BE/ECE 247B - Bioelectronics
CGE for Biowarfare Detection
March 13, 2007

David Drake
Lisa Eisenman
Nathan Shepard
Joseph Wargo
Outline – Overview

- Outline
- Why Biodefense and Why Capillary Gel Electrophoresis
- Biowarfare Detection Procedure
  - Primer Design
  - Sample Prep
  - Capillary Gel Electrophoresis
  - Signal Processing
  - Detection Logic
- Strengths and Limitations of Technology
- Future of the Technology
- Conclusion and Q&A
Biowarfare agents (BWA): microorganisms or toxic biological products that can cause injury to humans
- Bacteria: anthrax, plague, tularemia
- Viruses: smallpox, ebola, marburg
- Toxins: botulinum, ricin

Why BWA detection?
- 2001 anthrax attack made bioterrorism a realistic threat to the US
- US spent $4.2B for civilian biodefense in 2006 alone\(^1\)

\(^2\) http://www.rap.ucar.edu/projects/shield/technology/mei_2004urb.htm
\(^3\) http://www.eqecat.com/abscorporatesolutions/TRIA.html
Capillary Gel Electrophoresis Technology

- Charged particles in solution can be separated based on charge and size
  - DNA has net negative charge

- DNA length determines effective charge and radius
  - Small DNA fragments have high electrophoretic mobility
  - Large DNA fragments have low electrophoretic mobility

- Gel properties can be manipulated to allow only chosen DNA fragments to migrate

\[
\mu_i = \frac{Q_{\text{eff}}}{6\pi\eta R}
\]

\(\mu_i\) effective electrophoretic mobility [cm\(^2\cdot s^{-1} \cdot V^{-1}\)]
\(Q_{\text{eff}}\) effective charge of the ion [C]
\(R\) total radius of the ion [cm]
Strengths and Limitations of CGE

Strengths

- High efficiency
  - Hundreds of components can be simultaneously separated
- Easily automated
- Low sample requirement
- Low reagent requirement
- Low Joule heating
- Quantitative measurement

Strengths and Limitations of CGE

Limitations

- Error due to Electrodispersion
  - Difference in sample zone and buffer conductivity
- Single gel that separates a wide range of different length DNA strands has not been developed
- Only molecules that can carry a charge can be separated
- Solute-wall interactions

Sample to Detection Process

Primer Design

Data Analysis

Sample Preparation

CGE Process

Save the World
Clever Primer Design – The Foundation for Detection

Pre filter all of the genomic data for the target organisms

Divide the remaining sequence into all possible primers

Primer Filter
- Tm melting range
- G’s or C’s in a row
- GC Content
- Self dimer checks
- Hairpin check
- Differences between primer ends
- Primer length

Binding Filter
- $48 \leq Tm_{\text{mismatch}} \leq 62$
- Hamming $\leq 2$
- 3’base match = 4
- Coverage $> 50$

Multiplex
- Choose primers that can be multiplexed together successfully

Coverage
- Design primer sets that allow for the greatest coverage of target organism and AVOID background

Final Choice
- Computer validation
- Manual analysis
- Chemical validation.

Pairing Filter
- $300 < \text{prod length} < 400$
- Amplicon checks-
  - primer dimer
  - secondary structure
Gives Specificity Needed for Detection

- **PCR Primer Design**
  - Primers designed to specifically amplify only intended target organisms
  - Entire set of primers designed for simultaneous multiplexing
  - All amplicons designed to be within the CGE detection “window” (~300 – 400 bp)
  - Multiple amplicons are designed to target each threat
  - Each primer pair is chosen to produce an amplicon length within the threat window
Sample Preparation for CGE

- Air collection done on filter
- Lysis done to expose DNA material within spores
- DNA purification done to prepare sample for PCR
- Capillary Gel Electrophoresis is not sensitive enough to work with raw sample – must be amplified through PCR
- Multiple CGE can be ran in parallel due to small input volume to improve detection results (repeats of data)
**Capillary Gel Electrophoresis**

- DNA fragments are sent through a capillary filled with gel
- Movement driven by electrical current
- Output data are the lengths of the masses of PCR amplicons

**Fluorescent Peaks Represent Masses of PCR Amplicons**
DNA Migration Theory – Ferguson Plot

- Fragment migration follows two main theoretical models based on size.
- Ogston model: fragments are small enough to fit through the pores in the gel.
- Biased reptation model: fragments are too large to fit through the pores in the gel without a considerable amount of deformation and stretching of the molecules.
- Leads to two distinct regions in the Ferguson plot.
- Linear interpolation used to find migration times of unknown samples.
Error rates vary from 0% to 5% with no apparent systematic error

- For a particular amplicon, migration is constant, even if it doesn’t agree with actual length (run to run variability less than 1%)
- Though sequence analysis does not reveal trends, G, C, A, and T may migrate differently due to size
- Bioinformatic data is not complete or accurate
Proper peak identification and time data gives ability to create an equation that relates migration time to the size of the DNA.

Heating decreases viscosity which decreases migration time.

Calibration peaks picked out of the data.

