ADVANCES IN SINGLE-MOLECULE NUCLEIC ACID SEQUENCING

Thesis by

Christopher J. Lacenere

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2006

(Defended July 22, 2005)

Dedicated to the memory of Ann Romano, my grandma, my hero

Acknowledgements

First, I would like to sincerely thank Neil Garg, a former student in the Stoltz lab, for synthesizing the longer-linkered nucleotides used in much of this work. Neil is one of the most naturally talented and intense scientists that I have ever had the pleasure of knowing. Welcome to the world of the biologique—cheers.

Thanks the Ido Braslavsky for passing on all of his wisdom and teaching me how to sequence single molecules.

I'd also like to thank Hung Jin Lee and Brandon Birdwell for the construction of the microscope used in much of this work, as well as the Dervan lab for allowing me to use their Typhoon imager.

Thanks to my committee members, Dr. Ray Deshaies, Dr. Brian Stoltz, Dr. Michael Elowitz, and Dr. Stephen Quake, for their support and guidance. I'd especially like to thank Dr. Stephen Quake, my advisor, whose creativity has been an inspiration.

Thanks to my former adviser, Dr. Paul Sternberg, and former Sternberg lab member Tom Clandinin for their help and guidance.

I'll forever be grateful to Dean Michael Hoffman for his financial support.

Thanks to Dr. Costigan, Jessie, and the rest of his staff; and to Karen and all of the good people at Orthopedic Sports Care for getting me out of enough back pain to finish writing my thesis. I absolutely never could've finished this without you.

Thanks to my dear friend Dr. Padmaja Juvvadi for encouraging me to go back to school and finish my Ph.D.

Thanks to current and former members of the Quake Lab: Curly, for his help throughout the years, Josh (and Buster), for their laughter and for keeping it real, Fibonacci, for endless hours of wonderful conversation, and Little Flappy, for keeping me laughing. Cue the circus music.

Liz Ayala, Gwen Murdock, and Bill Lease—ya'll are great. Thanks for everything!

Finally, last but not least, a big, big, thanks to Demo, my closest friend here at Caltech, for his friendship, generosity, and support with more things, both scientific and personal, than I could ever name here. I don't know how I will ever repay him.

What a long, strange trip it's been…

Abstract

The ability to quickly and accurately obtain sequence information from single molecules of DNA and RNA has far-reaching implications for our understanding of biology. In the work presented here, we have made several advances in the area of single-molecule DNA and RNA sequencing. Specifically, in attempting to increase the read length of DNA polymerase, we have assayed several custom synthesized fluorescent nucleotides containing longer dye–base linkers. We have validated the efficacy of these nucleotides at both bulk and single-molecule levels. Furthermore, we have screened several commercially available DNA polymerases for their ability to incorporate these nucleotides. We also show that reverse transcriptase is able to synthesize a complimentary DNA strand of 28 bases in length from an RNA template, using solely fluorescently labeled nucleotides. Additionally, we show that reverse transcriptase is able to incorporate a fluorescently labeled nucleotide into an RNA template at the single-molecule level. Finally, we demonstrate automated reagent exchange for our single-molecule sequencing system.

Table of Contents

List of Figures

1

Introduction

1.1 Introduction

1.1.1 What Is the Sequence of My Genome?

Towards the later half of the 20th century, scientists from around the world rallied behind a single unified goal: to obtain the entire sequence of the human genome (DAVIS 1990). "What is the sequence of *our* genome?" they asked, referring to the species *Homo sapiens*. In 2001, roughly four years ahead of schedule, this goal was realized at a cost of approximately $3 billion (LANDER *et al.* 2001; VENTER *et al.* 2001). The completion of the Human Genome Project ahead of schedule was, in part, due to technological advances in sequencing protocols, computational algorithms, miniaturization, and automation. However, another aspect of its rapid expedition was due to the open culture amongst scientists involved in the project. Data was freely exchanged between groups almost instantaneously, and improvements and refinements to sequencing protocols were openly shared (Figure 1.1).

Only five years into the 21st century, we are asking a different question:

"What is the sequence of *my* genome?"

What was obtained by the Human Genome Project was a consensus sequence—an averaged approximation of the sequence of DNA samples taken from many subjects. The consensus human genome sequence was an essential undertaking in order to understand what it is that makes us, *Homo sapiens*, human. But with that in hand, we now want to

understand what makes each and every one of us, as individuals, unique. Obtaining the sequence of one's own DNA, a personal genome, has broad implications with respect to medical care, including disease prevention and customized pharmaceutical treatment.

The push for personal genomes is, for the most part, limited by cost. Sequencing costs have gone down drastically since the beginning of the Human Genome Project. The current cost of having an entire human genome sequenced runs in the millions of dollars; a price tag far higher than the average person can afford.

Most of the technological advances in sequencing over the past 25 years have involved streamlining the Sanger sequencing method (SANGER *et al.* 1977) rather than inventing entirely new sequencing methods. Although sequencing costs continue to fall, it is likely that an entirely new sequencing technology is needed in order to bring the cost down to the NIH goal of $1000 per genome (http://grants1.nih.gov/grants/guide/rfa-files/RFA-HG-04-003.html).

Here we describe advances in an entirely new sequencing technology: single-molecule sequencing (SMS). SMS offers many potential advantages over current sequencing technologies, including speed, miniaturization, and sparing use of reagents. All of these advantages ultimately translate into reduced costs, edging us closer to the goal of a $1000 genome.

A description of specific aspects and principles of SMS follows.

1.2 Overview

SMS experiments performed in the Quake lab provided the first demonstration that sequence information could be obtained from single molecules of DNA (BRASLAVSKY *et al.* 2003). Rather than relying on chain termination using dideoxynucleotides, this entirely new sequencing technique follows the sequencing-by-synthesis paradigm. Sequencing-by-synthesis methodologies obtain sequence information from template DNA as each nucleotide is added by the DNA polymerase. In the system described below, the incorporation of fluorescently labeled nucleotides by DNA polymerase into a template DNA strand is observed using total internal reflection microscopy (Figure 1.2).

1.2.1 Minimizing Background Fluorescence

In order to obtain single-molecule sensitivity it is absolutely essential to keep background fluorescence to a minimum. A noisy system, high in background fluorescence, will easily swamp out the fluorescent signal of individual molecules. Several factors have been taken into consideration to minimize background fluorescence in our system. First, by using a total internal reflection microscope, an evanescent wave illuminates only approximately 150 nm above the coverslip surface. Second, a highly negatively charged surface is created on the coverslip using polyelectrolyte multilayering (KARTALOV *et al.* 2003). Polyelectrolyte multilayering consists of alternately applying a positive polyelectrolyte, followed by a negative polyelectrolyte, numerous times to the coverslip surface. When finished, the surface is much more negatively charged than silica alone, resulting in fluorescently labeled nucleotides being repelled, since they too are negatively

charged. Furthermore, this highly negatively charged surface prevents template DNA (also negatively charged) from lying down onto the glass.

1.2.2 Sequencing Chamber and Template DNA

The sequencing chamber consists of a flexible plastic adhesive flow cell, less than a millimeter in height, which has been applied on the top of a low-autofluorescence microscope coverslip. The sequencing chamber holds approximately 80 μL of fluid. Template DNA is anchored to the surface via the strong binding properties of biotin and streptavidin. Biotin, a modified version containing an amine group, is first covalently bonded to the carboxyl groups of polyacrylic acid, the negatively charged polyelectrolyte, using the catalyst EDC. Next, streptavidin is bound to biotin on the coverslip surface. Template DNA, which has been biotinylated at its 5′ end, is finally applied onto the streptavidin. Thus, streptavidin acts as a bridge between the surface biotin and the biotin of the template DNA. The template DNA has previously been annealed to a Cy3-labeled primer.

1.2.3 Imaging the Template DNA

Once the template DNA is tethered to the sequencing chamber surface, one may begin sequencing. First, the sample is illuminated using a green laser, and an image is acquired using a Cy3 emission filter. The features, or spots, seen in this image result from the fluorescence of the Cy3-labeled primer annealed to the template DNA. Oxygen scavenger is used while acquiring this image, allowing each template's fluorescent signal

to persist long enough for image acquisition. Acquired images are 16-bit grayscale with a typical resolution of 2048 × 2048 pixels, representing approximately a 200 μ × 200 μ field of view.

1.2.4 Imaging Incorporations

In the next step, the fluorescent signal of the template DNA is photobleached by illuminating the sample with the green laser in the absence of oxygen scavenger. It is important to note that these features bleach in a single step, supporting our claim that these are indeed single molecules. DNA polymerase along with a fluorescently labeled nucleotide is added next, and allowed to incubate for a short period. After a series of washing steps, another image is acquired using a red laser (we will assume that the nucleotide is labeled with Cy5). This entire process may be repeated any number of times.

1.2.5 Image Processing (Figure 1.3)

After obtaining the images of template DNA and of subsequent incorporation events, we next determine which molecules of template DNA have had a fluorescent nucleotide incorporated into them. This is done using custom software routines written in the Interactive Data Language (IDL). These routines may be found in Appendix I. Brief descriptions of their applications follow.

The images are first denoised using a band-pass wavelet noise reduction routine. This is followed by the identification of circular features, presumably corresponding to our template DNA in one image, and corresponding to subsequent incorporations into the template in another image. After these features are identified, an algorithm approximates the center of each feature. At this point, we are now working with coordinates in two-dimensional space (with each pair of coordinates, or point, representing a feature center), rather than a complex 16-bit grayscale image.

In the final step of data processing, the second set of coordinates is shifted over the first set of coordinates, along both the x and y axes. At each step, the number of correlating features is noted, resulting in a correlogram. A correlogram is a visual representation of the best alignment of the two images. After aligning the images, correlated features are inferred to be template DNA molecules that have had a fluorescent nucleotide incorporated into them by DNA polymerase. Based on Watson-Crick base pairing rules, and because we are adding one nucleotide into the sequencing chamber at a time, we can logically assume the identity of the first base following the primer in these template DNA molecules. Theoretically, the entire sequence of the template DNA can be determined by sequentially repeating this process with all four nucleotides.

1.3 Concluding Remarks

To summarize, we have developed an entirely new DNA sequencing technology that may potentially lead to the realization of the $1000 genome. The intrinsic small scale of single-molecule experimentation consumes only a sparing amount of reagents, several

orders of magnitude less than current DNA sequencing methodology. Perhaps the most significant advantage of our SMS system is its unmatched level of parallelism. A 1″ × 1″ area is capable of holding approximately 12 million template DNA molecules. We estimate that one microscope, running continuously, will be able to obtain five billion raw base pairs per day. In order to get 10X coverage of the human genome, consisting of approximately three billion base pairs, we estimate that this SMS platform will need to run for 24 days straight, with a total cost of approximately $15,000; orders of magnitude less than the millions of dollars price tag that current sequencing methodology can offer.

Many obstacles still exist, however, in turning this technology into a practical one. For instance, one of the biggest problems is read length. We are currently able to obtain a sequence of no more than five base pairs in length. Furthermore, given the calculations above with respect to the human genome, the automation of reagent exchange steps is necessary. In the chapters that follow, several advances and improvements in these areas are presented.

1.4 Figure Legends

Figure 1.1

1.4.1 Computing and Sequencing Growth

From SHENDURE *et al.* 2004. The dark blue line shows the doubling of computer instructions per second per U.S. dollar that has been occurring approximately every 18 months since 1900 in accordance with Moore's law. The magenta line shows similar growth in the number of base pairs of accurate DNA sequence per US dollar. To some extent, this growth mirrors Moore's law because of the sequencing technologies' dependence upon computers. The orange line depicts the growth of web sites (doubling time for months), and shows how quickly a technology can spread within a given infrastructure. The turquoise line depicts the growth of fluorescent in situ sequencing with polonies (MITRA *et al.* 2003) in base pairs per minute and has a doubling time of one month.

Figure 1.2

1.4.2 Single-Molecule Sequencing and Image Acquisition

The microscope used for single-molecule sequencing is in the center of the figure. It is a homemade, through-the-objective, total internal reflection apparatus equipped with two lasers, red and green, shown on the left. Since the microscope employs total internal reflection, the source laser light reflects entirely, resulting in an evanescent wave that

illuminates only the area just above the coverslip's surface (~150 nm), as depicted in the zoomed area in the upper left. The sequencing chamber consists of a flexible adhesive flow cell (~80 µl) adhered to a low-autofluorescence coverslip. DNA samples are anchored to the coverslip surface via biotin and streptavidin as shown in the zoomed area in the upper right. The template DNA's primer is labeled with Cy3. Using the green laser, an image of these template DNA molecules is first acquired (Image 1). After image acquisition, the primers' fluorescence is photobleached. DNA polymerase, along with a fluorescent nucleotide, is added next, and after a short incubation, a second image is acquired (Image 2), which represents molecules that have potentially had a fluorescent nucleotide incorporated into them.

Figure 1.3

1.4.3 Image Processing

Acquired images of single molecules of DNA are first denoised using a band-pass wavelet denoising algorithm (bpass.pro, Appendix I). Denoised images are shown in step 1. Next, features are identified, as shown in step 2 (feature.pro, Appendix I). A two-dimensional coordinate representing the center of each feature is determined (step 3). Finally, the second images is shifted over the first, and the number of correlations at each step of the shift is noted (step 4, compare_pts5.pro, Appendix I), producing a correlogram (step 5).

Figure 1.1

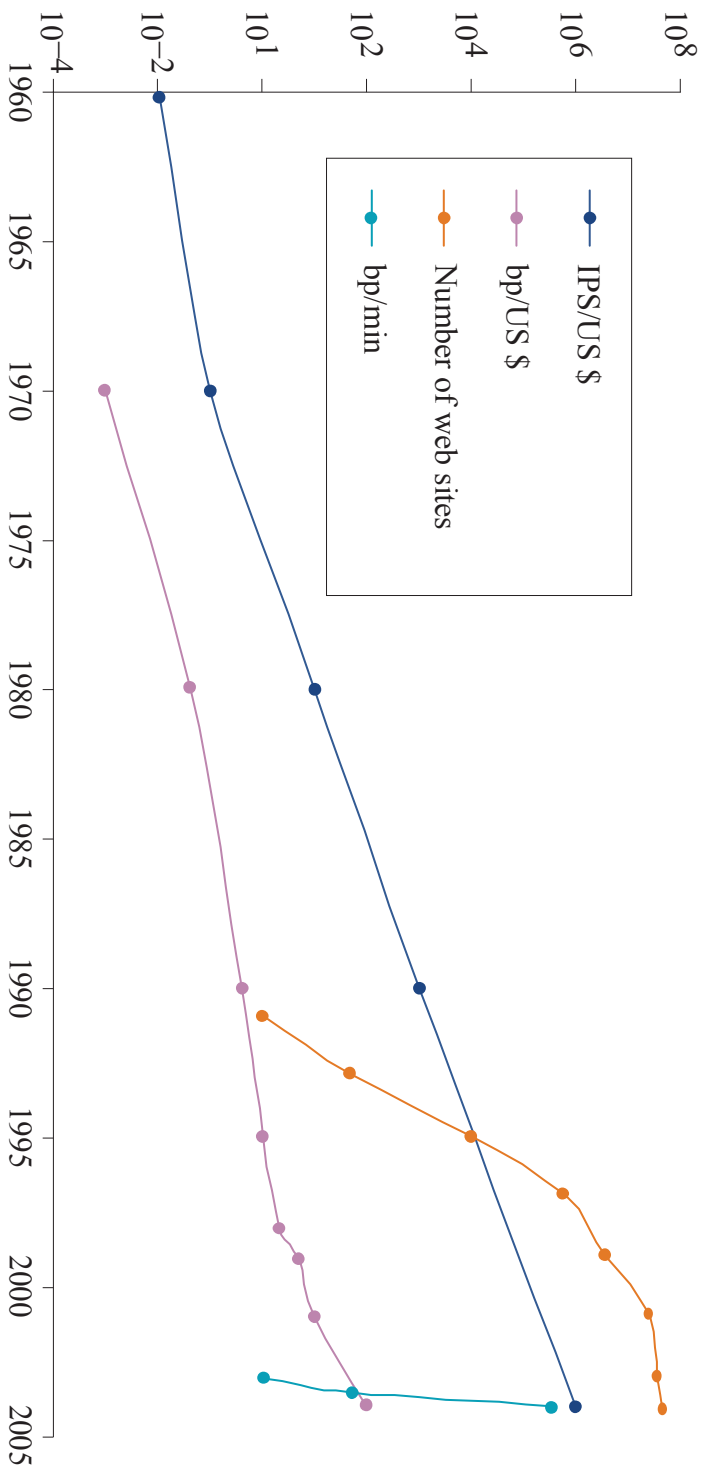

10^8
10^6
10^4
10^2
10^1
10^{-2}
10^{-4}
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
IPS/US $
bp/US $
Number of web sites
bp/min

Figure 1.2

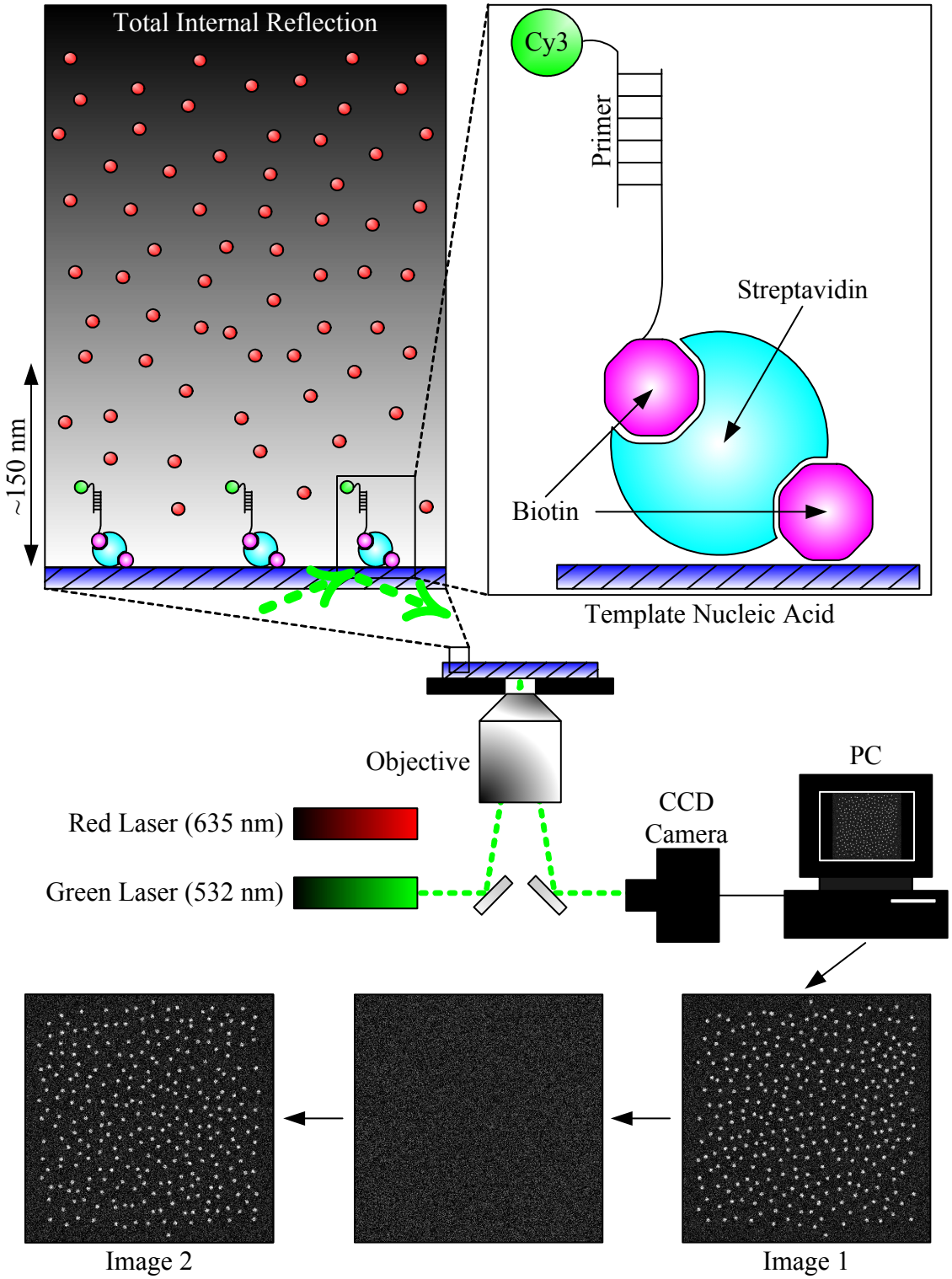
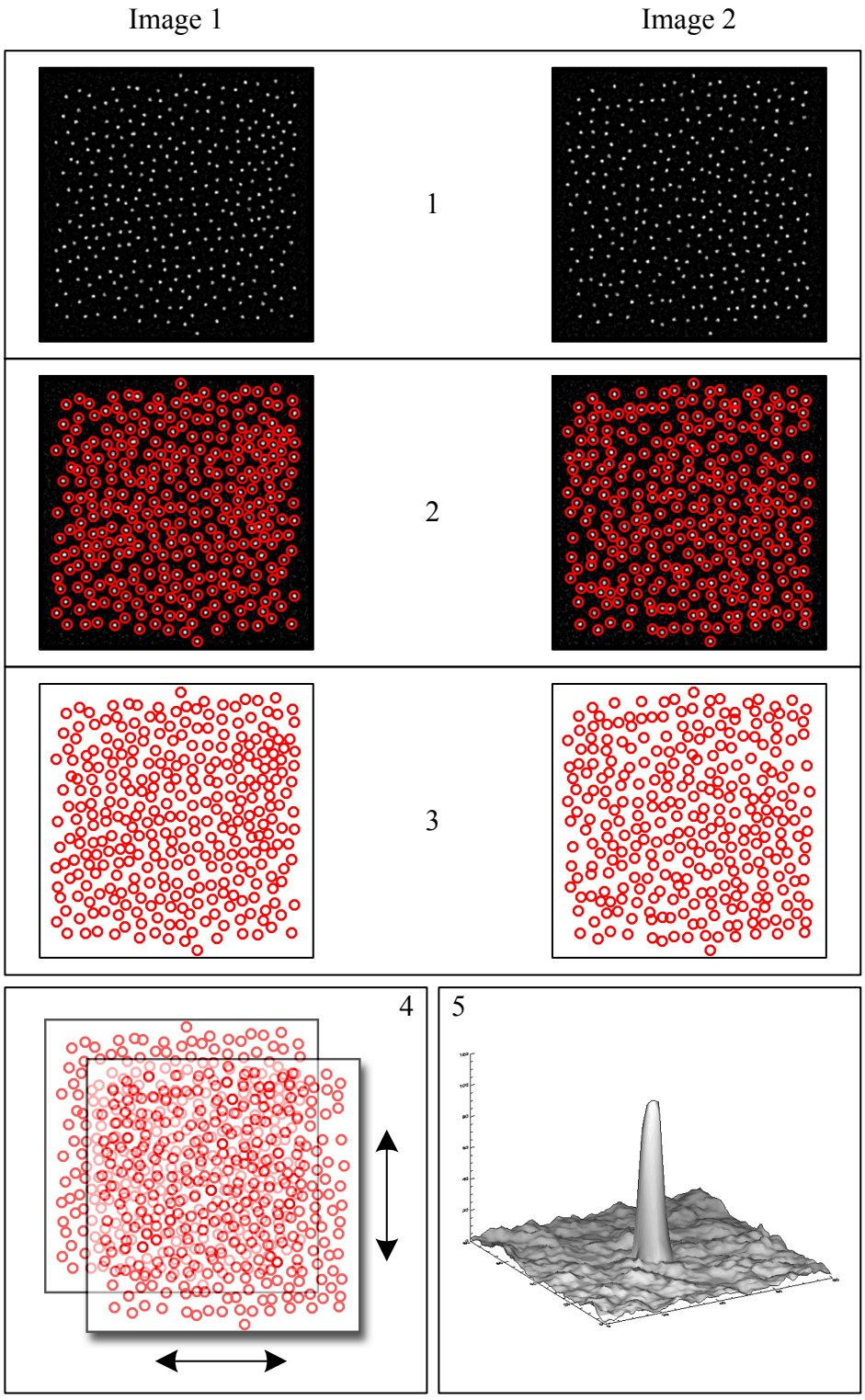

Total Internal Reflection
~150 nm
Cy3
Primer
Streptavidin
Biotin
Template Nucleic Acid
Objective
Red Laser (635 nm)
Green Laser (532 nm)
CCD
Camera
PC
Image 2
Image 1

Figure 1.3



1.5 References

BRASLAVSKY, I., B. HEBERT, E. KARTALOV and S. R. QUAKE, 2003. Sequence information can be obtained from single DNA molecules. Proc Natl Acad Sci USA **100:** 3960–3964.

DAVIS, B. D., 1990. The human genome and other initiatives. Science **249:** 342–343.

KARTALOV, E. P., M. A. UNGER and S. R. QUAKE, 2003 Polyelectrolyte surface interface for single-molecule fluorescence studies of DNA polymerase. Biotechniques **34:** 505–510.

LANDER, E. S., L. M. LINTON, B. BIRREN, C. NUSBAUM, M. C. ZODY *et al.*, 2001. Initial sequencing and analysis of the human genome. Nature **409:** 860–921.

MITRA, R. D., J. SHENDURE, J. OLEJNIK, EDYTA-KRZYMANSKA-OLEJNIK and G. M. CHURCH, 2003. Fluorescent in situ sequencing on polymerase colonies. Anal Biochem **320:** 55–65.

SANGER, F., S. NICKLEN and A. R. COULSON, 1977. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA **74:** 5463–5467.

SHENDURE, J., R. D. MITRA, C. VARMA and G. M. CHURCH, 2004. Advanced Sequencing Technologies: Methods and Goals. Nature Reviews Genetics **5:** 335–344.

VENTER, J. C., M. D. ADAMS, E. W. MYERS, P. W. LI, R. J. MURAL *et al.*, 2001. The sequence of the human genome. Science **291:** 1304–1351.

2

Custom Nucleotides: Longer-Linkered Derivatives

2.1 Introduction

Sequencing-by-synthesis, a paradigm by which sequence information is obtained from template DNA by identifying incorporated bases sequentially as they are added by the polymerase, offers a promising alternative to standard dideoxy sequencing (SANGER *et al.* 1977). However, when fluorescently labeled nucleotides are used, read length is limited. This limited read length is presumably due to steric hindrance effects between the relatively bulky dye molecules and DNA polymerase, dye-dye interactions, or both.

Using a single-molecule detection strategy, we have been able to obtain sequence information from individual DNA molecules (BRASLAVSKY *et al.* 2003). However, our read length is limited to approximately 5 bases.

2.1.1 Cleavable Nucleotides

Modified nucleotides offer an attractive solution to the read length problem. For instance, nucleotides containing a cleavable linker arm between the dye and the nucleobase may be cleaved after incorporation, resulting in a less sterically encumbered system. Three main classes of cleavable nucleotides are diagrammed in Figure 2.1.

Theoretically, the cleavage of the linker arm may be triggered by a wide variety of catalysts including chemicals, pH, enzymes, or light. Recently, three photocleavable fluorescently labeled nucleotides (BAI *et al.* 2004; LI *et al.* 2003; SEO *et al.* 2004) were used to identify 7 bases in a DNA template (SEO *et al.* 2005). These results demonstrated

that these photocleavable nucleotide analogues could be incorporated accurately into a growing DNA strand and that the fluorophore was able to be efficiently cleaved using near-UV irradiation, allowing the continuous identification of the template sequence.

A cleavable nucleotide variant, termed a reversible terminator, offers, perhaps, an ideal solution to the read length problem by placing the fluorescent dye off of the 3′ hydroxyl group of the nucleotide's sugar. Such a nucleotide offers other benefits as well, namely, the ability to sequence repeat regions of the genome. Since only one nucleotide at a time is able to be incorporated, subsequent incorporations are only possible after the 3′ linker is cleaved, releasing the dye. It is essential that, after cleavage, the 3′ hydroxyl group is returned so that DNA polymerase may add the next nucleotide.

Another attractive property of reversible terminators is that all four nucleotides may be added together into the sequencing chamber if each nucleotide is labeled with a unique fluorescent dye. As described above, only one nucleotide at a time is able to be incorporated and subsequent incorporation events are contingent upon exposure of the 3′ hydroxyl group.

Seminal work by Sarfati (CANARD *et al.* 1995; CANARD and SARFATI 1994; RASOLONJATOVO and SARFATI 1999) using reversible terminators with chemically or enzymatically cleavable linkers resulted in premature cleavage of the linker by DNA polymerase. Premature cleavage was shown to depend on the sequence of the template DNA. Furthermore, this effect was observed using various DNA polymerases.

More recently, a hybrid reversible terminator was constructed containing a photocleavable fluorescent dye stemming from the nucleobase, and a chemically removable allyl group stemming from the 3′ hydroxyl position of the sugar (RUPAREL *et al.* 2005). It was reasoned that the relatively small allyl group would be better tolerated by DNA polymerase. This hybrid reversible terminator was shown to be a good substrate for DNA polymerase, being incorporated and cleaved successfully. Furthermore, removal of the allyl group initiated DNA synthesis as expected. Further work is still needed in this area.

2.1.2 Longer-Linkered Nucleotides

Before exploring the possibility of using cleavable nucleotides in our single-molecule sequencing system, we first wanted to test longer-linkered nucleotide variants. Our reasoning was that using a cleavable nucleotide could potentially add at least one extra reagent exchange step to our protocol. Furthermore, much less is known regarding the fidelity and tolerance of these cleavable nucleotides.

PCR studies by Waggoner (ZHU *et al.* 1994) revealed that longer-linkered Cy3-labeled nucleotides were apt to be present in the PCR product in greater number than shorter linkered Cy3-labeled nucleotides. However, when longer-linkered nucleotides were used, the overall yield of the PCR reaction was reduced. The authors conclude that PCR yield, and, thus, tolerance by DNA polymerase, is inversely proportional to the length of the nucleotide's linker.

In collaboration with the Stoltz group, here at Caltech, we have synthesized two longer-linkered variants of Cy5-dUTP. The results of our validation experiments involving these longer-linkered nucleotides is presented below.

2.2 Results and Discussion

2.2.1 Longer-Linkered Cy5-dUTP Derivatives

Three variants of Cy5-dUTP are shown in Figure 2.2. Nucleotide A is identical to commercially available Cy5-dUTP, and in order to avoid confusion regarding linker length, it will be referred to henceforth as Cy5-10-dUTP. Nucleotide B is a longer-linkered variant of Cy5-dUTP, henceforth referred to as Cy5-17-dUTP, in which the linker arm between Cy5 and uracil has been extended by seven atoms. Nucleotide C is also a longer-linkered variant of Cy5-dUTP, henceforth referred to as Cy5-24-dUTP, which has had its linker arm extended by an additional 14 atoms.

2.2.2 Bulk Nucleotide Validation

We first sought to validate the efficacy of the synthesized nucleotides by performing a primer extension assay. The results of this assay are shown in Figure 2.3. Briefly, a primed poly(A) DNA template was extended using DNA polymerase and the appropriate nucleotide. After an incubation period, the reactions were run through a gel filtration column and imaged for Cy5 in a clear-bottomed 96-well plate. We reasoned that if an

incorporation of a Cy5 nucleotide had indeed occurred, we would be able to detect Cy5 fluorescence after the reactions had been run through the gel filtration column since the column retains small molecules, such as unincorporated nucleotides. However, larger molecules, such as the template DNA, pass through the column. These results show that with regard to incorporation into our DNA template, the synthesized Cy5-10-dUTP and Cy5-17-dUTP are indistinguishable from commercially available Cy5-dUTP. Furthermore, the incorporation of these nucleotides is polymerase dependent as well as primer and template dependent, as one would expect.

2.2.3 Single-Molecule Nucleotide Validation

We next sought to validate the efficacy of the longer-linkered Cy5-dUTP derivatives at the single-molecule level. The following set of experiments addresses the efficacy of these longer-linkered nucleotides under the experimental conditions used during single-molecule DNA sequencing. The same template used in the bulk primer extension assay described in Figure 2.3 was used for the single-molecule experiments presented below.

The results of an experiment designed to test Cy5-17-dUTP incorporation under single-molecule sequencing conditions is shown in Figure 2.4. A maximum of 77 matches were found in one position. This, represented by the obvious peak displayed in the center of the correlogram, strongly suggests that Cy5-17-dUTP is capable of being incorporated by DNA polymerase into the surface-bound template DNA of the sequencing chamber under the experimental conditions required for single-molecule sequencing.

The resulting correlogram from an identical experiment using Cy5-24-dUTP instead of Cy5-17-dUTP is shown in Figure 2.5. A maximum of 16 matches were found at one position. The peak within this correlogram, while apparent, is markedly smaller than the peak observed for Cy5-17-dUTP in Figure 2.4. This suggests that Cy5-24-dUTP is also capable of being incorporated into the surface-bound template DNA of the sequencing chamber under the experimental conditions necessary for single-molecule DNA sequencing.

The negative control for our tests of incorporation at the single-molecule level is shown in Figure 2.6. This experiment was identical to the two single-molecule experiments described above, however, the incorrect nucleotide, Cy5-dCTP, was used instead of Cy5-dUTP. A maximum of seven matches were found at two positions in the correlogram. Since a peak is not apparent in this correlogram, and since a maximum number of matches were found at multiple positions, these seven matches likely represent chance correlations of nonspecific binding of Cy5-dCTP to the sequencing chamber surface.

Because a poly(A) template was used in each of these single-molecule experiments, it is unknown whether these correlations represent a single or multiple incorporations.

In conclusion, these results suggest that the two longer-linkered variants of Cy5-dUTP, Cy5-17-dUTP and Cy5-24-dUTP, are able to be incorporated by DNA polymerase using our single-molecule DNA sequencing scheme. Furthermore, it appears that as the

nucleotide linker length increases, the efficiency of incorporation under these conditions decreases.

2.3 Concluding Remarks

To summarize, we set out to increase the read length of DNA polymerase using custom synthesized nucleotides in which the linker between the fluorescent dye and the nucleobase was extended. In collaboration with the Stoltz group here Caltech, several longer-linkered variants of Cy5-dUTP were synthesized. Before measuring read length, we first tested the purity and efficacy of these nucleotides using various incorporation assays at both bulk and single-molecule levels. The results presented here suggest that the longer-linkered nucleotides are able to be successfully incorporated by DNA polymerase at each of these levels. Based on our single-molecule experiments, it appears that incorporation efficiency is inversely proportional to linker length. Further work examining Cy5-24-dUTP's effect on read length is presented in Chapter 3.

2.4 Materials and Methods

2.4.1 Nucleotides

dTTP was purchased from Roche (Indianapolis, IN). Cy5-dUTP and Cy5-dCTP were purchased from Amersham Biosciences (Piscataway, NJ). Cy5-17-dUTP and Cy5-24-dUTP were prepared in a manner similar to the protocol developed by Waggoner (ZHU *et al.* 1994) as follows: Cy5-10-OSu was coupled to 6-amino-hexanoic acid (MIDURA-NOWACZEK *et al.* 1995) under standard conditions to produce Cy5-17-OH. After activation to the succinimidyl ester, Cy5-17-OSu was reacted with AP-dUTP to afford Cy5-17-dUTP. Cy5-10-OSu was coupled to 6-(6-amino-hexanoylamino)-hexanoic acid under standard conditions to produce Cy5-24-OH. After activation to the succinimidyl ester, Cy5-24-OSu was reacted with AP-dUTP to afford Cy5-24-dUTP. ESI-MS: [M-1] 1384.3, [M+4Na] 1473.3.

2.4.2 Template Preparation

Synthetic DNA oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA). Annealing of primer with template was performed by mixing 3 nmol of Cy3-labeled DNA primer (5′-Cy3-GTCTGGGCTTTTGGTTTGTGGG-3′) with 3 nmol of biotinylated template DNA (5′-Biotin-$[A]_{30}$CCCACAAACCAAAAGCCCAGAC-3′) in 50 µl annealing buffer (150 mM NaCl, Tris-HCl, pH 7.2), heating the mixture for 5 minutes at 100°C, and cooling to room temperature over 1 hour. Annealed duplex was then purified using a P30 gel filtration column (BioRad).

2.4.3 Primer Extension

Primer extension reactions were performed using 15 pmol of annealed duplex in 50 µl reaction buffer (Tris-HCl, pH 7.5, 5 mM $MgCl_2$, 12.5 mM dithiothreitol) containing 25 µM of the appropriate nucleotide, and 1 unit of Klenow exonuclease minus DNA polymerase. Reactions were allowed to incubate for 2 hours at 37°C after which they were stopped by the addition of 2 µl 100 mM EDTA. Excess nucleotides were removed from each reaction using a P30 gel filtration column (BioRad).

2.4.4 Product Analysis

Primer extension reaction products were initially diluted 1 to 4 in high-purity H_20, followed by three 1 to 10 serial dilutions. 100 µl was transferred to a clear-bottom 96-well plate and imaged on a Typhoon 8600 variable mode imager (Amersham Biosciences) at normal sensitivity and 200 micron resolution. The focal plane of the imager was set to 3 mm above the scanner bed. Cy5 fluorescence was detected using the red (633 nm) laser (PMT setting at 800 V) as excitation source and a 670 nm BP 30 emission filter. Images were processed and denoised using Typhoon's IQ Solutions software package.

