**HIGH SPEED DNA SEQUENCE ANALYSIS
BY MATRIX-ASSISTED
LASER DESORPTION MASS SPECTROMETRY**

Final Report

for Period February 15, 1991 – February 14, 2001

Lloyd M. Smith

University of Wisconsin-Madison
Department of Chemistry
1101 University Avenue
Madison, WI 53706

February 2001

DOE Patent Clearance Granted

Mark P. Dvorscak

Mark P. Dvorscak

(630) 252-2393

E-mail: mark.dvorscak@ch.doe.gov

Office of Intellectual Property Law

DOE Chicago Operations Office

Apr 23, 2002

Date

Prepared for

**THE U.S. DEPARTMENT OF ENERGY
AWARD NO. DE-FG02-91ER61130**

This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the United States Department of Energy, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product or process disclosed or represents that its use would not infringe privately-owned rights.

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

FINAL PROGRESS REPORT

This grant had as its focus three major areas, as follows:

- i) to develop chemistry and enzymology to permit the enzymatic synthesis of 2' fluoro modified Sanger sequencing reactions, which would be resistant to fragmentation during the MALDI process.
- ii) to develop rapid MALDI analyses of DNA sequence polymorphisms using Peptide Nucleic Acid (PNA) DNA analogs in conjunction with solid phase chemistry.
- iii) to study the fundamental mechanisms occurring in the MALDI analysis of nucleic acids.

With respect to goal i), we made substantial headway on this problem. Our work was described in a publication,

Ono, T., Scalf, M. and Smith, L.M. 1997. 2'-fluoro modified nucleic acids: polymerase-directed synthesis, properties, and stability to analysis by matrix-assisted laser desorption/ionization mass spectrometry. Nucleic Acids Res., **25** (22), 4581-4588.

However, as this work proceeded, we became concerned that this was not actually the right approach to the effective mass spec-based sequencing of DNA, for two reasons: first, it turned out to be quite difficult to enzymatically synthesize DNA which was completely modified at every non-T nucleotide with fluorine. It was fairly straightforward to put in single fluorine-modified nucleotides, or even two, but long contiguous stretches of fluorine-modified nucleotides required very long enzymatic synthesis times, and were not PCR-compatible because of those long synthesis times as well as due to the increased stability of the fluorine-modified duplexes in DNA hybridization. It was difficult to obtain even modest proof-of-principle results as reported in the above publication, and it appeared that to extend this to a practical technology would be very difficult indeed. This problem is not unique to our efforts. Similar issues have arisen in the work of Keller and colleagues at LANL, who wished to enzymatically synthesize DNA molecules with substituent dyes at every nucleotide, and Barbara Ramsey-Shaw at Duke, who wished to incorporate boronated nucleotides at every position. It is hard to convince polymerases to make DNA molecules which are so egregiously modified.

Second, the basic hypothesis of this approach, that the mass range limitation of MALDI sequencing was due to fragmentation of the DNA, began to be called into question. Our

lab obtained results indicating that the mass range limitation, using the matrix 3-HPA, was not due to fragmentation in contrast to our previous thinking. These results were obtained in studies of both synthetic and enzymatically synthesized 2' fluoro-modified DNA.

As we were struggling with these issues, we began to explore a possible new approach to solving the mass range problem, which was to use electrospray ionization (ESI) in conjunction with a new technology developed in our laboratory referred to as CREMS, for "charge reduction electrospray mass spectrometry." The idea here was to generate DNA ions by ESI, rather than by MALDI, and reduce their charge states to singly charged ions. We were able to obtain sufficiently promising preliminary results with this approach to write a successful grant proposal to NIH to provide further support for the work. This work has led so far to four papers, and we believe offers a very promising and exciting new route to the analysis of complex biomolecular mixtures. The papers are

Mouradian, S., Skogen, J., Dorman, F.D., Zarrin, F., Kaufman, S.L., Smith, L.M., 1997. DNA analysis using an electrospray scanning mobility particle sizer. Analytical Chemistry, **69** (5), 919-925.

Scaif, M., Westphall, M.S., Krause, J., Kaufman, S.L. and Smith, L.M. 1999. Controlling charge states of large ions. Science, **283** (5399) 194-197.

