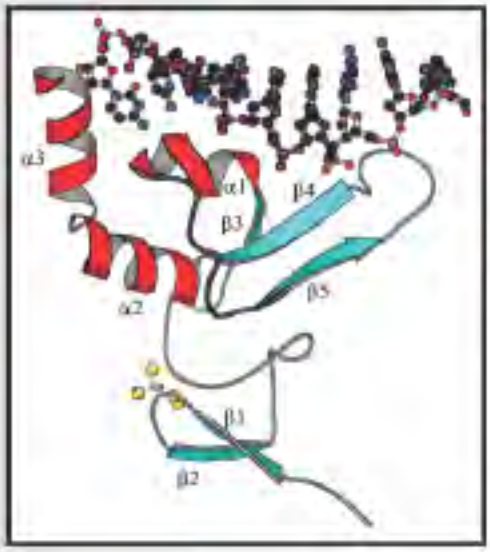
William R. Wiley

EMSL

Environmental Molecular Sciences Laboratory



Annual Report 2000

Macromolecular Structure and Dynamics

June 2001

Pacific Northwest National Laboratory

Operated by Battelle for the U.S. Department of Energy

Cover Photo: ***A molecular model of a single-stranded DNA 9-mer (ball and stick chain) bound to XPA-MBD oriented in the 3′ → 5′ direction that illustrates how the elements of secondary structure are organized in relation to the DNA. The XPA-DNA interactions were characterized by NMR spectroscopy.***

Other post-genomic-era research activities within the Macromolecular Structure and Dynamics Program include structural genomics, DNA recognition and repair, protein dynamics/interactions, and proteomics. The molecular level understanding of biological processes and the impacts of anthropogenic activities on these processes provide fundamental new insights into the intricacies of complex biological machinery.

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 Battelle Memorial Institute, 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, or Battelle Memorial Institute. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-ACO6-76RLO183O

Printed in the United States of America
Available to DOE and DOE contractors from the
Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062;
ph: (865) 576-8401
fax: (865) 576-5728
email: reports@adonis.osti.gov

Available to the public from the National Technical Information Service,
U.S. Department of Commerce, 5285 Port Royal Rd., Springfield, VA 22161
ph: (800) 553-6847
fax: (703) 605-6900
email: orders@ntis.fedworld.gov
online ordering: http://www.ntis.gov/ordering.htm

This document was printed on recycled paper.
(8/00)

PNNL-13626
UC-400

Annual Report 2000

Macromolecular Structure and Dynamics

D. W. Koppenaal, Associate Director
and the Staff of the Macromolecular Structure and Dynamics Program

July 2001

Prepared for the U.S. Department of Energy
under Contract DE-AC06-76RL01830

Contents

1. Introduction

2. Magnetic Resonance Research Staff Reports

3. Magnetic Resonance Research Users Report

4. Mass Spectrometry Research Staff Reports

5. Mass Spectrometry Research User Reports

6. Appendix

1. Introduction

This annual report describes the research and accomplishments in 2000 of the Macromolecular Structure and Dynamics (MS&D) Program, one of six research programs at the William R. Wiley Environmental Molecular Sciences Laboratory (EMSL) - a multidisciplinary, national scientific user facility and research organization (Figure 1.1). EMSL is operated by Pacific Northwest National Laboratory (PNNL) for the U.S. Department of Energy's (DOE's) Office of Biological and Environmental Research. The resources and opportunities within the facility are an outgrowth of DOE's commitment to fundamental research on the physical, chemical, and biological processes that are the foundation for understanding and resolving environmental and other critical scientific issues. Currently, over 100 instrument systems are available to the scientific community, which includes EMSL's resident research staff and their collaborators as well as external users.

The EMSL complex evolved from a 1986 vision by the late Dr. William R. Wiley, Director of Pacific Northwest National Laboratory, to build an innovative multipurpose user facility providing synergism between the physical, mathematical, and life sciences. Dr. Wiley and others believed that molecular-level research was essential for attacking problems in environmental cleanup, energy efficiency, health, and other areas. From this belief grew a concept for a center for molecular science that would bring together theoreticians with expertise in computer modeling of molecular processes and experimentalist from the physical and life sciences.

EMSL's resident research staff (Figure 1.2) conducts fundamental research in areas relevant to DOE missions and affords expert assistance to users, who include university professors, graduate and undergraduate students, and postdoctoral fellows, and scientists from industry, national laboratories, and DOE contractors. Their research programs and collaborative efforts are providing a scientific basis for evaluating a number of social and technical factors associated with environmental issues—factors such as transformation and migration of contaminants in soils and groundwater; cell response to environmental contaminants; reaction of important tropospheric molecules—and a technical basis for waste retrieval and processing.

The MS&D Program is helping to advance DOE interest in bioremediation, environmental effects, and human health by providing fundamental new insights into the intricacies of complex biological machinery. Research focuses on developing a molecular level understanding of biological processes and the impacts of anthropogenic activities on these processes. MS&D's highly specialized staff are currently engaged in post-genomic-era research, including structural genomics, DNA recognition and repair, protein dynamics/interactions, and proteomics. Biological systems and models under investigation encompass both bacterial and mammalian organism. MS&D resources in support of these activities include premier magnetic resonance and mass spectrometric instrumentation systems for structural and mass characterization of biological molecules such as nucleic acids and proteins.

Figure 1.1. As a national scientific user facility, the EMSL supports both open and proprietary research. More information and specific procedures for becoming a user are available on the EMSL Website at http://www.emsl.pnl.gov/.

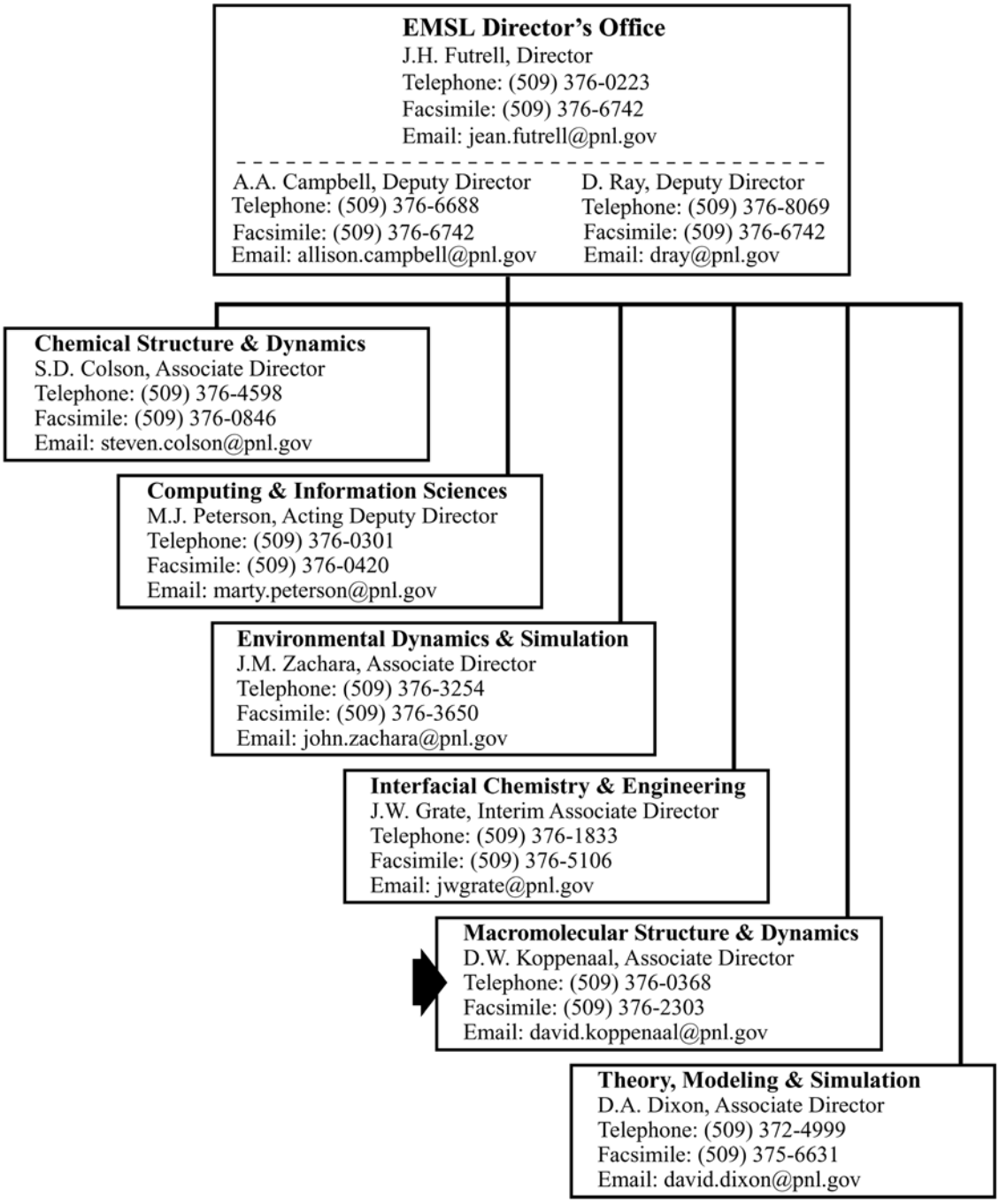


Figure 1.2. Contact information for EMSL research organizations.

Macromolecular Structure and Dynamics Program

The MS&D Program focuses on developing a fundamental, molecular-level understanding of biochemical and biological systems and their response to environmental effects. A secondary focus is in aspects of materials science and catalysis and the chemical mechanisms and processes operative in these areas. The directorate also operates two of the EMSL's high-volume user facilities - the High Field Magnetic Resonance Facility and the High Field Mass Spectrometry Facility (see inserts) - in support of these and other user research programs. Over 160 external scientists and users have been afforded access to these facilities since the EMSL opened in late 1997.

The research interests of the MS&D scientific staff include some of the most exciting areas in modern molecular biology and biochemistry.

- Structural Genomics – Determination of three-dimensional structures of DNA, RNA, proteins and enzymes, and their bonding and intermolecular associations. Particular interests and collaborations exist relative to protein fold classification and sequence-structure-fold relationships.

- Functional Genomics – Determination of the bonding and interaction domains among DNA, RNA, and proteins, with particular emphasis on DNA damage recognition and repair processes.

- Proteomics – Determination and study of the protein repertoire (the *proteome*) expressed by a cell, organism, or tissue; and its dependence on environmental factors, physiology, age, or disease state.

- Biological Imaging – Acquisition of imaging and corresponding chemical information at the cellular, organism, and small animal scales, with particular interest in development of combined magnetic resonance and optical spectroscopy techniques to observe and elucidate biological processes.

- Measurement Science and Instrumentation Development - Development and application of novel and unique nuclear magnetic resonance (NMR) and mass spectrometry (MS) instrumentation techniques for biological and environmental problems.

High Field Magnetic Resonance Facility - Instrumentation & Capabilities

NMR
- 900-MHz NMR (in production)
- 800-MHz NMR
- 750-MHz NMR
- 600-MHz NMR (2 systems)
- 500-MHz WB NMR
- 500-MHz NMR (2 systems)
- 400-MHz WB NMR
- 300-MHz WB NMR (2 systems)

EPR
- EPR Spectrometer with ENDOR/ELDOR capability

(Additional details and applications of these capabilities are provided in this report.)

High Field Mass Spectrometry Facility - Instrumentation & Capabilities

Fourier Transform Ion Cyclotron MX
- 3-T FTICR-MS
- 7-T FTICR-MS
- 11.5-T FTICR/MS
-

Other MS
- Triple-quadrupole MS
- quadrupole-TOF MS
- Ion Trap MS (3)
- Sector-TOF MS

(Additional details and applications of these capabilities are provided in this report.)

This annual report describes our research and accomplishments in 2000. Individual papers representing both staff- and user-initiated research are provided in four sections, listed below, that correspond to our two technical groups and associated facilities.

- Magnetic Resonance Group
- High Field Magnetic Resonance Facility
- Biological Separations and Mass Spectrometry Group
- High Field Mass Spectrometry Facility

Additional information provided within this report includes compilations of facility users and collaborators, facility and instrumental capability descriptions, staff profiles, and research accomplishments (publications, presentations, awards, and honors) that resulted from all MS&D research endeavors.

2000 Highlights

The turn of the millennium brought several notable developments to MS&D staff and capabilities. Activities and achievements of note are described below.

- A 900 MHz narrow-bore NMR system (see Figure 1.3) was demonstrated by Oxford Instruments and Varian in August 2000. Oxford and Varian are under contract to deliver a 900 MHz wide-bore instrument to EMSL, and realization of this system is a significant and promising milestone towards EMSL's high field NMR magnet system. Protein and DNA samples from EMSL were among the first demonstration samples to be run on the new 900 MHz narrow-bore NMR system.

- Proteomics research continued to be the primary focus of the Mass Spectrometry & Biological Separations technical group, headed by Dr. Richard D. Smith. The group successfully completed a major DOE demonstration project involving the global measurement of the proteome of the radiation-resistant bacterium *Deinoccus radiodurans*. A multi-million dollar follow-on project is now underway with funding from the DOE Office of Biological and Environmental Research. Based on these developments, PNNL was made a partner in the DOE's Joint Genome Institute. A new high-throughput 9.4-T FTICR-MS instrument has been ordered to initiate an ambitious proteomics capability expansion (delivery expected April 2001).

Figure 1.3. Varian/Oxford 900-0MHz, narrow-bore NMR system. EMSL will receive a 900-MHz, wide-bore NMR system in September 2001.

- The second EMSL Workshop on Structural Genomics was held in July 2000. Over 60 scientists attended, representing research institutions worldwide. Two outcomes of the meeting included 1) a strong consensus that NMR spectroscopy would play an increasingly significant role in the determination of bio-molecular structure determination and 2) a need exists for a dedicated, national-level NMR facility for structural genomics research.

- The EMSL NMR capability and expertise were part of a successful Structural Genomics Pilot Center grant application to the National Institutes of Health (NIH). The Northeast Center for Structural Genomics, headed by Professor Guy Montelione of Rutgers University, includes Mike Kennedy of EMSL as a co-Principal Investigator (PI). This accomplishment follows a focused MS&D effort to become a player in the structural genomics field.

- Increased levels of research funding were received from NIH, complementing and augmenting MS&D's DOE funding base. During 2000, NIH funding for MS&D research totaled nearly $3M. This demonstrates success in MS&D's strategy to expand proteomics and structural/functional genomics research with NIH support.

- Dr. Richard D. Smith received two internal awards during 2000. He was named both Battelle Inventor of the Year (see Figure 1.4), and also PNNL Mentor of the Year (student and post-doctoral category). Dr. Smith was also nominated for the U.S. DOE's E.O. Lawrence Award. These awards and nominations recognize Dr. Smith's research and prolific accomplishments.

Figure 1.4. Richard D. Smith, (MS&D senior staff scientist), center, receiving 1999 Battelle Inventor of the Year award. Adrian Roberts (PNNL), left, and Mark Kontos (Battelle), right, also pictured.

- Success continues in our magnetic resonance microscopy research endeavors. A combined magnetic resonance and optical microscopy capability is being developed and demonstrated using EMSL's 500-MHz wide-bore NMR instrument. A second-generation combined microscope was designed and built in 2000. Concurrent NMR and confocal microscopy images can be obtained now on single cells.

2. Magnetic Resonance Research Staff Reports

Combined Optical and Magnetic Resonance Microscopy Development and Implementation for Cellular Research

R. A. Wind, E. J. Ackerman, G. R. Holtom, P. D. Majors, and K. R. Minard

Supported by PNNL Cellular Observatory funding (Project No. 90001)

During the past year, we successfully developed and tested the world's first combined scanning confocal fluorescence (OM) and magnetic resonance microscope (MRM) for the simultaneous, real-time analysis of living cells (Wind et al. 2000). We summarize our initial combined microscopic imaging results and their application for the enhancement of MRM resolution and contrast using OM as *a priori* information.

Introduction

OM and MRM are highly complementary modalities: OM of fluorescently stained cells provides high spatial and temporal image resolution of cell structures, while MRM provides a wealth of chemical and biophysical information in the form of spin densities, chemical shifts, diffusion rates, and NMR T_1 and T_2 relaxation times. These NMR measureables can be obtained alone or in the form of image contrast, albeit with lower resolution than OM. Further, MRM provides complementary information in opaque or partially opaque cell systems, where OM is of limited use.

Simultaneous application of OM and MRM allows for the chemical and biophysical characterization of different cell populations and organelles. OM and MRM also can be combined synergistically (i.e., the data obtained with one technique can be used to guide and optimize measurements with the other methodology). For example, the high-resolution optical images can be used to select volumes of interest to perform local MR spectroscopy and to increase the resolution and contrast in the MR images. Therefore, the integrated OM/MRM will significantly improve the capabilities of both methodologies for cellular research.

Combined OM/MRM Instrument

Figure 2.1 shows a sketch of the combined OM/MRM instrument. Housed at the EMSL NMR User Laboratory at PNNL, it utilizes an 89-mm diameter, vertical-bore, 11.7-T Oxford magnet (WB500), into which a bottom-loading OM probe and a top-loading MRM probe are inserted. The sample compartment contains an optical window and a solenoid NMR coil wound onto a horizontally aligned, 0.82-mm ID silica capillary sample tube. The sample tube is part of a perfusion system used to transport the cells to the test section and to maintain cell viability. (The confocal beam enters the sample transverse to the coil and is partially attenuated by the coil turns; therefore, the coil design is a compromise between NMR sensitivity and correctable optical distortion.)

For the OM component, an isotropic resolution in the xy-plane of 1 micron can be achieved over a 2x2-mm^2 field of view, whereas the point-spread function in the z-direction perpendicular to the plane has a half-intensity width of about 33 microns. The typical isotropic MRM resolution is 20 microns.

Combined Microscopy of *Xenopus* Oocytes

Initial testing of the combined microscope was performed using *Xenopus laevis* oocytes. These amphibian cells are studied intensively with light and confocal microscopy because of their unique biology and relevance to human cells. In the current study, they were ideal for testing because their sizes (0.2 to 1.3 mm) enabled detailed imaging with MRM.

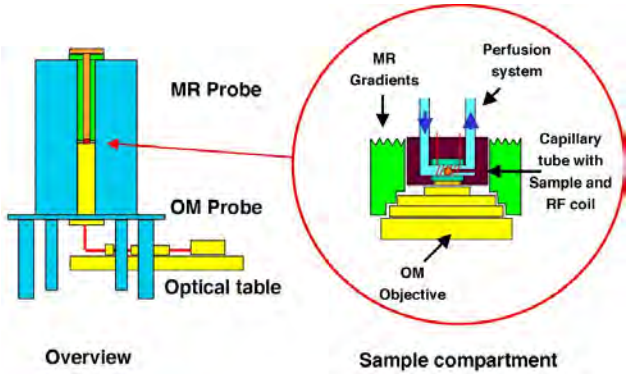


Figure 2.1. Sketch of the combined OM/MR microscope.

Confocal fluorescence images and MR water and lipid images were obtained on *Xenopus laevis* oocytes of different growth stages. Prior to the image experiments, the oocytes and their surrounding follicle particles were stained with rhodamine-123, a nontoxic fluorescent dye selective for active mitochondria. Then the stained oocytes were injected into the perfusion system, filled with Barth's medium, and the flowing medium transported the oocytes into the sample chamber. In order to register and calibrate the OM and MRM image spaces, a 0.53-mm translucent polystyrene bead was injected prior to insertion of the oocyte cells. In Figure 2.2, two-dimensional OM and MRM images of a same plane through the bead and a 0.62-diameter, stage-3 oocyte are shown. With MRM, the distribution of both water and mobile lipids were imaged. In the OM image, the right side of the bead and the top right side of the oocyte are poorly defined due to the presence of three small oocytes in the optical path and below the OM plane. These small oocyte produce shadows in the plane of interest. The high OM intensity at the oocyte boundary is believed to arise from the mitochondria in the surrounding follicle cell layer, while the interior is optically opaque. Additional features include stained connective tissue near the oocyte and a fluorescing layer along the inner wall of the sample tube. We conclude that excellent image registration of both images is obtained. Figure 2.3 shows similar confocal and MR images obtained on a smaller (0.38-mm-diameter) stage-2 transparent oocyte. In contrast to results obtained on the larger oocyte, the confocal image clearly shows the mitochondrial cloud; whereas, in the water MR image, only an inside layer of enhanced water intensity is detected. The results clearly illustrate that combined microscopy provides significantly more information than that obtained with each of the techniques individually. The MR images provide detailed information about the intracellular structure of the larger opaque oocyte that is not observed with OM. On the other hand, in the smaller transparent oocytes, the high-resolution optical images can be used to complement the relatively low-resolution MR images.

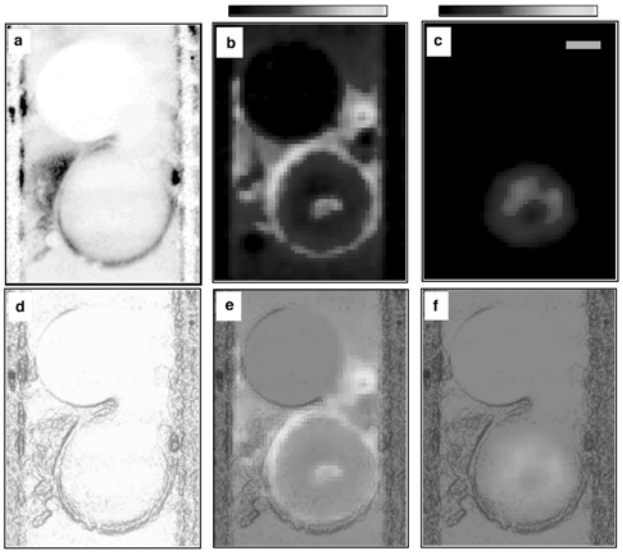


Figure 2.2. Two-dimensional combined OM and MRM images of a 0.53-mm-diameter polystyrene bead (top object) and a 0.62-mm-diameter, stage-3 *Xenopus laevis* oocyte (bottom object) in an 0.82-mm ID glass capillary tube. (a) an OM image; (b) a water-selective MR image; (c) a lipid-selective MR image; (d) a OM relief contour image, obtained from image (a); (e) an overlay of the water MR image with the OM relief image; and (f) an overlay of the lipid MR image with the OM relief image. The scale bar shown in (c) is 0.2 mm in length.

First OM/MRM Application: Sharpening of the MR images. The *a priori* knowledge provided by the confocal image can be used to improve the boundary resolution and the contrast in the MR images, as illustrated in Figure 2.3. By overlaying the high-resolution optical contour plot, given in Figure 2.3b, with the relatively low-resolution MR image shown in Figure 2.3a, image pixels can be identified in the MR image containing the boundary between the oocyte and the surrounding medium. Then the average intensity in each of these pixels can be redistributed into each compartment inside these pixels. Figure 2.3c shows the resulting MR image. It follows that both the boundary resolution and contrast are significantly enhanced. Hence integrated OM/MRM can be used to produce images in which the optical spatial resolution is combined with the MR contrast, which could be important for a variety of applications including the diagnosis of diseased cells.

References

Wind, R. A., K. R. Minard, G. R. Holtom, P. D. Majors, and E. J. Ackerman. 2000. "An Integrated Confocal and Magnetic Resonance Microscope for Cellular Research." *J. Magn. Reson.*, **147**, 371-377.

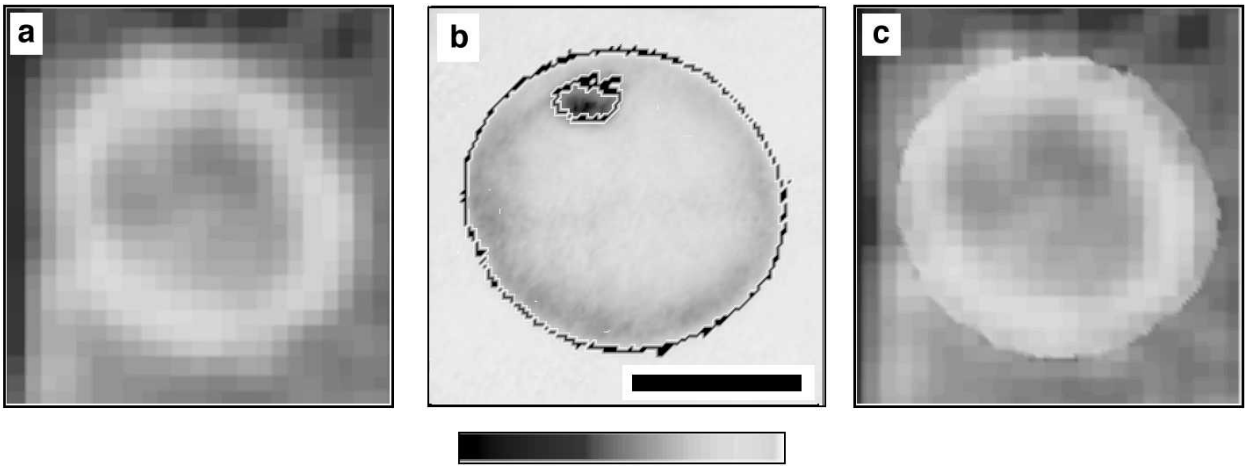


Figure 2.3. Application of confocal image data to enhance the resolution and contrast of MR images at object boundaries. (a) MR water image of the stage-2 oocyte; (b) OM image of the same oocyte, and the contour plot, obtained from this image; and (c) MR image with enhanced resolution and contrast near the contour boundaries. The scale bar shown in (b) is 0.2 mm in length.

Integrated Optical and Magnetic Resonance Microscopy for Mammalian Cell Cultures

R. A. Wind, D. S. Daly, G. R. Holtom, P. D. Majors, K. R. Minard, and B. D. Thrall[(a)]

Supported by National Cancer Institute (NCI) Funding

(a) PNNL Molecular Bioscience Resource

We recently demonstrated a prototype, integrated optical/magnetic resonance microscope (OM/MRM) for the study of live cells *in vitro* (Wind et al. 2000). This instrument provides simultaneous, complementary, three-dimensional OM and MR microscopy information for large single cells or mammalian cell agglomerates (Minard et al. 1998) in a perfused test tube environment. We are currently developing a second OM/MRM to study live mammalian cell cultures adhering to a flat microscope slide (i.e., with a configuration similar to that used with traditional white light microscopy). Optical images of selectively stained cells will be used to identify individual healthy and diseased cells, and (in conjunction with statistical methods) resolve the MR spectra into healthy versus diseased contributions, even when the different cell types are intermixed. (MR methods provide a wealth of biochemical and biophysical information about the cells.) Further, we will assess whether MR images alone can be used to discriminate diseased from surrounding healthy cells. Then the high-resolution optical images can be used to sharpen the lower-resolution MR images (Wind et al. 2000), thus combining the optical resolution with the MR contrast. This may result in an enhanced diagnostic capability to detect diseased cells (e.g., in pap smears and biopsies) and to follow the effects of therapy with enhanced specificity.

A new MR detection coil and sample chamber was developed that orients the microscope slide (containing live cells) perpendicular to the axis of the magnet. This has the advantage of allowing the optical probe already developed for the first combined microscope to be used for examining monolayer cell cultures, while preserving optimum MR sensitivity. The new coil is a microscopic variant of the so-called butterfly coil used in medical MR imaging. Figure 2.4 shows a photograph of the butterfly coil. The dashed circle denotes the area in which the MR sensitivity varies by less than 10%. The diameter of the circle is 2.2 mm. Assuming a confluent cell density of roughly 2000 cells per square mm, this means that the coil can measure up to 8000 cells in a single monolayer.

Figure 2.5 shows the sensitivity profile for the butterfly coil. This profile was measured by placing a water sample next to the coil surface and performing an MR image. The y-axis denotes distance from the coil surface in microns. Analysis of the image intensity along the y-axis suggests that even at 500 microns

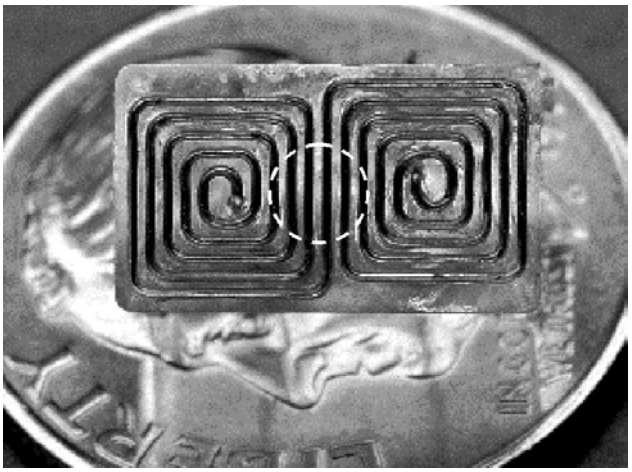


Figure 2.4. The butterfly coil on a U.S. dime.

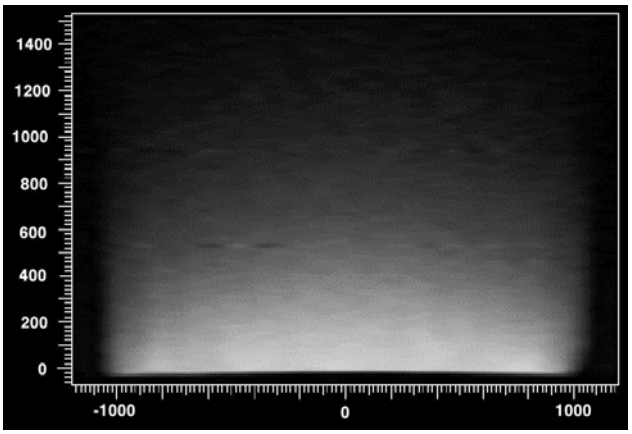


Figure 2.5. MR response of butterfly coil.

away from the coil the sensitivity is still about 50% of its maximal value. This means that at least two neighboring sample slides (cell monolayers) can be studied simultaneously.

Figure 2.6 shows a microscopic proton MR-only image for a small colony of mouse-skin epithelial (JB6) cells containing about 4000 cells, collected with the butterfly MR coil. This amount of cells is at least three orders of magnitude less than used in a traditional MR probe, thus illustrating the sensitivity of the coil. The cells were attached to a 130-micron-thick glass sample slide and placed adjacent to the coil. The image shown in Figure 2.6 was obtained in 30 minutes and has a planar resolution of 20x80 squared microns. The proton MR spectrum shown in Figure 2.7 was collected as an average of 1000 measurements to improve the signal-to-noise ratio. In addition to the

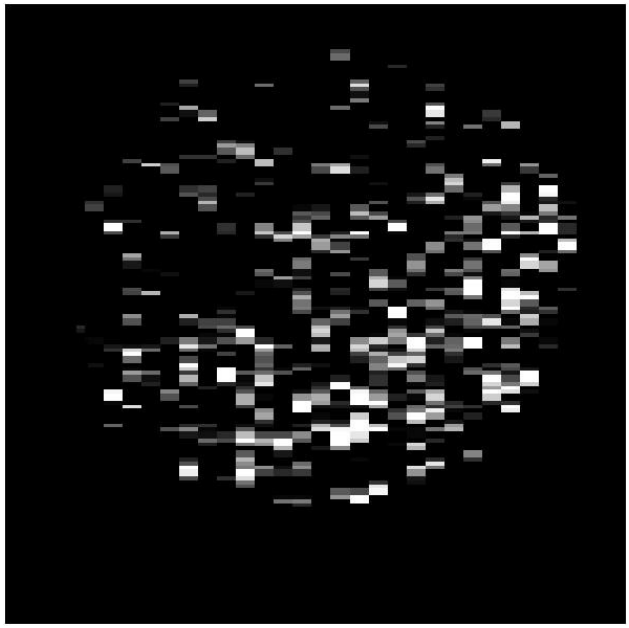


Figure 2.6. Proton image of JB6 cells.

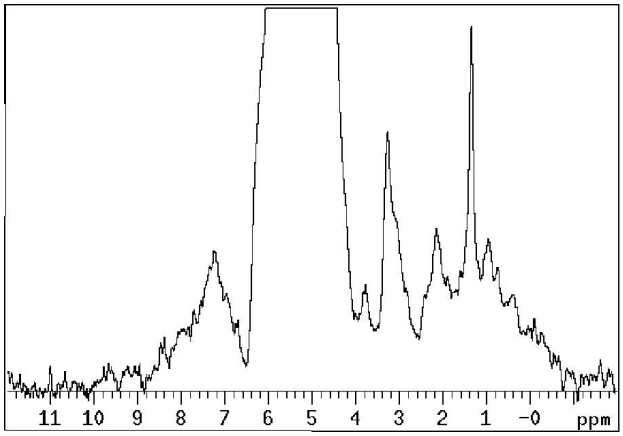


Figure 2.7. Proton MR spectrum of JB6 cells.

large water line, several metabolites and other mobile compounds are observed, including choline, creatine, and mobile lipids. Both the image and the spectrum have signal-to-noise ratios that are sufficiently large for the study of small cell populations in a reasonable period of time.

Figure 2.8 shows a sketch of the sample chamber. This chamber has perfusion and variable temperature capabilities, and cells can be placed into the chamber via the two removable microscope slides E and F. The top and bottom windows are mounted permanently. The top window contains a coating on

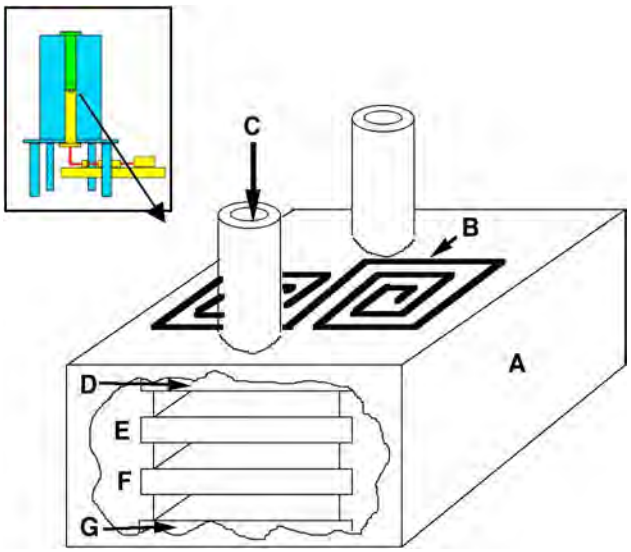


Figure 2.8. Sketch of the sample chamber. A) sample holder; B) NMR butterfly coil; C) perfusion inlet; D) fixed top window; E, F) removable sample slides; G) fixed bottom window. The inset shows an overall sketch of the combined microscope, consisting of a superconducting magnet with a top-loading MR insert, a bottom-loading optical subsystem, and the sample chamber at the interface.

its bottom, into which a grid pattern is laser-etched. Figure 2.9 shows a portion of this etched pattern, which serves as a fiducial marker to calibrate the optics space and the NMR imaging space. In practice, the fiducial system is imaged by both OM and MR, and used to accurately determine the spatial transformations (translations, rotations, and rescaling) needed to coalign the images. These spatial transformations are subsequently applied to the OM and MR data for the microscope slide cell samples, which are situated adjacent to (in front of) the fiducial system. At present this combined microscope is approaching its final stage of construction, and we hope to start testing the instrument in the near future.

References

Minard, K. R., X. Guo, and R. A. Wind. 1998. "Quantitative 1H MRI and MRS Microscopy of Individual V79 Lung Tumor Spheroids." *J. Magn. Reson.*, **133**, 368-73.

Wind, R. A., K. R. Minard, G. R. Holtom, P. D. Majors, and E. J. Ackerman. 2000. "An Integrated Confocal and Magnetic Resonance Microscope for Cellular Research." *J. Magn. Reson.*, **47**, 371-377.

Figure 2.9. Optical (left), MR (right), and combined (1:1 intensity overlay; center) microscopy images of a 20% portion of the fiducial system. The optical resolution is 2 microns per pixel and the MR resolution is 15 microns per pixel. The letters in EMSL are 80 microns tall.

Determining Particulate Matter Dosimetry and Clearance in Live Mice Using High Resolution Magnetic Resonance Imaging

K. Minard, R. Corly,[(a)] D. Rommereim, C. Timchalk,[(b)] and R. Wind

Supported by PNNL Laboratory Directed Research and Development Funding

(a) PNNL Molecular Bioscience Resource
(b) PNNL Chemical Dosimetry

Numerous epidemiological studies have linked airborne particulate matter (PM) with human morbidity and mortality. To assess potential health risks, biological response models are needed for extrapolating PM toxicity data derived from animal studies to relevant human exposures. This, however, requires large amounts of data for defining model parameters and for validation.

Unfortunately, current methods for determining PM dosimetry involve large numbers of animals and serial sacrifice. Moreover, localization of PM in specific regions of the respiratory tract can be problematic because of the elaborate, costly, and time-intensive sample preparation required. In addition, it currently is not possible to follow deposition and clearance of particles in a living animal unless particles are tagged with a gamma-emitter, and this approach has several limitations. Thus, there is a critical need to develop analytical techniques for determining PM deposition and, if possible, clearance in living, normal animals as well as in animals that are predisposed to cardiopulmonary disease. The focus of this project is to

determine whether MR imaging can be employed for this purpose. If successful, this knowledge will provide significant improvements in the design and cost of whole-animal PM toxicity studies and, therefore, aid in the development and validation of biological models to predict localized "dose" in the human respiratory tract.

Particle Testing

In clinical applications of MR imaging, observed contrast is routinely manipulated by introducing paramagnetic molecules for altering the relaxation properties of surrounding water in biological tissues. In this study, we are testing whether MR imaging can be used for visualizing the deposition and clearance of paramagnetic carbon black particles, particularly, since these particles are known to be relatively benign when inhaled in small amounts.

In NMR experiments performed in EMSL, T_1 relaxation measurements were performed using carbon black particles suspended in water. These experiments showed that particles filtered with a 20-micron mesh had an average relaxivity of ~10 sec^{-1} per milligram particles per milligram water. The experiments also showed that the phagocytosis of these particles by aveolar macrophages leads to a dose-dependent T_1 for intracellular water as shown in Figure 2.10.

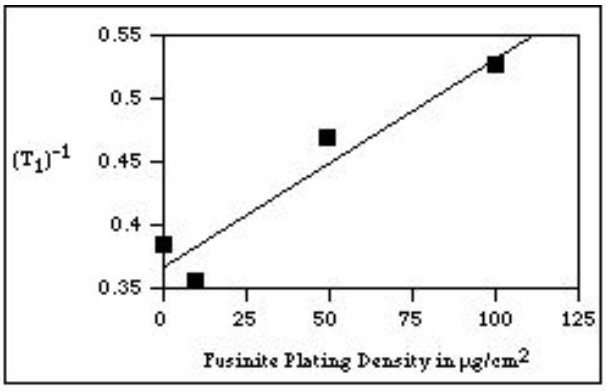


Figure 2.10. Measured $(T_1)^{-1}$ for intracellular water in aveolar macrophages as a function of particle plating density.

Imaging Technology

To quantify the concentration of paramagnetic particles in the lung parenchyma, appropriate methods were developed for mapping the T_1 of water in lung tissue. This is illustrated in Figure 2.11 where a T_1 map of the lung in a dead mouse is shown along with the raw image data from which it was derived. The

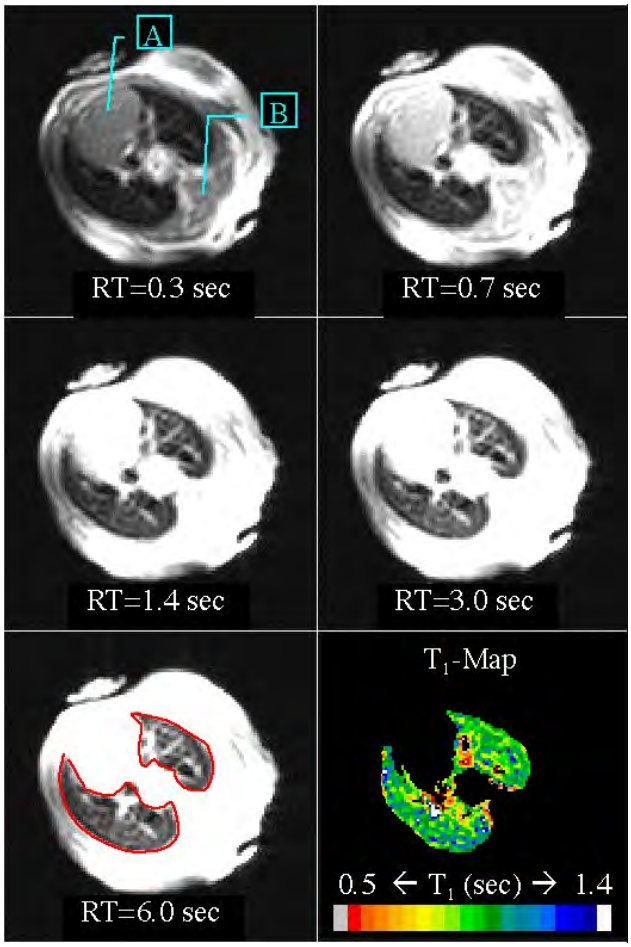


Figure 2.11. A spatial map of the T_1 for water in lung tissue is created by first collecting a series of five T_1-weighted MR images at 2.0 Tesla. Each was acquired using a saturation-recovery sequence run with a different recovery time (RT). All images show the same 2.0-mm-thick slice in which the heart (A), spinal column (B), and lung parenchyma (outlined in red) are clearly visible. By post-processing the raw image data, a synthetic T_1-map of the lung parenchyma is created where contrast only depends on the measured T_1 in each pixel of the image data. In the example shown, planar resolution is 0.5 x 0.5 squared millimeters and raw image data required 40 minutes to collect.

results suggest that maps of T_1 with submillimeter spatial resolution can be derived from MR image data requiring only about an hour to collect. Interestingly, the average T_1 for lung tissue shown in Figure 2.11 isin close agreement with reported values determined *in vivo* in studies involving human subjects. This finding suggests that measured T_1 values in lung tissue vary little across species and are insensitive to physiological status. In this context, studies performed in the mouse should be transferable to higher order species, and particle concentrations derived from changes in T_1 therefore will not likely be influenced by small differences in physiology.

Performance Expectations

Assuming the relaxivity for carbon black particles in lung tissue is the same as in bulk water and that the T_1 for lung tissue with no particles is about 1 second, 11 milligrams of particles per gram of tissue water would be required to alter the T_1 by 10%. Given that the average weight of a rat lung is about one gram, 11 milligrams of particles would be needed to observe only a 10% change in measured T_1. Importantly, this is significantly higher than the approximately 2.0 milligrams known to exceed most clearance mechanisms. Therefore, further development is required before maps of T_1 like that shown in Figure 2.11 can be used for determining particle concentrations at physiologically responsive doses.

Particle Refinement

To reduce the concentration of carbon black required for shortening T_1s in lung tissue, the relaxivity of inhaled particles needs to be increased by about a factor of 10 to roughly 100 sec^{-1}. So far, two approaches for increasing particle relaxivity have been identified. The first approach involves the use of an acid washing procedure to increase the free radical concentration of each particle. The second approach recognizes that the short range of the dipole-dipole interaction normally thought to drive the reduction in T_1 makes it largely a surface phenomenon. Since the surface area for a given particle mass increases inversely with particle size, smaller particles should exhibit a stronger relaxivity than larger particles. Importantly, relaxivity measurements performed so far were done using carbon particles filtered by a 20-micron mesh. Assuming much of the particle mass is dominated by particles with sizes close to 20 microns, significant improvements may be realized if, for example, particles are sorted and only those around 2 microns in size are used. In future work, the influences of both particle size and acid washing on particle relaxivity will be examined.

Molecular Basis of Beryllium Sensitivity

D. F. Lowry, A. S. Lipton, P. D. Ellis, and G. G. Burrows[(a)]

Supported by DOE Office of Biological and Environmental Research

(a) EMSL User, Oregon Health Sciences University

Chronic beryllium disease is a lung disease associated with occupational exposure to beryllium by individuals with a certain genetic makeup. The disease symptoms resemble tuberculosis and is caused by inhaling beryllium metal or insoluble salts such as beryllium sulfate. Individuals who are particularly sensitive to beryllium were found to have a single amino acid mutation in one of the immune response proteins (MHC class II HLA-DPB1*0201). The amino acid change was from glutamine 39 in non-sensitive individuals to glutamate 39 in sensitive individuals (Saltini et al. 1998). It was suggested that, because glutamate is negatively charged and can directly bind cations such as beryllium, we might be able to detect a difference in beryllium binding to these two forms of the protein, and this difference could be part of the reason why individuals with the glutamate mutation are sensitive. Especially since the amino acid at position 69 forms part of the antigen binding pocket of the HLA antibody (Figure 2.12).

The HLA antibody cannot be generated recombinantly by normal protocols of protein production. Professor G. G. Burrows of Oregon Health Science University has come up with an ingenious method of producing large quantities of the HLA antibody for structural studies. The beta-sheet platform/anti-parallel alpha-helix domain binds antibody. This domain is made up from two separate polypeptide chains, red and blue in Figure 2.12. Expressed separately, they do not fold into a functional globular domain. However, when the amino-terminus of one polypeptide chain to the carboxyl-terminus of the other chain. This results in a folded antigen binding domain of HLA. This

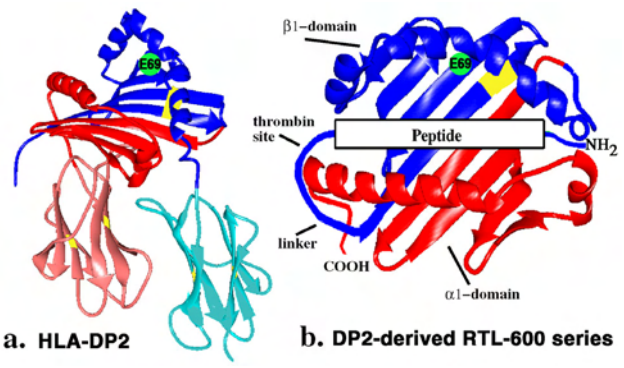


Figure 2.12. Design of the HLA peptides under study: a) the full human HLA-DP2 structure as homology modeled on the structure of human DR1, for which a structure is available. The red and blue ribbons show the portions of the two polypeptide chains that form the antigen binding domain. b) The recombinant version of the antibody binding domain where red and blue chains are coupled to form a single ribbon, resulting in a stable globular domain. The mutation site at position 69 is shown as a green circle.

construct, with either glutamine or glutamate at position 69, is the subject of our solid-state NMR investigations.

This study is part of a larger effort to directly detect metals in proteins. Briefly, ^{9}Be represents a 100% abundant isotope of beryllium. Its sensitivity relative to ^{13}C is ~80. That is, it is 80 times more sensitive that carbon. The ^{9}Be NMR experiment represents a novel and direct probe of the binding of Be^{2+} to these antibody fragments. Indeed, we have succeeded, for the first time in the history of NMR, in recording the solid-state NMR signal of beryllium bound to a protein (recombinant HLA domain). The spectrum is shown in Figure 2.13, and we are currently attempting to identify the ligands by comparison with solid-state beryllium spectra of model compounds. Also depicted in Figure 2.13 is the solid-state NMR spectrum of beryllium sulfate (bottom).

Reference

Saltini C., M. Amicosante, A. Franchi, G. Lombardi, and L. Richeldi. 1998. "Immunogenetic basis of environmental lung disease: Lessons from the berylliosis model." *J. Eur Respir.*, **12(6)**, 1463-75.

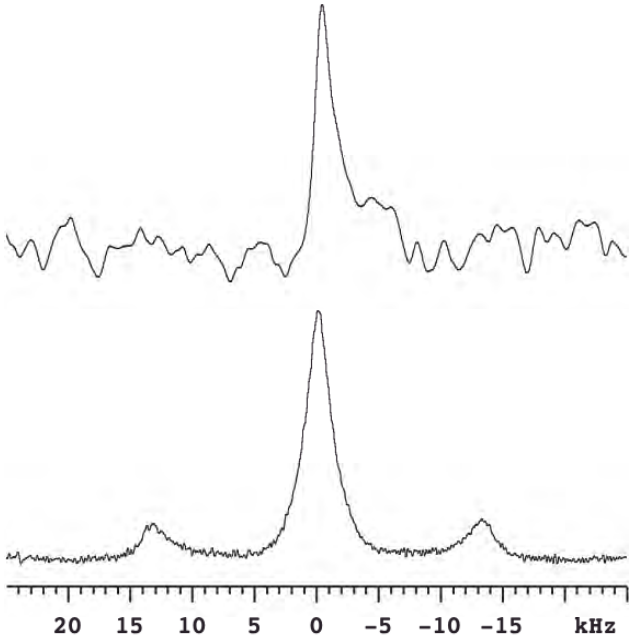


Figure 2.13. Solid-state NMR spectra of ^{9}Be bound to DP2-derived RTL-600 peptide and beryllium sulfate.

^{67}Zn NMR Investigation of the Consequences of Bound Water

A. S. Lipton, M. D. Smith,[(a)] R. D. Adams,[(a)] and P. D. Ellis

Supported by National Institutes of Health, Federal Grant GM26295F

(a) Department of Chemistry and Biochemistry, University of South Carolina

From a recent survey of zinc coordination in protein structures (Alberts et al. 1998), it was shown that the metals are predominantly tetrahedral, with a fair number of pentacoordinated sites in these systems. Those proteins that use zinc catalytically also tend to have water as at least one of the ligands. It is therefore interesting to note the effects water has on the electronic environment of the metal that make it so useful. As examples of metal centers that have both waters of hydration and an anhydrous site we have chosen zinc formate dihydrate. The crystal structure shows there are two sites: the first site has two bridging formyl groups and four bound waters, and

the second site has six bridging formyl groups. This structure forms a polymeric network that has sheets of zinc atoms connected by formates and the parallel sheets connected by columns of the hydrated sites. This network is depicted in Figure 2.14. The quadrupole coupling constants have been determined for each site of zinc formate and were found to be different, 9.6 and 6.1 MHz (Larsen et al. 1999). The challenge then is twofold: first to assign the EFG tensors to each site and then to orient the tensors in the frame of the metal. We have assigned the two sites using cross polarization and selective isotopic labeling, and then have performed a single crystal NMR experiment to determine the alignment of the EFG tensors in the crystal frame.

The two different sites of zinc formate dihydrate have different quadrupole coupling values as one might expect due to the different coordination environments; however, the isotropic chemical shifts are practically degenerate (Larsen 1999). In order to determine which lineshape corresponded to the hydrated species a combination of isotopic labeling and cross polarization (CP) was used. First, the magnetization buildup due to CP was measured on the isotopically normal (^{67}Zn enriched) material. Then, the sample was recrystallized from D_2O, which effectively removed the proton source of the bound waters from the CP experiment. The results of the magnetization buildup from each sample is shown if Figure 2.15. One can

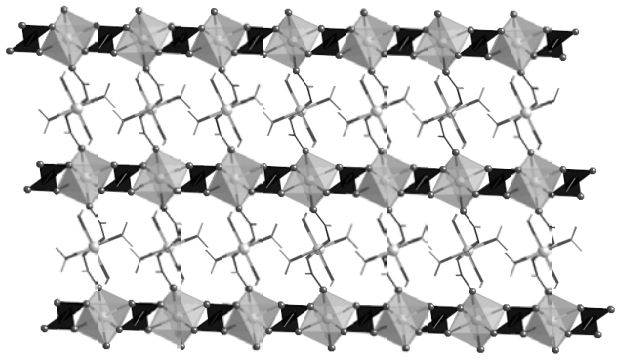


Figure 2.14. View of multiple unit cells of zinc formate dihydrate where the polyhedra represent the anhydrous sites.

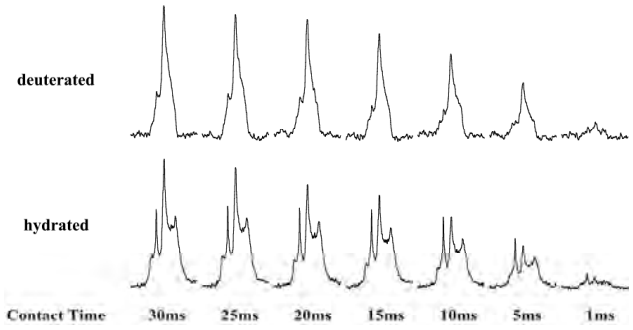


Figure 2.15. Static CP with varying contact time for hydrated and deuterated zinc formate dihydrate.

see that the sharp feature at the left-hand side of the lineshape (~21kHz), while present in both data sets, builds up rapidly to equilibrium in the hydrated sample. This feature belongs to the broader of the two lineshapes (Cq = 9.6 MHz and $\eta_q = 0.45$), which means that this zinc is the closest to the bound waters.

The other piece of evidence supporting this conclusion is the relative ratios of the two lineshapes in each of the samples. In the hydrated material, the broad lineshape is pronounced over the same one in the deuterated sample, which implies that it received more magnetization from the waters of hydration.

