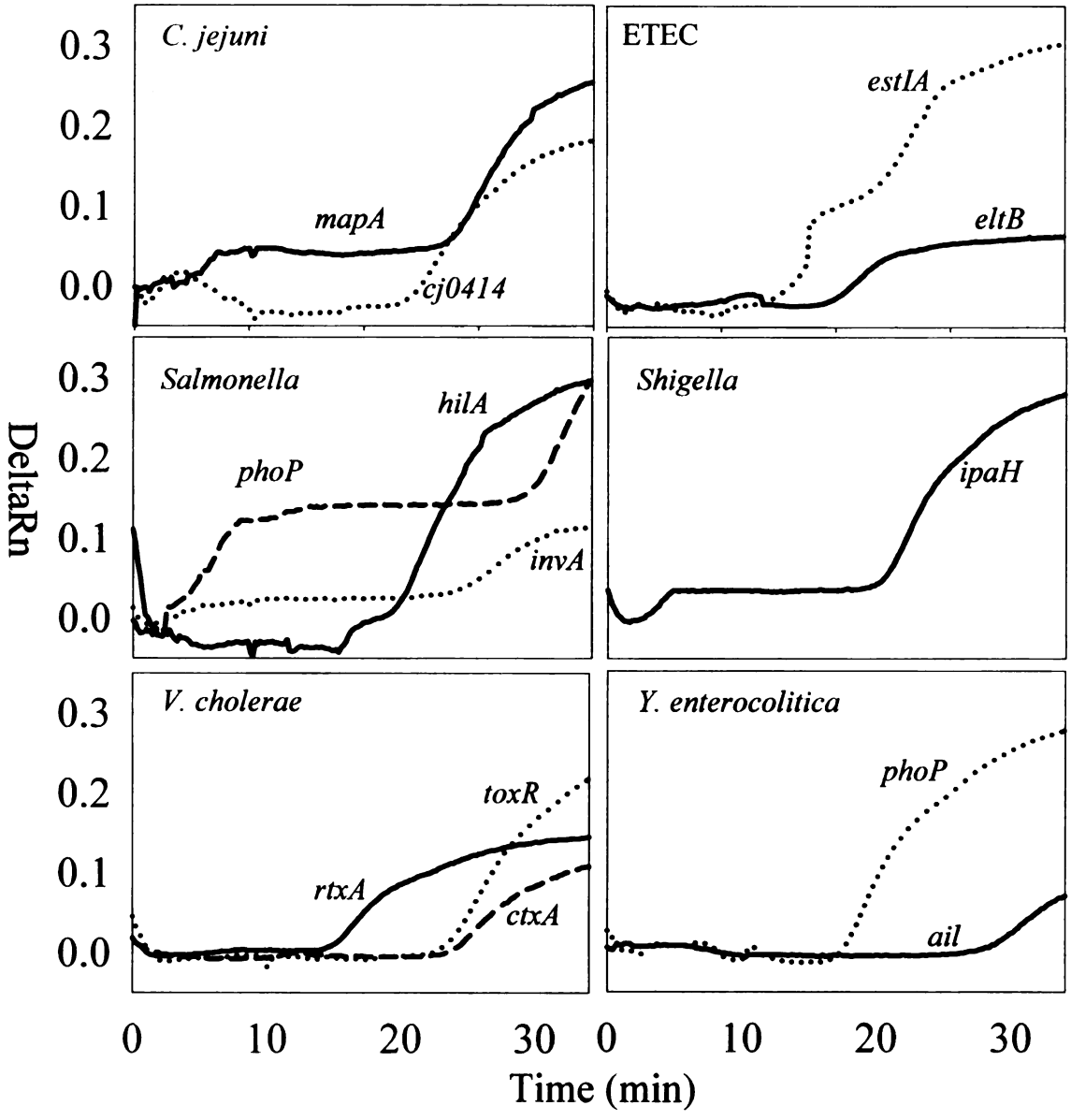


FIGURE 6.5 Amplification plots as measured using the developed instrument. The severe baseline fluctuations are most likely due to air bubble formation when the microvalves are not effectively closed.

