MGH Microarray Oligonucleotide Set NHLBI- Program in Genomics Application

Mason W. Freeman, M.D.
Lipid Metabolism Unit
Massachusetts General Hospital

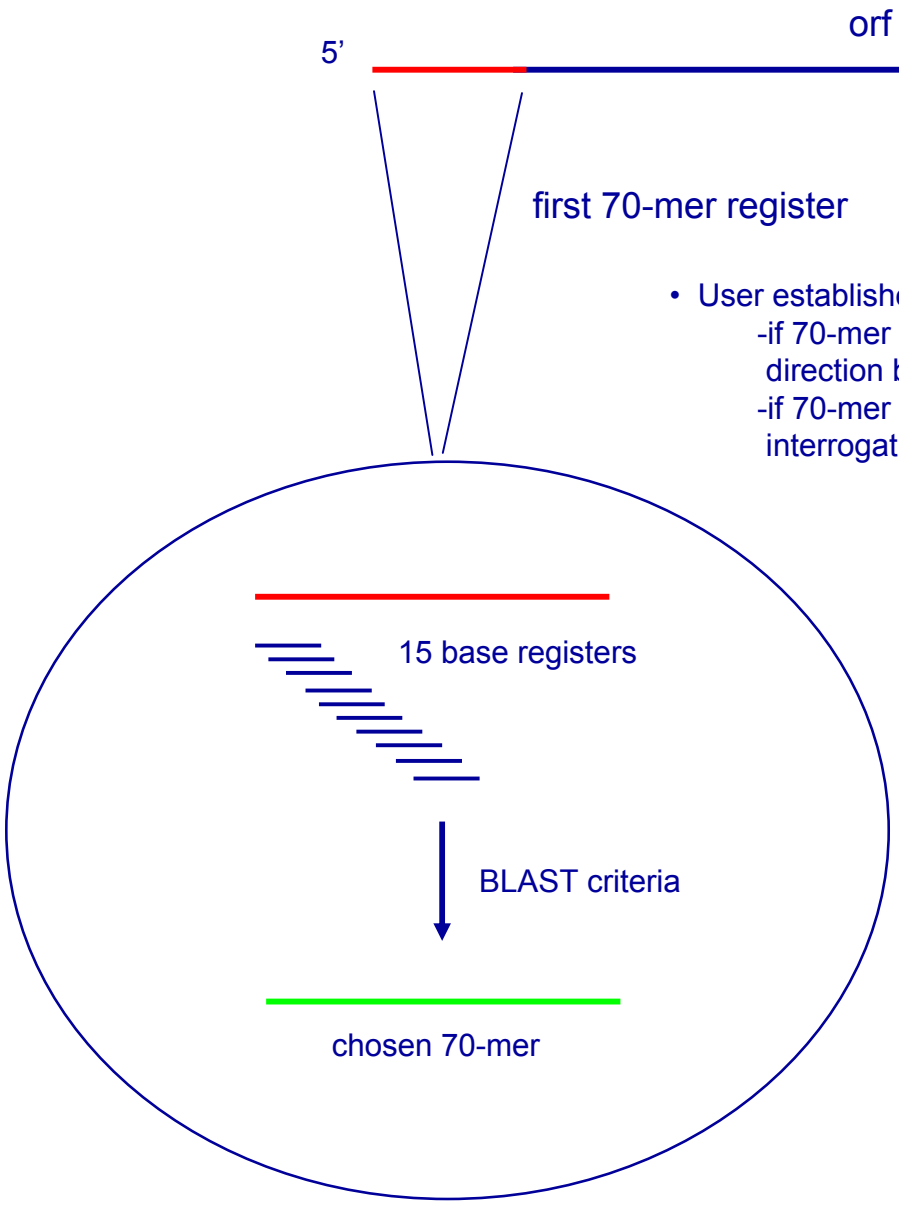
Oligonucleotide choices

- **Most oligos have been selected at 3' UTR**
  - **Oligo dT priming of RNA populations**
  - **Sequence divergence is typically greater in this region**
  - **However, 3' UTR variability is substantial**
  - **Tools for predicting prevalence of alternative 3' ends are still primitive**
- **5' ends also have drawbacks**
  - **Transcription units lacking TATA boxes have end heterogeneity**
  - **Estimates of the distribution of 5' ends are rare**
- **As most investigators want to use microarray profiling to serve as a surrogate for assaying protein prevalence, coding regions are likely to be best surrogate**

Random priming of RNA

- **Permits detection of coding region isoform differences**
- **Will favor 5’ sequences in detection, as primer extension is always toward the 5’ end**
- **? Impact of priming on non mRNA species**

Oligonucleotide Selection Algorithm



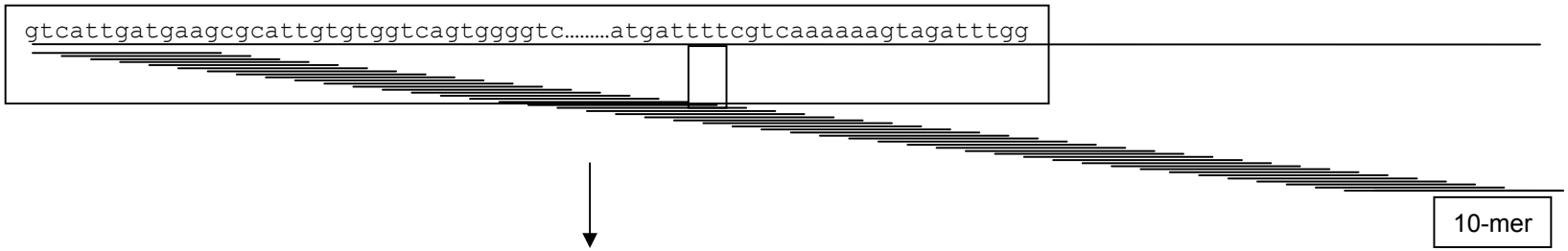
- User established Tm criteria: Tm ≥ 74 °C
  - -if 70-mer fails, move the register in a 5’ direction by one base
  - -if 70-mer passes, proceed to 15 base interrogation