2.4.5 Single-Molecule Detection and Data Analysis

A homemade microscope configured for through-the-objective, total internal reflection (TIR) illumination served as a platform for the single-molecule experiments. Two laser beams, 635 nm (Hitachi) and 532 nm (Intellite), with nominal powers of 30 and 50 mW,

respectively, were circularly polarized by quarter-wave plates. The laser beams were focused on the back aperture of the objective (PlanApo, 60X NA 1.45 oil, Olympus) to create an epi-illuminated area of 200 × 200 μm. TIR was achieved by displacing the beams from the objective's center, toward the objective's edge with a mirror mounted to a translation stage. The relatively large numerical aperture of the objective allowed the beams to be focused at angles greater than the critical angle (~62.5°) at the cover glass/water interface, allowing the evanescent wave to extend ~150 nm into the specimen chamber. The specimen chamber made from a low-autofluorescence microscope coverslip (Schott) and an adhesive hybridization well (Sigma). An image splitter (Optical Insights) directed the light through two band-pass filters (630dcxr, HQ585/80, HQ690/60; Chroma Technology) to a scientific-grade charge-coupled device camera (ORCA-ER, Hamamatsu), which recorded adjacent images of a 200 × 100 μm section of the surface in two colors. Typically, several exposures of 0.5 sec each were taken of each field of view to compensate for possible intermittency in the fluorophore emission. Custom IDL software (Crocker and Grier 1996) was modified to analyze the locations and intensities of fluorescence objects in the intensified charge-coupled device pictures. We inspected the resulting traces to determine incorporation events on the primed template sequences.

2.4.6 Single-Molecule Coverslip and Sequencing Chamber Preparation

Surface chemistry based on polyelectrolytes (DECHER 1997; KARTALOV *et al.* 2003) and biotin-streptavidin bonding was used to anchor the DNA molecules to the coverslip surface of the hybridization chamber and to minimize nonspecific binding of the

nucleotides to the surface. Coverslips were sonicated in 2% MICRO-90 soap (Cole–Parmer) for 20 minutes and then cleaned by immersion in boiling RCA solution (6:4:1 high-purity H_2O/30% NH_4OH/30% H_2O_2) for 1 hour (UNGER *et al.* 1999). They were then immersed alternately in polyallylamine (positively charged) and polyacrylic acid (negatively charged; both from Aldrich) at 2 mg/ml and pH 8.0 for 10 minutes each and washed intensively with high-purity water in between. The carboxyl groups of the last polyacrylic acid layer served to prevent the negatively charged labeled nucleotide from binding nonspecifically to the surface. In addition, these functional groups were used for further attachment of a layer of biotin. The coverslips were incubated with 5 mM biotin-amine reagent (Biotin-EZ-Link, Pierce) for 10 minutes in the presence of 1-[3-(dimethylamino)propyl]-3-ethylcarbodiimide hydrochloride (EDC, Sigma) in MES buffer, pH 5.5, followed by incubation with streptavidin (Neutravidin, Pierce) at 0.1 mg/ml for 15 min in Tris buffer, pH 8.0. The biotinylated DNA template was deposited onto the streptavidin-coated chamber surface at 25 pM for 10 minutes in Tris buffer, pH 8.0. For incorporations, the reaction solution contained Klenow exonuclease minus DNA polymerase (New England Biolabs) at 100 units/ml in the reaction buffer (EcoPol reaction buffer, New England Biolabs) and Cy5-dUTP derivatives at 0.5 μM. Reaction incubation times were 10 minutes. To reduce bleaching of the fluorescence dyes, an oxygen scavenging system (YILDIZ *et al.* 2003) was used during all illumination periods, with the exception of the bleaching of the primer's Cy3 tag.

2.4.7 Reagent Exchange Sequence

The positions of the anchored Cy3-primed DNA were recorded, and then the tags were

bleached by green laser illumination. Cy5-dUTP derivative and DNA polymerase were introduced and washed out. An image of the surface was then analyzed for incorporated Cy5-dUTP derivative using red illumination.

2.5 Figure Legends

Figure 2.1

2.5.1 Cleavable Nucleotides

Three broad classes of cleavable nucleotides. A) Class I, in which the linker between the dye and the nucleobase is cleavable. B) Class II, a reversible terminator, in which linker between the dye and the 3′ position of the nucleotide's sugar is cleavable. Upon cleavage, the 3′ OH group is restored, allowing for the next incorporation event. C) Class III, a reversible terminator, in which the linker between the dye and the nucleobase, and the linker between a protecting group and the 3′ position of the nucleotide's sugar, is cleavable. Upon cleavage, the 3′ OH group is restored, allowing for the next incorporation event. Each linker may be induced to cleave by either the same factor or by different factors, depending on the design of the cleavable moiety.

Figure 2.2

2.5.2 Longer-Linkered Cy5-dUTP Derivatives

Three synthesized Cy5-dUTP variants as follows: A) Cy5-10-dUTP, a commercially available standard cy5-dUTP derivative. B) Cy5-17-dUTP, a longer-linkered Cy5-dUTP derivative in which the arm joining the dye to the nucleotide has been extended by 7 atoms. C) Cy5-24-dUTP, a longer-linkered Cy5-dUTP derivative in which the arm joining the dye to the nucleotide has been extended by 14 atoms. Longer-linkered dye-

labeled nucleotide derivatives may increase read length by potentially reducing steric hindrance of the bulky dye side chains on adjacent nucleotides and/or DNA polymerase.

Figure 2.3

2.5.3 Validation of Synthetic dUTP Derivatives

Bulk primer extension assay of dTTP (I), commercially available Cy5-10-dUTP (II), synthesized Cy5-10-dUTP (III), and Cy5-17-dUTP (IV). Prior to imaging, all reactions were run through a P30 gel filtration column allowing detection of fluorescent signal only if the nucleotide had been incorporated into the relatively large template DNA molecule. Extension reactions are shown in columns numbered 1, negative controls lacking polymerase are shown in columns numbered 2, negative controls lacking DNA template are shown in columns numbered 3, and negative controls lacking primer are shown in columns numbered 4. Rows A–D represent 1 to 10 serial dilutions of the reactions, respectively. The polymerase dependent and template dependent fluorescence signals strongly suggest that the synthesized nucleotides are capable of acting as substrate for DNA polymerase. Further support is lent by the fact that commercially available Cy5-10-dUTP is nearly indistinguishable from the synthesized derivatives.

Figure 2.4

2.5.4 Cy5-17-dUTP Correlogram

Correlogram of the incorporation of Cy5-17-dUTP into the first position of a primed

DNA template by exonuclease deficient Klenow fragment DNA polymerase. A maximum of 77 matches were found at one position with an average vector shift of 2.500000 pixels along the x-axis and -18.812500 pixels along the y-axis. This initial shift has been compensated for in the correlogram.

Figure 2.5

2.5.5 Cy5-24-dUTP Correlogram

Correlogram of the incorporation of Cy5-24-dUTP into the first position of a primed DNA template by exonuclease deficient Klenow fragment DNA polymerase. A maximum of 16 matches were found at one position with an average vector shift of -7.812500 pixels along the x-axis and 21.812500 pixels along the y-axis. This initial shift has been compensated for in the correlogram.

Figure 2.6

2.5.6 Negative Control Cy5-dCTP Correlogram

Negative control correlogram of the incorporation of Cy5-dCTP, the incorrect nucleotide, into the first position of a primed DNA template by exonuclease deficient Klenow fragment DNA polymerase. A maximum of seven matches were found at two positions with an average vector shift of -8.0000000 pixels along the x-axis and 21.750000 pixels along the y-axis.

Figure 2.1

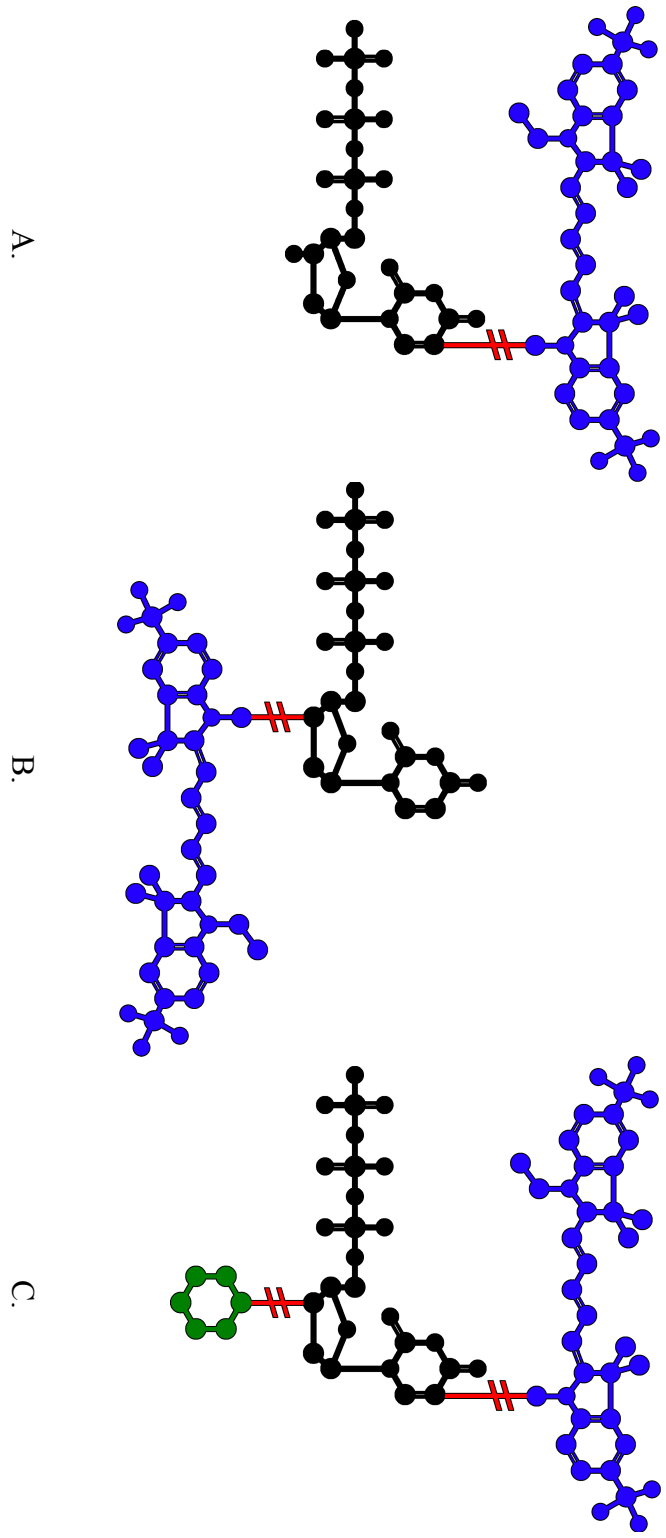


Figure 2.2

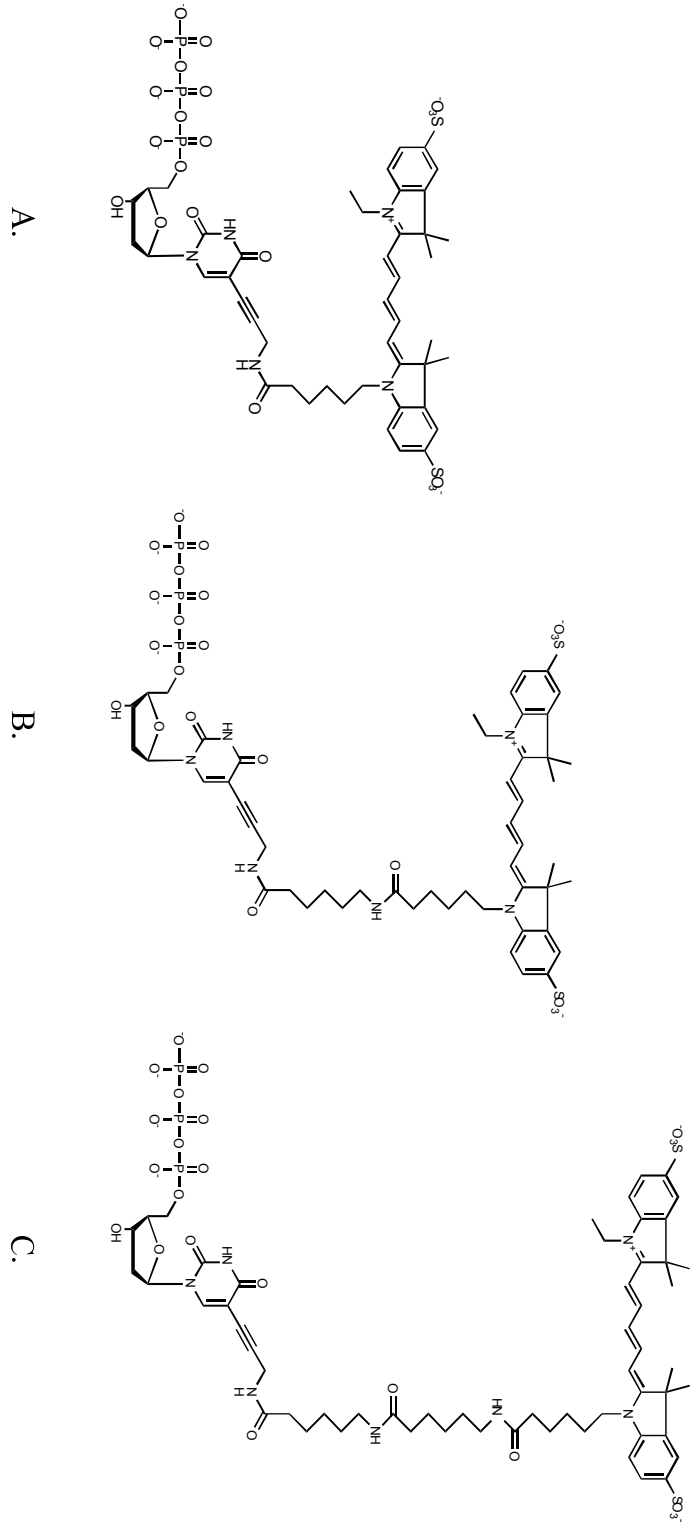


Figure 2.3

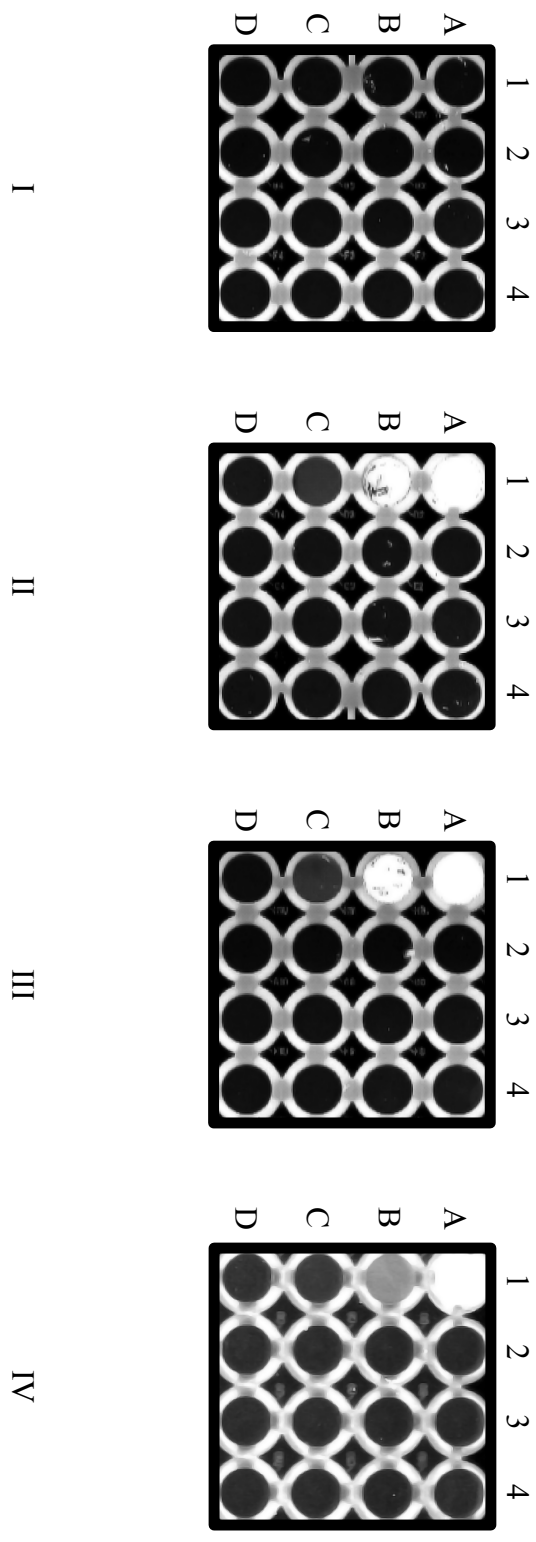


Figure 2.4

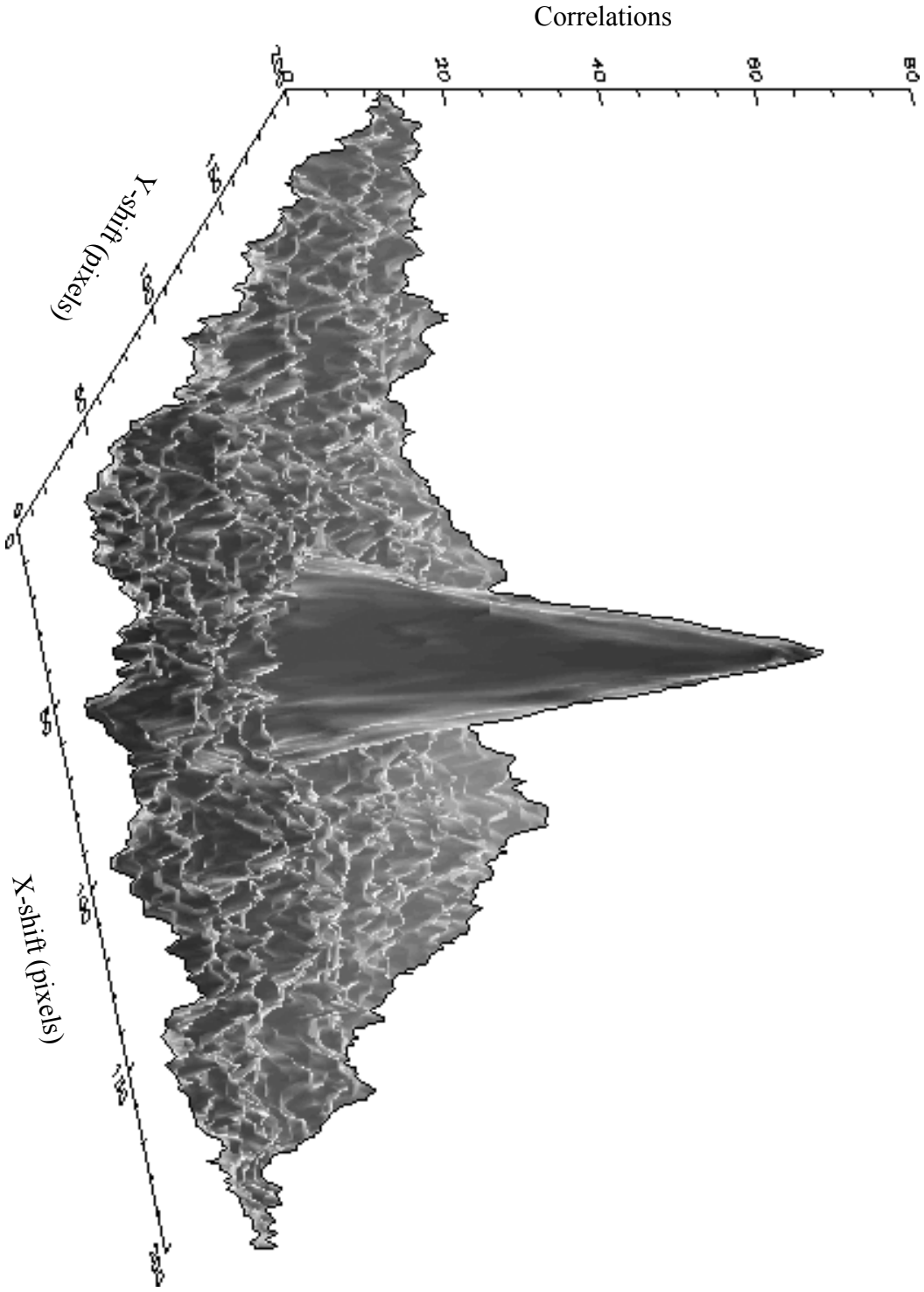


Figure 2.5

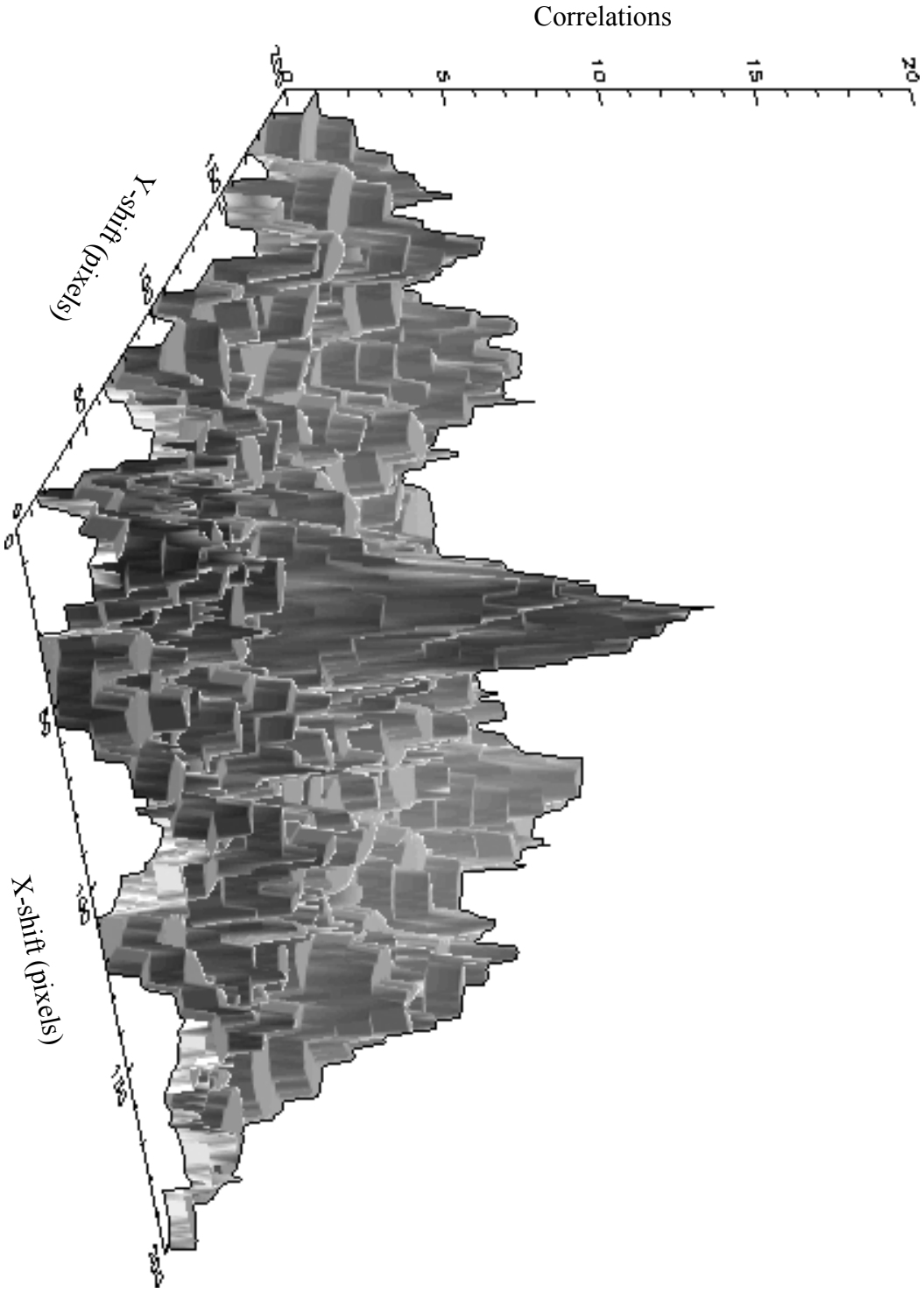
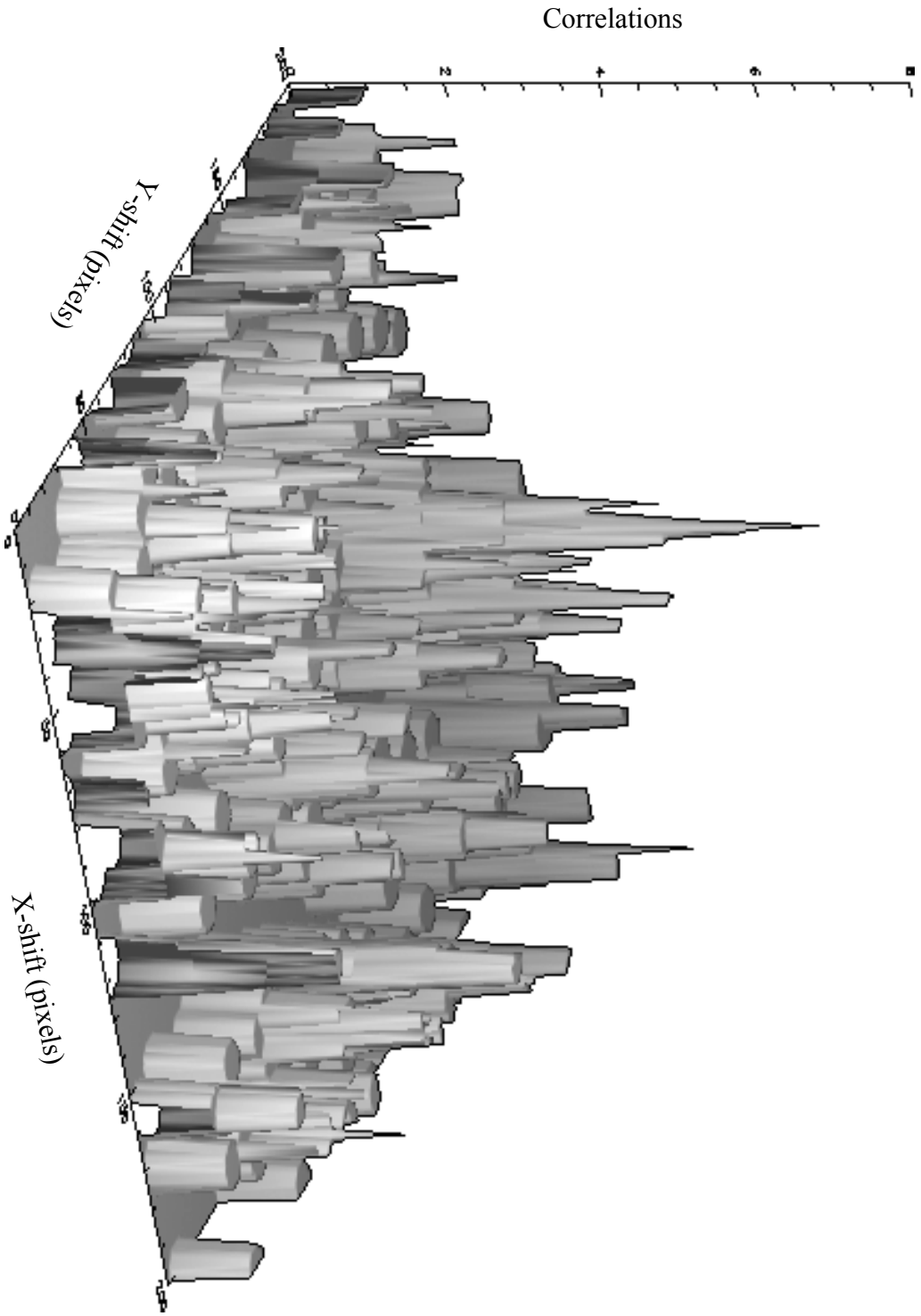


Figure 2.6



2.6 References

BAI, X., S. KIM, Z. LI, N. J. TURRO and J. JU, 2004. Design and synthesis of a photocleavable biotinylated nucleotide for DNA analysis by mass spectrometry. Nucleic Acids Res **32:** 535–541.

BRASLAVSKY, I., B. HEBERT, E. KARTALOV and S. R. QUAKE, 2003. Sequence information can be obtained from single DNA molecules. Proc Natl Acad Sci USA **100:** 3960–3964.

CANARD, B., B. CARDONA and R. S. SARFATI, 1995. Catalytic editing properties of DNA polymerases. Proc Natl Acad Sci USA **92:** 10859–10863.

CANARD, B., and R. S. SARFATI, 1994. DNA polymerase fluorescent substrates with reversible 3′-tags. Gene **148:** 1–6.

CROCKER, J. C., and D. G. GRIER, 1996. Methods of Digital Video Microscopy for Colloidal Studies. Journal of Colloid and Interface Science **179:** 298.

DECHER, G., 1997. Fuzzy Nanoassemblies: Toward Layered Polymeric Multicomposites. Science **277:** 1232–1237.

KARTALOV, E. P., M. A. UNGER and S. R. QUAKE, 2003. Polyelectrolyte surface interface for single-molecule fluorescence studies of DNA polymerase. Biotechniques **34:** 505–510.

LI, Z., X. BAI, H. RUPAREL, S. KIM, N. J. TURRO *et al.*, 2003. A photocleavable fluorescent nucleotide for DNA sequencing and analysis. Proc Natl Acad Sci USA **100:** 414–419.

MIDURA-NOWACZEK, K., I. BRUZGO, J. POPLAWSKI, W. ROSZKOWSKA-JAKIMIEC and K. WOROWSKI, 1995. Synthesis and activity of N alpha-(epsilon-aminocaproyl)-

alanines and N epsilon-(epsilon-aminocaproyl)-caproic acid. Acta Pol Pharm **52:** 505–507.

RASOLONJATOVO, I., and S. R. SARFATI, 1999. Development of a new DNA sequencing method: 3′-ester cleavage catalyzed by Taq DNA polymerase. Nucleosides Nucleotides **18:** 1021–1022.

RUPAREL, H., L. BI, Z. LI, X. BAI, D. H. KIM *et al.*, 2005. Design and synthesis of a 3′-O-allyl photocleavable fluorescent nucleotide as a reversible terminator for DNA sequencing by synthesis. Proc Natl Acad Sci USA **102:** 5932–5937.

SANGER, F., S. NICKLEN and A. R. COULSON, 1977. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA **74:** 5463–5467.

SEO, T. S., X. BAI, D. H. KIM, Q. MENG, S. SHI *et al.*, 2005. Four-color DNA sequencing by synthesis on a chip using photocleavable fluorescent nucleotides. Proc Natl Acad Sci USA **102:** 5926–5931.

SEO, T. S., X. BAI, H. RUPAREL, Z. LI, N. J. TURRO *et al.*, 2004. Photocleavable fluorescent nucleotides for DNA sequencing on a chip constructed by site-specific coupling chemistry. Proc Natl Acad Sci USA **101:** 5488–5493.

UNGER, M., E. KARTALOV, C. S. CHIU, H. A. LESTER and S. R. QUAKE, 1999. Single-molecule fluorescence observed with mercury lamp illumination. Biotechniques **27:** 1008–1014.

YILDIZ, A., J. N. FORKEY, S. A. MCKINNEY, T. HA, Y. E. GOLDMAN *et al.*, 2003. Myosin V walks hand-over-hand: single fluorophore imaging with 1.5-nm localization. Science **300:** 2061–2065.

ZHU, Z., J. CHAO, H. YU and A. S. WAGGONER, 1994. Directly labeled DNA probes using fluorescent nucleotides with different length linkers. Nucleic Acids Res **22:** 3418–3422.

3

Effects of a Modified Dye-Labeled Nucleotide Spacer Arm on Incorporation by Thermophilic DNA Polymerases

(In Press, Nucleosides, Nucleotides, & Nucleic Acids)

3.1 Abstract

The ability of eight commercially available thermophilic DNA polymerases to sequentially incorporate fluorescently labeled nucleotides was analyzed by a gel based primer extension assay. Cy5-dUTP or a variant nucleotide, in which the linker between the dye and the nucleobase was lengthened by 14 atoms , were compared. We found that the Cy5-dUTP with a longer linker resulted in longer primer extension lengths. Furthermore, some of the assayed polymerases are capable of extending the primer to the full-length or near full-length of 30 nucleotides using dye-labeled nucleotides exclusively.

3.2 Introduction

The ability of DNA polymerase to sequentially incorporate dye-labeled nucleotides has profound implications in the field of DNA sequencing, especially in the single-molecule sequencing-by-synthesis paradigm (BRASLAVSKY *et al.* 2003; MITRA *et al.* 2003). Sequencing-by-synthesis generally entails the sequential addition of fluorophore-labeled deoxynucleotide triphosphates (dNTPs) by DNA polymerase. If all dNTPs are labeled with the same fluorophore, it is necessary to incorporate each dNTP separately to the template in order to obtain sequence information. If each dNTP is labeled with a unique fluorophore, it may be possible to add all four dye-labeled dNTPs into the reaction at once and observe the polymerase extending the template sequence through time.

In each of these scenarios, the number of bases that the polymerase is able to extend a

template sequence solely using dye-labeled dNTPs is limited, likely because the bulk of the fluorophore side chain stemming off of the nucleotide sterically inhibits the polymerase from extending the template. In support of this theory, dNTPs containing cleavable linkers allow template extension in an unhindered manner (BAI *et al.* 2004; LI *et al.* 2003; SEO *et al.* 2004). Others have shown that by increasing the length of this side chain, DNA polymerase is able to incorporate more fluorophore-labeled dNTPs into the growing primer (ZHU *et al.* 1994). Scattered reports of conditions under which DNA polymerase is able to synthesize DNA exclusively from labeled nucleotides exist (BRAKMANN and NIECKCHEN 2001; FOLDES-PAPP *et al.* 2001; TASARA *et al.* 2003), however an established protocol capable of generating high product yield independent of template sequence has remained elusive. The efficiency of labeled nucleotide incorporation by DNA polymerase depends on many factors and is influenced by both the dye label as well as the nucleobase substrate (GILLER *et al.* 2003).

Although a correlation between fluorophore-dNTP linker length and the ability of DNA polymerase to incorporate a greater number of dye-labeled nucleotides into a given template DNA molecule has been qualitatively established, the extent of this linker length effect on subsequent nucleotide incorporations is unknown. We designed a gel based assay to directly examine the extent of primer extension by DNA polymerase in the presence of either commercially available Cy5-dUTP (Cy5-10-dUTP) or a Cy5-dUTP variant in which the linker has been lengthened by 14 atoms (Cy5-24-dUTP). We assayed eight thermophilic DNA polymerases using either Cy5-10-dUTP or Cy5-24-dUTP as substrate. All of the assayed polymerases generated products of greater length

when provided with Cy5-24-dUTP as substrate. The extent of increased product length varied for individual DNA polymerases.

3.3 Materials and Methods

3.3.1 Nucleotides

dTTP was purchased from Roche (Indianapolis, IN). Cy5-10-dUTP was purchased from Amersham Biosciences (Piscataway, NJ). Cy5-24-dUTP was prepared in a manner similar to the protocol developed by Waggoner (ZHU *et al.* 1994) as follows: Cy5-10-OSu was coupled to 6-(6-amino-hexanoylamino)-hexanoic acid (MIDURA-NOWACZEK *et al.* 1995) under standard conditions to produce Cy5-24-OH. After activation to the succinimidyl ester, Cy5-24-OSu was reacted with AP-dUTP to afford Cy5-24-dUTP. ESI-MS: [M-1] 1384.3, [M+4Na] 1473.3.

3.3.2 Template Preparation

Synthetic oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA). Annealing of primer with template was performed by mixing 3 nmol of Cy3-labeled primer (5′-Cy3-GTCTGGGCTTTTGGTTTGTGGG-3′) with 3 nmol of template (5′-$[A]_{30}$CCCACAAACCAAAAGCCCAGAC-3′) in 50 µl annealing buffer (150 mM NaCl, Tris-HCl, pH 7.2), heating the mixture for 5 minutes at 100°C, and cooling to room temperature over 1 hour. In order to remove any unannealed DNA, the mixture was then treated with 50 units of Exonuclease I (New England Biolabs, Beverly,

MA) and allowed to incubate at 37°C for 2 hours. Annealed duplex DNA was purified from the reaction using the QiaQuick Nucleotide Removal Kit (Qiagen, Valencia, CA).

3.3.3 Primer Extension

Primer extension reactions were performed using 15 pmol of annealed duplex DNA in 50 µl reaction buffer (Tris-HCl, pH 7.5, 5 mM $MgCl_2$, 12.5 mM dithiothreitol) containing 25 µM of the appropriate nucleotide and 1 unit of the appropriate DNA polymerase: LA Taq (Takara Mirus Bio, Madison, WI), Q-BioTaq (Qbiogene, Irvine, CA), Vent Exo$^-$ (New England Biolabs), Sequitherm (Epicentre, Madison, WI), Deep Vent Exo$^-$ (New England Biolabs), ThermalAce (Invitrogen, Carlsbad, CA), Thermus (Chimerx, Milwaukee, WI), Taq (Qiagen, Valencia, CA). Reactions were allowed to incubate for 3 hours at 60°C after which they were stopped by the addition of 2 µl 100 mM EDTA. Excess nucleotides and DNA polymerase were removed from each reaction using the QiaQuick Nucleotide Removal Kit. Extension reaction product was eluted from the kit's column using 50 µl 90% formamide.