The next phase of the analysis was to perform a single crystal experiment to project the principal axes of the electric field gradient tensor onto the crystal frame. Figure 2.16 depicts the line positions plotted against rotation angle for each site along with the respective fit of the data. For the hydrated site, the tensor was determined to lie along the vector made by the formyl groups or orthogonal to the plane containing the waters. The site containing only bridging formyl groups also has the V_{zz} element aligned nearly along a Zn-O_{formyl} vector. This is in the *bc* plane (orthogonal to the a^* axis), which consists of a sheet of bridging formates, as shown in Figure 2.16. This result, in conjunction with the single crystal result from zinc acetate dihydrate (Vosegaard et al. 1999), demonstrates the level of sensitivity of the zinc electric field gradient to complexation with the neutral ligand water.

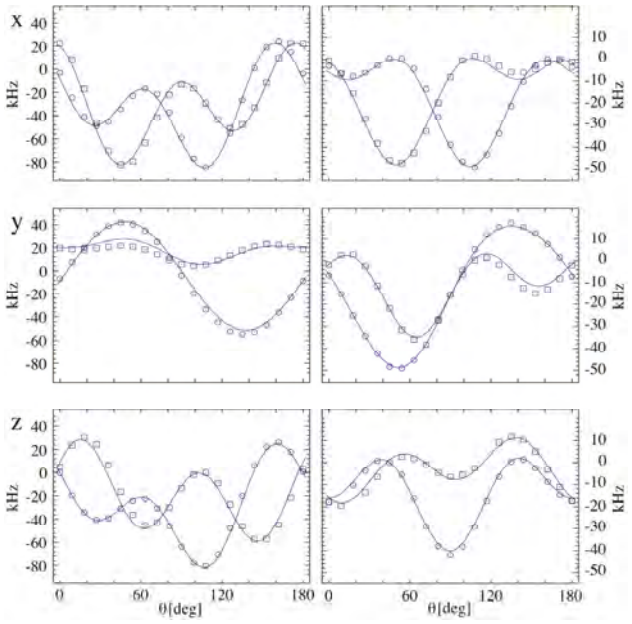


Figure 2.16. Single crystal rotation plots for zinc formate dihydrate and their respective fits.

References

Alberts, I. L., K. Nadassy, and S. J. Wodak. 1998. *Protein Science*, **7**, 1700-1716.

Larsen, F. H., A. S. Lipton, H. J. Jakobsen, N. C. Nielsen, and P. D. Ellis. 1999. *J. Am. Chem. Soc.*, **121**, 3783-3784.

Vosegaard, T., U. Andersen, and H. J. Jakobsen. 1999. *J. Am. Chem. Soc.*, **121**, 1970-1971.

^{67}Zn and ^{25}Mg NMR of Species with Large Quadrupole Couplings

A. S. Lipton, R. W. Heck, and P. D. Ellis

Supported by National Institutes of Health, GM26295 and DOE Office of Biological and Environmental Research KP11-01-01: 24931 and 41055

Recently a general strategy for the NMR observation of half integer quadrupolar nuclides has been presented (Larsen et al. 1999) and applied to the first direct observation of a ^{67}Zn resonance in a zinc dependent metalloprotein (Lipton et al. 2001). The strategy entails the utilization of low temperatures (4 to 25 K) to enhance the Boltzmann factor in combination with cross polarization (CP) (Pines et al. 1972) from ^{1}H to ^{67}Zn and Fourier transformation of a train of Carr-Purcell, Meiboom-Gill echoes (Carr and Purcell 1958, Meiboom and Gill 1958). Critical to this strategy is the management of the ^{1}H spin lattice relaxation times in the laboratory and rotating frames, T_1 and $T_{1\rho}$, respectively. The value of T_1 limits the overall rate in which the experiment can be repeated, and the value of $T_{1\rho}$ dictates the important aspects of the efficiency of the magnetization transfer from the ^{1}H bath to the rare spin ^{67}Zn nuclide in the rotating frame. In the previous examples, rapid rotation of the methyl groups present in the systems being investigated was exploited as a means to shorten the ^{1}H T_1. The methyl groups provide a relaxation sink for the remainder of the ^{1}H spins system *via* the methyl group's rapid internal rotational motion and efficient spin diffusion within the whole ^{1}H spin bath. However, the requirement of methyl groups in the chemical system of interest is limiting. We are exploring the use of a paramagnetic dopant as a means of managing ^{1}H spin lattice relaxation times. We have applied this methodology to the characterization of the ^{67}Zn NMR parameters of a series of 4- and 6-coordinate zinc poly (pyrazolyl) borate complexes. The details will be presented in a paper that is in preparation (Wright et al., in preparation). In this writeup, we describe a procedure for the acquisition of broad lineshapes that are wider than the normal excitation profile of the probe in a single experiment. The following illustrates this approach using zinc complex of a 4-coordinate pyrazolylborates.

As mentioned above, these experiments utilized a combination of low temperature (4 to 25 K), CP, and a train of CPMG spin echoes. The combined approach is termed CP/QCPMG pulse sequence. The experimental aspects associated with the determination of the lineshape fall into two discrete categories that depend on the overall width (baseline to baseline) of the lineshape. The first category is when the observed width of the lineshape is less than the excitation profile associated with the CP sequence. In typical experiments the excitation profile corresponds to ~50 kHz. In this case the spectrum is recorded normally and then analyzed. This is the situation for 6-coordinate polypyrazolylborate complexes. However, when the width of the lineshape exceeds this nominal value, the spectrometer frequency must be arrayed using a step size that is consistent with the excitation profile. The number of steps in the array depends on the width of the spectrum. This latter

procedure is necessary for 4-coordinate pyrazolylborates. Since such broad lineshapes are not a common situation in normal NMR experiments, we will outline in some detail our method for obtaining the resulting lineshape.

The first step is to gain an understanding of the excitation bandwidth of the probe. For the selective observation of the ±1/2 transition for a quadrupolar nuclide with nuclear spin of I under conditions of CP, the excitation bandwidth can be estimated from the following equation

$$\omega_{1S}/2\pi = \omega_{1H}/[2\pi\ (I + 1/2)] \sim 1/[(I + 1/2)\ 4\tau^{H}_{90}].$$

In this equation, $\pm\omega_{1S}/2\pi$ is the approximate excitation width in Hz, and τ^{H}_{90} is the ^{1}H $\pi/2$ pulse width in μs. It would not be unreasonable to anticipate that an unsymetrical 4-coordinate zinc or magnesium compound would have a lineshape that would span a range of frequencies ≥ 300 kHz. Under such circumstances the lineshape can also be attenuated by the excitation bandwidth of the NMR probe (i.e., the "Q" of the probe). Once the attenuation as a function of frequency is established, it can be utilized to correct the amplitudes of the observed spectrum. To get a sense as to how one would determine such a "transfer function," it important to summarize how we would obtain the uncorrected spectra. With some experience it is easy to estimate the center of the lineshape. At this point, the probe is tuned and matched at that frequency. The whole lineshape then is divided into bins, each bin corresponding to an excitation bandwidth. The spectrometer frequency is then arrayed and the sub-spectra collected. The transfer function is determined using a slightly different procedure.

The resulting amplitude and phased corrected subspectra, after appropriate re-referencing are then displayed in a "two-dimensional" format and subsequently projected along an axis of the display. Such a display for the ^{67}Zn spectrum of $[H_2B(pz)_2]_2Zn$ ("pz" denotes the pyrazolyl ring) is illustrated in Figure 2.17. The projections used are sky projections. The extraction of the quadrupole and shielding parameters, in these cases, is based on the resulting

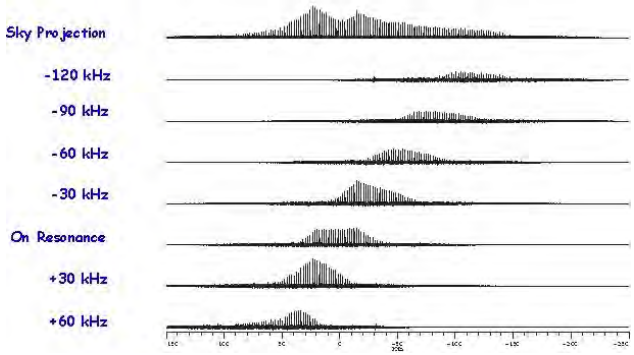


Figure 2.17. This is an illustration of the seven subspectra needed to define the ^{67}Zn lineshape for $[H_2B(pz)_2]_2Zn$ ("pz" denotes the pyrazolyl ring). The sky projection denotes the full lineshape. The approximate value of the quadrupole coupling constant, Cq, and asymmetry parameter, η_q, are 10.9 MHz and 0.8, respectively.

amplitude corrected lineshape depicted in these sky projections. This situation underscores the importance of high magnetic fields. The data reported here was obtained on a 9.4-T magnet. At 18.8-T, the number sub-spectra would be cut in half, and each sub-spectrum could be acquired an order of magnitude faster. We are developing the necessary probes to perform these experiments at 18.8-T.

Reference

Carr, H. Y. and E. M. Purcell. *Phys. Rev*., **94**, 630 (954).

Larsen, F. L, A. S. Lipton, H. J. Jakobsen, C. N. Nielsen, and P. D. Ellis. 1999. *J. Amer. Chem. Soc*., **121**, 3783-3784.

Lipton, A. S., G. W. Buchko, J. A. Sears, M. A. Kennedy, and P. D. Ellis. 2001. *J. Am. Chem. Soc*., **123**, 992.

Lipton, A. S., J. A. Sears, and P. D. Ellis, *J. Magn. Reson*., in press.

Meiboom, S. and D. Gill. 1958. *Rev. Sci. Instrum*., **29**, 688.

Pines, A., M. G. Gibby, and J. S. Waugh. 1972. *J. Chem. Phys*., **56**, 1776.

Wright, T. A., A. S. Lipton, M. K. Bowman, R. A. Wind, D. L. Reger, and P. D. Ellis, in preparation.

Spatial Distribution of Trapped 5-Thymyl Radicals in Low-Temperature Acid Glasses

M. K. Bowman, H. M Mottaz, A. F. Fuciarelli,[(a)] and J. D. Zimbrick[(b)]

(a) N. W. Toxicology/Bioanalysis
(b) School of Health Sciences, Purdue University

Ionizing radiation deposits its energy inhomogeneously in living cells. In DNA, this pattern of deposition results in multiply damaged sites (MDS) consisting of one or more free radicals localized in a microscopic volume on the order of several tens of nanometers in diameter. The complexity of damage in these sites and their spatial density is thought to vary with the energy and type of radiation being used. In this project we are attempting to obtain experimental measurements of the spatial characteristics of MDS by using pulsed electron paramagnetic resonance (EPR) methodologies to examine dipolar interactions between radicals trapped in the MDS at low temperatures. Such measurements can yield values of average spacing between radicals and average local concentrations of radicals as well as average MDS dimensions. Experimental values of these parameters may be important as inputs to various models of radiation damage to the genome from low doses of radiation. Other facets of the project being conducted at Purdue University involve the identification and quantitation of the stable products found in the DNA after it has been irradiated with radiations of various types and energies.

There is very little published work on experimentally determined spatial properties of MDS. Consequently, as a first approach in the project, we are using pulsed EPR techniques to study a simplified model system consisting of trapped DNA base radicals (5-thymyl radicals) produced by hydrogen-atom reactions with thymine in an irradiated acid glass at 77 K. In this system, we find a linearly increasing bulk yield of trapped radicals per unit mass of sample over an extended range of doses. By varying the dose we can vary the spacing between MDS so that they range from being spatially isolated to being so close that they overlap.

Figure 2.18 presents a schematic of one of the pulsing sequence used in these EPR experiments. Pulses at

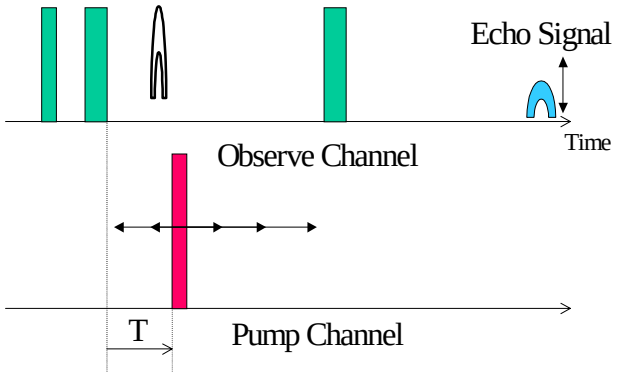


Figure 2.18. Double electron electron resonance (DEER) pulse sequence.

one microwave frequency are used to observe the radicals in one portion of the EPR spectrum while a second pump frequency flips radicals in another part of the spectrum, thereby changing the dipolar interaction between them. The DEER effect is seen as a change in the intensity of the observed radical echo signal as the time delay of the pumping pulse is changed. The stronger the dipolar interaction, the faster the observed echo signal changes with the time delay. The dipolar interaction varies as $1/r^3$ where r is the vector distance between two radicals. Consequently the interaction is much stronger between radicals that are close together, thus allowing a distinction to be made between radicals trapped in the same MDS versus those in separate MDS. A two pulse sequence known as instantaneous diffusion (ID) measures the same quantity more conveniently for hydrogen atoms. Both the ID and the DEER decay data can be used to derive values for r. Figure 2.19a below presents data reflecting the relationship between local radical density and yield of 5-thymyl radicals, while Figure 2.19b shows radical density versus yield for both trapped H-atoms and D-atoms. At doses below about 700 Gray, the radical density of both species remains relatively constant even though the total bulk radical concentration is increasing in this dose range (solid lines, both figures). These data support the hypothesis that both radical species are trapped in isolated MDS that are spatially separated at doses below 700 Gray. At doses above 700 Gray, the local radical density begins to increase with dose, and we hypothesize that we have reached the point at which MDS begin to overlap. The data would suggest an approximate MDS radius of 25 nm for both the trapped H-atoms and 5-thymyl radicals

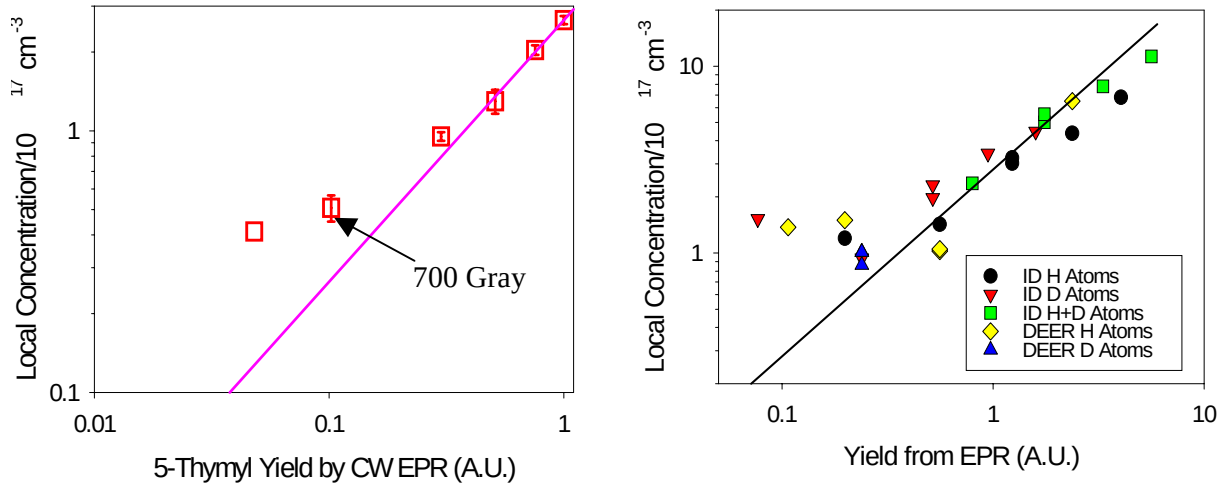


Figure 2.19. Dependence of local radical concentration versus average radical concentration for 5-Thymyl radicals (left) and hydrogen deuterium atoms (right). The transition from isolated spurs to homogenous radical distribution occurs around 700 Gray as marked on the plot on the left.

High Resolution EPR Spectroscopy

M. K Bowman, D. M. Kramer,[(a)]
U. Diebolt,[(b)] *and L. Kispert*[(c)]

Supported by DOE-OBER Funding

(a) Washington State University
(b) Tulane University
(c) University of Alabama

Description

In condensed phases, the electronic wavefunction provides a wealth of information about the structure, chemical reactivity, and physical properties. EPR spectroscopy offers a unique view of the electronic wavefunction of free radicals, which are molecules with an unpaired electron spin and include many important reaction intermediates, catalytic sites, and metal ions. The higher the resolution in the EPR measurement, the greater the detail determined about the species observed. As high resolution EPR is applied to problems in biology, chemistry, and materials, new methods to increase resolution and to improve analysis of the resulting data are needed. Thus, scientific applications are driving the development and application of new EPR techniques.

Technical Accomplishments

New Methods

Development has been completed on two new pulsed EPR methods. One a multiple resonance method called electron nuclear quadruple resonance (ENQOR) that provides higher resolution information about the extent of the unpaired electron wavefunction and its overlap with nearby nuclei. The second is a graphical method to analyze non-axial, hyperfine spectra obtained using hyperfine sublevel correlation spectroscopy (HYSCORE). In a large class of cases, the elements of the hyperfine tensor describing the interaction of a nucleus with the unpaired electron wavefunction can be directly read from the transformed spectra. Both methods are being applied to scientific applications.

Two new DEER techniques are under development for the measurement of distances within proteins and protein complexes and for the characterization of free radical distribution functions in solid materials. The first is a variant of the refocused echo DEER method, but it uses a composite pumping pulse at nearly the same frequency as the observing frequency. This approach is advantageous because it allows measurements on radicals with either narrow spectra or very large splittings.

The second method is a two-dimensional version of the single-frequency technique for refocusing dipolar couplings (SIFTER). The conventional SIFTER measurement produces responses from weak hyperfine couplings as well as from the desired dipolar couplings. The two-dimensional SIFTER resolves these responses into different quadrants of the resulting two-dimensional spectrum and allows measurement of dipolar couplings that are nearly degenerate with frequencies of hyperfine-coupled nuclei.

Cytochrome b_6f Complex

The cytochrome b_6f complex is a large, intrinsic membrane protein complex that, with the closely related cytochrome bc_1 complex, is the major energy transduction apparatus in most organisms. It couples the flow of electrons to the pumping of protons across cellular or mitochondrial membranes. Recent x-ray crystal structures suggest a large-scale conformational change in the complex as part of its catalytic cycle, changing the location of the Rieske protein headgroup by 11Å or more (Figure 2.20).

We found a metal ion-binding site on cytochrome *f* and used it as a paramagnetic probe of some of these events. It is well removed from the Rieske protein, yet it shows changes in the structure of the protein complex associated with changes not only of the

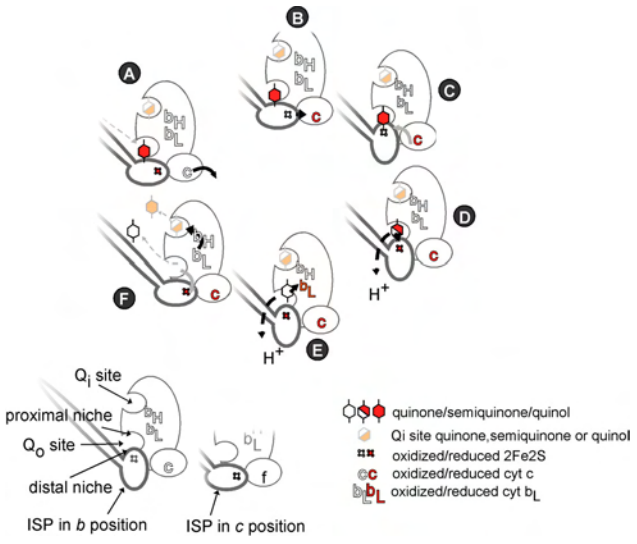


Figure 2.20. Working model for the catalytic cycle of the cytochrome bc_1 complex, illustrating the motion of the Rieske protein headgroup and some of the molecular events hypothesized to be responsible. In the cytochrome b_6f complex, cytochrome *c* is replaced by cytochrome *f*.

Rieske protein conformational, but also in pH and molecules occupying the quinone substrate-binding site. Our study of the metal site in the cytochrome b_6f complex prompted the Berry group to locate two analogous metal sites in the x-ray diffraction structure of the related cytochrome bc_1 complex.

The physical basis for the different EPR spectral forms of the Rieske protein is being examined in order to learn how its conformation is changed and controlled *in vivo*. Difference Mims ENDOR spectra of the Rieske isolated protein exchanged into deuterated solvent reveal two exchangeable protons located near the active site. Two different conformations of the 2Fe2S active site are reported in all of the x-ray diffraction structures of the Rieske protein. The position of the protons is consistent with only one of these conformations, suggesting that we now have a spectroscopic signature for that conformation in complexes that cannot be crystallized and *in vivo*.

Distance Measurements

The accurate measurement of distances within a non-crystalline solid on a nanometer scale is a challenge in any areas of research (e.g., structural biology) and in understanding the energy deposition and effects of ionizing radiation. Our approach has been to use the magnetic dipolar interaction between two free radicals to probe their distance of separation. The dipolar interaction between two spins can be calculated from first principles and varies from 52 to 0.1 MHz at distances from 1 to 8 nm. The challenge is how to measure this interaction using free radicals whose spectrum is often in excess of 200 MHz. Our approach is to use the superior resolving power of multiple resonance methods in EPR, such as DEER and SIFTER. This approach is being tested on two model biradicals that contain nitroxide free radical groups at well defined distances as shown in Figure 2.21.

The refocused echo DEER spectrum of biradical II is shown in Figure 2.22. The spectrum is a characteristic Pake doublet and the vertical lines mark the dipolar splitting in this spectrum, which is 0.94 MHz. These appear to be the first measurements in which the outer pair of lines are actually observed with good intensity. Their location shows that the exchange interaction between the nitroxide groups is negligible and does not interfere with determination of the distance between the nitroxides. From the measured dipolar interaction of 0.94 MHz, a distance between nitroxides of 3.81 nm is calculated from first principles and

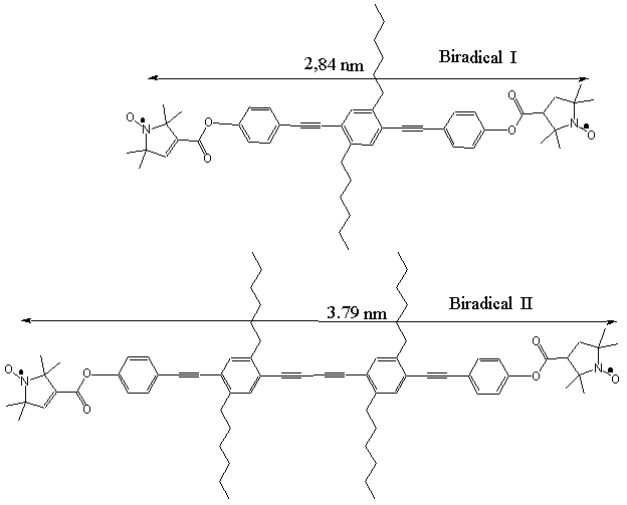


Figure 2.21. Two shape persistent nitroxide biradicals with well defined distances between the radicals of 2.8 and 3.8 nm.

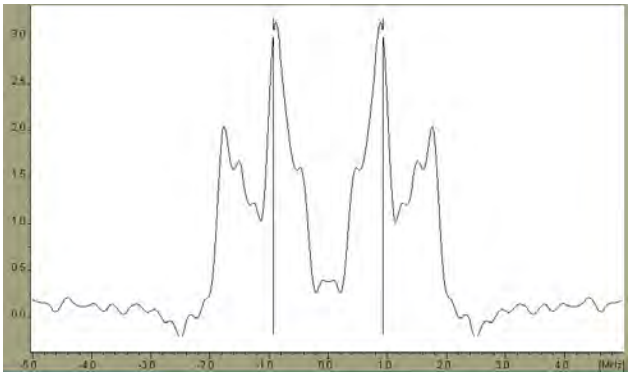


Figure 2.22. The refocused echo DEER spectrum of biradical II in frozen solution of toluene at 30 K. The vertical lines mark the dipolar splitting in the spectrum, which is 0.94 MHz, corresponding to 3.81 nm.

corresponds well with a distance of 3.84 nm reported from molecular mechanics. This methodology is directly applicable to the measurement of distances within proteins or protein complexes using the powerful techniques of site-directed spin labeling.

TiO_2

TiO_2 is an important material in its crystalline rutile and anatase phases. Its conductive, dielectric, and optical properties can be altered by slightly altering its stoichiometry and by introducing small amounts of defects or impurities. Many of these defects consist of Ti(II) and are paramagnetic. Crystals of rutile were made oxygen deficient by heating in an oxygen-depleted atmosphere. More than 20 different paramagnetic defects were observed, and their characteristics changed as the heat treatment temperature increased. Based on the symmetry elements of the EPR spectra relative to the crystal symmetry, most of the defects could be classified as either point defects (vacancies or substitutions) or as extended defects (such as shear planes). We found a strong correlation between the type of defect and the formation of complex structures on the surface during reoxidation of the crystals. Most of the defects in samples heated to lower temperatures could be identified as substitutional or interstitial Ti(III) point defects. In crystals heated to higher temperatures, extended shear plane defects were found.

Carotenoid Radical Cations

Carotenes and carotenoids are important biomolecules found in every organism that lives in the presence of oxygen and light. These compounds perform the vital role of protecting the organism against singlet oxygen, excited triplet states, and a variety of free radicals. They also have a role in the collection of light for photosynthesis and in some biomimetic photoconversion systems.

The cation radicals of carotenoids are readily formed and can be stabilized for study on metal oxide surfaces. The EPR spectra consist of a single line, but hyperfine interactions from many nuclei can be observed but are poorly resolved in ENDOR and HYSCORE spectra. The newly developed analysis methods for HYSCORE spectra enabled the full hyperfine tensors for some of the methyl groups and α-protons to be extracted from the HYSCORE spectra (Table 2.1). These measurements reveal far greater detail about the carotenoid cation radicals than did previous liquid phase ENDOR measurements. In particular, several conformations of the carotenoid radicals are evident from variations in the hyperfine tensors. These forms may be caused either by specific interactions between the carotenoid radical and the solid support they were generated on, or by isomerization around any one of the double bonds in the carotenoid. This later possibility can be tested by quantum chemical calculations of hyperfine tensors for hypothesized isomer cations.

Table 2.1. Principal values (MHz) of the hyperfine tensors of carotenoid cation radicals obtained by HYSCORE analysis.

Carotenoids / Protons	Simulated HYSCORE A_{XX}	A_{YY}	A_{ZZ}	A_{iso}	Simulated [a] CW ENDOR	Calculated [c] A_{iso}^{RHF}	A_{iso}^{INDO}
Canthaxanthin							
1) C(10,10')H_α	-5.38	-5.38	14.65	1.30	1.1	1.1	3.28
2)C(13,13′)CH_3	7.43	7.43	14.74	9.87	13.0; 13.0; 13.0	13.1;13.4	21.43
3) C(15,15′)H_α	2.30	2.30	7.16	3.92	3.9; 5.3; 4.7	-5.2; -5.3	10.46
4)C(13,13′)CH_3	6.47	6.47	16.47	9.80	13.0; 13.0; 13.0	13.1;13.4	21.30
5) C(11,11′)H_α	4.86	4.86	11.29	7.00	7.4	-7.0; -7.2	11.47; 12.94
Apo-carotenal		6.1	13.67				
1) C(15′)H_α	6.1	0.07	6.05	8.62	-	- 9.06	20.70
2) C(12)H_α	0.07	6.76	18.67	2.10	-	-2.25	7.87
3)		6.01	21.65				
4) C(9)CH_3	6.76	8.25	18.46	10.73	11.0;11.0;11.2		
5)	6.01			11.22		12.5	22.60
	8.25			11.65			

Publications

Bowman M. K. and A. M. Tyryshkin. "Electron Nuclear Quadruple Resonance for Assignment of Overlapping Spectra." *Journal of Magnetic Resonance*, **144**, 74-84 (2000).

Dikanov S., A. M. Tyryshkin, and M. K. Bowman. "Intensity of Cross-Peaks in HYSCORE Spectra of S=1/2, I=1/2 Spin Systems." *Journal of Magnetic Resonance*, **144**, 228-242 (2000).

Elder S. H., F. M. Cot, Y. Su, S. M. Heald, A. M. Tyryshkin, M. K. Bowman, Y. Gao, A. G. Joly, M. L. Balmer, and A. C. Kolwaite. "The Discovery and Study of Nanocrystalline TiO_2-(MoO_3) Core-Shell Materials." *Journal of the American Chemical Society*, **122,** 5138-46 (2000).

Li M., W. Hebenstreit, U. Diebold, A. M. Tyryshkin, M. K. Bowman, G. C. Dunham, and M. A. Henderson. "The Influence of the Bulk Reduction State on the Surface Structure and Morphology if Rutile TiO_2 (110) Single Crystals." *Journal of Physical Chemistry*, B **104**, 4944-50 (2000).

Rao B.K.S., A. M. Tyryshkin, A. G. Roberts, M. K. Bowman, and D. M. Kramer. "Inhibitory Copper Binding Site on the Spinach Cytochrome b_6f Complex: Implications for Q_o Site Catalysis." *Biochemistry*, **39**, 3285-3296 (2000).

Samoilova R. I., A. A. Shubin, M. K. Bowman, J. Hütterman, and S. A. Dikanov. "Observation of Two Paramagnetic Species in Electron Transfer Reactions within Cesium Modified X and Y Zeolites." *Chemical Physics Letters*, **316**, 404-410 (2000).

Geo-Sourcing of Controlled Substances by Rare-Isotope NMR Spectroscopy

H. Cho, D. W. Koppenaal, and P. Hays[(a)]

Supported by the U.S. Department of Energy, Office of Nonproliferation and National Security, Special Technology Programs

(a) Special Testing and Research Laboratory, U.S. Drug Enforcement Administration

NMR spectroscopy of rare isotopes is an effective way to infer points-of-origin of plant-derived 1992 a,b; Guillou et al. 1992; Hays et al. 2000). The method is based on the finding that $^2H/1H$, $^{13}C/^{12}C$, $^{18}O/^{16}O$, and $^{15}N/^{14}N$ ratios in organic molecules can deviate measurably from their natural terrestrial ratios depending on local environmental conditions and processing history. These differences are explained

by isotope effects on the kinetics of the relevant biochemical reactions; provided reaction pathways or substances such as drugs and food (Martin et al. conditions vary from place to place, two otherwise identical chemicals may be differentiated and assigned to different locales according to site-resolved isotope ratios.

This approach is being tested as a way to determine the origins of controlled substances in a collaborative effort with the U.S. Drug Enforcement Administration (DEA). A primary goal of this work is to devise NMR procedures that reduce the amount of material and the experiment time needed for an accurate isotopic profile analysis.

A rapid and discriminating profile of the hydrogen isotope distribution in an organic molecule can be extracted from solution-state ^{2}H NMR spectra (Figure 2.23). The relative populations of deuterium atoms at different sites of the molecule are individually quantified from the integrated intensities of resolved ^{2}H NMR lines and then compared with the ratios expected from the stoichiometry of the molecule. A catalog of the deviations from the stoichiometric ratios provides a detailed profile by

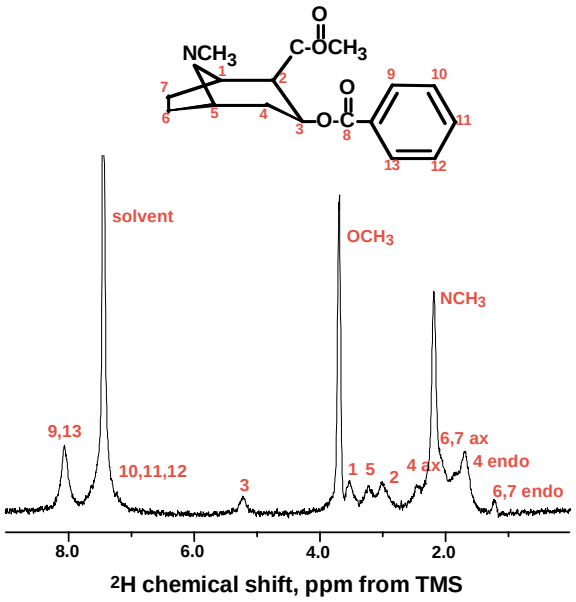


Figure 2.23. ^{2}H NMR spectrum (ν = 76.74 MHz, T = 313 K) and assignments of resonances of free base cocaine (Sigma) dissolved in chloroform. The linewidths are due to the short spin-lattice (T_1) relaxation times of deuterium, which also make it impossible to resolve ^{1}H-^{2}H *J* couplings.

which chemical points-of-origin can be identified. With a combined MS and NMR approach, both precise (bulk) isotope ratio and molecular isotope distribution fingerprints are obtained. A compilation of location/isotope profiles is currently being developed with the use of DEA samples of known provenance.

References

Guillou, C., G. Remaud, and G. Martin. 1992. *Trends Food Sci. Technology*, **3**, 197.

Hays, P. A., G. S. Remaud, E. Jamin, and Y. -L. Martin. 2000. *J. Forensic Sci*., **45**, 552.

Martin, G. J., M. L. Martin, and B. L. Zhang. 1992. *Plant Cell Environ.*, **15**, 1037.

Martin, G. J., X. Y. Sun, C. Guillou, and M. L. Martin. 1992. *Tetrahedron*, **41**, 3285.

High Throughput Structural Geonomics Using NMR

M. A. Kennedy, J. R. Cort,[(a)] and T. Terwilliger[(b)]

(a) Associated Western Universities
(b) Los Alamos National Laboratory

Project Description

A Structural Genomics pilot project focused on *Pyrobaculum aerophilum* is led by Dr. Tom Terwilliger at Los Alamos National Laboratory. Our group has entered into collaboration with Dr. Terwilliger to explore the utility of NMR as part of an overall high-throughout approach for Structural Genomics. Here we present a solution-state NMR structure of a disimilatory sulfite reductase from the *P. aerophilum* organism.

Technical Accomplishments

The solution structure of an archaeal homolog of the gamma subunit of dissimilatory sulfite reductase has been determined by NMR spectroscopy. This 11.4-kDa protein from the hyperthermophilic archaeon *P aerophilum* adopts a novel fold consisting of an up-down-up-down helical bundle with a small beta-hairpin attached to one side. The protein contains two disulfide bonds but remains folded upon reduction of these bonds with only a small decrease in stability as measured by thermal denaturation. The

structure appears to contain a helix-turn-helix DNA-binding motif, suggesting a role in the regulation of dissimilatory sulfite reductase or its siroheme or [4Fe-4S] cluster cofactors.

The dsrC gene was cloned from *P. aerophilum* genomic DNA and inserted into the pET28b vector, which provides a N-terminal 6X-His tag for affinity purification. Cells were grown in M9 minimal media supplemented with vitamins and minerals and 20 μg/ml kanamycin at 37°C. Samples of DsrC labeled with ~99% 15N, ~99% 13C/15N, and 10% 13C were prepared using U-13C-glucose and 15N NH4Cl as carbon and nitrogen sources in minimal media. Expression was induced with 1-mM IPTG when cultures reached OD_{600}=0.6. After induction, cultures were allowed to grow for 4 to 5 hours. Cells were harvested by centrifugation for 25 minutes at 6700 g and resuspended in 25 mL of lysis buffer (50-mM sodium phosphate pH 8.0, 300-mM NaCl, and 10-mM imidazole). PMSF (12.5 mg in 1-mL ethanol) was added to the suspension, and it was passed three times through a French press at 10,000 psi. Insoluble material was removed from the pressate by centrifugation for 12 minutes at 17,500 g. Protamine sulfate (12.5 mg in 1-mL lysis buffer) was added to the supernatant to precipitate nucleic acids, and the solution was clarified by centrifugation again. The supernatant was further clarified by ultracentrifugation for 1 hour at 60,000 rpm in a Ti90 rotor (Beckman). The ultracentrifugate was loaded on a 10-cm by 1-cm Ni-NTA column for purification. The column was washed at a rate of 1 mL per minute with 60 mL of lysis buffer and 60 mL of wash buffer (lysis buffer with 20-mM imidazole), and then bound protein was released with elution buffer (300-mM imidazole). SDS-PAGE gels were used to confirm that the appropriate fractions contained protein, and these fractions were concentrated to 1 mL in centricon-3 ultrafiltration devices (Millipore). Concentrated protein was exchanged into NMR buffer (25-mM phosphate pH 7.4, 100-mM NaCl, 10% D_2O) with a PD-10 column (Pharmacia) and concentrated to the appropriate volume. Reduced protein was generated by exchange into buffer that contained 10 to 25 mM DTT. For amide proton exchange measurements, NMR samples were lyophilized and redissolved in D_2O.

DsrC expressed in *E. coli* at a yield of approximately 30 mg/L following purification by Ni affinity chromotography. The protein was soluble to at least 22 mg/mL, with the 6X-His tag still attached.

Backbone and sidechain ^{1}H, ^{13}C, and ^{15}N chemical shifts were completely assigned. Chemical shift values for the beta carbons of the four assigned Cys residues (a fifth Cys residue is the second to last in the sequence and was not located in the NMR spectra) suggested that they were not in a reduced state (Sharma & Rajarathnam). Because the original sample did not contain reducing agent, another sample was prepared with 10-mM DTT present. The ^{15}N-^{1}H HSQC spectrum of this additional sample showed that several residues displayed small changes in their amide nitrogen and proton chemical shifts and that these changes were concentrated in the regions of the sequence containing the Cys residues. A ^{15}N-edited NOESY of the reduced sample showed little change in the identities and intensities of the crosspeaks when compared to the same experiment conducted on the non-reduced sample. Even residues whose amide proton and nitrogen had moved considerably could be easily assigned by analogy. Small shifts were observed for sidechain protons also.

Circular dichroism experiments showed the non-reduced and reduced protein samples had similar thermal unfolding characteristics. Because this protein comes from a thermophilic organism, complete unfolding could not be observed, however.

DsrC was also found to exhibit a stable, low-pH form. The HSQC spectrum of a sample at pH 3.0 displayed a wide dispersion of proton and nitrogen chemical shifts, similar to that observed at pH 7.4, and some of the peaks appeared to have not moved as a result of the pH change. The structure of this form was not characterized.

Using the assigned NMR spectra of non-reduced DsrC, 1000 unambiguous distance restraints were derived from the Noesy spectra and used in structure calculations (Figure 2.24). 200 dihedral restraints and 100 hydrogen bond restraints were also included. The final ensemble of 20 structures had a backbone rmsd of 0.65 Å for all elements of secondary structure. The structure contains a N-terminal beta hairpin followed by five helical segments arranged in an orthogonal bundle. The hairpin lies flat against the first helix (H1). The protein is somewhat elongated, with dimensions of approximately 20 x 30 x 40 Å. Eight

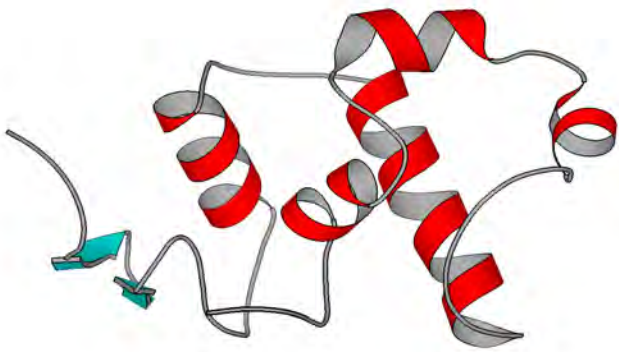


Figure 2.24. Molscript rendition of the solution-state NMR structure of the DSrC protein.

residues at the C-terminus appear to protrude from the globular protein and probably are not structured, as they were not found in the NMR spectra. A notable feature of the structure is a helix-turn-helix (HTH) structural motif centered on Gly 80. No similar folds were identified with Dali, although several HTH containing proteins were identified with this program. In these cases, only the HTH motif and the preceding helix (H2) aligned with the returned structures.

Publications

Cort, J. R., E. J. Koonin, P. A. Bash, and M. A. Kennedy. "A Phylogenetic Approach to Target Selection for Structural Genomics." *Nucleic Acids Research,* **27**(20), 4018-4027 (1999).

NMR Based Structural Genomics

M. A. Kennedy, J. R. Cort,[(a)] P. A. Bash,[(b)] E. J. Koonin,[(c)] A. D. Edwards,[(d)] and C. H. Arrowsmith[(d)]

(a) Associated Western Universities
(b) Northwestern University
(c) National Center for Biotechnology Information
(d) University of Toronto

Project Description

Structural genomics presents an enormous challenge with up to 100,000 protein targets in the human genome alone. Two facets of structural genomics were addressed in the project. First, estimates of the required number of structures needed to either experimentally represent all desired structures or to provide a basis for reliable homology modeling range on the order of 10,000 structures. Target selection remains a central focus of Structural Genomics projects to ensure that the novel sequence fold information is gained with each new experimentally determined protein structure. In this project, a phylogenetic approach to target selection was examined using the National Center for Biotechnology Information database of Clusters of Orthologous Groups (COGS). The strategy is designed so that each new protein structure is likely to provide novel sequence-fold information. Second, the role that NMR will play in structural genomics has remained an open question. This project was in part aimed at demonstrating the utility and potential role that NMR can play in structural genomics.

Technical Accomplishments

Structure Determination and Fold Classification of E. coli YciH

The solution-state structure of YciH was determined using NMR methods (Figure 2.25). The structural classification databases CATH, Dali/FSSP, and SCOP were utilized for comparative analysis of the YciH

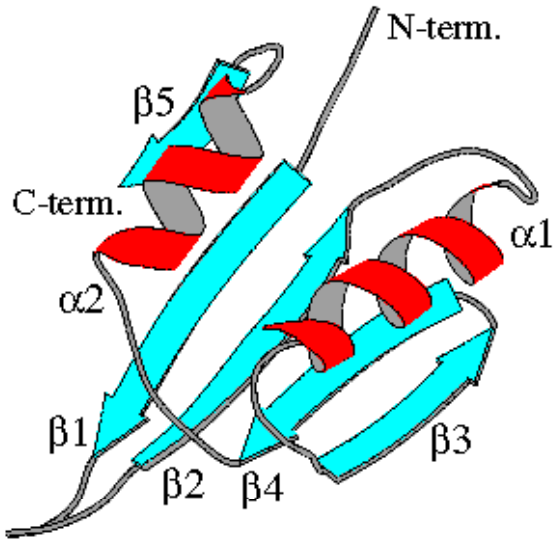


Figure 2.25. Molscript rendition of the secondary structure of *E. coli* YciH. The structure contains the ββαββαβ topology that folds into a two-layer α+β open sandwhich. The YciH sequence-fold defines a new homologous superfamiliy within this common fold family.

structure. The fold of YciH clearly resembles many others with the open-faced (or 2-layer) β-sandwich form, a fold in the α+β class. A Dali search returned 10 unique domains (including eIF-1) with Z scores greater than 3.0, indicating significant structural similarity. Thirty-two additional structures had scores between 2.0 and 3.0; in a Dali search, scores above 2.0 are considered evidence of structural similarity. Most have the same or a quite similar topology, or ordering of secondary structure elements within the sequence, as YciH. In the SCOP classification of protein structures, the structures found in the Dali search are members of one of two fold categories (similar to topology in CATH): the ferredoxin-like fold (23 superfamilies) and the DCoH-like fold (2 superfamilies). In the CATH classification, all the classification (32 homologous superfamilies), which encompasses both of the SCOP folds. The two proteins in SCOP with DCoH-like topology, the arginine binding domain of the arginine repressor, and dimerization cofactor of hepatocyte nuclear factor-1(DCoH), have the identical core topology as YciH: ββαββα. A domain of DNA gyrase that was located by Dali also has this topology but does not appear in CATH similar structures fall into the α-β plait topology or SCOP. Structures with ferredoxin-like topology are slightly permuted: βαββαβ. Both contain the same placement of secondary structure elements—two helices running antiparallel to each other packed on the same side of a four-stranded antiparallel β-sheet. Furthermore, they have the same organization of the strands in the core of the sheet: β2, β3, and β4 in the DCoH topology correspond to β1, β2, and β3 in the ferredoxin topology. In other words, removal of the N-terminal and C-terminal β-strands, respectively, from the DCoH and ferredoxin topologies yields exactly the same structure. YciH has a fifth beta strand as well, which is parallel to β1 and thus its topology is ββαββα(β). In YciH, if β1 were removed and β5 moved sideways slightly to adjoin (antiparallel to) β2, the resulting topology would be βαββαβ, the ferredoxin-like fold.

Solution-State Structure and Structure-Based Functional Classification of a Hypothetical Protein, MTH538, from Methanobacterium thermoautotrophicum.

The structure of a hypothetical protein, MTH538, from *M. thermoautotrophicum*, has been determined by NMR spectroscopy (Figure 2.26). MTH538 is one of numerous structural genomics targets selected in a genome-wide survey of uncharacterized sequences

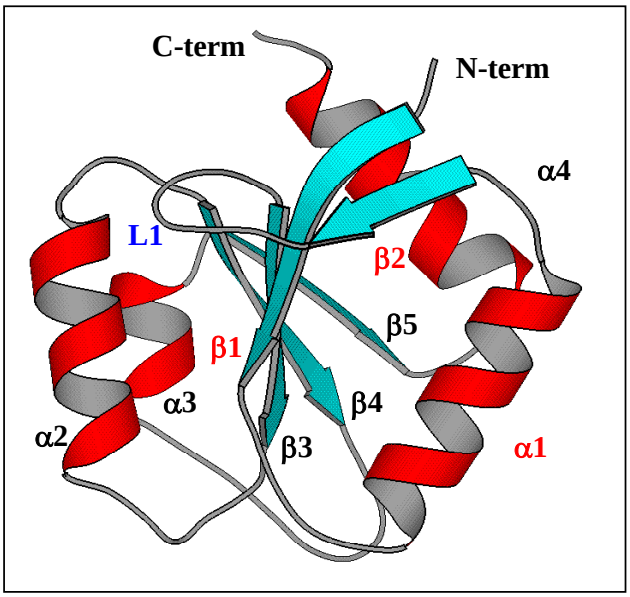


Figure 2.26. Molscript rendition of the secondary structure of *M. therm* MTH538. The structure contains the ββαββαβ topology that folds into a two-layer α+β open sandwhich. MTH538 is a singleton in sequence space. Structure-based functional classification indicates a possible role as a phosphorylation-independent response-regulator signaling protein in a two-component signal transduction system.

from this organism. MTH538 is not unequivocally similar to any other (known) sequences. The structure of MTH538 closely resembles those of the receiver domains from two component response regulator systems, such as CheY, and also is similar to structures of flavodoxins and GTP-binding proteins. Tests for characteristic activities of CheY and flavodoxin (phosphorylation in the presence of acetyl phosphate and flavin mononucleotide binding, respectively) were negative. However, a role in a two component response regulator system may still be plausible.

A PSI-BLAST search with the MTH538 sequence reveals weak (18 to 23%) similarity between a 95 residue segment of MTH538 and a common segment from members of a family of proteins annotated as putative ATPases or kinases (Figure 2.26). This family (COG1618 in the clusters of orthologous groups) contains another *M. thermoautotrophicum* ΔH protein, MTH1068 along with representatives from other archaebacterial species, including *Methanococcus jaanaschi* and the primitive eubacterium *Thermotoga maritimium*. Two nucleotide-triphosphate binding motifs, GXXXXGK(S/T) and DXXG, are conserved

in COG1618 but are not present in MTH538. No structural information is available for any proteins in this family.

Solution-State Structure and Fold-Classification of a Hypothetical Protein, MTH1175, from Methanobacterium thermoautotrophicum.
The solution structure of MTH1175, a 124-residue protein from the archaeon *M. thermoautotrophicum* has been determined by NMR spectroscopy (Figure 2.27). MTH1175 is part of a family of from all complete archaeal genomes and the archaeal gene-rich bacterium *Thermotoga maritima*. Sequence similarity indicates this protein family may be related to the nitrogen fixation proteins NifB and NifX. MTH1175adopts an α/β topology with a single mixed conserved hypothetical proteins (COG1433) with unknown functions which contains multiple paralogs β-sheet, and contains two flexible loops and an unstructured C-terminal tail. The fold resembles that of Ribonuclease H and similar proteins, but differs from these in several respects, and is not likely to have a nuclease activity.

A PSI-BLAST search with the amino acid sequence of MT1175 indicated similarity to the COG1433 proteins and some additional Archaeal orthologs, as well as the bacterial nitrogen fixation proteins NifX and a

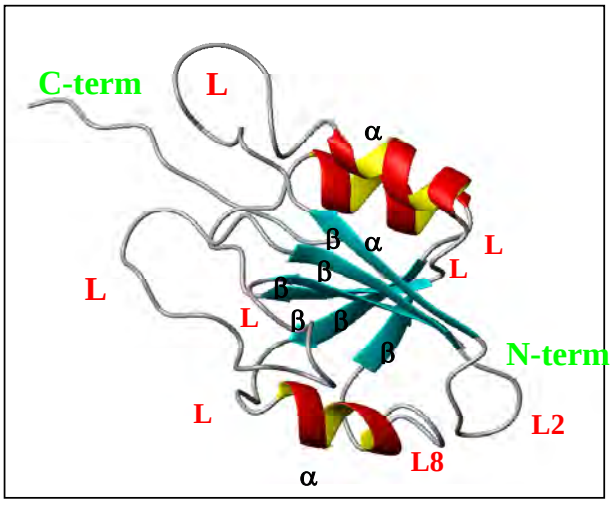


Figure 2.27. Molscript rendition of the secondary structure of *M. therm* MTH1175. The structure contains the ββαββαβ topology that folds into a two-layer α+β open sandwhich. MTH1175 is a member of a conserved family of mainly archaeal proteins whose function remains unknown.

C-terminal segment of NifB. The sequence similarity between MTH1175 and NifX or NifB is approximately 20 to 25% over a span of ~80 residues. Similarity to two Archaeal NifX proteins, from *M. thermoautotrophicum* and *Methanococcus maripaludis* also was revealed by the PSI-BLAST search. However, Archaeal NifB proteins from *M. thermoautotrophicum* and *Methanococcus jannaschii* lack the C-terminal segment that is similar to COG1433 proteins and NifX. In a cluster dendrogram (not shown) of the aligned sequences, the expanded COG1433, NifX, and NifB families appear to constitute three distinct, but related, families. The limited sequence similarity between proteins from different families probably is not sufficient to unequivocally call them homologues, without knowing whether they are structurally similar. The cluster dendrogram also indicates that Archaeal NifX proteins from *M. thermoautotrophicum* and *M. maripaludis* do not clearly fall into any one of these families, in terms of sequence similarity, although the branch point in the dendrogram lies closest to the cluster of NifB sequences.

Summary and Conclusions

The aim of this project was to 1) demonstrate the effectiveness of a phylogenetic approach to target selection for structural genomics and 2) to demonstrate the utility and potential role for NMR in structural genomics.

The YciH sequence was targeted using the phylogenetic target selection approach with the goal in mind of identifying a new sequence-fold relationship. In this example, we found that the YciH structure corresponded to a well known and highly represented fold class in protein fold space, albeit this relationship between sequence and fold could not be identified in the absence of the structure using the sequence information alone. This method was successful at identifying a new homologous superfamily in a known fold class (i.e., a new sequence-fold relationship was discovered). In this sense, this one example indicates that a phylogenetic approach to target selection for structural genomics would be effective.

The two remaining protein structures determined in this project were part of a pilot demonstration to illustrate the effectiveness of using NMR methods for protein structure determination for structural

genomics. Each protein structure determined in the project required about 4 to 6 weeks of NMR data collection. With new technological advances occurring in NMR science, it is possible that that data collection times can be reduced by another order of magnitude and the molecular weight of proteins for which structures can be determined using NMR methods is likely to be increased significantly over the current limit of ~40 kD. Furthermore, in this project, we demonstrated that NMR methods likely provide a complementary approach to x-ray crystallographic methods by allowing rapid functional testing for substrate or ligand binding and a means of characterizing dynamic and unstructured regions in proteins that have a potentially important function and that are difficult or impossible to characterize by other methods.

Publications

Cort, J. R., E. J. Koonin, P. A. Bash, and M. A. Kennedy. "A Phylogenetic Approach to Target Selection for Structural Genomics." *Nucleic Acids Research*, **27**(20), 4018-4027 (1999).

Cort, J. R., A. Yee, A. M. Edwards, C. H. Arrowsmith, and M. A. Kennedy. "Structure-based Functional Classification of Hypothetical Protein MTH538 from Mehtanobacterium thermoautotrophicum." *J. Mol. Biol.*, **302**, 189-203 (2000).

Cort, J. R., A. Yee, A. M.Edwards, C. H. Arrowsmith, and M.A. Kennedy. "NMR Structure Determination and Structure-Based Functional Characterization of a Conserved, Hypothetical Protein MTH1175 from *Methanobacterium thermoautotrophicum*." *J. Structural and Functional Genomics*, **1**, 15-25 (2000).