Scaif, M., Westphall, M.S. and Smith, L.M. 2000. Charge reduction electrospray mass spectrometry. Analytical Chemistry, **72** (1), 52-60.

Ebeling, D., Westphall, M.S., Scaif, M. and Smith, L.M. 2000. Corona discharge in charge reduction electrospray mass spectrometry. Analytical Chemistry, **72** (21), 5158-5161.

We also collaborated with researchers at LANL during this period to develop new methods for the analysis of DNA and proteins using stable isotopic mass tagging, which resulted in two publications,

Chen, X., Fei, Z., Smith, L.M., Bradbury, E. M. and Majidi, V. 1999. Stable isotope assisted MALDI-TOF mass spectrometry for accurate determination of nucleotide compositions of PCR products. Analytical Chemistry, **71** (15), 3118-3125.

Chen, X., Smith, L.M., Bradbury, E.M. 2000. Site-specific mass tagging with stable isotopes in proteins for accurate and efficient protein identification. Analytical Chemistry, 72 (6), 1134-1143.

With respect to goal ii), we had a very productive and fruitful effort. The basic idea of using PNA probes to analyze SNPs worked very well and resulted in two publications describing our work:

Griffin, T.J., Tang, W., and Smith, L.M. 1997. Genetic analysis by peptide nucleic acid affinity MALDI-TOF mass spectrometry. Nature Biotechnology, 15 (13), 1368-1372.

Griffin, T. J. and Smith, L. M. 1998. An approach to predicting the stabilities of PNA:DNA duplexes. Analytical Biochemistry, 260, 56-63.

We also explored two other approaches, one which required PCR amplified DNA as the input, as described in the publication below,

Fei, Z., Griffin, T.J and Smith, L.M. 2000. Analysis of single-nucleotide polymorphisms by primer extension and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Rapid Commun. Mass Spectrom., 14, 950-959.

and a second which could be employed directly upon human genomic DNA, as described in the following two papers:

Griffin, T.J., Hall, J.G., Prudent, J.R. and Smith, L.M. 1999. Direct genetic analysis by matrix-assisted laser desorption/ionization mass spectrometry. Proc. Natl. Acad. Sci. USA, 96, 6301-6306.

Griffin, T.J. and Smith, L.M. 2000. Genetic identification by mass spectrometric analysis of single-nucleotide polymorphisms: Ternary-encoding of genotypes. Analytical Chemistry, 72, (14), 3298-3302.

With respect to the third goal, of increasing our fundamental understanding of the MALDI process, this is an ongoing interest and concern. Although we certainly cannot claim to understand all aspects of the MALDI process, our understanding is far deeper and more detailed than it used to be. We have published a number of studies in this area, including

Tang, W., Nelson, C.M., Zhu, L., and Smith, L. M. 1997. Positive ion formation in the ultraviolet matrix-assisted laser desorption/ionization analysis of oligonucleotides using 2,5-dihydroxybenzoic acid. J. Am. Soc. Mass Spec., **8**, 218-224.

Tang, W., Zhu, L., and Smith, L.M. 1997. Controlling DNA fragmentation in MALDI-MS by chemical modification. Analytical Chemistry, **69** (3), 302-312.

Tang, W., Krause, J., Zhu, L., and Smith, L. M. 1997. Factors influencing oligonucleotide stability in matrix-assisted laser desorption/ionization mass spectrometry. Int. J. Mass Spec. & Ion Proc., **169/170**, 301-311.

Krause, J., Scalf, M. and Smith, L.M. 1998. High resolution characterization of DNA fragment ions produced by UV MALDI using linear and reflecting time-of-flight mass spectrometry. J. Am. Soc. Mass Spec., **10** (5), 423-429.

In summary, we continue to be fascinated by the unique challenge posed by mass spectrometric DNA sequencing, and although the problem is not yet solved we continue to make progress and find new avenues for exploration. Our work on Mass spec-based SNP analysis was very fruitful and resulted in the development of powerful yet straightforward new MS-based methods for SNP analysis.

We are grateful to the DOE Genome Program for its staunch support over the years in tackling these fascinating, important, and challenging problems at the interface of genetics and chemistry and look forward to a continued relationship.