3.3.4 Product Analysis

Electrophoretic separation of reaction products was performed using denaturing 15% (w/v) polyacrylamide TBE-urea gels (Invitrogen, Carlsbad, CA). 2 µl of denatured GeneScan LIZ120 (Applied Biosystems, Foster City, CA) was used as a size standard. 10 µl of extension reaction product (approximately 2 pmol) was mixed with 10 µl of 2X TBE-urea preparative buffer (Invitrogen) containing no dyes. This mixture was heated to

100°C for 5 minutes to denature primer product from template and immediately transferred to an ice slurry. Samples were loaded onto the gel and run in 1X TBE at constant 180 V for 38 minutes at 55°C. Gels were transferred to 500 ml of fixing solution (10% acetic acid, 10% methanol) and gently shaken for 1 hour after which they were washed 3 times (15 minutes each wash) with 500 ml of ultrapure water. Gels were imaged on a Typhoon 8600 variable mode imager (Amersham Biosciences) at high sensitivity and 100 micron resolution. Cy3 fluorescence was visualized using the green (532 nm) laser as excitation source (PMT setting at 700 V) and a 555 nm BP 20 emission filter. Cy5 and LIZ were detected using the red (633 nm) laser (PMT setting at 800 V) as excitation source and a 670 nm BP 30 emission filter. Images were processed and channel cross-contamination was removed using the Typhoon's IQ Solutions software package.

3.4 Results and Discussion

We designed a primer extension assay to screen several commercially available thermophilic DNA polymerases for their ability to incorporate a Cy5-labeled dNTP. Additionally, we sought to compare the extent to which these polymerases were able to extend the primer using a Cy5-dUTP variant that contained a longer spacer arm between the dye and the dNTP.

To validate that the buffer conditions were appropriate for the given polymerases, we first performed the primer extension assay using unlabeled dTTP as polymerase substrate. As shown in Figure 3.2 (lanes 3-10), under these conditions, each of the tested polymerases

extended all detectable primer to full-length or near full-length (52 bases). Slight differences in product size may be due to template independent extension of the fully extended strand by one or a few bases at its 3′ end, a known property of some thermophilic DNA polymerases exploited in molecular subcloning of PCR products into plasmid vectors (ZHOU and GOMEZ-SANCHEZ 2000). Figure 3.2 (lane 2) shows the unextended primer of 22 bases under identical conditions in the absence of polymerase.

The assay was repeated with both commercially available Cy5-dUTP (Cy5-10-dUTP) and a longer-linkered variant (Cy5-24-dUTP) as substrate (Figure 3.3). When provided with Cy-10-dUTP as substrate, most polymerases were able to extend the primer to varying degrees by approximately 5–10 bases. Notably, several polymerases appear to convert at least a portion of primer to full-length or near full-length product: LA Taq, Figure 3.3A (lane 3), ThermalAce Figure 3.3B (lane 5), and Thermus, Figure 3.3B (lane 7).

When Cy-24-dUTP is instead provided as substrate, the length of the longest extended primer for each polymerase is greater. LA Taq (Figure 3.3A, lane 4), Q-Bio Taq (Figure 3.3A, lane 6), Deep Vent Exo$^-$ (Figure 3.3B, lane 4), Thermal Ace (Figure 3.3B, lane 6), and Thermus (Figure 3.3B, lane 8) appear to extend a portion of the primer to full-length or near full-length. Strikingly, ThermalAce produces product ranges between approximately 30 and 52 bases (primer extended by 8–30 nucleotides, respectively).

We were surprised to observe that a large fraction of primer remained unextended in the assay when labeled-dUTP was provided as substrate. With the exception of Q-Bio Taq and Thermal Ace, the majority of primer appears to run the same length as negative controls containing no polymerase (Figure 3.3A and 3.3B, lane 2). Although sufficient time was given to fully extend the template with dTTP, a longer extension time may be needed with labeled dNTPs. Alternatively, labeled dNTPs may interfere with polymerase docking or the initiation of synthesis for our chosen template.

We originally sought to quantitate extension product by labeling our primer with Cy3. However, as seen in Figure 3.3, although incorporated Cy5 nucleotides are readily detected, we were unable to detect any significant Cy3 fluorescence from the extended products. One possibility is that multiple incorporation events enhance Cy5 fluorescence levels stoichiometrically. Additionally, Cy5 is an acceptor for Cy3 fluorescence energy. Although 22 bases (the distance between the Cy3 tag and the first incorporation site) exceeds the Forster radius for the Cy3/Cy5 fluorescent resonant energy transfer (FRET) pair, the sheer number of Cy5 molecules on the extended product and within the three dimensional space of the gel may absorb this energy and mask Cy3 fluorescence.

In conclusion, we compared the ability of eight commercially available thermophilic DNA polymerases to extend a primer utilizing only Cy-10-dUTP or Cy-24-dUTP as substrate. Our findings show that using the longer-linkered Cy5-24-dUTP results in greater extension lengths as well as in an increase of product extended to full-length or near full-length.

3.5 Figure Legends

Figure 3.1

3.5.1 Cy5-Labeled Nucleotides

Commercially available Cy5-dUTP (Cy5-10-dUTP) and longer-linkered derivative Cy5-24-dUTP. Cy5-24-dUTP contains two additional aminohexanoic acid fragments relative to Cy5-10-dUTP. These extra groups increase the distance between Cy5 and the deoxynucleoside triphosphate by 14 atoms.

Figure 3.2

3.5.2 dTTP Primer Extensions

Size separation of primer extension reactions on a 15% (w/v) denaturing polyacrylamide gel imaged on a Typhoon 8600 scanning imager. Lane 1 shows the GeneScan LIZ120 size standard (red). Numbers on the left indicate size in bases. Lanes 2–10 are primer extension reactions using dTTP as polymerase substrate in which green represents fluorescence from the Cy3-labeled primer. Lane 2 shows the negative control in which no polymerase was added to the reaction. Lanes 3–10 are positive controls and represent primer extension reactions in which the following thermophilic DNA polymerases were used: LA Taq (lane 3), Q-BioTaq (lane 4), Vent Exo$^-$ (lane 5), Sequitherm (lane 6), Deep Vent Exo$^-$ (lane 7), ThermalAce (lane 8), Thermus (lane 9), and Taq (lane 10).

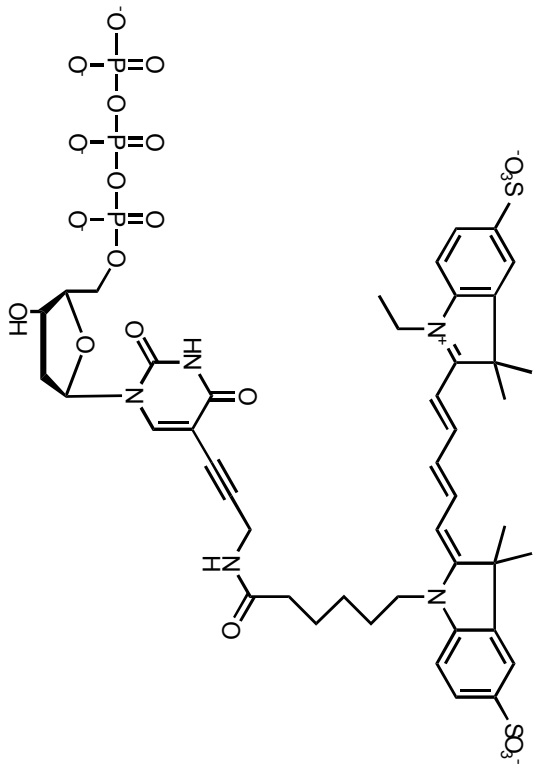
Figure 3.3

3.5.3 Cy5-10-dUTP and Cy5-24-dUTP Primer Extensions

Size separation of primer extension reactions using Cy5-10-dUTP and Cy5-24-dUTP as polymerase substrates. Extension products were separated on 15% (w/v) denaturing polyacrylamide gels and imaged using a Typhoon 8600 imaging system. For both A and B, Lane 1 shows the GeneScan LIZ120 size standard (red). Numbers on the left indicate size in bases. A) Lane 2 shows the negative control in which no polymerase was added to the reaction containing Cy5-10-dUTP. Lanes 3, 5, 7, and 9 used Cy5-10-dUTP as polymerase substrate while lanes 4, 6, 8, and 10 used Cy5-24-dUTP as polymerase substrate. Red represents fluorescence from incorporated Cy5-labeled nucleotides and green represents fluorescence from the Cy3-labeled primer. The following DNA polymerases were used: LA Taq (lanes 3 and 4), Q-BioTaq (lanes 5 and 6), Vent Exo$^-$ (lanes 7 and 8), and Sequitherm (lanes 9 and 10). B) Lane 2 shows the negative control in which no polymerase was added to the reaction containing Cy5-24-dUTP. Lanes 3, 5, 7, and 9 used Cy5-10-dUTP as polymerase substrate while lanes 4, 6, 8, and 10 used Cy5-24-dUTP as polymerase substrate. The following DNA polymerases were used: Deep Vent Exo$^-$ (lanes 3 and 4), ThermalAce (lanes 5 and 6), Thermus (lanes 7 and 8), and Taq (lanes 9 and 10).

Figure 3.1

A.



B.

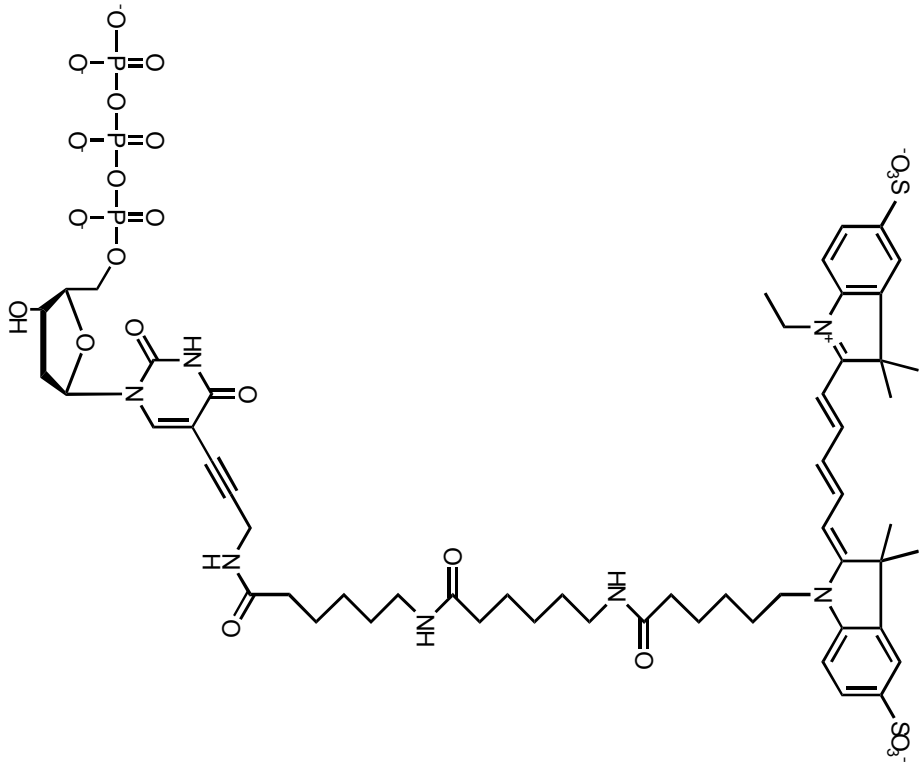


Figure 3.2

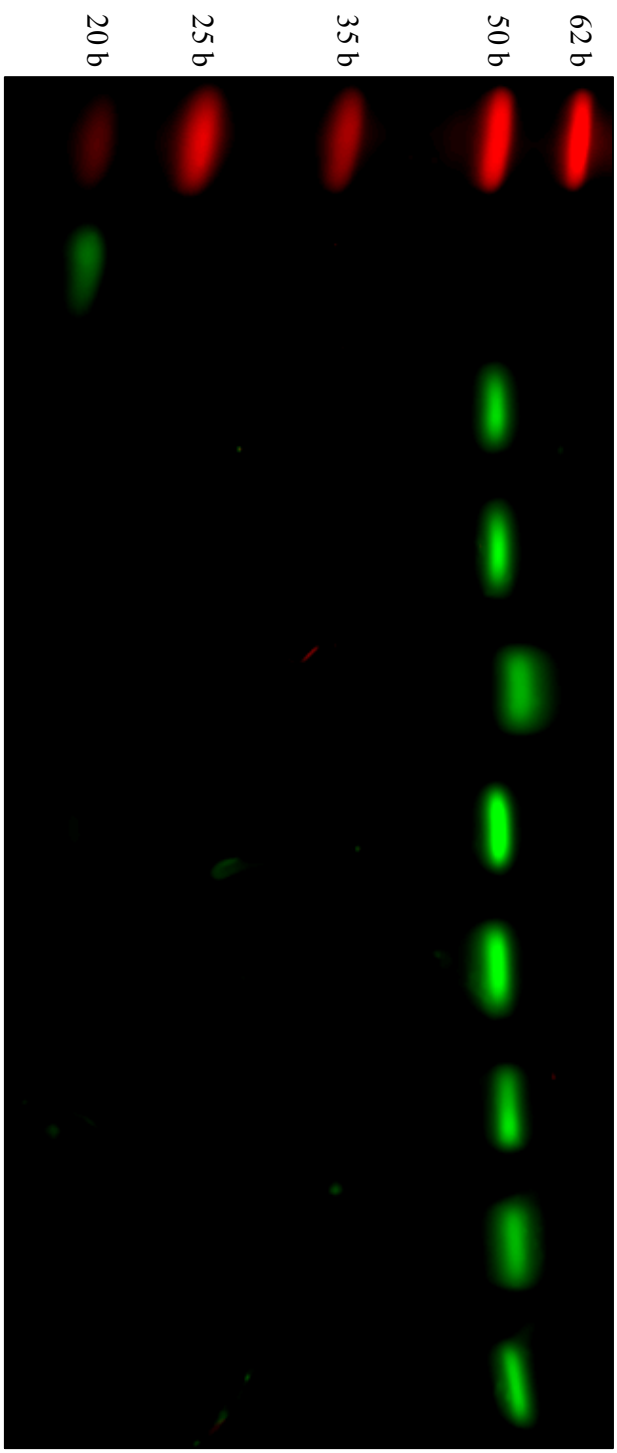
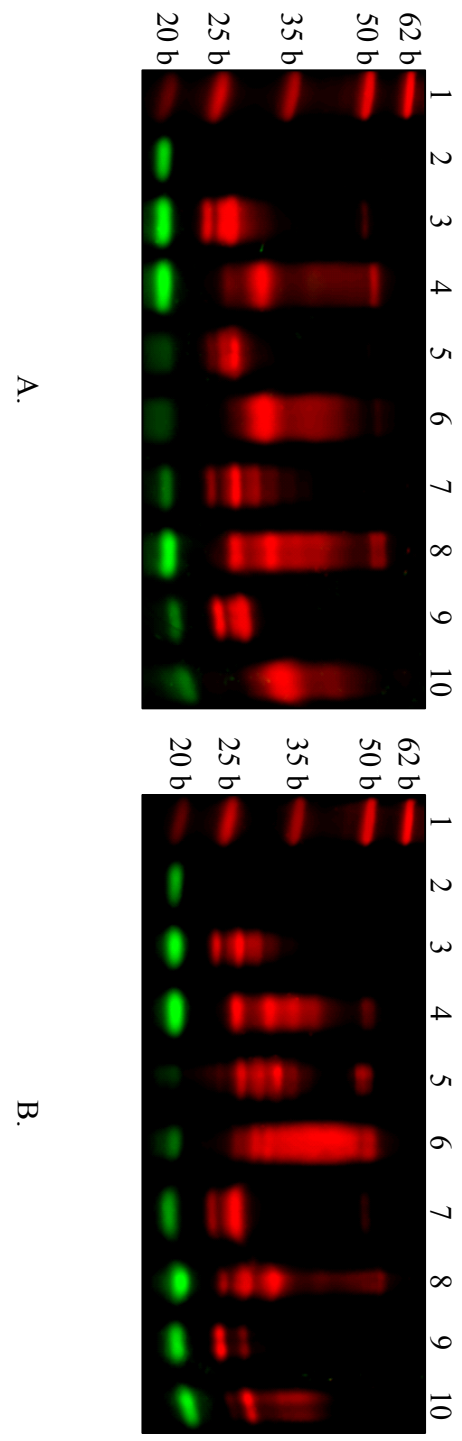


Figure 3.3


62 b
50 b
35 b
25 b
20 b
1 2 3 4 5 6 7 8 9 10
A.
62 b
50 b
35 b
25 b
20 b
1 2 3 4 5 6 7 8 9 10
B.

3.6 References

BAI, X., S. KIM, Z. LI, N. J. TURRO and J. JU, 2004. Design and synthesis of a photocleavable biotinylated nucleotide for DNA analysis by mass spectrometry. Nucleic Acids Res **32:** 535–541.

BRAKMANN, S., and P. NIECKCHEN, 2001. The large fragment of *Escherichia coli* DNA polymerase I can synthesize DNA exclusively from fluorescently labeled nucleotides. Chembiochem **2:** 773–777.

BRASLAVSKY, I., B. HEBERT, E. KARTALOV and S. R. QUAKE, 2003. Sequence information can be obtained from single DNA molecules. Proc Natl Acad Sci USA **100:** 3960–3964.

FOLDES-PAPP, Z., B. ANGERER, W. ANKENBAUER and R. RIGLER, 2001. Fluorescent high-density labeling of DNA: error-free substitution for a normal nucleotide. J Biotechnol **86:** 237–253.

GILLER, G., T. TASARA, B. ANGERER, K. MUHLEGGER, M. AMACKER *et al.*, 2003. Incorporation of reporter molecule-labeled nucleotides by DNA polymerases. I. Chemical synthesis of various reporter group-labeled 2′-deoxyribonucleoside-5′-triphosphates. Nucleic Acids Res **31:** 2630–2635.

LI, Z., X. BAI, H. RUPAREL, S. KIM, N. J. TURRO *et al.*, 2003. A photocleavable fluorescent nucleotide for DNA sequencing and analysis. Proc Natl Acad Sci USA **100:** 414–419.

MIDURA-NOWACZEK, K., I. BRUZGO, J. POPLAWSKI, W. ROSZKOWSKA-JAKIMIEC and K. WOROWSKI, 1995. Synthesis and activity of N alpha-(epsilon-aminocaproyl)-

alanines and N epsilon-(epsilon-aminocaproyl)-caproic acid. Acta Pol Pharm **52:** 505–507.

MITRA, R. D., V. L. BUTTY, J. SHENDURE, B. R. WILLIAMS, D. E. HOUSMAN *et al.*, 2003. Digital genotyping and haplotyping with polymerase colonies. Proc Natl Acad Sci USA **100:** 5926–5931.

SEO, T. S., X. BAI, H. RUPAREL, Z. LI, N. J. TURRO *et al.*, 2004. Photocleavable fluorescent nucleotides for DNA sequencing on a chip constructed by site-specific coupling chemistry. Proc Natl Acad Sci USA **101:** 5488–5493.

TASARA, T., B. ANGERER, M. DAMOND, H. WINTER, S. DORHOFER *et al.*, 2003. Incorporation of reporter molecule-labeled nucleotides by DNA polymerases. II. High-density labeling of natural DNA. Nucleic Acids Res **31:** 2636–2646.

ZHOU, M. Y., and C. E. GOMEZ-SANCHEZ, 2000. Universal TA cloning. Curr Issues Mol Biol **2:** 1–7.

ZHU, Z., J. CHAO, H. YU and A. S. WAGGONER, 1994. Directly labeled DNA probes using fluorescent nucleotides with different length linkers. Nucleic Acids Res **22:** 3418–3422.

4

Obtaining Sequence Information from Single Molecules of RNA

4.1 Introduction

The ability to quickly obtain large amounts of gene expression data is of central importance to many areas of biology. The identification of genes expressed at unique points both spatially and temporally during embryogenesis and throughout life, as well as their regulatory elements, is crucial in our understanding of the molecular basis of life (WELLE 2002). In cancer biology, determining differences in gene expression profiles between normal cells and cancerous cells may yield valuable drug targets (BOON and RIGGINS 2003; DECHERING 2005; LIU 2004) by focusing on cellular pathways that are altered during oncogenesis. Finally, the ability to compare single-molecule transcript mRNA sequences against their parent genes' DNA sequences, particularly at exon/intron boundaries, is a powerful tool in understanding the complex nature of post-transcriptional gene regulation and mRNA modification (ZHU *et al.* 2003). Several common quantitative, high-throughput techniques currently available to analyze gene expression are briefly introduced below, followed by a proposal for a new tool to analyze gene expression and a set of results supporting such a proposal.

4.1.1 Microarrays

Microarray hybridization technology is a parallel technique used to measure gene expression levels (DUGGAN *et al.* 1999; HACIA 1999; SCHENA *et al.* 1995). It relies on the detection of the hybridization of labeled oligonucleotides or cDNA clones, via complementary binding, to an array spotted with thousands to over one million unique DNA sequences. Microarray technology is a closed-ended system that demands a good

deal of planning and investment prior to experimentation. In order to produce a hybridization array, one must first possess knowledge of the genome of interest. Quantitation of gene expression using microarray hybridization relies on comparing fluorescent dye ratios or spot intensities and therefore offers an analog profile of the transcriptome. Furthermore, cross hybridization of closely related sequences may be problematic.

4.1.2 SAGE

SAGE (Serial Analysis of Gene Expression) (EL-MEANAWY *et al.* 2003; LINDLÖF 2003; VELCULESCU *et al.* 1995) is an open-ended method to analyze gene expression levels that requires no prior knowledge of the sample genome before experimentation. The SAGE technique does not rely on hybridization, but instead, directly obtains the sequence of a short region of cDNA. This tag is identified through database searches. Briefly, mRNA is isolated from a sample and its corresponding cDNA is produced. Short, unique regions of this cDNA are concatamerized together into a single clone for sequencing. In doing so, one is able to efficiently sequence 20 to 30 short tags at once in a single sequencing gel lane using standard dideoxy sequencing protocols. Gene expression profiling is done by comparing each short signature sequence of 9 to 13 base pairs, each of which represents a unique transcript, against a signature sequence tag database. Quantitation of gene expression is determined by the frequency at which a sequence tag is detected and is therefore, unlike microarrays, a digital characterization of the transcriptome. SAGE counting statistics are well-modeled by the Poisson distribution. The precision of

transcript abundance estimations may be increased by increasing the length of each sequence tag or the number of sequence tags counted (AUDIC and CLAVERIE 1997).

4.1.3 MPSS

MPSS (Massively Parallel Signature Sequencing) (BRENNER *et al.* 2000) is a relatively new tool for performing expression analysis. Like SAGE, MPSS is an open-ended technique—no prior knowledge of the genes being studied is needed before experimentation. Quantitation of gene expression is digital, with sensitivity levels of a few molecules of mRNA per cell. Briefly, mRNA is isolated and, after an RT-PCR step, its corresponding cDNA is "cloned" into microbeads of approximately 5 μ in diameter. Each bead carries a unique species of cDNA. Those beads containing tethered cDNA are isolated using a cell sorter. These loaded beads are placed as a monolayer into a flow cell and the sequence of their tethered cDNA is interrogated by a complex battery of ligations and cleavages using type IIs restriction enzymes. Up to one million beads may be assayed in a single experiment. Ultimately, a short signature sequence of 16–20 bases is obtained and identified through database searches.

4.1.4 RAMPAGE

We currently propose a novel method of gene expression analysis, tentatively termed RAMPAGE (Rapid and Massively Parallel Analysis of Gene Expression), that may obtain sequence information from single molecules of nucleic acid, either directly from mRNA or from its corresponding cDNA. Single-molecule sensitivity affords the

possibility of forgoing all amplification steps between transcript isolation and transcript identification. That is, mRNA transcripts may be identified directly by sequencing individual mRNA molecules with reverse transcriptase. A schematic is shown in Figure 4.1. Hybridization techniques may also be used to probe the RNA samples directly for a specific transcript with a labeled probe. Alternatively, as diagrammed in Figure 4.2, cDNA may be synthesized from its corresponding mRNA. Single cDNA molecules may be sequenced directly using DNA polymerase, eliminating all cloning steps and sample amplification that typically is required of other techniques lacking single-molecule sensitivity.

Like microarrays and MPSS, RAMPAGE is highly parallel. Theoretically, 12 million individual molecules can be sequenced in a 1″ × 1″ square area. This unmatched degree of parallelism translates to an ultrafast technique with which one may characterize the transcriptome of one or a few cells in its entirety. Furthermore, such an analysis may use only a fraction of the reagents that other techniques require because of the inherently small scale of space and reagent volume requirements of single-molecule detection systems and because an entire transcriptome may be analyzed in one shot without intermediate cloning and amplification steps.

Furthermore, like SAGE and MPSS, RAMPAGE is an open-ended system. That is, expression data may be acquired *de novo*, without prior knowledge of the sample's genome. A unique sequence tag read length of 10–20 base pairs should be sufficient to find a match by performing a homology search against one or more of the many available

databases that contain sequence tag information. This sequence tag length has been used successfully in other techniques, including SAGE. Longer read lengths, or selective priming, may allow for the identification of alternative splice forms of a transcript by sequencing across exon/intron boundaries.

Given that most RNA or cDNA molecules successfully adhere to the slide surface, and that transcript degradation is minimized, quantitation of gene expression should be directly reflected in the abundance of any particular sequence tag within the population of molecules and therefore offers a digital profiling of gene expression. Poisson statistics should apply to counting models using RAMPAGE in a similar manner as they do to SAGE. Furthermore, RAMPAGE should facilitate the detection of rare transcripts—in theory, a single transcript molecule—because it relies on a single-molecule detection scheme. Also, because RAMPAGE does not employ any amplification steps, abundant transcripts are not able to exponentially swamp out rarer ones, as is typically observed during cDNA cloning and subsequent amplification.

Direct sequencing, in bulk, of cDNA without recourse to cloning has previously been reported (KITAMURA and WIMMER 1980) and the groundwork for sequencing single molecules of DNA has recently been elegantly demonstrated (BRASLAVSKY *et al.* 2003). Furthermore, examples exist of direct RNA sequencing by reverse transcriptase (HAMLYN *et al.* 1978; HAMLYN *et al.* 1981; ZIMMERN and KAESBERG 1978) using a modified Sanger protocol (SANGER *et al.* 1977) and chain-terminators. Since all of these examples were of bulk sequencing of RNA, we chose to explore the possibility of

directly sequencing single molecules of RNA using reverse transcriptase, fluorescently labeled nucleotides, and the sequence-by-synthesis paradigm.

4.2 Results and Discussion

In order to test the feasibility of directly sequencing single molecules of RNA, we sought to address the following three fundamental questions:

1. Is reverse transcriptase able to incorporate dye-labeled nucleotides into a DNA-primed RNA template, and if so, what sort of read length is observed?

2. Are RNA template molecules stable enough to be successfully detected after being anchored to the coverslip surface of the sequencing chamber?

3. Can we detect the incorporation of a dye-labeled nucleotide by reverse transcriptase into DNA-primed RNA template molecules that have been anchored to the coverslip surface of the sequencing chamber?

4.2.1 Primer Extension Assays of an RNA Template

Figure 4.3 shows the results of a primer extension assay in which a primed RNA template containing a stretch of 28 consecutive A's was extended by Sensiscribe reverse transcriptase. Surprisingly the resulting product of this extension, when Cy5-dUTP is provided as nucleotide, appears to be full or near full length when compared to the

resulting extension product when dTTP is provided as nucleotide. When Sensiscribe reverse transcriptase is withheld from the reaction the resulting product appears to be the same size as the primer alone. Therefore, production of apparently full length product is dependent upon reverse transcriptase.

A more detailed analysis of primer extension of an RNA template by reverse transcriptase is shown in Figure 4.4. In this experiment, four commercially available reverse transcriptases were assayed against an RNA template containing a stretch of 28 consecutive A's, as in Figure 4.3. For each reverse transcriptase assayed, an apparent full length product of correct size was produced when Cy5-dUTP was given as nucleotide. Also, in each case, when reverse transcriptase was withheld, the resulting product was of the same size observed of the primer alone. Thus, with respect to the labeled nucleotide Cy5-dUTP and a poly(A) RNA template, reverse transcriptase appears to be able to extend an annealed DNA primer by approximately 28 bases solely using a labeled nucleotide.

A possible explanation for the observed ability of reverse transcriptase to extend a primer annealed to an RNA template to such a degree using only a labeled nucleotide may lie in the three-dimensional structure of its active site. Figure 4.5 shows a three-dimensional model of T7 DNA polymerase complexed with a duplex DNA template. The active site of T7 DNA polymerase, through which the growing template strand must pass, appears relatively narrow compared to the active site of HIV reverse transcriptase complexed with an RNA pseudoknot (Figure 4.6). Thus, the bulky dye side chain of Cy5-dUTP

may impose less of a steric burden upon the active site of reverse transcriptase, allowing it to utilize the dye-labeled nucleotide better than DNA polymerase.

Encouraged by these results, we next wanted to test the stability of an RNA template under the conditions typically used to sequence single molecules of DNA. Additionally, we sought to detect incorporation of Cy5-dUTP into an RNA template at the single-molecule level.

4.2.2 Single-Molecule Cy5-dUTP Incorporation into an RNA Template

The relative instability of RNA, largely due to degradation by RNAse enzymes present in the surrounding environment, necessitated the validation of our ability to detect single RNA molecules and, subsequently, the incorporation of a dye-labeled nucleotide (Cy5-dUTP) into these molecules by reverse transcriptase. Below is a brief description of these experiments.

The same template used in the bulk primer extension assay described above was used for our single-molecule experiments described here. It was a Cy3-labeled DNA oligonucleotide annealed to an RNA template containing a stretch of 28 consecutive A's. This template was adhered to the surface of the sequencing chamber as described in the Materials and Methods section.

First, we imaged the surface of the sequencing chamber using a green laser (532 nm) and a Cy3 emission filter. The fluorescent features detected in this image represent

prospective template RNA molecules. After acquiring this image, we bleached all fluorescence from the field of view using the green laser in the absence of oxygen scavenger. We next added a buffer containing reverse transcriptase and Cy5-dUTP. After a brief incubation, we acquired a second image using a red laser (635 nm) and a Cy5 emission filter. The fluorescent features detected in the second image represent, potentially, the Cy5-dUTP nucleotides incorporated into the surface-bound template RNA molecules. Sterile technique was adhered to at all times, and all solutions, including those used during wash steps, contained RNAse inhibitor.

Using custom IDL software, these images were denoised and the coordinates of the fluorescent features observed in each were noted. A correlogram was produced by shifting the second set of coordinates over the first. At each step during shifting, the number of correlating features was determined.

The resulting correlogram is shown in Figure 4.7. Forty-two matches were found at a single position, as represented by the obvious peak in this three-dimensional representation. The off-center shift of this peak is a result of the second field of view being slightly offset from the first. This is likely due to a shift of the microscope stage or the sequencing chamber itself. However, this result strongly suggests that the surface-bound template RNA molecules were able to persist through the experiment and that Cy5-dUTP had been successfully incorporated into the surface-bound template RNA molecules. Because our template contained a series of A's, it is not known whether these

incorporations represent a single Cy5-dUTP incorporation or multiple ones. In theory, it may be possible to determine this by comparing relative fluorescence intensities.

As a negative control, the identical procedure described above was repeated in the absence of reverse transcriptase. The results of this experiment are shown in Figure 4.8. A maximum of 17 matches were found at six positions and suggests that, as one may expect in the absence of reverse transcriptase, no incorporations into the template RNA molecules were detected.

Taken together, these experiments demonstrate that our template RNA is stable enough to be detected at the single-molecule level and, that once tethered to the sequencing chamber surface, it is accessible to reverse transcriptase. Furthermore, we have demonstrated, at the single-molecule level, the ability to detect the incorporation of Cy5-dUTP into our template RNA using reverse transcriptase.

4.3 Concluding Remarks

In summary, we set out to determine the feasibility of directly sequencing RNA using fluorescently labeled nucleotides and reverse transcriptase. We first performed bulk primer extension assays to assess the ability of reverse transcriptase to incorporate dye-labeled nucleotides and to determine what sort of read lengths resulted. Our results suggest that using a poly(A) RNA template and Cy5-dUTP as the dye-labeled nucleotide, four commercially available reverse transcriptases are able to extend a DNA primer by approximately 28 bases at the least.

Second, we set out to determine whether an RNA template was stable enough to be detected at the single-molecule level. Furthermore, we wished to detect the incorporation of Cy5-dUTP by reverse transcriptase at the single-molecule level. We conclude that the stability of our template RNA molecule is sufficient to be detected at the single-molecule level and that incorporation of Cy5-dUTP by reverse transcriptase is possible within the sequencing chamber.

Although these promising results give hope to the possibility of single-molecule transcriptome analysis, much more work is needed. Diverse templates and other dye-nucleotide combinations must be rigorously examined. Protocols to tether genuine mRNA templates to the sequencing chamber surface must be validated and the bioinformatics of transcript identification must still be worked out. The potential scientific payoff for developing a practical, feasible, working RAMPAGE technology is enormous and warrants further investigation.

4.4 Materials and Methods

4.4.1 Nucleotides

dTTP was purchased from Roche (Indianapolis, IN). Cy5-dUTP was purchased from Amersham Biosciences (Piscataway, NJ).

4.4.2 Template Preparation

Synthetic DNA and RNA oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA). Annealing of primer with template was performed by mixing 3 nmol of Cy3-labeled DNA primer (5′-Cy3-GTCTGGGCTTTTGGTTTGTGGG-3′) with 3 nmol of biotinylated RNA template (5′-Biotin-$[A]_{28}$CCCACAAACCAAAAGCCCAGAC-3′) in 50 µl annealing buffer (150 mM NaCl, Tris-HCl, pH 7.2), heating the mixture for 3 minutes at 100°C, and cooling to room temperature over 1 hour. Annealed duplex was then purified using an RNAse-free P30 gel filtration column (BioRad).

4.4.3 Primer Extension

Primer extension reactions were performed using 15 pmol of annealed duplex in 50 µl reaction buffer (Tris-HCl, pH 7.5, 5 mM $MgCl_2$, 12.5 mM dithiothreitol) containing 25 µM of the appropriate nucleotide, 50 units of RNAsin RNAse inhibitor (New England Biolabs), and 250 units of the appropriate reverse transcriptase: Sensiscribe (Qiagen), M-MuLV (Roche), AMV (Roche), or Superscript III (Invitrogen). Reactions were allowed

to incubate for 2 hours at 50°C after which they were stopped by the addition of 2 µl 100 mM EDTA. Excess nucleotides were removed from each reaction using an RNAse-free P30 gel filtration column (BioRad).

4.4.4 Product Analysis

Electrophoretic separation of reaction products was performed using denaturing 12% (w/v) polyacrylamide TBE-urea gels (Invitrogen, Carlsbad, CA). 2 µl of denatured GeneScan ROX350 (Applied Biosystems, Foster City, CA) was used as a size standard. 10 µl of extension reaction product (approximately 2 pmol) was mixed with 10 µl of 2X TBE-urea preparative buffer (Invitrogen) containing no dyes. This mixture was heated to 100°C for 5 minutes to denature primer product from template and immediately transferred to an ice slurry. Samples were loaded onto the gel and run in 1X TBE at constant 180 V for 50 minutes at room temperature. Gels were imaged on a Typhoon 8600 variable mode imager (Amersham Biosciences) at normal sensitivity and 200 micron resolution. Cy3 fluorescence was visualized using the green (532 nm) laser as excitation source (PMT setting at 700 V) and a 555 nm BP 20 emission filter. ROX fluorescence was detected using the green (532 nm) laser as excitation source (PMT setting at 700 V) and a 610 nm BP 30 emission filter. Cy5 fluorescence was detected using the red (633 nm) laser (PMT setting at 800 V) as excitation source and a 670 nm BP 30 emission filter. Images were processed, denoised, and channel cross-contamination was removed using the Typhoon's IQ Solutions software package.

4.4.5 Single-Molecule Detection and Data Analysis

A homemade microscope configured for through-the-objective, total internal reflection (TIR) illumination served as a platform for the single-molecule experiments. Two laser beams, 635 nm (Hitachi) and 532 nm (Intellite), with nominal powers of 30 and 50 mW, respectively, were circularly polarized by quarter-wave plates. The laser beams were focused on the back aperture of the objective (PlanApo, 60X NA 1.45 oil, Olympus) to create an epi-illuminated area of 200 × 200 μm. TIR was achieved by displacing the beams from the objective's center, toward the objective's edge with a mirror mounted to a translation stage. The relatively large numerical aperture of the objective allowed the beams to be focused at angles greater than the critical angle (~62.5°) at the cover glass/water interface, allowing the evanescent wave to extend ~150 nm into the specimen chamber. The specimen chamber made from a low-autofluorescence microscope coverslip (Schott) and an adhesive hybridization well (Sigma). An image splitter (Optical Insights) directed the light through two band-pass filters (630dcxr, HQ585/80, HQ690/60; Chroma Technology) to a scientific grade charge-coupled device camera (ORCA-ER, Hamamatsu), which recorded images of a 200 × 200 μm section of the surface. Typically, several exposures of 0.5 sec each were taken of each field of view to compensate for possible intermittency in the fluorophore emission. Custom IDL software (CROCKER and GRIER 1996) was modified to analyze the locations and intensities of fluorescence objects in the intensified charge-coupled device pictures. We inspected the resulting traces to determine incorporation events on the primed template sequences.