Christendat, D., A. Yee, A. Dharamsi, Y. Kluger, A. Savchenko, J. R. Cort, V. Booth, C. D. Mackereth, V. Saridakis, I. Ekiel, G. Kozlov, K. L. Maxwell, N. Wu, L. P. McIntosh, K. Gehring, M. A. Kennedy, A. R. Davidson, E. F. Pai, M. Gerstein, A. M. Edwards, and C. H Arrowsmith. "Structural Proteomics of M. thermoautotrophicum: A global survey of non-membrane protein expression, solubility, and structure." *Nature Structural Biology*, **7**, 903-909 (2000).

Acknowledgments

EVK is grateful to L. Aravind for his help in the analysis of the COGs for target selection. J. R. Cort is a 1999 recipient of an American Cancer Society postdoctoral fellowship.

Ultrahigh Field NMR Applications of Stable Isotopes

M. A. Kennedy, L. A. Silks III,[(a)]
C. J. Unkefer,[(a)] K. McAteer,[(b)] and N. Isern

(a) Los Alamos National Laboratory
(b) Associated Western Universities

Project Description

This year our research focused on the use of residual dipolar couplings as a means of improving the quality of DNA structures determined by NMR methods.

Abstract

Restrained molecular dynamics is widely used to calculate DNA structures from NMR data. Results of an *in silico* experiment show that the force field can be significant when compared to the NMR restraints in driving the final structures to converge. Specifically, we observed that 1) the influence of the force field leads to artificially tight convergence within final families of structures and 2) the precision and character of resulting structures depend on the choice of force field used in the calculations. A canonical B-DNA model was used as a target structure. Distances, dihedral angles, and simulated residual dipolar couplings (RDCs) were measured in the target structure and used as restraints. X-PLOR and Discover, which use CHARMM and AMBER force fields, respectively, were tested and found to produce different final structures despite the use of identical distance and dihedral restraints. Incorporation of residual dipolar coupling restraints in X-PLOR improved convergence with the target structure indicating that the force field dependence can potentially be overcome if RDC restraints are employed.

Technical Accomplishments

Restrained molecular dynamics with relaxed bounds. Calculations were carried out with X-PLOR (CHARMM) and Discover (AMBER) using restraints with ranges typical of those used in NMR structure calculations. Using B-DNA as a starting structure these calculations produce families of structures that superimpose on the canonical B-DNA target structure, with average RMSD values of 0.93 Å and 1.10 Å, respectively (Figure 2.28 A and B). The six lowest energy structures from within the final family of structures calculated in X-PLOR converge with an average RMSD value of 1.01 Å (Figure 2.28 C) and the convergence for the final family of Discover-generated structures is 0.41 Å (Figure 2.28 D). Notably, the lowest energy structures from X-PLOR and Discover calculations superimpose on each other with an average RMSD value of 1.45 Å (Figure 2.28 E), which is comparably larger than their respective RMSD values with the target structure. For final families of structures calculated from A-DNA starting structures, convergence with the canonical B-DNA target structure is 1.26 Å for X-PLOR and 1.22 Å for Discover. Convergence of a family of structures is commonly used as a criterion for evaluating whether or not a structure is adequately restrained. However, these results indicate that the convergence within the individual families produced by X-PLOR (1.01 Å) or Discover (0.41 Å) is significantly tighter compared to convergence with the families of structures produced by the alternative force field (1.45 Å). This result indicates that the force field has a greater influence compared to the distance and dihedral angle restraints, in driving the convergence within individual families of structures to such small values.

Restrained molecular dynamics with tight bounds. To determine if the dependence on the force field could be overcome by tightening the range of restraints, calculations were repeated using artificially tight restraint sets. The results shown in Figure 2.29 indicate that different force fields produce different final average structures, even when identical NOE and dihedral angle restraints with exceptionally tight bounds are used during the refinement procedure. The lowest energy structures from X-PLOR and from Discover calculations using canonical B-DNA as a starting structure superimpose on canonical B-DNA target structure with average RMSD values of 0.62 Å and 1.01 Å, respectively. The convergence between a family of structures starting from B-DNA was 0.68 Å with X-PLOR and 0.32 Å with Discover. The average RMSD values between the target structure, B-DNA, and final structures that were calculated using A-DNA as the starting structure were 1.20 Å for X-PLOR and 1.20 Å for Discover. The convergence between a family of structures starting from A-DNA was 0.69 Å for X-PLOR and 0.26 Å for Discover. Again, these results indicate that the convergence within the individual families produced by X-PLOR or Discover is significantly tighter compared to convergence with either the target structure or with the family of structures produced by the alternative force field. Since satisfactory convergence with the target function was achieved with both programs starting from a B-DNA starting structure, one might expect that the final structures from both X-PLOR and Discover would superimpose with an acceptable RMSD value of < 1.0 Å. However, even with artificially 'tight' restraints the X-PLOR structure superimposes on the Discover structure with an average RMSD value of 1.26 Å (Figure 2.29 E), indicating that the force field, and not the NMR data, dominates in driving the structures to converge to artificially small values.

Restrained molecular dynamics with RDC restraints. Residual dipolar couplings can be accurately measured, but to date they have had limited use in DNA structure determination. Using relaxed restraints, addition of a single RDC restraint (J_{err} 0.6 Hz) per base (the N3-H3 vector on thymine, the N1-H1 vector on guanine, the C6-H6 vector on cytosine, and the C8-H8 vector on adenine) to NOE and dihedral angle restraints improved the RMSD value between the final structure and the target structure from 0.93 Å to 0.78 Å, using B-DNA as the starting structure. Under these conditions the average RMSD value between the lowest energy structures from three different rMD calculations improved from 1.01 Å to 0.70 Å. Convergence with the target structure also improved from 1.26 Å to 1.05 Å for final structures calculated from a canonical A-DNA starting structure. Additional RDC restraints (five per residue) resultedin an average RMSD value between the lowest energy structure and the target structure of 0.75 Å and an average RMSD value of 0.57 Å between the six lowest energy structures using B-DNA as the starting structure. Figure 2.30 shows the progressive improvement in final structures as the number and classes of restraints are increased in the rMD calculation, starting from the canonical B-DNA model. The RMSD value between

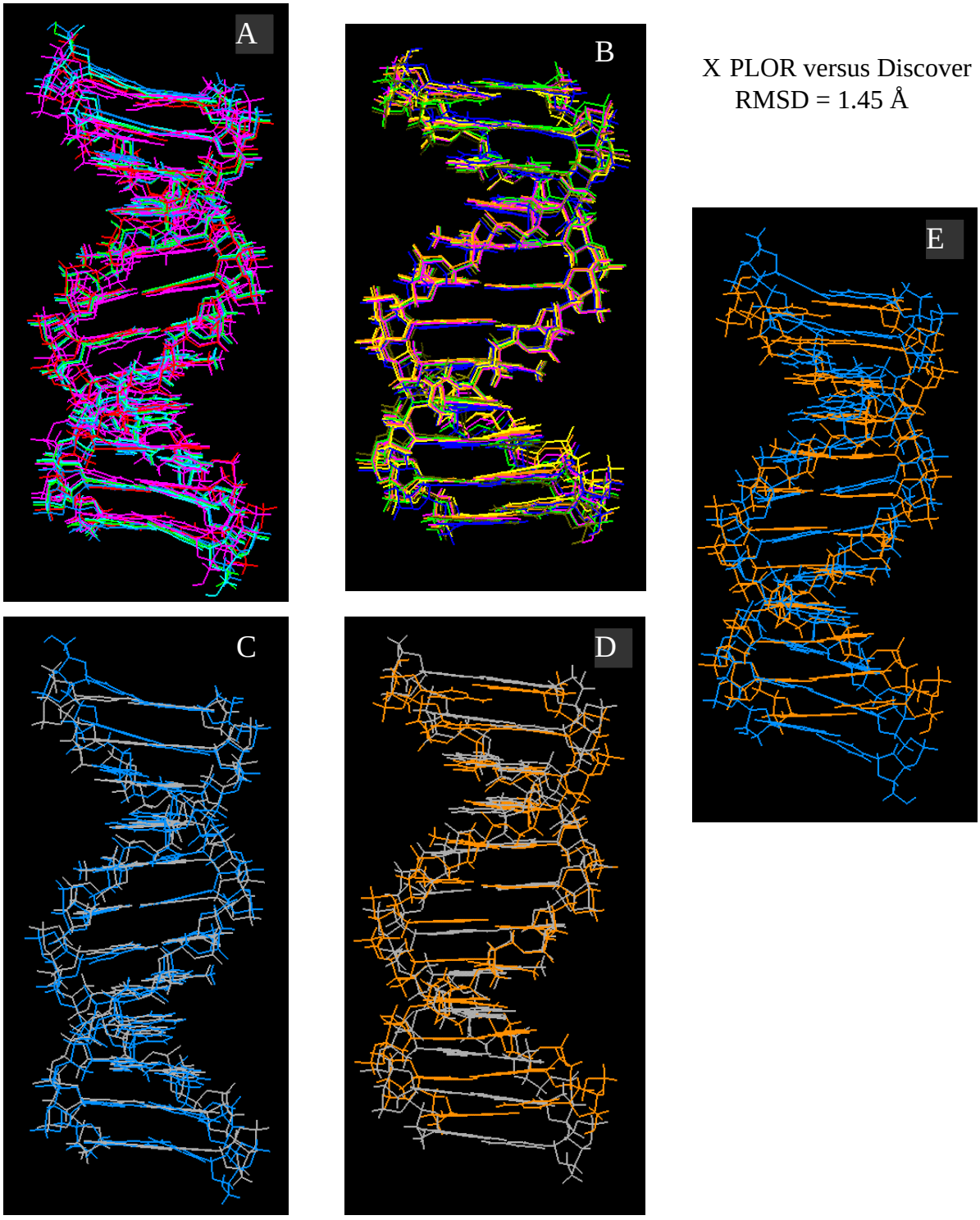


Figure 2.28. Average structures of [d(CGCGAATTCGCG)$_2$] calculated with identical relaxed restraints (see text) during rMD in X-PLOR and Discover. (A) Family of 6 X-PLOR structures, RMSD = 1.01 Å. (B) Family of 6 Discover structures, RMSD = 0.41 Å. (C) X-PLOR structure superimposed on B-DNA (gray), RMSD = 0.93 Å. (D) Discover structure superimposed on B-DNA (gray), RMSD = 1.10 Å. (E) X-PLOR structure (blue) superimposed on the Discover structure (orange), RMSD = 1.45 Å.

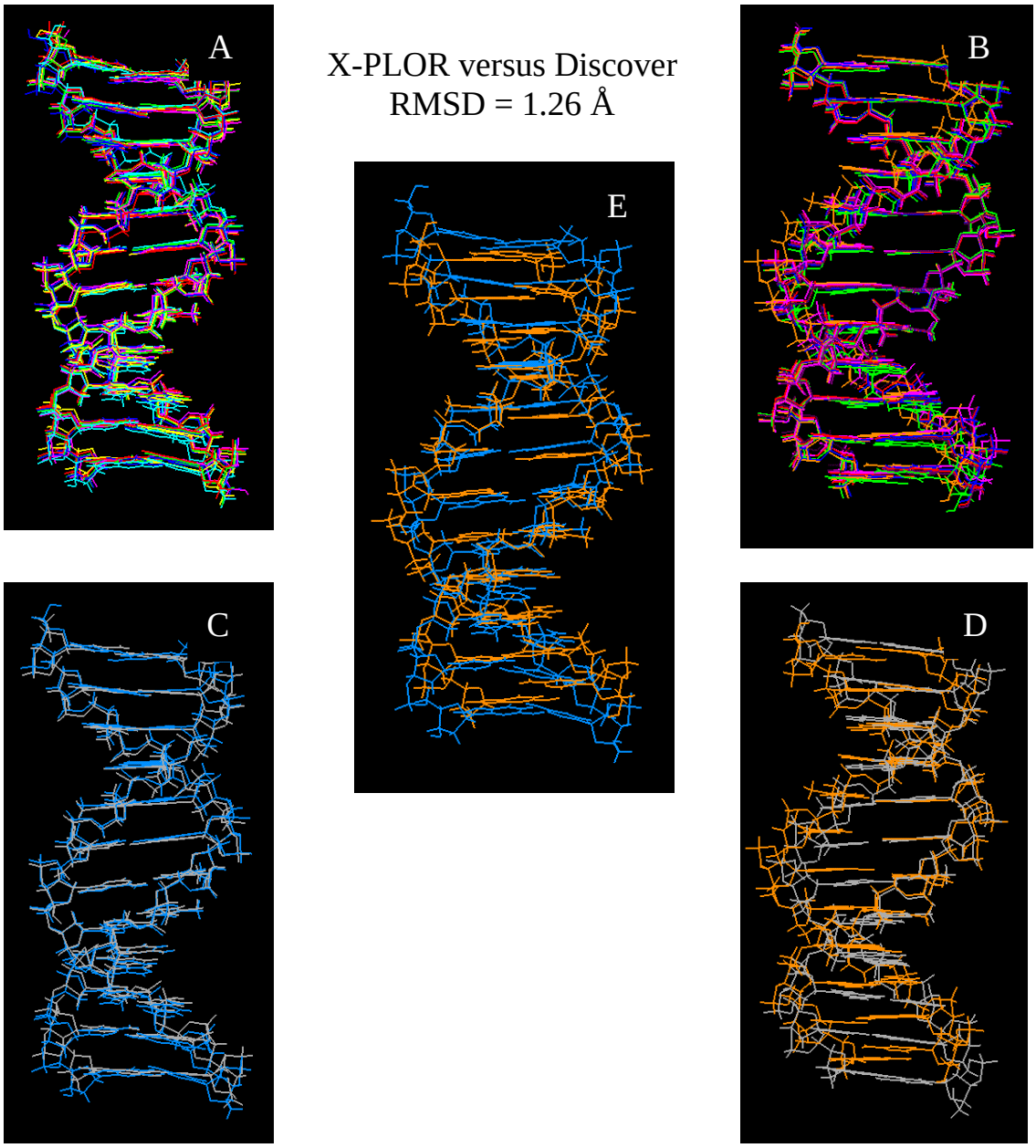


Figure 2.29. Average structures of [d(CGCGAATTCGCG)$_2$] calculated with identical tight restraints during rMD in X-PLOR and Discover. (A) Family of 6 X-PLOR structures, RMSD = 0.68 Å. (B) Family of 6 Discover structures, RMSD = 0.32 Å. (C) X-PLOR structure superimposed on B-DNA (gray), RMSD = 0.62 Å. (D) Discover structure superimposed on B-DNA (gray), RMSD = 1.01 Å. (E) X-PLOR structure (blue) superimposed on the Discover structure (orange), RMSD = 1.26 Å.

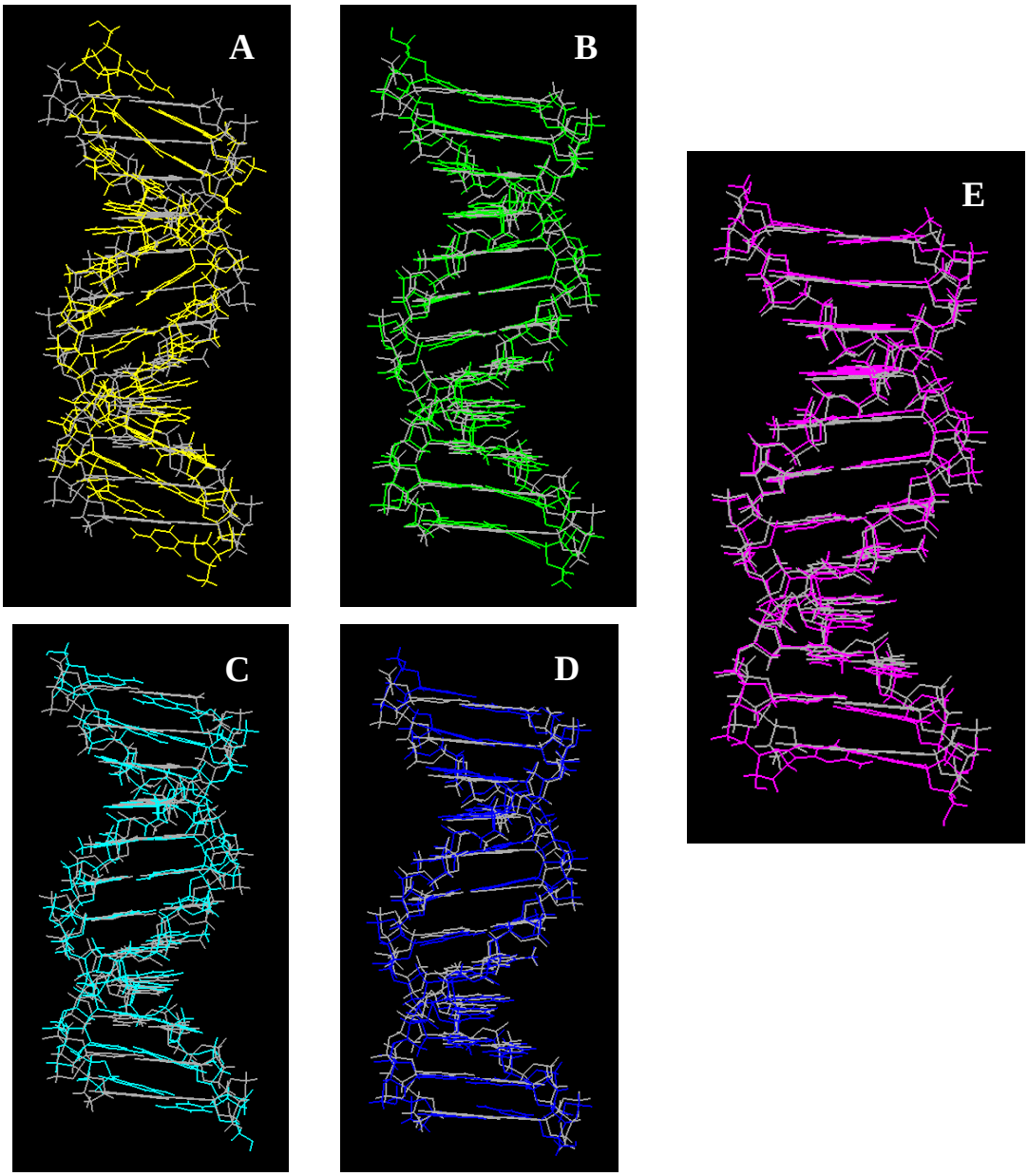


Figure 2.30. Average structures of [d(CGCGAATTCGCG)$_2$], calculated with X-PLOR, superimposed on minimized B-DNA (gray). (A) NOEs only (relaxed), RMSD = 2.94 Å. (B) NOEs (relaxed) + dihedrals (relaxed), RMSD = 0.89 Å. (C) NOEs (relaxed) + residual dipolar coupling restraints (1/base, relaxed), RMSD = 1.13 Å. (D) NOEs (relaxed) + dihedrals (relaxed) + residual dipolar coupling restraints (1/base, relaxed), RMSD = 0.78 Å. (E) NOEs (relaxed) + dihedrals (relaxed) + all residual dipolar coupling restraints (relaxed), RMSD = 0.75 Å.

a final structure refined with distance restraints only (relaxed bounds) and the B-DNA target is 2.94 Å (Figure 2.30 A). Incorporation of dihedral angle restraints (relaxed bounds) reduces this value to 0.89 Å (Figure 2.30 B). When distance restraints and one RDC restraint per base are used, the resulting final structure superimposes on the target structure with a value of 1.13 Å (Figure 2.30 C). Adding dihedral angle restraints improves the super imposition to a value of 0.78 Å (Figure 2.30 D). Furthermore, when five RDC restraints per residue and distance restraints with relaxed bounds are included in the calculations, the RMSD value between the final average structure and canonical B-DNA is 0.77 Å. Adding dihedral angle restraints with relaxed bounds improves convergence on the target structure to 0.75 Å (Figure 2.30 E). These data suggest that when RDC restraints are used in a refinement procedure starting from a canonical B-DNA model, highly converged families of structures can be generated even in the absence of dihedral angle restraints. For A-DNA starting structures, convergence improved significantly when five RDC restraints were used in the calculations. The average RMSD value with the target structure improved from 1.26 Å to 0.99 Å. The average RMSD value between the five lowest energy structures improved from 1.40 Å to 0.72 Å, and the value between structures from different calculations improved from 1.20 Å to 0.43 Å. Clearly, using RDCs in NMR-based DNA structure refinement results in improved final structures using the X-PLOR software. Unfortunately it was not possible to incorporate RDCs into Discover calculations. Although the RMSD numbers for structure refinement with the Discover program are acceptable, it would be interesting to know if RDC restraints improve convergence and produce structures that have greater similarity to those resulting from X-PLOR calculations than in the cases described above.

Conclusions

The results of the *in silico* experiment presented here highlight the strong influence that the force field can have in NMR-based rMD DNA structure calculations in the absence of RDC restraints. Specifically we observed that the convergence within the individual families produced by X-PLOR (CHARMM) or Discover (AMBER) is significantly tighter compared to convergence with either the target structure or with the family of structures produced by the alternative force field. This observation indicates that while the NMR restraints confine the structure to a reasonable geometry that is near the solution, the final precision with which a family of structures is determined, measured by the RMSD value, is ultimately driven by the force field used in rMD calculations and depends on the choice of force field. The use of artificially tight restraints was unable to overcome this influence of the force field. A practical consequence of this observation is that one should not depend on families of structures generated by a single force field for interpretation of the fine details of DNA structures in the absence of residual dipolar couplings. A more prudent approach would be to calculate structures using both AMBER and CHARMM or other force fields in order to determine the structural features that persist notwithstanding which force field is used. This realization is significant since the fine details of NMR-based DNA structures are frequently analyzed with the goal of understanding damage recognition and other DNA-protein interactions. An alternative approach, when possible, is to measure RDCs in labeled DNA samples and to use these to provide additional restraints in the structure calculations. Incorporation of simulated RDC restraints into X-PLOR calculations significantly improves convergence of the final average structure with the target structure, and has the potential to overcome the dependence of NMR-derived DNA structures on the force field used in restrained molecular dynamics calculations.

Acknowledgements

The authors would like to thank Dr. Nico Tjandra for providing the patch to include residual dipolar couplings into X-PLOR calculations.

Invited Talks

Michael A. Kennedy, Kathleen McAteer, Ryszard Michalczyk, Jurgen Schmidt, Nancy Isern, Louis Silks, and Cliff Unkefer "Measurement of Residual Dipolar Couplings in ^{15}N-labeled and ^{13}C-labeled DNA aligned in Liquid Crystalline Bicelles using NMR Spectroscopy", 41st Experimental Nuclear Magnetic Resonance Conference, April 9-14, 2000, Asilomar Conference Center, Asilomar, California.

Development of Novel Means to Significantly Reduce the Dielectric Losses Associated with Biological NMR Samples

P. D. Ellis, M. A. Kennedy, and R. Wind

Project Description

Dielectric losses associated with biological NMR samples at high magnetic fields (>17 T) ultimately determine the potential NMR signal-to-noise (S/N) ratio. These losses have a direct effect of lengthening (quadratically) the time required to acquire NMR data sets required to deduce the three-dimensional structure of a protein in solution. The goal of this research is to develop technical means for minimizing dielectric losses in biological NMR experiments. If successful, this could lead to a significant reduction in NMR data collection times leading to more rapid structure determinations. One of the most innovative new ideas for reducing the dielectric losses in biological NMR samples is to encapsulate proteins in reverse micelles dissolved in low dielectric bulk solvents. Preparing and characterizing such systems by small angle x-ray scattering and NMR spectroscopy will be the primary focus of this project.

Introduction

In the post-genomic era, the scientific focus is changing from determining the complete DNA sequence of the human genome to characterizing their gene products. The expected number of proteins from the human genome is ~100,000. Proteomics, defined as the study of the entire complement of proteins expressed by a particular cell, organism, or tissue type at a given time for a given disease state or a specific set of environmental conditions, promises to bridge the gap between genome sequence and cellular behavior. An integral challenge involves the characterization of the biological functions of these proteins and analysis of the corresponding three-dimensional structure. This is referred to as functional and structural genomics, respectively. Ultimately, the goal of structural genomics is to determine the structure of a sufficiently large subset of the ~100,000 proteins such that "basis" structures could formulated. These basis structures would, in turn, be used as predictors of the remaining structures, while simultaneously providing a rationale for the observed function all of the gene products.

Currently, x-ray crystallography requires diffraction-quality crystals, selenium-labeled proteins, and access to high-intensity light sources available at the synchrotrons located at the U.S. national laboratories. Given these conditions, x-ray data sets can be collected in a matter of 2 to 4 hours and virtually have no limitation with respect to protein size. However, dynamic or statically disordered regions of proteins are invisible to x-ray crystallography. NMR spectroscopy has several unique capabilities that are complementary to x-ray crystallography: 1) proteins can be examined under physiological solution-state conditions; 2) dynamic regions of the proteins can be well characterized; and 3) intermolecular complexes can be easily studied as a function of pH, ionic strength, etc. However, NMR currently has two significant limitations when compared to x-ray crystallography: 1) data sets currently take on the order of 60 days to collect and 2) the size of proteins amenable to NMR structure determination is currently limited to ~50 kDa.

If the determination of the structure of large subset of the ~100,000 proteins is what is needed, then x-ray crystallography data collection would require maximally between 24 and 45 years of synchrotron time assuming protein crystals are obtainable at high enough quality in all cases. On the other hand, for this same subset, given the current technology, NMR spectroscopy would require approximately 16,438 years of NMR time. Estimates of the cost per structure with current technologies are running around $50,000 to $100,000 each. The cost of such an effort would be between $5 billion and $10 billion. Ideally, one could imagine dividing the task of structural genomics equally between x-ray crystallography and NMR spectroscopy, but, at the current state of technology, it is anticipated that NMR would play a limited role. If NMR spectroscopy is to contribute significantly to this effort, then the data collection time for the NMR experiment must be dramatically reduced, and the applicable molecular weight range must be expanded to >100 kD.

Given the existence of high-field (>17-T magnets) NMR spectrometers, why does it take so long to collect the critical data sets? A significant recent advance is the introduction of cryoprobe technology in which the NMR coil and preamplifier are cooled to near liquid helium temperatures. This advance, which

can theoretically increase NMR sensitivity by a factor of 3 to 4, holds the prospect of reducing NMR data collections times by as much a factor of 10 in time. However, cryoprobe performance will reach a practical limit at high fields because of the consequences of the dielectric losses of the sample have on NMR sensitivity. If the sample losses could be significantly reduced, the sensitivity would increase and optimal cryoprobe performance could be obtained. Virtually all proteins require water as a solvent in the presence of varying concentrations of salts (typically < 200 mM). This solvent/salt/buffer combination produces a conductive sample that, in turn, generates inductive losses that arise from the formation of eddy currents in the sample. Under normal circumstances, these sample losses and not losses associated with the coil itself, dominate the Q of the receiver coil of the probe. The consequences of these losses increase strongly with the frequency of the spectrometer and the sample diameter. Hence, the higher the field or the larger the sample at a fixed field, the worse the problem becomes. If these losses could be eliminated, then the length of time required for data collection could be significantly reduced. In this project, research will be carried out to determine if a general approach for encapsulating proteins in reverse micelles dissolved in low dielectric bulk solvents as a way of minimizing dielectric losses in during NMR data collection can be realized.

Results and Accomplishments

The primary technical task for the project was to assemble and characterize reverse micelle systems containing encapsulated proteins. A systematic approach was taken to explore the range of solution conditions under which stable and well behaved, spherical, non-aggregated, or non-complex lipid structures are formed. The primary characterization methods was small-angle, x-ray scattering. Preliminary experiments were carried out and were successful in demonstrating that small-angle, x-ray scattering was an effective technique for characterizing the dimensions of reverse micelles that contained either solvent or in the presence of encapsulated proteins. A single anionic micelle system AOT and pentane was explored. SAXS proved to be a senstive means of characterizing reverse micelle structure in terms of dimensions and geometrical organization.

Summary and Conclusions

The objective of this project was to establish methods and systems that will facilitate encapsualtion of proteins in reverse micelles under a broad range of solution conditions (i.e., pH, salt, temperature, and pressure) so that this method can be generally applied to protein structure determinations. The proposed program has the potential to produce a dramatic reduction in the experimental time needed to acquire NMR data for protein structure determinations. If these methods are successful, they could lead to significant time savings and could dramatically incrase the molecular weight of proteins whose structures can be determined by NMR. Furthermore, the structure of transmembrane proteins integrated into model membranes may be addressable by these methods.

The work carried out in this project demonstrated that the combination of small-angle, x-ray scattering as a screening method for the preparation of encapsulated protein samples in reverse micelles for NMR data collection is an effective strategy, and if systematically pursued, it could provide an efficient means of developing a broad range of possible reverse micelles systems for use in protein structure determination by NMR.

Related Publications

Fulton, J. L. and R. D. Smith. "Reverse Micelle and Microemulsion Phases in Supercritical Fluids." *J. Phys. Chem.*, **92**, 2903-2907 (1988).

Fulton, J. L., J. P. Blitz, J. M. Tingey, and R. D. Smith. "Reverse Micelle and Microemulsion Phases in Supercritical Xenon and Ethane: Light Scattering and Spectroscopic Probe Studies." *J. Phys. Chem*, **93**, 4198-4204 (1989).

Wand, A. J., M. R. Ehrhardt, and P. F. Flynn. "High Resolution NMR of Encapsulated Proteins Dissolved in Low-Viscosity Fluids." *Proc. Natl. Acad. Sci.*, **95**, 15299-15302 (1998).

Second EMSL Workshop on Structural Genomics

M. A. Kennedy

The sequencing of the human genome is complete, and the number of completely sequenced genomes for other organisms is rapidly growing. Scientists now face a significant challenge in determining the function for all the proteins encoded by the various genomes.

Proteomics and Structural Genomics: The task of proteomics is more complicated than that of genomics. Even the term proteomics has taken on several different meanings. One commonly used definition of proteomics means to quantify the all the proteins expressed at any given time in a cell. This definition is rather limited in comparison to the definition analogous to of that of genomics, the goal of which is to determine the complete DNA sequence for all the genes contained in an organism's complete genome. In a direct analogy, proteomics might be defined as the determination of the function of all the proteins encoded by the organism's entire genome. This idea has been referred to as "functional genomics," but it is sometimes referred to as "functional proteomics." Achieving either the former or the latter goal is substantially more complicated than achieving the goal of genomics. Living cells have the potential of experiencing an infinite variety of conditions over the course of a lifetime, and so characterization and quantitation of protein expression levels must interpreted in the context of this variability. On the other hand, understanding protein function is complicated by the fact that protein function is carried out through interactions with other proteins, DNA, small molecules, metal ions, cofactors etc. A single protein can have multiple functions that can only be identified given the knowledge and context of its interacting partners. Furthermore, understanding the function of any protein at the molecular level ultimately requires knowledge of the protein structure. This endeavor has been coined "structural genomics," but is sometimes referred to as "structural proteomics." In many cases, even knowledge of the structure of a protein will not be sufficient to understand its function since many protein functions are carried out in dynamic regions of proteins that may not have a well defined structure. Clearly, proteomics by any definition will require a substantial effort well into the forseeable future and will require effort on several fronts, including protein structure determination on a genomic scale, coordinated efforts to characterize biochemical function and efforts aimed at determining protein expression levels.

Objectives

The Second EMSL Workshop on Structural Genomics, held July 28 and 29, 2000, had several objectives.

1. To bring representatives from DOE labs together to build new interactions and collaborations to meet the challenge of structural genomics.

2. To bring together members of the NMR community from around the country and world to discuss a) the role that NMR spectroscopy will play in structural and gunctional genomics, b) the technical challenges NMR faces for structural genomics and c) to report progress in meeting those challenges.

3. To bring together scientists from the Pacific Northwest region to build new interactions and collaborations.

4. To bring principal investigators of pilot projects in structural genomics to the meeting to provide updates on progress and stimulate exchange of information.

5. To discuss the need for dedicated nationally supported NMR facilities for structural genomics.

Summary of the Workshop

The workshop was held over two days and included five scientific sessions and a round-table discussion. The agenda included 22 speakers. The program included speakers and attendees from Brookhaven National Laboratory, Los Alamos National Laboratory, Lawrence Berkeley National Laboratory, Argonne National Laboratory, Oak Ridge National Laboratory, and Pacific Northwest National Laboratory. The program also included many leaders in the NMR and x-ray community from around the country and the world including Cheryl Arrowsmith

from Canada, Shigeyuki Yokoyama from Japan, Iain Campbell from England, Hartmut Oschkinat from Germany, and Michal Linial from Israel. The primary funding agencies were represented by Jana Werhrle (NIH-NIGMS), Abraham Levy (NIH-NCRR) and Roland Hirsch (DOE). The meeting was attended by 65 registered attendees, 47 of which were external visitors.

The attendees and speakers created a stimulating environment for the meeting that allowed for substantial cross-fertilization of ideas. The EMSL facility was put on display and advertised to a diverse group of potential users. The entire meeting was video-taped, and we are in the process of making each talk available as a quick-time movie on our web-site for the meeting. A formal summary of the workshop is under preparation. The greatest significance of the meeting was the effective communication between scientists and the funding agency representatives concerning 1) the significant role that NMR can be expected to play in the field of structural genomics over the coming years and 2) the need for national dedicated NMR facilities for structural genomics.

From a non-scientific perspective, the Tri-City community was favorably displayed to the attendees. All meals were catered to maximize time for the scientific program. The Friday evening session was held at the Hedges Winery offering a wine tasting and with entertainment by local musician Dan Myers who played piano, flute, and classical guitar. Saturday evening included a barbecue and drinks provided by the Ice Harbor Brewery and entertainment by the Classical Guitar Society of the Tri Cities. The meeting was well supported by five commercial vendors that made much of the special programs possible within the allowed budget.

Table 2.2. 2nd Annual EMSL Workshop on Structural Genomics Scientific Program

William R Wiley Environmental Molecular Sciences Laboratory
Pacific Northwest National Laboratory JULY 28-29, 2000

FRIDAY MORNING, JULY 28

8:00am-9:00am Check-in, Badging, Coffee and Donuts, EMSL High Field Magnetic Resonance Facility Open House

SESSION I: Welcome and Opening Remarks by EMSL Leadership

9:00am-9:10am David Koppenaal, EMSL Associate Director, Pacific Northwest National Laboratory

Statements by Funding Agency Representatives
9:10am-9:30am Janna Wehrle, National Institutes of General Medical Sciences
9:30am-9:50am Abraham Levy, National Center for Research Resources
9:50am-10:10am Roland Hirsch, Department of Energy
10:10am-10:30am Refreshment Break

DOE Programs in Structural Genomics

10:30am-11:10am Andrzej Joachimiak, Argonne National Laboratory
11:10am-11:50am Geoff Waldo, Los Alamos National Laboratory
11:50am-12:30pm Michael Kennedy, Pacific Northwest National Laboratory
12:30pm-1:30pm Lunch Break Catered Lunch by Tri-Tech Catering

FRIDAY AFTERNOON, JULY 28

SESSION II: Computational Tools for Structural Genomics

1:30pm-2:10pm Michal Linial, The Hebrew University, Jerusalem, Israel
2:10pm-2:50pm David Baker, The University of Washington
2:50pm-3:10pm Refreshment Break
3:10pm-3:50pm Ying Xu, Oak Ridge National Laboratory
3:50pm-4:30pm Keith Dunker, Washington State University

FRIDAY EVENING, JULY 28

4:30pm-5:30pm Bus Transportation to Hedges Winery
5:30pm-7:00pm Wine Tasting at Hedges Winery
6:00pm-7:30pm Catered Dinner by Leo's Catering/Blue Moon Restaurant
Entertainment by Dan Myers: Piano, Guitar, Flute

SESSION III: International Programs in Structural Genomics
7:30pm-8:30pm Iain Campbell, University of Oxford

SATURDAY MORNING, JULY 29

8:30am-9:00am Coffee and Donuts, EMSL High Field Magnetic Resonance Facility Open House

SESSION IV: International Programs in Structural Genomics (Continued)
9:00am-9:40am Shigeyuki Yokoyama, University of Tokyo
9:40am-10:20am Cheryl Arrowsmith, University of Toronto
10:20am-11:00am Hartmut Oschkinat, Protein Structure Factory, Berlin

Ultra-High Field NMR in the United States
11:00am-11:40am Tim Cross, National High Magnetic Field Laboratory
11:40am-12:40pm Lunch Break Catered Lunch by Casa Chapala

SATURDAY AFTERNOON, JULY 29

SESSION V: NMR Technical Challenges for Structural Genomics Research

12:40pm-1:20pm David Cowburn, The Rockefeller University
1:20pm-2:00pm John Markley, University of Wisconsin
2:00pm-2:40pm Tom Szyperski, SUNY Buffalo
2:40pm-3:00pm Refreshment Break
3:00pm-3:40pm Frank Delaglio, The National Institutes of Health
3:40pm-4:20pm Rich Withers, Bruker Instruments
4:20pm-5:00pm Gaetano Montelione, Rutgers University
5:00pm-6:30pm EMSL Catered Barbecue Dinner
Catered by Leo's Catering/Blue Moon Restaurant
Beer by Ice Harbor Brewing
Entertainment by the Classical Guitar Society of the Tri Cities

SATURDAY EVENING, JULY 29

SESSION VI: Discussion of Workshop Topics
6:30pm-7:00pm NMR Challenges for Structural Genomics:
Discussion Leaders-David Cowburn and Gaetano Montelione
7:00pm-7:30pm Role of NMR in Structural/Functional Genomics
Discussion Leader-Cheryl Arrowsmith
7:30pm-8:00pm National NMR Center for Structural Genomics
Discussion Leader-Michael Kennedy
8:00pm-8:30pm Organization of NMR Community
Discussion Leader-John Markley
8:30pm-9:00pm Open Discussion of Topics
Moderator-Michael Kennedy
9:00pm Adjourn

A.- D. Workshop attendees listening to the proceedings.

E. Some of the speakers (clockwise from top left): Michal Linial, Shigeyuki Yokoyama, Tim Cross, Geoffrey Waldo.

F. Friday evening wine tasting at Hedges Winery, Benton City, Washington.

List of Attendees:

Name	Organization
Arrowsmith, Cheryl	University of Toronto
Bair, Raymond A.	PNNL
Baker, David	University of Washington
Bowers, Peter M.	University of Washington
Buchko, Garry	University of Vermont/PNNL
Campbell, Iain	University of Oxford
Cho, Herman	PNNL/EMSL
Cort, John	PNNL/EMSL
Cotten, Myriam	University of Washington
Cowburn, David	The Rockefeller University
Cross, Tim	National High Magnetic Field Laboratory
Daughdrill, Gary W.	University of Idaho
Delaglio, Frank	National Institutes of Health
Dunker, Keith	Washington State University
Ellis, Paul D.	PNNL/EMSL
Everts, Sarah	University of British Columbia
Ford, Joe	PNNL/EMSL
Gao, Xiaolian	University of Houston
Guilfoy, Gerry	Cambridge Isotope Laboratories, Inc.
Hess, Nancy	PNNL
Hirsch, Roland	Department of Energy
Joachimiak, Andrezj	Argonne National Laboratory
Joshi, Manish	University of British Columbia
Kennedy, Michael	PNNL/EMSL
Kim, Hak Jun	Washington State University
Koppenaal, David	PNNL/EMSL
Larson, Gary	Isotec, Inc.
Lee, Gregory M.	University of British Columbia
Lee, Min S.	Molecumetics, Ltd.
Levy, Abraham	National Institutes of Health
Linial, Michal	The Hebrew University of Jerusalem
Lowry, David F.	PNNL/EMSL
Mackereth, Cameron D.	University of British Columbia
Manrao, Suraj P.	Martek Biosciences
Markley, John	University of Wisconsin-Madison
McAteer, Kate	PNNL/EMSL
Medvick, Patricia A.	PNNL
Montelione, Guy	Rutgers University
Nakanishi, Hiroshi	Molecumetics, Ltd.
Nieman, Ronald A.	Arizona State University
Oschkinat, Hartmut	Protein Structure Factory, Forschungsinstitut fuer Molekulare Pharmakologie
Resat, Haluk	PNNL
Rohl, Carol	University of Washington
Rommereim, Donald	PNNL/EMSL
Schaerpf, Manuela E.	University of British Columbia, UBC
Schaufler, Lawrence	University of Washington
Sekharan, Monica R.	University of Washington
Song, Jianxing	Biotechnology Research Institute, National Research Council of Canada
Stauffer, Mellisa E.	Washington State University
Strauss, Charlie	University of Washington/Los Alamos National Lab
Szyperski, Thomas	State University of New York at Buffalo
Tian, Fang	Complex Carbohydrate Research Center
Ulrich, Eldon L.	University of Wisconsin-Madison
Veenstra, Timothy D.	PNNL/EMSL
Vise, Pam	University of Idaho
Waldo, Geoffrey S.	Los Alamos National Laboratory
Walters, Ronald A.	PNNL
Wehrle, Janna P.	National Institutes of Health, National Institute of General Medical Sciences
Withers, Richard	Bruker Instruments, Inc.
Xu, Ying	Oak Ridge National Laboratory
Yee, Adelinda A.	University of Toronto
Yokoyama, Shigeyuki	University of Tokyo, RIKEN Genomic Sciences Center
Young, John K.	Washington State University

Corporate Sponsors:

Varian Associates
Bruker Instruments
Cambridge Isotopes, Inc.
Martek Biosciences Corporation
Isotec, Inc.

3. Magnetic Resonance Research Users Report

Use of NMR Microscopy to Examine the Sub-Pore-Scale Structure of Porous Media with Multiple Liquid/Solid Phases

B. D. Wood,[(a)] K. Minard,
M. Oostrom,[(b)] and E. M. Murphy[(c)]

(a) PNNL Interfacial Geochemistry Group
(b) PNNL Hydrology Group
(c) PNNL Fundamental Science Division

Chemical, physical, and microbiological processes in natural porous media are manifest at a hierarchy of scales from the molecular scale to the scales associated with geological features. In many instances, processes that occur at the sub-pore scale have a dramatic influence on the so-called macroscopic behavior measured over a representative volume of porous media. One of the most fundamental challenges for understanding the physics of transport phenomena in porous media is to determine how processes that occur at disparate scales can be formally connected.

Methods for upscaling the microscopic physics to obtain a macroscopic representation of multiphase media, such as in the method of volume averaging, have been in existence for some time. The goal of such approaches is to begin with a description of a multiphase system written at the continuum level (i.e., at a level where individual phases are resolved at scales above the statistical-dynamical scale, but well below the spatial correlation length of individual phases), and develop the macroscopic conservation equations via a formal averaging procedure.

In order to connect the microscopic and macroscopic scales using methods such as volume averaging, one must have information about the microscopic scale properties in a representative volume of the multiphase system. The measurement of such microscale properties is now within the reach of NMR microscopy. These techniques offer the potential to make the critical connection between theory and measurement in complex multiphase porous systems. Such measurements have the potential to direct research in multiphase porous media towards a laboratory scale, which has been unattainable previously.

We have been allotted small segments of NMR time to conduct several proof-of principal experiments for using NMR microscopy to elucidate porous media structure and processes at the sub-pore scale. Recently, we were able to successfully resolve the fluid-filled pore space of a granular porous medium; a 5 x 5 x 5-mm cube of porous medium (0.3-mm average grain diameter) was imaged at a resolution of 62 microns, thus demonstrating the potential to achieve a resolution of 31 microns for such samples. An example of the porous media sample is shown in Figure 3.1.

We are unaware of any other studies that have obtained satisfactory results on the structure of porous media at this scale using NMR. A rendering of the structure of the porous media that was used in our study appears as Figure 3.2.

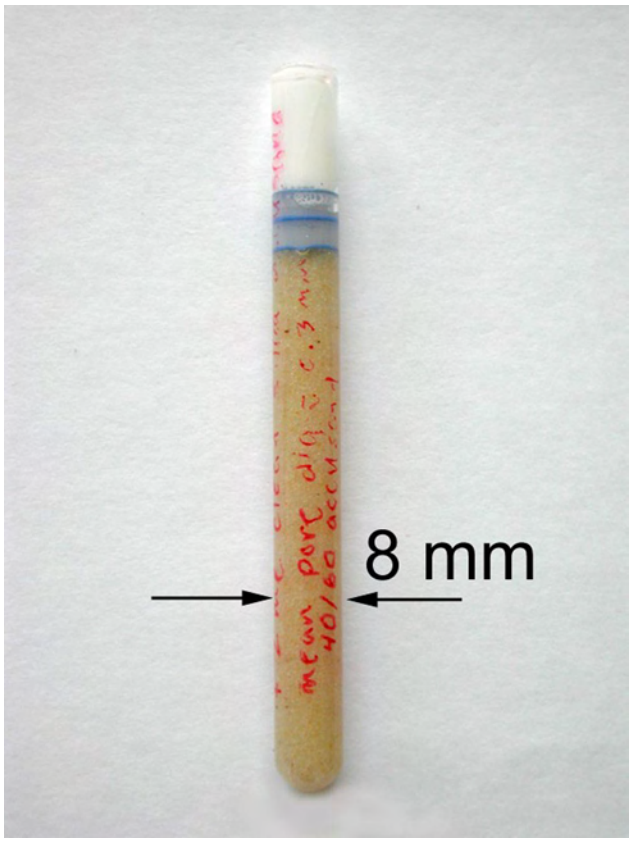


Figure 3.1. A sample of porous media used for NMR microscopy study. The average grain size of this media is 0.3 mm. A 5 x 5 x 5-mm cube of porous media was imaged, representing a region containing over 4000 particles (see Figure 3.2).

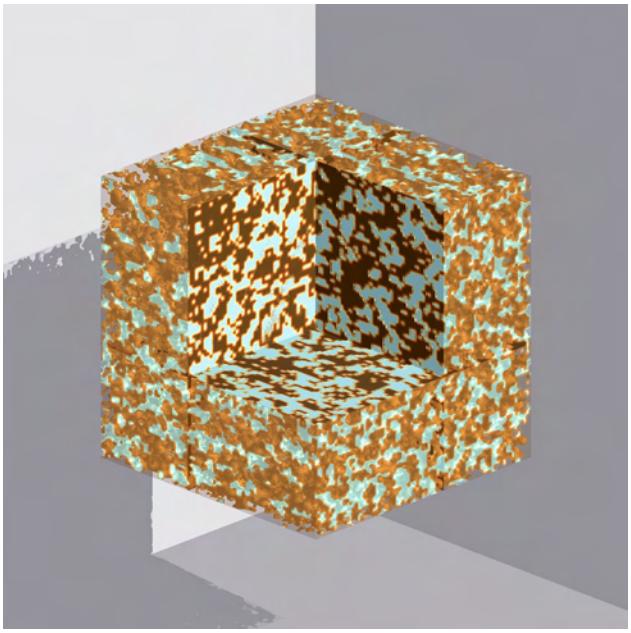


Figure 3.2. A rendering of the NMR image generated for the porous medium illustrated in Figure 3.1. Brown (black in cross-section) regions represent solids; blue regions represent liquid. The scale of the cube is 5 x 5 x 5 mm.

Currently, we are proposing to extend this work to problems involving 1) multi-phase fluids and 2) microbial biomass distributed within a porous network at a resolution of 31 microns. Our proposed research will be organized as follows.

Measurement of the Fluid Phase Distribution of a Two-Phase System

For this work, NMR microscopy will be used to develop a three-dimensional map of the distribution of water and oil phases occupying the pore space of a porous medium. For this preliminary study, only a single media type will be imaged under static hydrodynamic conditions. A 0.3-mm grain diameter material will be prepared at a pre-determined oil and water saturation. Then, a volume of 8 x 8 x 8-mm will be imaged at a 30-μm resolution to determine the distribution of the solid, oil, and water phases in the system.

Some attempts have been made to develop macroscopic representations of the equations for conservation of oil and water in such three-phase (two fluid) systems. Although there is still much work to be done, some preliminary results have suggested that it is not the oil and water phase volume fractions that control the effective properties (as conventionally thought) but, in fact, the area of the phase contacts. If NMR microscopy is able to resolve these phase contacts, this method will provide a strong justification for proposing research of multi-phase fluid flows in porous media at the sub-pore scale.

Measurement of the Distribution of Microorganisms in Porous Media

For this work, NMR microscopy will be used to develop a three-dimensional map of the distribution of microbial mass in the pore space of a porous medium. Again, only a single media type will be imaged under static hydrodynamic conditions. A 0.3-mm grain diameter material will be prepared under conditions where a biofilm (*Shewanella spp*.) has been grown to fill part of the void space through the media. A volume of 8 x 8 x 8-mm will be imaged (possibly using a diffusion-weighted NMR method (Potter et al. 1996) or relaxation method (Hoskins et al. 1999) at a 31-μm resolution to determine the distribution of the distributions of attached biomass in the system.

The distribution of microorganisms at the pore scale is a problem of interest in bioremediation, microbial transport, and medical biotechnolgy. If NMR microscopy is able to resolve these phase contacts, this method will provide a strong justification for proposing additional research regarding the micron-scale distributions of organisms in porous media.

References

Hoskins, B. C., L. Fevang, P. D. Majors, M. M. Sharma, and G. Georgiou. Selective imaging of biofilms in porous media by NMR relaxation. *Journal of Magnetic Resonance*, **139**, 67-73 (1999).

Potter, K., R. L. Kleinberg, F. J. Brockman, and E. W. McFarland. Assay for bacteria in porous media by diffusion-weighted NMR. *Journal of Magnetic Resonance*, Series B, **113**, 9-15 (1996).

Conformational and Dynamic Studies of Human Salivary Histatin Bound on Hydroxyapatite by Solid-State NMR

M. Cotten, G. P. Drobny,[(a)] and P. S. Stayton[(b)]

(a) University of Washington
(b) Department of Chemistry and Department of Bioengineering

Biomineralization, a field that studies the formation of mineral tissues by organisms, offers structural biologists many opportunities to address some core issues of structure-function relationships and molecular recognition at interfaces. Some important implications of better understanding biomineralization include some new advances in materials science using biomimetics. To learn more about biomineralization phenomena, we have been studying human salivary polypeptides found in the acquired enamel pellicle. In this project, we focus on histatin-5, a polypeptide that is histidine-rich (see Figure 3.3) and possesses at least two important functions: control of hydroxyapatite (HAP, calcium phosphate) crystal growth and antimicrobial activity. Previous studies have characterized functionally important regions of the peptide sequence, and some secondary conformation analysis in solution have been done. Very little is known about specific histatin-5/ surface interactions and the conformation of the HAP-bound peptide. This knowledge nevertheless is necessary to better understand molecular recognition and structure-function relationships. Using solid-state NMR, our primary goals have been to characterize the conformation of histatin-5 both free (lyophilized samples) and bound to HAP crystals (lyophilized and hydrated samples). Moreover, we have been interested in identifying interactions between the peptide and HAP. In our laboratory at the University of Washington, solid-state NMR experiments that measure internuclear distances to sub-Angstrom accuracy (DRAWS, Dipolar Recoupling with A Windowless Sequence) are being performed to determine distances between two ^{13}C carbonyl labels in adjacent amino acids (residues i and i+1) and, thereby, constrain the secondary structure of the peptide bound to HAP.

At EMSL, we have started to use REDOR (Rotational Echo DOuble Resonance) to complement the DRAWS structural constraints with additional distance

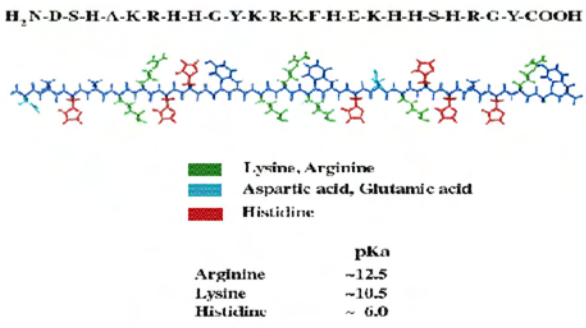


Figure 3.3. Histatin-5, the polypeptide studied in this project.

constraints. REDOR is a technique that has been successfully and widely used for characterizing $^{13}C/^{15}N$ distances in peptides. Here, the goals have been 1) to perform REDOR to characterize distances in specifically labeled (i/i+4) $^{13}C/^{15}N$ spin pairs in the peptide backbone and 2) to study the dynamic properties of the aromatic side chains. Interest in this subject developed as the first goal was investigated last year at EMSL. We used the 500-MHz, 89-mm bore, Varian Unity spectrometer at EMSL, and a triple resonant Doty probe. Some highlights of the work follow.