CONCLUSIONS

In this study, a fully functional prototype of a low-cost, robust and compact instrument for microfluidic LAMP was developed. In a single instrument run, four purified DNA samples can be analyzed for up to 15 pathogens in disposable microfluidic

chips. At the end of the user, the sample can be readily dispensed in the chip using a common pipettor, after which the chip is placed in the heated chip holder with integrated optics. Using an iPod Touch, the user then sets the required amplification time and temperature and initiates the instrument wirelessly through Wi-Fi. At the end of the instrument, automated data analysis then calculates the TTP for each reaction well and estimates the target copy number based on pre-defined calibration parameters. Preliminary validation of the developed LAMP primer sets for six major diarrheal bacteria showed that the primers, microfluidic chips and optical sensor performed well. However, air bubble formation led to substantial fluctuations in baseline signal intensity, which complicated estimation of TTP. In this prototype, closing of the valves proved difficult due to the use of thumb screws rather than toggle clamps and the flexibility of the closing lid. The design of a robust closing lid therefore needs to be given due consideration for the next prototype.

Finally, the use of a cell phone as user-interface is a salient feature of the device since the ability of 'exchanging information between the site of the analysis and a central location staffed by experts, with real-time interpretation of the assay by those experts (i.e., telemedicine)' is a crucial requirement of analytical devices for developing nations (Martinez et al., 2008). This interest is sparked by the widespread availability of cell phones even in developing nations, with several researchers exploiting this to develop low-cost diagnostics tools for use in remote settings, such as camera phones for clinical microscopy (Breslauer et al., 2009) and digitizing colorimetric assay readouts (Martinez et al., 2008).

REFERENCES

Antikainen, J., E. Tarkka, K. Haukka, A. Siitonen, M. Vaara and J. Kirveskari, 2009. New 16-plex PCR method for rapid detection of diarrheagenic *Escherichia coli* directly from stool samples. European Journal of Clinical Microbiology & Infectious Diseases, 28: 899-908.

Black, R.E., S. Cousens, H.L. Johnson, J.E. Lawn, I. Rudan, D.G. Bassani, P. Jha, H. Campbell, C.F. Walker, R. Cibulskis, T. Eisele, L. Liu, C. Mathers and U. Who, 2010. Global, regional, and national causes of child mortality in 2008: a systematic analysis. Lancet, 375: 1969-1987.

Breslauer, D.N., R.N. Maamari, N.A. Switz, W.A. Lam and D.A. Fletcher, 2009. Mobile Phone Based Clinical Microscopy for Global Health Applications. PLOS One, 4.

Kosek, M., C. Bern and R.L. Guerrant, 2003. The global burden of diarrhoeal disease, as estimated from studies published between 1992 and 2000. Bulletin of the World Health Organization, 81: 197-204.

Li, X.F., S. Zhang, H.W. Zhang, L.H. Zhang, H.T. Tao, J. Yu, W.J. Zheng, C.H. Liu, D. Lu, R. Xiang and Y. Liu, 2009. A loop-mediated isothermal amplification method targets the *phoP* gene for the detection of *Salmonella* in food samples. International Journal of Food Microbiology, 133: 252-258.

Li, Y.J., M. Jiang, W. Liu, L.H. Zhang, S. Zhang, X.P. Zhao, R. Xiang and Y. Liu, 2010. Loop-mediated isothermal amplification method targets to the *phoP* gene for detection of *Yersinia enterocolitica*. Molecular and Cellular Probes, 24: 68-71.

Martinez, A.W., S.T. Phillips, E. Carrilho, S.W. Thomas, H. Sindi and G.M. Whitesides, 2008. Simple telemedicine for developing regions: Camera phones and paper-based microfluidic devices for real-time, off-site diagnosis. Analytical Chemistry, 80: 3699-3707.

O'Leary, J., D. Corcoran and B. Lucey, 2009. Comparison of the EntericBio Multiplex PCR System with Routine Culture for Detection of Bacterial Enteric Pathogens. Journal of Clinical Microbiology, 47: 3449-3453.

Okada, K., S. Chantaroj, T. Taniguchi, Y. Suzuki, A. Roobthaisong, O. Puiprom, T. Honda and P. Sawanpanyalert, 2010. A rapid, simple, and sensitive loop-mediated isothermal amplification method to detect toxigenic *Vibrio cholerae* in rectal swab samples. Diagnostic Microbiology and Infectious Disease, 66: 135-139.

