Final Technical Report - - - - - - - DOE, Grant #: DE-FG03 94 ER 61817

Title: "Advanced Sequencing Technology"
Period of Support: 02/01/94 to 01/31/97 (extension: 2/1/97 to 10/31/98)
Recipient: Univ. of Utah, Sch. of Medicine, Dept. of Human Genetics
Principal Investigator: Dr. Raymond F. Gesteland

This "Advanced Sequencing Technology" project evolved considerably on a number of fronts during the course of its three years (1994-97) plus 1.5 years extension (1997-98).

A major effort throughout was development of instrumentation (hardware and software) for automated, mulitiplexed probing of membranes. This has resulted in a number of very sophisticated machines that are able to probe and image large membranes with one probe after another so that the multiplexed DNA samples on the membrane can each be interrogated sequentially. The membrane to be probed is on the outside surface a drum that is 24 inches in diameter. This drum rotates within a second drum so that its bottom is bathed in a buffer puddle. The buffer composition is programmed to run through cycles of hybridization – washing – detection – imaging – stripping and back to the next round with the next probe. The instrument will operate unattended for 20, eight- hour cycles of hybridization to a membrane with up to a total of 1900 samples, all with automated collection of data. We discovered a novel detection method that made this process possible. Alkaline phosphatase, linked to the probe, acts on fluorogenic substrate whose product binds to the membrane maintaining the spatial localization of the probe (patent issued). The fluorescent pattern of bands is imaged while the drum rotates by excitation with a band of light and capture of the signal on a CCD. All aspects of the process are computer controlled; including temperatures, fluid deliveries, times and scan rate. In addition failsafe systems will sense problems or low fluid levels and interrupt the process without loss.

This automated probing technology was developed as a general tool with a number of applications. Multiplexed sequencing was realized. Efficient base calling software was developed , (this is now being used by Molecular Dynamics). Suitable read lengths and throughput were achieved and were competitive with early ABI machines. The 1.908 Mb genome of *P. furiosus* and a 300kb of a human BAC were sequenced with a combination of multiplex and ABI sequencing. However, the advantage of multiplexing was marginalized by the advent of longer reads and denser sample loading of the ABI and of capillary sequencers.

DISCLAIMER

The probe chambers are extensively used for mapping of transposon inserts, and sets of vectors ideally suited for multiplex mapping and cloning are in use.

The probe chambers have really come into their own for large-scale genotyping. At our center Mark Leppert has genotyped STR markers in 250,000 samples from 1400 people in disease studies by automated, multiplexed probing. Each membrane has 3600 genotyping samples. This has proven to be very effective and is in daily use. The annual throughput of one chamber is several million genotypes/year. A number of instrument units have been built. Complete specs and plans are available.

We investigated the possibility of using abasic DNA for sequencing, especially with the possibility of simplifying and substantially improving applicability of mass spectrometry for sequencing. The idea was that if the DNA bases could be removed from Sanger sequence products, all the information about sequence is still present in the families of fragment lengths. Analysis of these by gel electrophoresis for size would, in principle, be easier since structure complications (e.g. compression) would disappear. Also with the simpler chemistry and smaller mass, the fragments might be much more suitable for analysis by mass spectrometry. After extensive efforts, it is clear that complete removal of all bases is very difficult – neither chemical nor enzymatic tricks were good enough. Secondly, abasic DNA is surprisingly unstable and not very amenable to manipulations. This project is on hold waiting for a new insight.

We developed mass spectrometry for determining genotypes. The idea was that for known polymorphisms it should be possible to distinguish between the possibilities by accurate determination of the masses of the two strands of PCR products. Using an electrospray quadropole mass spec, we worked out conditions to measure masses of DNA strands up to 120 nucleotides long with an accuracy of +/- 4 Da (or 60 nucleotides, +/- 2 Da). This is sufficient to distinguish which of the 4 bases is present at a single base polymorphic site. For genotyping applications aca. 50 nt sequence is amplified from genomic DNA with the polymorphic site in the middle between the two primers. After minimal clean up, masses of the 4 products (2 strands from two chromosomes) are determined by electrospray MS. Both alleles are seen and the two strands give redundant information, which increases accuracy. The front end of the MS is automatable with use of 96 well dishes so that unattended collection of data is possible. In addition we have been successfully multiplexing PCR products such that as many as 5 polymorphic sites can be determined simultaneously. Our genetic center has implemented this system for genotyping thousands of people as part of disease gene finding projects. The simplicity, speed and inherent accuracy make this a very favorable method for SNP analysis.

We have also been developing MS technology for proteome analysis. Our interest has been driven by our study of "recoding" – mechanisms by which certain mRNAs use special sequence signals to redirect reading of the genetic code. Examples of recoding include programmed frameshifting, redefinition of stop codons (e.g. to insert selenocysteine) or to bypass internal mRNA sequences. These processes often result in more than one protein coming from one mRNA, adding to the proteome diversity and to points at which levels of protein expression are controlled. While we know a lot about a few recoding examples, we know little about the overall contribution this makes to the proteome. This prompted us to develop methods to ask how many proteins really come from each gene, including recoding, modifications, etc. We are testing technology on specific mammalian genes of interest and on the proteome complement of yeast. The goal is to have efficient technology that will allow global monitoring of proteome expression profiles.

Referenced Papers:

1. Brown, J.R., Robb, F.T., Weiss, R., and Doolittle, W.F. 1997. Evidence for the early divergence of tryptophanyl- and tyrosyl-tRNA synthetases. J. Mol. Evol., Jul;45(1):9-16

2. Riera, J., Robb, R.T., Weiss, R., and Fontecave. 1997. Ribonucleotide reductase in the archaeon Pyrococcus furiosus: a critical enzyme in the evolution of DNA genomes? Proc. Natl. Acad. Sci. U.S.A. 94:475-478.