Set-up for REDOR

The XY8 version of REDOR was used since it includes a phase cycling that minimizes experimental imperfections such as off resonance and finite pulse width effects. Two samples were used to set up REDOR: 1) a glycine sample labeled on the carbonyl carbon and amide nitrogen (2.5 Å distance) and 2) α-helical peptide ("LK" peptide, sequence: Ac-LKK[^{13}C]LLKL[^{15}N]LKKLLKL, 4.2 Å distance). The resulting dephasing curve for this latter case (i.e., the S/S_0 ratio as a function of rotor cycle number) is shown in Figure 3.4 (S and S_0 are the signal intensities of the ^{13}C isotropic chemical shift when the ^{15}N pulses are on and off, respectively). The experimental data just bracketed between the 4 and 4.5 Å distances, which is consistent with the 4.2 Å distance expected for a α-helical peptide.

REDOR Experiments on Histatin-5 Samples

Next, the REDOR experiment was applied to three samples of labeled histatin-5: [^{13}C]Ala_4-[^{15}N]His_8

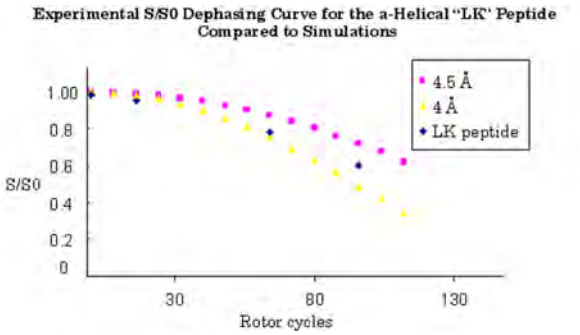


Figure 3.4: REDOR dephasing curves. Experimental data for the α-helical peptide, [$^{13}C_1$]Leu$_4$–[^{15}N]Leu$_8$ "LK", are compared to simulations curves for distances of 4 and 4.5 Å. The S/S$_0$ ratios for the ^{13}C isotropic signal from the labeled carbonyl carbon are plotted as a function of REDOR rotor cycles. The intensities are corrected for the background signal from natural abundance ^{13}C carbonyls. The 180° pulse widths for ^{13}C and ^{15}N were 8.6 and 17.5 μsec, respectively. The resonance frequency for ^{13}C was 125.8 MHz. The spinning rate was 6 kHz (i.e., each rotor cycle lasted 166.7 μsec). The simulations took into account finite pulse width effects.

hsn-5, [^{13}C]Tyr$_{10}$-[^{15}N]Phe$_{14}$ hsn-5, and [^{13}C]His$_{19}$-[^{15}N]Gly$_{23}$ hsn-5. Because of their physiological relevance, hydrated samples have been of the highest interest. However, at room temperature, the signal sensitivity is very poor, which indicates that some dynamic phenomena occurring on the NMR time scale prevents efficient cross polarization and hence signal detection. We had hoped that cooling the samples to change the motional time scale would solve this problem.

It was a great advantage to be able to run low temperature experiments (T=-37°C) on the 500-MHz, wide-bore at EMSL since we recovered signal from the hydrated samples. It was therefore possible to perform REDOR not only a dry sample ([^{13}C]His$_{19}$-[^{15}N]Gly$_{23}$) but also on all three hydrated samples mentioned above. The results we have acquired to date indicate that the ^{13}C/^{15}N spin pair dipolar interactions, and thereby distances, in all three peptides are close to or beyond the detectable limits of REDOR. Consequently, the peptides do not appear to adopt a α-helical secondary structure when they are bound to HAP and hydrated.

Dynamics Information Obtained on Histatin-5

Interesting observations have been made about the peptide dynamics by comparing spectra obtained for various conditions: room temperature versus T=-37°C, short versus longer delays between ^{13}C transients, and dry versus hydrated conditions. For instance, by increasing the pulse delay from 4 to 10 seconds, some additional signals appear in the spectrum recorded at T=-37°C (Figure 3.5). Their positions are consistent with side bands calculated from published isotropic chemical shifts for the aromatic carbons in the histidine, phenylalanine, and tyrosine side chains. This suggests that due to their large number, the natural abundance ^{13}C in the aromatic side chains of histatin-5 gives rise to significant signals. The fact that these side chains have

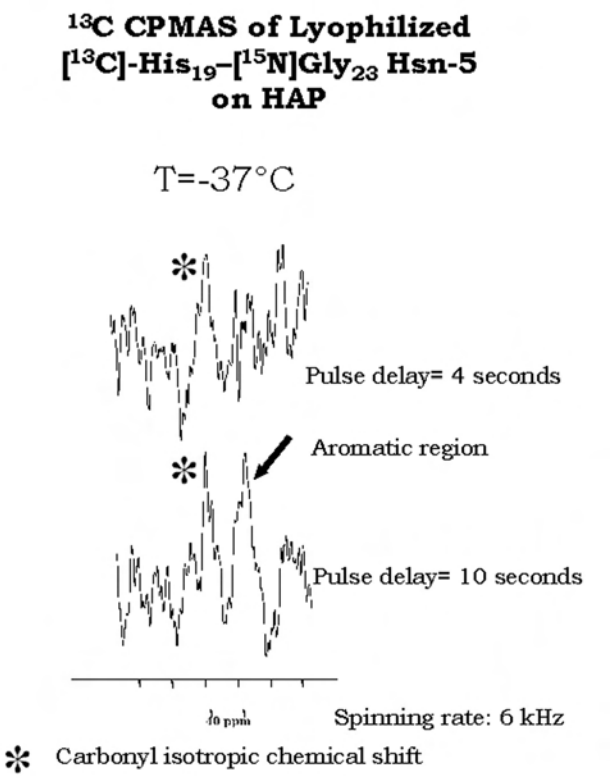


Figure 3.5. Effects of the pulse delay times on the ^{13}C signal intensities. The spectra display ^{13}C CPMAS signals from lyophilized [^{13}C]His$_{19}$-[^{15}N]Gly$_{23}$ hsn-5 bound on HAP using a pulse delay of 4 (top) and 10 (bottom) seconds between ^{13}C transients. The data was collected at 125 MHz using a spinning rate of 6 kHz.

longer relaxation times than the backbone carbonyl carbons suggests these aromatic side chains are motionaly constrained compared to the backbone. This may indicate that they are in contact with the HAP surface, thereby loosing motional freedom.

In light of these exciting preliminary studies, we want to study further the conformation and dynamics of histatin-5 when it is bound to HAP and hydrated.

Related Publications

Gregory, et al. *Chem. Phys. Letters*., **246**, 654-663 (1995).
Gullion, T. and J. Schaefer. *J. Mag. Reson.*, **81**, 196-200 (1989).
Fu, R. and T. A. Cross. *Ann. Rev. Biophys. Biomol. Struct.*, **28**, 235-268 (1999).

NMR Relaxation Reveals Fast Internal Motions and Conformational Heterogeneity for Flexible Protein Fragments of *E. coli* Thioredoxin

G. W. Daughdrill, P. D. Vise, M. L. Tasayco, and D. F. Lowry

Background

For many proteins and protein domains, the amino acid sequence is the only information needed to form a compact rigid structure. In contrast, the amino acid sequences of many other proteins and protein domains specify partially collapsed, flexible structures. For proteins that adopt compact rigid structures the average positions of individual atoms are relatively fixed and can often be accurately defined, as evidenced by the more than 12,000 empirically derived structures of compact rigid proteins contained in the protein data bank (PDB). The formation of these compact rigid structures tend to be dominated by the presence of intramolecular interactions between amino acid residues that are far apart in the linear sequence (long-range interactions). For proteins that adopt partially collapsed, flexible structures, the average positions of individual atoms are not fixed and are best described by an ensemble of rapidly interconverting structures. Proteins that form partially collapsed, flexible structures tend not to have long-range interactions between amino acids. Further, the ensemble of rapidly interconverting structures depends on the allowable Ramachandran space for adjacent residues and the propensity of a given sequence to form helical or sheet structures. While the presence and importance of partially collapsed, flexible structures in nature is well established, it is unclear by what mechanism the fragmentation of a protein that normally forms a compact rigid structure leads to the formation of a partially collapsed, flexible structure. To clarify this problem, we have developed a family of flexible, complementary fragments derived from oxidized *E. coli* Thioredoxin (Trx) as a model system to study the partially collapsed, flexible state. This report summarizes a few results of NMR relaxation studies on two of these fragments corresponding to the N-terminal 73 amino acids (trx1-73) and the C-terminal 35 amino acids (trx74-108) of *E. coli* Thioredoxin. The data presented in this report were collected at the EMSL High Field Magnetic Resonance Facility.

Results and Discussion

Figure 3.6 shows the temperature dependence of the heteronuclear Overhauser effect (NH-NOE) for trx1-73. The NH-NOE is a rapid, sensitive measure of the rotational correlation time in proteins. As temperature is increased for the trx1-73 fragment, many of the amide ^{1}H-^{15}N pairs undergo a transition from a positive value (blue peaks) to a negative value (green peaks), indicating an increase in the rotational correlation time for those residues. We know from previous structural studies that trx1-73 does not contain any of the long-range NOEs that are observed in full-length thioredoxin. Therefore, trx1-73 does not form a compact rigid structure. However, the temperature dependence of the NH-NOE suggests the presence of a partially collapsed, flexible structure that is temperature sensitive. The numerous positive peaks at the lower temperature, 15°C, indicate the presence of correlated motion in those residues. As the temperature is increased to 35°C the motion for most of the residues in trx1-73 becomes uncorrelated, as indicated by the negative peaks. A few peaks remain positive at the higher temperature. The majority of these peaks are between residues I23-K36. The fact that they remain positive at the higher temperature suggests the motion is still correlated. This result is consistent with differential scanning calorimetry (DSC) measurements indicating the trx1-73 fragment contains buried hydrophobic surface area in a temperature range of 20 to 40°C. With the exception of F12, all of the aromatic

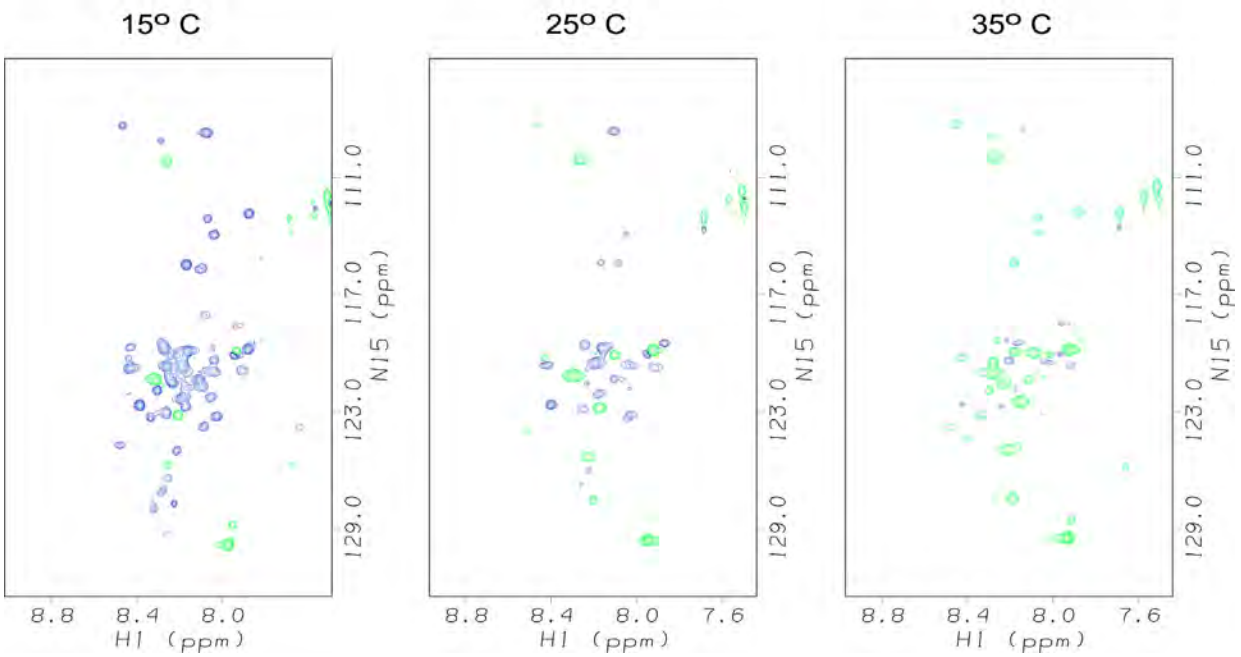


Figure 3.6. Spectra showing the heteronuclear Overhauser effect between the amide proton and its attached nitrogen, NH-NOE, for trx1-73 at 15°, 25°, and 35°C. Proton chemical shifts are plotted on the horizontal axis and Nitrogen chemical shifts are plotted on the vertical axis. As temperature is increased several the NH-NOE for several amide proton-nitrogen pairs goes from positive (blue peaks) to negative (green peaks) indicating and decrease in rate of molecular tumbling. Many of the amide proton-nitrogen pairs that remain positive are thought to form a 'folding nucleus' in full-length trx.

residues in the trx1-73 fragment are between residues I23-K36 (F27, W28, and W31). Figure 3.7 shows a Rasmol image of the coordinates of residues 1-73 from the crystal structure of full-length *E. coli* Thioredoxin. Side chains are shown for residues that have positive NH-NOE values at 35°C. The distribution of residues in Figure 3.7 suggests some residual native structure in the trx1-73 fragment. In particular, residues in β2 and subset of rapidly inter-converting structures with an α2 (V25, D26, F27, W28, W31, and K36) and α2 and β3 (D47, Y49, K52, and L53) are likely to exist as a average that is similar to the native structure. However, the fast exchange that these residues undergo limits their ability to form stable, long-range interactions and thus explains the absence of long-range NOE's for the trx1-73 fragment. Future studies of trx1-73, and other thioredoxin fragments will help us better understand the requirements for forming compact rigid structures or partially collapsed flexible structures. This work would not have been possible without the expertise of the EMSL staff. In particular the authors wish to thank Dave Hoyt, Joe Ford, and Nancy Isern for assistance with sample handling and experiments.

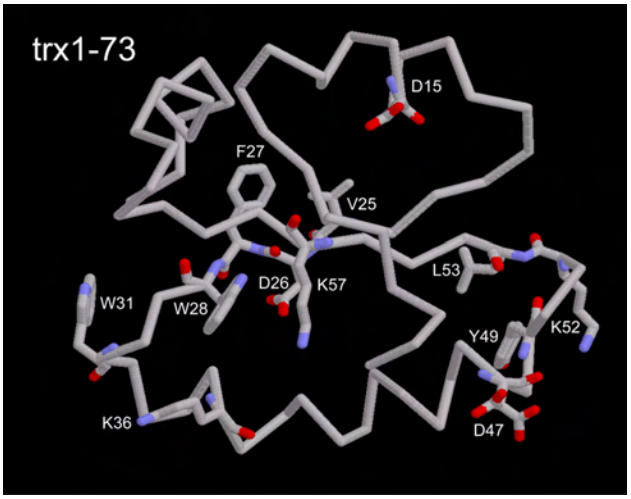


Figure 3.7. Rasmol image of the coordinates of residues 1-73 from the crystal structure of full-length *E. coli* Thioredoxin. Side chains are shown for residues that have positive NH-NOE values at 35°C.

Solution Structures of the Dead Ringer and Mu Repressor Protein-DNA Complexes

R. T. Clubb,[(a)] J. Iwahara,[(a)] and J. Wojciak[(a)]

(a) Department of Chemistry and Biochemistry and the UCLA-DOE Laboratory of Structural Biology and Molecular Medicine, University of California, Los Angeles, California

We are using NMR spectroscopy to elucidate the structural basis of DNA recognition by two novel binding domains: 1) the AT-rich interaction domain (ARID) from the Dead Ringer protein and 2) the "winged-helix" binding domain of the Mu Repressor protein. During the past year, access to the 750- and 800-MHz NMR spectrometers at the EMSL played an important role in advancing these projects. These instruments afforded increased resolution and sensitivity to unambiguously identify the set of intermolecular contacts within each protein-DNA interface and enabled the collection of intramolecular distance restraints to characterize the structure of the DNA component. The data has enabled us to determine the solution structure of the Mu Repressor-DNA complex to a high degree of precision, which has recently been published (Wojciak et al. 2001). It is also being used in the structure elucidation of the Dead Ringer-DNA complex (in progress).

Project 1: Structure of a Novel Protein-DNA Complex: The AT-Rich Interaction Domain-DNA Complex

The Dead ringer protein from *Drosophila melanogaster* is a transcriptional regulatory protein required for early embryonic development. It is the founding member of a large family of DNA-binding proteins that interact with DNA through a highly conserved domain called the AT-rich interaction domain (ARID). The ARID motif is unrelated to any of the common domains used to recognize DNA (homeodomains, zinc fingers, etc.) and presumably interacts with the minor-groove. ARID containing proteins are present in a variety of eukaryotic organisms and have been shown to participate in several biologically significant processes. Members of the family include among others: the yeast SWI1 protein, from the SWI/SNF complex involved in global transcriptional activation; Bright, a B-cell-specific trans-activator of IgH transcription; RBP1 and RBP2, retinoblastoma binding factors; PLU-1, a protein that is up-regulated in breast cancer cells and the mammalian proteins Jumonji, SMCx, SMCy, Mrf-1, and Mrf-2. Dead ringer is itself highly conserved, with specific orthologs in mammalian genomes (mouse and human), suggesting that it plays a fundamental role in embryonic development.

We have collected high-field NMR data on the specific complex formed between the Dead Ringer protein (residues Gly262 to Gly398) and its cognate DNA binding site (total molecular weight 25.8 kilodaltons). Specifically, we have performed 1) a three-dimensional 13C F1-filtered, F3-edited NOESY experiment for identification of intermolecular NOEs, 2) a two-dimensional 13C/15N-filtered NOESY experiment for the assignment and structure determination of the DNA in the complex, and 3) a three-dimensional 15N-edited NOESY experiment for the structure determination of the bound protein.

The EMSL data has been useful in defining the structure of the DNA and in the assignment of intermolecular distance restraints to define the interface of the complex. Figure 3.8 shows representative data collected on the 800-MHz NMR spectrometer at the EMSL. At present, we have generated a medium resolution structure of the complex derived from data collected at the EMSL and at UCLA. Although still in progress, we anticipate submitting a manuscript describing this work in the near future. The structure will reveal for the first time how the ubiquitous ARID class of eukaryotic binding domains recognizes DNA.

Project 2: Structure of the 'Winged Helix'-DNA Complex Involved in DNA Transposition

Transposition of the Mu bacteriophage serves as a paradigm for the reactions that transpose DNA, since the movement of this large and efficient element has been extensively characterized *in vitro*. The phage encoded transposase (MuA) recombines DNA in the context of several stable higher order nucleoprotein complexes that are built by the assembly of catalytically inactive MuA monomers. A key step in protein assembly relies upon interactions between an

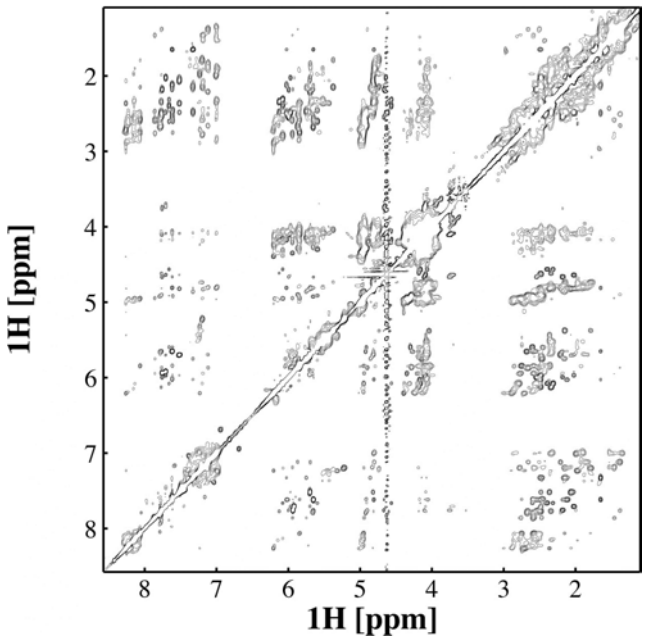


Figure 3.8. Two-dimensional 13C/15N-filtered NOESY spectrum of the Dead Ringer-DNA complex. All cross-peaks originate from the DNA component of the complex enabling its structure to be determined.

N-terminal DNA-binding domain in the MuA protein and DNA sequences within an enhancer element (also called the internal activation sequence, IAS), facilitating transposition in a distance-independent manner. Interestingly, a highly homologous IAS-binding domain is also present in the Mu repressor protein (MuR), but its binding promotes lysogeny of the phage by repressing the expression of genes required for lytic growth and by directly blocking MuA access to the IAS.

We have used data from the EMSL to determine the structure of the Mu repressor-DNA complex. Figure 3.9 shows representative data collected at the EMSL on a complex between unlabeled Mu repressor protein and its cognate DNA site labeled C13 and N15. The data clearly demonstrates strong contacts between the H2 protons of the minor groove and the side chains of lysines-53 and -56 of the protein. This data along with a three-dimensional C13 F1-filtered, F3-edited NOESY experiment were collected at the EMSL to identify intermolecular NOEs. In addition, we also recorded a two-dimensional N15 HSQC data set at 800 MHz to derive small magnitude residual dipolar couplings that result from the anisotropic magnetic sucessibility of the complex. The modest

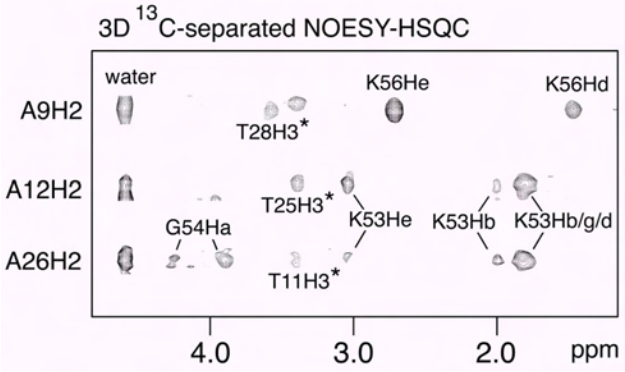


Figure 3.9. Representative three-dimensional C13-edited NOESY-HSQC NMR spectra collected on the 800-MHz spectrometer at the EMSL. The data shows that the protein interacts with the minor groove of the DNA. Cross-peaks from folded imino protons are dashed and marked with an asterisk.

degree of field dependent alignment enabled the DNA and protein components of the complex to be positioned relative to one another, thereby increasing the accuracy of the final three-dimensional structure.

Using this data the solution structure of the complex between the winged-helix enhancer-binding domain of the Mu repressor protein and its cognate DNA site was determined (Figure 3.10) and has recently been published (Wojciak et al. 2001).

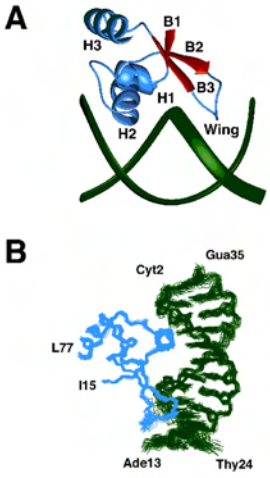


Figure 3.10. Three-dimensional structure of the Mu Repressor-DNA complex. a) Ribbon diagram of the complex. b) Ensemble of 20 conformers of the complex calculated from NMR data obtained at UCLA and the EMSL.

The structure reveals an unusual use for the 'wing,' which becomes immobilized after deeply penetrating the minor groove for intermolecular hydrogen bonding. Although the wing is mobile in the DNA-free state, it partially negates the large entropic penalty associated with its burial by maintaining a small degree of structural order in the absence of DNA. Extensive contacts also are formed from the recognition helix, which reads the major groove of a highly conserved region of the binding site through a single base-specific hydrogen bond and van der Waals contacts.

References

Wojciak J. M., J. Iwahara, and R. T. Clubb. "The Mu repressor-DNA complex contains an immobilized 'wing' within the minor groove." *Nature Structural Biology*, **8**(1):84-90 (2001).

The Apo-Form of the Substrate-Binding Domain of DnaK, an Hsp70 Chaperone

J. F. Swain and L. M. Gierasch

Departments of Biochemistry and Molecular Biology and Chemistry, Lederle Graduate Research Tower, University of Massachusetts, Amherst, Massachusetts

Of critical importance to many problems in cell biology and medicine is an understanding of how proteins fold in the cell. In the past decade, it has become clear that folding and maturation of proteins *in vivo* are facilitated by molecular chaperones. The most ubiquitous family of chaperones characterized to date is the Hsp70 family, representatives of which occur in all compartments of eukaryotic cells, and in prokaryotes (Ellis and Harthl 1999). Hsp70 chaperones are required intracellularly both for survival under stress conditions and for a wide variety of normal cellular functions, including folding of newly synthesized proteins, translocating proteins into organelles, disrupting macromolecular complexes, regulating transcription factors, and proteolytic degradation (Rüdiger et al. 1997). The wide variety of roles they play implies that Hsp70s can display both selective and promiscuous polypeptide substrate recognition. These two-domain proteins assist proper folding by binding to exposed hydrophobic stretches in unfolded protein chains in an ATP-regulated manner. With ADP bound to the N-terminal nucleotide-binding domain, the C-terminal substrate-binding domain has high affinity for substrates; whereas, with ATP bound, substrates bind with low affinity and faster on/off rates.

High-resolution crystal structures of each of these domains in isolation have provided great insight into the overall architecture of Hsp70s. The nucleotide-binding domain forms a bi-lobed structure with a deep cleft where nucleotide binds (Flaherty et al. 1990, Harrison et al. 1997). The substrate-binding domain co-crystallized with a heptapeptide (NRLLLTG, termed NR) consists of a twisted β-sandwich capped by a α-helical bundle (Zhu et al. 1996). Substrate peptide is nestled in an extended conformation between loops at one edge of the β-sandwich, and is covered by a hydrophobic arch of loop side chains, as well as the helical bundle. The peptide-Hsp70 binding interaction is primarily composed of backbone-backbone H-bonds, with a deep hydrophobic pocket in the β-sandwich accommodating the central leucine of the NR peptide.

Understanding the peptide-bound state of the substrate-binding domain only allows us to describe one state of this allosteric protein. If the conformational change necessary to produce the empty state were known, it would be possible to develop a model for allosteric signaling in the Hsp70s. However, attempts to study the empty conformation of the substrate-binding domain have been stymied by intramolecular association with C-terminal sequences that mimic substrate (Morshauser et al. 1999, Wang et al. 1998). In the case of a truncated DnaK substrate-binding domain (corresponding to residues 389-561), the tail sequence ^{540}DHLLHSTR^{547} is inserted into the substrate-binding pocket in the same orientation as the NR peptide in the crystal (Zhu et al. 1996, Wang et al. 1998). In this study, the tail sequence could not be competed out of the binding site with excess NR peptide, presumably due to the entropic advantage of the tail being a tethered ligand.

Our work with a similar fragment of DnaK (corresponding to residues 387-552) has shown that simply reducing the buffer pH from 7 to 5 opens a window on the as yet uncharacterized empty state. We hypothesized that this pH-dependent shift might be caused by

release of the C-terminal tail residues from the substrate-binding site upon protonation of the histidines flanking leucines 542-543. Titration of the protein with NR peptide at pH 5 using NMR indeed reveals that NR can bind at this pH, and does so in the slow exchange regime.

Based on the exciting possibility that we could observe the empty state of the DnaK substrate-binding domain at pH 5, we collected high-resolution [^{1}H,^{15}N] and [^{1}H,^{13}C] NOESY-HSQC spectra on the 800-MHz Varian Inova NMR spectrometer at EMSL in order to generate distance constraints for structure calculation of this 19-kDa protein. Soon thereafter, however, assignment of the NR-bound state revealed that our conclusions were premature. By far, the residues that undergo the largest chemical shift change when NR binds to the protein are those in the tail sequence that was observed to bind intramolecularly in the substrate-binding pocket at pH 7 (Figure 3.11, c.f. His541 in panel c) (Wang et al. 1998). This unexpected result illustrated very clearly that this C-terminal tail sequence still occupies the binding site most of the time at pH 5, which is contrary to our earlier conclusions.

Rather than reduce the pH to further favor the empty state, we mutated the two leucines in the C-terminal substrate-mimicking sequence to a more hydrophilic sequence (L542Y/L543E) in order to strongly disfavor tail-binding and generate the true apo-form. Whereas in the original protein in which the NR peptide bound with low affinity and the binding isotherm did not saturate, the new construct binds NR strongly and stoichiometrically at NMR concentrations (Figure 3.11). Furthermore, the C-terminal tail residues are insensitive to NR-binding, indicating that they do not occupy the substrate-binding site in the mutated protein (Figure 3.11d, c.f. His 541).

Recently, an NMR structure was published reporting the empty conformation of a DnaK substrate-binding domain fragment lacking the entire helical subdomain (Pellechia et al. 2000). This β-domain fragment is characterized by substantial protein dynamics on the intermediate time-scale, as well as a large conformational change in edge strand β3 compared to its NR-bound form. Based on the conformational change in β3 of the β-domain, a model for allosteric coupling in the Hsp70 proteins was proposed in which strand

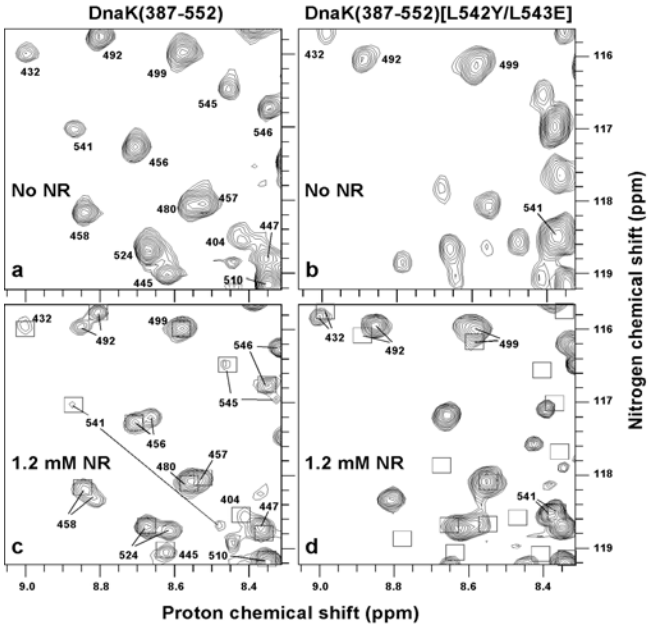


Figure 3.11. The L542Y/L543E mutant of DnaK(387-552) does not bind its C-terminal tail sequence, as shown by the fact its C-terminal residues are unaffected by NR binding. The binding affinity for NR is also increased substantially, as 1.2 mM NR is sufficient to saturate the mutant version, whereas only ~40% of the parent construct bound NR under the same conditions. [^{1}H,^{15}N]HSQC spectra of 0.4 mM DnaK(387-552) (a, c) or DnaK(387-552)[L542Y/L543E] (b, d) were collected at pH 5 and 35 C in the absence (a, b) or presence (c, d) of 1.2 mM NR peptide. Boxes in c and d indicate the peak positions when NR is not bound.

β3 communicates information about occupancy of the substrate-binding site to the nucleotide-binding domain. Strikingly, in our construct which retains most of the helical sub-domain, the chemical shift changes in strand β3 when NR binds are extremely small, and thus do not support any role of β3 in allosteric coupling to the nucleotide-binding domain. Assignment of the true empty state of the DnaK substrate-binding domain is currently underway, and will be crucial for understanding the mechanism of allostery in the Hsp70 chaperones.

References

Ellis, R. J. and F. -U. Hartl. *Current Opinion in Structural Biology*, **9**, 102-110 (1999).

Flaherty, K. M., C. Deluca-Flaherty, and D. B. McKay. *Nature*, **346**, 623-628 (1990).

Harrison, C. J., M. Hayer-Hartl, M. Di Liberto, F. U. Hartl, and J. Kuriyan. *Science*, **276**, 431-435 (1997).

Morshauser, R. C., W. Hu, H. Wang, Y. Pang, G. C. Flynn, and E.R.P. Zuiderweg. *Journal of Molecular Biology*, **289**, 1387-1403 (1999).

Rüdiger, S., A. Buchberger, and B. Bukau. *Nature Structural Biology*, **4**, 342-349 (1997).

Pellecchia, M., D. L. Montgomery, S. Y. Stevens, C. W. Vander Kooi, H. -P. Feng, L. M. Gierasch, and E.R.P. Zuiderweg. *Nature Structural Biology*, **7**, 298-303 (2000).

Wang, H., A. V. Kurochkin, Y. Pang, W. Hu, G. C. Flynn, and E.R.P. Zuiderweg. *Biochemistry*, **37**, 7929-7940 (1998).

Zhu, X., X. Zhao, W. F. Burkholder, A. Gragerov, C. M. Ogata, M. E. Gottesman, and W. A. Hendrickson. *Science*, **272**, 1606-1614 (1996).

Structure of the PBX Homeodomain Bound to a 14mer DNA Duplex

T. Sprules,[(a)] N. Green,[(b)]
M. Featherstone,[(a,b)] and K. Gehring[(a,c)]

(a) Department of Biochemistry, McGill University
(b) McGill Cancer Centre, McGill University
(c) Montreal Joint Centre for Structural Biology, McGill Univeristy

Homeodomain proteins are an important class of DNA binding proteins that act as transcription co-factors. The molecular features important for homeodomain DNA binding and complex formation have been studied extensively. This research project involves investigating the interaction between the PBX homeodomain, its DNA binding site, and its co-factors using NMR spectroscopy.

PBX (pre-B-cell transformation related gene) is a proto-oncogene. It was first identified due to its involvement in pre-B cell leukaemias. In these cancers, a 1:19 translocation occurs in which the 5' half of the E2A gene, containing its DNA-binding and dimerization motifs, is replaced by sequences from PBX, including its homeodomain. PBX and its homologues, *Drosophilia extradenticle* (Exd) and *Caenorhabditis elegans ceh*-20, the PBC family of proteins, belong to the three amino acid loop extension (TALE) superclass of homeodomains. The extra amino acids, located in the loop between helices one and two, are important for co-operative interactions with HOX homeodomains.

Full-length monomeric PBX is unable to interact with DNA, and its isolated homeodomain binds weakly. Maximal monomer and cooperative binding of PBX to DNA are seen when the 15 amino acids C-terminal to its homeodomain are present. A highly conserved YPWMR/K pentapeptide motif N-terminal to the HOX homeodomain is required for the co-operative interaction with PBX. HOX-derived peptides containing the YPWM motif and its flanking amino acids increase the affinity of PBX for DNA. These peptides also prevent the formation of HOX-PBX heterodimers.

We have solved the solution structure of the free PBX homeodomain. The protein consists of three α-helices separated by two loops. The helices fold around a well-defined hydrophobic core. The structure of the extended PBX homeodomain bound to a 14 base pair DNA duplex is currently being solved (Figure 3.12).

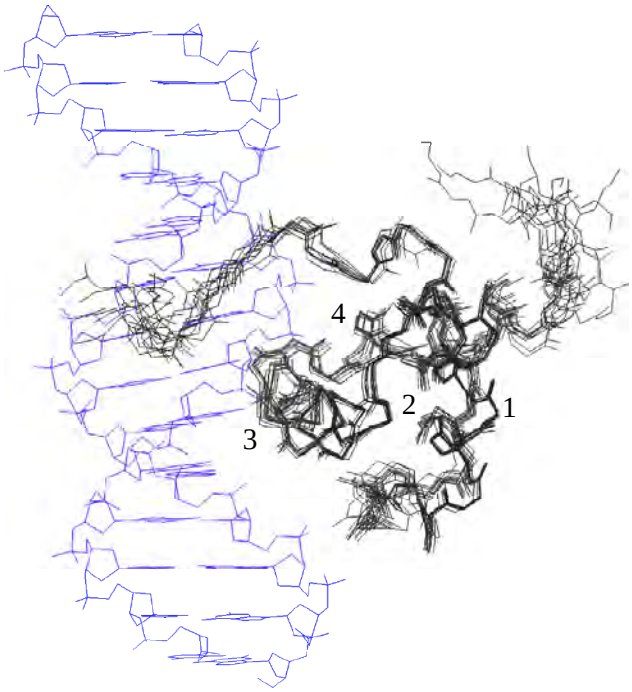


Figure 3.12. Structure of the PBX-DNA complex. The superposition of 10 lowest energy structures is shown. The positions of the four α-helices are labeled.

The extended PBX homeodomain includes the 15 amino acids immediately C-terminal to the homeodomain and demonstrates an increased affinity for DNA in comparison to the homeodomain alone. In the absence of DNA the C-terminal extension is unstructured; however, upon binding to DNA, it forms a fourth α-helix, which folds across the homeodomain, stabilizing the position of the third helix in the major groove of the DNA.

To resolve ambiguities in the assignment of NOESY spectra filtered experiments were carried out on a sample of doubly ^{13}C,^{15}N labeled PBX homeodomain bound to an unlabelled DNA duplex. Approximately 40 intermolecular NOEs were assigned in the NOESY spectrum of the complex with a ^{13}C filter applied in the ν2 dimension (Figure 3.13). This aided greatly in positioning the PBX homeodomain on the DNA duplex.

In order to complete the structure of the PBX homeodomain-DNA complex, experiments to determine the structure of the DNA are ongoing.

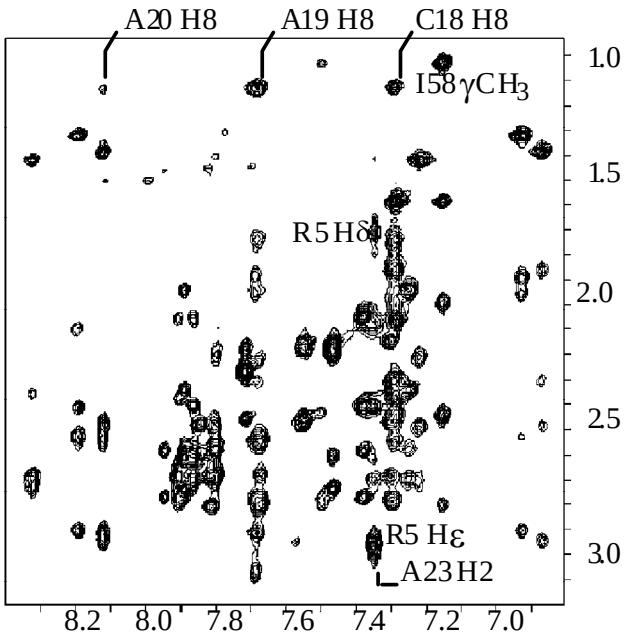


Figure 3.13. B: two-dimensional NOESY with ^{13}C-filtering in ν2. Sample concentration was 2 mM in 100% D_2O. Spectra were recorded at 750 MHz; with a NOESY mixing time of 150 ms. Total measuring time was 19.5 hours.

NMR Structural Investigations of the Breast Cancer Susceptibility Protein, BRCA1

P. Brzovic and R. E. Klevit

Biomolecular Structure Center, University of Washington

This project is aimed at gaining structural insights into the biological function and dysfunction of the breast cancer susceptibility protein known as BRCA1. Accumulating evidence implicates the involvement of BRCA1 in cellular response to DNA damage in addition to other postulated functions in homologous recombination and transcriptional regulation. A full understanding of the cellular role of BRCA1 requires a detailed understanding of both the function and the structure of the protein and its interacting partners.

BRCA1 is a large and complicated protein that is undoubtedly comprised of a multiplicity of functional domains. We are focusing initially on the N-terminal region of BRCA1, which contains a conserved pattern of cysteine and histidine residues found in the RING-finger family of proteins. The RING-finger is a Zn^{+2}-binding motif found in a wide variety of proteins that are diverse both in function and origin. We have found that the RING-finger motif in BRCA1 is part of a larger N-terminal structural domain. This region is the site of numerous mutations found in families genetically predisposed to breast and ovarian cancer. The N-terminal, RING-finger domain of BRCA1 interacts with the RING-finger region of a second protein, BARD1, to form a specific heterodimer complex. Thus, structural studies of the BRCA1/BARD1 heterodimer complex promise to yield important information regarding the structure of RING-finger proteins in general and the nature of the BRCA1 and BARD1 interaction in particular.

Based on domain mapping strategies, we designed bacterial expression systems to produce constructs of BRCA1 and BARD1 that each contain the first ~110 N-terminal residues of each protein. The size of the resulting heterodimeric complex is ~26kD, which presents certain challenges for NMR structure determination due primarily to problems associated with spectral overlap and signal sensitivity due to unfavorable relaxation properties of larger molecules. Since BRCA1/BARD1 is a heterodimer, we are able

to reduce the spectral overlap problem significantly through the use of heterodimers in which only one protein component is isotopically labeled. Thus, each half of the heterodimer may be studied independently, thus affording substantial spectral simplification. The price for this simplification, however, is that we must collect each experiment twice: once with the isotopic labels in the BRCA1 subunit and again with the BARD1 component labeled.

Previous work at PNNL and on our own 500-MHz instrument provided data sufficient for complete backbone and sidechain assignments for each subunit of the BRCA1/BARD1 heterodimer complex. During the past year we have taken advantage of the increased sensitivity and resolution afforded by the 600 and 800-MHz instruments at PNNL to collect several high quality NOESY-based spectra. Two four-dimensional ^{13}C-edited NOESY spectra acquired at 600 MHz, were critical for alleviating spectral overlap and allowing determination of the three-dimensional structure for each subunit in the heterodimer complex. ^{13}C-edited/filtered NOESY spectra with excellent signal-to-noise ratios collected at 800 MHz have been invaluable for determining intersubunit interactions and defining the dimerization interface for the complex. Without these data, determination of the three-dimensional solution structure of the BRCA1/BARD1 heterodimer would have been nearly impossible.

The past year has seen us progress to the final stages of structure determination for the BRCA1/BARD1 complex. Initially, we focused on solving the structures of the individual subunits. This process is nearly complete and preliminary structures for both the BRCA1 and BARD1 subunits have been calculated. The current model for the BRCA1 domain is shown in Figure 3.14. With data derived from the 13 C-edited/filtered NOESY spectra, we are in a position to move toward final determination of the entire heterodimer complex. Though not fully completed, current models have been invaluable for understanding other ongoing studies in our lab and that of our biological collaborators concerning the consequences of mutations in the RING-domain of BRCA1 that are cancer predisposing.

Figure 3.14. Current model of the three-dimensional structure for the BRCA1 subunit of the BRCA1/BARD1 heterodimer complex.

Publications

Brzovic, P. S., J. Meza, M. C. King, and R. E. Klevit. "The Cancer-Predisposing Mutation C61G Disrupts Homodimer Formation in the N-terminal BRCA1 RING-finger Domain." *J. Biol. Chem.*, **273**, 7795-7799 (1998).

Meza, J., P. Brzovic, M. C. King, and R. E. Klevit. "Characterization of the BRCA1/BARD1 Heterodimer." J. Biol. Chem., 274, 5659-5665 (1999).

Relaxation of ^{113}Cd Complexes with Aquatic Fulvic Acid

W. Otto,[(a)] S. Burton,
W. R. Carper,[(b)] and C. K. Larive[(a)]

(a) Department of Chemistry, University of Kansas
(b) Department of Chemistry, Wichita State University

Humic substances are a heterogeneous mixture of decomposition products of natural organic matter. They are the predominate source of dissolved organic carbon and are important environmentally because they play a major role in the transport mechanisms of various organic pollutants and metal ions. Fulvic acids are a subclass of humic substances, operationally defined by their solubility after extraction. In general, aquatic fulvic acids are a heterogeneous mixture of components with an average molecular weight

of 800 Daltons, relatively high oxygen content, and low nitrogen and sulfur content. The specific chemical characteristics of a fulvic acid are affected strongly by the environment from which it originates. On average there are 4 to 5 carboxylate moieties per fulvic acid, which are important determinants of their metal binding properties. We have examined the binding of the environmentally important toxic heavy metal ion Cd(II) to the Swannee River fulvic acid (SRFA) using ^{113}Cd NMR relaxation times. Prior studies have utilized ^{113}Cd NMR chemical shifts to study the complexation of Cd(II) with aquatic and soil fulvic acids. Binding constants were measured previously for several fulvic acids and the predominate binding site was determined to be the fulvic acid carboxylate moieties on the fulvic acids.

Relaxation times were measured for a series of model compounds to give a more complete understanding of the effect of different binding motifs on the relaxation rates of ^{113}Cd(II). The model compounds were chosen to represent a series of carboxylate binding sites. The first model compounds, acetate and salicylate, represent a single and a bidentate carboxylate-hydroxyl binding, respectively. A carboxylate site with a nitrogen donor is represented by nitrilotriacetate (NTA). A polydentate carboxylate binding site is modeled using 1,2,3,4 *cis-cis-cis-cis* cyclopentanetetra-carboxylic acid (CPTA).

To determine the T_1 relaxation times, a series of ^{113}Cd inversion recovery spectra were measured for each sample as a function of the relaxation delay, τ. The series of spectra obtained for a D_2O solution of 3.45-mM Cd(II) (95% enriched with ^{113}Cd) and 3 mg/mL SRFA are shown in the figure below. The intensity of the ^{113}Cd resonance was measured as a function of τ and fit to obtain the T_1 relaxation time using the equation $I_t=I_o(1-Bexp(-\tau/T_1)$. Our results are shown in Figure 3.15. The good fit of a single exponential function for such a heterogeneous sample suggests a single average T_1 resulting from fast exchange on the relaxation time scale. ^{113}Cd T_1 and T_2 relaxation times measured for D_2O solutions containing 3.45-mM ^{113}Cd(II) at pD 6.4 as a function of the SRFA concentration are reported in Table 3.1. These relaxation times are much shorter than the value obtained for free Cd(II) (~60s).

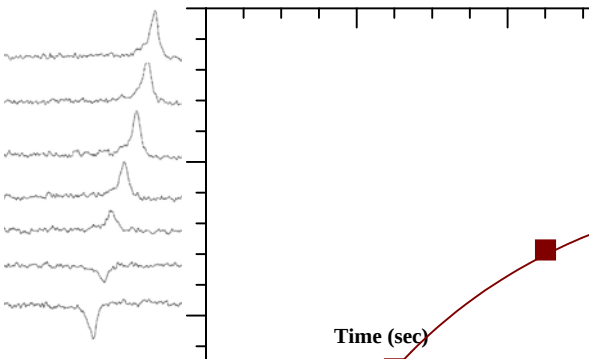


Figure 3.15. T_1 relaxation of SRFA-bound ^{113}Cd resonance.

Table 3.1. Dependence of the ^{113}Cd relaxation times on SFRA concentration.

SRFA mg/mL	δ_{obs}	T_1 (s)	T_2 (ms)
0.50	-0.84	1.8	9.4
0.60	-4.48	1.6	4.6
0.75	-5.93	1.5	3.9
1.0	-6.17	1.1	3.3
1.2	-8.52	0.66	2.3
1.7	-10.57	0.56	2.0
2.0	-15.76	0.33	1.3
3.0	-17.40	0.26	1.2
5.0	-19.06	0.22	0.84
10	-20.35	0.17	0.60

The relaxation rates ($R_1 = 1/T_1$) for samples with a constant ^{113}Cd(II) concentration are plotted as a function of the inverse of the SRFA concentration in the Figure 3.16. From this plot, there are two observable regimes. The first regime, observed at high fulvic acid concentration, shows a large decrease in relaxation rate until the ratio of 1/[SRFA] is 1.5. This regime corresponds to the region where most of the Cd(II) is bound. Since the SRFA is a heterogeneous mixture, the binding sites occupied in this concentration regime correspond to the strongest binding sites. A second regime in the plot of relaxation rates is observed at lower fulvic acid concentrations, where Cd(II) binding should occur at the weaker fulvic acid binding sites, or where on average there is now more than one Cd(II) bound.

Transverse relaxation rates ($R_2 = 1/T_2$) measured using the standard spin echo experiment were found to be equal within error to those determined from the

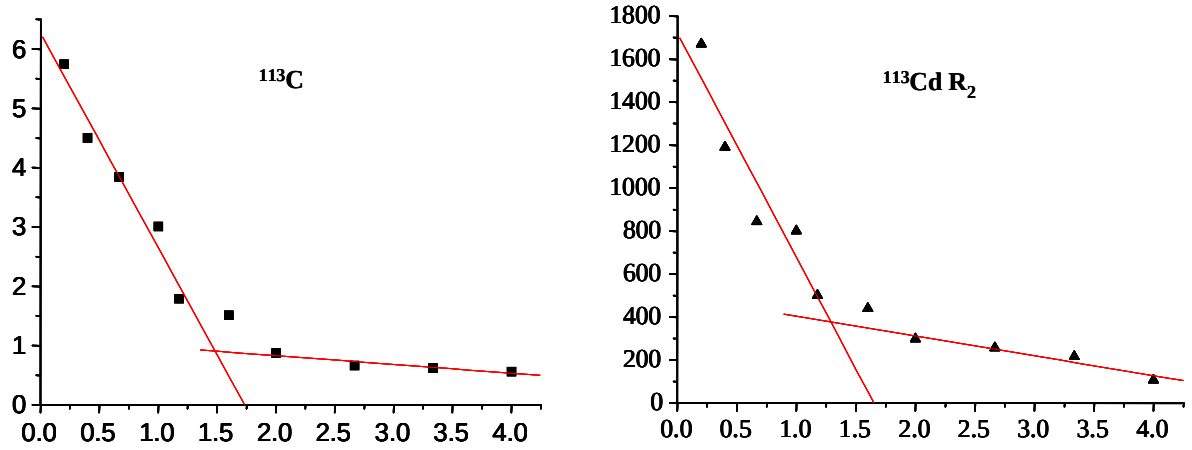


Figure 3.16. Inverse SFRA concentration.

resonance line widths. The R_2 relaxation rates are plotted as a function of the inverse SFRA concentration in Figure 3.16. Again, two regimes are noted with an intersection at a 1/[SRFA] of approximately 1.5. The strongest binding sites corresponding to the highest SRFA concentration produce the larger R_2 values, and the weaker binding sites correspond to the lower R_2 values.

Figure 3.16 shows that the strongest binding sites correspond to fastest relaxation rates. This would likely arise from the efficient relaxation mechanisms provided by the binding site such that the Cd(II) is hindered sterically in rotation. For example, a multidentate binding site in which the fulvic acid effectively surrounds the Cd(II) should provide efficient relaxation.

A series of model compounds also was studied to probe the dependence of the type of Cd(II) binding site on the ^{113}Cd relaxation times, shown in Table 3 2. These solutions were prepared so that the primary species in solution would be the 1:1 ligand-Cd(II) complex. The chemical shifts listed in Table 3.2 indicate that the carboxylate binding sites are all near -20 ppm. The effect of nitrogen is observed in the chemical shift of the Cd-NTA complex, 7.6 ppm. The Cd-CPTA complex has a much shorter T_1 than the other model compounds due to the four carboxylate donors that are present in the all *cis* CPTA complex. This binding site then restricts the mobility of the Cd(II) causing a more efficient mechanism of relaxation. The relaxation times observed for the other three model compounds are all at least an order of magnitude higher.

From our study of ^{113}Cd relaxation times in SRFA complexes it is apparent that two types of Cd(II) binding sites exist. The first is a stronger binding site that provides an efficient relaxation mechanism for Cd(II). This binding site is best represented by the Cd-CPTA model. Such a complex could be considered to be an inner sphere complex in which these stronger binding sites would have a much higher affinity for a nucleus such as Cd(II) rather than a cation such as Ca(II). This type of binding mechanism would explain some of the differences determined by Otto et al. (in press) in their measurement of Ca(II) competition. In this paper, we observed a small number of fulvic acid binding sites that preferentially bind Cd(II). These binding sites would correspond to the stronger binding sites observed here based on the Cd:FA ratios examined. The second type of binding site could then be a binding to the exterior of the fulvic acid, where it is more superficially bound to carboxylate functionalities, corresponding to an outer sphere complex.

Table 3.2. Parameters Determined for 1:1 Ligand–^{113}Cd Complexes

Model Compound	δ_{obs}	T_1 (s)	T_2 (ms)
Acetate	-20.5	30.4	24
Salicylate	-13.5	19.5	11
NTA	7.6	11.5	1.9
CPTA	-19.0	1.8	10

References

Otto, W. H., W. R. Carper, and C. K. Larive. *Environ. Sci. Technol.*, in press (2001).

Larive, C. K., A. Rogers, M. Morton, and W. R. Carper. *Environ. Sci. Technol.*, **30**(9), 2828-2831 (1996).

Li, J., E. M. Perdue, and L. T. Gelbaum. *Environ. Sci. Technol.*, **32**(4), 483-487 (1998).

Kowalewski, J., G. C. Levy, L. F. Johnson, and L. Palmer. *J. Magn. Reson.*, **26**, 533-536 (1977).

Sass, M. and D. Ziessow. *J. Magn. Reson.*, **25**, 263-276 (1977).

Use of Paramagnetic NMR Information for Resonance Assignment and Refinement of Metalloprotein Structures

S. L. Alam

Department of Biochemistry, Institute of Human Genetics, University of Utah

The foresight for using paramagnetic contributions to NMR observables as structural restraints dates back to when the first NMR structures were being determined. This utility has only been moderately pursued. These potential structural constraints arise from the many effects that paramagnetism has on NMR parameters. Of course, for these to be useful, nearly complete resonance assignments are required. Described below are our efforts to use standard resonance assignment strategies in conjunction with other characteristics of unassigned signals to aid in their identity.