- BLAST each 15 base register for uniqueness against input sequence set
  - -if each 15 base register is unique within a 70-mer, the oligo is chosen

reation of an oligo database containing all possible 10-mer

A

Gene Index = 1000



gtcattgatg → **(1000,1)**,(343, 22),(4442, 201),(4599, 890),(18949,1900), …
tcattgatga → **(1000,2)**,(10225, 455),(14567, 890),(20021,12), …
cattgatgaa → **(1000,3)**,(23,444),(2265, 211),(7895, 2110), …
attgatgaag → **(1000,4)**,(6679,3451), …
ttgatgaagc → **(1000,5)**,(7865, 67), …
tgatgaagcg → **(1000,6)**,(4599,895),(9899,22), …
……

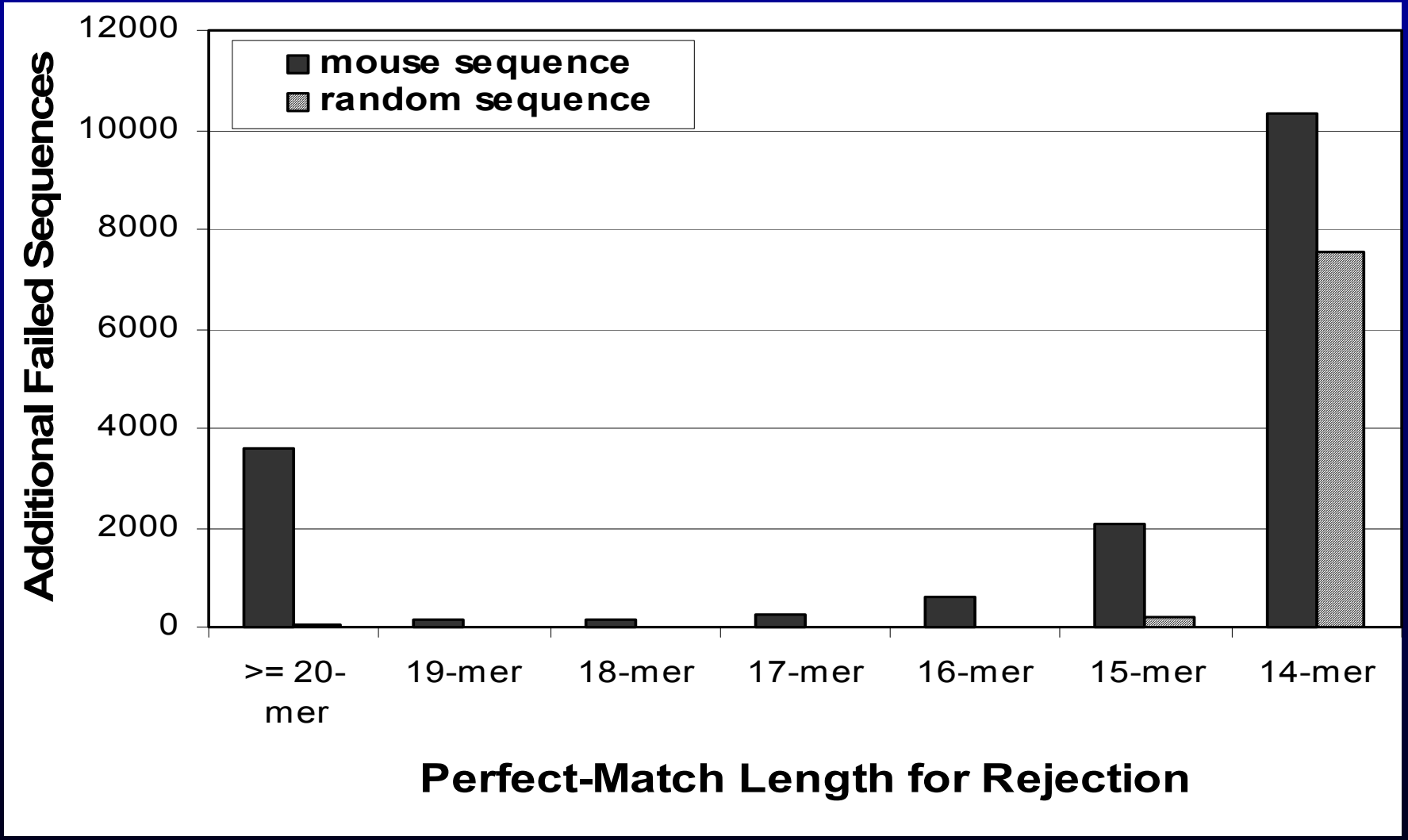
B

gtcattgatgaagcg (15-mer)