Suo, B., Y.P. He, G. Paoli, A. Gehring, S.I. Tu and X.M. Shi, 2010. Development of an oligonucleotide-based microarray to detect multiple foodborne pathogens. Molecular and Cellular Probes, 24: 77-86.

Wang, S.J., 2008. Rapid and specific detection of enterotoxigenic *Escherichia coli* and *Salmonella* strains by multiplex PCR systems. Journal of Food and Drug Analysis, 16: 81-87.

Yamazaki, W., M. Taguchi, M. Ishibashi, M. Kitazato, M. Nukina, N. Misawa and K. Inoue, 2008. Development and evaluation of a loop-mediated isothermal amplification assay for rapid and simple detection of *Campylobacter jejuni* and *Campylobacter coli*. Journal of Medical Microbiology, 57: 444-451.

Yano, A., R. Ishimaru and R. Hujikata, 2007. Rapid and sensitive detection of heat-labile I and heat-stable I enterotoxin genes of enterotoxigenic *Escherichia coli* by Loop-Mediated Isothermal Amplification. Journal of Microbiological Methods, 68: 414-420.

You, Y.H., C.X. Fu, X. Zeng, D.S. Fang, X.M. Yan, B.C. Sun, D. Xiao and J.Z. Zhang, 2008. A novel DNA microarray for rapid diagnosis of enteropathogenic bacteria in stool specimens of patients with diarrhea. Journal of Microbiological Methods, 75: 566-571.

CHAPTER SEVEN

CONCLUSIONS AND ACCOMPLISHMENTS

Microbial diagnostics using NAAT has tremendous potential for addressing the global threat of infectious diseases because of its rapidity, high accuracy and suitability for multiplexed detection. However, due to its high cost and complexity, NAAT is currently available only in centralized laboratories in developed nations. Therefore, a need exists for devices that are low-cost and simple-to-use so that NAAT can also be used in low-resource settings, at point of care, or any other settings where a threat exist for human exposure to microbial pathogens. In this project, we leveraged the significant advances in the fields of genomics, low-cost microfluidics and molecular amplification techniques to develop such an instrument.

The importance of screening for multiple genetic markers was demonstrated in light of the extensive genetic diversity that exists among isolates of a given pathogens, both in terms of occurrence of virulence genes and their allelic variability. Using this information and an in-house database for hundreds of virulence genes, the suitability of a combined format of multiplex PCR and microarray hybridization was shown by parallel detection of multiple waterborne pathogens. While the cost of this detection method was, at the time of development, considered reasonably low on a per pathogen basis, more recent advances in low cost microfluidics have drastically changed this notion. Therefore, development of a microfluidic chip for pathogen detection was the logical next step for this project.

Isothermal techniques in general and LAMP in particular are much more promising for reducing the cost and complexicity of a device for NAAT than PCR, which remains the most widely used technique for NAAT. Aside from eliminating the need for thermal cycling, LAMP also provides a tremendous benefit for simplifying the optics necessary for real time monitoring of the reaction because of its high amplification yield. Furthermore, by careful selection of the fluorescent dye, a very low-cost sensor can be build that has identical performance in terms of rapidity and accuracy as an expensive real time PCR machine.

Towards the development of the low cost NAAT device, the following key tasks were accomplished:

- Developed a simple and versatile technique for rapid and inexpensive prototyping of polymer microfluidic chips using hot embossing, inexpensive molds with optimized design, lamination, and xurography;
- Designed and fabricated a 64-channel microfluidic chip for LAMP that is easy-to-use in the hands of untrained personnel in terms of sample dispensing and sealing of the chip;
- Developed an inexpensive and compact optical sensor for fluorogenic LAMP that employs a single photodiode to read 64 reaction wells, without the need for mechanically moving components;
- Designed LAMP primers for six bacteria linked to the majority of cases of infectious diarrhea in developing nations;

- Validated the performance of these primers in a fully functional prototype of the device.

Despite tremendous progress in academic laboratories, the number of commercially available devices for NAAT that employ microfluidic chips is very limited. The device developed here fills this gap. Furthermore, due to the simplicity of the chips and the robustness, low cost, and compactness of the instrument, it has significant potential to be used in a variety of settings, most notably in developing nations. By virtue of this accomplishment, this project addresses the need for 'development of technologies that allow assessment of individuals for multiple conditions or pathogens at point-of-care', which was identified as one of the Grand Challenges in Global Health by the Bill and Melinda Gates Foundation.