Our group plans to use the monomeric Hb from *Glycera dibranchiata* and sperm whale myoglobin to validate this utility. The structures of both proteins have been completed, and several assignments are available for the ferric-CN forms of both proteins. Since this form is believed to be isostructural with the ferrous-CO ligation state, they can be used to substantiate the use of both relaxation and chemical shift data as structural restraints during refinement. The ferric unligated form of each protein also will be assigned because they have significantly different paramagnetic properties compared to the CN-forms (five unpaired electrons versus one and differences in contributions to NMR parameters) and will help assess this method at differing levels of paramagnetism.

Our approach is twofold. First, we will obtain as many unambiguous NMR assignments as possible using standard triple-resonance and edited experiments. Many assignments will be missing for various reasons including the heightened nuclear relaxation due to association near a paramagnetic center. Figure 3.17 gives an indication of the efficacy of standard methods for SWMB in the ferric form (no column in the ferric slope data indicates no assignment). There are approximately 15 resonances that are in a simple HSQC that give partial or no triple-resonance correlations. It is this subgroup that may be assigned with additional information.

From calculations of residual dipolar couplings (Figures 3.17 and 3.18) for each amide, and from the paramagnetic contributions to chemical shifts (Figure 3.18), the magnetic susceptibility tensor can be established

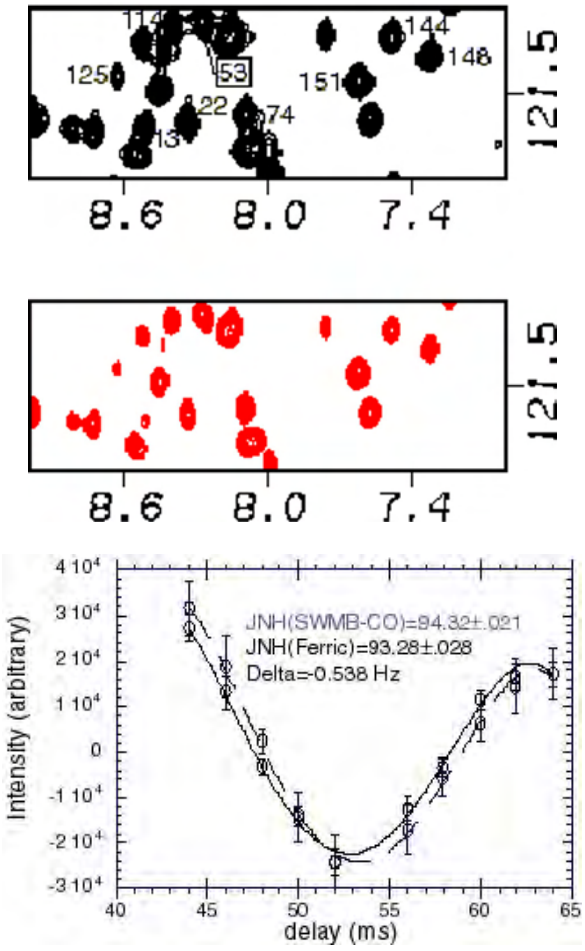


Figure 3.17. Measurement of residual dipolar couplings in SWMB. Differences in fit J_{NH} coupling constants at different field strengths (750 MHz and 500 MHz). HSQC-J experiments with different coupling delay times and the fitting of JNH coupling constants.

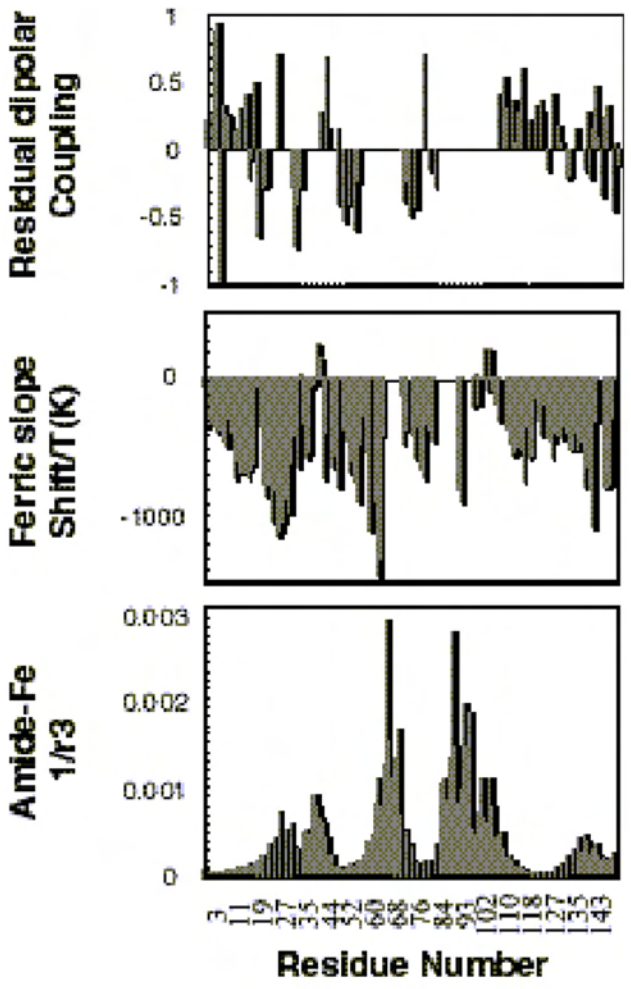


Figure 3.18. Amide proton chemical shift temperature dependencies and residual dipolar couplings for ferric SWMB. A relative magnitude for the dipolar.

and cross-validated by the two techniques for each protein. The measurement of residual dipolar couplings relies on the partial alignment of the molecule of interest with the magnetic field. Alignment is dependent on the molecule's properties and field strength of the NMR magnet (higher fields result in more alignment). The alignment tensors will be established from the subset of unambiguous assignments (~90%). This property will be useful, in combination with the current structures of each protein, to test the classical theories that describe paramagnetic contributions to chemical shifts. In effect, we foresee the ability to predict the paramagnetic contributions to chemical shifts for the unassigned, but observable, subset of NMR resonance's.

Once restraints are generated, the ferric unligated and ferric-CN structures of these proteins will be refined. This will help validate use of these restraints in protein structures where internal NOE-restraints are sparse due to the quickened relaxation of the nuclei near the iron of the heme group. In addition, the alignments tensor describes many electronic properties of the active site metal.

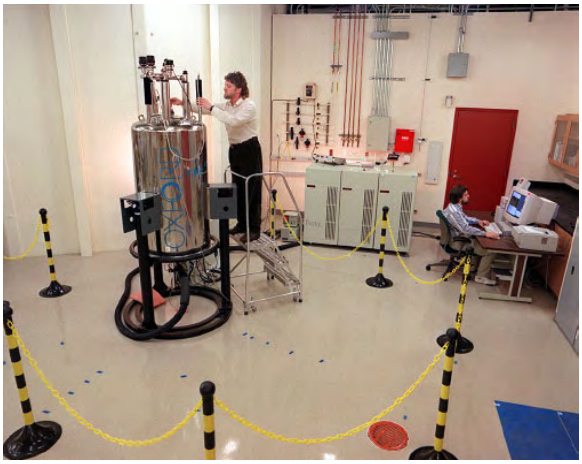
Magnetic Resonance Research Capabilities

Varian INOVA 800

The Varian 800 is an INOVA-based spectrometer and an Oxford 18.8-T magnet with a 63-mm room temperature bore. This system is capable of high-resolution-liquid and solid-state NMR. There are four RF channels with waveform generators and pulsed-field gradients. The wideline ADCs run at 5 MHz and the narrow ADCs have a maximum rate of 500 kHz. We currently have a 5-mm HCN probe with Z-gradient for liquids and a single tuned, 5-mm MAS probe. Under construction are a 5-mm orthogonal powder probe optimized for low gamma nuclides, a 5-mm HX MAS probe (12-kHz spinning), and a 4-mm HXY MAS probe (25-kHz spinning).

Varian INOVA 750

The Varian 750 is an INOVA-based spectrometer and an Oxford 17.6-T magnet with a 51-mm room temperature bore. This system is capable of high-resolution-liquid, solid-state NMR and μ-imaging. There are four RF channels with waveform generators and pulsed field gradients. The narrow ADCs have a maximum rate of 500 kHz. We currently have two 5-mm HCN probes (Z-gradient), an 8-mm HCN probe (Z-gradient), a 5-mm HCP probe (Z-gradient), a 5-mm HX MAS probe (10 kHz spinning), and an μ-imaging probe.

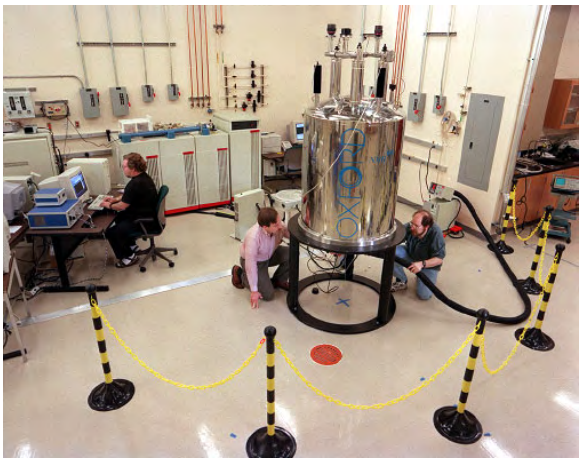


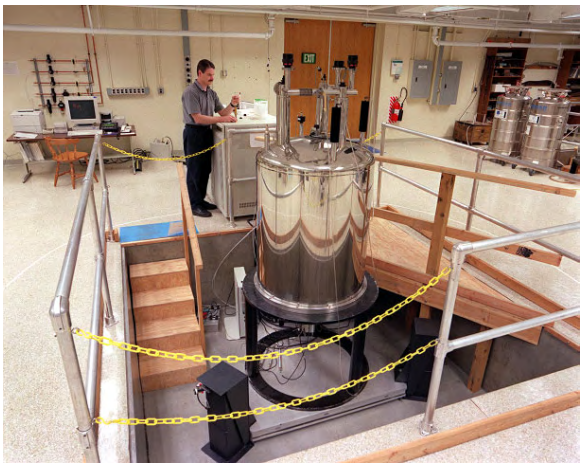
Varian INOVA 600

The Varian 600 is an INOVA-based spectrometer and an Oxford 14.1-T magnet with a 51-mm room temperature bore. This system is capable of high-resolution-liquid NMR. There are four RF channels with waveform generators and pulsed-field gradients. The narrow ADCs have a maximum rate of 500 kHz. We currently have a 5-mm HCN probe with Z-gradient and a 5-mm HX probe.

Varian Unity 600

Varian 600 is a Unity-based spectrometer and an Oxford 14.1-T magnet with a 51-mm room temperature bore. This system is capable of high-resolution-liquid NMR. There are three RF channels with waveform generators and pulsed-field gradients. The narrow ADCs have a maximum rate of 500 kHz. We currently have a 5-mm HCN probe with Z-gradient and a 5-mm HX probe.





Varian/Chemagnetics Infinity 500

The Varian 500 is a Unity-based spectrometer and an Oxford 11.7-T magnet with an 89-mm room temperature bore. This system is capable of solid-state NMR, μ-imaging, and small animal MRI. There are three RF channels with waveform generators. The wideline ADCs run at 5 MHz. We currently have a 5-mm orthogonal powder probe, 5-mm HXY MAS probe (15-kHz spinning), a 7-mm HX MAS probe (10-kHz spinning), an HX single crystal probe, a ^{1}H CRAMPS probe, a micro-coil imaging probe, and a 40-mm imaging probe. Under construction are a combined μ-imaging and optical microscope probe and a static HX low temperature probe (2 - 300 K).

Varian/Chemagnetics Infinity 500

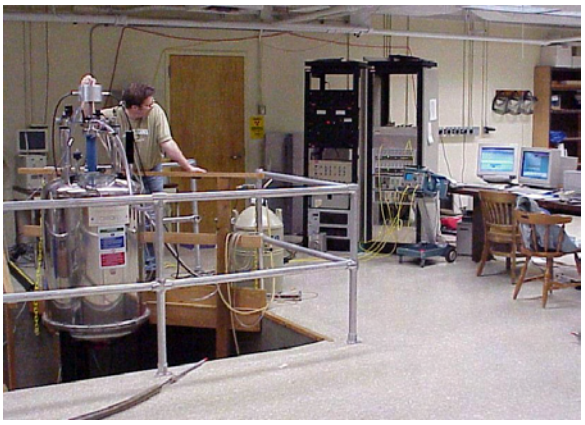
The Chemagnetics 500 is an Infinity-based spectrometer and an Oxford 11.7-T magnet with a 51-mm room temperature bore. This system is capable of high-resolution-liquid and solid-state NMR. There are three RF channels. We currently have a 5-mm HCN probe, a 10-mm HX probe, and a 5-mm HX MAS probe.

Varian Unity^{+} 500

The Varian 500 is a Unity-based spectrometer and an Oxford 11.7-T magnet with a 51-mm room temperature bore. This system is capable of high-resolution liquids NMR. There are three RF channels with waveform generators and pulsed-field gradients. We currently have a 5-mm HCN probe with Z-gradient and a 10-mm HX probe.

Homebuilt/Infinity Plus 400

The homebuilt 400-MHz NMR spectrometer is based on Varian Infinity Plus console and receiver section, and an Oxford 9.4-T magnet with an 89-mm room temperature bore. This system is capable of solid-state NMR. There are two RF channels and 2-MHz ADCs. We currently perform low temperature (5-25 K) experiments using an Oxford Crystal and homebuilt probes.



Varian/Chemagnetics Infinity 300

The Chemagnetics 300 is an Infinity-based spectrometer and an Oxford 7.04-T magnet with 89-mm room temperature bore. This system is capable of high-resolution-liquid and solid-state NMR. There are three RF channels. We currently have a 5-mm HX probe, a 10-mm HX probe, a 7.5-mm HX MAS probe, and a 5-mm orthogonal powder probe.

Varian Unity^{+} 300

The Varian 300 is a Unity-based spectrometer and an Oxford 7.04-T magnet with an 89-mm room temperature bore. This system is capable of solid-state NMR, μ-imaging, and small animal MRI. There are two RF channels with wideline ADCs running at 5 MHz. We currently have a 7-mm HX MAS probe (10-kHz spinning), an HX single crystal probe, a ^{1}H CRAMPS probe, a single-tuned DOR probe, an HX 5-mm, low-temperature MAS probe (35-300K, 12-kHz spinning), a 7-mm HX high temperature probe (-100 - 500°C, 7-kHz spinning), a microscopy probe, and a 40-mm imaging probe.

Bruker Pulsed EPR/ENDOR/ELDOR Spectrometer

The EPR/ENDOR work focuses on the development and application of pulsed, two-dimensional methods for electronic structure and distance determinations in biological macromolecules and materials. The spectrometer consists of a Bruker 380E Pulsed EPR Spectrometer with pulsed ENDOR and has been modified to include pulsed ELDOR capabilities.

Supercritical Fluids Accessory

A potential limitation to liquid-state NMR experiments is when the rotational correlation time of the molecule of interest becomes so long that significant line broadening occurs which in turn limits resolution and sensitivity. Hence, it is of interest to develop solvent systems that can extend the limits of rotational correlation times. Supercritical solvents represent such a system. These solvents also have the advantage that they process novel solvent chemistry and catalytic properties. To support DOE sponsored activities, we have developed a system to allow characterization of solutes by NMR and EPR under supercritical conditions. This capability is being applied to develop an understanding of the field dependence of dynamic nuclear polarization experiments as they relate to signal enhancement methods for characterization of homogeneous and heterogeneous catalysts. The EMSL system is based on fused silica capillaries, and we employ two pumps. The first pump is a low-pressure pump (ISCO) capable of 10,000 psi (~0.67 kbar) and an operating volume of 100 mL. The second pump is a high-pressure pump (High Pressure Equipment Corporation) capable of 60,000 psi (~4 kbar) and an operating volume of 10 mL.

Magnetic Resonance Users and Collaborators

Users External to PNNL

Steven L. Alam
University of Utah
Dynamics in Myoglobins and Hemoglobins

Cheryl H. Arrowsmith
University of Toronto
Structural Genomics: Deducing Protein Function From Three Dimensional Structure

Paul Bash
Northwestern University
Genome Directed Structural Biology

William F. Bleam
University of Wisconsin-Madison
2-Dimensional NOESY Studies of Metal-Binding sites in Humic

Jeremy L. Bretherton
University of British Columbia
Very HF/High Resolution Solid State NMR Studies Quadrupole

David L. Bryce
Dalhousie University
Characterization of Boron Nuclear Magnetic Shielding Tensors in Some Model Compounds

William Carper
University of Arkansas
Fulvic Acid Complexes

Anna Cavinato
Eastern Oregon University
Remote 600-MHz NMR Experiment Over the Internet for Undergraduate Students

Robert Clubb
University of California, Los Angeles
The Solution Structures of the Dead Ringer and Mu Repressors Protein-DNA-Complexes

Myriam Cotten
University of Washington
Conformational Studies of Human Salivary Histatin Bound on Hydroxyapatite by Solid State NMR

Mamadou S. Diallo
Howard University/Cal Tech
Characterization of Chelsea Humic Acid by 1-D/2-D Solution ^{13}C and 1H NMR Spectroscopy

Gary P. Drobny
University of Washington
Conformational Studies of Human Salivary Histatin Bound On Hydroxyapatite by Solid State NMR

Bruce R. Dunlap
University of South Carolina
Structural Determination of Sperm Binding Protein

Kathryn R. Ely
The Burnham Institute
Molecular Structure of BAG Protein
Molecular Structure of a New Protein Motif: PR

Jeremy N.S. Evans
Washington State University
NMR Structure Determination of EPSP Synthase

Theresa W.-M. Fan
University of California-Davis
Plant Rhizosphere Effects on Metal Mobilization and Transport

Juli Feigon
University of California, Los Angeles
Structure of Nucleolin RBD12 in Complex With Its RNA Target

Joanna L. Feltham
University of Massachusetts
The Apo-Form of the Substrate Binding Domain of DnaK, an Hsp70 Chaperone

John J. Fitzgerald
South Dakota State University
Solid State ^{93}Nb, ^{25}Mg, and ^{67}Zn NMR…Solid-Solutions and Related Relaxor Ferroelectrics.
Solid State $^{47,49}Ti$ NMR Studies of the Surface and Catalytic Chemistry of Titanium Oxides and Titanosilicates

Colin A. Fyfe
University of British Columbia
Very HF/High Resolution Solid State NMR Studies Quadrupole
Connectivities in the Solid State

Kalle Gehring
McGill University
Structural Studies of a Ternary PBX/DNA/Peptide Complex
Structural Changes in the PBX Homeodomain Mediated by YPWM Containing Peptides

John V. Hanna
CSIRO
Very HF (Static) Broadline NMR Detection of Quadrupolar Nuclei

James F. Hinton
University of Arkansas
Very High Magnetic Field Effects on Ion Transport Through Biological Channels

Stefan Hunger
University of Delaware
Mechanisms of Phosphorus Stabilization in the Soil Environment

Eun Gyeong Kim
University of Wisconsin-Madison
2-Dimensional NOESY Studies of Metal-Binding Sites in Humic

Rachel E. Klevit
Washington State University
NMR Structural Investigations of BRCA1

Cynthia K. Larive
University of Arkansas
Fulvic Acid Complexes

Rudy L. Luck
Michigan Technology University
Variable Temperature ^{13}C MAS NMR Studies on MoN (OtBu)3

Gaetano T. Montelione
Rutgers University
Structural Genomics of Strongly Conserved Gene Families
Structural Genomics of Eukaryotic Model Organisms

Wendy Shaw
University of Washington
^{1}T and ^{1}Tro Measurements of Carbons

Louis A. Silks
Los Alamos National Laboratory
DNA and Protein Molecules

Tara Sprules
Structural Changes in the PBX Homeodomain Mediated by YPWM Containing Peptides

Santhana Velupillai
Los Alamos National Laboratory
High-Throughput Structural Proteomics

Roderick Wasylishen
Dalhousie University
Characterization of Boron Nuclear Magnetic Shielding Tensors in Some Model Compounds

Marc S. Wold
University of Iowa
Human Replication Protein A (hsRPA)

John D. Zimbrick
Purdue University
Spatial Properties of Clustered Free Radicals Produced in DNA and Biodosimeters by Ionizing Radiation

Users (within PNNL)

Eric J. Ackerman
MRI Imaging and NMR/Confocoal Imaging of Xenopus Germ Cells and Early Embryos

Richard A. Corley
In Vivo Particulate Matter NMR Imaging/Quantitation in Rat Lung
In Vitro Particulate Matter NMR Imaging/Quantitation in Alveolar Macrophases

Timothy Hubler
Determination of Structure in Solution Phase Binding of Lantharide-Tetraaklyldiamide Complexes

Liang Liang
NMR Analysis of Composite Materials

Greg Lumetta
Determination of Structure in the Solution Phase Binding of Lantharide-Tetraaklydiamide Complexes

Bruce McNamara
Determination of Structure in the Solution Phase Binding of Lantharide-Tetraaklyldiamide Complexes

Brian Rapko
Determination of Structure in the Solution Phase Binding of Lantharide-Tetraaklydiamide Complexes

Charles Timchalk
In Vitro Particulate Matter NMR Imaging/Quantitation in Alveolar Macrophases
In Vivo Particulate Matter NMR Imaging/Quantitation in Rat Lung

Li-Qiong Wang
High Resolution NMR Investigation of the Nano-Materials
Solid State NMR Studies of Nano-Materials

Brian D. Wood
Use of NMR Microscopy to Examine the Sub-Pore-scale Structure of Porous Media, and Porous Media, and Porous Media with Multiple Liquid Phases

Collaborators

Eric J. Ackerman
PNNL

4. Mass Spectrometry Research Staff Reports

Proteomic Analysis of Deinococcus Radiodurans R1

M. S. Lipton, L. Paša-Tolić,
G. A. Anderson, H. R. Udseth,
T. D. Veenstra, and R. D. Smith

Supported by the Office of Biological and Environmental Research at DOE

With the advent of the near completion of the human genome and the availability of many complete microbial genomes, a new paradigm has emerged in the biological sciences that focuses on the characterization of the gene products: the messenger RNAs and proteins. This shift involves understanding the much greater complexity associated with the dynamic nature of the gene products in the context of whole biological systems. Proteomics, the study of the array proteins expressed by a particular cell population at a given time or under a specific set of environmental or internal conditions, is seen as an essential approach to understanding the operation of biological systems.

A pilot project at PNNL supported by the Office of Biological and Environmental Research has applied advanced instrumentation to study the proteome of the microorganism *Deinococcus radiodurans*. The most significant defining characteristic of *D. radiodurans* is its ability to resist the lethal effects of DNA damaging agents such as ionizing and UV radiation, hydrogen peroxide, and desiccation. The bacterium has been shown to survive ionizing radiation exposures a thousand times greater than many other organisms, apparently because of unusually efficient DNA repair mechanisms. The instrumentation is centered around a unique high magnetic field Fourier transform ion cyclotron resonance mass spectrometer (FTICR-MS), available in the EMSL facility, to efficiently identify each of the proteins in the organism.

The identification of the proteins is based on the correlation of sequence related data obtained from data-dependent tandem mass spectrometry (MS/MS) using conventional mass spectrometric technologies, liquid-chormatography (LC) elution time, and the observation of the intact peptide by FTICR-MS. Peptides are initially characterized by data-dependent MS/MS where the peptide, once eluted from the LC into the mass spectrometer is observed, fragmented, and the MS/MS spectra obtained. The results from the MS/MS spectra are searched against SEQUEST or MASCOT using the appropriate database. Once a confident identification has been made, the exact mass is calculated and used to search the FTICR data for a peptide match. Once the peptide has been identified in both sets of data, the peptide is matched back to the parent protein and the protein identification is verified.

An illustration of the identification of this process is shown in Figure 4.1 for a peptide from the protein elongation factor Tu in *D. radiodurans*. The peptide, created by a global tryptic digest of a *D. radiodurans* lysate, was analyzed by LC-MS on both a Finnigan LCQ ion-trap mass spectrometer and an 11.5-tesla FTICR-MS. The ion trap equipped with data-dependent MS/MS capabilities, yielded a spectra with many y and b fragmentation ions (spectrum A). The information contained in the spectra, when searched by SEQUEST, identified the peptide as part of the protein elongation factor Tu. Concurrently, the same peptide mixture was analyzed by LC-FTICR using an

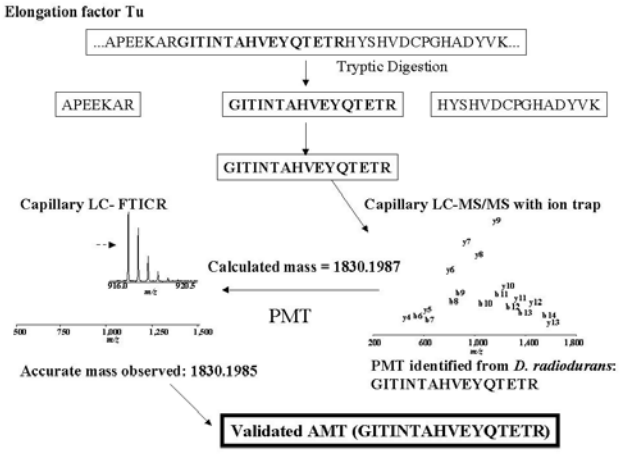


Figure 4.1. An illustration of the identification of this process is shown for a peptide from the protein elongation factor Tu in *D. radiodurans.* Identification of the protein elongation factor Tu by the combination of tandem mass spectrometry data and high resolution mass measurements. The peptide in a central region of the protein was analyzed and the MS/MS fragmentation data obtained (spectrum A). The resulting data was used to search SEQUEST revealing a possible identification. The accurate mass was calculated and used along with elution data to search the FTICR data (spectrum B). Once a match was observed, the peptide and protein were considered to be identified.

identical liquid chromatographic system. A mass of 1830.1987 was calculated for the peptide from the sequence predicted by SEQUEST and that mass was used as an identifier when searching the FTICR data. A peptide with the same mass (within 1 ppm mass accuracy) and eluting at the same time (within 20%) was found in the LC-FTICR chromatogram (spectrum B). Since the peptide was found in both analyses with the same mass and elution time, the identity of the peptide and therefore the presence of the protein in the sample was confidently confirmed.

The pilot-study research team identified about 40% of the all the possible gene products that might be expressed by *D. radiodurans*, a significantly higher fraction than actually detected for any other organism to date. They also demonstrated the capability to measure the expression levels for all proteins in a single experiment, and are beginning to apply this approach to determine the expression patterns of proteins for the organism after exposure to ionizing radiation. The results from this research will have important implications in both the understanding of DNA repair processes as well as for the development of bioremediation agents for use in contaminated areas. In a broader context, this new capability to study the array of all proteins expressed by a cell under a given condition, and to precisely measure their changes in expression, will propel efforts aimed at developing an understanding of complex biological systems, and ultimately will enable us to predict the cellular response to an environmental perturbation.

Proteomics Analysis

M. Powers,[(a)] N. Tolic,
G. Kiebel,[(b)] and D. Clark[(b)]

Supported by the Office of Biological and Environmental Research at DOE

(a) PNNL Applied Information Systems Group
(b) PNNL Electronic Systems Group

Proteomics research at PNNL centers on the study of the complete protein complement expressed by living cells using mass spectrometry. This includes investigating the differential expression of proteins under varying stress conditions with the goal of better understanding the function of proteins in biological processes. The basic approach is to extract the protein complement, enzymatically digest the proteins, separate the mixture with high-pressure liquid chromatography (HPLC), and detect the peptides with mass spectrometry. This is a multi-year effort, and it involves many thousands of mass spectrometer runs on many thousands of biological samples. Only computer automated data analysis makes this approach feasible.

Experiments were performed on two types of mass spectrometers: quadrupole ion trap mass spectrom eters (Finnegan LCQ) and FTICR spectrometers (custom FTICR spectrometers designed and built at PNNL). The LCQ data were processed through SEQUEST software using protein databases obtained from the Institute for Genomic Research (www.tigr.org). Peaks that scored 2 or higher were accepted as potential mass tags (PMT). To increase the confidence of the assignments in the PMT database the experiments were repeated on the FTICR instruments. The FTICR spectra have a mass measurement accuracy (MMA) of 1 ppm, and PMTs found in these spectra are promoted to accurate mass tags (AMT). To achieve high MMA, internal calibration was implemented through a mass lock function. This function utilized peptide masses from housekeeping proteins distributed throughout the run that were determined with high confidence level. Masses from the FTICR scans were corrected based on presence of the locker masses. Additional concept of the unique mass class (UMC) and supporting functions were developed to extend the correctional influence of the lockers on as many scans as possible and completely automate peak identification in the FTICR spectra. This procedure is shown in Figure 4.2

Expression studies are conducted by using isotopic tags, either isotopically depleted growth media or isotopically coded affinity tags (ICAT), to mark proteins derived from systems under stress that are then mixed with proteins from a standard system. The results are presented in a fully interactive two-dimensional display with point and click features connecting directly to the data sets. The interpretation of the datasets can be enhanced by importing Codon Adaptive Index (CAI) data for each gene. Also, DNA start location, DNA stop location, and direction, can be imported into the database and the associated gap and operons calculated.

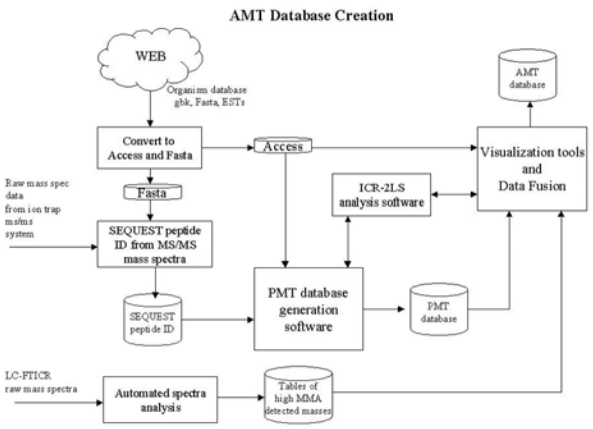


Figure 4.2. Illustrates the data flow required for the generation of AMTs. Development of the AMTs are a critical part of the data analysis, and this requires fusing data from different types of mass spectrometers with different types of analysis and different data quality. Our DMS efforts are designed to fully automate this entire process.

The complete list of functions added to the two-dimensional display application in direct support of the OBER project follows:

- Unique Mass Classes concept and visualization
- Mass Locking
- MS, MS/MS search of protein database (using ICR-2LS)
- PMT database search and analysis
- ICAT PMT search and analysis
- $^{14}N/^{15}N$ pair's analysis
- ORF and AMT databases visualization as two-dimensional gels
- Mass distribution visualization and analysis
- Data comparisons and results statistics.

A new system has been put into place to organize and automate management of the large amounts of data created and used by proteomics research in the EMSL. This Data Management System (DMS) acquires, stores, and tracks mass spectrometer raw data and analysis results and works with EMSL infrastructure systems to handle gigabytes of new information produced daily and terabytes of accumulated information. DMS consists of a tracking database, data file storage, several supporting programs (called "managers"), and a web-based end user interface. DMS is implemented on a primary server computer and several auxiliary server computers, as shown in Figure 4.3.

DMS also manages information that is associated with data sets. Experimental records keep information about the biological samples that are analyzed by the mass spectrometer runs. This includes the name of the organism, the type of media on which it was grown, the protocols used to prepare the sample, the stresses applied to it, and so forth. Each data set is directly associated with exactly one experiment, but an experiment may be associated with many data sets. DMS also keeps campaign records. A campaign is a particular line of investigation that involves one or more experiments. An experiment is can be associated with many experiments. The tracking information for the data sets and their associated experiments and campaigns can be searched and used to find a particular data set of interest, or related groups of data sets.

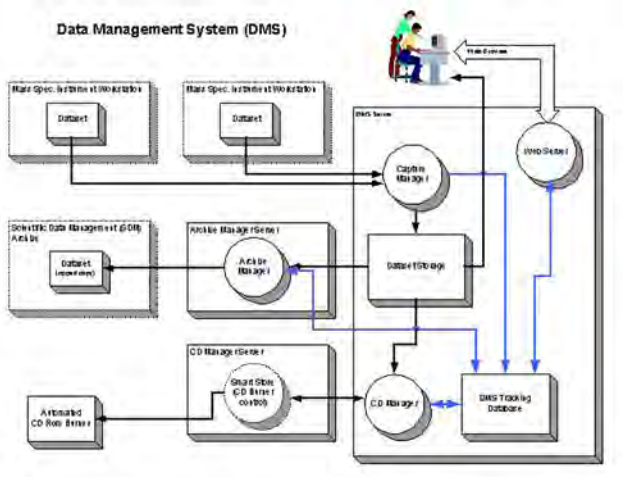


Figure 4.3. Block Diagram of DMS.

The present functions of DMS are:

- Capture raw data from MS instrument workstations
- Store data in "well known" and accessible locations
- Collect and maintain tracking information about data
- Burn CD-ROM copies of data for backup and offsite distribution
- Copy the data and its metadata to the EMSL Scientific Data Management Archive
- Collect results of various analyses performed on data
- Manage storage space.

End users are comfortable with the Web-based interface. It works with all their workstations and with current versions of both Netscape and Microsoft Explorer. Distribution, installation, and upgrades of the user interface software are fully supported. This allows the developers to encourage feedback from the end users, respond to it, and to place the improvements back in their hands quickly.

Rapid and Ultra-Sensitive Analysis of Global Protein Expression in *Shewanella Putrefaciens*

M. S. Lipton, K. K. Peden,
P. K. Jensen, G. A. Anderson,
Y. Gorby, M. F. Romine,
T. D. Veenstra, and R. D. Smith

Supported by the Natural and Accelerated Bioremediation Research (NABIR) Program in the Office of Science at DOE

Shewanella putrefaciens is a key organism in bioremediation, due to its ability to reduce and precipitate a diverse range of heavy metals and radionuclides. Developing the most efficient organisms for bioremediation requires an understanding of the key cellular pathways and proteins involved. An improved understanding of these organisms can likely be obtained from study of their proteome (the entire protein complement of the cell expressed under a given set of conditions). A single genome can exhibit many different proteomes depending on stage in cell cycle, cell differentiation, response to environmental conditions (nutrients, temperature, stress, etc.), or the manifestation of disease states. This requires faster, more sensitive and quantitative capabilities for the characterization of cellular constituents.

We are currently developing technologies that integrate protein separation and digestion methods with advanced mass spectrometric methods. Capillary isoelectric focusing on-line with Fourier transform ion cyclotron resonance mass spectrometry (CIEF-FTICR) provides a powerful tool to study the changes in expression for hundreds to potentially thousands of proteins simultaneously. First one must be able to identify the proteins of interest. To accomplish this, the proteins are digested with a protease and the resulting peptides are first analyzed by capillary LC-MS/MS on an LCQ (ion trap) mass spectrometer to obtain accurate mass tags (AMT), or "biomarkers," for each protein. The use of MS/MS provides additional sequence information, that, when combined with the accurate mass (AM) of the parent peptide, can enable protein identification. Therefore, the same, or similar, proteomes are then analyzed using LC-FTICR MS. The AMTs obtained in the first step are then compared to the AMs obtained from the FTICR. Due to the high MMA provided by FTICR, these proteolytic fragments can often be used as AMTs for identification of many proteins, which can then be used in all subsequent studies of the same organism.

Examination of changes in the proteome of *S. putrefaciens* initially concentrated on the soluble proteins located in the periplasmic region, mainly due to the significant proportion of metal reductase activity and related c-type cytochromes that have been located in this region. Figure 4.4 shows the LC-FTICR and LC-MS/MS analyses with one of the peptides found in both experiments, which identified a peptide that belongs to fumarate reductase flavoprotein subunit precursor. Initial studies show that *S. putrefaciens* grown under increasingly anaerobic conditions start producing more cytochrome-type proteins in order to complete its metabolic pathways. Further analysis of the samples will be done using CIEF-FTICR, which allows us to obtain high MMA for the intact proteins, as well as information on their isoelectric points. These combined technologies will

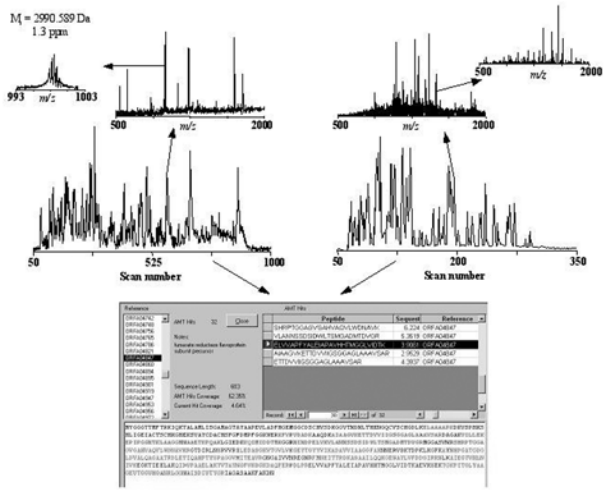


Figure 4.4. Capillary LC-FTICR (left) and capillary LC-IT-MS/MS (right) analysis of tryptic digest of *S. putrefaciens*. Insets for FTICR data show the scan for a selected peptide and the high-resolution spectrum for the 3+ charge state. Insets for the LCQ (IT) data show the scan containing the same peptide found in the FTICR data, and the MS/MS spectrum of the 2+ charge state. This peptide allows identification of the protein as fumarate reductase flavoprotein subunit precursor, along with many other proteins in the LC-IT-MS/MS analysis, and enables their use as AMTs for further studies where MS/MS will not be required.

enable ultrasensitive proteome–wide expression profiling to evaluate changes in the complete proteome of the iron reducing bacterium *S. putrefaciens* strain MR-1 induced by switching from aerobic to anaerobic respiration with heavy metals and radionuclides.

Quantitative Analysis of Bacterial and Mammalian Proteomes Using a Combination of Cysteine Affinity Tags and 15N Metabolic Labeling

T. P. Conrads, K. Alving, W. Chrisler,[(a)] T. D. Veenstra, G. A. Anderson, M. E. Belov, B. Thrall,[(a)] and R. D. Smith

(a) PNNL Molecular Toxicology Group

Proteomics, the analysis of the entire complement of proteins expressed by a cell, tissue, or organism, is rapidly becoming a necessary complement to genomics. While genomics provides the blueprint describing the proteins an organism may express, the goal of proteomics is to actually describe the proteins being expressed at a given time and under a specific set of conditions. A necessary component of effective proteomics is quantification of protein abundances. It is also desirable to measure changes in protein abundances in a high throughput manner so that the effect of many "perturbations" upon, or changes to, a cell type, tissue, or organ, can be determined in a reasonable time period. The predominant method of measuring changes in protein expression levels using two-dimensional polyacrylamide gel electrophoresis (O'Farrell 1975) proteomic technology is to compare protein spot intensities (Newsholme et al. 2000). This strategy fails to detect subtle changes in protein expression levels and lower abundance proteins, and the precision of the measurements is much less than desired. It is also difficult to compare results between different labs despite increased reproducibility through standardization and automation (Celis et al. 1999). Despite more than a decade of significant effort, this technique has not been shown to be amenable to high throughput automation, and it is unlikely to meet the needs of future proteomic studies in which multiple perturbations are compared.

Efforts in our lab have been directed to establishing high throughput proteomics methods that allow for multiple perturbations of a cell system to be analyzed in a short period of time. The general strategy taken has employed a high-resolution separation technique such as CIEF for intact proteins or capillary reverse phase LC for peptides, coupled on-line with FTICR-MS (Paša-Tolić et al. 1999). To measure differences in relative protein abundances, two isotopically distinct versions of a proteome are combined and analyzed. Our initial demonstration of isotopic labeling strategies for whole proteomes was the analysis of intact proteins using CIEF-FTICR to examine the cadmium (Cd^{2+}) stress response in *E. coli*. In these studies, *E. coli* was grown in both normal (i.e., natural isotopic abundance) and rare-isotope (^{13}C, ^{15}N) depleted media. To measure changes in proteins relative expression, aliquots were removed from the unstressed (normal medium) and stressed (depleted medium) cultures at different time intervals after Cd^{2+} addition, and the cells were mixed prior to sample processing and analysis by CIEF-FTICR. Combining the cells at this early time eliminates the experimental variables associated with cell lysis, separation, and

MS analysis. In other stable isotope-labeling approaches, cells have been cultured in ^{15}N-enriched medium and combined with cells cultured in normal medium (Gao et al. 2000). The proteins extracted from these combined populations are typically proteolytically digested and analyzed at the peptide level by capillary LC/MS.

Another approach for high throughput proteome quantitation, recently reported by Gygi et al. (1999), introduces the isotopic label after the proteins are harvested from the proteomes being studied. This approached, termed the ICAT method, involves the affinity selection of cysteine-containing polypeptides (Cys-polypeptides) by modifying the proteins with a Cys-specific reagent, biotinyliodoacetamidyl-4,7,10 trioxatridecanediamine (BIT) that contains a biotin group as well as a linker arm connecting the thiol-reactive group and biotin moiety. The reagent has been produced in a heavy isotopic version by replacing eight hydrogen atoms in the linker arm with deuterium atoms. Derivatization of two distinct proteomes with the light and heavy versions of BIT provides the basis for proteome quantitation as well as providing an additional Cys-constraint that can be used to increase the confidence of peptide identification.

We have combined ^{15}N-metabolic labeling with a commercially available Cys-affinity tag to derivatize and isolate Cys-polypeptides and allow the quantitation of relative peptide abundances from two separate proteomes. Initial studies focused on using this technique with the highly radioresistant bacterium *D. radiodurans* and mouse B16 epidermal cells. After equal numbers of cells grown in either normal or ^{15}N-enriched media were combined, the extracted proteins were derivatized with PEO-biotin, digested with trypsin, and the Cys-polypeptides isolated using immobilized avidin. The resultant Cys-polypeptide mixture was analyzed by capillary LC-FTICR. The results (Figure 4.5) show this to be an effective method to determine relative Cys-polypeptide abundances from proteomes grown in isotopically distinct media. In addition to a Cys-constraint, the combined use of a Cys-affinity tag and ^{15}N metabolic labeling in conjunction with high-resolution FTICR also provides an identification constraint based on the number of nitrogens present in the peptide.

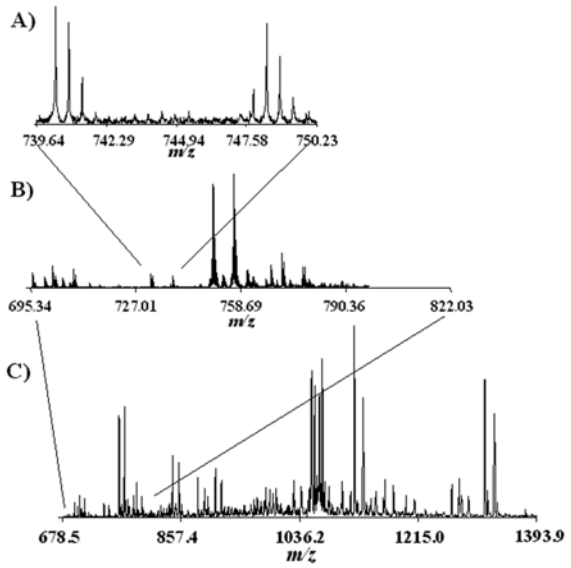


Figure 4.5. High density of information obtainable by combining high-resolution capillary LC separations with the high dynamic range and sensitivity of FTICR. Low abundance pairs of Cys-polypeptides (A) are readily seen in a spectrum containing more highly abundant species (B and C).

References

Celis, J. E. and P. Gromov. *Curr. Opin. Biotechnol.*, **10**, 16-21 (1999).

Gao, H., Y. Shen, T. D. Veenstra, R. Harkewicz, G. A. Anderson, J. E. Bruce, L. Paša-Tolić, and R. D. Smith. *J. Micro. Sep.*, **12**, 383-390 (2000).

O'Farrell, P. H. *J. Biol. Chem.*, **250**, 4007-4021 (1975).

Gygi, S. P., B. Rist, S. A. Gerber, F. Turecek, M. H. Gelb, and R. Aebersold. *Nature Biotech.*, **17**, 994-999 (1999).

Gygi, S. P., G. L. Corthals, Y. Zhang, Y. Rochon, and R. Aebersold. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 9390-9395 (2000).

Newsholme, S. J., B. F. Maleeff, S. Steiner, N. L. Anderson, and L. W. Schwartz. *Electrophoresis*, **21**, 2122-2128 (2000).

Paša-Tolić, L., P. K. Jensen, G. A. Anderson, M. S. Lipton, K. K. Peden, S. Martinovic, N. Tolic, J. E. Bruce, and R. D. Smith. *J. Am. Chem. Soc.*, **121**, 7949-7950 (1999).

Phosphoprotein Isotope-Coded Affinity Tag (PhIAT) Approach for Isolating and Quantitating Phosphopeptides

M. B. Goshe, T. P. Conrads,
E. A. Panisko, N. Angell,
T. D. Veenstra, and R. D. Smith

The availability of complete genome sequences is moving biological research to an era where cellular systems are analyzed as whole systems rather than as individual components. While genome sequences and global gene expression measurements at the mRNA level open the door to important biological advances, much of the understanding of cellular systems and the roles of the constituents will arise from proteomics. Proteomics, the analysis of the entire complement of proteins expressed by a cell, tissue, or organism, provides the most informative characterization of the cell since proteins are the primary players that carry out nearly all processes within the cell. A key aspect to successful proteomic measurements is the ability to precisely measure protein abundance changes in a high throughput manner to allow the effects of many "perturbations" upon, or changes to, a cell type, tissue or organ, to be determined in a rapid fashion (Patterson 2000). An inherent goal of proteomic studies is to provide a greater understanding of the function of proteins in a global, cellular context, along with the more conventionally delineated molecular function. A greater understanding at the level of cellular systems will provide a stronger basis for understanding complex biological pathways, the nature of diseases, and will permit a starting point to begin developing predictive capabilities for modeling the effects of cellular insults. The global understanding of cellular systems provided by proteomic investigations will provide numerous opportunities unlikely to originate from the present paradigm of "single" protein characterization methodologies.

While methods to make global measurements of gene expression at the mRNA level have been developed, such methods do not provide direct measurements of protein abundance (Gygi et al. 2000). Accordingly, global analysis of gene expression at the protein level provides the most informative target about the roles of individual gene products and their involvement in cellular pathways. Gygi et al. (1999) recently reported an approach for high throughput proteome quantitation, termed the ICAT™ method. This method involves the affinity isolation of Cys-polypeptides by modifying the proteins with a Cys-specific reagent, BIT, a molecule that contains a linker arm connecting a thiol-reactive group to a biotin moiety. The ICAT™ reagent is used in both a "light" and a "heavy" isotopic version where eight hydrogen atoms in the linker arm of the light reagent (D_0) have been substituted by eight deuterium atoms in the heavy reagent (D_8). Derivatization of two distinct proteomes with the light and heavy versions of the ICAT™ reagent provides a basis for proteome quantitation while providing an additional Cys-constraint that serves to increase the confidence of peptide identification.

While the ICAT™ approach promises great advances in determining protein abundance changes, delineation of protein function solely from abundance changes will be limited since numerous vital activities of proteins are modulated by post-translational modifications that may not be reflected by changes in protein abundance. The most important post-translational protein modification that modulates protein activity is phosphorylation, and it is by far the most common method cells use to propagate signals within cellular pathways and networks. Studies estimate that as many as one-third of all cellular proteins derived from mammalian cells are phosphorylated. Cellular processes ranging from protein kinase activation, cell cycle progression, cellular differentiation, and transformation, development, peptide hormone response, and adaptation are all regulated by changes in the state of protein phosphorylation.

As mentioned, proteomic studies to date have predominantly focused on measuring differences in protein abundance as a means to identify differences in the activities of distinct cellular systems. The activities of proteins, however, can be regulated at the level of synthesis and phosphorylation, as well as through protein-protein interactions. There are important examples of regulation at the level of protein synthesis with only parallel (proportional) changes in the level of protein phosphorylation as well as examples of regulating protein function by phosphorylation without corresponding changes in the level of protein synthesis. Therefore, the ability to identify changes in the phosphorylation state of a protein may lead to

discoveries related to protein activity, regardless of whether a given protein is differentially expressed.

The predominant method of studying protein phosphorylation is by labeling the proteins with ^{32}P-labeled inorganic phosphate ($^{32}P_i$). To measure differences in relative abundances of phosphorylation, ^{32}P-labeled proteomes are analyzed by two-dimensional polyacrylamide gel electrophoresis, and the relative spot intensities are compared (11,12). The use of $^{32}P_i$ to label proteins does not lend itself to high-throughput, proteome-wide analysis due to the problems with handling radioactive compounds and the associated radioactive contamination of instrumentation. It would be valuable to identify other methods that pose less of a risk than $^{32}P_i$, yet are still able to effectively identify phosphorylated proteins and quantitate the extent of phosphorylation. Two methods using heavy isotope metabolic labeling of proteins or derivatization of phosphoserine and phosphothreonine containing peptides with isotopically distinct reagents after protein extraction from the cell have been proposed (Oda et al 1999, Wechwerth et al 2000). While both of these strategies are effective at quantitating phosphopeptides, they lack the ability to isolate phosphopeptides from complex mixtures. Immuno- and metal-affinity columns designed to enrich mixtures for phosphopeptides often result in the isolation of many non-phosphorylated peptides through non-specific interactions. These non-specifically bound components complicate downstream analysis by introducing uncertainty about the nature of the sample.

To address this need, we have developed a method to tag phosphoserine and phosphothreonine residues so that peptides containing these residues can be isolated and quantitated efficiently. The scheme is shown in Figure 4.6. In the first step (i) the phosphate group is removed from the serine or threonine residue via a hydroxy-mediated β-elimination. The remaining electrophilic dehydroalanine (Figure 4.6b) is then reacted with ethanedithiol (EDT) via a Michael addition reaction (ii) to generate a free sulfhydryl group on the peptide. Iodoacetyl PEO-biotin is then added to the reaction mixture (iii) and binds to the sulfhydryl group to incorporate the biotin group onto the phosphopeptide. Immobilized avidin is then used to isolate the biotinylated serine and threonine phosphopeptides. The key is the use of a light (EDT-D_0) and heavy (EDT-D_4) isotopic version of EDT to modify the sites of serine or threonine phosphorylation after β-elimination of the phosphate group.

To demonstrate the ability of the PhIAT method to isolate and quantitate phosphopeptides, two equivalent samples of β-casein were modified with either the light or heavy isotopic version of EDT, followed by derivatization with iodocetyl PEO-biotin and peptide isolation using immobilized avidin chromatography. Analysis of the resulting sample by LC/MS shows the presence of pairs of peptides separated by 4 Da corresponding to the mass difference between EDT-D_0 and EDT-D_4. A pair of these peptides has a mass corresponding to the phosphorylated peptide FQS^{P}EEQQQTEDELQDK in which the serine residue has been modified with EDT (D_0 or D_4) and iodoacetyl PEO-biotin (Figure 4.6c). All of the expected phosphopeptides from b-casein were observed in the LC/MS analysis, showing the efficacy of the PhIAT approach for quantitating and isolating this class of peptides.

References

Gygi, S. P., G. L. Corthals, Y. Zhang, Y. Rochon, and R. Aebersold. "Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology." *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 9390-9395 (2000).

Gygi, S. P., B. Rist, S. A. Gerber, F. Turecek, M. H. Gelb, and R. Aebersold. "Quantitative analysis of complex protein mixtures using isotope-coded affinity tags." *Nat. Biotechnol.*, **17**, 994-999 (1999).

Oda, Y., K. Huang, F. R. Cross, D. Cowburn, and B. T. Chait. "Accurate quantitation of protein expression and site-specific phosphorylation." *Proc. Natl. Acad. Sci. U.S.A.*, **96**, 6591-6596 (1999).

Patterson, S. D. Proteomics: "The industrialization of protein chemistry." *Curr. Opin. Biotechnol.*, **11**, 413-418 (2000).

Wechwerth, W., L. Willmitzer, and O. Fiehn. "Comparative quantitation and identification of phosphoproteins using stable isotope labeling and liquid chromatography/mass spectrometry." *Rapid Comm. Mass Spectrom.*, **14**, 1677-1681 (2000).