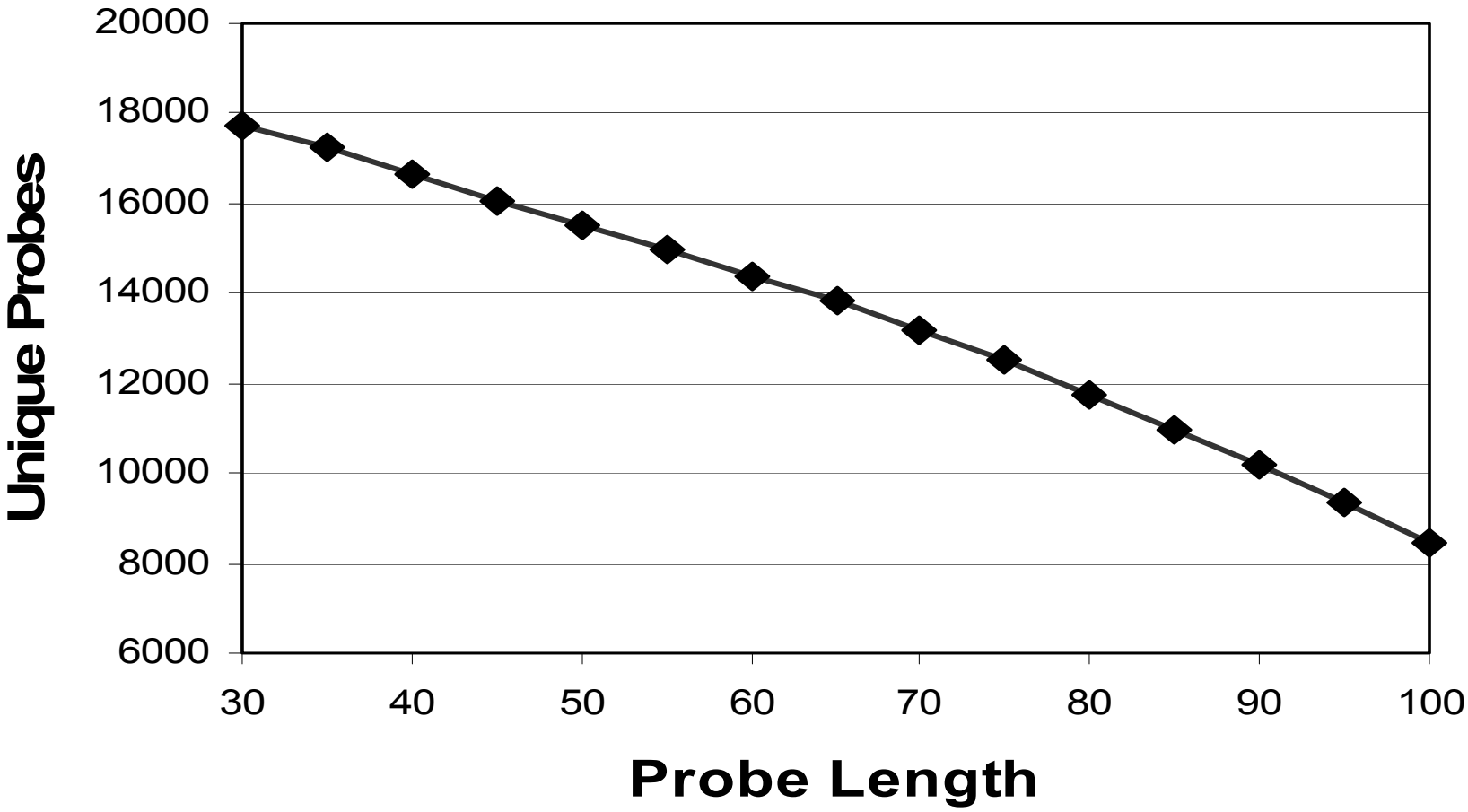
gtcattgatg → (1000,1),(343, 22),(4442, 201),**(4599, 890)**,(18949,1900), …

tgatgaagcg → (1000,6),**(4599,895)**,(9899,22), …

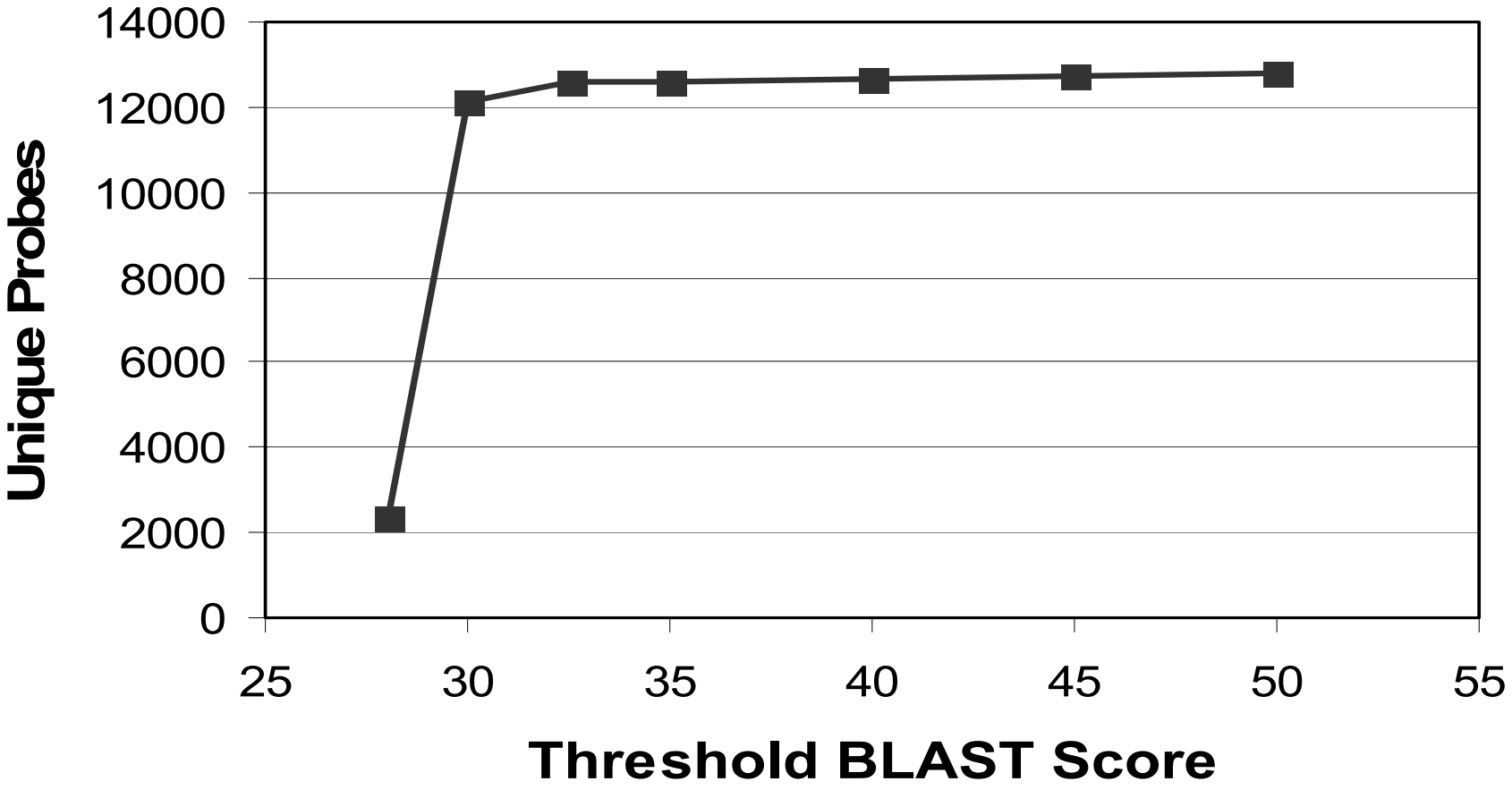
Effect of exclusion length on number of failed oligos