4.4.6 Single-Molecule Coverslip and Sequencing Chamber Preparation

Surface chemistry based on polyelectrolytes (DECHER 1997; KARTALOV *et al.* 2003) and biotin-streptavidin bonding was used to anchor the RNA molecules to the coverslip surface of the hybridization chamber and to minimize nonspecific binding of the nucleotides to the surface. Coverslips were sonicated in 2% MICRO-90 soap (Cole–Parmer) for 20 minutes and then cleaned by immersion in boiling RCA solution (6:4:1 high-purity H_2O/30% NH_4OH/30% H_2O_2) for 1 hour (UNGER *et al.* 1999). They were then immersed alternately in polyallylamine (positively charged) and polyacrylic acid (negatively charged; both from Aldrich) at 2 mg/ml and pH 8.0 for 10 minutes each and washed intensively with high-purity water in between. The carboxyl groups of the last polyacrylic acid layer served to prevent the negatively charged labeled nucleotide from binding nonspecifically to the surface. In addition, these functional groups were used for further attachment of a layer of biotin. The coverslips were incubated with 5 mM biotin-amine reagent (Biotin-EZ-Link, Pierce) for 10 minutes in the presence of 1-[3-(dimethylamino)propyl]-3-ethylcarbodiimide hydrochloride (EDC, Sigma) in MES buffer, pH 5.5, followed by incubation with streptavidin (Neutravidin, Pierce) at 0.1 mg/ml for 15 min in Tris buffer, pH 8.0. The biotinylated RNA templates were deposited onto the streptavidin-coated chamber surface at 25 pM for 10 minutes in Tris buffer, pH 8.0 and RNAsin RNAse inhibitor (New England Biolabs) at 1 unit/µl. From this point onward, all solutions including washes contained RNAsin RNAse inhibitor at 1 unit/µl. For incorporations, the reaction solution contained Superscript III RNAse H minus reverse transcriptase (Invitrogen) at 5 units/ µl in the reaction buffer (Superscript III reaction buffer, Invitrogen) and Cy5-dUTP or Cy5-dCTP (Amersham Biosciences) at 0.5

μM. Reaction incubation times were 10 minutes. To reduce bleaching of the fluorescence dyes, an oxygen scavenging system (YILDIZ *et al.* 2003) was used during all illumination periods, with the exception of the bleaching of the primer's Cy3 tag.

4.4.7 Reagent Exchange Sequence

The positions of the anchored Cy3-primed RNA were recorded, and then the tags were bleached by green laser illumination. Cy5-dUTP or Cy5-dCTP and reverse transcriptase were introduced and washed out. An image of the surface was then analyzed for incorporated Cy5-labeled nucleotides using red illumination.

4.5 Figure Legends

Figure 4.1

4.5.1 Schematic Diagram of mRNA RAMPAGE

This diagram outlines a theoretical approach towards utilizing a modified single-molecule sequencing protocol to rapidly profile a transcriptome by directly sequencing individual mRNA transcript molecules. In the first step, total RNA is isolated from the cell or population of cells to be analyzed. The poly(A) tail possessed by most mRNA transcript molecules is used to hybridize individual mRNA molecules to the slide surface via complementary binding of a poly(T) DNA oligonucleotide that has already been tethered to the surface of the sequencing chamber. In scheme 1, this poly(T) DNA oligonucleotide serves as a primer for reverse transcriptase. Sequence tags are obtained by directly sequencing individual mRNA molecules using dye-labeled nucleotides and reverse transcriptase. Short read lengths, of approximately 10 to 20 bases, are presumably sufficient to identify a homolog in various public databases. Longer read lengths spanning exon/intron boundaries may permit the identification of transcriptional splice variants. In scheme 2, rather than directly sequencing the mRNA to obtain a sequence tag, a gene-specific labeled oligonucleotide probe is hybridized to the surface-bound mRNA transcript molecules in order to survey the population for a specific tag. This process may be repeated with another specific labeled oligonucleotide probe, and so on, to provide rapid and specific expression profiling when it is unnecessary to profile the entire population of molecules. Both schemes have a significant advantage in detecting rare transcripts because they are characterizing individual molecules. It is an open-ended

system, requiring no prior knowledge of the genome being analyzed. Furthermore, transcriptome quantitation is digital; the abundance of any given transcript is reflected by the number of sequence tags corresponding to that transcript. Although biotin and streptavidin are shown in the current example to facilitate surface binding, the protocol is not exclusively limited to this method of surface attachment. Any number of techniques may be used to tether the mRNA or the poly(T) primer oligonucleotide to the slide surface directly, or via intermediates. For example, the poly(T) oligonucleotide may be synthesized with an amine group stemming from its 3′ end and coupled directly to the carboxyl groups of polyacrylic acid, which are present on the outer layer of the sequencing chamber surface, using EDC.

Figure 4.2

4.5.2 Schematic Diagram of cDNA RAMPAGE

This diagram outlines a variant protocol for transcriptome profiling, using cDNA instead of mRNA, and forgoing all cloning and amplification steps that typically lie between cDNA production and sequence tag identification. This protocol exploits many of the features described in the mRNA RAMPAGE protocol (Figure 4.1). First, total RNA is isolated from the cell or population of cells to be analyzed and converted into cDNA using reverse transcriptase. Two schemes are presented for anchoring cDNA samples to the surface of the sequencing chamber. Scheme 1 involves ligation of a short double-stranded oligonucleotide, which in the current illustration arbitrarily contains series of A-T base pairs, to both ends of each cDNA molecule. Scheme 2 requires the enzyme

terminal deoxynucleotidetransferase (TDNT), which catalyzes the template-independent addition nucleotides to the 3′ end of double-stranded DNA. In the current example, incubation of cDNA with TDNT and dATP produces poly(A) 3′ ends. This poly(A) tail may be anchored to the sequencing chamber surface via complementary binding to a poly(T) DNA oligonucleotide much the same way as described in Figure 4.1. Furthermore, the poly(T) DNA oligonucleotide serves as a primer for DNA polymerase. Sequence tags are obtained by directly sequencing individual cDNA molecules using dye-labeled nucleotides and DNA polymerase.

Figure 4.3

4.5.3 Primer Extension Assay: Qiagen's Sensiscribe

Size separation of poly$(A)_{28}$ RNA primer extension reactions on a 12% (w/v) denaturing polyacrylamide gel imaged on a Typhoon 8600 scanning imager. Qiagen's Sensiscribe reverse transcriptase was used for all extension reactions containing reverse transcriptase. Fluorescent signal from the Cy3-labeled primer is detected. Lane 1 shows the unextended primed RNA template only. Lane 2 shows the positive control, a primer extension reaction using dTTP as reverse transcriptase substrate. Lane 3 shows the negative control in which no reverse transcriptase was added to the dTTP reaction. Lane 4 is a primer extension reactions using Cy5-dUTP as reverse transcriptase substrate. Lane 5 shows the negative control in which no reverse transcriptase was added to the Cy5-dUTP reaction.

Figure 4.4

4.5.4 Primer Extension Assay: Four Commercially Available Reverse Transcriptases

Size separation of poly$(A)_{28}$ RNA primer extension reactions on a 12% (w/v) denaturing polyacrylamide gel imaged on a Typhoon 8600 scanning imager. Four commercially available reverse transcriptases were assayed. Lane 1 shows the GeneScan ROX350 size standard (red). Numbers on the left indicate size in bases. Lanes 2–9 are primer extension reactions in which green represents fluorescence from the Cy3-labeled primer. Lanes 2–5 are negative controls and represent identical reaction conditions as those shown in Lanes 6–9, respectively, except that reverse transcriptase was withheld from the reaction. Lanes 6–9 show primer extension reactions using Cy5-dUTP as reverse transcriptase substrate. The following reverse transcriptases were used: AMV (lane 6), M-MuLV (lane 7), Sensiscribe (lane 8), and Superscript III (lane 9). All assayed reverse transcriptases appear to be able to extend the poly$(A)_{28}$ RNA template to full or near full length.

Figure 4.5

4.5.5 T7 DNA Polymerase 3-D Surface Model

Structural three-dimensional surface model of T7 DNA polymerase complexed with primer/template duplex DNA and ddATP (left). The active site of the enzyme (boxed, right), through which the template DNA must pass, appears much narrower than that of

reverse transcriptase (Figure 4.6). The relatively narrow active site of DNA polymerase may hinder the enzyme's ability to sequentially incorporate dye-labeled nucleotides due to steric interference with the growing DNA strand. The model was constructed using coordinates obtained through the Protein Data Bank (PDB accession code 1SKR).

Figure 4.6

4.5.6 HIV Reverse Transcriptase 3-D Surface Model

Structural three-dimensional surface model of HIV reverse transcriptase complexed with a 33-nucleotide RNA pseudoknot (left). The active site of the enzyme (boxed, right), through which the template nucleic acid must pass, appears wider than that of DNA polymerase (Figure 4.5). The more spacious active site of reverse transcriptase may facilitate the ability to sequentially incorporate dye-labeled nucleotides to a greater degree than that of DNA polymerase due to a reduced steric hindrance of the enzyme on the growing strand of DNA. The model was constructed using coordinates obtained through the Protein Data Bank (PDB accession code 1HVU).

Figure 4.7

4.5.7 Single-Molecule RNA Correlogram

Correlogram of the incorporation of Cy5-dUTP into a poly$(A)_{28}$ template by Superscript III reverse transcriptase. A maximum of 42 matches were found at one position with an average vector shift of 22.0000000 pixels along the x-axis and -26.0000000 pixels along

the y-axis.

Figure 4.8

4.5.8 Single-Molecule RNA Negative Control Correlogram

Negative control correlogram of the incorporation of Cy5-dUTP into a poly(A) template in the absence of reverse transcriptase. A maximum of seventeen matches were found at six positions with an average vector shift of -5.8333333 pixels along the x-axis and 3.1666667 pixels along the y-axis.

Figure 4.1

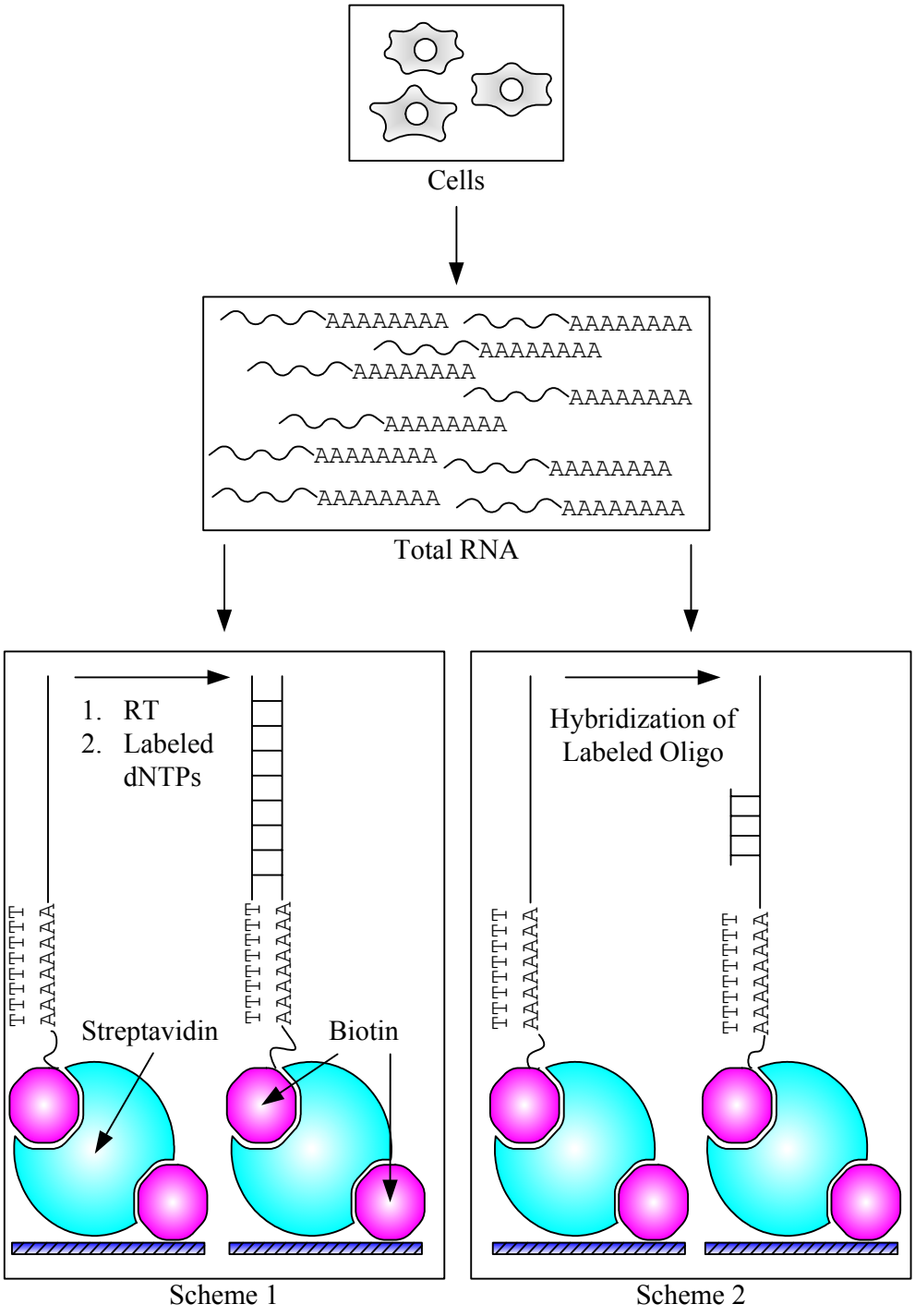

Cells
AAAAAAAA
Total RNA
1. RT
2. Labeled dNTPs
TTTTTTTT
AAAAAAAA
Streptavidin
Biotin
Scheme 1
Hybridization of Labeled Oligo
Scheme 2

Figure 4.2

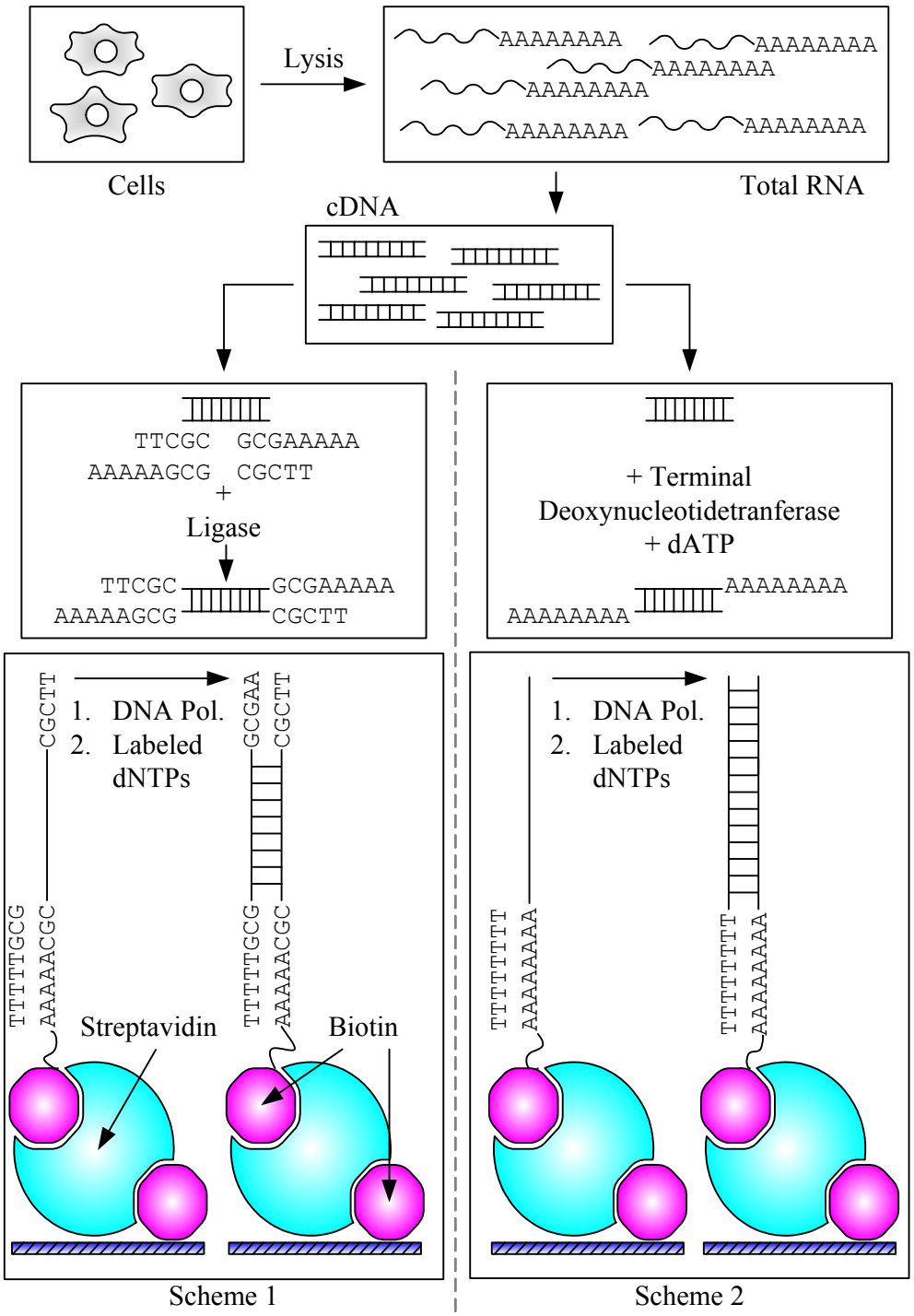

Lysis
AAAAAAAA
AAAAAAAA
AAAAAAAA
AAAAAAAA
AAAAAAAA
Cells
Total RNA
cDNA
TTCGC GCGAAAAA
AAAAAGCG CGCTT
+
Ligase
TTCGC GCGAAAAA
AAAAAGCG CGCTT
+ Terminal
Deoxynucleotidetranferase
+ dATP
AAAAAAAA
AAAAAAAA
1. DNA Pol.
2. Labeled
dNTPs
CGCTT
TTTTTGCG
AAAAACGC
GCGAA
CGCTT
TTTTTGCG
AAAAACGC
Streptavidin
Biotin
Scheme 1
1. DNA Pol.
2. Labeled
dNTPs
TTTTTTTT
AAAAAAAA
TTTTTTTT
AAAAAAAA
Scheme 2

Figure 4.3

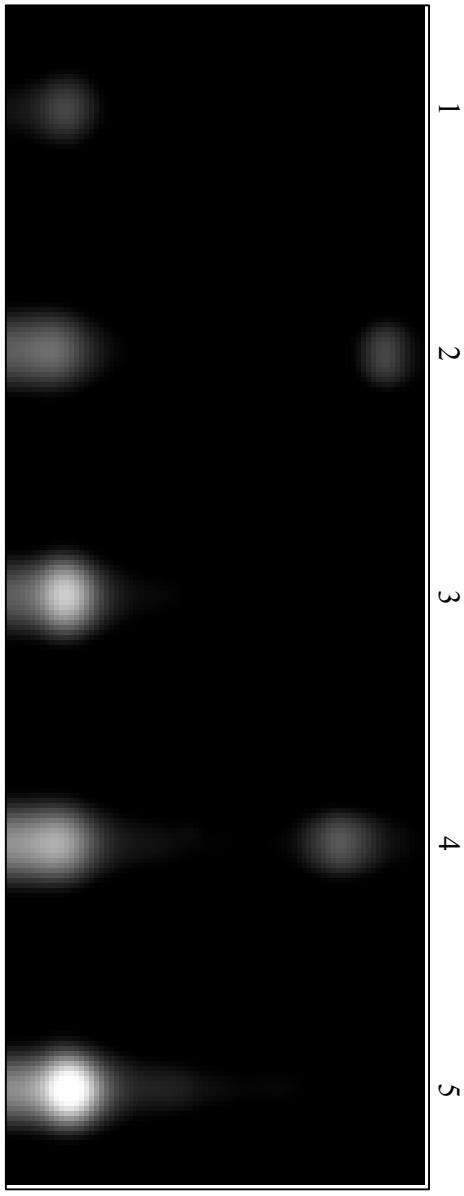


Figure 4.4

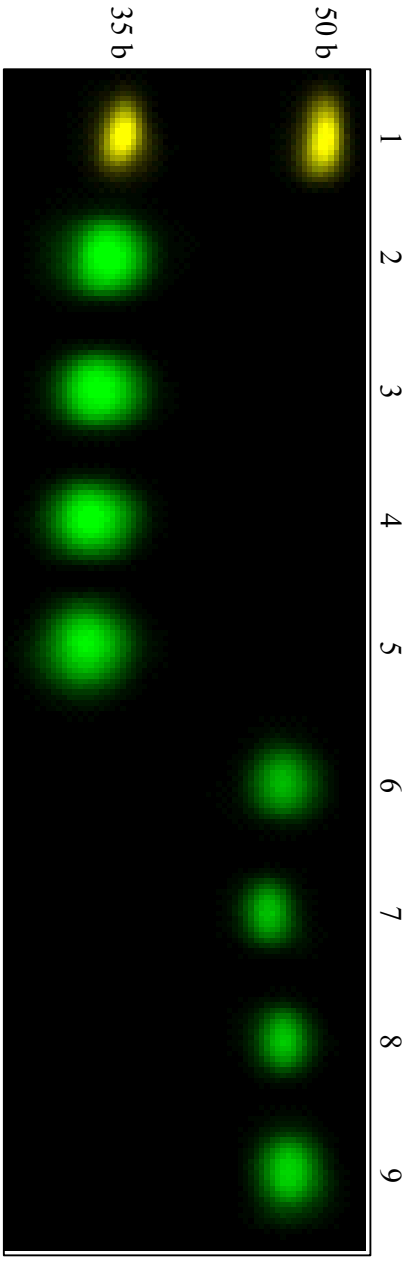


Figure 4.5

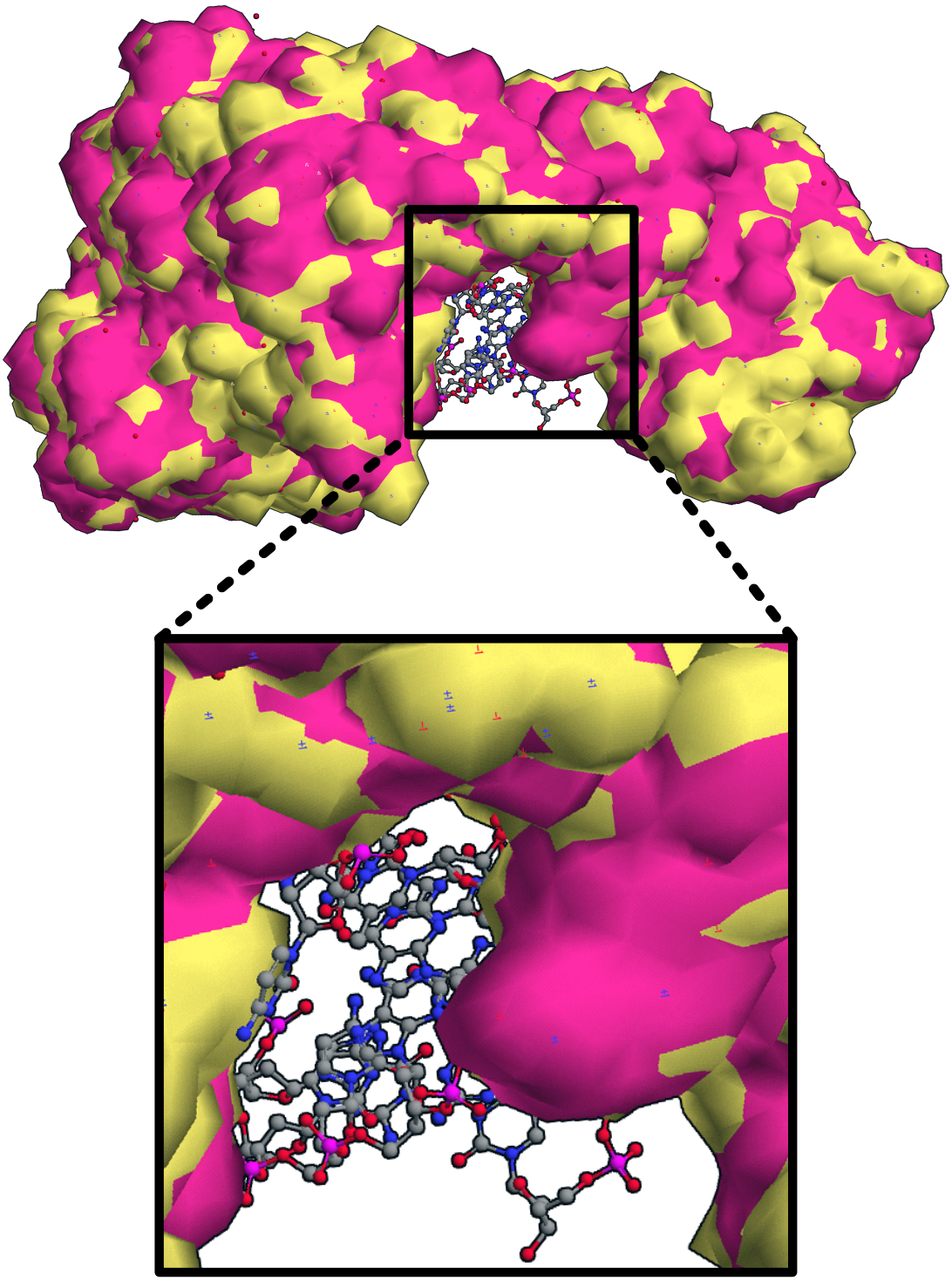


Figure 4.6

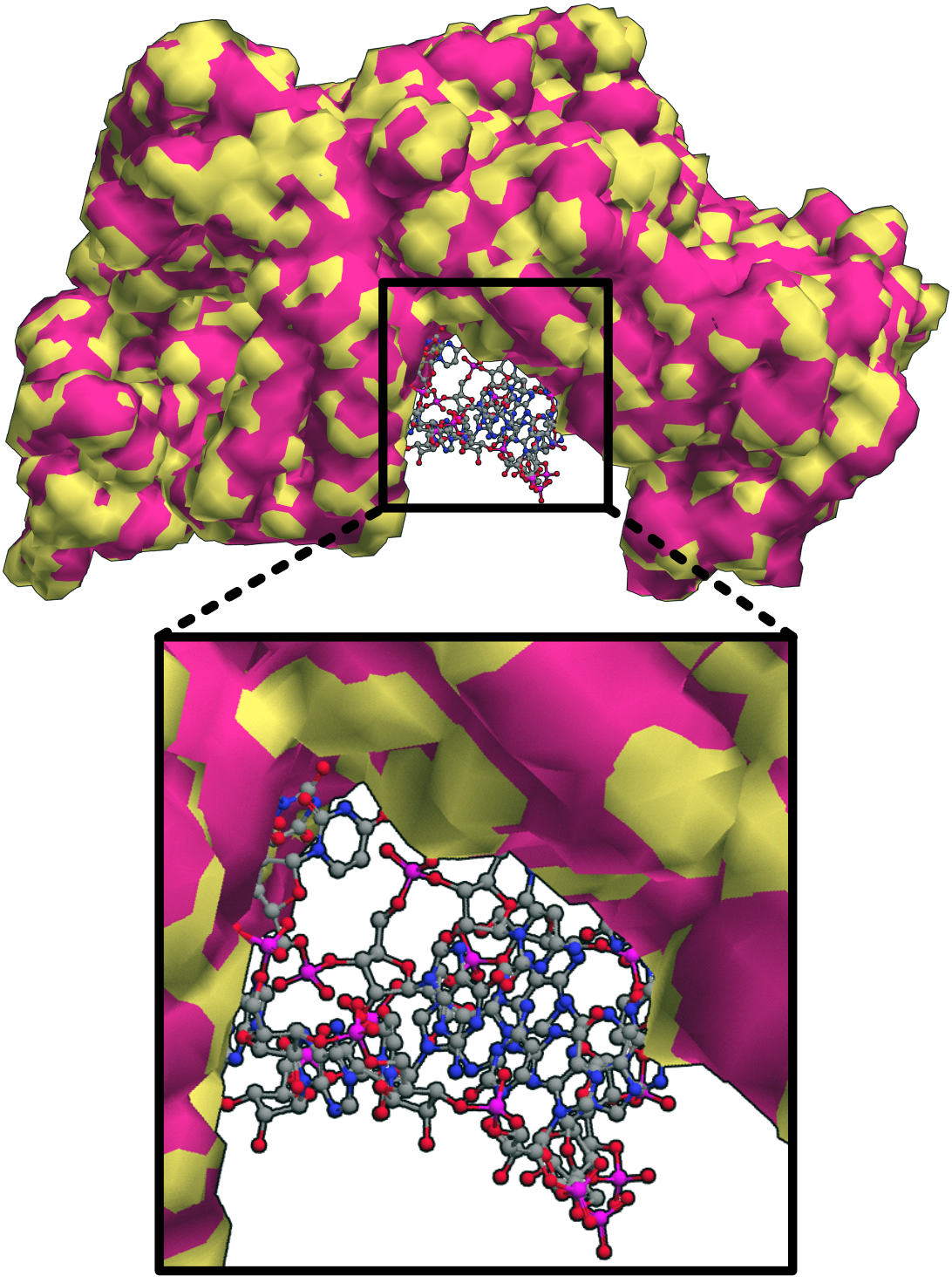


Figure 4.7

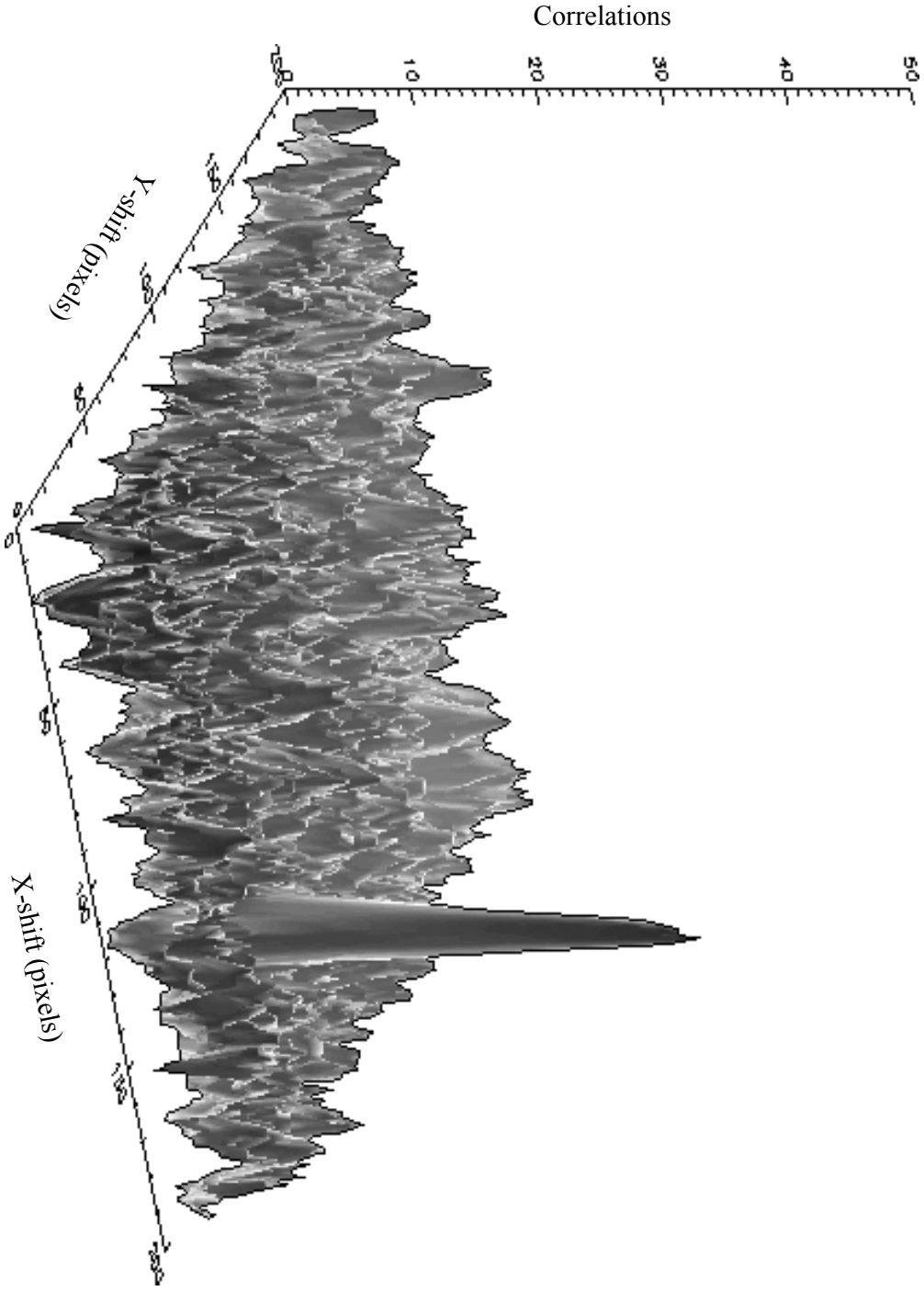
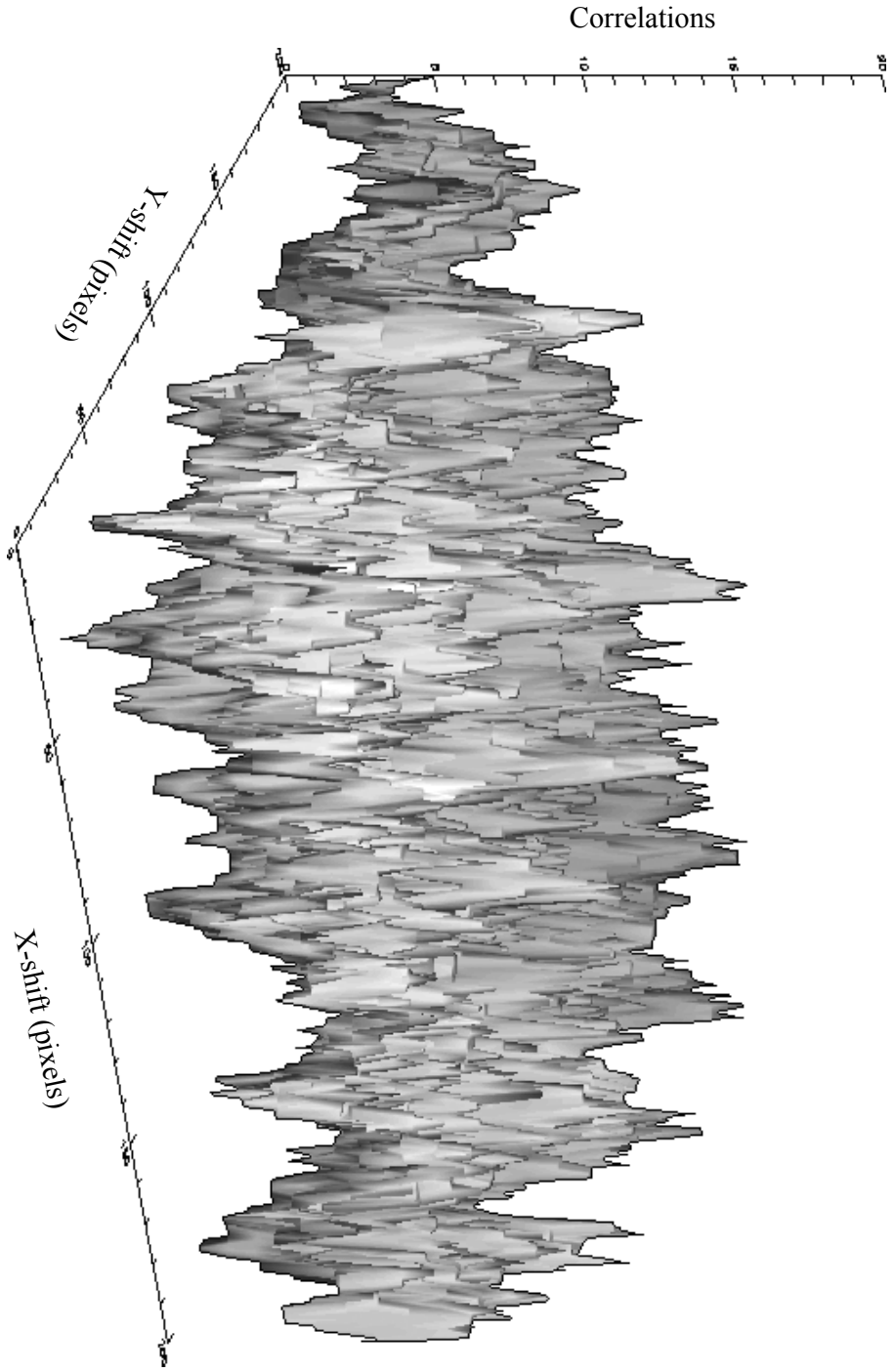


Figure 4.8



4.6 References

AUDIC, S., and J. M. CLAVERIE, 1997. The significance of digital gene expression profiles. Genome Res **7:** 986–995.

BOON, K., and G. J. RIGGINS, 2003. SAGE as a strategy to isolate cancer-related genes. Methods Mol Biol **222:** 463–479.

BRASLAVSKY, I., B. HEBERT, E. KARTALOV and S. R. QUAKE, 2003. Sequence information can be obtained from single DNA molecules. Proc Natl Acad Sci USA **100:** 3960–3964.

BRENNER, S., M. JOHNSON, J. BRIDGHAM, G. GOLDA, D. H. LLOYD *et al.*, 2000. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nat Biotechnol **18:** 630–634.