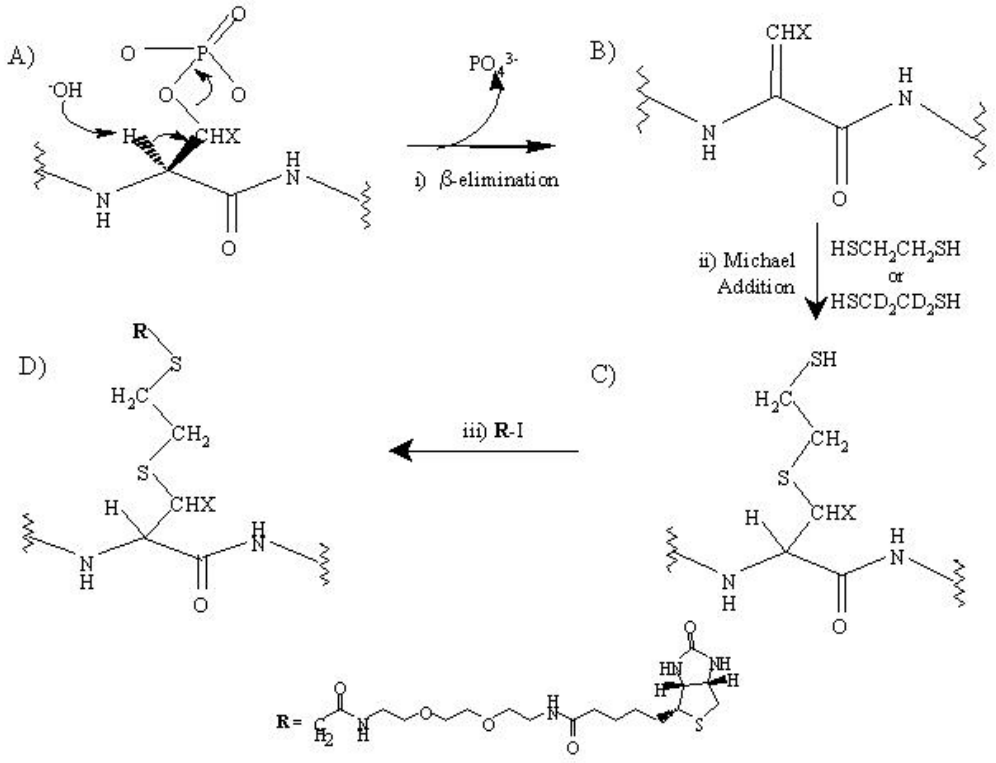


Figure 4.6. Reaction scheme showing the derivatization of the phosphoserine (X = H) or phosphothreonine (X = CH_3) residues of proteins to ultimately produce biotinylated peptides that can be affinity isolated from non-phosphorylated peptides via avidin affinity chromatography. The ability to quantitate the extent of phosphorylation between two identical peptides extracted from different sources is based on the use of a light ($HSCH_2CH_2SH$) and heavy ($HSCD_2CD_2SH$) isotopic versions of ethanedithiol (steps B to C).

Measuring and Understanding the Differences in DNA Damage Resulting from Normal Oxidative Processes and Low Levels of Ionizing Radiation

M. S. Lipton, O. L. Blum, D. S. Wunschel,[(a)] C. D. Masselon, and R. D. Smith

Supported by The Low Dose Radiation Research Program from The Office of Biological and Environmental Research at DOE

(a) PNNL Advanced Organic Analytical Methods Group

Introduction

Every day, each cell in the human body is constantly exposed to toxic oxidative agents that are damaging to DNA, yet 10^4-10^5 lesions/cell/day that result from endogenous oxidative reactions are repaired through normal processes (Beckman and Ames 1997). The human body also is constantly exposed to low levels of natural background radiation, and although the damage is similar to normal oxidative damage, the question remains whether the biological significance is the same (Wallace 1998). Currently, the health risks associated with exposure to low levels of ionizing radiation are poorly understood, in part because of the lack of appropriate analytical tools for characterizing the distinctive damage products that may be produced. It is possible that the small fraction of unique and unrepaired damage types resulting from exposure to ionizing radiation may lead to deleterious effects (Wallace 1998). It is important therefore to understand the real differences and similarities between DNA damage induced by normal oxidative processes versus low doses of ionizing radiation (Henle et al. 1996, Dizdaroglu 1992). Recent dramatic advances in analytical methods (Smith et al. 1993)

are now providing significantly more sensitive capabilities for efficiently and quantitatively detecting and characterizing the distinctive and chemically complex damage products from ionizing radiation (Hofstadler et al. 1995).

Methods and Instrumentation

New ultra-high sensitive FTICR-MS methods now allow the study of damage in DNA segments of much greater length (e.g., 25 to 30 mers) than previously practical for extremely low-level DNA damage. The damage products can thus be characterized and quantified in the context of small DNA segments instead of mono-nucleotides.

A novel "comparative display" approach is being applied to identify and quantitate the differences in damage products by the use of mass labeled set of DNA segments. This approach involves the simultaneous characterization of damaged DNA resulting from both ionizing radiation and normal oxidative processes, and also provides the basis for high precision in the quantitation of the damaged products by using one set of DNA products as effective "internal calibrants."

The labeled DNA segments are exposed to H_2O_2 oxidative stress to produce lesions that mimic normal oxidative stress while unlabeled segments are irradiated in a ^{60}Co gamma ray source at exposure doses designed to produce similar levels to major oxidative damage products. After damage the labeled and unlabeled samples are mixed together so all subsequent experimental variables are identical for both the radiation and oxidative damage products.

Fractions containing the intact DNA as well as fractions containing smaller DNA pieces created by radiation (e.g., induced double strand breaks) are collected after IP-RP-HPLC (Huber et al. 1995) and concentrated, before being subjected to a high-resolution capillary LC separation coupled to ESI-FTICR.

Preliminary Data and Results

Preliminary damage studies were performed on palindromic 12-mer oligonucleotide (~1 μmole, 5'-CAGTTCGAACTG-3') and on the 12-mer with a 5-methyl dC incorporated in position 1 or 6 as a 14-Da mass tag (see Figure 4.7). We initially used high levels of DNA damage agents to produce extensive damage for study using an LCQ ion trap mass spectrometer. The methylated 12-mer oligonucleotide was exposed for 3 hours to 3 mM H_2O_2, and the unlabeled 12-mer oligonucleotide was irradiated with 100 Gy of ionizing radiation (IR). Paired damage peaks were easily identified, allowing unique unpaired types of damage products from IR to be identified. The LC-MS results (Figure 4.7) have also shown that damage products can be enriched for the reverse phase separation, thus making them easier to identify in the context of excess undamaged oligonucleotides. Peaks corresponding to the damaged species appeared after the undamaged product, eluted and dominated the spectra, which aided in their characterization. Further experiments utilized uniformly mass labeled nucleotide bases ($^{13}C/^{15}N$) from PCR amplification of a 160 bp DNA template. DNA segments (9 to 37 bp) were obtained by a restriction digestion of both labeled and unlabeled 160-mer PCR products. After damage, a fractionation step was performed using reverse phase high performance liquid chromatography (RP-HPLC) to enrich for the damage species before MS analysis. The ability to examine each fragment from a large PCR product will allow further

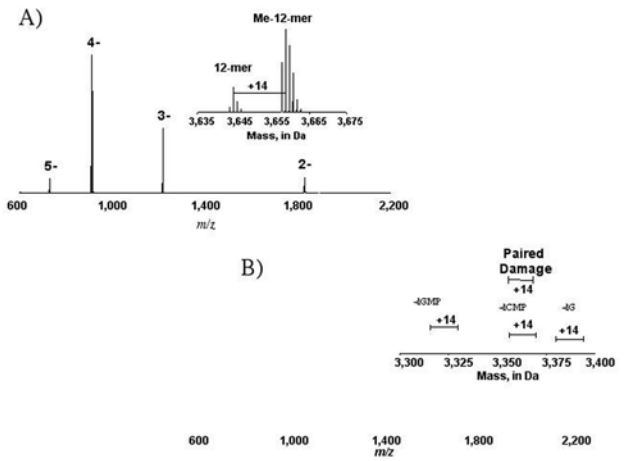


Figure 4.7. 7-T FTICR-MS spectra containing a mixture of unlabeled and 5-methyl-dC labeled 12-mers damaged by 3 mM and 0.03 mM H_2O_2, respectively. In A) the remaining undamaged 12-mers dominate the mass spectrum (inset contains mass transformed view). In B) ejection of undamaged species prior to detection allows the numerous damage species to be observed (inset contains mass transformed view of the 3 pairs of damaged species).

investigations into damage effects along the length of the DNA in the context of its packaging into nucleosomes or the effectiveness of DNA repair.

A set of consecutive experiments will be performed in the attempt to decrease gradually the level of damaging radiations to truly study "low" levels of IR (1 to 0.1 Gy) and its damage products.

Conclusion

Preliminary extensive damage studies on 12-mer oligonucleotide performed by ion trap mass spectrometry, showed forms of DNA damage common to both damage agents, as well as forms unique to ionizing radiation. The ultra high sensitivity of ESI-FTICR-MS instrumentation will be needed in conjunction with the novel comparative display approach to identify and quantitate previously undescribed forms of DNA damage arising from truly low levels of ionizing radiation.

References

Beckman, K. B. and B. N. Ames. "Oxidative decay of DNA." *J. Biol. Chem.*, **272**, 19633-19636 (1997).

Dizdaroglu, M. "Oxidative damage to DNA in mammalian chromatin." *Mutat. Res,*. **275**, 331-342 (1992).

Henle, E. S. et al. "Fe^{2+}, Fe^{3+} and oxygen react with DNA-derived radicals formed during iron-mediated Fenton reactions." *J. Biol. Chem*., **271**, 21,177-21,186 (1996).

Hofstadler, S. A. et al. "CITP ESI-FTICR for the study of X-irradiation damaged tetranucleotides." Atlanta, Georgia, 449 (1995).

Huber, C. G. et al. "Rapid and accurate sizing of DNA fragments by ion-pair chromatography on alkylated nonporous poly(styrene-divinylbenzene) particles." *Anal. Chem.*, **67.3**, 578-585 (1995).

Smith, R. D. et al. "Capillary electrophoresis/ mass spectrometry." *Anal. Chem.*, **65.13**, A574-A584 (1993).

Wallace, S. S. "Enzymatic processing of radiation-induced free radical damage in DNA." *Rad. Res.*, **150**, S60-S79 (1998).

Terascale Data Management for Proteomics Research

P. Cowley,[(a)] K. Swanson, and G. Kiebel[(b)]

Supported by the Office of Biological and Environmental Research at DOE

(a) PNNL Data and Knowledge Engineering Group
(b) PNNL Electronic Systems Group

Automated Acquisition and Management of FTICR Mass Spectrometry Data

PNNL has successfully completed the first year of a multiyear program to develop and implement high throughput proteomics. The centerpiece of the project is mass spectral analysis using FTICR-MS. The FTICR spectra have a MMA of 1 ppm, and over 1000 such scans can be produced in a single LC-MS analysis.

The raw data obtained from a single scan is on the order of 1 megabyte, and in a full production mode, approximately 500 such scans may be obtained per hour from a single FTICR machine. The near-term goal is to automate the entire data collection process such that four FTICR machines may obtain mass spectra 24 hours per day, 7 days a week. Since a single sample can generate over a thousand spectra, it is expected that the experiments will generate about 3.6 terabytes of data per year. Analysis of the spectra is performed using both custom and commercially available software, including SEQUEST originally developed at the University of Washington and commercially available from Finnigan Corporation; the Windows-based ICR-2LS software suite developed by the EMSL Instrument Development Laboratory; and a software package called "2D Gels."

As the proteomics project started, researchers quickly found that data volumes exceeded what they could store on personal computers, and they began storing the raw spectra on recordable CDs (CD-Rs), each of which can store a few hundred megabytes of data. Although this was a workable interim solution, at the projected production rates, one would consume over 60,000 such disks per year. In addition to the approximately 20 minutes required to write a single CD, the data contained on the disk is essentially

irreproducible; a loss or failure of a single disk could be a significant setback for the experiment. It became increasingly clear that we needed:

- A more formalized approach to data processing
- Common, integrated working storage for the large volumes of data being processed and analyzed
- Long-term archiving for data for which analysis is complete.

Work began on a data management system in order to formalize and automate the routine data processing previously done by hand. A database application was implemented to track the location and progress of data as it is acquired, processed, analyzed, and archived; to generate metadata (i.e., data about data); and to associate the data files with experiments and campaigns. An AMT database was also developed to provide researchers with information about the results of analyzing the data. File naming conventions were formalized and adopted.

To provide integrated working space, a Windows NT server was designated as the initial data archive site. The server was originally configured with 113 gigabytes of storage. Through the rest of the fiscal year, another 560 gigabytes of storage space was added to the server but even this was not adequate to accommodate the amount of data being generated.

To solve the problem of long-term data archiving, researchers are using NWArchive and its associated SDM software. NWArchive—EMSL's hierarchical storage management system—provides safe, reliable storage for EMSL's scientific and technical data. Archived data is initially stored on NWArchive's disk array and then written to tape. Each file is written to two tapes so the failure of a single tape does not lead to data loss. NWArchive handles this transfer without user intervention being required.

SDMExplorer is the primary SDM software being used to access the Archive. SDMExplorer was developed while the FTICR group's requirements were being identified and much of the functionality it provides was developed to accommodate the needs of the EMSL FTICR group. SDMExplorer exploits the Archive's ability for shared group-level access to the data, and thus allows simultaneous PC-based access to the data for viewing and analysis, thus eliminating the group's dependency on writing and exchanging CD-Rs. In addition, the tracking database can take advantage of SDMExplorer's ActiveX programmatic interface to automatically store the raw experimental data and its associated metadata in NWArchive. These linkages are shown schematically in Figure 4.8.

Furthermore, from their office PCs, researchers are able to query the tracking database and the AMT database and to use their own local copy of SDMExplorer to access and retrieve data for analysis on a shared, "as-needed" basis.

Efforts over the past year have provided the FTICR group with more reliable storage of their data, improved operator efficiency, significantly enhanced capability to locate relevant data, and dramatically improved ability to share and collaborate on data analysis. Because of NWArchive's reliability, the data is always safely stored and, through searches on the associated metadata, can be retrieved and reanalyzed for years to come.

Over the next year, more work will be required before we are able to manage and process data on a terabyte scale when four FTICR machines are generating data 24 hours per day, seven days per week. We are examining researchers' requirements and our current tools and processes to determine how we need to enhance what we have and what new capabilities we require. We are planning to add a new, larger dual-processor NT server with two terabytes of fiber channel RAID disks. This processor will provide a "sliding window" to data of immediate interest, which

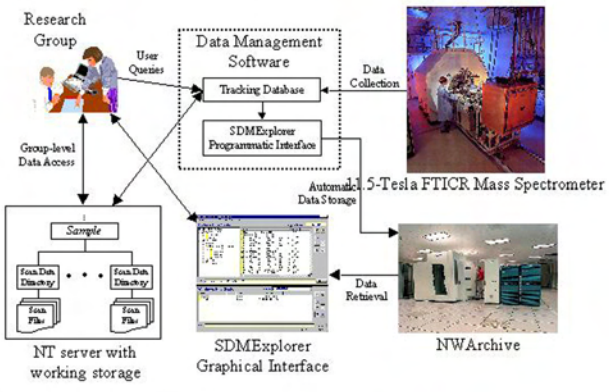


Figure 4.8. Proteomics research data management component.

includes data being processed for the first time and older data brought back online from NWArchive for further processing and analysis.

Data Dependent Control

G. Anderson

Portions of this work supported by NCI and NCRR

FTICR-MS can routinely deliver ppm mass measurement accuracy. For some cases this does not provide unique identification of the detected ion, and further fragmentation (MS/MS) of the ion is required. For complex spectra, such as those produced in proteomics research, selecting multiple sets of ions for MS/MS analysis on the fly and based on a user defined criteria is essential. Such a capability, called Data Dependent Control, is under development at PNNL. It is accomplished using the Finnigan ODESSEY Data Station coupled with a PC. The Data Station is used to generate the experiment scripts and all trigger signals. The Data Station stores all spectra for post-experiment data analysis. The PC records data in parallel with the data station and performs analysis of this data during the experiment. The raw data is transformed to the m/z domain and processed. This processing allows the PC to implement the user defined selection criteria and to generate accumulation and sustained of-resonance irradiation (SORI) collisionally induced dissociation (CID) waveforms that are a function of the MS data from the FTICR. The computer architecture is shown in Figure 4.9.

The Finnigan Data Station is used to generate all of the timing signals and the TTL trigger signals needed to control the experiment and to synchronize the PC with the Data Station. The experiment consists of two sequences, MS followed by MS/MS. During the MS step, the PC is triggered to acquire raw data in parallel with the data station. After the data is acquired, the PC converts the raw data into a m/z spectrum and then processes this result. Both isolation and SORI CID waveforms are calculated from this MS spectrum. After these waveforms are calculated they are downloaded to the waveform generators. The signals are output when the Data Station triggers the waveform generators using TTL trigger signals. The FTICR cell connections are shown in Figure 4.10.

Relay 1 and relay 2 are used to change the configuration of the detect and excite signals. These relays are under computer control; TTL signals from the data station determine the state of these relays, and they can be changed during the experimental script. The relays are shown in the default position, which supports normal excite/detect. Relay 1 controls the signal that is sent to the excite amplifier. In its default position, the data station excite signal is sent to the excite amp and then to the detection cell. When this relay is energized, the SWIFT isolation signal from the National Instruments DAQ5411 is sent to the Excite amp and on to the cell. Relay 2 is used to connect the Cell to the excite amp and pre-amp or to a transformer box used for QE and SORI. The transformer arrangement shown will allow simultaneous application of QE and SORI signals to the FTICR Cell. This relay box also is controlled using TTL trigger signals from the data station. The control software is shown in Figures 4.11 and 4.12.

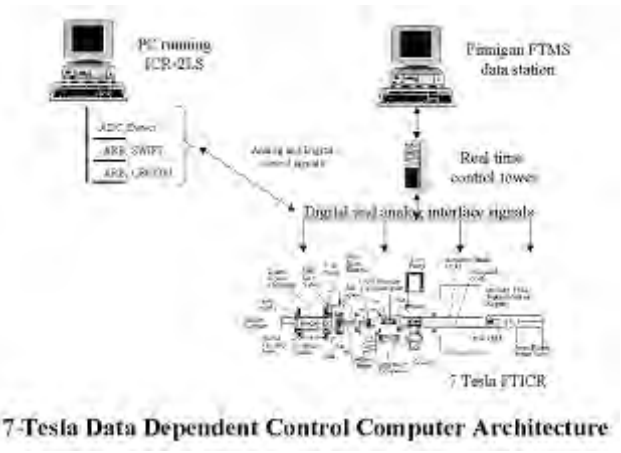


Figure 4.9. Computer architecture of the data dependent control system.

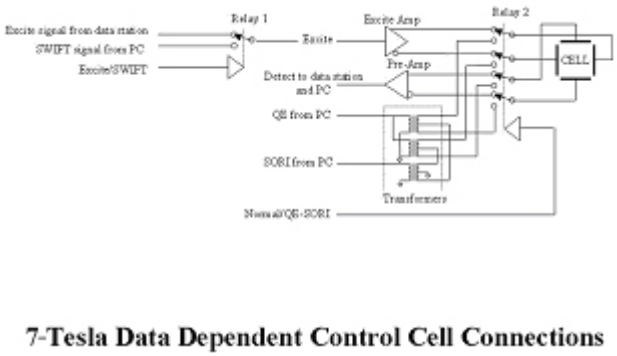


Figure 4.10. The FTICR cell connections used for Data Dependent Control.

The PC hardware is controlled by ICR-2LS, a software package developed in our laboratory for data analysis and instrument control. The user can define the number of data points and the sampling rate the ADC will use to acquire the raw data. After the data is acquired ICR-2LS will apply a macro to the raw data. This macro allows the user to define the processing steps desired for this raw data. These steps can include advanced options such as selecting peaks to use for SWIFT and SORI waveform generation. ICR-2LS also saves in a log file a record of the analysis steps; this record is used in post experiment data analysis. In the example of Figure 4.12, four peaks are selected and a SWIFT waveform is generated for the isolation of these four peaks. This operation is done using the data collected and displayed in the scope alone with the Excite Parameters shown in the dialog box. This dialog box allows the researcher to select multiple

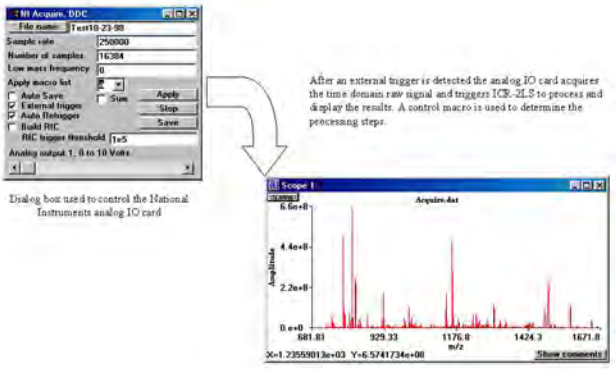


Figure 4.11. The dialog box used to control the acquisition of data by ICR-2LS.

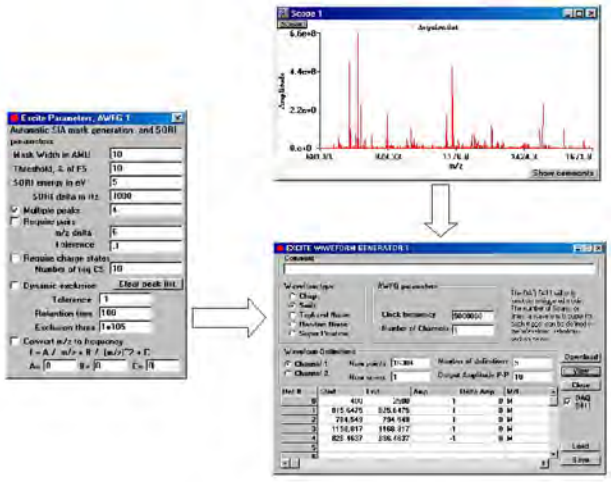


Figure 4.12. The user defined macro applied to the raw data after its acquisition.

peaks, and to use exclusion lists to prevent reanalysis of the same parent ion. In addition to generation of SWIFT isolation waveform, SORI CID waveforms are generated by selecting Super Position as the waveform type. SORI waveforms can be generated using an offset from the selected m/z peak in Hz or by defining the energy in eV. If the energy mode is selected the researcher can fix the voltage, frequency shift, or m/z shift for the generated waveform.

Data-Dependent Multiplexed Tandem FTICR Mass Spectrometry for High Throughput Proteomic Analysis

L. Li, C. D. Masselon, G. A. Anderson, S. Lee, L. Paša-Tolić, Y. Shen, T. P. Conrads, S. Berger, T. D. Veenstra, R. Zhao, and R. D. Smith

Portions of this work supported by PNNL Laboratory Directed Research and Development, NCI, and NCRR

Tandem mass spectrometry (MS/MS) plays an important role in the unambiguous identification and structural elucidation of biomolecules. In contrast to conventional MS/MS approaches for protein identification in which individual polypeptides are sequentially selected and dissociated, we have developed a multiplexed MS/MS approach where several peptides are selected simultaneously and dissociated using either infrared multiphoton dissociation (IRMPD) or multiple frequency SORI CID (Masselon et al. 2000). The high MMA and resolution afforded by FTICR coupled with database searching allows the fragments arising from several different parent ions to be correctly assigned and multiple peptides to be identified. Herein we discuss the application of multiplexed MS/MS in conjunction with on-line separations to rapidly dissociate and, hence, identify peptides present in a complex mixture (i.e., whole cell lysate). We have software developed that allows for “on-the-fly” data-dependent peak selection of polypeptides from each FTICR-MS acquisition. In the subsequent MS/MS acquisitions, several co-eluting peptide species can be fragmented simultaneously using either

SORI-CID or IRMPD. This MS/MS approach effectively allows numerous peptides from protein tryptic digests to be identified, thus enabling high-throughput proteomics.

The multiplexed MS/MS approach has been extended to rapid analysis of peptides present in complex mouse and yeast protein digests. In combination with the high-resolution capillary LC separation, multiplexed tandem mass spectrometric analysis allowed a large number of peptides and proteins to be identified in a high-throughput manner. As an illustration of the high density of information available, >100,000 polypeptides were detected, and over 800 proteins were uniquely identified using a single LC-FTICR multiplexed MS/MS analysis of yeast whole cell lysate (Figure 4.13). To reduce the complexity of the mouse proteome, a cysteine (Cys) specific biotin affinity tag (i.e., iodoacetyl PEO-biotin, PEO) was used to isolate Cys-containing peptides. In the example of an on-line LC-FTICR analysis of Cys-containing tryptic peptides isolated from mouse astrocytes shown in Figure 4.14, four parent masses were selected after an MS acquisition and simultaneously dissociated (using SORI-CID) in the subsequent MS/MS acquisition. The Cys-constraint, combined with high MMA, aids protein identification by reducing the number of potential database hits. These results highlight the significant utility of LC-FTICR multiplexed MS/MS approach for high throughput proteomic analysis.

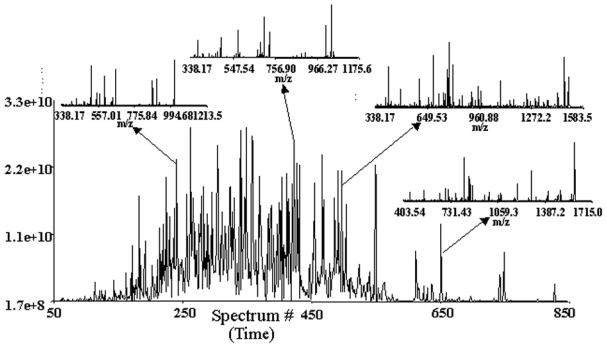


Figure 4.13. Total ion chromatogram reconstructed from FTICR multiplexed MS/MS spectra obtained during LC separation of yeast cellular tryptic digest, along with representative examples of multiplexed SORI-CID spectra obtained during various stages of LC separation.

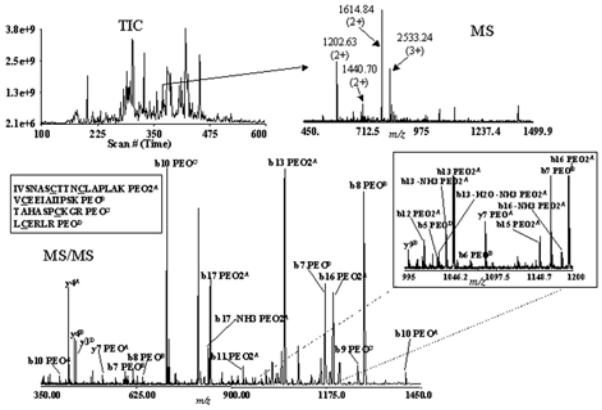


Figure 4.14. Example of a SORI-CID multiplexed MS/MS spectrum of four cysteine-containing peptides isolated from mouse astrocytes.

Reference

Masselon, C., G. A. Anderson, R. Harkewicz, J. E. Bruce, L. Paša-Tolić, and R. D. Smith. "Accurate Mass Multiplexed Tandem Mass Spectrometry for High-Throughput Polypeptide Identification from Mixtures." *Anal. Chem.*, **72**(8), 1918-1924 (2000).

Radial Stratification of Ions as a Function of *m/z* Ratio in Collisional Cooling RF Multipoles Used as Ion Guides or Ion Traps

A. Tolmachev, R. Harkewicz, K. Alving, C. Masselon, G. Anderson, V. Rakov, L. Paša-Tolić, E. Nikolaev, M. Belov, H. Udseth, and R. D. Smith

Collisional cooling radio frequency (RF) multipole ion guides are used in mass spectrometry to transmit ions from an intermediate pressure region of an ion source into a higher vacuum region of a mass analyzer (Douglas 1998). In practice, most collisional multipole ion guides do not use axial DC gradients, and space charge provides the main driving force for axial ion motion after collisional cooling. A rapidly developing application of the collisional RF ion guides is for the external accumulation of ions prior to mass

analysis in FTICR or other non-continuous mass analyzers. In these applications, the ion population may approach a level having significant space charge influence.

We have investigated the spatial m/z dependent distribution of ions in collisional multipoles. Computer simulations have revealed that, after sufficient cooling, the ion cloud takes the form of concentric cylindrical layers with each layer composed of ions having the same m/z ratio; the higher the m/z, the larger the radial position occupied by the ions. The cross section of such a structure is illustrated on the Figure 4.15a. Such a behavior follows from the fact that the effective RF focusing is stronger for ions having lower m/z, pushing these ions closer to the axis. Consider a quadrupole ion guide having an inscribed radius ρ, filled by several species of ions having masses m_i and charges ez_i, indexed in order of increasing m/z . Using the effective potential approximation, we can obtain the following radial positions of the layers:

$$R_{k,i} = \rho\left(\sum_{j=1}^{i+k-1} Q_j / q_i Q_{\max}\right)^{1/2} ; \quad q_i = \frac{4ez_i V_{RF}}{m_i \omega^2 \rho^2}; \quad n_i = \frac{4\varepsilon_0 V_{RF}^{\ 2}}{m_i \rho^4 \omega^2} \ \text{for } R_{0,i} < r < R_{1,i}$$

The index k=0 is for the inner radius, and k=1 is for the outer radius. Q_i, q_i, and ni are respectively the linear charge density, Mathieu parameter and ion number density for the ion species i, V_{RF} is the RF voltage, ω is the angular frequency, and $Q_{max}=\pi\varepsilon_0 V_{RF}$ is the maximum charge capacity of the RF quadrupole (Tolmachev et al. 2000). The ion density distribution described by these relations is shown by the dashed lines in Figure 4.15b; the ion species have the same mass m=1500 and charges 3+, 2+, and1+. Each charge state creates a cylindrical layer, separated from the adjacent layer by a zero population gap. The ion number density is the same inside each layer and is equal to $n_0=4\varepsilon_0 V_{RF}^{\ 2}/(m\omega^2\rho^4)$. The structure expands radially when the linear charge increases, but the number densities, n_i, stay the same. Similar relationships have been derived for a multipole having N pairs of rods; for N>2, the ion number densities are dependent on radius: $n_i(r) \propto r^{2(N-2)}$.

This idealized picture does not account for smearing of the layers due to the random component of ion motion and to the fast RF oscillations of ions. A more detailed description has been obtained by direct computer simulations, thus taking into account space charge and ion-molecule collisions. The simulations have shown that the stratified structure establishes itself when the collisional relaxation is complete, after a time interval of ~0.2 ms for a pressure of 10^{-2} Torr, or ~20 ms for 10^{-4} Torr. The stationary distribution obtained from simulations is shown on Figure 4.15b. The boundaries of the layers are smeared, but radial positions and number densities roughly correspond to above estimates. The smearing effect may be estimated as $\Delta r=4(\varepsilon_0 kT/e^2z^2n)^{1/2}$ (Tolmachev et al. 2000), thus the higher the ion number density n, the more distinct the stratified structure. The multiply charged ions produce more distinct radial layers, since $\Delta r \propto 1/z$. Simulations for higher order multipoles (e.g. hexapole and octapole device) also show a similar stratified radial ion structure.

The stratified structure of the ion cloud may result in ejection of the highest m/z ions, which have the largest radius. We have experimentally investigated the m/z discrimination of ions in RF-only quadrupoles used as external ion guide/traps on FTICR instruments. Figure 4.16a shows mass spectra obtained for insulin at different external accumulation times (t_{acc}). For t_{acc} shorter than 0.3 s, the intensities of all the peaks increased proportionally with t_{acc}. For 0.3 s $<t_{acc}<$0.5 s, a sharp decrease of the 3+ charge state has been observed, supporting distinct stratification and loss for the 3+ ions. The third spectrum of Figure 17a (t_{acc}=1s) shows similar discrimination for the 4+ peak.

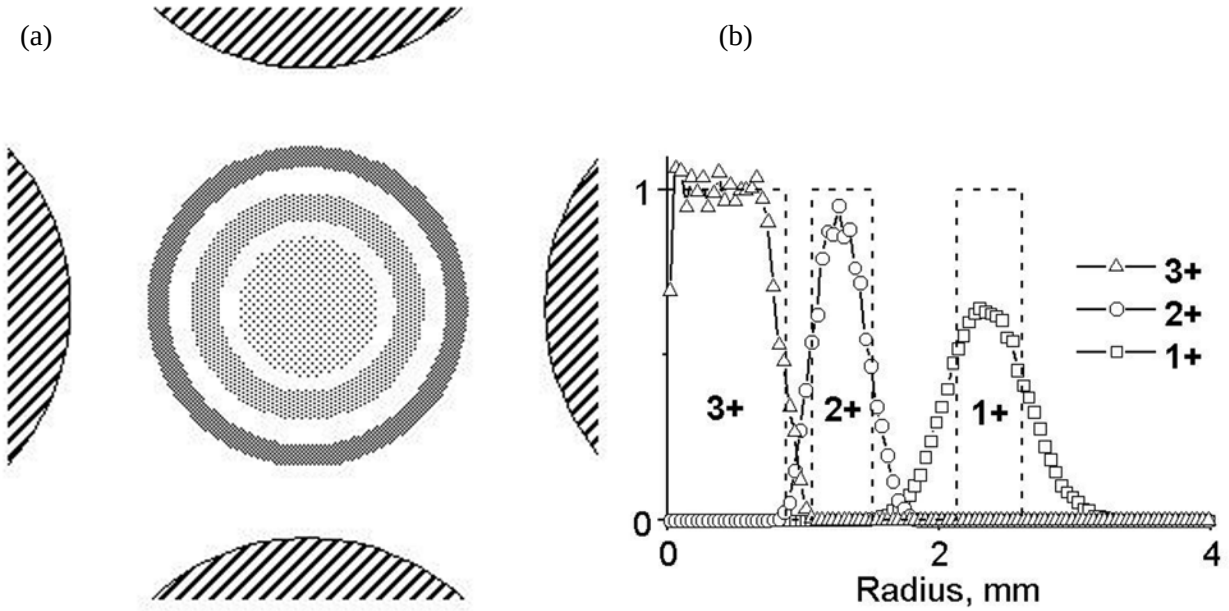


Figure 4.15. Radial ion density distribution in the collisional RF quadrupole: (a) a cross section view (radial statification and (b) ion number density of radial distribution for three charge states of ions m=1500, z=1+2+ and 3+. Effective potential results (dashed line) and direct computer simulation (symbols).

Another example of the high m/z discrimination has been observed in an RF-only quadrupole, which is used to couple an electrospray ion source to our 7-T FTICR spectrometer (see Figure 4.16b). In this case, ions are not trapped, but rather the DC ion current is transmitted through the ion guide at relatively high pressure, ~0.2 Torr. Modeling of the ion transport in RF quadrupoles has demonstrated that a considerable space charge can arise near the entrance and a stratified radial structure can be established. The ion cloud expands axially to help transmit ions, but also can expand radially up to quadrupole rods, resulting in the high m/z discrimination. This situation may occur when using pressures $>\sim 10^{-2}$ Torr, exceptionally high ion currents (such as obtained with an ion funnel interface), long ion guide lengths, and weak RF focusing (e.g., low V_{RF}). The understanding of these effects is essential for realizing optimum instrumental performance.

References

Douglas, D. J. *J. Am. Soc. Mass Spectrom.*, **9**, 101-113 (1998).

Tolmachev, A. V, H. R. Udseth, and R. D. Smith. *Anal. Chem.*, **72**, 970-978 (2000).

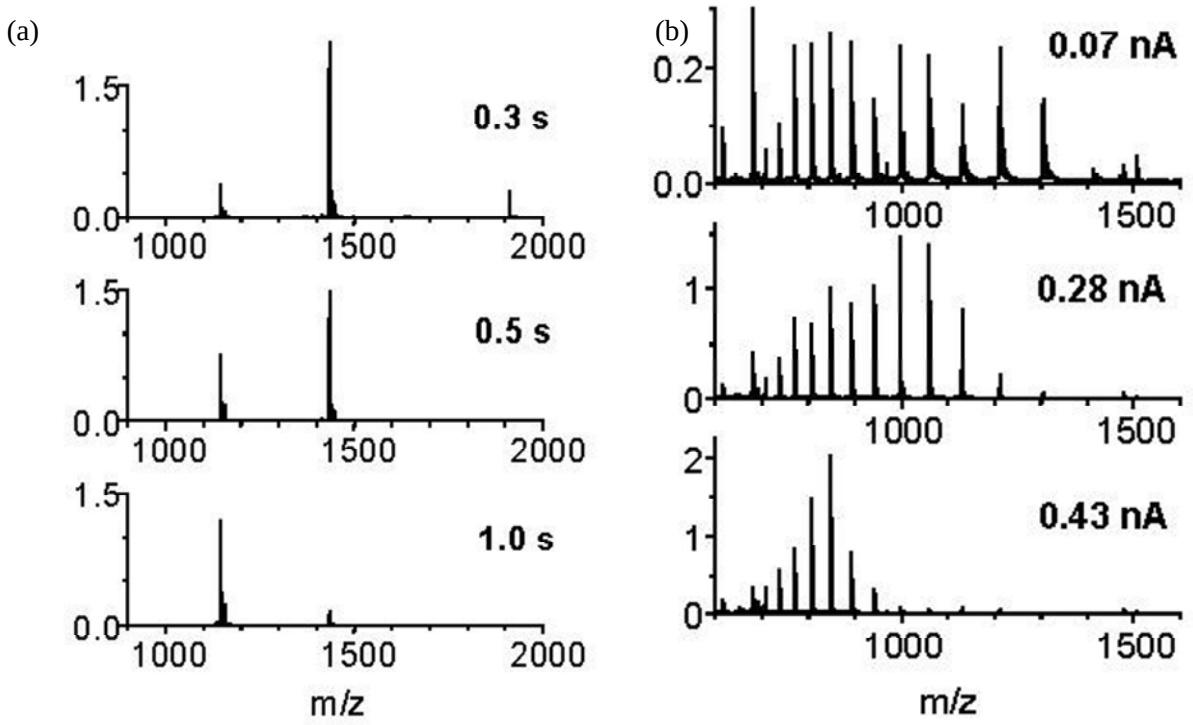


Figure 4.16. High m/z discrimination resulting from stratification of ions in RF quadrupoles. (a) Mass spectra of insulin obtained for different external accumulation times (radial statification and (b) transmission of myoglobin through collisional RF ion guide at elevated pressure, ~0.2 Torr (radial stafification).

Improving Mass Spectrometric Sensitivity Using Micro-Fabricated Multiple Electrospray Arrays, a Heated Multi-Capillary Inlet, and an Electrodynamic Ion Funnel Interface

K. Tang, T. Kim,
A. V. Tolmachev, D. C. Prior,
G. A. Anderson, H. R. Udseth,
and R. D. Smith

Supported by PNNL Internal Research and Development and National Center for Research Resource

Electrospray ionization mass spectrometry (ESI-MS) has become an important analytical tool, providing an effective means of analysis ranging from small organic and inorganic compounds to large biomolecules. A major limitation on the sensitivity of ESI-MS is the relatively inefficient transmission of ions in the first differential pumping stage of the mass spectrometer. Here ions formed at atmospheric pressure are transferred into the mass spectrometer via a conductance limit (e.g., capillary, sampling cone, or orifice) to the low vacuum region of the mass analyzer (0.5 to 10 Torr). Gas dynamic jet expansion and charge-charge repulsion led to a significantly expanded ion cloud and decreased transmission to the lower pressure region. Conventional electrostatic ion optics devices become increasingly ineffective at focusing the ion-cloud in this pressure region.

In this work, a heated multi-capillary inlet and an electrodynamic ion funnel interface, coupled with a micro-fabricated, multi-electrospray array, were developed to improve the ionization efficiency and the ion transmission efficiency through the first vacuum region of the mass spectrometer. The use of the heated multi-capillary inlet, consisting of seven stainless steel capillaries (Figure 4.17a), allows us to significantly increase the ion transmission from atmospheric pressure electrospray source to the first vacuum region without reducing the ion desolvation efficiency. The electrodynamic ion funnel (Figure 4.17b), installed in the first vacuum

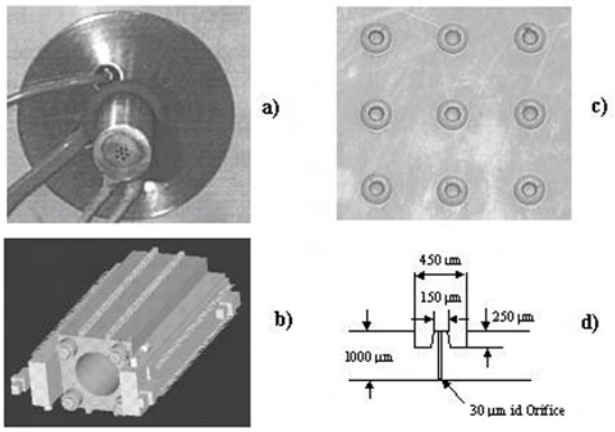


Figure 4.17. a) Heated multi-capillary inlet; b) electrodynamic ion funnel; c) microfabricated nine electrospray emitters arranged in a 3 x 3 configuration; d) dimensions for each spray emitter in the array.

chamber, effectively collects, focuses, and transmits the ion cloud into the high vacuum region of the mass spectrometer. To further improve the electrospray ionization efficiency, the multi-electrospray array (Figure 4.17c and Figure 4.17d), fabricated by the laser ablation method on a 1-mm thick polycarbonate sheet, is used to generate stable multi-electrosprays. This approach significantly increases the total spray current at a given liquid flow rate.

The technologies developed through this project were implemented and tested on a Sciex API 3000 triple quadrupole mass spectrometer. As shown in Figure 4.18, the standard electrospray source and curtain gas-skimmer interface of the API 3000 was replaced with the micro-fabricated multi-electrospray source, the heated multi-capillary inlet, and ion funnel interface. Mass spectra were collected using different test compounds, namely 5-fluorouracil, minoxidil, taurocholic acid, and reserpine, on both the standard and modified API 3000 for different operation modes (e.g., ESI $^{+}$, ESI $^{-}$, MS, and MS/MS modes) for the sensitivity comparison. It was found experimentally that a sensitivity enhancement of a factor of 20 to 30 could be achieved using the modified API 3000 for all the operating modes compared to the standard API 3000. Even greater increases of overall sensitivity and dynamic range are expected if the operation of the multi-electrosprays can be optimized and the space charge effect on ion transmission through the bottom of the ion funnel can be reduced effectively.

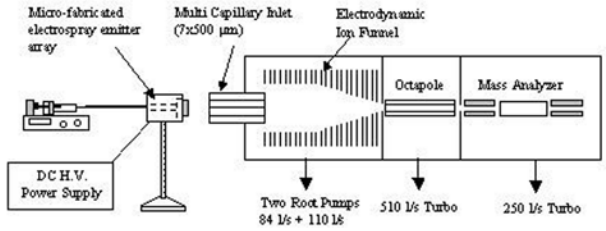


Figure 4.18. A modified API 3000 triple quadrupole MS using micro-fabricated multiple Electrospray Arrays, a heated multi-capillary inlet, and an electrodynamic ion funnel interface for significantly improved sensitivity.

Publications

Kim, T., A. V. Tolmachev, R. H. Harkewicz, D. C. Prior, G. Anderson, H. R. Udseth, R. D. Smith, T. H. Bailey, S. Rakov, and J. H. Futrell. *Anal. Chem.*, **72**, 2247-2255 (2000).

Kim, T., H. R. Udseth, and R. D. Smith. *Anal. Chem*., **72**, 5014-5019 (2000).

Kim, T., K. Tang, H. R. Udseth, and R. D. Smith. *Anal. Chem*., submitted.

Tang, K., Y. Lin, D. W. Matson, T. Kim, and R. D. Smith. *Anal. Chem*., submitted.

Investigation of Physical/ Chemical Separation Phenomena in the Electrospray Ionization Process

K. Tang and R. D. Smith

Supported by DOE-Office of Biological and Environmental Research

Electrostatic spraying, a liquid atomization process relying solely on the electric force in its formation, can generate highly charged droplets over a wide size range. A stable electrospray is often observed to form droplets initially with a bimodal size distribution consisting of larger primary droplets and smaller satellite droplets. The satellite droplets are quickly separated from the primary droplets in the electrospray due to their smaller inertia and higher charge density by the mutual repulsion between the primary and satellite droplets. A two zone structure of electrospray is often observed which consists of a

Of electrospray is often observed that consists of a core of the larger monodisperse primary droplets and a shroud, confined in a narrow periphery region, of the smaller satellite droplets.

As droplets in the electrospray move downstream, evaporation of neutral solvent continuously increases the surface charge density of the droplet. Eventually, the droplet charge approaches the Rayleigh limit (60 to 90%) at which point the droplet becomes unstable. The subsequent coulombic fission generates progeny droplets having significantly smaller sizes. The cascade of such droplet evaporation-coulombic fission processes eventually results in the generation of multiply charged gas phase ions. This electrospray ionization process has become one of the most widely used ion sources in mass spectrometry (ESI-MS) due to its gentleness, efficiency, and nearly universal applicability. The progeny droplets resulting from this highly asymmetric fission process are ejected most often from the parent droplet in the sidewise direction toward the periphery of the electrospray. Thus, it is expected that the progeny droplets will also tend to be separated from their parent droplets and move towards the periphery of the electrospray.

Both the initial droplet formation and subsequent fission processes involve a substantial deformation of the liquid surface. For a multi-component solution containing species preferentially located at the droplet surface, it is expected that the large deformation of the liquid surface in these processes will result in the enrichment of surface-active species in the satellite and the progeny droplets. This combined physical and chemical separations will likely result in a spatial partitioning of electrospray in which surface-active compounds are enriched in the periphery of the electrospray.

Two separate experiments were used to evaluate the possible physical/chemical separations in electrospray. The first experiment was performed on a simple experimental arrangement (Figure 4.19) using a solution of the hydrophilic methyl green dye and a brown surfactant, Fluorad (25 μg/ml methyl green plus 0.4% Fluorad by volume in deionized water). Stable electrospray at constant flow rate was generated and collected on the ground electrode an extended period of time (~2 hrs). The electrospray

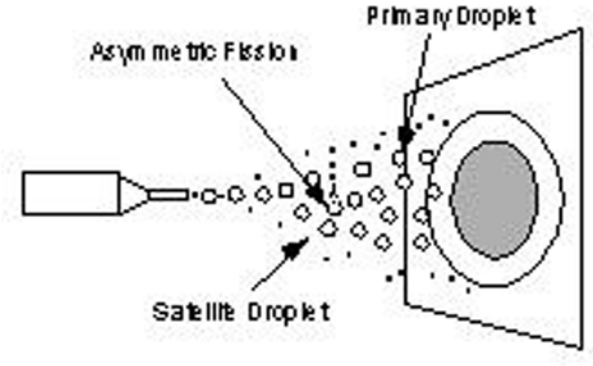


Figure 4.19. Experimental setup for the investigation of compound separation in the droplet formation and asymmetric fission processes.

deposition (Figure 4.20) shows a striking two zone structure of different colors; a center region that is predominantly green color and increasingly for brown color at larger radii, and a brown ring. We attribute the inner green region to the larger droplets and the transition to increasing brown due to progeny droplets from droplet asymmetric fission, and the fact that droplet breakup can likely occur at any distance from the emitter. The sharply defined outer ring is attributed to the smaller satellite droplets.

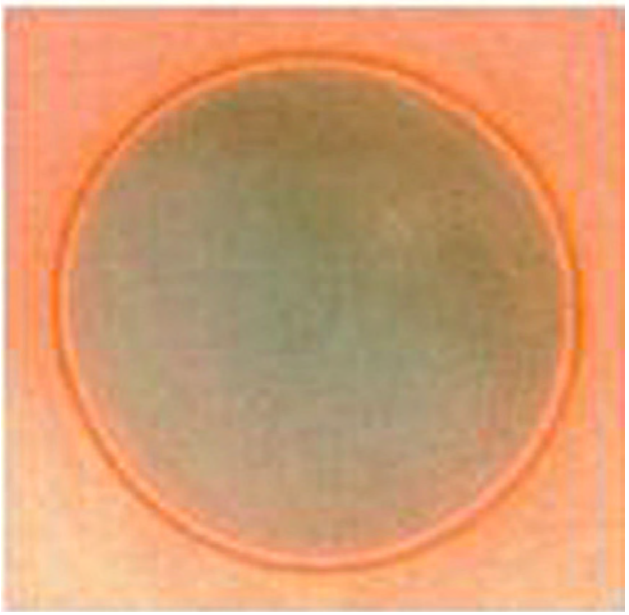


Figure 4.20. Electrospray deposition pattern collected on the ground metal plate.

The physical/chemical separation in the electrospray was further examined using direct mass spectrometric analysis at different spray radial positions using a triple quadrupole mass spectrometer. Mass spectra were collected at different radial positions for a solution containing 2 μg/μl cytochrome c and 0.58 ng/μl sodium dodecyl sulfate (SDS), an ionic surfactant, in deionized water. Figures 4.21a and b show the detailed 9+ charge state region for mass spectra obtained at the center of the electrospray and 4 mm laterally displaced from the spray center. The relative abundance of surfactant adduct peaks at 4 mm off center (Figure 4.21b) are significantly enriched compared to those at the spray center (Figure 4.21a). The results confirm again the proposed separation mechanism.

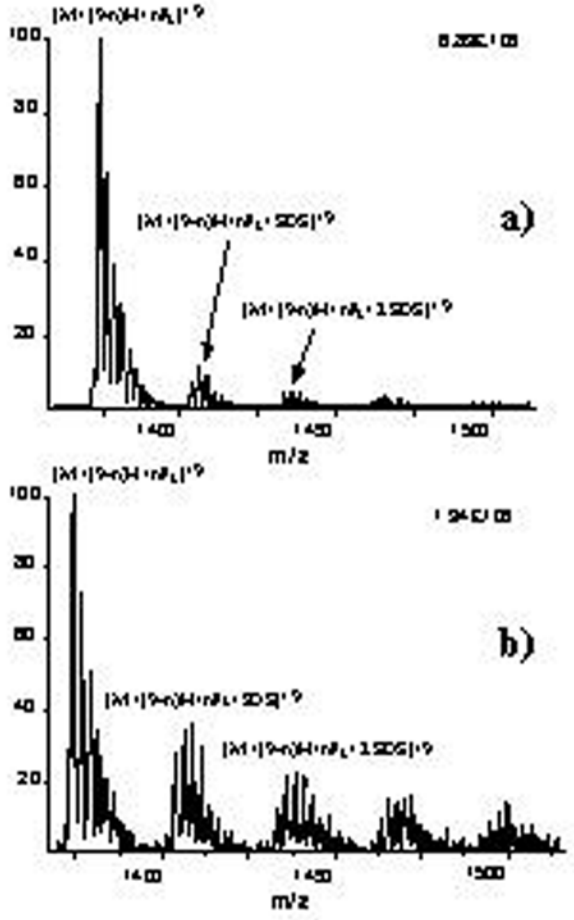


Figure 4.21. Mass spectra showing 9+ charge state region for ESI-MS of a mixture of cytochrome c and SDS obtained a) at the center of the electrospray, and b) at the edge of the electrospray.

The combined physical/chemical separation in electrosprays may have significant implications for the practice of ESI-MS. The ESI-MS analysis of biomolecules in surfactant containing solution is generally known to be problematic, and the surfactant signal can render many biomolecules undetectable. The analysis of such mixture is shown to be very sensitive to the electrospray position relative to the MS sampling orifice. A common approach is to sample electrospray at "off-axis" locations where higher signal-to-noise ratios can be typically obtained. The separation phenomenon shown in this study indicates that the optimum location to sample the electrospray will depend on the nature of the liquid solution. For solutions containing surfactants, the spray center may be a more favorable sampling location.

Related Publications

Gomez, A. and K. Tang. *Phys. Fluids*, **6**, 404-414 (1994).

Taflin, D. C., T. L. Ward, and E. J. Davis. *Langmuir*, **5**, 376-384 (1989).

Tang , K. and R. D. Smith. *J. Am. Soc. Mass Spectrom.*, in press.

Efficient Ion Transmission Using a Multi-Capillary Inlet and an Ion Funnel Interface

T. Kim, K. Tang, A. Tolmachev, H. R. Udseth, and R. D. Smith

In atmospheric pressure ionization (API) sources for mass spectrometry, a differential pumping system involving several stages is commonly used to achieve high transmission efficiencies. The most significant ion loss occurs between atmospheric pressure and the first vacuum stage. To increase the ion transmission efficiency through the first stage, a larger capillary inlet aperture can be used, but the ion current through the second stage does not increase significantly due to incomplete desolvation (Kim et al. 2000). We report on the improved ion transmission obtained using a new interface, a multi-capillary inlet, and an ion funnel interface.

The multi-capillary inlet provides more surface area than a single capillary with equivalent conductance

and results in better desolvation. A seven-capillary inlet was made by inserting seven stainless steel tubes into a central guiding hole of a cylindrical stainless steel heating block. The tubes and the heating block were silver soldered together (Figure 4.22a). A single capillary also made by a similar technique and used for comparison. An RF ion funnel has been demonstrated to provide improved transmission efficiency for a pressure typical of the first vacuum stage (~1 Torr) (Shaffer et al. 1997), and was used in this region to effectively capture ions from the multiple inlets. The experiments were carried out using a triple quadrupole mass spectrometer (SCIEX API 3000) with a multi-capillary inlet and ion funnel interface.