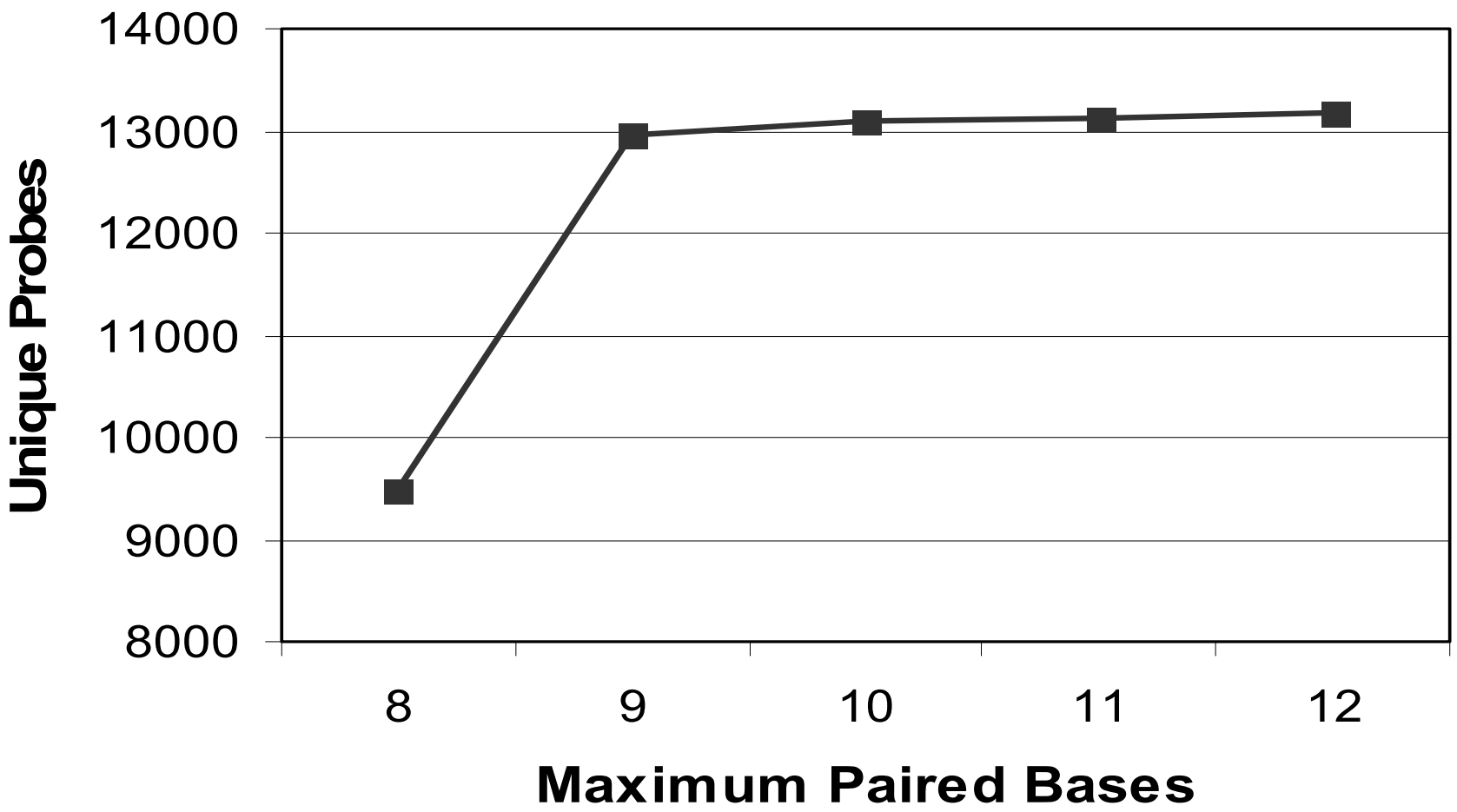
Effect of probe length on ability to pick unique oligos