CROCKER, J. C., and D. G. GRIER, 1996. Methods of digital video microscopy for colloidal studies. Journal of Colloid and Interface Science **179:** 298.

DECHER, G., 1997. Fuzzy nanoassemblies: toward layered polymeric multicomposites. Science **277:** 1232–1237.

DECHERING, K. J., 2005. The transcriptome's drugable frequenters. Drug Discov Today **10:** 857–864.

DUGGAN, D. J., M. BITTNER, Y. CHEN, P. MELTZER and J. M. TRENT, 1999. Expression profiling using cDNA microarrays. Nat Genet **21:** 10–14.

EL-MEANAWY, M. A., S. BARATHAN, P. S. HAYDEN, S. K. IYENGAR, J. R. SCHELLING *et al.*, 2003. Serial analysis of gene expression. Methods Mol Med **86:** 257–274.

HACIA, J. G., 1999. Resequencing and mutational analysis using oligonucleotide microarrays. Nat Genet **21:** 42–47.

HAMLYN, P. H., G. G. BROWNIEE, C. C. CHENG, M. J. GAIT and C. MILSTEIN, 1978. Complete sequence of constant and 3′ noncoding regions of an immunoglobulin mRNA using the dideoxynucleotide method of RNA sequencing. Cell **15:** 1067–1075.

HAMLYN, P. H., M. J. GAIT and C. MILSTEIN, 1981. Complete sequence of an immunoglobulin mRNA using specific priming and the dideoxynucleotide method of RNA sequencing. Nucleic Acids Res **9:** 4485–4494.

KARTALOV, E. P., M. A. UNGER and S. R. QUAKE, 2003. Polyelectrolyte surface interface for single-molecule fluorescence studies of DNA polymerase. Biotechniques **34:** 505–510.

KITAMURA, N., and E. WIMMER, 1980. Sequence of 1060 3′-terminal nucleotides of poliovirus RNA as determined by a modification of the dideoxynucleotide method. Proc Natl Acad Sci USA **77:** 3196–3200.

LINDLÖF, A., 2003. Gene identification through large-scale EST sequence processing. Appl Bioinformatics **2:** 123–129.

LIU, E. T., 2004. Expression genomics and cancer biology. Pharmacogenomics **5:** 1117–1128.

SANGER, F., S. NICKLEN and A. R. COULSON, 1977. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA **74:** 5463–5467.

SCHENA, M., D. SHALON, R. W. DAVIS and P. O. BROWN, 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science **270:** 467–470.

UNGER, M., E. KARTALOV, C. S. CHIU, H. A. LESTER and S. R. QUAKE, 1999. Single-molecule fluorescence observed with mercury lamp illumination. Biotechniques **27:** 1008–1014.

VELCULESCU, V. E., L. ZHANG, B. VOGELSTEIN and K. W. KINZLER, 1995. Serial analysis of gene expression. Science **270:** 484–487.

WELLE, S., 2002. Gene transcript profiling in aging research. Exp Gerontol **37:** 583–590.

YILDIZ, A., J. N. FORKEY, S. A. MCKINNEY, T. HA, Y. E. GOLDMAN *et al.*, 2003. Myosin V walks hand-over-hand: single fluorophore imaging with 1.5-nm localization. Science **300:** 2061–2065.

ZHU, J., J. SHENDURE, R. D. MITRA and G. M. CHURCH, 2003. Single molecule profiling of alternative pre-mRNA splicing. Science **301:** 836–838.

ZIMMERN, D., and P. KAESBERG, 1978. 3′-terminal nucleotide sequence of encephalomyocarditis virus RNA determined by reverse transcriptase and chain-terminating inhibitors. Proc Natl Acad Sci USA **75:** 4257–4261.

5

Reagent Exchange Automation

5.1 Introduction

We set out to automate the exchange of reagents into the sequencing chamber of our single-molecule sequencing apparatus. It became clear that in order to transform our system into a practical technology, a programmable, user-free approach to reagent exchange would have to be implemented. Justification for such a system may be gleaned by taking a closer look at the reagent exchange steps of our current protocol for sequencing DNA.

First, after template DNA has been laid out onto the surface of the sequencing chamber, DNA polymerase and the labeled nucleotide are added and left to incubate for approximately 10 minutes. Next, the polymerase and nucleotide are washed out, followed by the addition of oxygen scavenger. After image acquisition, the oxygen scavenger is washed out and the dye molecules are bleached, for a total of four reagent exchange steps. It is important to note that we are only adding a single nucleotide at a time. Thus, 16 reagent exchange steps are required to interrogate the sequence of at least one base on each template molecule, given a diverse population of DNA templates.

Furthermore, if cleavable nucleotides are used, for example, a chemically cleavable nucleotide, two additional reagent exchange steps are added for each nucleotide. In this example, 24 reagent exchange steps are required to interrogate the sequence of at least one base on each template molecule, given a diverse template population.

Assuming that a read length of at least 30 bases from each template molecule is needed for sufficient overlap during shotgun sequencing assembly, up to 720 uninterrupted reagent exchange steps are necessary over a series of several days, in order to obtain all of the information needed to reconstruct an entire or partial genome. This large time requirement, along with the potential to introduce human error at each step, prompted us to design a system in which the entire process could be run automatically.

5.2 Results and Discussion

In designing the automated reagent exchange system, several factors needed to be taken into consideration. First, dead volume was to be kept to a minimum in order to minimize cross-contamination of reagents into the sequencing chamber as well as to efficiently utilize the full volume of each reagent as much as possible. Second, the reagents would likely need to be refrigerated. Additionally, we wanted to introduce as little change as possible into our sequencing system, keeping the flow cell and coverslip as is. Finally, the system needed to be able to exchange all reagents rapidly, switching between each reagent as needed.

We chose to use a series of syringe pumps connected to a multiport switch valve. A photograph of the reagent exchange system is shown in Figure 5.1. This series of syringe pumps were daisy chained together and were entirely programmable and controllable from a computer that was connected to the pumps via a serial port (RS-232) connection. Likewise, the switch valve was entirely programmable and controllable from the same computer using another, independent, serial port (RS-232) connection. Each circuit of

the system has a dead volume of less than 100 µl. Furthermore, the syringe pumps are small enough to fit into a small refrigerator equipped with tubing exit holes, if needed. Tubing was connected to the sequencing chamber using homemade press-on connectors fabricated from PDMS as described in the Materials and Methods. Software to control both the pumps and the switch valve was authored in LabVIEW. The graphical user interfaces of this software are shown in Figure 5.2.

As a proof-of-principle, we first sought to demonstrate the automation of reagent exchange using dilutions of simple food coloring dyes. The results of this demonstration are shown in Figure 5.3. In this demonstration, we were able to successfully introduce red food coloring into the sequencing chamber, followed by a wash step. We next introduced green food coloring into the sequencing chamber, which was also followed by a wash step. Each step flushed total volume of 500 µl through the flow cell at a rate of 1 ml per minute.

Cross-contamination can clearly be seen in frames 3 and 8 of Figure 5.3. This is likely due to the outlet port of the switch valve picking up a small volume of contaminating reagent, between its starting position and destination position, as it turns to align with the correct reagent port. Increasing the speed at which the switch valve module turns may reduce the amount of cross-contamination. Alternatively, introducing pressure regulated control valves in line with the switch valve and syringe pumps may prevent the residual accumulation of reagents into the inlet ports of the switch valve.

5.3 Concluding Remarks

In summary, we have demonstrated an automated reagent exchange system compatible with our single-molecule sequencing system. The results presented here are promising and represent a proof-of-principle demonstration using diluted food coloring dye. Further work is necessary in order to validate the automated reagent exchange system using genuine reagents during an actual sequencing experiment.

5.4 Materials and Methods

Syringe pumps were purchased from New Era Pumps, model NE-1000 and were controlled by a computer via a serial port (RS-232) connection. The switch valve was purchased from Hamilton (MVP, multivalve positioner) and was controlled by a computer via another, independent serial port (RS-232) connection. High-purity Teflon tubing (0.0625″ outer diameter, 0.003″ inner diameter) was purchased from Upchurch Scientific. The sequencing chamber, a Hybriwell flow cell, was purchased from Sigma. Press-on connectors, used to connect the switch valve to the inlet of the sequencing chamber and the outlet of the sequencing chamber to a waste container, were fashioned out of PDMS, nitrocellulose, and adhesive as described by KHOURY *et al.* 2002. Luer connectors were purchased from Upchurch Scientific and were used to connect each standard 10 ml syringe (B&D) to an inlet port of the switch valve. Fluid exchange of food coloring was performed using software authored in LabVIEW (Appendix II) for 30 seconds at each step and a nominal flow rate of 1 ml per minute.

5.5 Figure Legends

Figure 5.1

5.5.1 Automated Sequencing Apparatus

This photo is of the apparatus used for sequencing single molecules of nucleic acids. It is a through-the-objective, total internal reflection microscope equipped with a series of automated syringe pumps (right). Each pump connects to a unique inlet port of an automated switch valve (center), which in turn connects to the sequencing chamber (left). The valves as well as the syringe pumps are controlled via a personal computer using two independent serial port (RS-232) cables.

Figure 5.2

5.5.2 Graphical User Interfaces for Reagent Exchange Automation

Two GUIs (graphical user interfaces) written in LabVIEW control all aspects of reagent exchange. The GUI on the left programs specific parameters into the syringe pumps (e.g., flow rate, etc.) and initializes the switch valve. The GUI on the right is programmed with the sequence of commands needed to carry out a complex series of reagent exchange tasks. These include which pump or pumps are to fire and for how long, which outlet ports of the switch valve will open, which direction the valve turns in order to minimize cross-contamination, and how much time elapses before the next round of commands begins.

Figure 5.3

5.5.3 Demonstration of Automation

Images are extracted movie frames of a demonstration of the automated setup using diluted food coloring. A standard sequencing chamber was hooked up to the switch valve using homemade press-on connectors as described in the Materials and Methods. Automation was controlled by programming the pumps and valve using software authored in LabVIEW (Appendix II). Each exchange was for 30 seconds at a flow rate of 1 ml per minute.

Figure 5.1

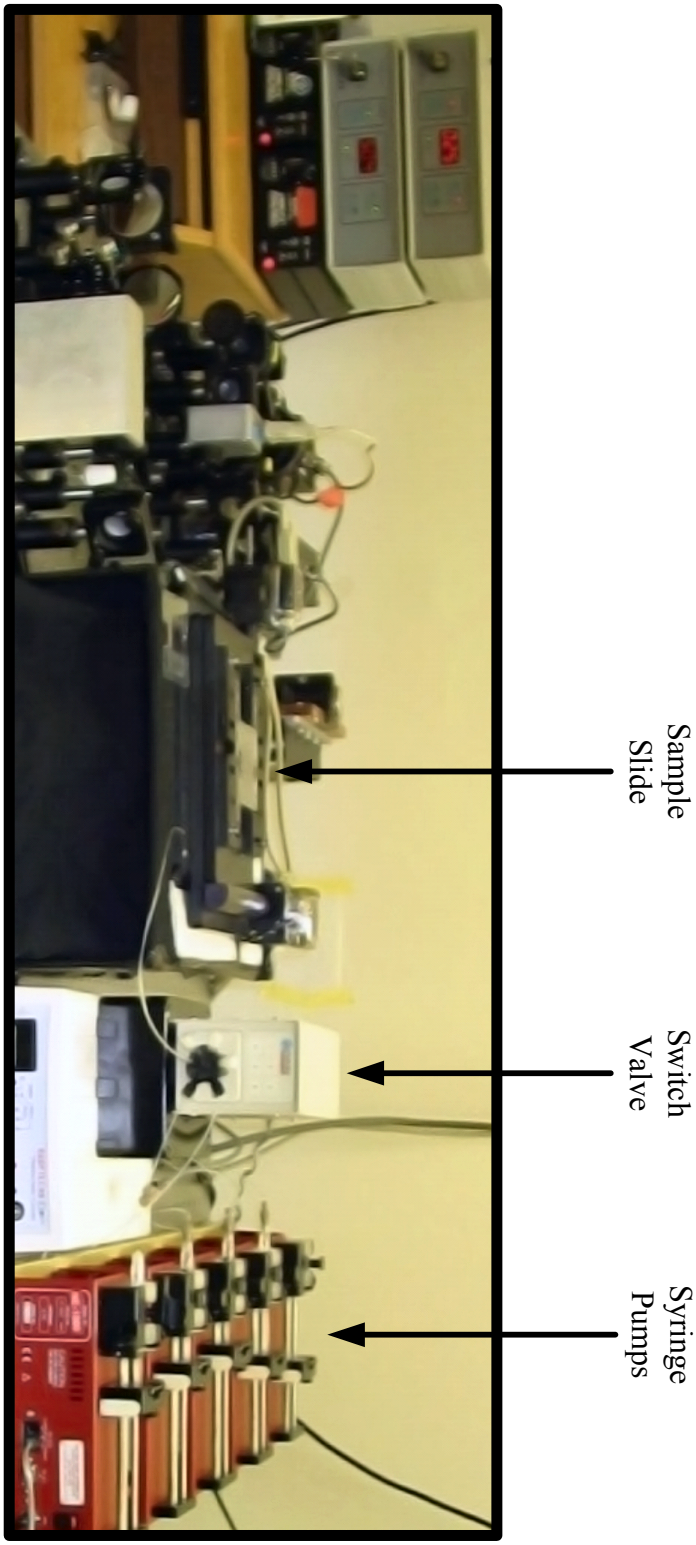


Figure 5.2

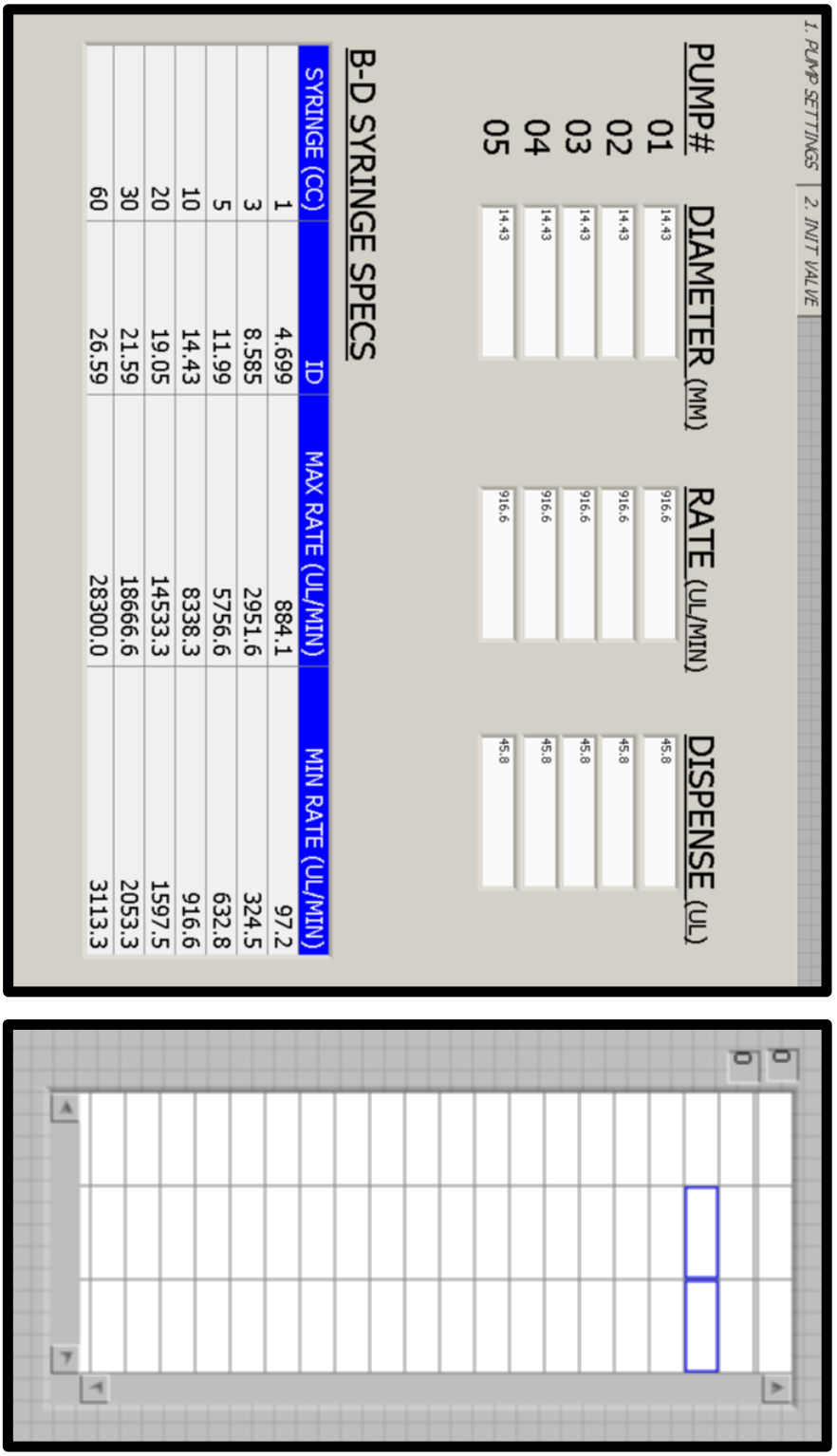
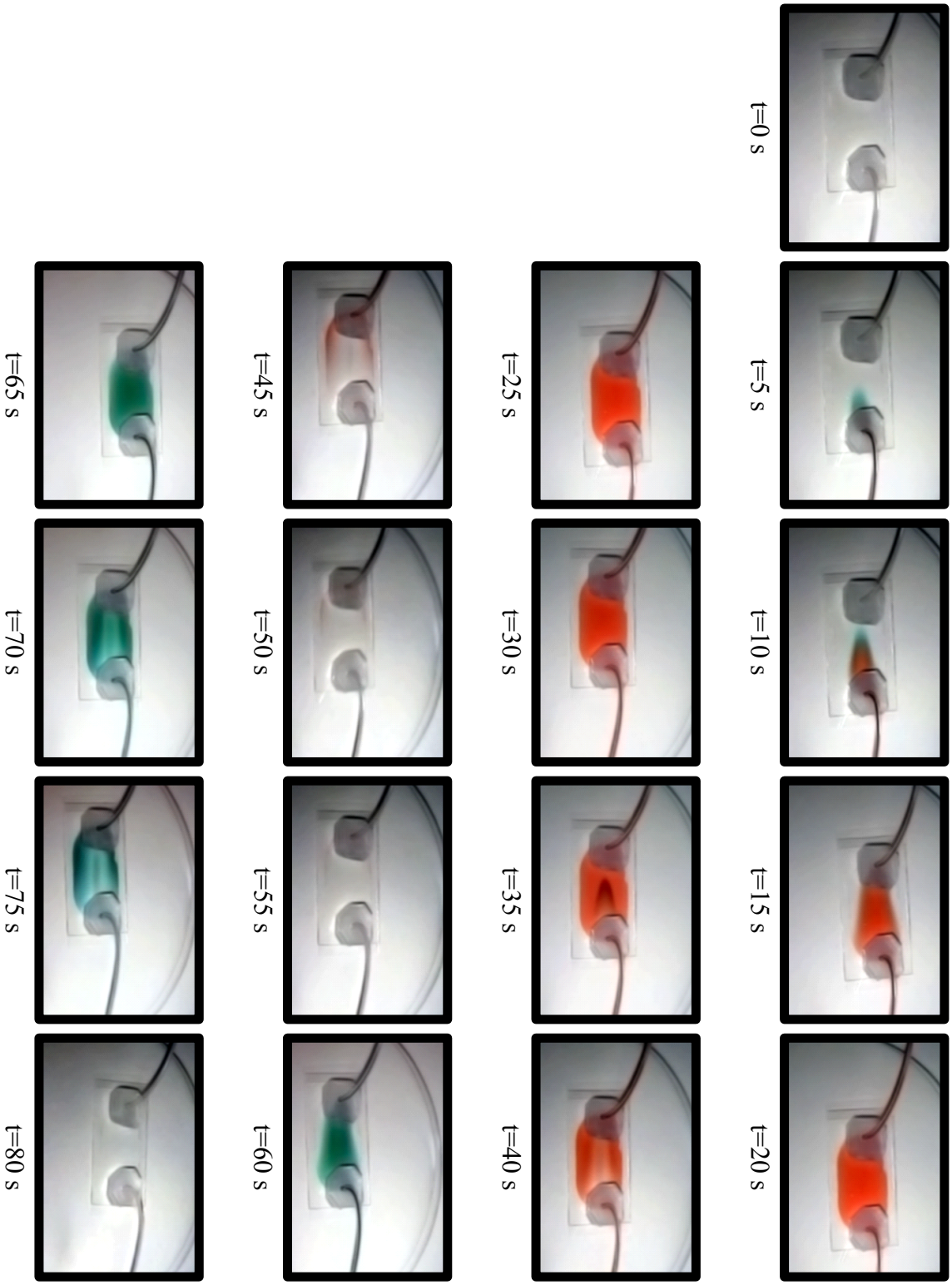

1. PUMP SETTINGS
2. INIT VALVE
PUMP#
DIAMETER (MM)
RATE (UL/MIN)
DISPENSE (UL)
01
02
03
04
05
14.43
916.6
45.8
B-D SYRINGE SPECS
SYRINGE (CC)
ID
MAX RATE (UL/MIN)
MIN RATE (UL/MIN)
1 4.699 884.1 97.2
3 8.585 2951.6 324.5
5 11.99 5756.6 632.8
10 14.43 8338.3 916.6
20 19.05 14533.3 1597.5
30 21.59 18666.6 2053.3
60 26.59 28300.0 3113.3

Figure 5.3


t=0 s
t=5 s
t=10 s
t=15 s
t=20 s
t=25 s
t=30 s
t=35 s
t=40 s
t=45 s
t=50 s
t=55 s
t=60 s
t=65 s
t=70 s
t=75 s
t=80 s

5.6 References

KHOURY, C., G. A. MENSING and D. J. BEEBE, 2002. Ultra rapid prototyping of microfluidic systems using liquid phase photopolymerization. Lab Chip **2:** 50–55.

6

Appendix I: IDL Routines

6.1 Summary

The IDL software routines necessary for single-molecule image processing and data analysis are presented here. Included are the following routines:

- all-in-one.pro
- bpass.pro
- feature.pro
- compare_pts5.pro

In most cases, a brief explanation can be found at the beginning of each routine regarding its parameters and use in IDL.

Aside from all-in-one.pro, these routines, as well as a helpful tutorial, may be found in their original form at: http://www.physics.emory.edu/~weeks/idl/

6.1.1 all-in-one.pro

```
;**********TYPICAL PARAMETERS**********
;**********COMPILE & RUN IN A SINGLE STEP**********
;**********CJL**********

;********** READ IMAGES **********

A0=READ_TIFF("C:\A.TIF")
B0=READ_TIFF("C:\B.TIF")

IMG1=DOUBLE(A0)
IMG2=DOUBLE(B0)

;********** BPASS **********

BPIMG1=BPASS(IMG1,1,11)
BPIMG2=BPASS(IMG2,1,11)

;********** GET FEATURES **********

FEAT1=FEATURE(BPIMG1,11, MASSCUT=7000)
FEAT2=FEATURE(BPIMG2,11, MASSCUT=7000)

;********** COMPARE FEATURES **********

COMPARE=COMPARE_PTS5(FEAT1[0:1,*],FEAT2[0:1,*],100,100,NSTEPS=200,RAD=3
)

END
```

6.1.2 bpass.pro

```
; NAME:
;     BPASS
; PURPOSE:
;     IMPLEMENTS A REAL-SPACE BANDPASS FILTER WHICH SUPPRESS
;     PIXEL NOISE AND SLOW-SCALE IMAGE VARIATIONS WHILE
;     RETAINING INFORMATION OF A CHARACTERISTIC SIZE.
;
; CATEGORY:
;     IMAGE PROCESSING
; CALLING SEQUENCE:
;     RES = DGFILTER( IMAGE, LNOISE, LOBJECT )
; INPUTS:
;     IMAGE:   THE TWO-DIMENSIONAL ARRAY TO BE FILTERED.
;     LNOISE: CHARACTERISTIC LENGTHSCALE OF NOISE IN PIXELS.
;      ADDITIVE NOISE AVERAGED OVER THIS LENGTH SHOULD
;      VANISH. MAY ASSUME ANY POSITIVE DOUBLEING VALUE.
;     LOBJECT: A LENGTH IN PIXELS SOMEWHAT LARGER THAN A TYPICAL
;      OBJECT. MUST BE AN ODD VALUED INTEGER.
; OUTPUTS:
;     RES: FILTERED IMAGE.
; PROCEDURE:
;     SIMPLE 'WAVELET' CONVOLUTION YIELDS SPATIAL BANDPASS FILTERING.
; NOTES:
; MODIFICATION HISTORY:
;     WRITTEN BY DAVID G. GRIER, THE UNIVERSITY OF CHICAGO, 2/93.
;     GREATLY REVISED VERSION DGG 5/95.
;     ADDED /FIELD KEYWORD JCC 12/95.
;     REVISED & ADDED 'STACK','VOXEL' CAPABILITY JCC 5/97.
;
;   THIS CODE 'BPASS.PRO' IS COPYRIGHT 1997, JOHN C. CROCKER AND
;   DAVID G. GRIER.  IT SHOULD BE CONSIDERED 'FREEWARE'- AND MAY BE
;   DISTRIBUTED FREELY IN ITS ORIGINAL FORM WHEN PROPERLY ATTRIBUTED.
;-
FUNCTION BPASS, IMAGE, LNOISE, LOBJECT, FIELD = FIELD, NOCLIP=NOCLIP

NF = N_ELEMENTS(IMAGE(0,0,*))

;ON_ERROR, 2          ; GO TO CALLER ON ERROR

B = DOUBLE( LNOISE )
W = ROUND( LOBJECT > (2. * B) )
N = 2*W + 1

R = (DINDGEN( N ) - W)/(2. * B)
XPT = EXP( -R^2 )
XPT = XPT / TOTAL(XPT)
FACTOR = ( TOTAL(XPT^2) - 1/N )

GX = XPT
GY = TRANSPOSE(GX)

BX = DBLARR(N) - 1./N
BY = TRANSPOSE(BX)
```

```
IF KEYWORD_SET( FIELD ) THEN BEGIN
    IF N MOD 4 EQ 1 THEN INDX = 2*INDGEN(W+1)$
         ELSE INDX = 1+ (2*INDGEN(W))
    GY = GY(INDX)
    GY = GY/TOTAL(GY)

    NN = N_ELEMENTS(INDX)
    BY = DBLARR(NN) - 1./NN
ENDIF

RES = DOUBLE(IMAGE)
; DO X AND Y CONVOLUTIONS
FOR I = 0,NF-1 DO BEGIN
    G = CONVOL( DOUBLE(IMAGE(*,*,I)), GX )
    G = CONVOL( G, GY )

    B = CONVOL( DOUBLE(IMAGE(*,*,I)), BX )
    B = CONVOL( B, BY )

    RES(*,*,I) = G-B
ENDFOR

IF KEYWORD_SET( NOCLIP ) THEN $
    RETURN,RES/FACTOR $
ELSE $
    RETURN,RES/FACTOR > 0

END

;*********** END OF BPASS.PRO
```