Calibration length-to-time line is shifted to go through the two calibration points of the data.
Best Fit Model – Improvement over Commercial Software

Step 1

- Spline fit the peaks found in the ladder file to create a length time calibiration curve – This file is done once for each set of runs (12 samples in one run)
Signal Processing – Agent Identification and Detection

- Signal Processing Algorithm must process resulting data – peak finding and length discrimination

Sample peaks: size of peaks given in bp (based on calibration to the marker peaks and a sizing ladder run with each chip)

- **Y-axis** = Fluorescent intensity of DNA (i.e. how much is there)
- **X-axis** = Migration time of DNA (i.e. how big it is)
- **Y-axis** = Fluorescent intensity of DNA (i.e. how much is there)
Best Fit Model – Improvement over Commercial Software

- **Step 2**
  - Find the two calibration peaks in the sample data
  - Use the migration time of those peaks to correct the length time calibration curve

- Corrects for variation seen between sample 1 and sample 12 of a single run

- Results are better than commercially available software
Data Examples

- 2 primer pairs tested alone for each surrogate organism
- 100 org copies

All expected primer lengths found within 5% error

<table>
<thead>
<tr>
<th>Surrogate</th>
<th>Primer Pairs</th>
<th>Expected Lengths</th>
<th>Color</th>
</tr>
</thead>
<tbody>
<tr>
<td>E. Coli K12</td>
<td>104 &amp; 114</td>
<td>400 &amp; 750</td>
<td>●</td>
</tr>
<tr>
<td>E. Coli O157:H7 EDL933</td>
<td>103 &amp; 113</td>
<td>230 &amp; 283</td>
<td>●</td>
</tr>
<tr>
<td>Bacillus cereus</td>
<td>99 &amp; 109</td>
<td>158 &amp; 482</td>
<td>●</td>
</tr>
<tr>
<td>Bacteriophage lambda</td>
<td>101 &amp; 111</td>
<td>134 &amp; 533</td>
<td>●</td>
</tr>
</tbody>
</table>

- Expected Lengths
- Unexpected Lengths

Single-plex data for 8 primer pairs, 4 surrogate organisms
Peak Finding

- To locate peak interpolate with second order polynomial fit (parabola)
- Improves Peak Finding Especially Noisy or Sparse Peaks
- Peak Closest to Interpolated Peak is Chosen as Peak Index
- Also improves time to length calibration curve

**Improved Peak Finder**

- Peak Based on Fit
- Peak Based On Index
- Peak On Indices Red Circle Interpolated Green

![Graph showing peak finding process with improved peak finder](image-url)
Peak Resolution – Peaks too close together

- Peak Resolution:
  - \( R_s = (t_1 - t_2) / (\text{Width Avg.}) \)

- Individual Peaks Resolved when \( R_s > 1.5 \)

- High SNR Required for Detailed Resolution

- Each peak may be modeled as an Individual Waveform for Analysis (Gaussian Waveform commonly used)
Detection Logic

- Detector takes in all the information – how many peaks were present within the detection range, and how likely is that to happen without a threat being present.

- Detection Logic gets more complex if organism discrimination is desired.

- Likelihood of seeing gel peak in correct bin when target is present is $P_d = 95\%$
  - Missed detections from PCR failures, too-low peaks, etc.
  - Likely conservative.

- Likelihood of seeing a peak in any given bin when intended target is not present is $P_{fa} = 1\%$
  - Caused by unintended cross-priming of background, system noise, air bubbles in gel, etc.
  - Can be improved through better primer design.
Approximate Pd / Pfa Trade to Identify Organisms

- Probability of detection (Pd) = likelihood of correctly detecting one target
- Probability of false alarm (Pfa) = likelihood of falsely detecting at least one target

System Pd (solid) and Pfa (dashed) at 2.5 % gel accuracy

Pd goal 95%
Pfa goal $10^{-7}$

Probability

# of hits

N = 1
N = 2
N = 3
N = 4
N = 5
N = 6
Strengths and Limitations of CGE for detection

**Strengths**

- **Speed**
  - Three hour sampling time
  - One hour processing time

- **Sensitivity and Accuracy**
  - Only 8µl of sample is needed for each test
  - 99% with >3ppb sampled from air with careful primer selection

- **Stability**
  - BP length invariant on one gel
  - Temperature invariant
  - Low space/energy footprint

**Limitations**

- **Speed**
  - Four hours between exposure and detection

- **Accuracy**
  - Background interference/noise
  - Peaks close together

- **Maintenance**
  - Gels are one-use
Future Possibilities

- Parallel capillaries to increase sample space
- Increase light path to increase sensitivity
  - Bends in capillary or rectangular capillaries
- Interfacing between CGE and mass spectrometry
- Interfacing with Micellar electrokinetic capillary chromatography
  - Extend biowarfare detection CGE to molecules that cannot carry a charge
Conclusions

- CGE alongside PCR amplification can be used for Bio-Agent detection
  - Accurate, Stable, Can be Automated

- Technology already exists to build and use a CGE bio-detection system
  - PCR, CGE, DSP chips, etc., only need is for integration

- Signal Processing is a powerful tool to account for and correct the chaotic factors inherent in biological testing
  - Flexible Technology that will only get stronger as computing/processing advances
References

- Patrick Camilleri (Editor) - Capillary Electrophoresis : Theory and Practice (New Directions in Organic and Biological Chemistry Series), Published by CRC Pr, ISBN: 084939127X, 1997
- Christoph Heller - Analysis of nucleic acids by capillary electrophoresis (Chromatographia CE Series Volume 1), Published by Vieweg Press, ISBN 352806871X, 1997
- James P. Landers (Editor) - Handbook of Capillary Electrophoresis, Published by CRC Pr, ISBN: 084932498X, 1997
- Kevin D. Altria (Editor), Capillary Electrophoresis Guidebook : Principles, Operation, and Applications (Methods in Molecular Biology, 52) Published, ISBN 0896033154, 1996