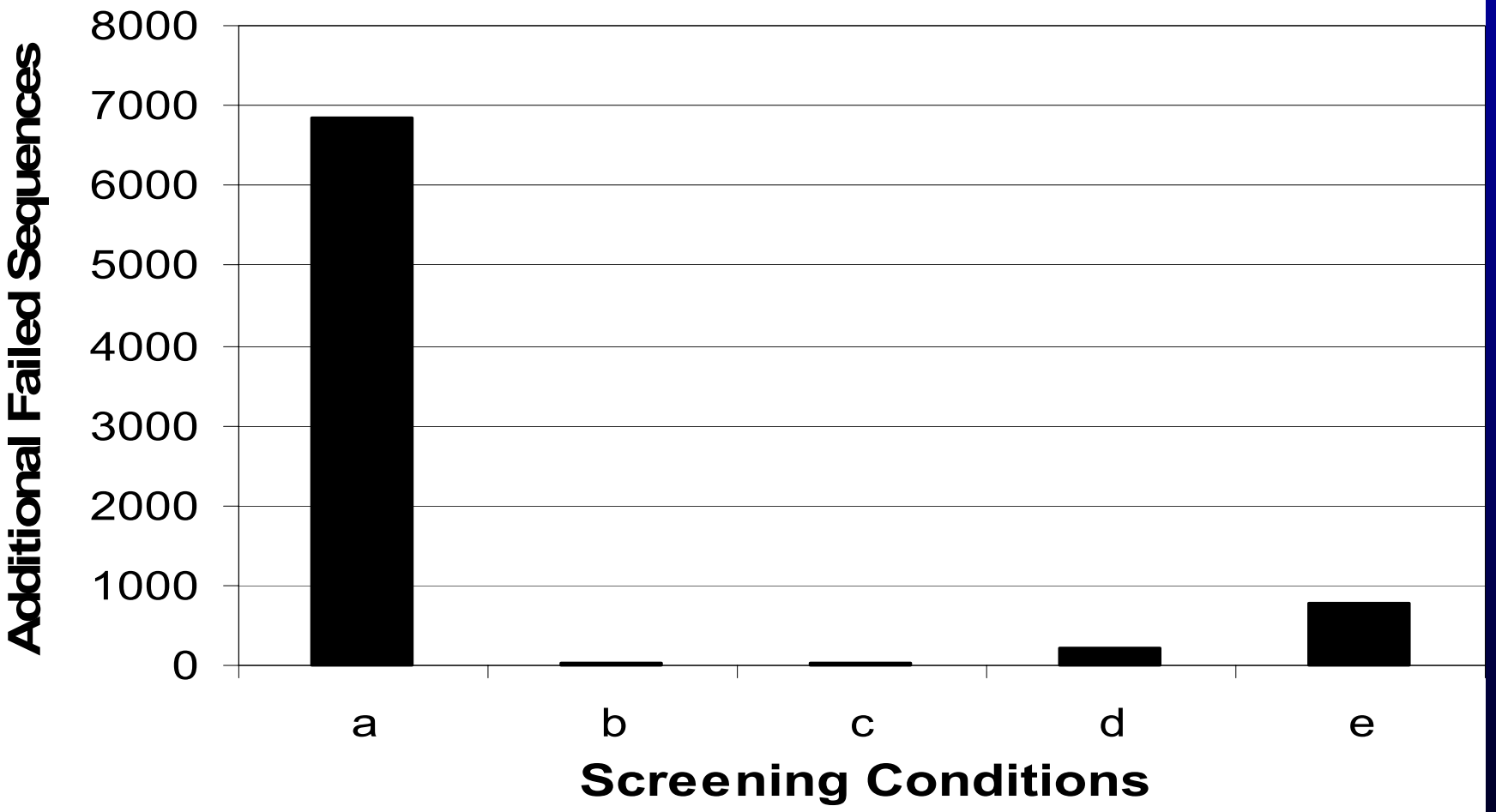
BLAST Score and unique oligos



Maximum Base Pairing and Unique Oligos



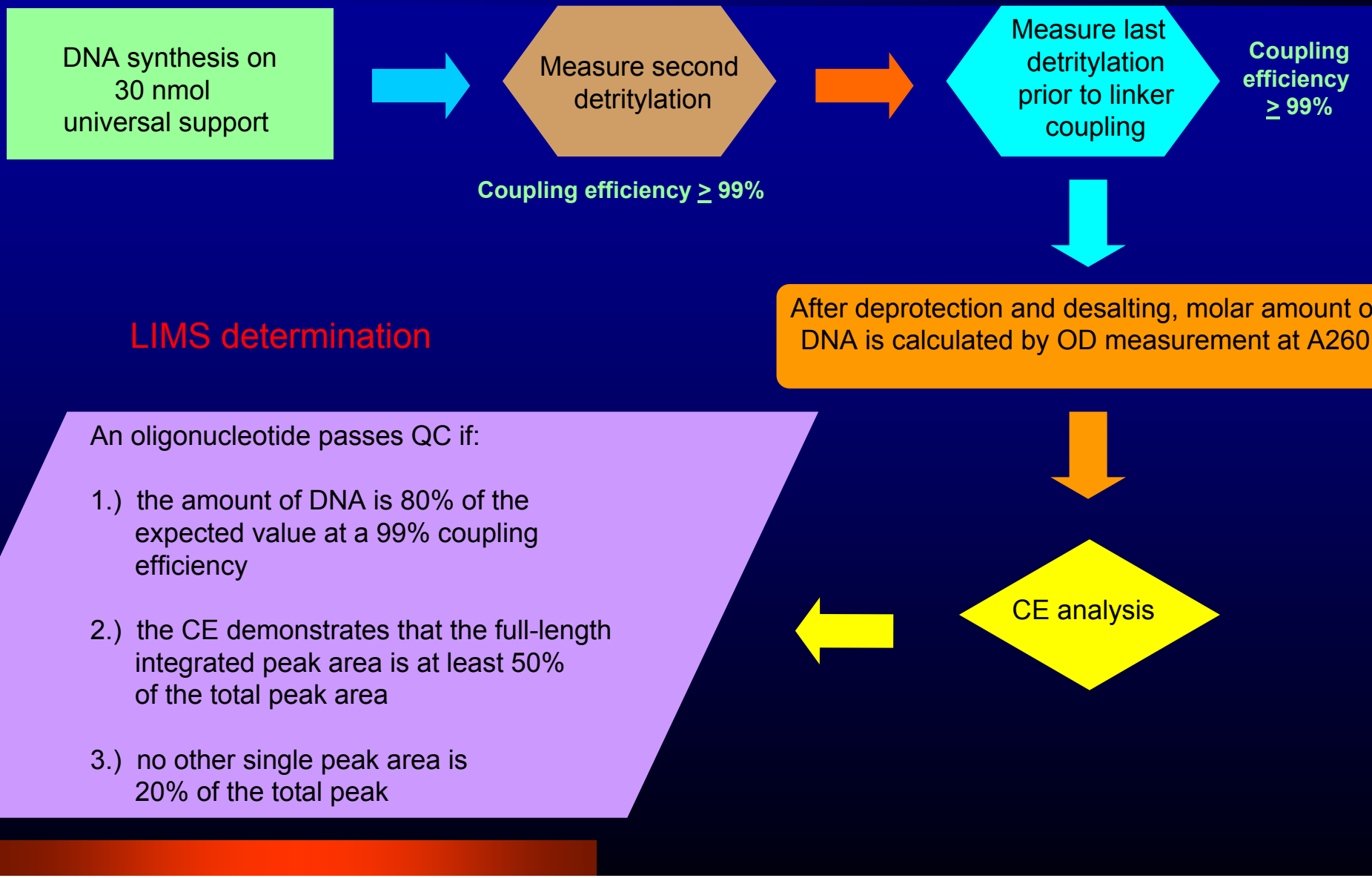
Screening parameters and oligo failures



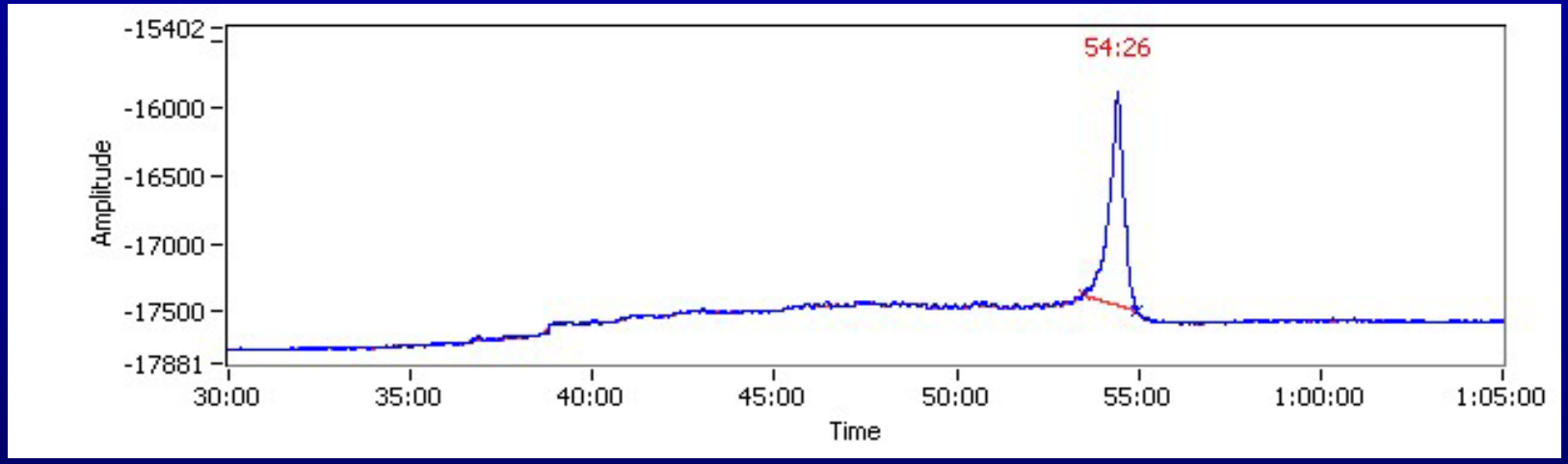
Oligonucleotide generation

- **70 mers synthesized at MGH and UTSW**
- **Quality control measures in DNA synthesis**
  - **Oligonucleotide OD**
  - **First and last trityl**
  - **Capillary electrophoresis**
- **5' amino linker for covalent coupling**
  - **Terminal deoxynucleotide transferase assay**
  - **Random nine-mer hybridization**