3. Ghosh, M., Grunden, A.M., Dunn, D.M., Weiss, R., and Adams, M.W. 1998. Characterization of native and recombinant forms of an unusual cobalt-dependent proline dipeptidase (prolidase) from the hyperthermophilic archaeon Pyrococcus furiosus. J. Bacteriol. Sep;180(18):4781-9.

4. Grentzmann, G., Ingram, J.A., Kelly, P.J., Gesteland, R.F., and Atkins, J.F. 1998. A dual-luciferase reporter system for studying recoding signals. RNA 4:479-486.

5. Nelson, C., Ward, K., Leppert, M., and Gesteland, R. 1998. Multiplex genotyping by analysis of PCR products using electrospray mass spectrometry. Proc. Amer. Soc. for Mass Spectrometry Conf., Orlando, FL.

6. Nelson, C., Nelson, L., and Ward, K. 1998. Genetic profiling of trinucleotide repeat regions and polymorphisms in DNA by electrospray mass spectrometry. Proc. Amer. Soc. for Mass Spectrometry Conf., Orlando, FL.

7. Nelson, C., Nelson, L., and Ward, K. 1998. Electrospray Mass Spectrometry of PCR Products: A powerful new technique for genetic screening and mutation detection. Amer. J. Obstetrics and Gynecology, p. 2, 178(1), Annual Meeting, Miami Beach, FL.

8. Nelson, L., Nelson, C., Cragun, D., and Ward, K. 1998. Identification of alanine repeat regions and polymorphisms in the HOXD13 gene by electrospray mass spectrometry in a synpolydactyly family. Amer. J. Human Genetics 63(4), suppl. to vol. 63, Progr. and Abstr.

9. Ward, K., Nelson, L., and Nelson, C. 1998. Mutiplex genotyping of angiotensinogen and thrombolitic mutations using electrospray mass spectrometry. Amer. J. Human Genetics 63(4), suppl. to vol. 63, Progr. and Abstr.

10. Maeder,D.L., Weiss, R.B., Dunn, D.M., Cherry, J.L., González, J.M., DiRuggiero, J., and Robb, F.T. 1999. Divergence of the Hyperthermophilic Archaea Pyrococcus furiosus and P. horikoshii inferred from complete genomic sequences. Genetics 152:1299-1305.

11. Nelson, C., Nelson, L., and Ward, K. 1999. Large-scale population screening of genetic variants using ESI/MS analysis of multiplex PCR products. Proc. Amer. Soc. for Mass Spectrometry Conf., Dallas, TX.

12. Roy, R., Mukund, S., Schut, G.J., Dunn, D.M., Weiss, R., and Adams, M.W. 1999. Purification and molecular characterization of the tungsten-containing formaldehyde ferredoxin oxidoreductase from the hyperthermophilic archaeon Pyrococcus furiosus: The third of a putative five-member tungstoenzyme family. J Bacteriol, Feb 15;181(4):1171-1180 .

Department of Energy
Patent Certification

INTERIM [ ] FINAL [x]

DOE Prime Contract / Grant No.: DE-FG03 94 ER 61817

Subcontract No.: N/A

Contractor / Grantee Name: Raymond Gesteland

Address: University of Utah, 6160 EIHG, Salt Lake City, UT

If the appropriate response is "NONE," that response must be entered in the space provided. A separate list may be used to supply the information required provided the list is specifically referenced in the space provided.

Contractor / Grantee hereby certifies that:

1 - All procedures for identifying and disclosing subject inventions as required by the patent clause of the contract / grant have been followed throughout the reporting period.

2 - The following is a complete list of technical reports prepared during the course of the work under this contract / grant and the DOE office to which the reports were sent:

3 - There were no subcontracts or purchase orders involving research, development and demonstration except as follows: (If such subcontract or purchase order was issued, please include a copy.)

4 - No inventions or discoveries were conceived or first actually reduced to practice in performance of work under this contract / grant other than the following (Certification _____ includes _____ does not include all subcontracts):

Contractor Docket No.	Title	Inventor(s)	Date Reported To DOE	Doe Docket No.
SEE ATTACHMENT				

5 - The completion date of this contract / grant is: 10/31/1998

6 - The following period is covered by this patent certification:

From: 7/5/1994 (Month / Day / Year) To: 6/15/1999 (Month / Day / Year)

Date: 6/16/99 Signature: Brent Brown

Please submit directly to:
William C. Daubenspeck, Patent Attorney
U.S. Department of Energy, Office of Counsel
Intellectual Property Law Division
P.O. Box 808, L-376
Livermore, California 94550
(510) 422-4367

Title: Licensing Associate

Printed Name: Brent Brown

Title: _____

Telephone No.: 801-581-7792

Contractor Docket No.	Title	Inventors	Date Reported to DOE	DOE Docket No
U-2079	Magnetic Dna Sequencing Device	Mark D. Stump	09-12-94	not received, a copy of report letter enclosed
U-2212	Rapid, Accurate Identification of Variants in DNA Sequence by Electrospray Mass Spectrometry	Chris E. Hopkins, Raymond F. Gesteland, Pam Crain and James McCloskey	07/19/95	S-84,590
U-2636	An Efficient Method to Characterize the Multiple Protein Products of Individual Genes	Raymond F. Gesteland and John F. Atkins	02/26/98	S-92,073