6.1.3 feature.pro

```
; NAME:
;     FEATURE
; PURPOSE:
;     FINDS AND MEASURES ROUGHLY CIRCULAR 'FEATURES' WITHIN
;     AN IMAGE.
; CATEGORY:
;     IMAGE PROCESSING
; CALLING SEQUENCE:
;     F = FEATURE( IMAGE, DIAMETER [, SEPARATION, MASSCUT = MASSCUT,
;      MIN = MIN, ITERATE = ITERATE, /FIELD, /QUIET ] )
; INPUTS:
;     IMAGE:   (NX,NY) ARRAY WHICH PRESUMABLY CONTAINS SOME
;      FEATURES WORTH FINDING
;     DIAMETER: A PARAMETER WHICH SHOULD BEE A LITTLE GREATER THAN
;      THE DIAMETER OF THE LARGEST FEATURES IN THE IMAGE.
;      DIAMETER MUST BE ODD VALUED.
;     SEPARATION: AN OPTIONAL PARAMETER WHICH SPECIFIES THE
;      MINIMUM ALLOWABLE SEPARATION BETWEEN FEATURE
;      CENTERS. THE DEFAULT VALUE IS DIAMETER+1.
;     MASSCUT:SETTING THIS PARAMETER SAVES RUNTIME BY REDUCING THE
;      RUNTIME WASTED ON LOW MASS 'NOISE' FEATURES.
;     MIN:     SET THIS OPTIONAL PARAMETER TO THE MINIMUM ALLOWED
;      VALUE FOR THE PEAK BRIGHTNESS OF A FEATURE. USEFUL
;      FOR LIMITING THE NUMBER OF SPURIOUS FEATURES IN
;      NOISY IMAGES.
;     FIELD:   SET THIS KEYWORD IF IMAGE IS ACTUALLY JUST ONE FIELD
;      OF AN INTERLACED (E.G. VIDEO) IMAGE. ALL THE MASKS
;      WILL THEN BE CONSTRUCTED WITH A 2:1 ASPECT RATIO.
;     QUIET:   SUPRESS PRINTING OF INFORMATIONAL MESSAGES.
;     ITERATE: IF THE REFINED CENTROID POSITION IS TOO FAR FROM
;      THE INITIAL ESTIMATE, ITERATIVELY RECALC. THE CENTROID
;      USING THE LAST CETROID TO POSITION THE MASK.  THIS
;      CAN BE USEFUL FOR REALLY NOISY DATA, OR DATA WITH
;      FLAT (E.G. SATURATED) PEAKS.  USE WITH CAUTION- IT
;      MAY 'CLIMB' HILLS AND GIVE YOU MULTIPLE HITS.
; OUTPUTS:
;     F(0,*):  THIS CONTAINS THE X CENTROID POSITIONS, IN PIXELS.
;     F(1,*): THIS CONTAINS THE Y CENTROID POSITIONS, IN PIXELS.
;     F(2,*): THIS CONTAINS THE INTEGRATED BRIGHTNESS OF THE
;      FEATURES.
;     F(3,*): THIS CONTAINS THE SQUARE OF THE RADIUS OF GYRATION
;      OF THE FEATURES.
;     F(4,*): THIS CONTAINS THE ECCENTRICITY, WHICH SHOULD BE
;      ZERO FOR CIRCULARLY SYMMETRIC FEATURES AND ORDER
;      ONE FOR VERY ELONGATED IMAGES.
; SIDE EFFECTS:
;     DISPLAYS THE NUMBER OF FEATURES FOUND ON THE SCREEN.
; RESTRICTIONS:
;     TO WORK PROPERLY, THE IMAGE MUST CONSIST OF BRIGHT,
;     CIRCULARLY SYMMETRIC REGIONS ON A ROUGHLY ZERO-VALUED
;     BACKGROUND. TO FIND DARK FEATURES, THE IMAGE SHOULD BE
;     INVERTED AND THE BACKGROUND SUBTRACTED. IF THE IMAGE
```

```
;     CONTAINS A LARGE AMOUNT OF HIGH SPATIAL FREQUENCY NOISE,
;     PERFORMANCE WILL BE IMPROVED BY FIRST FILTERING THE IMAGE.
;     BPASS WILL REMOVE HIGH SPATIAL FREQUENCY NOISE, AND
;     SUBTRACT THE IMAGE BACKGROUND AND THUS PROVIDES A GOOD
;     COMPLEMENT TO USING THIS PROGRAM. INDIVIDUAL FEATURES
;     SHOULD NOT OVERLAP.
; PROCEDURE:
;     FIRST, IDENTIFY THE POSITIONS OF ALL THE LOCAL MAXIMA IN
;     THE IMAGE ( DEFINED IN A CIRCULAR NEIGHBORHOOD WITH DIAMETER
;     EQUAL TO 'DIAMETER' ). AROUND EACH OF THESE MAXIMA, PLACE A
;     CIRCULAR MASK, OF DIAMETER 'DIAMETER', AND CALCULATE THE X & Y
;     CENTROIDS, THE TOTAL OF ALL THE PIXEL VALUES, AND THE RADIUS
;     OF GYRATION AND THE 'ECCENTRICITY' OF THE PIXEL VALUES WITHIN
;     THAT MASK. IF THE INITIAL LOCAL MAXIMUM IS FOUND TO BE MORE
;     THAN 0.5 PIXELS FROM THE CENTROID AND ITERATE IS SET, THE MASK
;     IS MOVED AND THE DATA ARE RE-CALCULATED. THIS IS USEFUL FOR
;     NOISY DATA. IF THE RESTRICTIONS ABOVE ARE ADHERED TO, AND THE
;     FEATURES ARE MORE THAN ABOUT 5 PIXELS ACROSS, THE RESULTING X
;     AND Y VALUES WILL HAVE ERRORS OF ORDER 0.1 PIXELS FOR
;     REASONABLY NOISE FREE IMAGES.
;
; *********        READ THE FOLLOWING IMPORTANT CAVEAT!
**********
;     'FEATURE' IS CAPABLE OF FINDING IMAGE FEATURES WITH SUB-PIXEL
;     ACCURACY, BUT ONLY IF USED CORRECTLY- THAT IS, IF THE
;     BACKGROUND IS SUBTRACTED OFF PROPERLY AND THE CENTROID MASK
;     IS LARGER THAN THE FEATURE, SO THAT CLIPPING DOES NOT OCCUR.
;     IT IS AN EXCELLENT IDEA WHEN WORKING WITH NEW DATA TO PLOT
;     A HISTOGRAM OF THE X-POSITIONS MOD 1, THAT IS, OF THE
;     FRACTIONAL PART OF X IN PIXELS.  IF THE RESULTING HISTOGRAM
;     IS FLAT, THEN YOU'RE OK, IF ITS STRONGLY PEAKED, THEN YOU'RE
;     DOING SOMETHING WRONG- BUT PROBABLY STILL GETTING 'NEAREST
;     PIXEL' ACCURACY.
;
;     FOR A MORE QUANTITATIVE TREATMENT OF SUB-PIXEL POSITION
;     RESOLUTION SEE:
;     J.C. CROCKER AND D.G. GRIER, J. COLLOID INTERFACE SCI.
;     *179*, 298 (1996).
;
; MODIFICATION HISTORY:
;     THIS CODE IS INSPIRED BY FEATURE_STATS2 WRITTEN BY
;      DAVID G. GRIER, U OF CHICAGO,         1992.
;     WRITTEN BY JOHN C. CROCKER, U OF CHICAGO, OPTIMIZING
;      RUNTIME AND MEASUREMENT ERROR,         10/93.
;     ADDED FIELD KEYWORD,                  4/94.
;     ADDED ECCENTRICITY PARAMETER,              5/95.
;     ADDED QUIET KEYWORD            12/95.
;     ADDED ITERATION, FIXED UP THE RADIUS/DIAMETER FIASCO AND
;     DID SOME DEBUGGING WHICH IMPROVES NON-CENTROID DATA.  4/96.
;
;   THIS CODE 'FEATURE.PRO' IS COPYRIGHT 1997, JOHN C. CROCKER AND
;   DAVID G. GRIER.  IT SHOULD BE CONSIDERED 'FREEWARE'- AND MAY BE
;   DISTRIBUTED FREELY IN ITS ORIGINAL FORM WHEN PROPERLY ATTRIBUTED.
;-
;
;   PRODUCE A PARABOLIC MASK
;
```

```
FUNCTION RSQD,W,H

IF N_PARAMS() EQ 1 THEN H = W
R2 = DBLARR(W,H,/NOZERO)
XC = DOUBLE(W-1) / 2.
YC = DOUBLE(H-1) / 2.
X = (DINDGEN(W) - XC)
X = X^2
Y = (DINDGEN(H)- YC)
Y = Y^2

FOR J = 0, H-1 DO BEGIN
    R2(*,J) = X + Y(J)
ENDFOR

RETURN,R2
END
;
;   PRODUCE A 'THETA' MASK
;
FUNCTION THETARR,W

THETA = DBLARR(W,W,/NOZERO)
XC = DOUBLE(W-1) / 2.
YC = DOUBLE(W-1) / 2.

X = (DINDGEN(W) - XC)
X(XC) = 1E-5
Y = (DINDGEN(W) - YC)

FOR J = 0, W-1 DO BEGIN
    THETA(*,J) = ATAN( Y(J),X )
ENDFOR

RETURN,THETA
END
;
;   THIS ROUTINE RETURNS THE EVEN OR ODD FIELD OF AN IMAGE
;
FUNCTION FIELDOF,ARRAY,ODD=ODD,EVEN=EVEN

SZ = SIZE(ARRAY)
IF SZ(0) NE 2 THEN MESSAGE,"ARGUMENT MUST BE A TWO-DIMENSIONAL ARRAY!"
IF KEYWORD_SET(ODD) THEN F=1 ELSE F=0

NY2 = FIX( (SZ(2)+(1-F))/2 )
ROWS = INDGEN(NY2)*2 + F
RETURN,ARRAY(*,ROWS)

END
;
;   BARREL "SHIFTS" A DOUBLEING POINT ARR BY A FRACTIONAL PIXEL AMOUNT,
;     BY USING A 'LEGO' INTERPOLATION TECHNIQUE.
;
FUNCTION FRACSHIFT,IM,SHIFTX,SHIFTY

IPX = FIX( SHIFTX )
```

```
IPY = FIX( SHIFTY )
FPX = SHIFTX - IPX
FPY = SHIFTY - IPY
IF FPX LT 0 THEN BEGIN
    FPX=FPX+1 & IPX=IPX-1
ENDIF
IF FPY LT 0 THEN BEGIN
    FPY=FPY+1 & IPY=IPY-1
ENDIF

IMAGE = IM

IMAGEX  = SHIFT( IMAGE,IPX+1,IPY   )
IMAGEY  = SHIFT( IMAGE,IPX  ,IPY+1 )
IMAGEXY = SHIFT( IMAGE,IPX+1,IPY+1 )
IMAGE   = SHIFT( IMAGE,IPX  ,IPY   )

RES   = ( (1. - FPX) * (1. - FPY) * IMAGE   ) + $
    ( (     FPX) * (1. - FPY) * IMAGEX  ) + $
    ( (1. - FPX) * (     FPY) * IMAGEY  ) + $
    ( (     FPX) * (     FPY) * IMAGEXY )

RETURN,RES
END
;
;   JOHN'S VERSION OF LOCAL_MAX2, WHICH SUPPORTS THE FIELD KEYWORD
;   AND IS OTHERWISE IDENTICAL.
;
FUNCTION LMX, IMAGE, SEP, MIN = MIN, FIELD = FIELD

RANGE = FIX(SEP/2)
A = BYTSCL(IMAGE)
W = ROUND( 2 * RANGE + 1 )  ; WIDTH OF SAMPLE REGION
S = RSQD( W )       ; SAMPLE REGION IS CIRCULAR
GOOD = WHERE( S LE RANGE^2 )
MASK = BYTARR( W, W )
MASK(GOOD) = 1B
YRANGE = RANGE
IF KEYWORD_SET( FIELD ) THEN BEGIN
    MASK = FIELDOF( MASK, /EVEN )
    YRANGE = FIX(RANGE/2.) +1
ENDIF

B = DILATE( A, MASK, /GRAY )    ; FIND LOCAL MAXIMA IN GIVEN RANGE

          ; BUT DON'T INCLUDE PIXELS FROM THE
          ; BACKGROUND WHICH WILL BE TOO DIM
IF NOT KEYWORD_SET( MIN ) THEN BEGIN
    H = HISTOGRAM( A )
    FOR I = 1, N_ELEMENTS(H) - 1 DO $
       H(I) = H(I) + H(I-1)
    H = DOUBLE( H ) / MAX( H )
    MIN = 0
    WHILE H(MIN) LT 0.64 DO $
       MIN = MIN + 1
    MIN = MIN + 1
    ENDIF
```

```
R = WHERE( A EQ B AND A GE MIN )

          ; DISCARD MAXIMA WITHIN RANGE OF THE EDGE
SZ = SIZE( A )
NX = SZ(1) & NY = SZ(2)
X = R MOD NX & Y = R / NX
X0 = X - RANGE & X1 = X + RANGE
Y0 = Y - YRANGE & Y1 = Y + YRANGE
GOOD = WHERE( X0 GE 0 AND X1 LT NX AND Y0 GE 0 AND Y1 LT NY,NGOOD )
IF NGOOD LT 1 THEN $
    RETURN,[-1]
R = R(GOOD)
X = X(GOOD) & Y = Y(GOOD)
X0 = X0(GOOD) & X1 = X1(GOOD)
Y0 = Y0(GOOD) & Y1 = Y1(GOOD)
          ; THERE MAY BE SOME FEATURES WHICH GET
          ; FOUND TWICE OR WHICH HAVE FLAT PEAKS
          ; AND THUS PRODUCE MULTIPLE HITS.  FIND
          ; AND CLEAR SUCH SPURIOUS POINTS.
C = 0B * A
C(R) = A(R)
CENTER = W * YRANGE + RANGE
FOR I = 0D, N_ELEMENTS(R) - 1D DO BEGIN
    B = C(X0(I):X1(I),Y0(I):Y1(I))

    B =  B * MASK   ; LOOK ONLY IN CIRCULAR REGION
    M = MAX( B, LOCATION )
    IF LOCATION NE CENTER THEN $
       C(X(I),Y(I)) = 0B
    ENDFOR
          ; IDEALLY, THE ABOVE ROUTINE WOULD SHRINK
          ; CLUSTERS OF POINTS DOWN TO THEIR CENTER.
          ; AS WRITTEN, THIS WILL LEAVE THE LOWER
          ; RIGHT (?) PIXEL OF A CLUSTER.

R = WHERE( C NE 0 )     ; WHAT'S LEFT ARE VALID MAXIMA.

RETURN,R         ; RETURN THEIR LOCATIONS

END
;
;   JOHN'S VERSION OF DGG'S FEATURE_STATS2.
;   WHICH: A) AVOIDS SOME UNNECESSARY COMPUTATION (CONVOLUTIONS)
;          B) USES FRACTIONAL SHIFT TECHNIQUES TO REDUCE PIXEL BIAS IN
M AND RG
;          C) HAS THE FIELD KEYWORD
;
FUNCTION FEATURE, IMAGE, EXTENT, SEP,  MIN=MIN, MASSCUT = MASSCUT,
FIELD = FIELD,$
     QUIET = QUIET, ITERATE = ITERATE, EXCENCUT = EXCENCUT, ONEFEATURE
= ONEFEATURE, $
     RADGCUT = RADGCUT

EXTENT = FIX(EXTENT)
IF (EXTENT MOD 2) EQ 0 THEN BEGIN
    MESSAGE,'REQUIRES AN ODD EXTENT.  ADDING 1...',/INF
```

```
    EXTENT = EXTENT + 1
    ENDIF

SZ = SIZE( IMAGE )
NX = SZ(1)
NY = SZ(2)
IF N_PARAMS() EQ 2 THEN SEP = EXTENT+1

;   PUT A BORDER AROUND THE IMAGE TO PREVENT MASK OUT-OF-BOUNDS
A = DBLARR( NX + EXTENT, NY + EXTENT )
A(EXTENT/2:(EXTENT/2)+NX-1,EXTENT/2:(EXTENT/2)+NY-1) = DOUBLE( IMAGE )
NX = NX + EXTENT

;   FINDING LOCAL MAXIMA
IF KEYWORD_SET( FIELD ) THEN $
    IF NOT KEYWORD_SET( MIN ) THEN LOC = LMX(A,SEP,/FIELD) ELSE $
       LOC=LMX(A,SEP,MIN=MIN,/FIELD) $
ELSE $
    IF NOT KEYWORD_SET( MIN ) THEN LOC = LMX(A,SEP) ELSE $
       LOC=LMX(A,SEP,MIN=MIN)

IF LOC(0) EQ -1 THEN RETURN,-1
X  = DOUBLE( LOC MOD NX )
Y  = DOUBLE( LOC / NX )

NMAX=N_ELEMENTS(LOC)
XL = X - FIX(EXTENT/2)
XH = XL + EXTENT -1
M  = DBLARR(NMAX)

;   SET UP SOME MASKS
RSQ = RSQD( EXTENT )
T = THETARR( EXTENT )
MASK = RSQ LE (DOUBLE(EXTENT)/2.)^2
MASK2 = MAKE_ARRAY( EXTENT , EXTENT , /DOUBLE, /INDEX ) MOD (EXTENT ) +
1.
MASK2 = MASK2 * MASK
MASK3= (RSQ * MASK) + (1./6.)
CEN = DOUBLE(EXTENT-1)/2.
CMASK = COS(2*T) * MASK
SMASK = SIN(2*T) * MASK
CMASK(CEN,CEN) = 0.
SMASK(CEN,CEN) = 0.

;   EXTRACT FIELDS OF THE MASKS, IF NECESSARY
IF KEYWORD_SET( FIELD ) THEN BEGIN
    SUBA = DBLARR(EXTENT , FIX(EXTENT/2) , NMAX)
    MASK = FIELDOF(MASK,/ODD)
    XMASK = FIELDOF(MASK2,/ODD)
    YMASK = FIELDOF(TRANSPOSE(MASK2),/ODD)
    MASK3 = FIELDOF(MASK3,/ODD)
    CMASK = FIELDOF(CMASK,/ODD)
    SMASK = FIELDOF(SMASK,/ODD)

    HALFEXT = FIX( EXTENT /2 )
    YL = Y - FIX(HALFEXT/2)
    YH = YL + HALFEXT -1
```

```
    YSCALE = 2
    YCEN = CEN/2

ENDIF ELSE BEGIN
    SUBA = DBLARR(EXTENT, EXTENT, NMAX)
    XMASK = MASK2
    YMASK = TRANSPOSE( MASK2 )

    YL = Y - FIX(EXTENT/2)
    YH = YL + EXTENT -1
    YSCALE = 1
    YCEN = CEN
ENDELSE

;   ESTIMATE THE MASS
FOR I=0,NMAX-1 DO M(I) = TOTAL( A(XL(I):XH(I),YL(I):YH(I)) * MASK )
REJECT=0
IF KEYWORD_SET( MASSCUT ) THEN BEGIN
    W = WHERE( M GT MASSCUT, NMAX )
    IF KEYWORD_SET(ONEFEATURE) THEN BEGIN
       W = WHERE( M EQ MAX(M), NMAX )
       IF NMAX EQ 0 THEN BEGIN
         MESSAGE,'NO FEATURES FOUND!',/INF
         RETURN,-1
       ENDIF
       XL = XL(W)
       XH = XH(W)
       YL = YL(W)
       YH = YH(W)
       X = X(W)
       Y = Y(W)
       M = M(W)
       IF (MAX(M) LE MASSCUT) THEN BEGIN
         REJECT=1
         PRINT,'BACK TO MIDDLE'
         XL = FIX(NX/2)-EXTENT/2
         XH = XL+EXTENT-1
         YL = FIX(NY/2)-EXTENT/2
         YH = YL+EXTENT-1
         X = [(NX-EXTENT)/2.]
         Y = [NY/2.]
         M = [TOTAL( A(XL:XH,YL:YH) * MASK )]
       ENDIF
    ENDIF ELSE BEGIN
       IF NMAX EQ 0 THEN BEGIN
         MESSAGE,'NO FEATURES FOUND!',/INF
         RETURN,-1
       ENDIF
       XL = XL(W)
       XH = XH(W)
       YL = YL(W)
       YH = YH(W)
       X = X(W)
       Y = Y(W)
       M = M(W)
    ENDELSE
ENDIF
```

```
IF NOT KEYWORD_SET(QUIET) THEN MESSAGE, STRCOMPRESS( NMAX ) + '
FEATURES FOUND.',/INF

;   SETUP SOME RESULT ARRAYS
XC = DBLARR(NMAX)
YC = DBLARR(NMAX)
RG = DBLARR(NMAX)
E  = DBLARR(NMAX)

;   CALCULATE FEATURE CENTERS
FOR I=0,NMAX-1 DO BEGIN
    XC(I) = TOTAL( A(XL(I):XH(I),YL(I):YH(I)) * XMASK )
    YC(I) = TOTAL( A(XL(I):XH(I),YL(I):YH(I)) * YMASK )
ENDFOR

;   CORRECT FOR THE 'OFFSET' OF THE CENTROID MASKS
XC = XC / M - ((DOUBLE(EXTENT)+1.)/2.)
YC = (YC / M - ((DOUBLE(EXTENT)+1.)/2.)) / YSCALE

;   ITERATE ANY BAD INITIAL ESTIMATE.
IF KEYWORD_SET( ITERATE ) THEN BEGIN
COUNTER = 0
REPEAT BEGIN
    COUNTER = COUNTER + 1

    W = WHERE( ABS(XC) GT 0.6, NBADX )
    IF NBADX GT 0 THEN BEGIN
      DX = ROUND( XC(W) )
       XL(W) = XL(W) + DX
       XH(W) = XH(W) + DX
       X(W) = X(W) + DX
    ENDIF
    W = WHERE( ABS(YC) GT 0.6, NBADY )
    IF NBADY GT 0 THEN BEGIN
      DY = ROUND( YC(W) )
       YL(W) = YL(W) + DY
       YH(W) = YH(W) + DY
       Y(W) = Y(W) + DY
    ENDIF

    W = WHERE( (ABS(XC) GT 0.6) OR (ABS(YC) GT 0.6), NBAD )
    IF NBAD GT 0 THEN BEGIN     ; RECALCULATE THE CENTROIDS FOR THE
GUYS WE'RE ITERATING

       FOR I=0,NBAD-1 DO M(W(I)) = TOTAL( A(XL(W(I)):XH(W(I)),$
         YL(W(I)):YH(W(I))) * MASK )
       FOR I=0,NBAD-1 DO BEGIN
         XC(W(I)) = TOTAL( A(XL(W(I)):XH(W(I)),YL(W(I)):YH(W(I))) *
XMASK )
         YC(W(I)) = TOTAL( A(XL(W(I)):XH(W(I)),YL(W(I)):YH(W(I))) *
YMASK )
       ENDFOR

       XC(W) = XC(W) / M(W) - ((DOUBLE(EXTENT)+1.)/2.)
       YC(W) = ( YC(W) / M(W) - ((DOUBLE(EXTENT)+1.)/2.) ) / YSCALE
```

```
    ENDIF
ENDREP UNTIL (NBAD EQ 0) OR (COUNTER EQ 10)
ENDIF

IF (REJECT NE 1) THEN BEGIN
    ;  UPDATE THE POSITIONS AND CORRECT FOR THE WIDTH OF THE 'BORDER'
    X = X + XC - EXTENT/2
    Y = ( Y + YC - EXTENT/2 ) * YSCALE
    ;  CONSTRUCT THE SUBARRAY
    FOR I=0,NMAX-1 DO SUBA(*,*,I) = FRACSHIFT(
A(XL(I):XH(I),YL(I):YH(I)), -XC(I) , -YC(I) )

    ;  CALCULATE THE 'MASS'
    FOR I=0,NMAX-1 DO M(I) = TOTAL( SUBA(*,*,I) * MASK )
ENDIF
;   CALCULATE RADII OF GYRATION SQUARED
FOR I=0,NMAX-1 DO RG(I) = TOTAL( SUBA(*,*,I) * MASK3 ) / M(I)

;   CALCULATE THE 'ECCENTRICITY'
FOR I=0,NMAX-1 DO E(I) = SQRT(( TOTAL( SUBA(*,*,I) * CMASK )^2 ) +$
    ( TOTAL( SUBA(*,*,I) * SMASK )^2 )) / (M(I)-SUBA(CEN,YCEN,I)+1E-6)

;OPTIONAL 'EXCENCUT' KEYWORD SET BY BEN (09/13/2001)
;TO ALLOW CUTTING WITH EXCENTRICITY
IF KEYWORD_SET( EXCENCUT ) THEN BEGIN
    W = WHERE( E LT EXCENCUT, NMAX )
    IF NMAX EQ 0 THEN BEGIN
       MESSAGE,'NO FEATURES FOUND!',/INF
       RETURN,-1
    ENDIF
    ;XL = XL(W)
    ;XH = XH(W)
    ;YL = YL(W)
    ;YH = YH(W)
    X = X(W)
    Y = Y(W)
    M = M(W)
    RG = RG(W)
    E = E(W)
ENDIF

;OPTIONAL 'RADGCUT' KEYWORD SET BY BEN (09/28/2001)
;TO ALLOW CUTTING WITH RADIUS OF GIRATION
IF KEYWORD_SET( RADGCUT ) THEN BEGIN
    W = WHERE( RG LT RADGCUT, NMAX )
    IF NMAX EQ 0 THEN BEGIN
       MESSAGE,'NO FEATURES FOUND!',/INF
       RETURN,-1
    ENDIF
    ;XL = XL(W)
    ;XH = XH(W)
    ;YL = YL(W)
    ;YH = YH(W)
    X = X(W)
    Y = Y(W)
    M = M(W)
    RG = RG(W)
```

```
    E = E(W)
ENDIF

PARAMS =
[TRANSPOSE(X),TRANSPOSE(Y),TRANSPOSE(M),TRANSPOSE(RG),TRANSPOSE(E)]
RETURN,PARAMS
END
```

6.1.4 compare_pts5.pro

```
;THIS VERSION 5.0 FINDS NEAREST NEIGHBORS AT EVERY SHIFT
;SO IT IS LONGER, BUT DEFINITIVELY BETTER!
;
;THIS FUNCTION COMPARES 2 ARRAYS OF POINTS AND FINDS THE
;CORRELATED ONES BY SHIFTING THE SECOND IMAGE BY A RANGE
;OF DIFFERENT VECTORS. THE INPUTS ARE:
;   PTT1: FIRST ARRAY OF POINTS (N,*) OF WHICH (0:1,*) IS
;        THE X AND Y POINTS COORDINATES
;   PTT2: SECOND ARRAY OF POINTS (M,*)
;   DELX: THE RANGE OF VECTOR SHIFT IN X (FROM -DELX/2.0
;        TO DELX/2.0)
;   DELY: THE RANGE OF VECTOR SHIFT IN Y (FROM -DELY/2.0
;        TO DELY/2.0)
;THE OPTIONAL KEYWORDS ARE:
;   NSTEPS: THE NUMBER OF VECTORIAL STEPS
;   RAD: THE RADIUS WITHTIN WHICH YOU LOOK FOR CORRELATED
;       POINTS
;   INISHIFT: AN INITIAL SHIFT YOU'D LIKE TO GIVE TO THE
;            FIRST DATA SET
;   CHECKPOS: A PARTICULAR TRANSLATION VECTOR AT WHICH YOU
;            WANT TO KNOW THE NUMBER OF MATCHES
;   RETVEC: RETURNS THE TRANSLATION VECTOR FROM SET2 TO SET1
;   HOWLONG: SET THIS KEYWORD TO GET AN ESTIMATE OF THE RUN
;            TIME OF THE ROUTINE (AND AN OPTION TO STOP!)
;   PRINTNEG: PRINTS THE NEGATIVES OF VECTORS, NEVER USE IT
;             IS ONLY MEANT FOR THE FUNCTION CALLING ITSELF
;             WHEN THE INPUTS NEED TO BE SWITCHED...
;THE OUTPUT IS:
;   RESARR: AN ARRAY (N+M,*) OF WHICH (0:1,*) IS THE X AND Y
;          COORDS OF THE CORRELATED POINTS FROM THE 1ST SET,
;          AND (N:N+1,*) IS THE X AND Y OORDS OF THE
;          CORRELATED POINTS FROM THE 2ND SET.
;
;
;***-----------------------------------------------------***
;NOTE:
;INISHIFT IS THE SHIFT THAT'S GOING TO BE ADDED TO THE FIRST
;DATA SET, WHILE THE OUTPUT (IF "RETVEC" IS SET TO 1) IS THE
;LOCATION OF THE MAXIMUM OF THE CORRELATION PEAK, AND IS GIVEN
;AS A SHIFT THAT'S GOING TO BE ADDED TO THE SECOND DATA SET
;***-----------------------------------------------------***
;

FUNCTION
COMPARE_PTS5,PTT1,PTT2,DELX,DELY,NSTEPS=NSTEPS,RAD=RAD,PRINTPROGRESS=PR
INTPROGRESS, $
          INISHIFT=INISHIFT, CHECKPOS=CHECKPOS, RETVEC=RETVEC,
HOWLONG=HOWLONG, PRINTNEG=PRINTNEG

IF NOT KEYWORD_SET(RAD) THEN RAD=2.0
IF NOT KEYWORD_SET(NSTEPS) THEN NSTEPS=100
IF NOT KEYWORD_SET(PRINTNEG) THEN PRINTNEG=0
IF NOT KEYWORD_SET(PRINTPROGRESS) THEN PRINTPROGRESS=0
```

```
IF NOT KEYWORD_SET(HOWLONG) THEN HOWLONG=0

LEN1=N_ELEMENTS(PTT1(*,0))
LEN2=N_ELEMENTS(PTT2(*,0))
NPT1=N_ELEMENTS(PTT1(0,*))
NPT2=N_ELEMENTS(PTT2(0,*))
NUMTOT=DBLARR(NSTEPS+1,NSTEPS+1,3)

SAVEPTT1=PTT1 ;THIS IS JUST SAVING THE ORIGINAL POINTS TO RESET THEM AT
THE END
SAVEPTT2=PTT2 ;SEE END OF PROGRAM FOR EXPLAINATION

IF (NPT2 GT NPT1) THEN BEGIN
    PRINT,'>>>WARNING: SHIFTING ORDER OF INPUTS'
    IF NOT KEYWORD_SET(RETVEC) THEN RETVEC=0
    IF NOT KEYWORD_SET(PRINTPROGRESS) THEN PRINTPROGRESS=0

    IF NOT KEYWORD_SET(INISHIFT) THEN BEGIN
       IF NOT KEYWORD_SET(CHECKPOS) THEN $

RESS=COMPARE_PTS5(PTT2,PTT1,DELX,DELY,NSTEPS=NSTEPS,RAD=RAD,HOWLONG=HOW
LONG, $

RETVEC=RETVEC,PRINTNEG=1,PRINTPROGRESS=PRINTPROGRESS) ELSE $

RESS=COMPARE_PTS5(PTT2,PTT1,DELX,DELY,NSTEPS=NSTEPS,RAD=RAD,HOWLONG=HOW
LONG, $
                        CHECKPOS=-
CHECKPOS,RETVEC=RETVEC,PRINTNEG=1,PRINTPROGRESS=PRINTPROGRESS)
    ENDIF
    IF KEYWORD_SET(INISHIFT) THEN BEGIN
       IF NOT KEYWORD_SET(CHECKPOS) THEN $

RESS=COMPARE_PTS5(PTT2,PTT1,DELX,DELY,NSTEPS=NSTEPS,RAD=RAD,HOWLONG=HOW
LONG, $
                        INISHIFT=-
INISHIFT,RETVEC=RETVEC,PRINTNEG=1,PRINTPROGRESS=PRINTPROGRESS) ELSE $

RESS=COMPARE_PTS5(PTT2,PTT1,DELX,DELY,NSTEPS=NSTEPS,RAD=RAD,HOWLONG=HOW
LONG, $
                        INISHIFT=-INISHIFT,CHECKPOS=-
CHECKPOS,RETVEC=RETVEC,PRINTNEG=1,PRINTPROGRESS=PRINTPROGRESS)
    ENDIF
    ;PRINT,RESS
    IF KEYWORD_SET(INISHIFT) THEN INISHIFT=-INISHIFT
    IF KEYWORD_SET(CHECKPOS) THEN CHECKPOS=-CHECKPOS
    IF (RETVEC) THEN RETURN,-RESS ;THIS TAKES CARE OF RETURNING THE
RIGHT VECTOR
    ;LEN=N_ELEMENTS(RESS(*,0))/2
    RESS2=RESS
    IF (RESS(0,0) NE 0.0 AND RESS(1,1) NE 0.0) THEN BEGIN
       RESS2(0:LEN1-1,*)=RESS(LEN2:LEN2+LEN1-1,*)
       RESS2(LEN1:LEN2+LEN1-1,*)=RESS(0:LEN2-1,*)
    ENDIF
    RETURN,RESS2
ENDIF
```

```
IF (HOWLONG EQ 1) THEN BEGIN
    TIM=(10.4784*NSTEPS-47.9380)*EXP((NPT1+NPT2)*0.5*(2.29491E-
005*NSTEPS+0.00129073))+(-13.6754*NSTEPS+101.674)
    ;ORIGINAL ESTIMATE: TIM=(13.8926*NSTEPS-
240.814)*EXP((NPT1+NPT2)*0.5*(7.40000E-006*NSTEPS+0.00203399))+(-
17.3876*NSTEPS+309.982)
    PRINT,'>>>WARNING: THIS ROUTINE IS ESTIMATED TO RUN FOR
'+STRTRIM(FIX(TIM)/60,1)+' MINUTES'
    PRINT,'>>>DO YOU WISH TO CONTINUE? (Y/N)'
    ANS=GET_KBRD(1)
    IF (ANS NE 'Y') THEN RETURN,[[0.0,0],[0,0]]
ENDIF

IF KEYWORD_SET(INISHIFT) THEN BEGIN
    IF (PRINTNEG) THEN PRINT,'INITIAL SHIFT HAS BEEN COMPENSATED FOR IN
THE RESULTS!: ',-INISHIFT $
                ELSE PRINT,'INITIAL SHIFT HAS BEEN COMPENSATED FOR IN
THE RESULTS!: ',INISHIFT
    PTT1(0,*)=PTT1(0,*)+INISHIFT(0)
    PTT1(1,*)=PTT1(1,*)+INISHIFT(1)
ENDIF


DISS=DBLARR(NPT1,NPT2,2)
FOR K=0,NPT1-1 DO DISS(K,*,0)=REPLICATE(PTT1(0,K),NPT2)-PTT2(0,*)
FOR K=0,NPT1-1 DO DISS(K,*,1)=REPLICATE(PTT1(1,K),NPT2)-PTT2(1,*)
CLO=DBLARR(N_ELEMENTS(PTT1(*,0)),NPT2)
DISTT=DBLARR(2,NPT1)

FOR I=0,NSTEPS DO BEGIN
    IF (PRINTPROGRESS EQ 1) THEN PRINT,''
    DX=I*DELX*1.0/NSTEPS-DELX/2.0
    IF (PRINTPROGRESS EQ 1) THEN PRINT,'I='+STRTRIM(I,1)+', PERCENT
DONE: 0%                                                    100%'
    IF (PRINTPROGRESS EQ 1) THEN PRINT,FORMAT='(A30,$)','            '
    FOR J=0,NSTEPS DO BEGIN
       DY=J*DELY*1.0/NSTEPS-DELY/2.0
       DISSTOT=[[[DISS(*,*,0)-DX]],[[DISS(*,*,1)-DY]]]
       DISSTOT=REFORM(DISSTOT(*,*,0))^2+REFORM(DISSTOT(*,*,1))^2
       NUMMATCH=0
       FOR K=0,NPT2-1 DO NUMMATCH=NUMMATCH+(MIN(DISSTOT(*,K)) LE RAD^2)
       NUMTOT(I,J,*)=[NUMMATCH,DX,DY]
       IF (PRINTPROGRESS EQ 1 AND (J MOD (NSTEPS/10)) EQ 0) THEN PRINT,
FORMAT='(A4,$)','****'
    ENDFOR
ENDFOR
IF (PRINTPROGRESS EQ 1) THEN PRINT,''

;THE FOLLOWING IS JUST A LOT OF PROCESSING OF THE DATA, TO FIND A
;SUITABLE ESTIMATE OF THE POSITION OF THE MAXIMUM
MAXF=MAX(NUMTOT(*,*,0))
W=WHERE( NUMTOT(*,*,0) EQ MAXF, NW )
YS=W/FIX(NSTEPS+1)
XS=W-YS*(NSTEPS+1)
RESARR=DBLARR(2,NW)
FOR I=0L,NW-1 DO
RESARR(*,I)=[NUMTOT(XS(I),YS(I),1),NUMTOT(XS(I),YS(I),2)]
```

```
NVECTS=N_ELEMENTS(RESARR(0,*))
IF (NVECTS NE 1) THEN VECT=[MEAN(RESARR(0,*)),MEAN(RESARR(1,*))] ELSE
VECT=RESARR(0:1)
IF KEYWORD_SET(INISHIFT) THEN VECT=[VECT(0)-INISHIFT(0),VECT(1)-
INISHIFT(1)]
IF (PRINTNEG) THEN PRINT,'THE COMPUTER FOUND A MAXIMUM OF
'+STRTRIM(FIX(MAXF),1)+' MATCHES, AT ' $
       ,STRTRIM(NVECTS,1),' POSITION(S), WITH THE AVERAGE VECTOR:
[',STRTRIM(-VECT(0),1),',', $
       STRTRIM(-VECT(1),1),'].' ELSE PRINT,'THE COMPUTER FOUND A
MAXIMUM OF '+STRTRIM(FIX(MAXF),1)+ $
       ' MATCHES, AT ',STRTRIM(NVECTS,1),' POSITION(S), WITH THE
AVERAGE VECTOR: [', $
       STRTRIM(VECT(0),1),',',STRTRIM(VECT(1),1),'].'

RESX=DBLARR(NSTEPS+1)
RESY=RESX
WN=WHERE(NUMTOT LE MAX(NUMTOT(*,*,0))/1.3)
NUMTOT2=NUMTOT & IF (WN(0) NE -1) THEN NUMTOT2(WN)=0.0
FOR I=0,NSTEPS DO RESX(I)=MEAN(NUMTOT2(I,*,0))
FOR I=0,NSTEPS DO RESY(I)=MEAN(NUMTOT2(*,I,0))
RESX=RESX-MIN(RESX)
RESY=RESY-MIN(RESY)
;WX=WHERE(RESX LE MAX(RESX)/1.3)
;IF (WX(0) NE -1) THEN RESX(WX)=0
;WY=WHERE(RESY LE MAX(RESY)/1.3)
;IF (WY(0) NE -1) THEN RESY(WY)=0
WINDOW,5,TITLE='X (SOLID) AND Y (DOTTED) PROJECTIONS OF TRANSLATION
SPACE IMAGE'
PLOT,NUMTOT(0,*,2),RESX
OPLOT,NUMTOT(0,*,2),RESY,LINESTYLE=1
MEANX=0.0
MEANY=0.0
TOTX=TOTAL(RESX)
TOTY=TOTAL(RESY)
FOR I=0,NSTEPS DO MEANX=MEANX+NUMTOT(0,I,2)*RESX(I)/TOTX
FOR I=0,NSTEPS DO MEANY=MEANY+NUMTOT(0,I,2)*RESY(I)/TOTY

DISSTOT=[[[DISS(*,*,0)-MEANX]],[[DISS(*,*,1)-MEANY]]]
DISSTOT=REFORM(DISSTOT(*,*,0))^2+REFORM(DISSTOT(*,*,1))^2
NUMMATCH=0
W1=0 & W2=0
FOR K=0,NPT2-1 DO BEGIN
    NUMMATCH=NUMMATCH+(MIN(DISSTOT(*,K)) LE RAD^2)
    WT=(WHERE(DISSTOT(*,K) LE RAD^2))(0)
    IF (WT NE -1) THEN BEGIN
       W1=[W1,WT]
       W2=[W2,K]
    ENDIF
ENDFOR
IF KEYWORD_SET(INISHIFT) THEN MEANVECT=[MEANX-INISHIFT(0),MEANY-
INISHIFT(1)] ELSE MEANVECT=[MEANX,MEANY]
IF (N_ELEMENTS(W1) NE 1) THEN W1=W1(1:N_ELEMENTS(W1)-1)
IF (N_ELEMENTS(W2) NE 1) THEN W2=W2(1:N_ELEMENTS(W2)-1)
IF (PRINTNEG) THEN PRINT,'SHIFT DATA BY : [',STRTRIM(-
MEANVECT(0),1),',',STRTRIM(-MEANVECT(1),1), $
```

```
       '] AND FOUND ',STRTRIM(NUMMATCH,1),' MATCHES' ELSE PRINT,'SHIFT
DATA BY : [', STRTRIM(MEANVECT(0),1), $
       ',',STRTRIM(MEANVECT(1),1),'] AND FOUND ',STRTRIM(NUMMATCH,1),'
MATCHES'
IF (N_ELEMENTS(W1) NE 1 AND N_ELEMENTS(W2) NE 1) THEN
XANDY=[PTT1(*,W1),PTT2(*,W2)] ELSE XANDY=[[0.0,0],[0,0]]
MAXF=NUMMATCH

;IF THE 'CHECKPOS' KEYWORD IS SET, THE COMPUTER WILL LOOK AT THAT
;SHIFT AND RETURN THE POSITIONS OF CORRELATED POINTS AT THAT SHIFT
IF KEYWORD_SET(CHECKPOS) THEN BEGIN
    IF KEYWORD_SET(INISHIFT) THEN DISSTOT=[[[DISS(*,*,0)-CHECKPOS(0)-
INISHIFT(0)]], $
         [[DISS(*,*,1)-CHECKPOS(1)-INISHIFT(1)]]] ELSE $
         DISSTOT=[[[DISS(*,*,0)-CHECKPOS(0)]],[[DISS(*,*,1)-
CHECKPOS(1)]]]
    DISSTOT=REFORM(DISSTOT(*,*,0))^2+REFORM(DISSTOT(*,*,1))^2
    NUMMATCH=0
    W1=0 & W2=0
    FOR K=0,NPT2-1 DO BEGIN
       NUMMATCH=NUMMATCH+(MIN(DISSTOT(*,K)) LE RAD^2)
       WT=(WHERE(DISSTOT(*,K) LE RAD^2))(0)
       IF (WT NE -1) THEN BEGIN
         W1=[W1,WT]
         W2=[W2,K]
       ENDIF
    ENDFOR
    IF (N_ELEMENTS(W1) NE 1) THEN W1=W1(1:N_ELEMENTS(W1)-1)
    IF (N_ELEMENTS(W2) NE 1) THEN W2=W2(1:N_ELEMENTS(W2)-1)
    IF (PRINTNEG) THEN PRINT,'FOUND ',STRTRIM(NUMMATCH,1),' MATCHES AT
CHECK POINT: [', $
       STRTRIM(-CHECKPOS(0),1),',',STRTRIM(-CHECKPOS(1),1),']' ELSE
PRINT,'FOUND ',STRTRIM(NUMMATCH,1), $
       ' MATCHES AT CHECK POINT:
[',STRTRIM(CHECKPOS(0),1),',',STRTRIM(CHECKPOS(1),1),']'
    IF (N_ELEMENTS(W1) NE 1 AND N_ELEMENTS(W2) NE 1) THEN
XANDY=[PTT1(*,W1),PTT2(*,W2)] ELSE XANDY=[[0.0,0],[0,0]]
    MAXF=NUMMATCH
ENDIF

WINDOW,8
SHADE_SURF,NUMTOT(*,*,0),BACKGROUND=-1,COLOR=0

WINDOW, 6, XSIZE=NSTEPS+1, YSIZE=NSTEPS+1, TITLE='FINDING PEAK NUMBER
OF MATCHES'
TV,NUMTOT(*,*,0)

;THIS SAVE IS FOR LATER RECOVERY OF THE NUMTOT ARRAY
TOSAVE=REFORM(NUMTOT(*,*,0))
SAVE,TOSAVE,FILENAME='TEMPORARY_ARRAY.SAV'

PTT1=SAVEPTT1 ;THIS IS ONLY BECAUSE FOR SOME UNKNOWN REASON THE
COMPUTER KEPT OVERWRITING
PTT2=SAVEPTT2 ;THE ORIGINAL PTT1 AS IF IT WERE A GLOBAL VARIABLE

IF (KEYWORD_SET(RETVEC) AND KEYWORD_SET(INISHIFT)) THEN RETURN,[MEANX-
INISHIFT(0),MEANY-INISHIFT(1)]
```

```
IF KEYWORD_SET(RETVEC) THEN RETURN,[MEANX,MEANY]
IF KEYWORD_SET(INISHIFT) THEN BEGIN
    XANDY(0,*)=XANDY(0,*)-INISHIFT(0)
    XANDY(1,*)=XANDY(1,*)-INISHIFT(1)
ENDIF
RETURN,XANDY
END
```

7

Appendix II: LabVIEW Code

7.1 Summary

The software authored in LabVIEW and used for reagent exchange is presented here. The following two virtual instruments, authored in LabVIEW 6*i*, are included:

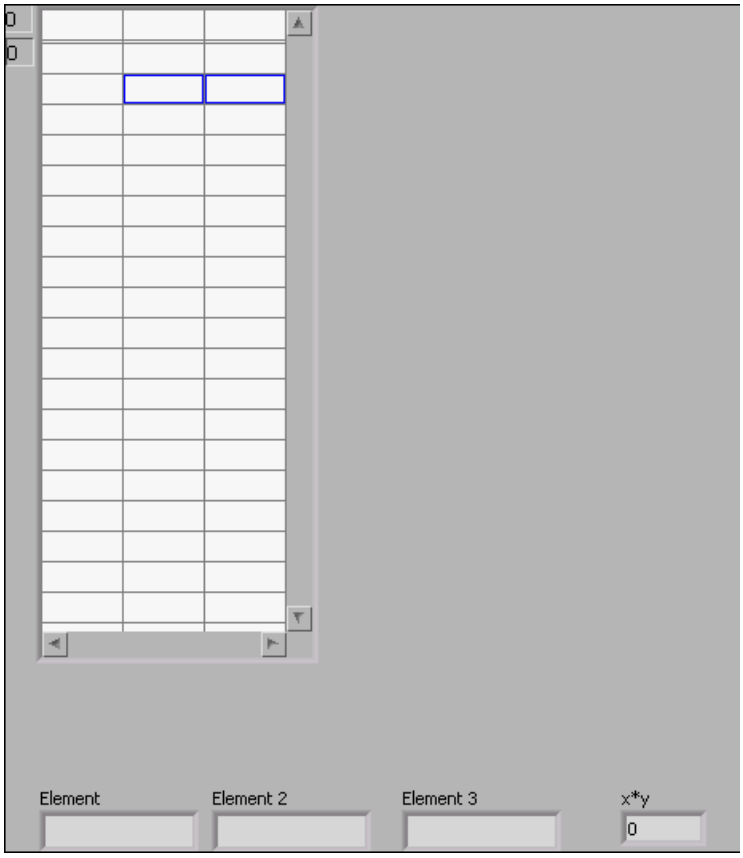
- good.vi
    - virtual instrument software used to automate a series of tasks for the both the syringe pumps and switch valve

- PUMPS INIT9.vi
    - virtual instrument software used to program parameters into the syringe pumps and initialize the switch valve