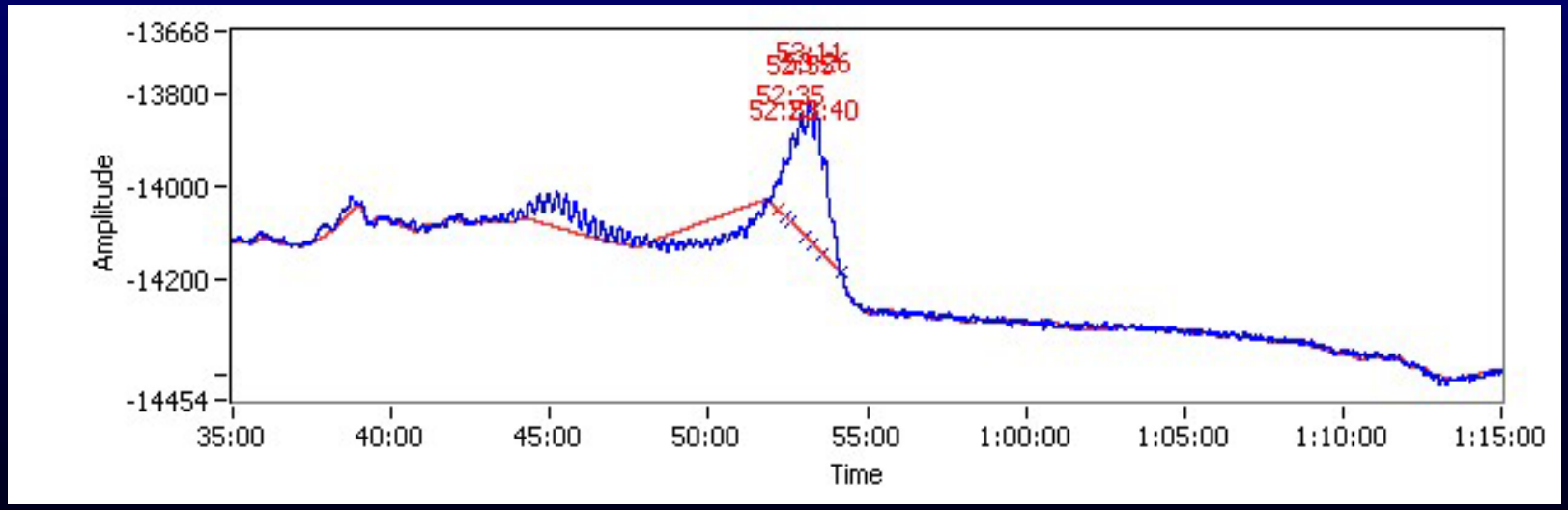
Oligonucleotide Quality Control Scheme



Probe Quality Control by Capillary Electrophoresis



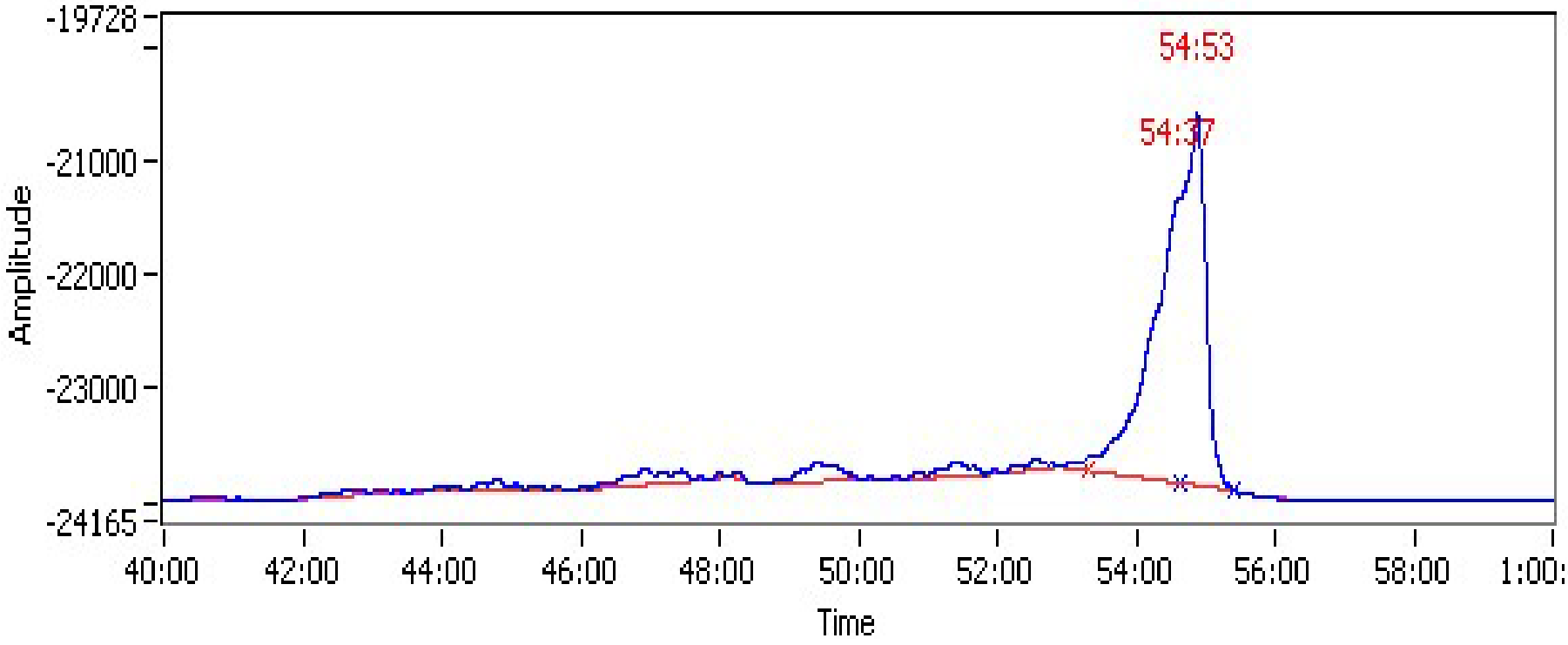
pass



fail

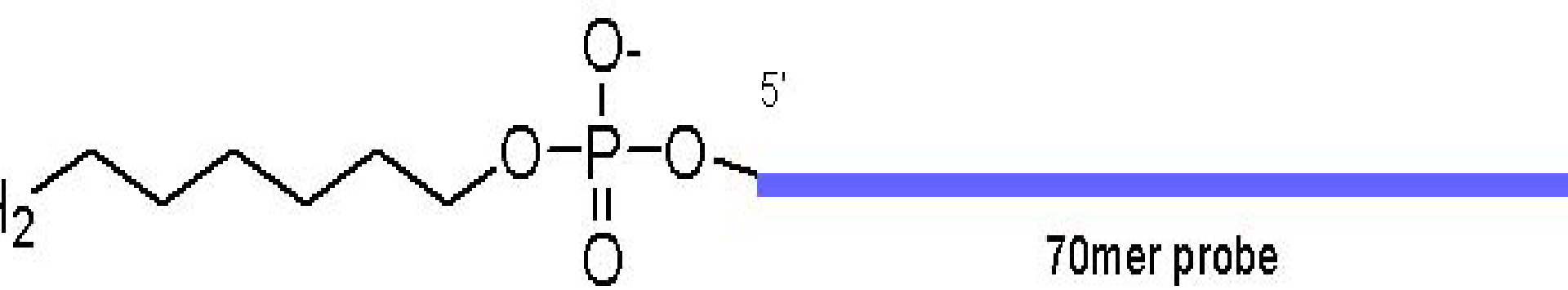