Results

Ion current transmission

An absolute ion transmission was measured with Dodecyltrimethyl-ammoniumbromide (DDTMA, C15H34NBr) in 100% acetonitrile. The ion transmission efficiency through the ion funnel was 65% for the seven capillary inlet.

Sensitivity Enhancement

Ion transmission efficiency also was measured by comparing the spectra of four different samples for ion spray source with the seven capillary inlet and ion funnel interface, and with the standard interface of API 3000. These results show that the detection efficiency was improved by factor of 4-15 with the multi-capillary inlet and ion funnel interface (Table 4.1).

Detection Efficiency

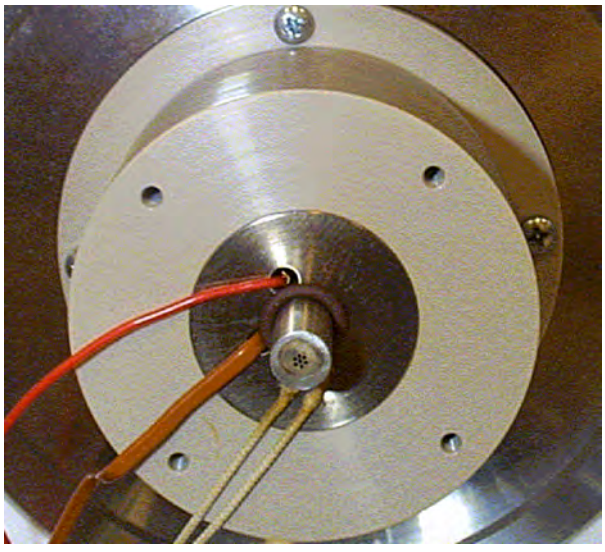
Ion detection efficiency was evaluated with a seven capillary inlet by monitoring the ion current after the analyzing quadrupole and by measuring the peak height in the spectrum (Table 4.2). A 2.9% overall detection efficiency was achieved with the new interface.

References

Kim, T., A. V. Tolmachev, R. Harkewicz, D. C. Prior, G. Anderson, H. R. Udseth, R. D. Smith, T. H. Bailey, S. Rakov, and J. H. Futrell. *Anal. Chem.*, **72**, 2247-2255 (2000).

Shaffer, S. A., K. Tang, G. Anderson, D. C. Prior, H. R. Udseth, and R. D. Smith. *Rapid Comm. In Mass Spectrom.*, **11**, 1813-1817 (1997).

(a)



(b)

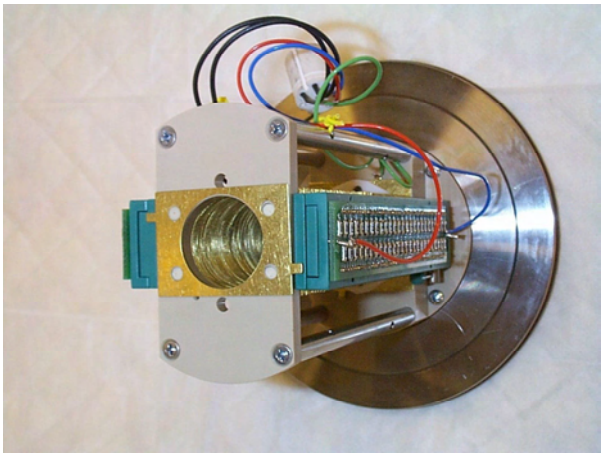


Figure. 4.22. The seven capillary inlet and the ion funnel. Efficient ion transmission (a) bmp and (b) efficient ion transmission.

Table 4.1. The major peak height in the mass spectrum with then standard interface of API 3000 and the seven capillary inlet and ion funnel (IF) interface.

Sample	Standard Interface [(a)]	Seven-Capillary and IF Interface [(b)]	Enhancement [(b/a)]
3.8 μM 5-Fluorouracil (m/z=129, 1-)	1.3 X 10^6	5.6 X 10^6	4.3
0.48 μM Minoxidil (210, 1+)	2 X 10^6	1.3 X 10^7	6.5
0.93 μM Taurochrolic (514, 1-)	3.3 X 10^5	4.8 X 10^6	14.5
0.16 μM Reserpine (609, 1+)	1.1 X 10^6	9.7 X 10^6	8.8

Table 4.2. Ion transmission efficiency through an analyzing quadrupole with unit mass resolution.

Sample	Infusion Rate (μL/min)	Analyte Mol. Infusion Rate[(a)] (A)	Iiq2[(b)] [peak height](m/z=228.3) (D)	Detection Efficiency (%) (Dtot [(c)]A X 100)
4.0 μM DDTMA	5.0	32 nA	1.0 nA	3.6
4.0 μM DDTMA	3.0	19.2 nA	0.58 nA	3.5
4.0 nM DDTMA	3.0	1.2 X 10^8 mols/sec	3.0 X 10^6 cps	2.9
0.16 μM Reserpine	1.0	1.6 X 10^9mols/sec	9.7 X 10^6 cps	1.0

(a) Calculated value from infusion rate assuming 100% ionization efficiency or molecular infusion rate.
(b) Ion current after the analyzing quadrupole or major peak intensity in the mass spectrum.
(c) Sum of ion currents at IQ2 with analyzing quadrupole set at major isotopic peaks (*m/z*=228.3 and 229.3) or sum of major isotopic peak heights in the spectrum.

All the Signal, All the Time: Focal Plane Mass Spectrometry Detector Development

D. W. Koppenaal,
C. J. Barinaga, M. B. Denton,[(a)]
G. M. Hieftje,[(b)] and
P. E. Miller[(c)]

Supported by the U.S. Department of Energy, Office of Nonproliferation and National Security (NN), Office of Research and Development (NN-20)

(a) Department of Chemistry, University of Arizona
(b) Department of Chemistry, Indiana University
(c) Lawrence Livermore National Laboratory

The objective of this DOE-NN Grand Challenge project is to develop, evaluate, and demonstrate an advanced MS ion detector based on a novel charge-integrating, multiplexer readout technology. The detector technology will provide for:

- simultaneous, wide mass range (~300 D) array providing detection of all the signal, all the time
- low noise level (< 5 e- read noise), high sensitivity (>10^{10} ions/sec) capability
- multiplexer-based destructive/non-destructive analog readout yielding inherent high dynamic range characteristics
- robust, expandable, flexible device design
- high-precision, high-accuracy isotope ratio performance.

The objective would represent a quantum leap in MS detector design and performance. Beyond the photoplate technology from the 1950s, no true multi-element, simultaneous, integrating detector exists for

MS applications. Only instruments that simultaneously monitor a small portion of the mass spectrum (4-9 masses) are available. For complex samples of limited size, or samples with time-varying concentrations, such instruments and techniques pose considerable constraints. The development of a detector with the above characteristics will enable and improve isotope measurement technology in terms of both technical and cost performance.

The research team assembled for this project is uniquely suited for the detector work. The Denton research group (University of Arizona) has pioneered the development and application of charge transfer device (CTD) technology to analytical optical spectroscopy instrumentation (Sweedler et al. 1994). This work, which commenced in the early 1980s, has resulted in the transformation of photon detection techniques from photomultiplier-based technology to solid-state detection. The emission spectroscopy market is a case in point virtually all ICP emission spectrometers today employ CTD devices as detectors. The group has also developed CTD technology for other optical techniques, including fluorescence, Raman, and x-ray diffraction spectroscopy (Hanley et al. 1996). The Denton group thus brings invaluable experience and contacts to the proposed research effort. In coordination with the other proposing principal investigators, this group in now turning its attention to solid- state detectors for mass spectrometry, the subject of this work.

Extensive prior experience also exists with focal plane mass spectrometry instrumentation. The Hieftje research group (Indiana University) has constructed and described a compact Mattauch-Herzog geometry instrument using a ICP as the ionization source (Burgoyne et al. 1997, Solyom et al. 1999). This instrument employs a short-array (~3 inches) MCP-CCD detector and is available for the proposed research. The PNNL research group has numerous sector mass spectrometers, including an available 10-cm focal length Mattauch-Herzog geometry instrument (formerly used for spark-source mass spectrometry using photographic detection). A new forward-geometry multi-collector ICPMS instrument for high-precision isotope ratio determinations has been added to the group in late CY 2000. The LLNL research group also possesses several sector field MS instruments, and has extensive isotope ratio analysis experience. Both national laboratory groups manage and operate nuclear detection and monitoring programs for DOE.

During the past year, work has focused on the design and development of the essential detector camera elements. The Arizona research group has acquired several candidate multiplexers and tested them for adequacy. Electronic readout systems have also been developed. The PNNL group has designed and developed a prototype Faraday "fingerboard," or ion detection element (IDEL) that will overlay the multiplexer device and complete the camera device. A diagram and photograph of the test IDEL assembly is shown in Figure 4.23. The Indiana University research group is developing a Mattauch-Herzog geometry ICPMS instrument (a smaller, ~3-inch, bench-top design) for the initial detector tests. To date, this group has developed an instrument with the following performance specifications: sensitivity – 10^9 cps/ppm; background – 100 cps; resolving power – 800 R; detection limits - 1 to 50 ppq; precision - ~5% rsd. The PNNL group is also readying a larger, 10-inch focal plane Mattauch-Herzog geometry ICPMS instrument for testing of a finished detector.

During 2001, the team will complete detector design and begin detector/camera testing on several group instruments.

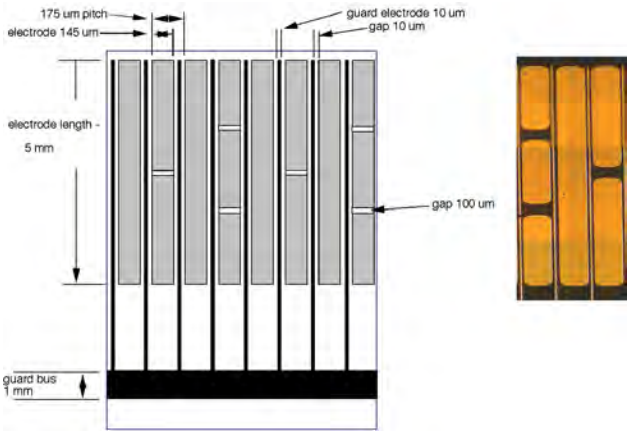


Figure 4.23. Diagram and photograph of a test IDEL. Shown are faraday strips (full length, 1/2 length, and 1/3 length) separated by guard electrodes designed to reduce scattered ions and channel cross talk. The strips are ~0.7-μ gold film on glass substrate.

References

Burgoyne, T. W., G. M. Hieftje, and R. A. Hites. *J. Amer. Soc. Mass Spectrom.*, **8**, 307 (1997).

Hanley, Q. S., C. W. Earle, F. M. Pennebaker, S. P. Madden, and M. B. Denton. *Analytical Chemistry*, **68**, 661A-667A (1996).

Solyom, D. A., T. W. Burgoyne, and G. M. Hieftje. *J. Anal. At. Spectrom.*, **14**, 1101-1110 (1999).

Sweedler, J. V., K. L. Ratzlaff, and M. B. Denton. *Charge-Transfer Devices in Spectroscopy*, VCH Publishers, New York (1994).

5. Mass Spectrometry Research User Reports

High Resolution FTICR-MS of the Explosive "Inert" Anion $CB_{11}(CF_3)_{12}$

B. T. King,[(a)] J. Michl,[(a)] C. Masselon, P. K. Jensen, L. Paša-Tolić, and R. D. Smith

Supported by PNNL Laboratory Directed Research and Development

(a) Department of Chemistry and Biochemistry, University of Colorado

Large size, absence of strongly basic sites, sterically protected delocalized charge, oxidation resistance, and chemical stability promise to make $CB_{11}(CF_3)_{12}$ dodecakistrifluoromethylcarba-closo-dodecaborate, a superior weakly coordinating counterion for extremely acidic, electrophilic, and/or oxidizing cations. However, its cesium salt is explosive, and the nearly spherical symmetry of the anion hinders crystalographic characterization of its salts.

The objective of this work was to measure the exact mass of an isotopically pure peak of the icosahedral anion $CB_{11}(CF_3)_{12}^-$ for determination of its elemental composition using high-resolution capabilities afforded by 11.5-T FTICR-MS (King and Michl 2000). This is a challenge for conventional mass spectrometers because they lack the sensitivity to measure isotopically pure peaks and the resolution to differentiate the more complex peaks (i.e., resolve isotopically fine structures). Attempts to resolve isotopically fine structures at lower magnetic fields (e.g., 7-T FTICR) were unsuccessful due to the phase locking of these closely spaced frequencies.

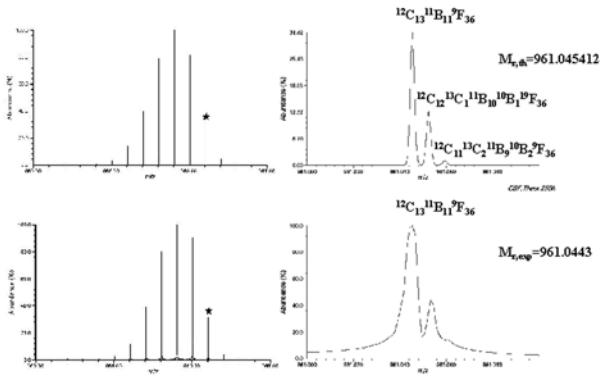


Figure 5.1. High-resolution negative ion ESI FTICR mass spectrum of $CB_{11}(CF_3)_{12}$. Experimental isotopic distribution (left bottom) and isotopic fine structure (right bottom) are shown in comparison with theoretical isotopic distribution (left top) and fine structure (right top).

Reference

King, B. T. and J. Michl. "The Explosive Inert Anion $CB_{11}(CF_3)_{12}$." *J. Am. Chem. Soc.*, **122**, 10,255-10,256 (2000).

Inactivation of Monomeric Sarcosine Oxidase by a Suicide Substrate

M. Jorns,[(a)] S. Martinovic, and R. D. Smith

Supported by Laboratory Directed Research and Development funding (STEP Initiative)

(a) M. C. P. Hahnemann University, Department of Biochemistry

Monomeric sarcosine oxidase (MSOX) is a monomeric protein containing FAD, which is covalently attached to a cysteine residue near the carboxy-terminus (cys315). Covalent attachment of FAD involves loss of 1 proton from cys315 and 1 proton from the 8-methyl group of FAD.

The kinetic mechanism of inactivation of monomeric sarcosine oxidase by N-(cyclopropyl) glycine (CPG) was recently determined. CPG is a sarcosine analog that acts as a suicide substrate and covalently modifies the flavin prosthetic group of the enzyme. We are now working on determining the chemical mechanism of inactivation. The modified flavin in CPG-modified MSOX is unstable since it cycles back to unmodified flavin via a 1,5-dihydroFAD intermediate. The modified flavin can be stabilized by further reduction with sodium borohydride. The borohydride-treated CPG-modified MSOX has been crystallized recently. A data set has been collected at 1.85 Å, but despite the high resolution, crystallographers still have difficulty in interpreting the new electron density near the flavin.

ESI-FTICR mass spectrometry was used to determine a molecular weight of the borohydride-reduced CPG-modified flavin and the molecular weight of unmodified MSOX as shown in Figure 5.2. Samples were analyzed under denaturing conditions to

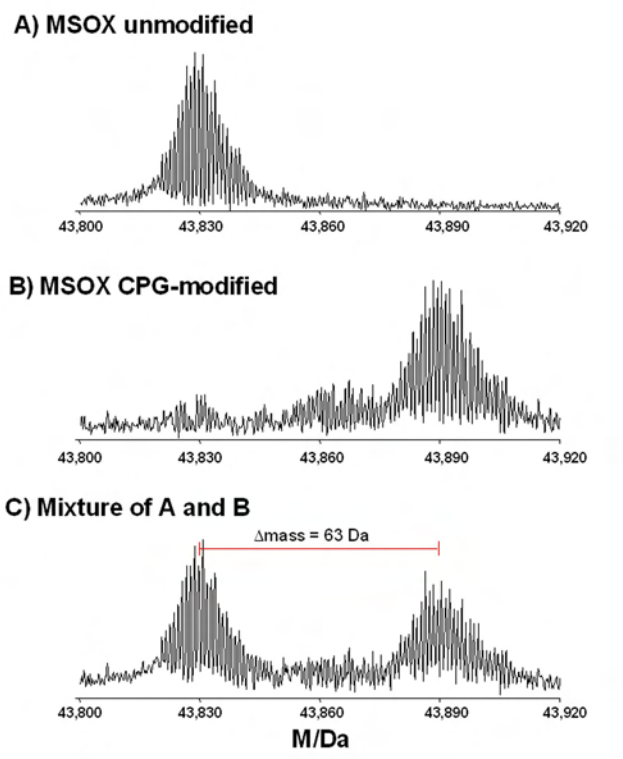


Figure 5.2. ESI-FTICR zero charge mass spectra of unmodified (A), CPG-modified (B) MSOX, and their mixture (C).

maintain covalent modifications. The modified flavin in this construct is "stable" upon denaturation with guanidine hydrochloride in the sense that it does not revert to unmodified oxidized flavin, but small spectral changes are observed over a period of hours after denaturation. The difference in measured molecular weight indicated the mass of covalent modification and gave rise to confirmation of the structure and mechanism of formation of CPG-modified FAD in MSOX after reduction with borohydride (see Figure 5.3). The proposed structure is consistent with the observed mass difference of 63 Da and with

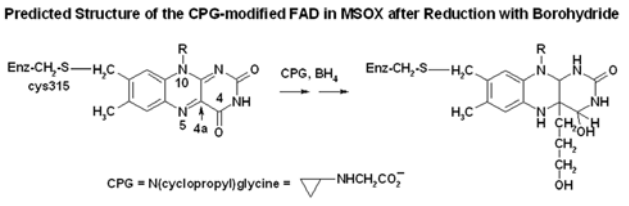


Figure 5.3. Proposed structure of CPG-modified FAD in MSOX after reduction with borohydride.

preliminary crystallographic results that show a tetrahedral structure for C(4) in the flavin ring and at least three atoms (not in a ring) attached to C(4A) of the flavin ring. Chemically, it is based on reaction of MSOX with CPG via an initial single electron transfer (SET) mechanism; borohydride is known to reduce imines and carbonyl functions, and we have evidence that borohydride reduces the C(4) carbonyl in unmodified MSOX.

Further, the modified MSOX will be analyzed by ESI-FTICR MS under native conditions in order to find out if there may be a fragment of the original inhibitor that is noncovalently bound and, therefore, lost under denaturing conditions. This information might be especially useful to the crystallographers who are still having problems completely interpreting the electron density.

Differential Isotopic Esterification for Quantitative Proteome Analysis and *de novo* Sequencing

D. R. Goodlett,[(a)] L. Li, L. Paša-Tolić, G. A. Anderson, and R. D. Smith

Supported by PNNL Laboratory Directed Research and Development

(a) The Institute for Systems Biology, Seattle, Washington

There has been significant interest in developing various isotopic labeling methodologies to measure differences in relative protein abundances between distinct proteomes. In the current study we have used differential isotopic per methyl-esterification of carboxylic acids in peptides as a method for quantitative analysis of identical proteins at different concentrations. Proteins were enzymatically digested to a mixture of peptides and esterified using either CH_3OH or CD_3OH. Peptides modified differentially with CH_3OH and CD_3OH were then mixed and analyzed by capillary reversed-phase LC-ESI-FTICR MS/MS. This chemical labeling scheme provides each peptide with an internal standard for normalization and thus relative quantitation. In addition, *de novo* sequencing is possible using the same data set used for quantitation by virtue of the fact that sets of tandem mass spectra for identical peptides are acquired during analysis (i.e., C-terminal [i.e., y ions] and N-terminal

[i.e., b-ions] ions are easily distinguishable, allowing the sequence to be derived readily within a series of consecutive ions).

To evaluate this labeling approach, we digested sperm whale myoglobin with trypsin and split the sample into two parts. One part was esterified with CH_3OH, and the other with CD_3OH, with the aliquot labeled with CH_3OH being twice the amount of the aliquot labeled with CD_3OH. The differentially labeled esterified tryptic digest was analyzed by LC-ESI-FTICR multiplexed MS/MS (Masselon et al. 2000). The PC-based ICR-2LS program controlled data processing in parallel with the Odyssey data station, rapidly converting the raw data to an m/z spectrum and generating the appropriate stored waveform inverse Fourier transform (SWIFT) and sustained off-resonance irradiation (SORI) waveforms to be applied during the ion isolation and dissociation preceding the next data acquisition. As shown in Figure 5.4, during the MS scan, two pairs of the most abundant parent ions were selected and adequate SWIFT and multiple frequency SORI waveforms were generated ("on the fly") and applied in the subsequent MS/MS data acquisition to dissociate all four parent species simultaneously. Since the two pairs of selected parent ions were different charge states of the same peptide, the searching of the MS/MS data against the sperm whale myoglobin database resulted in identification of peptide HGVTVLTALGAILK, with only the C-terminal Lys modified. This on-line multiplexed

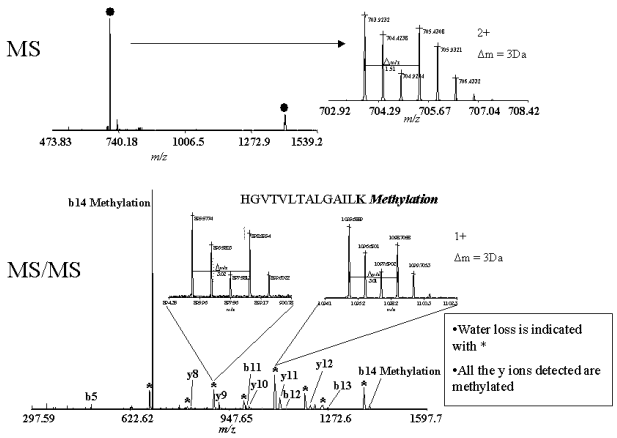


Figure 5.4. Example of multiplexed SORI-CID spectrum of tryptic peptide of sperm whale myoglobin modified with 2:1 CH_3OH:CD_3OH. The insert shows the isotopically resolved ion distribution of fragment ions.

MS/MS approach allowed the unambiguous identification of over 80% of the tryptic peptides resulting from this protein. The combination of differential isotopic labeling with high throughput multiplexed tandem FTICR-MS approach offers significant utility for proteomic studies.

References

Masselon, C., G. A. Anderson, R. Harkewicz, J. E. Bruce, L. Paša-Tolić, and R. D. Smith. "Accurate Mass Multiplexed Tandem Mass Spectrometry for High-Throughput Polypeptide Identification from Mixtures." *Anal. Chem.*, **72**(8), 1918-1924 (2000).

Identification of Nonpeptide Antigens for Human γδ T Cells

C. T. Morita,[(a)] L. Paša-Tolić, and R. D. Smith

Supported by the National Institutes of Health.

(a) Department of Internal Medicine, University of Iowa

T lymphocytes express either αβ or γδ T cell receptor heterodimers. Most T cells recognize antigenic peptides bound to major histocompatibility complex molecules, but the antigen recognition and biological function of γδ T cells are unknown. Human γδ T cells expressing Vγ2Vδ2 T cell receptors were shown to recognize nonpeptide antigens. Tanaka et al. (1995) identified natural antigens produced by mycobacteria recognized by human Vγ2Vδ2-bearing T cells as isopentenyl pyrophosphate and related prenyl pyrophosphate derivatives, compounds involved in the synthesis of complex polyisoprenoid compounds in microbial and mammalian cells. These results provided formal evidence that, in contrast to recognition of major histocompatibility complex-bound peptide antigens by αβ cells, human γδ cells can recognize naturally occurring small non-peptidic antigens.

The objective of this work is to measure the exact masses of isotopically pure peaks of non-peptidic antigens isolated from mycobacterial cultures for determination of their elemental compositions using high resolution and high mass measurement accuracy capabilities afforded by FTICR-MS. The studies at the EMSL successfully determined the chemical

composition of a number of phosphorylated compounds and led to their identification. Surprisingly, the M_r ~275 u compound, thought to be prenyl pyrophosphate compound (Tanaka et al. 1995), was identified as biologically inactive 6-phosphogluconate.

To determine the structure of the additional antigens from bacteria, we have purified the major bioactive antigen from *Mycobacterium smegmatis* by carbon filtration, ultrafiltration, anion exchange chromatography, and ion-pairing reverse phase chromatography. Based on the conventional tandem mass spectrometry (i.e., parent ion scan of 79 u for phosphate [PO_3^-] using triple-quadrupole), the antigen has a molecular mass of ~261 u and contains a pyrophosphate moiety. We have tentatively identified the M_r ~261 u compound as $^{12}C_5{}^{1}H_{11}{}^{31}P_2{}^{16}O_8$ based on ESI-FTICR spectra. (The identities of non-peptidic antigens will be further verified by tandem mass spectrometry.) The same antigen was also isolated from *Mycobacterium fortuitum*, and appears to be present in *Yersinia enterocolitica*, a Gram-negative rod, and *Staphylococcus aureus*, a Gram-positive cocci. The antigen is likely to be an unknown intermediate in the deoxyxylulose synthetic pathway for isopentenyl pyrophosphate that is present in eubacteria, apicomplexan protozoa, and plants. By recognizing a common biosynthetic intermediate, γδ T cells are using pattern recognition to determine when to mount an adaptive immune response.

Reference

Y. Tanaka, C. T. Morita, Y. Tanaka, E. Nieves, M. B. Brenner, and B. R. Bloom. "Natural and synthetic non-peptide antigens recognized by human γδ T cells." *Nature*, **375**, 155-158 (1995).

Structure Determination of Cyclic Siloxane

D. Nelson and J. Horner[(a)]

Supported by DOE, Office of Nonproliferation and National Security

(a) Currently at Thermoquest Corporation

A presumed cyclic Siloxane was observed after the synthesis of aromatic co-polymers, based upon fluorinated bisphenol A and a trisiloxane, prepared by hydrosilylation techniques. Some synthetic procedures produce up to 50% of the cyclic component compared to 4 to 9% obtained with other processes. To establish a rationale for the results, the identity of the cyclic product needs to be determined. IR suggested no active end-groups and, thus, a cyclic structure. NMR confirmed this finding and proposed a quite small structure with a molecular weight of approximately 900. Unfortunately, this does not allow a cyclic product incorporating both monomers with a unit molecular weight of 624. Further characterization by MS gave direct molecular weights to elucidate the reaction products and aid in determining their structure. Negative electrospray ionization MS was used to give parent ion molecular weights. This allowed determination of the degree of completion of the reaction and would identify any unwanted reaction products. Fragment ion spectra (MS/MS and MS/MS/MS) allowed identification of labile side chains and helped confirm the identity of the parent ions. It is projected that one further analysis will be required of a similar cyclic siloxane to complete the study. This work may also provide a synthetic route to avoid the cyclization during hydrosilylation. Such a route could be the basis for a patent.

Procedure

Roughly 10 uL of pure polymer isolated by HPLC was dissolved in 500 uL of acetonitrile containing 0.5% NH_4OH. The mixture was analyzed by negative ion electrospray (3.5 kV) on a Finnigan Classic ion trap mass spectrometer using SF_6 as a sheath gas. The solution was infused at 1.0 uL/min into a heated capillary regulated at a temperature of 200°C.

The full mass spectrum (shown in Figure 5.5) indicates two major components of mass 623.3 and 1247.1. These parent masses correspond to unreacted monomer and the desired dimer product ion thus indicating that the reaction is approximately 40% complete. The MS/MS analysis of the 623.3 Da component yields a major fragment at 607, which most likely corresponds to a loss of OH. MS/MS/MS analysis of the fragment at mass 607 yields fragments at 513 and 557. The MS/MS analysis of the 1247 Da component of the mass spectrum (Figure 5.6) yielded fragments at 1153.4 and 1227.3 Da. The major fragment from the MS/MS spectrum (1153.4 *m/z*) was analyzed by MS/MS/MS yielding a single major fragment at 1137.5, most likely again corresponding to the loss of OH.

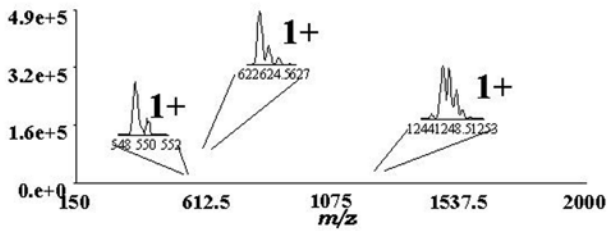


Figure 5.5. Negative ESI mass spectrum showing two major peaks corresponding to the monomer and product icon.

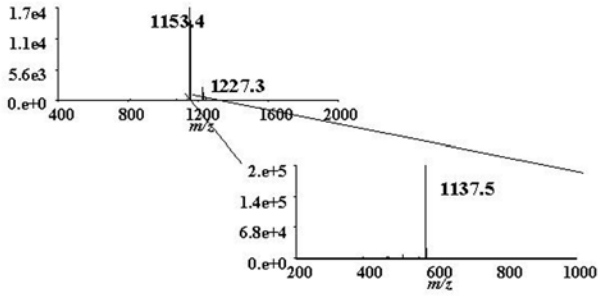


Figure 5.6. The upper spectrum is the MS/MS of parent mass 1247.1. It shows a primary peak corresponding to a loss of 94 daltons at 1153.4. The lower spectrum shows the fragmentation of this icon (MS/MS/MS spectrum). The only prominent feature is the icon at 1137.5, most likely corresponding to the loss of OH.

Sequence-Specificities of Hydrogen-Bonded Duplexes

H. Zeng,[(a)] S. Martinovic,
B. Gong,[(a)] and R. D. Smith

(a) Department of Chemistry, State University of New York, Buffalo

The construction of supramolecular architectures will be greatly facilitated if a diverse set of structural motifs leading to highly specific intermolecular interaction becomes available. Among the noncovalent interactions, hydrogen bonds have attracted the most attention in mediating the self-assembly of supramolecular structures.

This project involves the characterization of a new class of sequence-specific, H-bonded organic molecular duplexes (Zeng et al. 2000). Two oligoamide molecular strands, with complementary hydrogen-bonding sequences were found to form stable molecular duplexes in chloroform. These H-bonded duplexes eventually may be developed into a whole set of programmable molecular recognition units. Our purpose is to demonstrate the specificity within the given system.

We studied samples that contained six hydrogen bonding sites in which monomer 1 was fully complementary to monomer 2 and formed six attractive hydrogen-bonds in dimer 1•2 as shown in Figure 5.7. Monomer 5 self forms a self-complementary, six-hydrogen-bonded dimer 5•5 whose stability should be the same as dimer 1•2.

Electrospray ionization is a soft technique that can ionize intact noncovalent complexes. Complexes of H-bounded duplexes were detected using electrospray ionization from chloroform. The intact complexes were observed in the negative-ion mode as a Cl^--bound species using Ph_4PCl as the charge donor. ESI-FTICR mass spectrometry was used to determine a molecular weight of the ionized complexes. High-resolution of the FTICR instrument enables us to differentiate a specific dimer from nonspecific tetramer. M/z values for the (1-) charge state of the dimmer is the same as the (-2) charge state of the tetramer so the resolution of isotopic peaks is needed in order to differentiate the specific from non-specific complexes (Figure 5.8).

The next step includes quantifying the sequence-specificity of those duplexes by introducing mutant strands with mismatched H-bonding sites.

Related Publications

Zeng, H., R. S. Miller, R. A. Flowers II, and B. Gong. "A Highly Stable, Six-Hydrogen-Bonded Molecular Duplex." *J. Am. Chem. Soc.*, **122**, 2635-2644 (2000).

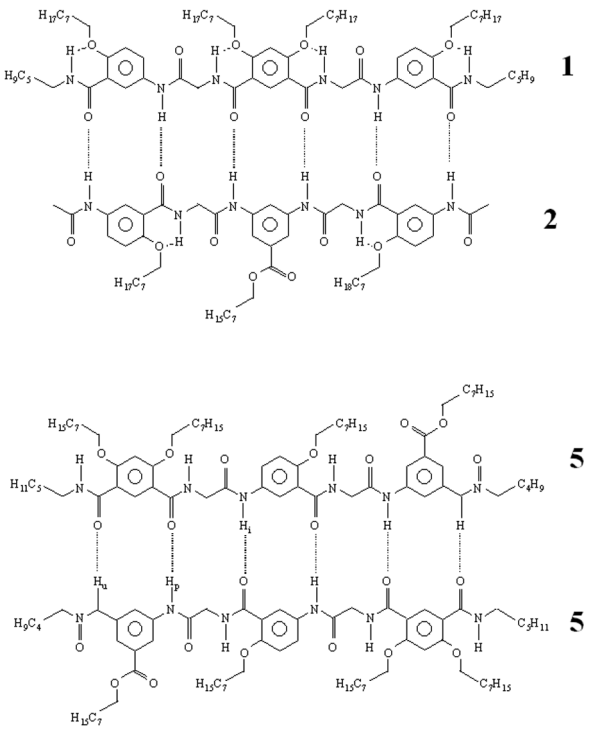


Figure 5.7. Structures of H-bonded molecular duplexes.

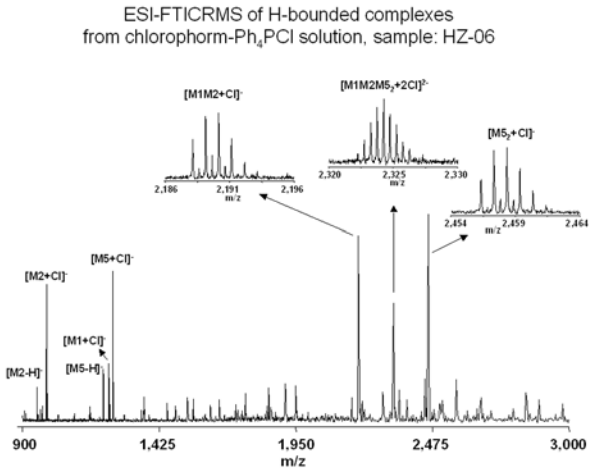


Figure 5.8. Negative ion mode ESI-FTICR mass spectrum of complex mixtures of 1•2 and 5•5 from chloroform.

Identification of Intrinsic Order and Disorder in the DNA Repair Protein XPA

L. M. Iakoucheva, A. L. Kimzey, C. D. Masselon, J. E. Bruce,[(a)] E. C. Garner,[(a)] C. J. Brown,[(a)] A. K. Dunker,[(a)] R. D. Smith,[(a)] and E. J. Ackerman

Supported by PNNL Laboratory Directed Research and Development.

(a) Washington State University, School of Molecular Biosciences (Currently with Merck Research Laboratories, West Point, Pennsylvania)

The DNA repair protein *Xeroderma pigmentosum group A* (XPA) is a crucial component of the nucleotide excision repair pathway (Friedberg 1995). The structure of full-length XPA (Figure 5.9) has proved intractable by both NMR spectroscopy and x-ray crystallography, and there is only limited structural knowledge for approximately one-third of the protein, the minimal binding domain (MBD). NMR solution structures of the MBD revealed that ~30% of its amino acids could not be assigned (Buchko et al. 1998, Ikegami et al. 1998). XPA is a potential example of an intrinsically unstructured protein whose flexibility facilitates complex interactions without sacrificing specificity.

To better characterize structural features of full-length XPA, we performed time-resolved trypsin proteolysis on active recombinant *Xenopus* XPA (xXPA) (Figure 5.10). The resulting proteolytic fragments were analyzed by electrospray ionization interface coupled to a FTICR-MS and SDS-polyacrylamide gel electrophoresis (PAGE). The molecular weight of the full-length xXPA determined by MS (30922.02 Da) was consistent with that calculated from the sequence (30922.45 Da). Moreover, the mass spectrometric data allowed the assignment of multiple xXPA fragments that were not resolvable by SDS-PAGE.

MEPEPEPEAN KEEEKILSAA VRAKIERNRQ RALMLRQARL ACRPYPTGEG
ISTVKAPPKV IDSGGGFFIE EEEAEEQHVE NVVRQPGPVL ECDYLICEEC
GKDFMDSYLS NHFDLAVCDS CRDAEEKHKL ITRTEAKQEY LLKDCDIDKR
EPVLKFILKK NPHNTHWGDM KLYLKAQVIK RSLEVWGSEE ALEEAKEVRK
DNRDKMKQKK FDKKVKELRR TVRSSLWKKE ASGHQHEYGP EEHVEEDSYK
KTCITCGYEM NYEKM

Figure 5.9. Sequence of the full-length xXPA with trypsin cutting sites bolded.

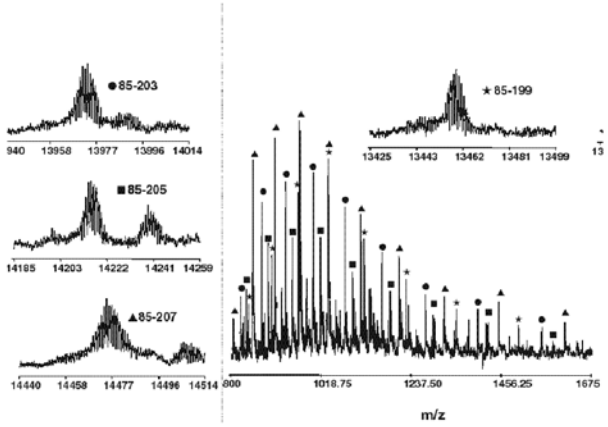


Figure 5.10. Example of ESI FTICR-MS spectrum of a fraction from the XPA digest. Insets show deconvoluted spectra for identified tryptic fragments.

The neural network program Predictor of Natural Disordered Regions (PONDR) applied to xXPA predicted extended-disordered N- and C-terminal regions with an ordered internal core. This prediction agreed with our partial proteolysis results (Figure 5.11), thereby indicating that disorder in XPA shares sequence features with other well-characterized intrinsically unstructured proteins.

Trypsin cleavages at 30 of the possible 48 sites were detected, and no cleavage was observed in an internal region (Q85-I179) despite 14 possible cut sites. For the full-length xXPA, there was strong agreement amongst PONDR, partial proteolysis data, and the NMR structure for the corresponding XPA fragment.

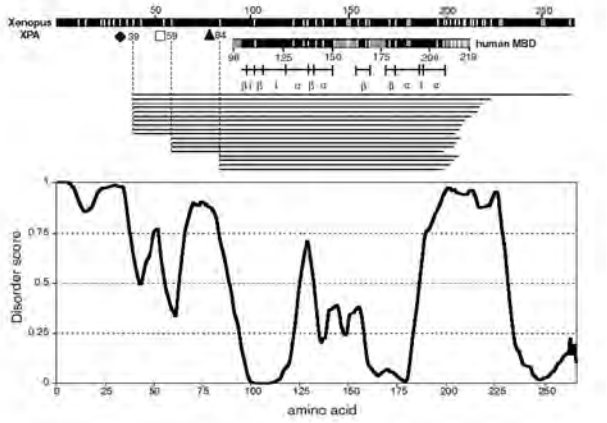


Figure 5.11. Comparison of proteolysis data with PONDR Predictions and NMR structure. (Top) Full-length *Xenopus* XPA depicted as a line, interspersed with trypsin sites as white vertical lines. The line below represents human XPA minimum binding domain (hMDB). Assigned structural regions (α = α-helix, β = β-sheet, t = turn) are depicted below hMBD with black vertical lines separating each region. Each of the experimentally observed trypsin:XPA proteolysis fragments corresponding to the three dominant bands on SDS gel are drawn as horizontal lines below. (Bottom) PONDR prediction of order/disorder in xXPA. Each residue is assigned a disorder score by the predictor based on the attributes of amino acids surrounding the residue. Predicted scores $\geq$ 0.5 signify disorder.

Related Publications

Buchko, G. W., S. Ni, B. D. Thrall, and M. A. Kennedy. "Structural features of the minimal DNA binding domain (M98-F219) of human nucleotide excision repair protein XPA." *Nucleic Acids Res*, **26**, 2779-2788 (1998).

Friedberg, E. C., G. C. Walker, and W. Siede. *DNA Repair and Mutagenesis*. ASM, Washington, D.C. (1995)

Ikegami, T., I. Kuraoka, M. Saijo, N. Kodo, Y. Kyogoku, K. Morikawa, K. Tanaka, and M. Shirakawa. "Solution structure of the DNA- and RPA-binding domain of the human repair factor XPA." *Nat. Struct. Biol.*, **5**, 701-706 (1998).

High-Throughout Proteomics Using High Efficiency Multiple-Capillary LC with On-Line High Performance ESI-FTICR Mass Spectrometry

Y. Shen, R. Zhao, N. Tolić, L. Paša-Tolić, L. Li, S. J. Berger, R. Harkewicz, G. A. Anderson, M. E. Belov, T. P. Conrads, T. D. Veenstra, M. S. Lipton, H. R. Udseth, and R. D. Smith

Supported by the National Cancer Institute under Grant CA 81654

We report on a capillary LC/ESI-FTICR approach to perform high-throughput proteome analyses. In this approach, whole protein extracts are proteolytically digested (e.g., with trypsin), and the resultant polypeptide mixture is analyzed in a single dimension of high-efficiency capillary LC and with high resolution, high sensitivity, and accurate mass measurements using FTICR. The single dimensional, high-efficiency capillary LC with long capillary columns containing small particles (≤3 μm) can provide a peak capacity of ~1000 for polypeptides at an operating pressure of 10,000 psi under conditions of coupling to FTICR-MS through an ESI source. The resolving power of FTICR can be as high as 10^5, and thus, the combined two-dimensional capillary LC/FTICR separation provides extremely high resolving power with a peak capacity of $>10^6$. The combined methodology holds the promise of proteome analysis with much greater throughput, sensitivity, dynamic range, and quantitation than presently possible.

We have designed a high-efficiency, multiple-capillary LC system for high-throughput proteome analysis. The multiple-capillary LC system was operated at the pressure of 10,000 psi using commercial LC pumps (ISCO Model 100DM) to deliver the mobile phase. Newly developed passive feedback valves were modified for switching the mobile phase flow and introducing samples for capillary LC (Figure 5.12). These valves (both 4-port and 6-port) have been effectively used for 10,000 psi capillary LC separations with lifetimes of >6 months (>400 times switching). These valves made it possible to implement a high-efficiency, multiple-capillary LC system equipped with small-particle

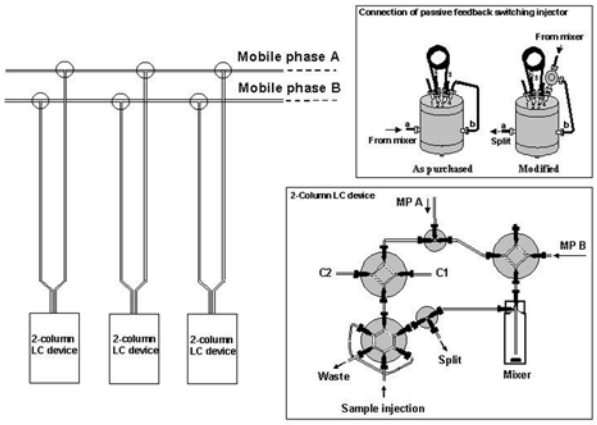


Figure 5.12. Schematic of the multiple-capillary LC system operated at 10,000 psi.

packed, long-capillary columns (e.g., 85-cm length, 3-μm particles). The multiple-capillary LC system was composed of several serially connected dual-capillary column devices. The dual-capillary column approach was designed to eliminate the time delay for re-generation (or equilibrium) of the capillary column after its use under the mobile phase gradient conditions (i.e., one capillary column was used in separation, and the other was washed using mobile phase A). The serially connected dual-capillary columns and ESI sources were operated independently, and could be used for either "backup" operation or with other mass spectrometer(s). Compared with the single-capillary LC in which the separation capillary was directly connected to the sample injector, no separation efficiency loss was found using the designed dual-capillary column device in the multiple-capillary LC system under the mobile phase gradient conditions. The capillary columns could be re-generated within the time period (~3 h) that required for completing high-efficiency separation of extremely complex cellular proteolytic polypeptides with the other column of the dual-capillary column device. This high-efficiency, multiple-capillary LC system uses switching valves for all operations and is highly amenable to automation. The separation efficiency of dual-capillary column design, optimal capillary dimensions (column length and packed particle size), and sample capacity, suitable mobile phases for electrospray, and the capillary re-generation were investigated. FTICR mass spectrometers were coupled on-line with this high-efficiency multiple-capillary LC system and used for analyzing cellular

(e.g., *Deinococcus radiodurans* and yeast) tryptic digests. Large numbers of cellular protein derived polypeptides (~100,000) could be reproducibly observed from the cellular proteolytic digests in single runs of 3 to 4 hours (Figures 5.13 and 5.14). We estimate the present dynamic range of measurements to be $>10^5$, and that this dynamic range can potentially be extended to $>10^7$.

Publications

Shen, Y., R. Zhao, M. E. Belov, T. P. Conrads, G. A. Anderson, G. Tang, L. Paša-Tolić, T. D. Veenstra, M. S. Lipton, and R. D. Smith. *Anal. Chem.*, submitted (2001).

Shen, Y., N. Tolic, R. Zhao, L. Paša-Tolić, L. Li, S. J. Berger, R. Harkewicz, G. A. Anderson, M. E. Belov, and R. D. Smith. *Anal. Chem.*, submitted (2001).

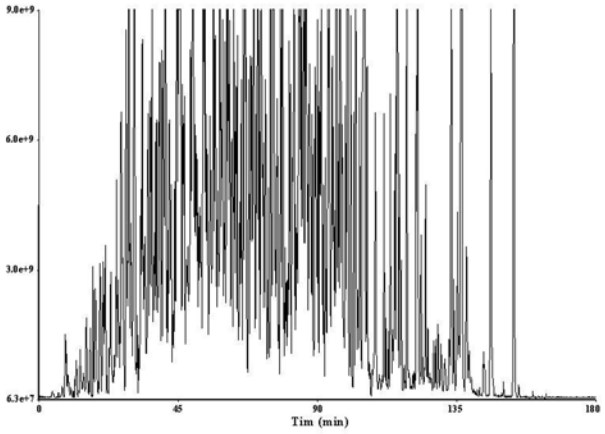


Figure 5.13. TIC of capillary LC/ESI FTICR-MS for cellular proteolytic polypeptides using the multiple-capillary LC system.

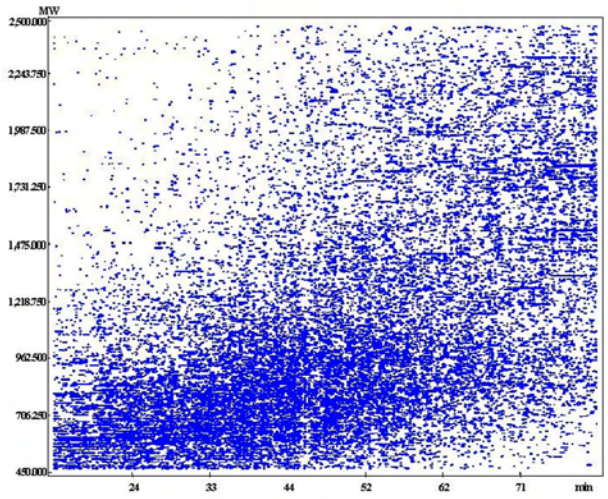


Figure 5.14. Capillary RPLC/ESI FTICR two-dimensional display for cellular proteolytic polypeptides.

Mass Spectrometry Research Capabilities

Triple Quadrupole Mass Spectrometer

The ThermoQuest TSQ 7000 is a triple quadrupole MS designed for use with an electrospray ionization source. The TSQ has a mass range of 50 - 4000 m/z with peak widths as low as 0.7 Da, and can be operated in MS or MS/MS mode using collision induced dissociation (CID). CID takes place in the second quadrupole at a nominal but variable gas pressure of 23 mTorr. The system can be operated in either positive or negative ion mode with the addition of SF_6 as a sheath gas.

Matrix-Assisted Laser Desorption/Ionization (MALDI) Mass Spectrometer

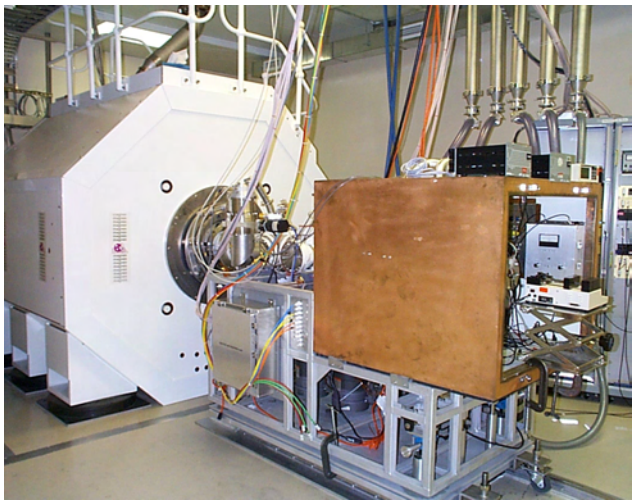
The PE Biosystems Voyager RP is a matrix-assisted laser desorption/ionization (MALDI) based time-of-flight (TOF) mass spectrometer. This instrument has a mass range up to 500 kDa, resolution to 2000, and accuracy of 10 ppm. The system is constructed with a 1-m flight tube fitted with a reflectron, allowing a 1-m (linear mode), or 2-m (reflectron mode) flight path. The system can be operated in post-source decay mode where the reflectron is used to analyze fragment ions from parents decaying in the first pass of the flight tube.

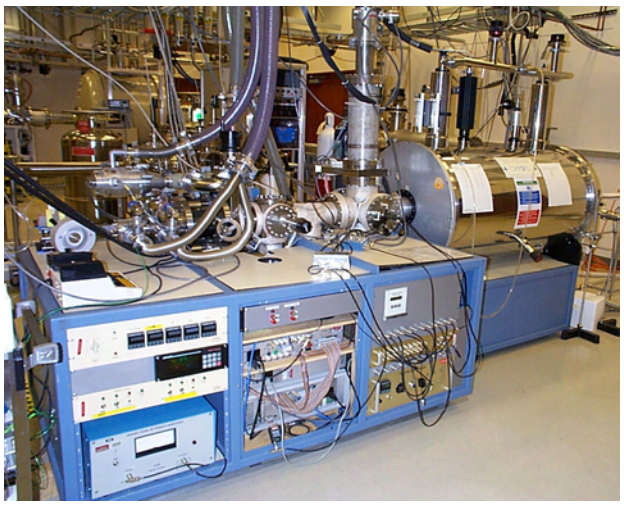
LC-Ion Trap Mass Spectrometers

Two ThermoQuest LCQ Classics and two DUOs are available. These are three-dimensional quadrupole ion traps designed for use with electrospray ionization sources. The mass range of this instrument is 150 - 2000 m/z, but for the classics, it can be extended to 4000 m/z for some applications. The LCQ has a maximum resolution of 10,000 in the zoom scan mode, and 4000 in full scan mode. In addition, the system is easily operated in either positive or negative ion mode with the addition of SF_6 as a sheath gas.

Fourier Transform Ion Cyclotron Resonance Mass Spectrometer

This ultrahigh performance FTCIR-MS uses a wide-bore, passively shielded 11.5-T superconducting magnet. The spectrometer is equipped with an electrospray ionization source and an ion funnel. The 11.5-T FTICR-MS has a resolution of about 150,000 and a mass accuracy approaching 1 ppm for protein samples with molecular weights ranging from 5000 to 20,000 Da. The 11.5-T FTICR can be fitted with an HPLC system and is equipped with an IR laser for multi-photon dissociation of samples for MS/MS in the ICR cell.



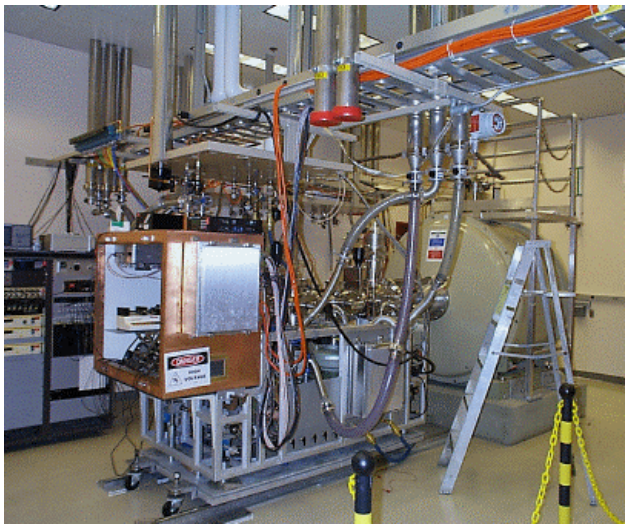


7-T FTICR Mass Spectrometer

This FTICR-MS is based on a 7-T superconducting magnet and is equipped with an external home-built ESI source. The 7-T FTICR has ultrahigh mass resolving power (e.g., $m/\Delta m$>2,000,000 has been obtained for insulin, while unit resolution is routinely achievable during on-line capillary isoelectric focusing (CIEF) experiments for species with M_r<30,000 u). Mass accuracy <5 ppm is typical for peptide/protein samples with molecular masses ranging from 500 to 30,000 u. A detection limit of ~10 attomoles has been obtained with online LC and CIEF separations.