7.1.1 good.vi

Connector Pane

Front Panel



Controls and Indicators

Table

Element

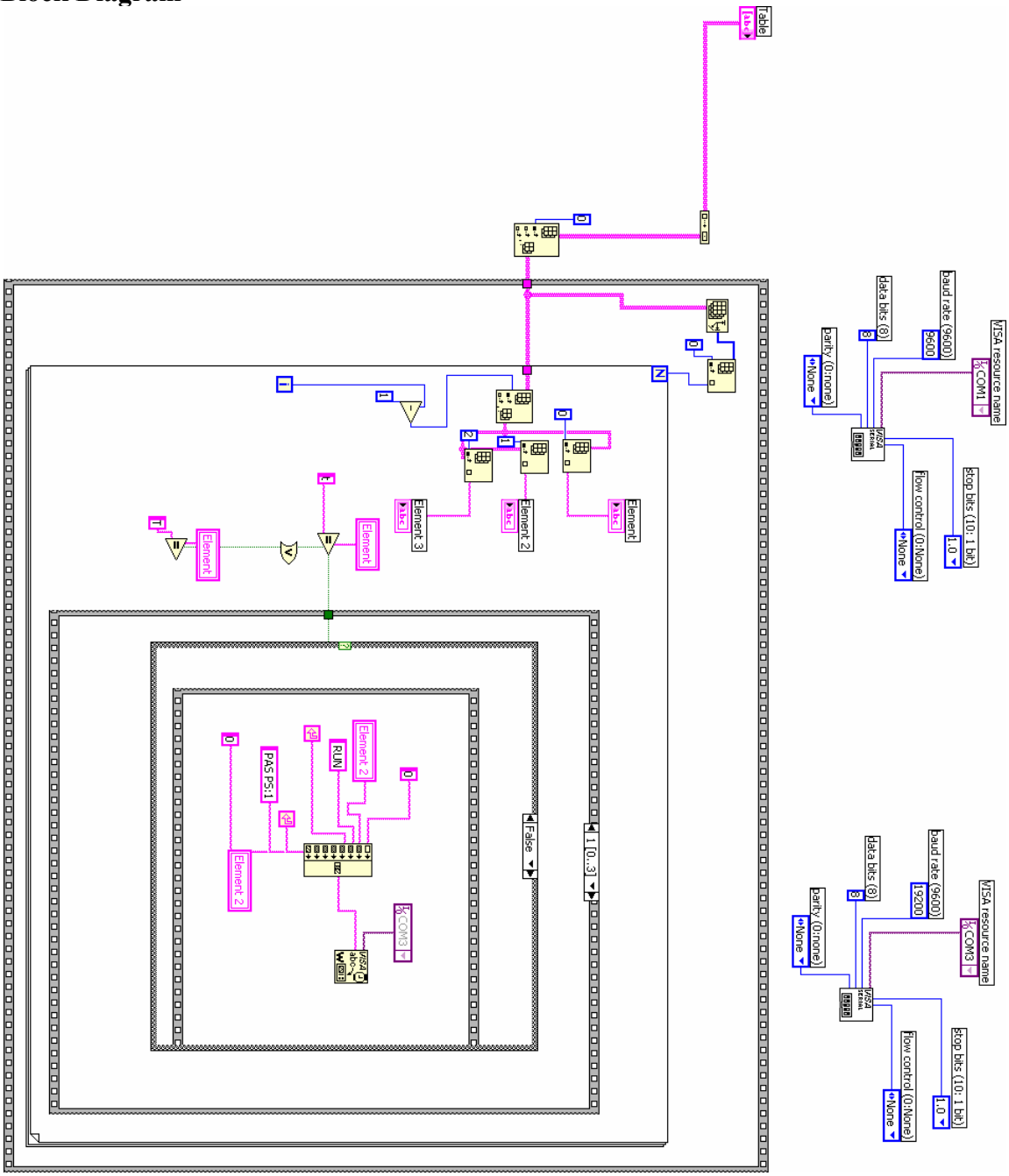
Element 2

Element 3

x*y

Element 4

Block Diagram



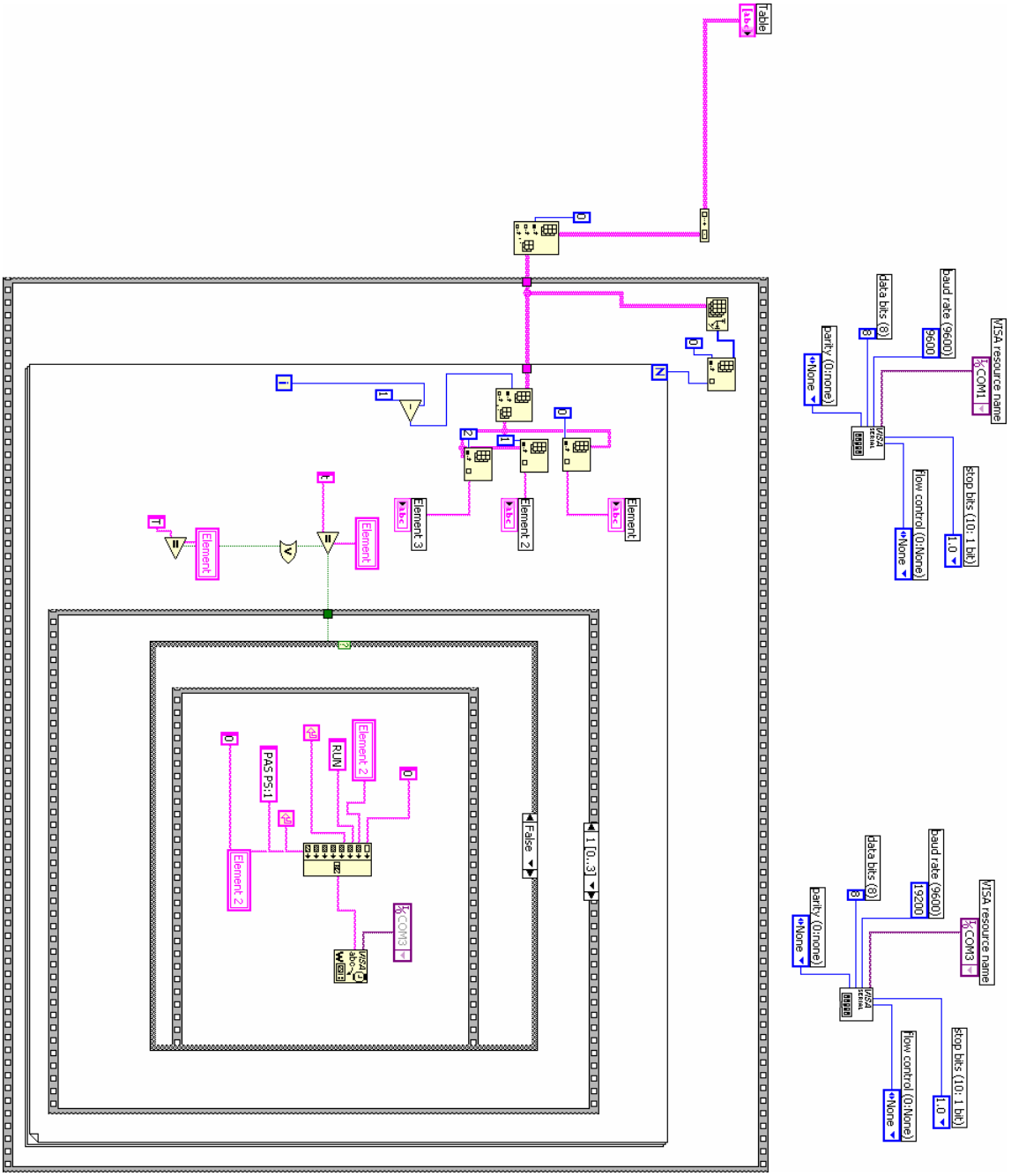
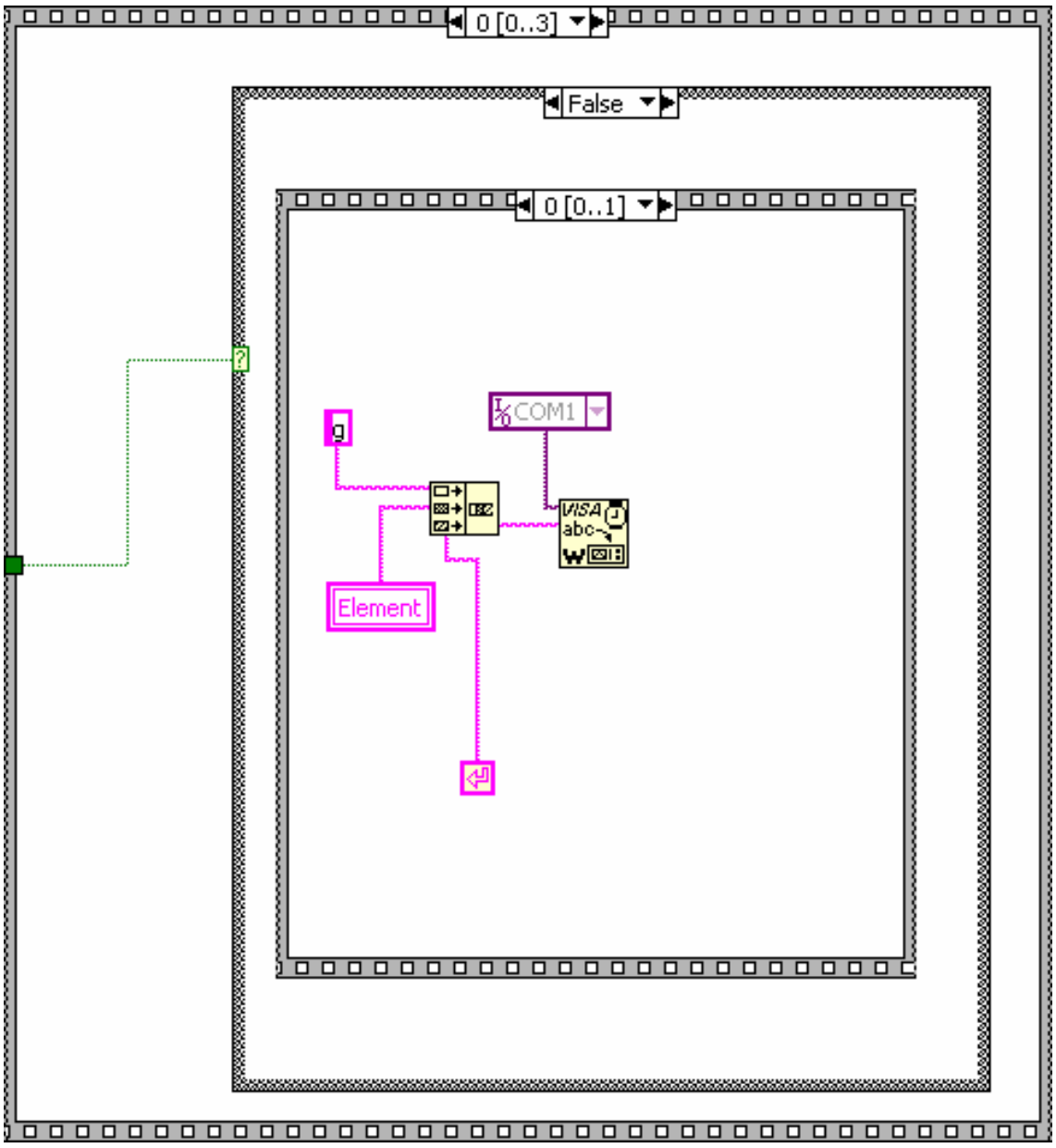
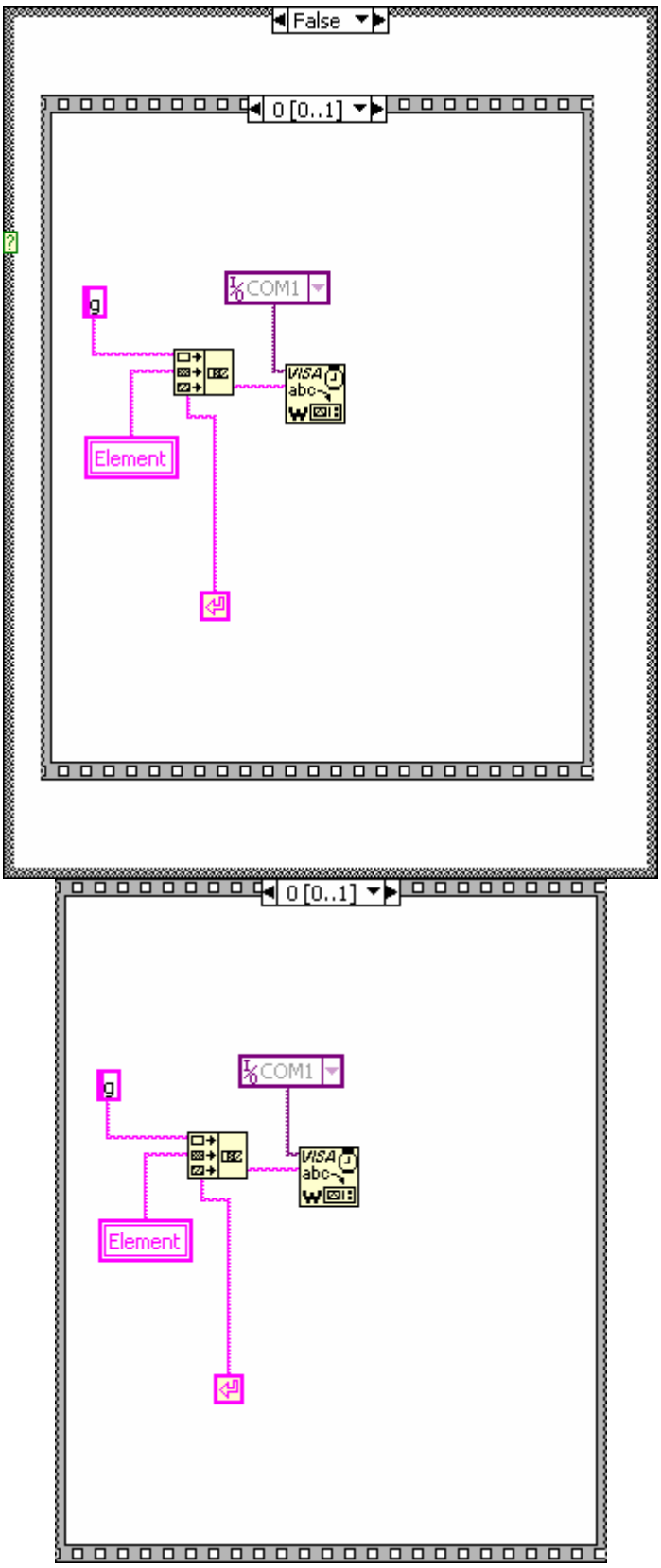
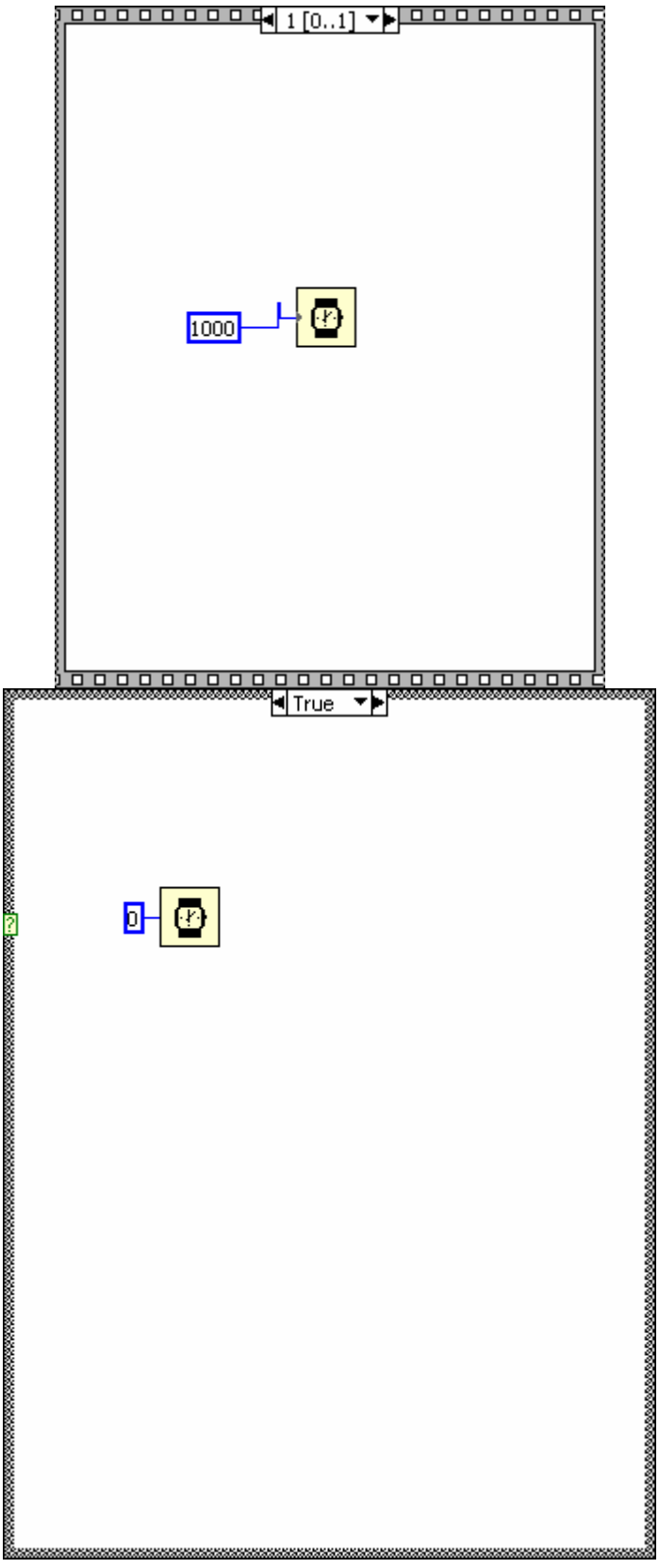
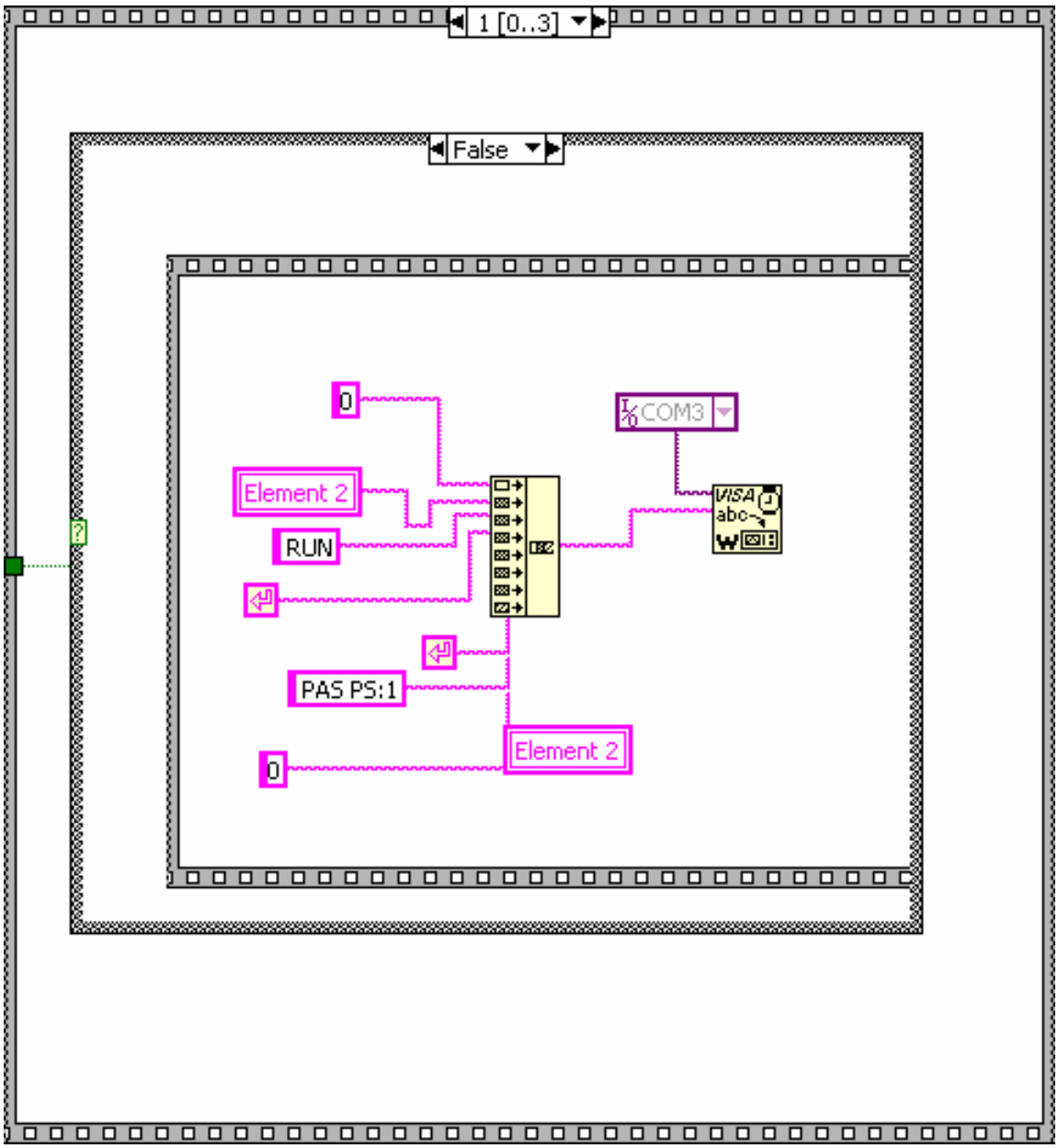


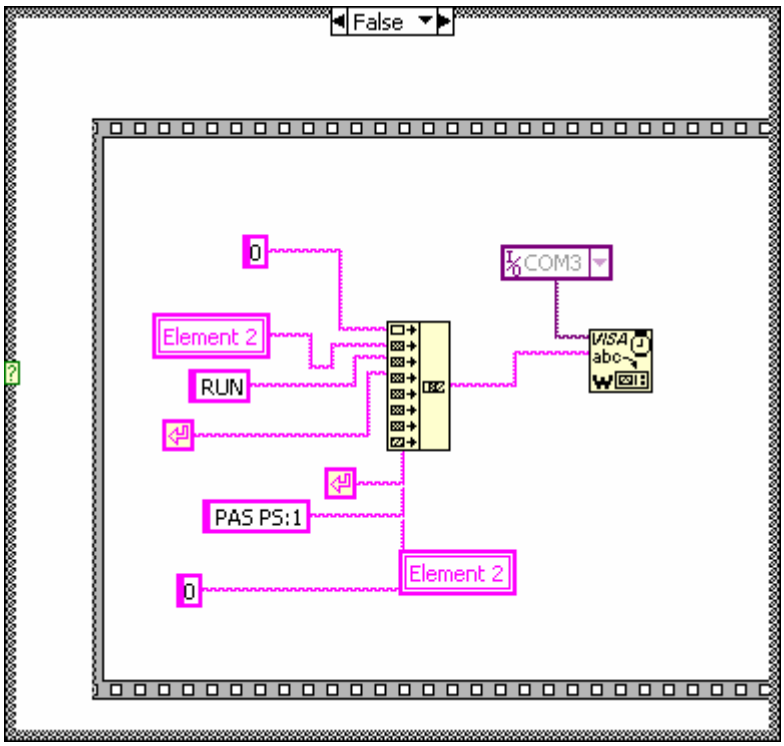
table
Element
Element 3
Element 2
Element
Element
T
F
PAS PS:1
RUN
Element 2
Element 2
0
0
COM3
False
1 [0..3]
VISA resource name
COM1
baud rate (9600)
9600
data bits (8)
8
parity (0:none)
None
stop bits (10: 1 bit)
1.0
flow control (0:None)
None
VISA resource name
COM3
baud rate (9600)
19200
data bits (8)
8
parity (0:none)
None
stop bits (10: 1 bit)
1.0
flow control (0:None)
None

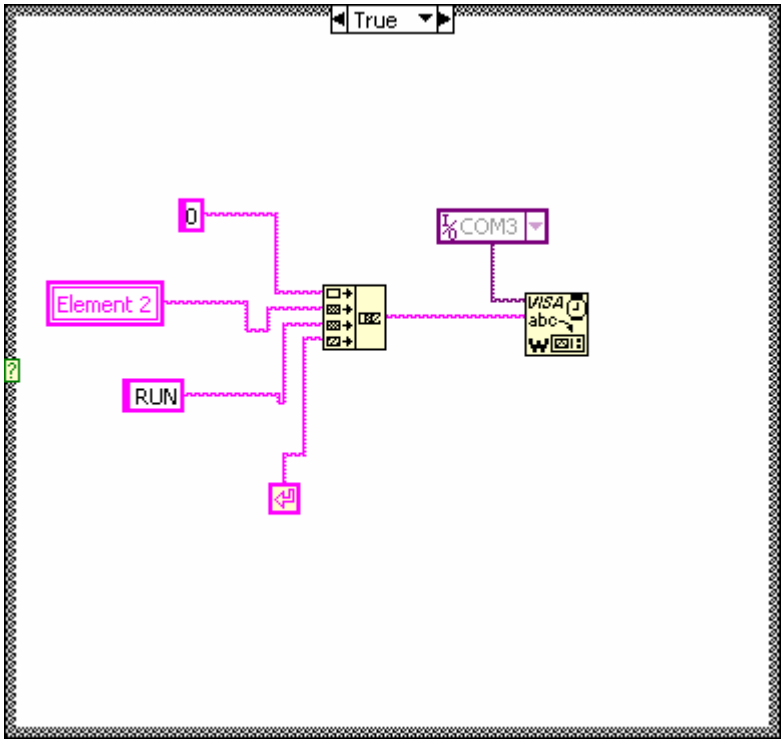

0 [0..3]
False
0 [0..1]
COM1
g
VISA
abc
Element

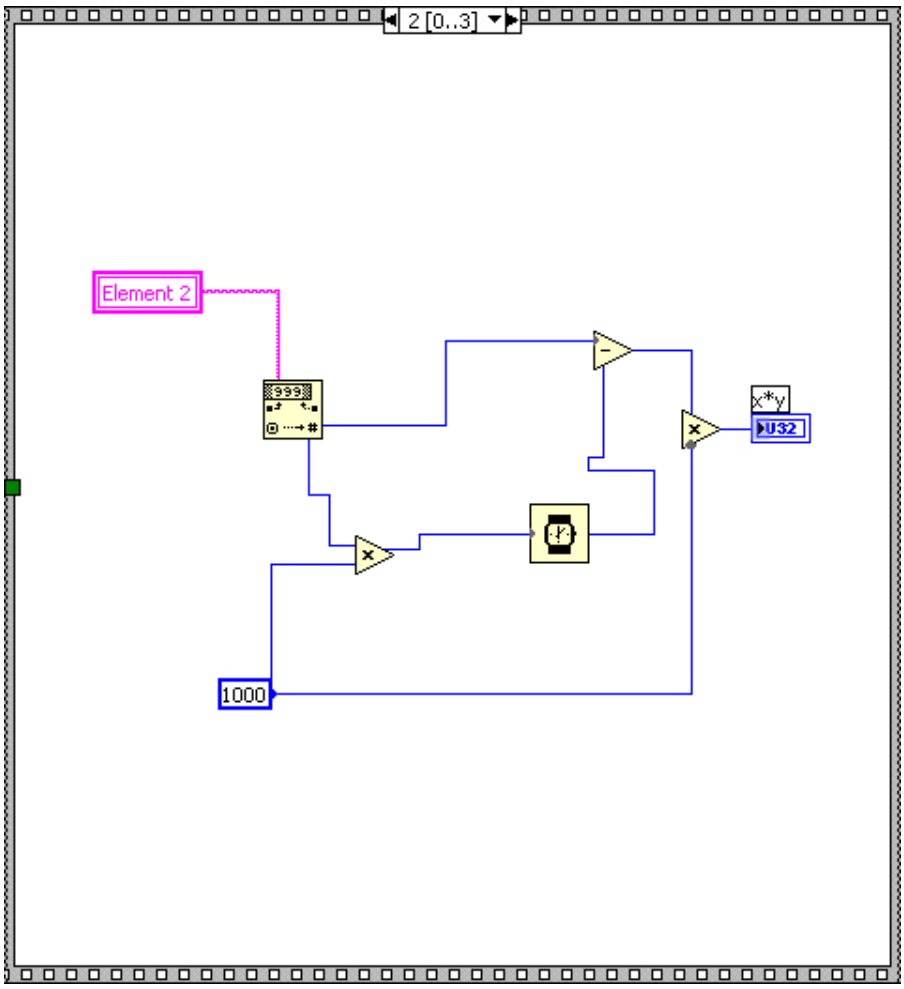

False
0 [0..1]
COM1
g
VISA
abc
Element
0 [0..1]
COM1
g
VISA
abc
Element

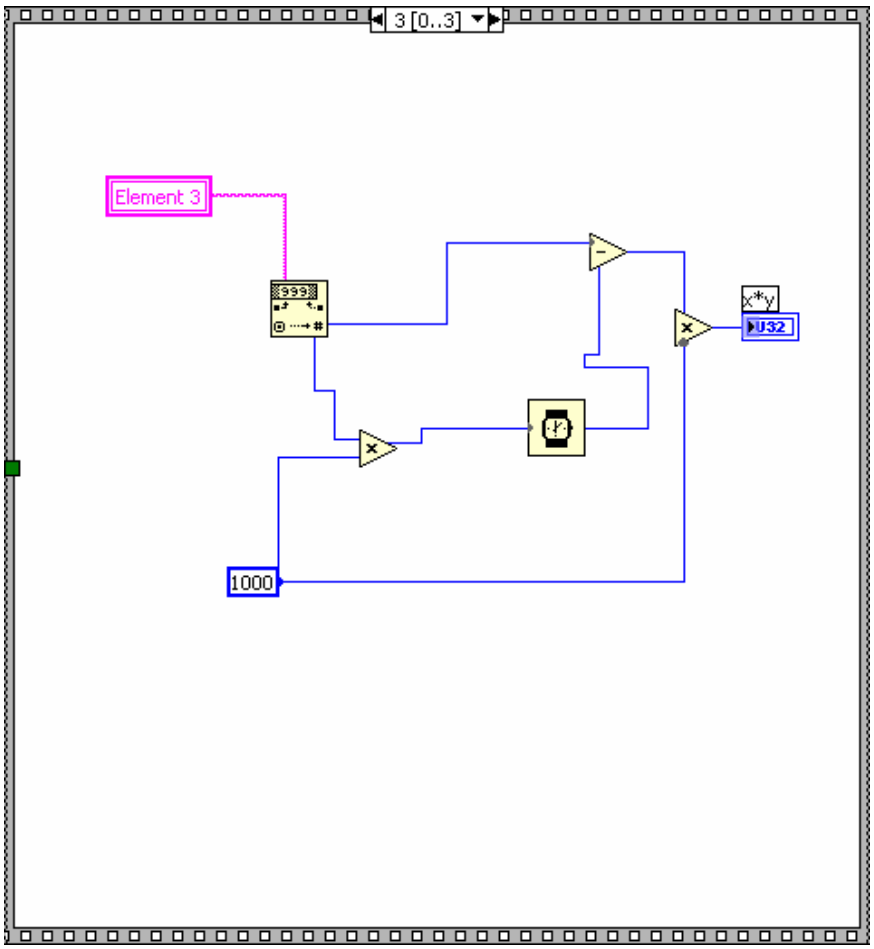

1 [0..1]
1000
True
0


1 [0..3]
False
0
COM3
Element 2
RUN
PAS PS:1
Element 2
0
VISA
abc


False
0
COM3
Element 2
RUN
VISA
abc
PAS PS:1
Element 2
0


True
0
COM3
Element 2
VISA
abc
RUN


2 [0..3]
Element 2
999
1000
x*y
U32



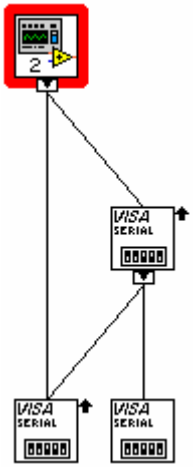
List of SubVIs and Express VIs with Configuration Information

VISA Configure Serial Port
C:\Program Files\National Instruments\LabVIEW 7.1\vi.lib\Instr_visa.llb\VISA Configure Serial Port

VISA Configure Serial Port (Instr).vi
C:\Program Files\National Instruments\LabVIEW 7.1\vi.lib\Instr_visa.llb\VISA Configure Serial Port (Instr).vi

VI Revision History
"good.vi History"
Current Revision: 8

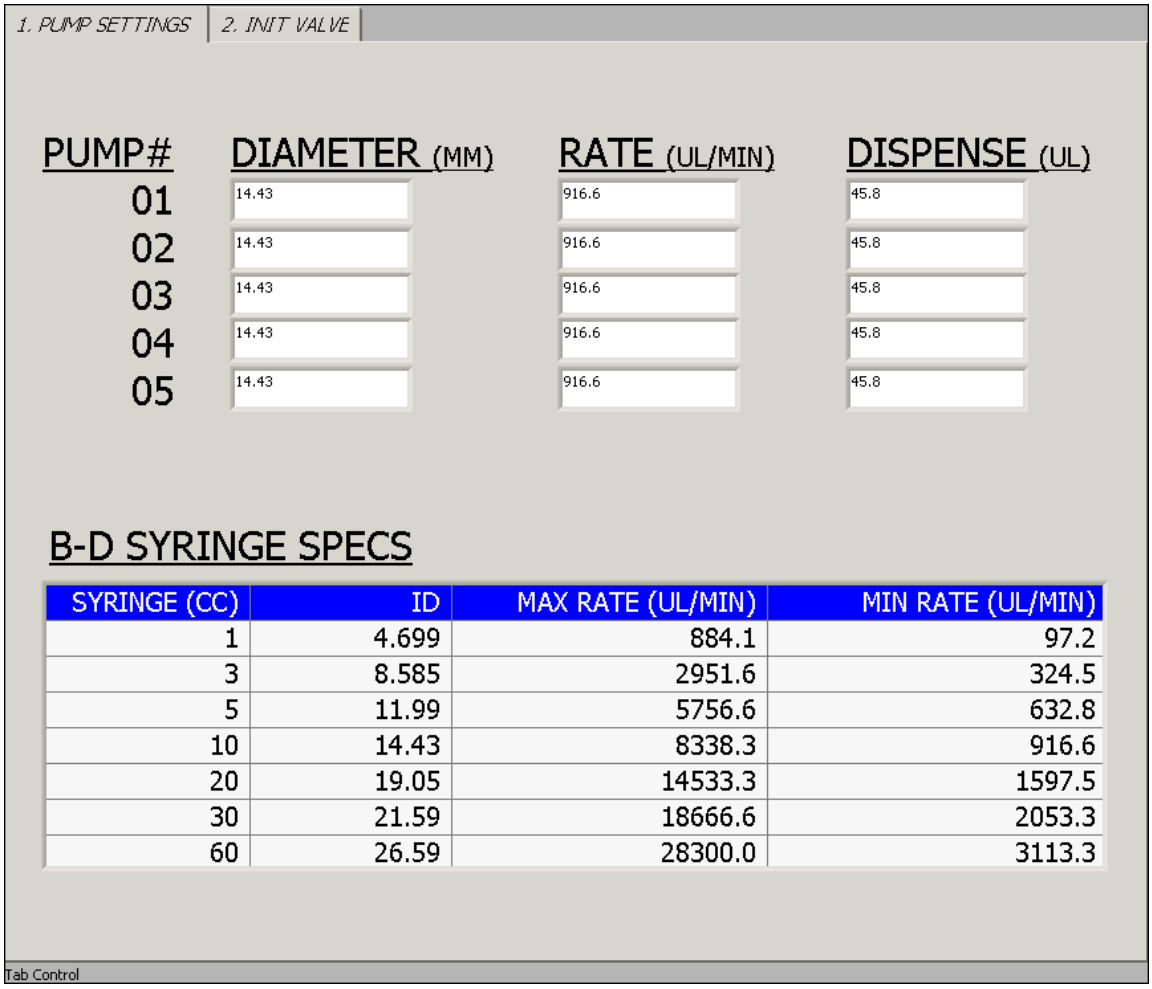
Position in Hierarchy



7.1.2 PUMPS INIT9.vi

Connector Pane

Front Panel



SYRINGE (CC)	ID	MAX RATE (UL/MIN)	MIN RATE (UL/MIN)
1	4.699	884.1	97.2
3	8.585	2951.6	324.5
5	11.99	5756.6	632.8
10	14.43	8338.3	916.6
20	19.05	14533.3	1597.5
30	21.59	18666.6	2053.3
60	26.59	28300.0	3113.3