Analysis performed on a CombiSep MCE 2000

CE analysis of equal molar mixture of 70-mer with and without C6-amino linker



Approximately 20 second differential in retention time

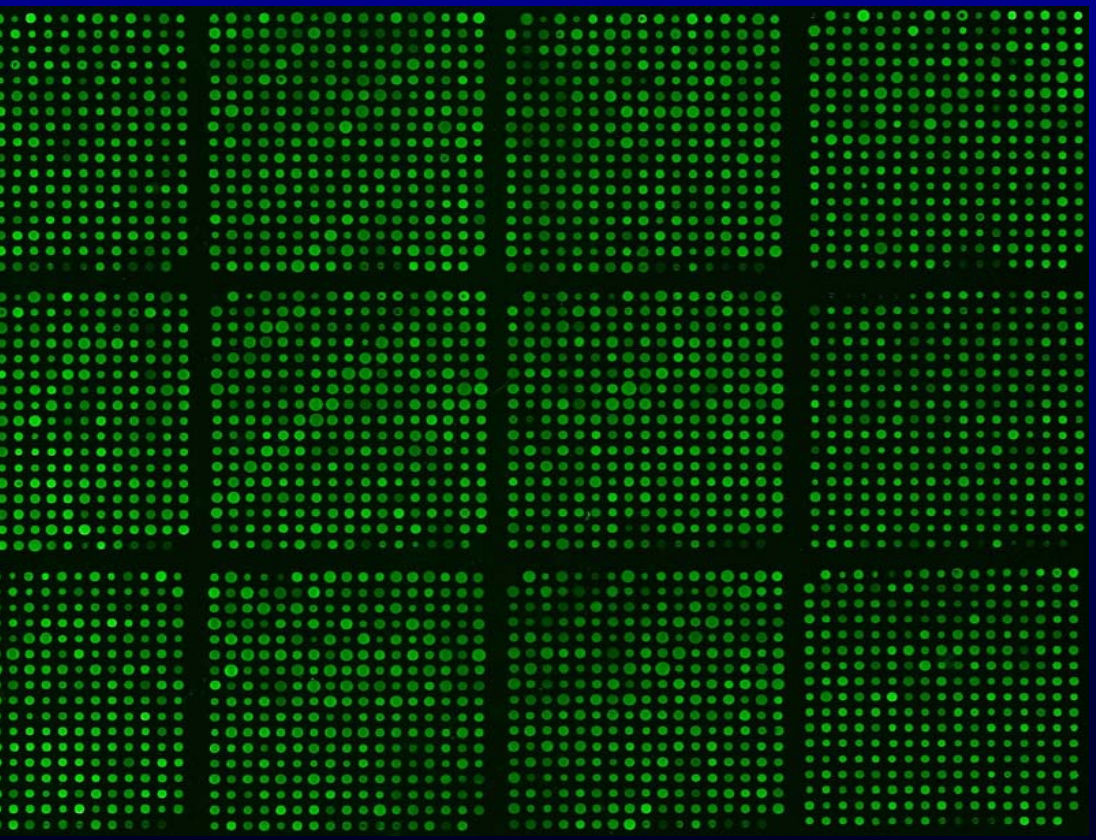
Chemical Structure of the C6-amino linker