Fourier-Transform Ion Cyclotron Resonance (FTICR) Mass Spectrometer

This high-performance FTICR-MS is coupled to a 3.5-T magnet and is equipped with an electrospray ionization source. The 3.5-T FTICR has a mass resolution of 50,000 to 150,000 and a mass accuracy of 3 to 7 ppm for protein samples with molecular weights ranging from 5000 to 20,000 Da. The detection limit of the instrument has recently been improved down to 3x10E-20 moles or about 18,000 molecules (based on sample consumption).



Mass Spectrometry Users and Collaborators

Users External to PNNL

Ruedi Aebersold
University of Washington
High Throughput Technology for Nervous System Proteome Analysis

Thomas H. Bailey
University of Delaware
Selective External Accumulation for FT-ICR-MS

Ninian Blackburn
Oregon Graduate Institute
Mass Spectrometry of PHM and its Reaction Products

Peter Brozovic
Washington State University
MS Verification of Non-Covalent Heterodimer Complex BRAC1-BARDI Za Binding Sites

Max Deinzer
Oregon State University
Characterization of the Recombinant Human Macrophage Colony Stimulating Factor β

Mamadou S. Diallo
Howard University/Cal Tech
Characterization of Chelsea Humic Acid and Lanchine Kerogen by Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry

Alvin Fox
University of South Carolina
Investigation of Tandem Mass Spectrometry for Identifying Single and Mass-Compensating Double Base Substitutions in p53 and Bacterial Ribosomal PCR Products

Ray Gestland
University of Utah
Characterization of Non-Peptide Antigens Recognized by Human T Cells

Steven Hofstadler
ISIS Pharmaceuticals
Evaluation of Ion Funnel for Drug Discovery

Bianca Hovey
University of Washington
Characterization of the A-subunit from Heat-Labile Enterotoxin, an AB5 Toxin from *E. coli*.

Benjamin T. King
University of Colorado-Boulder
High Resolution FT-ICR of CB11 (CF3)12-p

Larry L. Miller
University of Minnesota
Molecular Squares

Craig Morita
Iowa State University
Analysis of Protein Recoding in Yeast *Mito Chondira*

Bernd Moritz
Ruhr-Universitaet Bochum
FTICR-Mass Spectrometry

James J. Neitzel
Evergreen State College
Characterization of Bacteriophage T4 Early Gene Proteins

Carol L. Stone
Stevens Institute of Technology
Quantifying Human Liver Alcohol Dehydrogenase Isoenzymes

Ben L. Van Baar
TNO Prins Maurits Laboratory
High Resolution MS of Enzymatic Digest Fragments from a Sulfur Mustard and a Phosgene Protein Adduct

Michael S. Westphall
University of Wisconsin-Madison
Interfacing an Electrodynamic Ion Funnel with an Electrospray Ionization, Orthogonal Extraction, Time of Flight Mass Spectrometer for Improved Transmission of Single Stranded DNA Molecules

George M. Whitesides
Harvard University
Characterization of Novel Charge Ladders by Capillary Electrophoresis-FTICR Mass Spectrometry

Users (within PNNL)

David A. Nelson
Structure Determination of Cyclic Siloxane

Richard C. Zangar
Analysis of Tryptically Digested Proteins from 2D-PAGE using LC-MS to Determine Sequence Coverage

Collaborators

Eric Ackerman
PNNL

Ruedi Aebersold
University of Washington

Roger Brent
The Molecular Sciences Institute

Bruce A. Collings
MDS SCIEX

David R. Goodlet
University of Washington

Donald A. Griffin
Oregon State University

Richard Morrison
University of Washington

Tom Terwilliger
Los Alamos National Laboratory

Bruce A. Thomson
MDS SCIEX

Brian D. Thrall
PNNL

Kwong K. Wong
PNNL

6. Appendix

Macromolecular Structure and Dynamics Staff

Associate Director & Chief Scientist

David W. Koppenaal
(509) 376-0368
david.koppenaal@pnl.gov

B.S., Southwest Missouri State University 1974; Ph.D., University of Missouri-Columbia 1978; joined PNNL as Senior Research Scientist in 1988. Research Interests: Atomic mass spectrometry, laser ablation, ion trap-based instrumentation, inductively coupled plasma mass spectrometry.

Chief Scientists and Technical Group Leaders

Magnetic Resonance Research Group

Paul D. Ellis
(509) 372-3888
paul.ellis@pnl.gov

B.S., University of California at Davis 1966; Ph.D., University of California at Davis 1970; joined PNNL as Technical Group Leader for MS&D in 1993. Research Interests: Nuclear magnetic resonance spectroscopy, structural biology, and heterogeneous catalysis.

Biological Separations and Mass Spectrometry

Richard D. Smith
(509) 376-0723
rd_smith@pnl.gov

B.S., University of Massachusetts, Lowell 1971; Ph.D., University of Utah 1975; Postdoctoral Fellow, Naval Research Laboratory, 1975-1976; joined PNNL as Research Scientist in 1976. Research Interests: Development and application of advanced analytical methods and instrumentation and their biological applications, fundamentals and applications of mass spectrometry, and the development and application of new methods for probing the entire array of proteins expressed by a cell or organism (i.e., its "proteome").

Staff Scientists

Michael K. Bowman
(509) 376-3299
michael.bowman@pnl.gov

B.A., University of Kansas 1971; B.S., University of Kansas 1971; Ph.D., Wayne State University 1975; Postdoctoral Fellow, Argonne National Laboratory, 1976-1977; joined PNNL as a Staff Scientist for MS&D in 1992. Research Interests: Development and application of advanced pulsed EPR methodologies for elucidating structure and function in macromolecules, biological and nonbiological materials through high-resolution spectroscopy and magnetic relaxation.

Harold R. Udseth
(509) 376-3698
harold.udseth@pnl.gov

B.S., University of Minnesota 1969; Ph.D., University of Minnesota 1976; Postdoctoral Fellow, University of Utah, 1976-1978; Postdoctoral Fellow, Brookhaven National Laboratory, 1978-1981; joined PNNL as Senior Research Scientist in 1981. Research Interests: Novel work on both existing and new separation techniques, the development and construction of new interfaces between them and mass spectrometers, fundamental studies into ion molecule interactions and ion motion in the presence of collisions and electromagnet fields, and the development of electrospray ionization mass spectrometry and the liquid separations that can be coupled to a mass spectrometer by this ionization scheme.

Robert A. Wind
(509) 376-1119
robert.wind@pnl.gov

B.S., Delft University of Technology 1966; Ph.D., Delft University of Technology 1972; joined PNNL as a Staff Scientist for MS&D in 1993. Research Interests: Development of magnetic resonance microscopy instrumentation and methods with the largest possible sensitivity and spatial resolution, and applications of the methodology in projects of environmental and health-related significance.

Research Scientists

Mikhail E. Belov
(509) 376-0373
mikhail.belov@pnl.gov

M.S., Moscow Engineering Physics Institute, USSR 1988; Ph.D., General Physics Institute, Moscow, The Russian Federation 1993; Senior Research Associate, General Physics Institute, Russian Academy of Sciences, 1993-1994; Visiting Research Fellow, University of Burgundy, 1994; Research Fellow, University of Warwick, 1995-1997; Visiting Scientist, MS&D, 1998-1999; joined PNNL as a Senior Research Scientist for MS&D in 1999.

Sarah D. Burton
(509) 376-1264
sarah.burton@pnl.gov

B.S., Hartwick College 1985; Ph.D., University of Massachusetts 1991; joined PNNL as a Senior Research Scientist in 1992.

Herman M. Cho
(509) 376-2685
hm.cho@pnl.gov

B.A., Amherst College 1981; Ph.D. University of California at Berkeley 1987; joined PNNL as a Sr. Research Scientist in 1992.

Joseph J. Ford
(509) 376-2446
joseph.ford@pnl.gov

B.S., University of Notre Dame 1974; Ph.D., University of Wisconsin-Madison 1980; Postdoctoral Fellow, Carnegie-Mellon University, 1980-1981; Postdoctoral Fellow, University of Houston, 1981-1983; Research Associate, Baylor College of Medicine, 1983-1984; Research Assistant Professor, Baylor College of Medicine, 1985-1992; Applications Scientist, Otsuka Electronics USA, Inc., 1992-1998; Senior Installation Engineer, Varian Associates, Inc., 1998-1999; joined PNNL as a Senior Research Scientist II for MS&D in 2000.

Richard Harkewicz
(509) 376-1854
richard.harkewicz@pnl.gov

B.S., William Paterson University 1985; Ph.D., Michigan State University 1992; Postdoctoral Fellow, Argonne National Laboratory, 1992-1995; Staff Physicist, Michigan State University, 1995-1998; joined PNNL as a Senior Research Scientist for MS&D in 1998.

Julie A. Horner

B.S., University of British Columbia, Vancouver, BC, Canada 1988; M.S., Michigan State University 1991; Ph.D., Indiana University 1997; Postdoctoral Fellow, MS&D, 1997-1999; Senior Research Scientist, PNNL, 1999-2000; now at ThermoQuest, California.

David W. Hoyt
(509) 373-9825
david.w.hoyt@pnl.gov

B.A., Hartwick College 1985; Ph.D., University of Massachusetts 1991; Postdoctoral Fellow, University of Alberta, 1991-1995; Postdoctoral Fellow, MS&D, 1995-1997; joined PNNL as a Senior Research Scientist for MS&D in 1997.

Pamela K. Jensen

B.S., Northwest Missouri State University 1993; Ph.D., Iowa State University 1997; Postdoctoral Fellow, MS&D, 1998-1999; Senior Research Scientist, PNNL, 1999-2000; now at Monsanto, St. Louis, Missouri.

Michael A. Kennedy
(509) 372-2168
makennedy@pnl.gov

B.S., Muskingum College 1984; Ph.D., University of South Carolina 1989; Postdoctoral Fellowship, University of California at San Diego, 1989-1990; Postdoctoral Fellowship, MS&D, 1990-1993; joined PNNL as a Senior Research Scientist for MS&D in 1993.

Andrew S. Lipton
(509) 376-6633
as.lipton@pnl.gov
B.S., Syracuse University 1987; Ph.D., University of South Carolina 1993; Postdoctoral Fellowship, MS&D, 1993-1996; joined PNNL as a Senior Research Scientist for MS&D in 1996.

Mary S. Lipton
(509) 373-9039
mary.lipton@pnl.gov
B.S., Juniata College 1988; Ph.D., University of South Carolina 1993; Postdoctoral Fellow, MS&D, 1993-1996; Visiting Scientist, MS&D, 1997-1998; joined PNNL as Senior Research Scientist for MS&D in 1998.

David F. Lowry
(509) 373-0755
david.lowry@pnl.gov
B.S., Texas A&M University 1985; Ph.D., Bradeis University 1991; Postdoctoral Fellowship, University of Oregon, 1991-1995; joined PNNL as a Senior Research Scientist for MS&D in 1995.

Kevin Minard
(509) 373-9847
kevin.minard@pnl.gov
B.A., SUNY/Genesco 1988; M.A., Rice University 1992; Ph.D., Rice University 1995; Postdoctoral Fellowship, MS&D, 1995-1997; joined PNNL as a Senior Research Scientist for MS&D in 1997.

Ljiljana Paša-Tolić
(509) 376-8859
ljiljana.pasatolic@pnl.gov
B.S., University of Zagreb, Zagreb, Croatia 1986; M.S., University of Zagreb, Zagreb, Croatia 1989; Ph.D., University of Zagreb, Zagreb, Croatia 1992; Research Assistant, Rudjer Boskovic Institute, 1992-1993; Visiting Associate, Florida State University, 1993-1995; Postdoctoral Fellow, MS&D, 1995-1997; joined PNNL as a Senior Research Scientist for MS&D in 1997.

Donald N. Rommereim
(509) 376-2671
don.rommereim@pnl.gov
B.A., South Dakota State University 1972; M.A., South Dakota State University 1978; joined PNNL as a Senior Research Scientist in 1980.

Yufeng Shen
(509) 376-4528
yufeng.shen@pnl.gov
M.S., Dalian Institute of Chemical Physics, Chinese Academy of Science 1990; Ph.D., Brigham Young University 1998; Postdoctoral Fellow, MS&D, 1998-2000; joined PNNL as Senior Research Scientist for MS&D in 2000.

Keqi Tang
(509) 376-2124
keqi.tang@pnl.gov
M.S., University of Science and Technology of China, Hefei, Anhui, P.R. China 1987; M.S. and M.Ph., Yale University 1992; Ph.D., Yale University 1994; Postdoctoral Fellow, Yale University, 1994-1995; Postdoctoral Fellow, PNNL, 1995-1996; Senior Mechanical Engineer, YSI Inc. 1996-1997; Research and Development Scientist, Finnigan MAT, 1997-2000; joined PNNL as Senior Research Scientist for MS&D in 2000.

Timothy D. Veenstra
(509) 376-7233
timothy.veenstra@pnl.gov
B.S., Trent University, Ontario, Canada 1990; Ph.D., University of Windsor, Ontario, Canada 1994; Postdoctoral Fellow, Mayo Clinic/ Foundation, 1995-1997; Visiting Scientist, MS&D, 1998-1999; joined PNNL as Senior Research Scientist for MS&D in 1999.

Science & Engineering Associates

David J. Anderson
(509) 376-7898
david.anderson@pnl.gov
B.S., Eastern Washington University 1987; joined PNNL as a Technician in 1987.

Kim K. Hixson
(509) 373-6170
kim.peden@pnl.gov

B.S., Eastern Oregon University 1997; AWU fellowship, MS&D, 1997-1998; joined PNNL as Science & Engineering Associate for MS&D in 1998.

Nancy G. Isern
(509) 376-1616
nancy.isern@pnl.gov

B.A., Macalester College 1974; M.S., University of Missouri-Columbia 1986; M.S., Syracuse University 1996; joined PNNL as a Science and Engineering Associate for MS&D in 1994.

Ronald J. Moore
(509) 376-2721
ronald.moore@pnl.gov

B.S., Washington State University 1989; joined PNNL as a Technician in 1989.

Jesse A. Sears
(509) 376-7808
jesse.sears@pnl.gov

A.A.S Southeast Community College; joined PNNL as a Science and Engineering Associate for MS&D in 1997.

Visiting Scientists

Olivier Blum

B.S., University of Geneva, Switzerland 1992; Ph.D., University of Geneva, Switzerland 1999; Visiting Scientist, MS&D with Richard D. Smith, 1999-2000.

Evgueni N. Nikolaev

Ph.D., Moscow Institute of Physics and Technology, Moscow 1974; Minor Staff Scientist, The Institute of Chemical Physics Soviet Academy of Sciences, 1974-1979; Senior Staff Scientist, The Institute of Chemical Physics Soviet Academy, 1979-1983; Head of the Laboratory of Magnetic Processes, 1983-1987; Head of the Laboratory of Ion and Molecular Physics, The Institute for Energy Problems of Chemical Physics Russian Academy of Sciences, 1987-present; Visiting Scientist, MS&D with Richard D. Smith, 2000-.

Aleksey V. Tolmachev

Ph.D., Institute of Chemical Physics Academy of Sciences of the USSR 1984; Research Associate, Institute of Chemical Physics, 1976-1987; Senior Staff Scientist, Institute for Energy Problems of Chemical Physics, Russia, 1987-1998; Visiting Scientist, Technical University of Lodz, Poland, 1989-1990; Visiting Scientist, Giessen University, Germany, 1994 and 1997; Visiting Scientist, University of Manitoba, 1995-1996; Visiting Scientist, MS&D with Richard D. Smith, 1998-.

Postdoctoral Fellows

Kim Alving

M.S., Odense University 1994; Ph.D., University of Műnster 1999; Postdoctoral Fellow, MS&D with Richard D. Smith 1999-.

Nicolas H. Angell

B.S., University of North Carolina 1990; Ph.D., Imperial College of Science, Technology, and Medicine, London, UK 2000; Postdoctoral Fellow, MS&D with Richard D. Smith 2000-.

Ken J. Auberry
B.S., University of Tennessee 1993; Ph.D., Purdue University 1999; Postdoctoral Fellow, MS&D with Richard D. Smith, 1999-.

Scott J. Berger
B.S., Philadelphia College of Pharmacy and Science 1991; Ph.D., University of Colorado, Boulder 1998; Postdoctoral Fellow, MS&D with Richard D. Smith, 1998-.

Oleg V. Borisov
M.S., Moscow State University, Moscow, Russia 1992; Ph.D., Wayne State University 1997; Lawrence Berkeley National Laboratory, 1997-1999; Postdoctoral Fellow, MS&D with Richard D. Smith, 2000-.

Garry W. Buchko
B.S., University of Manitoba 1983; M.S., McMaster University 1986; Ph.D., University of Manitoba 1989; Postdoctoral Fellow, Centre d' Etudes Nucleaires de Grenoble, 1990; Postdoctoral Fellow, Cross Cancer Institute 1991-1993; Postdoctoral Fellow, Simon Fraser University 1993-1995; Postdoctoral Fellow, MS&D with Paul D. Ellis, 1995-.

Thomas P. Conrads
B.S., Washington State University, 1993; Ph.D., Ohio State University 1999; Postdoctoral Fellow, MS&D with Richard D. Smith, 1999-.

John R. Cort
B.A., Williams College 1991; Ph.D., University of Washington 1997; Postdoctoral Fellow, MS&D with Paul D. Ellis, 1997-.

Gary W. Daughdrill
B.S., University of Alabama 1992; Ph.D., University of Oregon 1997; Postdoctoral Fellow, MS&D with Paul D. Ellis, 1998-2000; now at University of Idaho, Moscow, Idaho.

Hongying Gao
B.S., Fudan University, Shanghai, P.R. China 1989; M.S., 1993; Ph.D., University of Cincinnati 1997; Postdoctoral Fellow, MS&D with Richard D. Smith, 1998-2000; now at Neurogen Corporation, Branford, Connecticut.

Michael B. Goshe
B.S., Walsh University 1992; Ph.D., Case Western Reserve University 1998; Postdoctoral Fellow, MS&D with Richard D. Smith, 2000-.

Taeman Kim
M.S., Chungbuk National University, Cheongju, S. Korea 1984; Ph.D., McGill University, Montreal, PQ 1997; Postdoctoral Fellow, MS&D with Richard D. Smith, 1998-.

Sang-Won Lee
B.S., Korea University 1990; M.S., Korea University 1992; Ph.D., California Institute of Technology 1999; Postdoctoral Fellow, MS&D with Richard D. Smith, 1999-.

Lingjun Li
B.S., Beijing Polytechnic University, Beijing, China 1992; Ph.D., University of Illinois at Urbana-Champaign 1999; Postdoctoral Fellow, MS&D with Richard D. Smith, 2000-.

Suzana Martinović
B.S., University of Zagreb, Zagreb, Croatia 1991; M.S., University of Zagreb, Zagreb, Croatia 1994; Ph.D., University of Metz, France 1997; Postdoctoral Fellow, ETH Zentrum, Zurich, Switzerland, 1997-1998; Postdoctoral Fellow, MS&D with Richard D. Smith, 1998-.

Christophe D. Masselon
B.S., University of Metz, France 1991; M.S., University of Metz, France 1993; Ph.D., University of Metz, France 1997; Postdoctoral Fellow, ETH Zentrum, Zurich, Switzerland, 1997-1998; Postdoctoral Fellow, MS&D with Richard D. Smith, 1998-.

Kathleen McAteer
B.S., University of Ulster 1990; Ph.D., University of South Carolina 1995; Postdoctoral Fellow, MS&D with Paul D. Ellis, 1996; Postdoctoral Fellow, MS&D with Paul D. Ellis, 1998-.

Ellen Panisko
B.S., Washington State University 1995; Ph.D., University of Texas Health Science Center 2000; Postdoctoral Fellow, MS&D with Timothy D. Veenstra, 2000-.

V. Sergey Rakov
B.S., Moscow Institute of Physics and Technology 1993; Ph.D., University of Delaware 1999; Postdoctoral Fellow, MS&D with Richard D. Smith, 1999-.

Nestor Rodriguez
M.Sc., Venezuelan Institute of Research, Caracas, Venezuela 1986; Ph.D., Stanford University 1999; Postdoctoral Fellow, University of California, 1999-2000; Postdoctoral Fellow, MS&D with Richard D. Smith, 2000-.

Yufeng Shen
M.S., Dalian Institute of Chemical Physics, Chinese Academy of Science 1990; Ph.D., Brigham Young University 1998; Postdoctoral Fellow, MS&D, 1998-2000; now Senior Research Scientist on permanent staff for MS&D.

Eric F. Strittmatter
B.S., Rutgers University 1995; Ph.D., University of California 2000; Postdoctoral Fellow, MS&D with Richard D. Smith, 2000-.

David S. Wunschel
B.S., Clemson University 1991; Ph.D., University of South Carolina 1997; Postdoctoral Fellow, MS&D with Richard D. Smith, 1997-2000; now at PNNL, Richland, WA.

Fan Xiang
M.S., Memorial University of Newfoundland, Canada 1992; Ph.D., Memorial University of Newfoundland, Canada 1997; Postdoctoral Fellow, MS&D with Richard D. Smith, 1998-2000; now at the University of California, Davis, California.

Jingdong Xu
B.S., California State University 1993; M.S., University of Washington 1996; Ph.D., University of Maryland 2000; Postdoctoral Fellow, MS&D with Richard D. Smith, 2000-.

Li-Rong Yu
B.S., Hangzhou Teachers College, Hangzhou, China 1991; M.S., Chinese Academy of Sciences, Wuhan, China 1994; Ph.D., Chinese Academy of Sciences, Shanghai, China 2000; Postdoctoral Fellow, MS&D with Richard D. Smith, 2000-.

Faculty

Rui Zhao
B.S., Northwest University 1977; M.S., Northwest University 1981; Lecturer and Vice Chairman of Chemistry Department, Yanan University, P.R. China, 1982-1988; Research Assistant, Loughbrough University, U.K., 1989-1994; Associate Professor and Vice Chairman of Chemistry Department, Northwest University, P.R. China, 1994-1999; Faculty, MS&D with Richard D. Smith, 1999-.

Laboratory Graduate

Deanna L. Auberry
B.A., University of Tennessee 1993; M.S., Purdue University 2000; AWU fellowship, MS&D with Mary S. Lipton, 2000-.

Undergraduate Students

Jennifer L. Ackerman
Stanford University; worked with David F. Lowry, 2000.

Katherine Ackerman
Calvin College; worked with Mary S. Lipton, 2000.

Nathan M. Brown
Washington State University; worked with Michael K. Bowman, 2000.

Sarah Marsing
College of Eastern Utah; worked with Nancy G. Isern, 2000.

Peter McLean
Hanford High School; worked with Mary S. Lipton, 2000

Hannah E. McMullen
University of Oregon; worked with Mary S. Lipton, 2000.

Eric Merkley
Brigham Young University; worked with Timothy D. Veenstra, 2000.

David Panther
Unaffiliated; worked with Timothy D. Veenstra, 2000.

Heidi Potter
Prosser High School; worked with Timothy D. Veenstra, 2000.

Yvonne G. Rodriguez
Washington State University; worked with Andrew S. Lipton, 2000.

Travis Sloughter
Hanford High School; worked with Mary S. Lipton, 2000.

Randy Studdard
Prosser High School; worked with Timothy D. Veenstra, 2000.

Adam S. Tenforde
Washington State University; worked with Michael A. Kennedy, 2000

Liesl Wegener
Hanford High School; worked with Pamela K. Jensen, 2000.

Susan Wierenga
Prosser High School; worked with Timothy D. Veenstra, 2000.

Elaine Wright
Prosser High School; worked with Timothy D. Veenstra, 2000.

Magnetic Resonance Research Group

Mass Spectrometry Research Group

Publications and Presentations from October 1999

Publications

M. E. Belov, M. V. Gorshkov, H. R. Udseth, G. A. Anderson, and R. D. Smith, "Zeptomole-Sensitivity Electrospray Ionization - Fourier Transform Ion Cyclotron Resonance Mass Spectrometry of Proteins," Analytical Chemistry 72(10), 2271-2279 (2000).

M. E. Belov, M. V. Gorshkov, H. R. Udseth, G. A. Anderson, A. V. Tolmachev, D. C. Prior, R. Harkewicz, and R. D. Smith, "Initial Implementation of an Electrodynamic Ion Funnel Technique with FTICR Mass Spectrometry," Journal of the American Society for Mass Spectrometry 11(1), 19-23 (2000).

M. K. Bowman and A. M. Tyryshkin, "Electron Nuclear Quadruple Resonance for Assignment of Overlapping Spectra," Journal of Magnetic Resonance 144(1), 74-84.

M. K. Bowman, "Intensity of Cross-Peaks In HYSCORE Spectra of S=1/2, I=1/2 Spin Systems," Journal of Magnetic Resonance 144(2), 228-242.

M. K. Bowman, R. I. Samoilova, A. A. Shubin, J. Huttermann, and S. S. Dikanov, "Observation of Two Paramagnetic Species In Electron Transfer Reactions Within Cesium Modified X and Y Zeolites," Chemical Physics Letters 316(5-6), 404-410.

J. E. Bruce, G. A. Anderson, M. V. Gorshkov, A. L. Rockwood, and R. D. Smith, "A Novel High Performance FTICR Cell for Improved Biopolymer Characterization," Journal of Mass Spectrometry 35(1), 85-94 (2000).

J. E. Bruce, G. A. Anderson, M. D. Brands, L. Paša-Tolić, and R. D. Smith, "Obtaining More Accurate FTICR Mass Measurements Without Internal Standards Using Multiply Charged Ions," Journal of the American Society for Mass Spectrometry 11(5), 416-421 (2000).

J. E. Bruce, L. Paša-Tolić, R. D. Smith, D. R. Goodlet, R. Aebersold, "Protein Identification with a Single Accurate Mass of a Cysteine-Containing Peptide and Constrained Database Searching," Journal of the American Chemical Society 72(6), 1112-1118 (2000).

G. W. Buchko, N. J. Hess, and M. A. Kennedy, "Cadmium Mutagenicity and Human Nucleotide Excision Repair Protein SPA: CD, EXAFS, and 1H/15N NMR Spectroscopic Studies of the Zinc (II) and Cadmium (II) Associated Protein," Carcinogenesis 21(5), 1051-1057.

G. W. Buchko, G. W. Daughdrill, R. de Lorimier, N. G. Isern, J. M. Lingbeck, J. S. Taylor, M. S. Wold, M. Gochin, L. D. Spicer, D. F. Lowry, and M. A. Kennedy, "Interactions of Human Nucleotide Excision Repair Protein XPA With DNA and RPA70C327: Chemical Shift Mapping and 15N NMR Relaxation Studies," Biochemistry (American Chemical Society) 38, 15116-15128.

G. W. Buchko, N. J. Hess, V. Bandura, S. S. Wallace, and M. A. Kennedy, "Spectroscopic Studies of Zinc(II) and Cobalt(II) Associated Escherichia coli Formidopyrimidine-DNA Glycosylase: Extended X-ray Absorption Fine Structure Evidence for a Metal-Binding Domain," Biochemistry 39, 12441-12449.

G. W. Buchko, N. J. Hess, and M. A. Kennedy, "Summary of EXAFS Studies On The Zn (II), Co (II), and Cd (II) Associated Minimal DNA-Binding Domain," Protein and Peptide Letters 7(1), 49-56.

G. W. Buchko, A. Rozek, P. Kanda, M. A. Kennedy, and R. J. Cushley, "Structural Studies of a Baboon (Papio sp) Plasma Protein Inhibitor of Cholesteryl Ester Transferase," Protein Science 8, 1548-1558.

H. M. Cho, "Hydrogen Bonding Interaction In Phase A (Mg7Si2O8(HO06) At Ambient and High Pressure," Physics and Chemistry of Minerals 27(4), 225-233.

H. M. Cho, "Off-Resonance Multiple-Pulse Dynamics In Solid-State NMR Spectroscopy: A Revised Coherent Averaging Theory Analysis," Journal of Magnetic Resonance 141, 164-179.

D. Christendat, A. Yee, A. Dharamsi, Y. Kluger, A. Savchenko, J. R. Cort, V. Booth, C. D. Mackereth, V. Saridakis, I. Ekiel, G. Kozlov, K. L. Maxwell, N. Wu, L. P. McIntosh, K. Gehring, M. A. Kennedy, A. R. Davidson, E. F. Pai, M. Gerstein, A. M. Edwards, and C. H. Arrowsmith, "Structural Proteomics Of An Archaeon," Nature Structural Biology 7(10), 903-909.

T. P. Conrads, G. A. Anderson, T. D. Veenstra, L. Paša-Tolić, and R. D. Smith, "Utility of Accurate Mass Tags for Proteome-Wide Protein Identification," Analytical Chemistry 72(14), 3349-3354 (2000).

J. R. Cort, E. V. Koonin, P. A. Bash, and M. A. Kennedy, "A Phylogenetic Approach To Target Selection for Structural Genomics: Solution Structure of YciH," Nucleic Acids Research 27(20), 4018-4027.

J. R. Cort, A. Yee, A. M Edwards, C. H. Arrowsmith, and M. A. Kennedy, "Structure-Based Functional Classification Of Hypothetical Protein MTH538 From Methanobacterium thermoautotrophicum," Journal of Molecular Biology 301(1), 189-203.

S. H. Elder, F. Cot, Y. Su, S. M. Heald, A. M. Tyryshkin, M. K. Bowman, Y. Gao, A. G. Joly, M. Balmer, A. C. Ebeling, K. A. Magrini, and D. M. Blake, "The Discovery and Study of Nanocrystalline TiO2-(MoO3) Core-Shell Materials," Journal of the American Chemical Society 122(21), 5138-5146.

B. Feng and R. D. Smith, "A Simple Nanoelectrospray Arrangement with Controllable Flowrate for Mass Analysis of Sub-microliter Samples," Journal of the American Society for Mass Spectrometry 11(1), 94-99 (2000).

H. Gao, Y. Shen, T. D. Veenstra, R. Harkewicz, G. A. Anderson, J. E. Bruce, L. Paša-Tolić, and R. D. Smith, "Two-Dimensional Electrophoretic/ Chromatographic Separations Combined with Electrospray Ionization FTICR Mass Spectrometry for High Throughput Proteome Analysis," Journal of Microcolumn Separations 12(7), 383-390 (2000).

J. Z. Hu, M. S. Solum, R. A. Wind, B. L. Nilsson, M. A. Peterson, R. J. Pugmire, and D. M. Grant, "^{1}H and ^{15}N Dynamic Nuclear Polarization Studies of Carbazole," Journal of Physical Chemistry A 104, 4413-4420.

P. K. Jensen, G. A. Anderson, M. S. Lipton, S. Martinovic, L. Paša-Tolić, K. K. Peden, and R. D. Smith, "Mass Spectrometric Detection for Capillary Isoelectric Focusing Separations of Complex Protein Mixtures," Electrophoresis 21(7), 1372-1380 (2000).

K. A. Keating, J. D. Myers, R. A. Bair, P. D. Ellis, J. G. Pelton, and D. E. Wemmer, "Development and Use Of A Virtual NMR Facility," Journal of Magnetic Resonance 143, 172-183.

M. A. Kennedy and K. McAteer, "NMR Evidence For Base Dynamics At All TpA Steps In DNA," Journal of Biomolecular Structure and Dynamics 17(6), 1001-1009.

T. Kim, A. V. Tolmachev, R. Harkewicz, D. C. Prior, G. A. Anderson, H. R. Udseth, R. D. Smith, T. H. Bailey, V. S. Rakov, and J. H. Futrell, "Design and Implementation of a New Electrodynamic Ion Funnel," Analytical Chemistry 72, 2247-2255 (2000).

T. Kim, A. V. Tolmachev, R. Harkewicz, D. C. Prior, G. A. Anderson, H. R. Udseth, R. D. Smith, T. H. Bailey, V. S. Rakov, and J. H. Futrell, "Properties of a New Electrodynamic Ion Funnel," Analytical Chemistry 72(10), 2247-2255 (2000).

M. T. Krahmer, J. J. Walters, K. F. Fox, A. Fox, K. E. Creek, L. Pirisi, D. S. Wunschel, R. D. Smith, D. L. Tabb, and J. R. Yates, "MS for Identification of Single Nucleotide Polymorphisms and MS/MS for Discrimination of Isomeric PCR Products," Analytical Chemistry 72, 4033-4040 (2000).

C. S. Maier, X. Yan, M. E. Harder, M. I. Schimerlik, M. L. Deinzer, L. Paša-Tolić, and R. D. Smith, " Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometric Analysis of the Recombinant Human Macrophage Colony Stimulating Factor B and Derivatives," Journal of the American Society of Mass Spectrometry 11, 237-243 (2000).

S. Martinovic, L. Paša-Tolić, C. D. Masselon, P. K. Jensen, C. L. Stone, and R. D. Smith, "Characterization of Human Alcohol Dehydrogenase Isoenzymes by Capillary Isoelectric Focusing-Mass Spectrometry," Electrophoresis 21(12), 2368-2375 (2000).

C. D. Masselon, G. A. Anderson, R. Harkewicz, J. E. Bruce, L. Paša-Tolić, and R. D. Smith, "Accurate Mass Multiplexed Tandem Mass Spectrometry for High-Throughput Polypeptide Identification from Mixtures," Analytical Chemistry 72(8), 1918-1924 (2000).

K. McAteer and M. A. Kennedy, "Evidence for Large Amplitude Base Dynamics at All TpA Steps in DNA," Journal of Biological Structure and Dynamics 17, 1001-1009.

J. H. Miller, K. R. Minard, R. A. Wind, G. A. Orner, L. B. Sasser, and R. J. Bull, "In vivo MRI Measurements of Tumor Growth Induced by Dichloroacetate: Implications for Mode of Action," Toxicology 145, 115-125.

Y. Shen, F. Xiang, T. D. Veenstra, N. Fung, and R. D. Smith "High-Resolution Capillary Isoelectric Focusing of Complex Protein Mixtures from Lysates of Microogranisms," Analytical Chemistry 71, 5348-5353 (1999).

Y. Shen and R. D. Smith, "High-Resolution Capillary Isoelectric Focusing of Proteins Using Highly Hydrophilic Substituted Cellulose-Coated Capillaries," Journal of Microcolumn Separations 12(3), 135-141 (2000).

Y. Shen, S. J. Berger, G. A. Anderson, and R. D. Smith, "High-Efficiency Capillary Isoelectric Focusing of Peptides," Analytical Chemistry 72(9), 2154-2159 (2000).

A. V. Tolmachev, H. R. Udseth, and R. D. Smith, "The Charge Capacity Limitations of Radio Frequency Ion Guides for Improved Ion Accumulation and Trapping in Mass Spectrometry," Analytical Chemistry 72(5), 970-978 (2000).

A. V. Tolmachev, H. R. Udseth, and R. D. Smith, "Radial Stratification of Ions as a Function of Mass to Charge Ratio in Collisional Cooling Radio Frequency Multipoles Used as Ion Guides or Ion Traps," Rapid Communications in Mass Spectrometry 14(20), 1907-1913 (2000).

A. M. Tyryshkin, A. G. Roberts, and M. K. Bowman, "Inhibitory Copper Binding Site On The Spinach Cytochrome b6f Complex: Implications For Qo Site Catalysis," Biochemistry (American Chemical Society) 39(12), 3285-3296.

T. D. Veenstra, C. Zhang, F. Xiang, L. Paša-Tolić, G. A. Anderson, and R. D. Smith, "Stepwise Mobilization of Focused Proteins in Capillary Isoelectric Focusing Mass Spectrometry," Analytical Chemistry 72(7), 1462-1468 (2000).

T. D. Veenstra, S. Martinovic, G. A. Anderson, L. Paša-Tolić, and R. D. Smith, "Proteome Analysis using Selective Incorporation of Isotopically Labeled Amino Acids," Journal of the American Society for Mass Spectrometry 11(1), 78-82 (2000).

J. Y. Wen, Y. Lin, F. Xiang, D. W. Matson, H. R. Udseth, and R. D. Smith, "Microfabricated Isoelectric Focusing Devices for Direct Electrospray Ionization-Mass Spectrometry," Electrophoresis 21(1), 191-197 (2000).

R. A. Wind, "1H DNP at 1.4 T of Water Doped With A Triarylmethyl-Based Paramagnetic Label," Journal of Magnetic Resonance 141(2), 347-354.

R. A. Wind, S. Bai, J. Z. Hu, M. S. Solum, D. M. Grant, and C. R. Yonker, "H1 Dynamic Nuclear Polarization In Supercritical Ethylene At 1.4 T3/01," Journal of Magnetic Resonance 143(1), 233-239.

D. S. Wunschel, L. Paša-Tolić, B. Feng, and R. D. Smith, "Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Analysis of Large Polymerases Chain Reaction Products," Journal of the American Society for Mass Spectrometry 11, 333-337 (2000).

D. S. Wunschel, L. Paša-Tolić, B. Feng, and R. D. Smith, "ESI-FTICR Analysis of Large PCR products," Journal of the American Society for Mass Spectrometry 11, 333-337 (2000).

F. Xiang, G. A. Anderson, T. D. Veenstra, M. S. Lipton, and R. D. Smith, "Characterization of Microorganisms and Biomarker Development from Global ESI-MS/MS Analyses of Cell Lysates," Analytical Chemistry 72(11), 2475-2481 (2000).

J. Yang, K. McAteer, L. A. Silks, R. Wu, N. G. Isern, C. J. Unkefer, and M. A. Kennedy, "A Comprehensive Approach for Accurate Measurement of Proton-Proton Coupling Constants in the Sugar Ring of DNA," Journal of Magnetic Resonance 146, 260-276.

Patents

R. D. Smith and S. A. Shaffer, "Method and Apparatus for Directing Ions and Other Charged Particles Generated at Near Atmospheric Pressures into a Region Under Vacuum," U.S. Patent 6,107,628 (granted August 22, 2000).

Presentations

M. E. Belov, E. N. Nikolaev, A. V. Tolmachev, and R. D. Smith, "Ion Instability in a Linear Segmented Quadrupole Ion Trap in the Presence of High Space Charge," 48th Annual ASMS Conference, Long Beach, California, June 2000.

M. E. Belov, E. N. Nikolaev, G. A. Anderson, K. J. Auberry, and R. D. Smith, "Selective Ion Accumulation Based on RF-Only Resonant Dipolar Excitation in the Presence of High Space Charge," 48th Annual ASMS Conference, Long Beach, California, June 2000.

M. E. Belov, E. N. Nikolaev, G. A. Anderson, M. V. Gorshkov, T. H. Bailey, H. R. Udseth, and R. D. Smith, "Selective External Accumulation of Electrospray-Generated Ions Coupled to Fourier Transform Ion Cyclotron Resonance Mass Spectrometry," 48th Annual ASMS Conference, Long Beach, CA, June 2000.

O. L. Blum, D. S. Wunschel, M. S. Lipton, and R. D. Smith, "Measuring and Understanding the Differences in DNA Damage Resulting from Normal Uxidative Processes and Low Levels of Ionizing Radiation," 48th Annual ASMS Conference, Long Beach, Califonia, June 2000.

L. J. Bond, M. S. Good, E. J. Ackerman, R. A. Wind, G. R. Holtom, and B. R. Tiltmann, "Development of a Combined Acoustic Microscopy and Magnetic Resonance Microscopy and/or Optical Confocal Microscopy Instrument," Environmental Health Initiative Technical Council Review, Richland, Washington, August 3, 2000.*

M. K. Bowman, "DEER Measurements of Inhomogeneous Distributions of Radicals In Irradiated Materials, Rocky Mountain Conference on Analytical Chemistry, Broomfield, Colorado, July 30, 2000.*

M. K. Bowman, "Long-Range Interactions In the Cytochrome b6f Complex," Chemistry Seminar, Tucson, Tempe, Arizona, April 6, 2000.*

M. K. Bowman, S. Rao, and D. Kramer, "Pulsed EPR and Heavy Metal Inhibition In the Cytochrome b6f Complex," Chemical Physics Seminar, Novosibirsk, Russia, December 12, 1999.*

M. K. Bowman, "Pulsed EPR Spectroscopy of Metal Centers In Proteins," Bioinorganic Seminar, Tucson, Arizona, April 11, 2000.*

H. M. Cho, "Magnetic Resonance Spectroscopy of Light and Heavy Nuclides In the Solid-State," Conference Seminar, Amherst, Massachusetts, April 5, 2000.*

H. M. Cho, "Magnetic Resonance Spectroscopy of Light and Heavy Nuclides In the Solid-State," Summer Conference, Amherst, Massachusetts, April 7, 2000.*

H. M. Cho, "Proton Chemical Shift Measurements and Hydrogen Bonding In Inorganic Solids," 35 th American Chemical Society Regional Meeting, Ontario, California, October 6, 1999.*

H. M. Cho, "Proton Chemical Shift Measurements and Hydrogen Bonding In Inorganic Solids," 35 th American Chemical Society Regional Meeting, Ontario, California, November 1, 1999.

O. T. Farmer, G. C. Eiden, C. J. Barinaga, D. W. Koppenaal, "Collision Cell ICPMS - Techniques and Applications," Federation of Analytical Chemistry and Spectroscopy Societies, Nashville, Tennessee, September 26, 2000.*

R. Harkewicz, G. A. Anderson, C. D. Masselon, L. Paša-Tolić, D. C. Prior, H. R. Udseth, and R. D. Smith, "FTICR Mass Spectrometry Employing Data-Dependent External m/z Ion Selection and Accumulation," 48th Annual ASMS Conference, Long Beach, California, June 2000.

J. A. Horner, K. K. Peden, and R. D. Smith, "Rapid Ultra Sensitive Characterization of Microorganisms Using ESI-MS/MS," National FACSS Meeting, Vancouver, British Columbia, October 1999.

D. W. Hoyt, S. D. Burton, J. J. Ford, A. S. Lipton, D. N. Rommereim, and P. D. Ellis, "High Field Magnetic Resonance Facility (HFMRF)," Environmental Nuclear Magnetic Resonance Conference, Pacific Grove, California, April 9, 2000.

P. K. Jensen, K. K. Peden, M. S. Lipton, G. A. Anderson, Y. A. Gorby, M. F. Romine, T. D. Veenstra, and R. D. Smith, "Rapid and Ultra-Sensitive Analysis of Global Protein Expression in Shewanella Putrefaciens Using Fourier Transform Ion Cyclotron Resonance Mass Spectrometry," 48 th Annual ASMS Conference, Long Beach, California, June 2000.

M. A. Kennedy, "Magnetic Resonance Strategies for Structural Genomics," Rutgers University, Piscataway, New Jersey, November 3, 1999.

M. A. Kennedy, K. McAteer, N. G. Isern, R. Michalcyzk, J. G. Schmidt, L. A. Silks, and C. Unkefer, "Measurement of Residual Dipolar Couplings in 15N-Labeled and 13C-Labeled DNA Aligned In Liquid Crystalline Bicelles Using NMR Spectroscopy, 41st ENC, Pacific Grove, California, April 9, 2000.

M. A. Kennedy, "NMR Based Structural Genomics," State University of New York, Albany, New York, March 16, 2000.*

M. A. Kennedy, "Structure-Based Classification of Putative DNA Repair Proteins from Deinococcus radiodurans," Environmental Molecular Sciences Symposia and 1st Users Meeting, Richland, Washington, June 21-23, 2000.

M. A. Kennedy, J. R. Cort, D. W. Hoyt, A. M. Edwards, and C. H. Arrowsmith, "Structure-Based Functional Classification of DNA Repair Proteins From D. radiodurans," EMSL User Symposium, Richland, Washington, June 23, 2000.*

M. A. Kennedy, "Structure-Based Functional Classification of Hypothetical Proteins," 2nd EMSL Workshop on Structural Genomics, Richland, Washington, July 28-29, 2000.*

M. A. Kennedy, J. R. Cort, A. M. Edwards, and C. A. Arrowsmith, "Structure-Based Functional Classification of Proteins for Structural Genomics," Structural Genomics: From Gene to Structure to Function, Cambridge, Great Britain, September 20, 2000.

D. W. Koppenaal, "Focal Plane Detectors In Atomic Mass Spectrometry-Past/Present/Future," Federation of Analytical Chemistry and Spectroscopy Societies, Nashville, Tennessee, September 25, 2000.*

D. W. Koppenaal, "Ion-Molecule Chemistry In Ion Traps and Ion Guides For ICPMS, Plasma Spectrochemistry Winter Conference 2000, Ft. Lauderdale, Florida, January 14, 2000.*

D. W. Koppenaal, G. C. Eiden, O. T. Farmer, and C. J. Barinaga, "Reaction Chemistry in Elemental Mass Spectrometry: Putting Chemistry To Work," American Chemical Society National Meeting, Washington, D.C., August, 21, 2000.*

D. W. Koppenaal, G. C. Eiden, O. T. Farmer, and C. J. Barinaga, "Reaction Chemistry in Elemental Mass Spectrometry: Putting Chemistry To Work," 10 th Biennial National Atomic Spectrometry Symposium, Sheffield, Great Britain, July 18, 2000.*

M. Liezers, O. T. Farmers, D. W. Koppenaal, and J. W. Grate, "Actinide Detection and Speciation at Attogram Concentrstions Using Quadrupole ICP/MS," Federation of Analytical Chemistry and Spectroscopy Societies (FACSS) Annual Meeting, Vancouver British Columbia, Canada, October 24, 1999.

Y. Lin, K. Tang, D. W. Matson, J. A. Horner, and R. D. Smith, "Microfluidic Devices for Rapid Sample Processing/Mass Spectrometric Characterization of Microbes," DOE Biomedical Engineering Contractors Meeting, Albuquerque, New Mexico, May 2000.*

Y. Lin, K. Tang, P. M. Martin, D. W. Matson, and R. D. Smith, "Microfluidic Devices for Rapid Sample Processing/Mass Spectrometric Characterization of Proteins," BioMEMS & Biomedical Nanotechnology World 2000, Columbus, Ohio, September 2000.

A. S. Lipton, J. A. Sears, and P. D. Ellis, "Solid-State NMR of Biologically Relevant Metals (Zn, Mg and Ca)," 41st ENC, Pacific Grove, California, April 9, 2000.

A. S. Lipton, "Solid-State NMR of Metals of Biological Interest, Rocky Mountain Conference on Analytical Chemistry, Denver, Colorado, August 4, 2000.

M. S. Lipton, T. D. Veenstra, and R. D. Smith, " Probing the Proteome with Capillary Isoelectric Focusing-ESI-FTICR Mass Spectrometry," The Department of Energy's Natural and Accelerated Bioremediation Research (NABIR) Program, Reston, Virginia, January 2000.

P. D. Majors, E. J. Ackerman, G. R. Holtom, K. R. Minard, and R. A. Wind, "Combined Optical and Magnetic Resonance Microscopy for Cellular Research, International Society for Magnetic Resonance in Medicine, Berkeley, California, March 27, 2000.

P. D. Majors, E. J. Ackerman, G. R. Holtom, K. R. Minard, and R. A. Wind, "First Experiences with Combined Confocal and Magnetic Resonance Microscopy," Magnetic Resonance Microscopy for Biological Research Symposium, EMSL 2000 User meeting, Richland, Washington, June 21-24, 2000.*

S. Martinovic, S. J. Berger, and R. D. Smith, "Characterization of Lentil Lectin Non-covalent Complex Isoforms by on-line CIEF-FTICR Mass Spectrometry," 48th Annual ASMS Conference, Long Beach, California, June 2000.

C. D. Masselon, A. V. Tolmachev, V. S. Rakov, L. Paša-Tolić, R. Harkewicz, G. A. Anderson, E. N. Nikolaev, and R. D. Smith, "Sub-ppm Local Frequency Shifts in FTICR Mass Spectrometry," 48th Annual ASMS Conference, Long Beach, California, June 2000.

K. R. Minard, "High-Field Performance of Proton NMR Microscopy In Cellular Research: Is High-Field the Holy Grail, EMSL 2000, June 24, 2000.*

L. Paša-Tolić, G. A. Anderson, D. J. Anderson, R. Harkewicz, M. S. Lipton, C. D. Masselon, K. K. Peden, Y. Shen, T. D. Veenstra, and R. D. Smith, "Probing Deinococcus radiodurans Proteome by Fourier Transform Ion Cyclotron Resonance Mass Spectrometry," 48th Annual ASMS Conference, Long Beach, California, June 2000.

Y. Shen and R. D. Smith, "High-Efficiency Capillary Isoelectric Focusing for Proteomics," HPLC-2000 Symposium, Seattle, Washington, June 2000.*

R. D. Smith, L. Paša-Tolić, G. A. Anderson, P. K. Jensen, and S. Martinovic, "Advanced Methods for the Study of Proteomes by Mass Spectrometry," AAPS Annual Meeting and Exposition, New Orleans, Lousiana, November 1999.*

R. D. Smith, L. Paša-Tolić, M. S. Lipton, P. K. Jensen, G. A. Anderson, and T. D. Veenstra, "Rapid Quantitative Measurements of Proteomes," Proteomics Conference, Stony Brook, New York, May 2000.*

R. D. Smith, L. Paša-Tolić, M. S. Lipton, G. A. Anderson, and T. D. Veenstra, "Quantitative Measurements of Proteomes for Biomarker Development," International Conference on Arctic Development, Pollution, and Biomarkers of Human Health, Anchorage, Alaska, May 2000.*

R. D. Smith, L. Paša-Tolić, M. S. Lipton, P. K. Jensen, G. A. Anderson, and T. D. Veenstra, "Rapid Quantitative Measurements of Proteomes," 4 th Genome to Proteome Meeting, Siena, Italy, September 2000.

A. V. Tolmachev, R. Harkewicz, C. D. Masselon, G. A. Anderson, V. S. Rakov, L. Paša-Tolić, E. N. Nikolaev, M. E. Belov, H. R. Udseth, and R. D. Smith, "Radial Stratification of Ions as a Function of m/z Ratio in Collisional Cooling Radio Frequency Multipoles Used as Ion Guides or Ion Traps," 48 th Annual ASMS Conference, Long Beach, California, June 2000.

T. D. Veenstra, S. Martinovic, G. A. Anderson, L. Paša-Tolić, and R. D. Smith, "High-Throughput Proteome Analysis Using Selective Incorporation of Isotopically Labeled Amino Acids," Asilomar Conference on Mass Spectrometry, Pacific Grove, California, October 1999.

R. A. Wind, E. J. Ackerman, D. G. Cory, G. R. Holtom, P. D. Majors, K. R. Minard, and X.-W. Tang, "Integrated Optical and Magnetic Resonance Microscopy for Cellular Research," 41 st ENC, Asilomar, California, April 9-14, 2000.

R. A. Wind and K. R. Minard, "Magnetic Resonance Microscopy At PNNL," Magnetic Resonance Workshop, Portland, Oregon, April 24, 2000.*

R. A. Wind, "Prospects of Magnetic Resonance Microscopy Combined With Dynamic Nuclear Polarization," Solid-State NMR Users Symposium, Denver, Colorado, January 7, 2000.*

N. Xu, Y. Lin, D. W. Matson, L. Paša-Tolić, J. E. Bruce, and R. D. Smith, "Microfabricated Devices for Biological Sample Cleanup and Separation in ESI-MS," University of Idaho, Boise, Idaho, August 2000.*

Honors and Recognition

Richard D. Smith was elected as Battelle Inventor of the Year.

Energy 100 Award presented to Richard D. Smith and group to honor one of the best scientific and technological accomplishments within DOE during the 20th century.

Mentor of the Year Awarded to Richard D. Smith for the mentoring of Post-docs.

Where MS&D Fits in PNNL

Pacific Northwest National Laboratory (L. J. Powell, *Director*)
- Energy Science and Technology Division (D. P. McConnell, *Associate Laboratory Director*)
- **Fundamental Sciences Division** (J. W. Rogers Jr., *Associate Laboratory Director*)
  - Atmospheric Sciences and Global Change Resources (W. T. Pennell)
  - Biogeochemistry Resources (H. J. Bolton Jr.)
  - Molecular Biosciences Resources (J. G. Pounds)
  - Statistics Resources (B. A. Pulsipher)
  - **W. R. Wiley Environmental Molecular Sciences Laboratory (EMSL)** (J. H. Futrell, *Director*)
    - Chemical Structure and Dynamics (CS&D) (S. D. Colson)
    - Interfacial Chemistry & Engineering (IC&E) (J. W. Grate, *interim*)
    - **Macromolecular Structure and Dynamics (MS&D)** (D. W. Koppenaal)
    - Theory, Modeling, and Simulation (TM&S) (D. A. Dixon)
    - Environmental Dynamics and Simulation (ED&S) (J. M. Zachara)
    - Computing and Information Sciences (C&IS) (M. J. Peterson, *acting*)
- Environmental Technology Division (W. J. Apley, *Associate Laboratory Director*)
- National Security Division (M. Kluse, *Associate Laboratory Director*)