Controls and Indicators

DIA2

DIA1

DIA4

DIA 3

DIA5

RATE5

RATE3

RATE4

RATE1

RATE2

RATE5 2

RATE3 2

RATE4 2

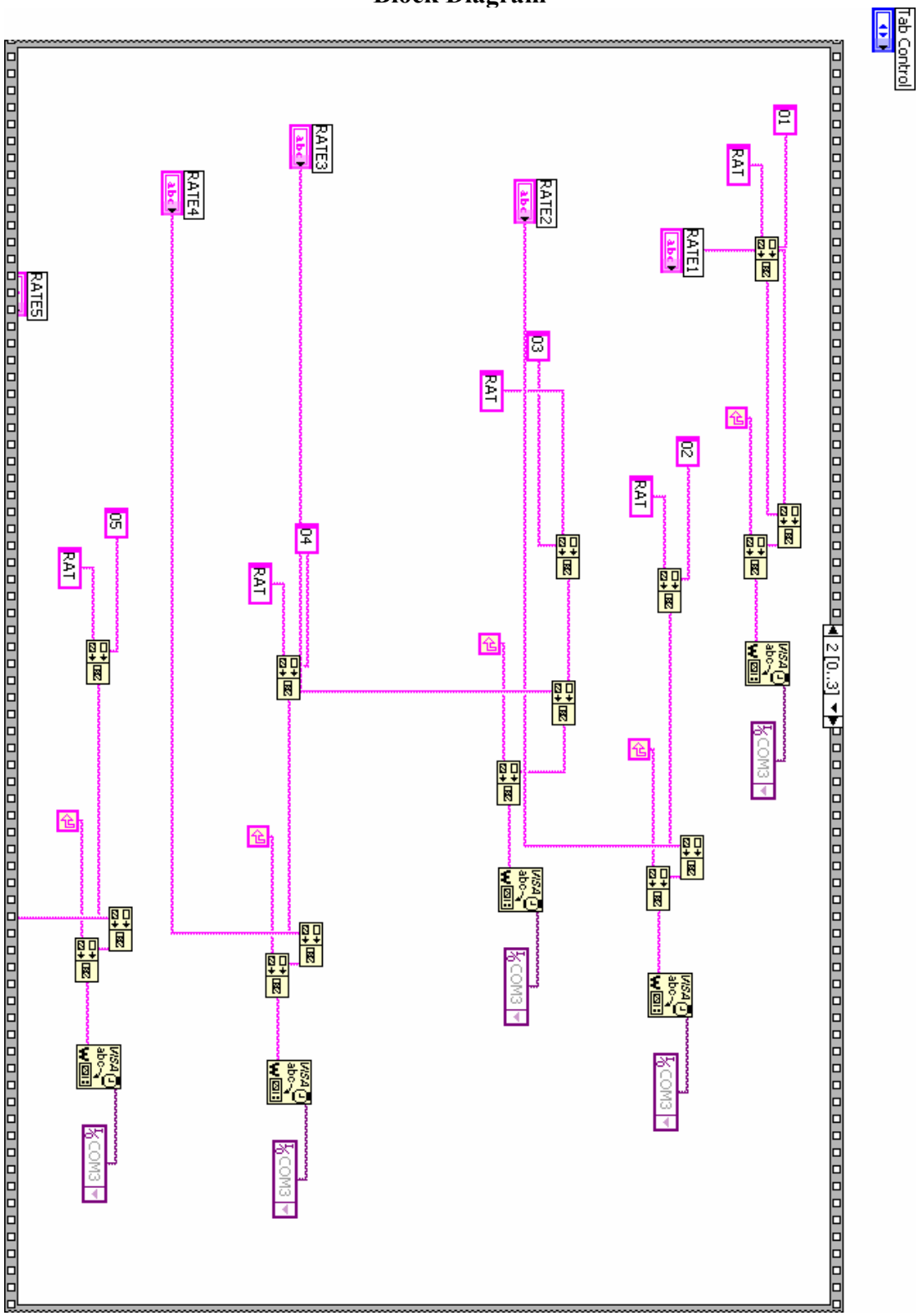
RATE1 2

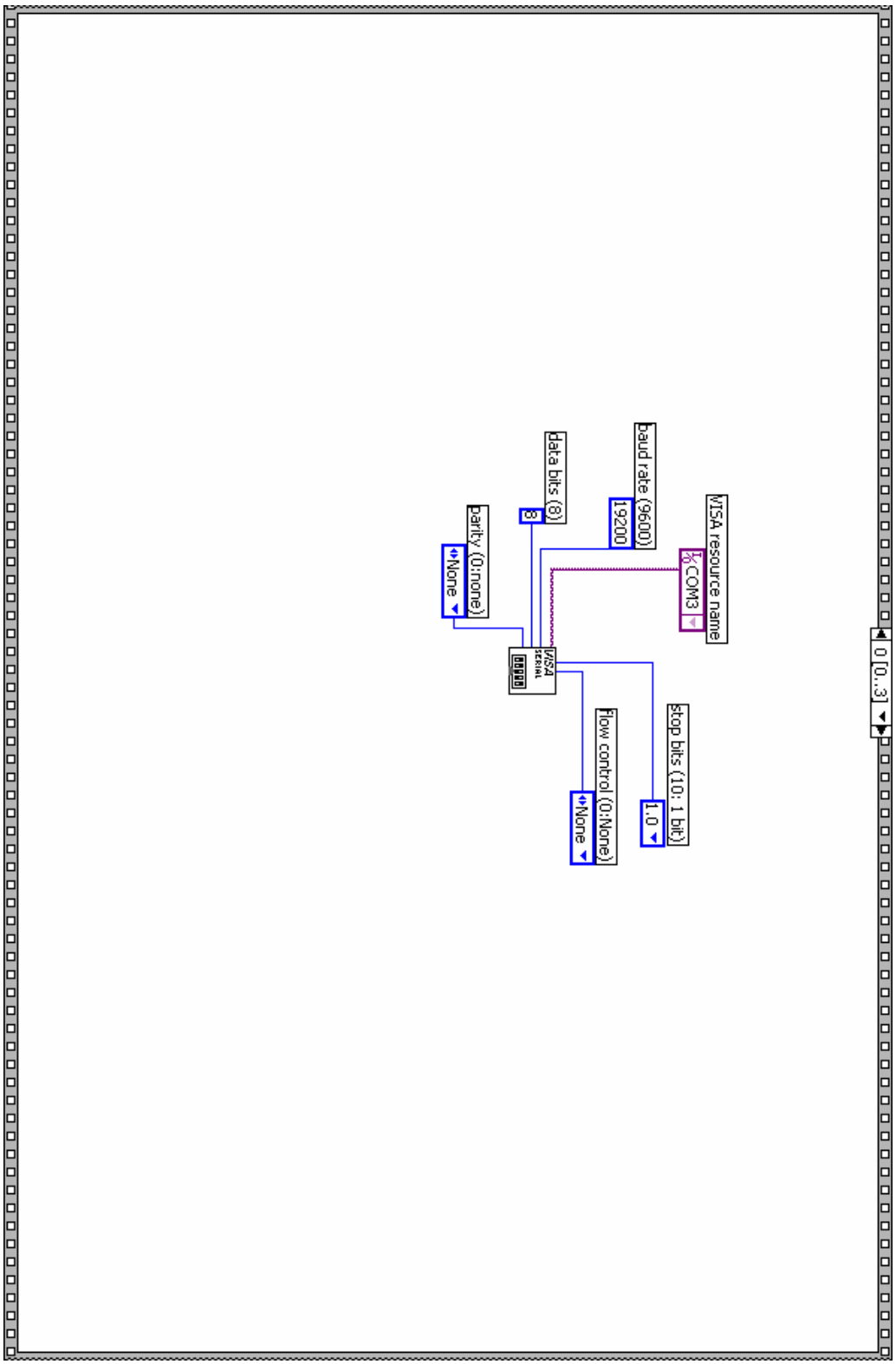
RATE2 2

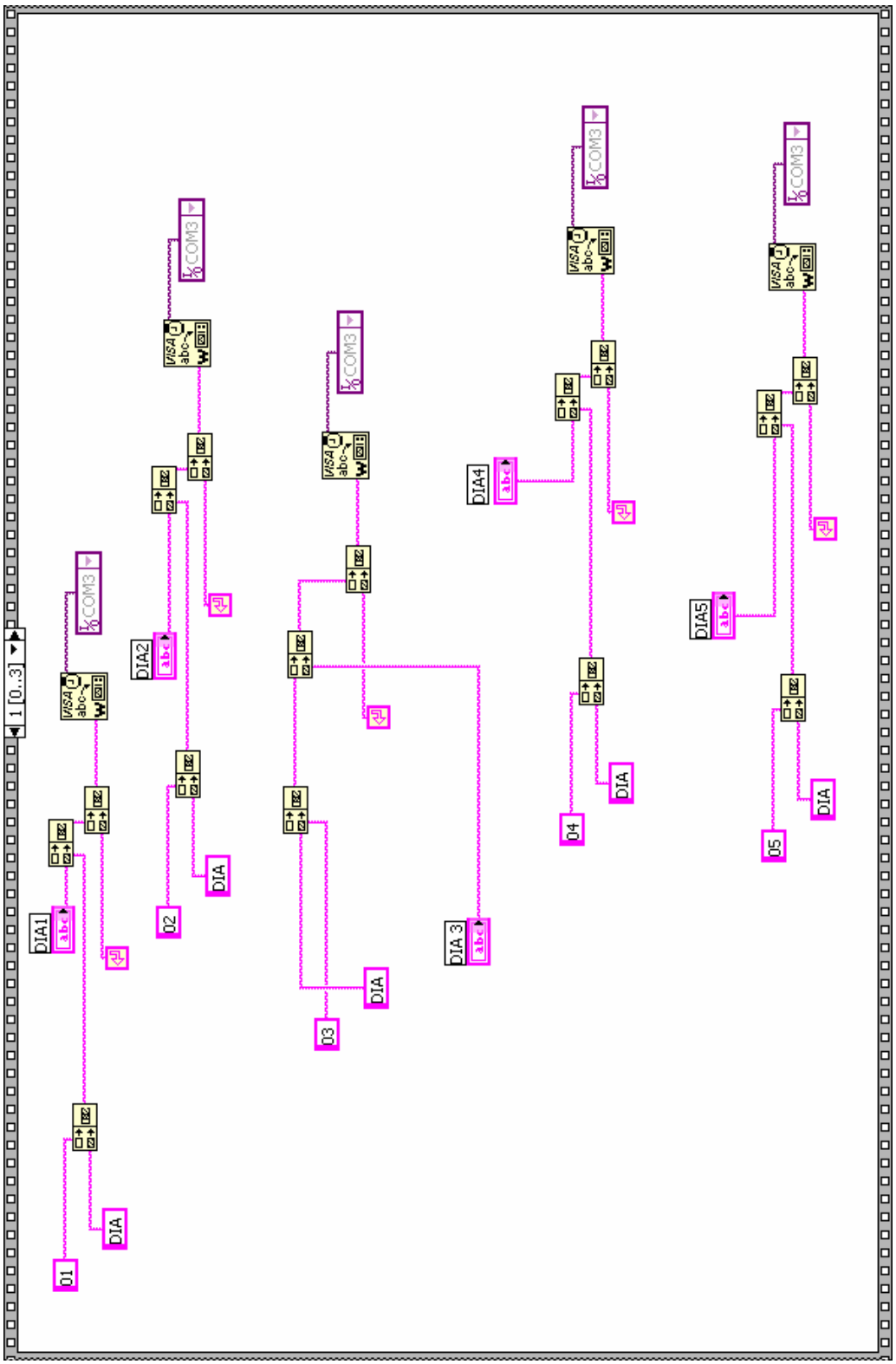
B-D SYRINGE SPECS

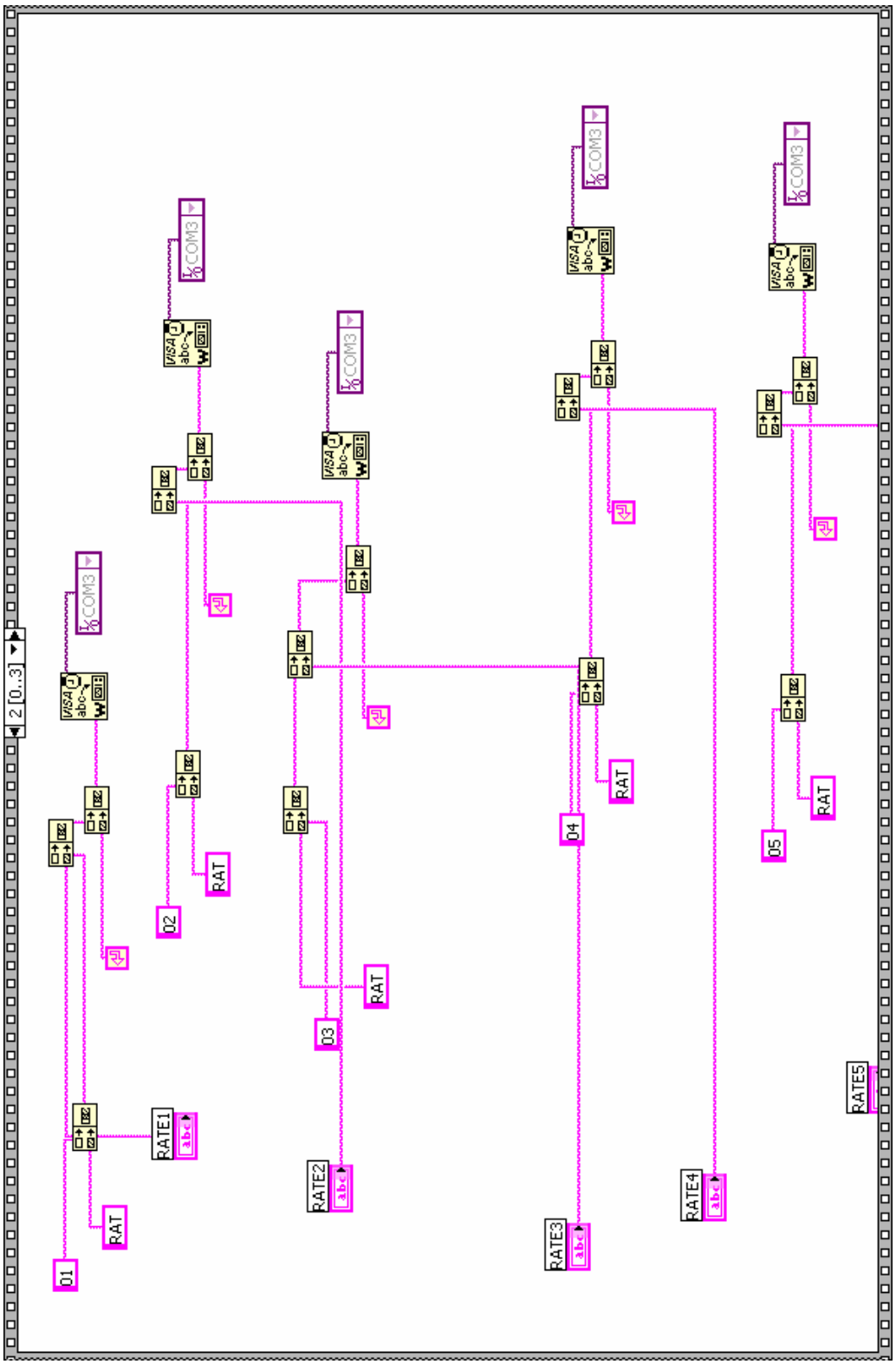
Tab Control

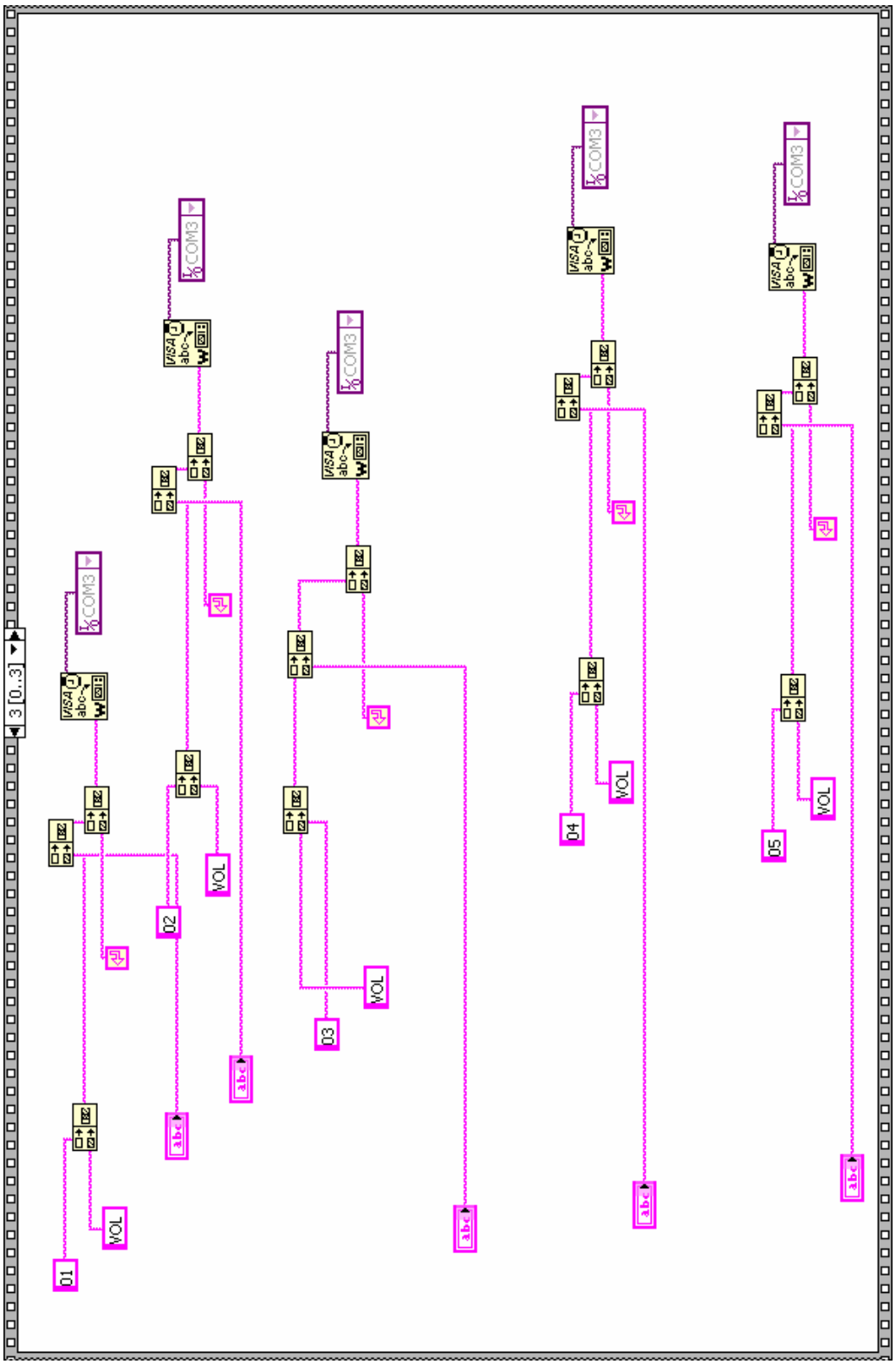
Block Diagram




VISA resource name
COM3
baud rate (9600)
19200
data bits (8)
8
parity (0:none)
None
stop bits (10: 1 bit)
1.0
flow control (0:None)
None
0 [0..3]


1 [0..3]
COM3
VISA
DIA1
DIA2
DIA 3
DIA4
DIA5
DIA
01
02
03
04
05


2 [0..3]
COM3
VISA
01
02
03
04
05
RAT
RATE1
RATE2
RATE3
RATE4
RATE5


3 [0..3]
COM3
VISA
01
02
03
04
05
WOL
abc

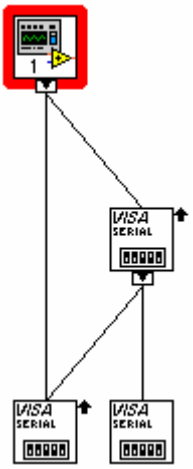
List of SubVIs and Express VIs with Configuration Information

VISA Configure Serial Port
C:\Program Files\National Instruments\LabVIEW 7.1\vi.lib\Instr_visa.llb\VISA Configure Serial Port

VISA Configure Serial Port (Instr).vi
C:\Program Files\National Instruments\LabVIEW 7.1\vi.lib\Instr_visa.llb\VISA Configure Serial Port (Instr).vi

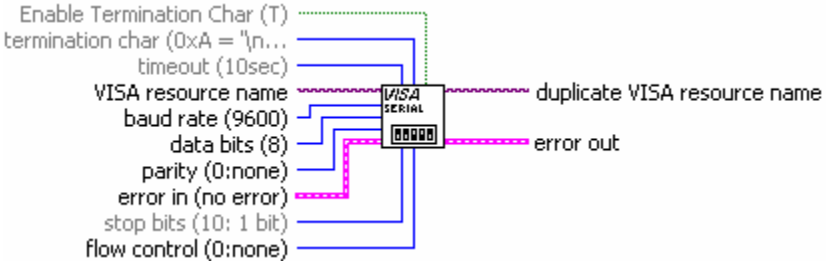
VI Revision History
"PUMPS INIT9.vi History"
Current Revision: 28

Position in Hierarchy



VISA Configure Serial Port
Initializes the serial port specified by **VISA resource name** to the specified settings. You can use this polymorphic VI to initialize a serial port using the Instr VISA class or Serial Instr VISA class. The VISA class specified in the **VISA resource name** determines the polymorphic instance to use.

Connector Pane



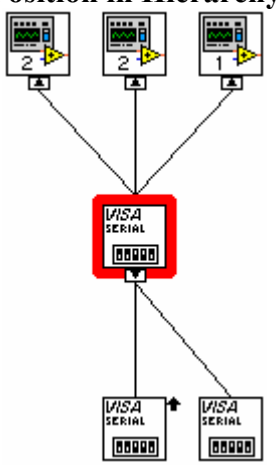
List of SubVIs and Express VIs with Configuration Information

VISA Configure Serial Port (Instr).vi
C:\Program Files\National Instruments\LabVIEW 7.1\vi.lib\Instr_visa.llb\VISA Configure Serial Port (Instr).vi

VISA Configure Serial Port (Serial Instr).vi
C:\Program Files\National Instruments\LabVIEW 7.1\vi.lib\Instr_visa.llb\VISA Configure Serial Port (Serial Instr).vi

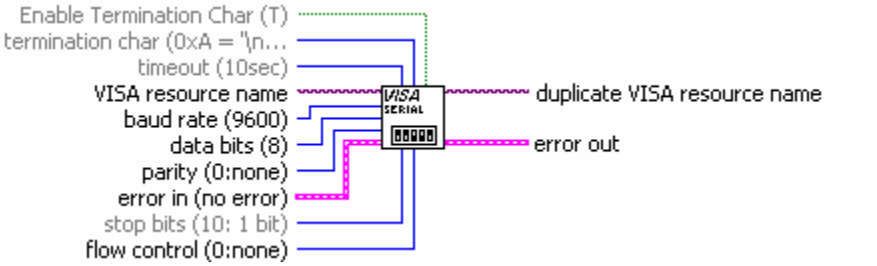
VI Revision History
"VISA Configure Serial Port History"
Current Revision: 4

Position in Hierarchy

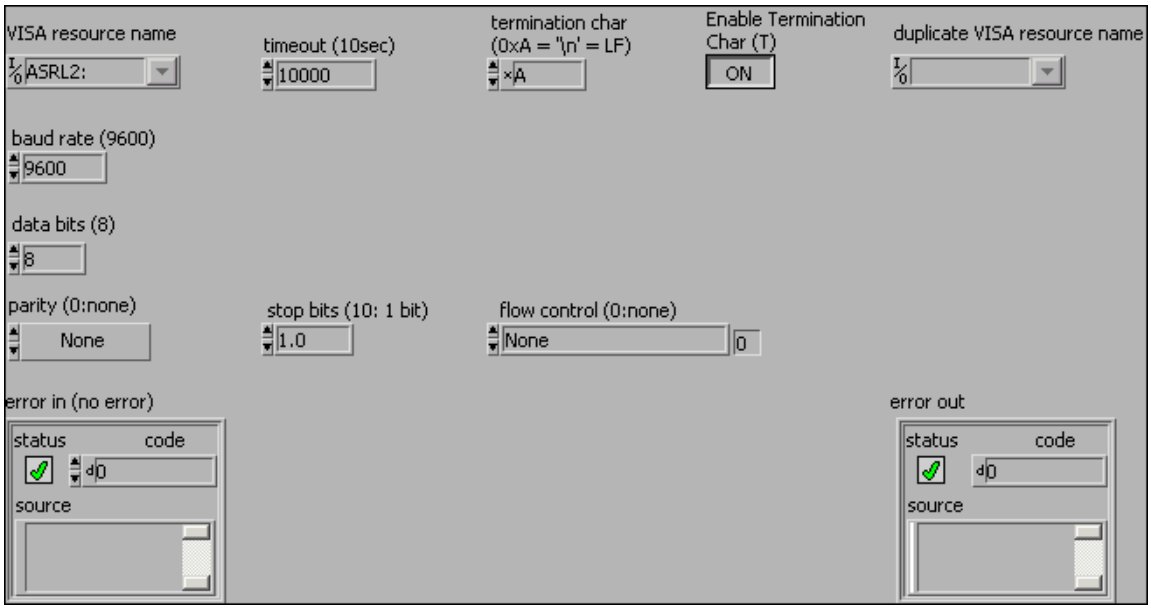


VISA Configure Serial Port (Instr).vi
Initializes the serial port specified by **VISA resource name** to the specified settings. You can use this polymorphic VI to initialize a serial port using the Instr VISA class or Serial Instr VISA class. The VISA class specified in the **VISA resource name** determines the polymorphic instance to use.

Connector Pane



Front Panel



Controls and Indicators

VISA resource name VISA resource name specifies the resource to be opened. This control also specifies the session and class.

timeout (10sec) timeout sets the timeout value for the write and read operations.

baud rate (9600) baud rate is the rate of transmission.

data bits (8) data bits is the number of bits in the incoming data.

parity (0:none) parity specifies the parity used for every frame to be transmitted or received.

termination char (0xA = '\n' = LF) termination char calls for termination of the read operation. The read operation terminates when the **termination char** is read from the serial device.

error in (no error) error in describes error conditions that occur before this VI or function runs.

- **status status** is TRUE (X) if an error occurred or FALSE (checkmark) to indicate a warning or that no error occurred.
- **code code** is the error or warning code.
- **source source** describes the origin of the error or warning and is, in most cases, the name of the VI or function that produced the error or warning.

stop bits (10: 1 bit) stop bits specifies the number of stop bits used to indicate the end of a frame.

Enable Termination Char (T) Enable Termination Char prepares the serial device to recognize **termination char**.

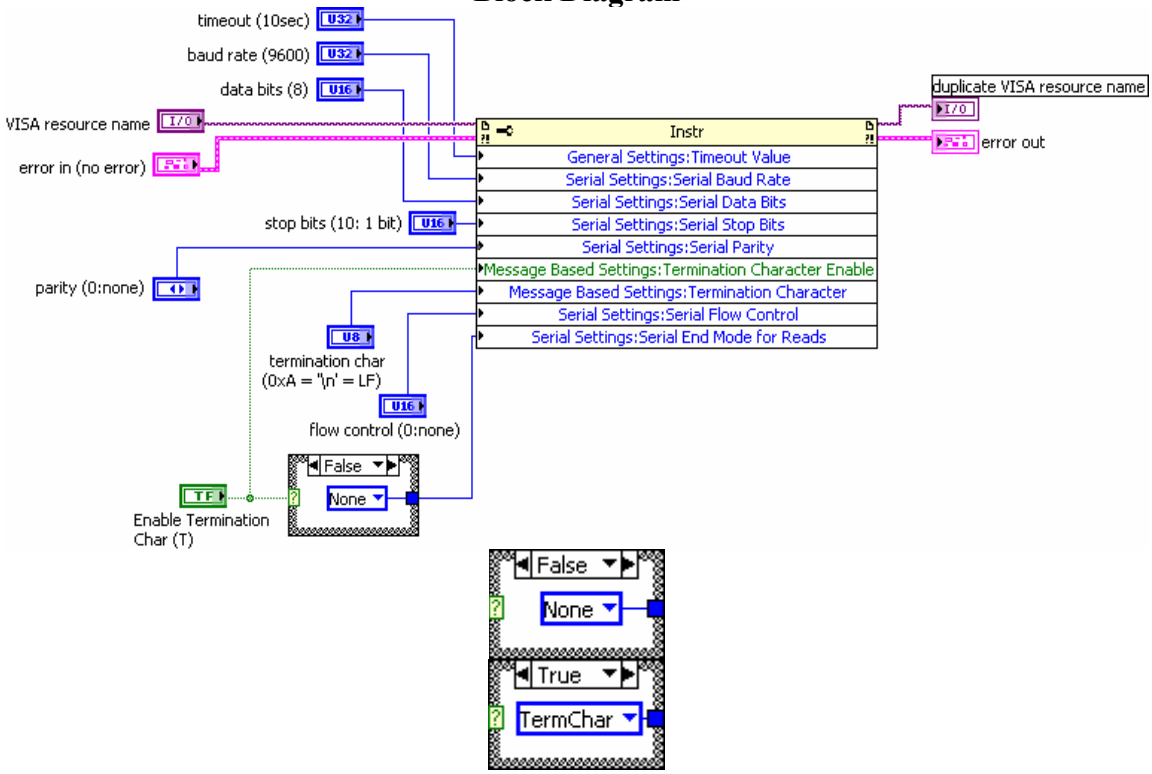
flow control (0:none) flow control sets the type of control used by the transfer mechanism.

error out error out contains error information. If **error in** indicates that an error occurred before this VI or function ran, **error out** contains the same error information. Otherwise, it describes the error status that this VI or function produces.

- **status status** is TRUE (X) if an error occurred or FALSE (checkmark) to indicate a warning or that no error occurred.
- **code code** is the error or warning code.
- **source source** describes the origin of the error or warning and is, in most cases, the name of the VI or function that produced the error or warning.

duplicate VISA resource name duplicate VISA resource name is a copy of the **VISA resource name** that is passed out of the VISA functions.

Block Diagram

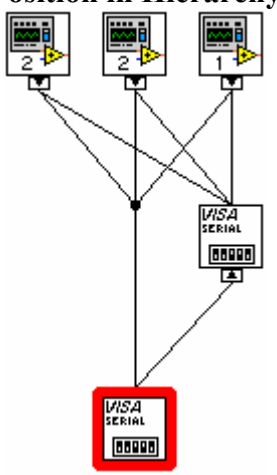


List of SubVIs and Express VIs with Configuration Information

VI Revision History

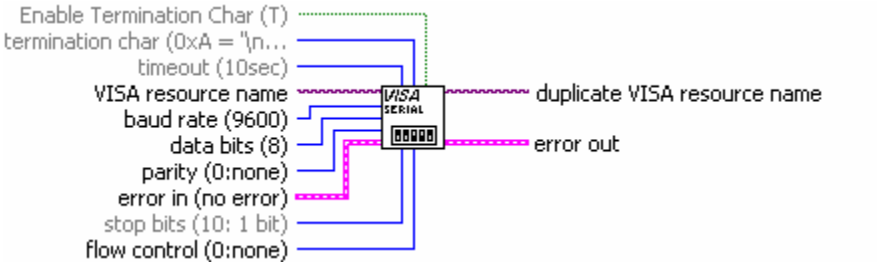
"VISA Configure Serial Port (Instr).vi History"
Current Revision: 103

Position in Hierarchy

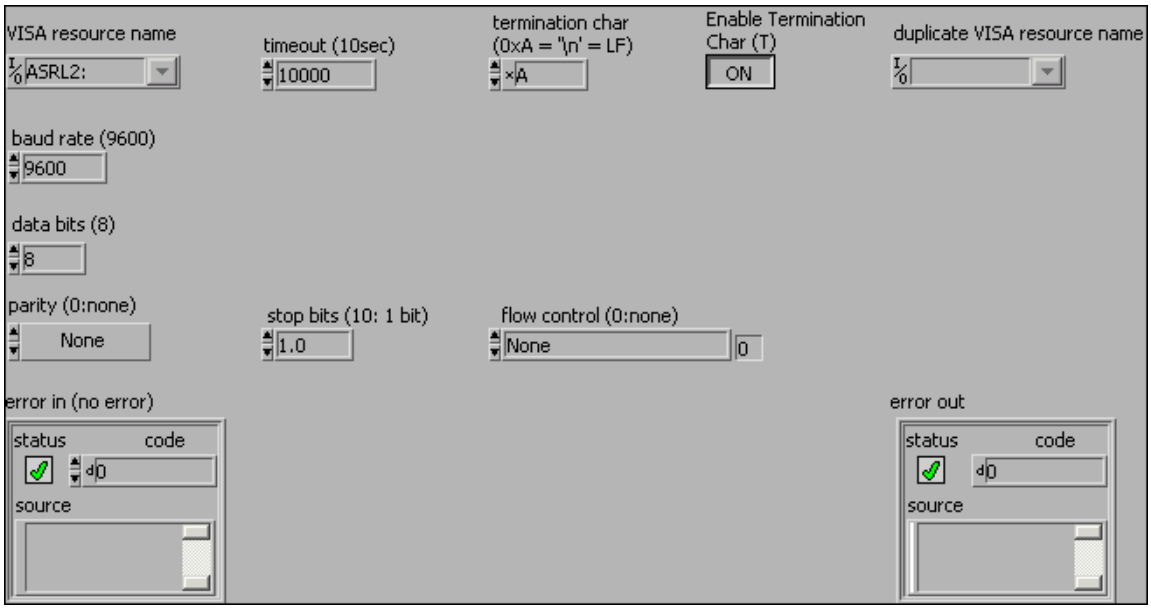


VISA Configure Serial Port (Serial Instr).vi
Initializes the serial port specified by **VISA resource name** to the specified settings. You can use this polymorphic VI to initialize a serial port using the Instr VISA class or Serial Instr VISA class. The VISA class specified in the **VISA resource name** determines the polymorphic instance to use.

Connector Pane



Front Panel



Controls and Indicators

VISA resource name VISA resource name specifies the resource to be opened. This control also specifies the session and class.

timeout (10sec) timeout sets the timeout value for the write and read operations.

baud rate (9600) baud rate is the rate of transmission.

data bits (8) data bits is the number of bits in the incoming data.

parity (0:none) parity specifies the parity used for every frame to be transmitted or received.

termination char (0xA = '\n' = LF) termination char calls for termination of the read operation. The read operation terminates when the **termination char** is read from the serial device.

error in (no error) error in describes error conditions that occur before this VI or function runs.

- **status status** is TRUE (X) if an error occurred or FALSE (checkmark) to indicate a warning or that no error occurred.
- **code code** is the error or warning code.
- **source source** describes the origin of the error or warning and is, in most cases, the name of the VI or function that produced the error or warning.

stop bits (10: 1 bit) stop bits specifies the number of stop bits used to indicate

the end of a frame.

Enable Termination Char (T) Enable Termination Char prepares the serial device to recognize **termination char**.

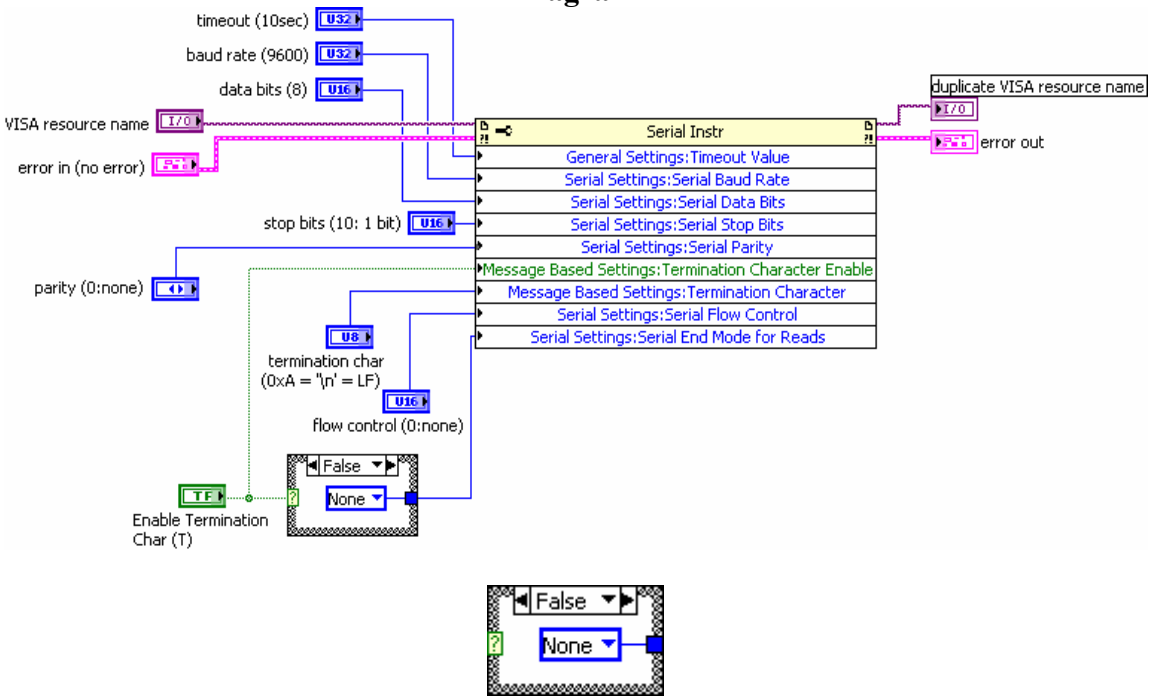
flow control (0:none) flow control sets the type of control used by the transfer mechanism.

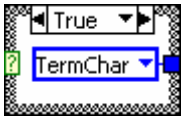
error out error out contains error information. If **error in** indicates that an error occurred before this VI or function ran, **error out** contains the same error information. Otherwise, it describes the error status that this VI or function produces.

- **status status** is TRUE (X) if an error occurred or FALSE (checkmark) to indicate a warning or that no error occurred.
- **code code** is the error or warning code.
- **source source** describes the origin of the error or warning and is, in most cases, the name of the VI or function that produced the error or warning.

duplicate VISA resource name duplicate VISA resource name is a copy of the **VISA resource name** that is passed out of the VISA functions.

Block Diagram





List of SubVIs and Express VIs with Configuration Information

VI Revision History
"VISA Configure Serial Port (Serial Instr).vi History"
Current Revision: 101

Position in Hierarchy