1) Linker is added at the end of synthesis process
2) Oligo covalently attaches to slide surface
3) Attachment is to N-hydroxysuccinimde surface
4) Chemistry compatible with any NH_2 attack on coated slide surface

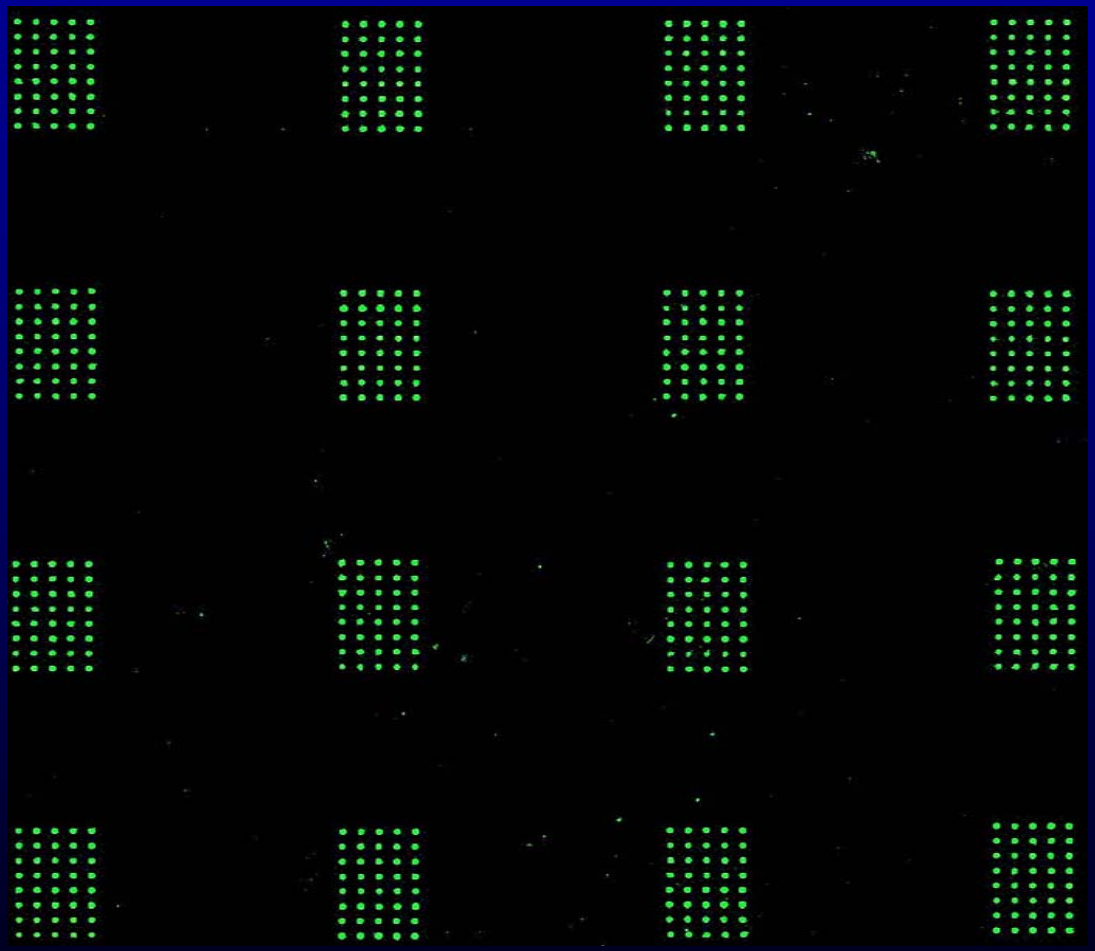
dT labeling of probes at 3’-termini as print QC metho

Mouse Operon Version 1.1 print QC



- After post-processing, slide 1 and 100 of the print-run are QC’d.
- Terminal deoxynucleotidyl transfer (24 units, 12 units/μL, Amersham), 2 μM Cy3-dCTP (Amersham)
- Reaction volume: 124 μL
- Slides are labeled on a GeneTAC hybridization station at 37 °C for 2 h with agitation
- After labeling, slides are washed consecutively with 2xSSC(0.1% SD followed by 2x SSC, and finally wa
- The slides are centrifuged dry and immediately scanned

MGH Mouse oligos 5’-labeled with dCTP-Cy3 by Tdt method



otential advantages of oligos from MGH

- **Cost**
  - **Our oligo sets will sell for ~$25-30,000 per 700 pmol of DNA for a 14,000 gene mouse set**
  - **This amount conservatively yields 3000 slides**
  - **DNA cost is therefore $8-10 per slide**
- **Open source oligo sequence database**
- **Improved annotation tools if set is adopted widely**
- **Incremental up-dating of sets**
- **Will include controls for printing and hybridization**
- **Availability- completion expected Feb 1, 2003**

Contributors

- Shukui Guan, Ph.D.
- Sean Quinlan, M.S.
- Brian Seed, Ph.D.
- Glenn Short, Ph.D.
- Temple Smith, Ph.D.
- Dan Stetson, M.S.
- Xiaowei Wang, Ph.D.
- Karen Armstrong
- Jocelyn Burke
- Jamie Gregory
- Tara Holmes
- Melinda Lee
- Najib El Messadi
- Christine Ordija
- Jing Wang