ELSEVIER

International Journal of Mass Spectrometry 200 (2000) 509–544

International Journal of Mass Spectrometry

Evolution of ESI–mass spectrometry and Fourier transform ion cyclotron resonance for proteomics and other biological applications

R.D. Smith*

Environmental Molecular Science Laboratory, Mail Stop K8-98, Pacific Northwest National Laboratory, P.O. Box 999, Richland, WA 99352, USA

Received 7 June 2000; accepted 5 September 2000

Abstract

A perspective is presented centered on the author's contributions to developments involving electrospray ionization–mass spectrometry, capillary electrophoresis–mass spectrometry, and Fourier transform ion cyclotron resonance (FTICR) mass spectrometry and their applications to biological systems, with a special emphasis on the study of noncovalent complexes and proteomics. (Int J Mass Spectrom 200 (2000) 509–544) © 2000 Elsevier Science B.V.

Keywords: Electrospray ionization; FTICR mass spectrometry; Proteomics; Capillary electrophoresis; Proteins

1. Introduction

The importance of mass spectrometry to the future of biological research is now clearly evident to all but the most skeptical observers. The last few years have seen mass spectrometry embraced by prominent leaders of the biological research community, with its role increasingly cited as an enabling tool for the future of biological research.

This great expansion in the role of mass spectrometry is being fueled by improvements in instrumentation, a rapid increase in the performance/cost ratio of instrumentation, and advances in on-line separation capabilities, computer control, and data processing. But without a doubt, the most important and crucial development has been the implementation of ionization methods for biopolymers that are broadly applicable, gentle, and enormously sensitive. The development of matrix-assisted laser desorption ionization (MALDI) and electrospray ionization (ESI) and their subsequent refinement and use for biological applications have been key. MALDI and ESI overlap significantly in their scope of application, and while the impact of ESI has been greater to this point, it is clear that both will continue to have expanding roles in the foreseeable future.

Already, ESI has substantially changed the practice of mass spectrometry because of its sensitivity, broad utility, facility for interfacing with separations, and applicability to large biomolecular species. In the last decade, instrumentation incorporating ESI sources has come to dominate many areas of mass spectrometry. One of the most confident predictions I can

* E-mail: rd_smith@pnl.gov

PII S1387-3806(00)00352-3

make is that the next decade will see many seminal developments and new understandings from biological research that will derive from, or involve the use of, ESI–mass spectrometry (ESI-MS).

This article is written from the perspective of someone whose research has explored and used (and contributed to the development of) ESI-MS, and these efforts will be the central focus here. I will briefly recap some of the developments I consider key and will emphasize the ongoing developments and new types of biological applications that are arising from the combination of ESI with the power of Fourier transform ion cyclotron resonance (FTICR) mass spectrometry, all with particular emphasis on the contributions from my group at Pacific Northwest National Laboratory (PNNL). In particular, I will conclude with a more detailed description of some of my group's present efforts that are specifically aimed at the use of ESI-FTICR for proteomics and that build on many of our earlier efforts.

My strong belief is that the developments and contributions to date in the field of mass spectrometry provide only the slightest hint of the enormous impact its use in biological research will have over the next few decades. These developments will likely be so large and so significant as to affect nearly everyone on our planet in one fashion or another.

2. A personal perspective and emphasis

It is necessary to emphasize that this will not be a balanced account of the development of ESI, FTICR, or their applications to biological research. This personal perspective (as well as space, energy, and time) necessarily precludes such balance, and many reviews on these subjects with varying degrees of objectivity have been published elsewhere at great length and from many different viewpoints. This is a more personal account. As such, the enormous contributions of the inventors and developers of ESI-MS (M. Dole; J. B. Fenn et al.) and FTICR (Comisarow and Marshall), for example, will not be a central focus. This will also be the case for the numerous and often seminal relevant contributions of K. Biemann, F. W. McLafferty, and others to the field of biological mass spectrometry.

It must also be emphasized that, particularly in recent years, the research and accomplishments I refer to are the products from the entire group of talented researchers that I have the pleasure of leading. A personal account such as this is obviously biased by one's individual blind spots and quirks (e.g., one's ego), and though I have made every effort to be correct with regard to what is presented as fact, I cannot be comprehensive, and I apologize in advance for perceived slights or omissions. I feel compelled to offer the following disclaimer: The views expressed in this article are only the views of the author. Although the intent is to be completely truthful, this does not mean that some of the more subjective recollections will not differ from those of others.

Looking back, it is usually quite easy to create a credible account and internally consistent story about how you got from there to here. The less subjective truth is, of course, much different, at least for me. From what (I hope) is an approximate midpoint in my career, the large majority of which has involved mass spectrometry, the general direction is clear to me. The detailed path followed, however, is something quite different. It seems that small things, decisions, and unexpected opportunities can cause either major diversions or leaps ahead in the direction of desired travel. Then there are, perhaps more than anything else, the sudden insights or ideas that appear while admiring (or puzzling over) the contributions of others. All this is leavened for me by the daily interactions with coworkers and postdoctoral associates and their contributions and insights, and constrained by the more prosaic concerns related to securing funding to support one's research.

3. Electrospray ionization

Although one can quibble, in my view ESI was developed by Malcolm Dole in the 1960s. Dole suffered the not uncommon fate of being so significantly ahead of his time that he could not obtain funding for the actual implementation of ESI with a

mass spectrometer [1,2]. Thus, although Dole's research with ESI continued into the 1970s, its possible significance for mass spectrometry went essentially unrecognized until the early 1980s.

That is not to say that the need for improved ionization methods was not recognized at that time. In fact, the time from the 1970s through the early 1980s was a period in which many ideas were being explored for the effective ionization of biopolymers as well as liquid chromatography (LC)–MS coupling. When I joined PNNL in 1976, it was already obvious that an area having enormous potential for future growth involved the biological application of mass spectrometry. At PNNL I was able to initiate several lines of research (not all of which involved mass spectrometry; in fact, my major funded project focused on the post-combustion chemistry of coal and the high-temperature chemistry of aromatic hydrocarbons). As a secondary project/hobby, in the late 1970s I designed and constructed a moving ribbon interface between a liquid chromatograph and a mass spectrometer that employed several new features, including spray deposition of the LC effluent and the use of a fast ion-bombardment ionization method [3,4]. The basic idea was to move the ribbon fast enough to limit ion flux damage to the surface and ionize intact molecular ions (earlier work had shown that under static SIMS conditions, with the much lower ion currents for surface bombardment, intact molecular ions could be produced).

However, the spectra produced proved to be highly complex, and it was clear that this approach was not the magic bullet that would enable broadly effective biological analyses. In addition, in the very early 1980s, both fast atom bombardment (FAB) for biopolymer analyses and thermospray ionization for LC-MS coupling appeared. But while these methods enormously increased the applicability of mass spectrometry for biological analyses, the reality was that, on an absolute basis, the impact of mass spectrometry on the overall enterprise of biological research remained negligible. Although FAB and thermospray, along with several other advances from that period, were important steps in the right direction, it was also clear that a huge gap existed between what was needed to deal with the complexities of biological samples and what could be accomplished using the existing technology and was not going to be bridged quickly. The gap, in my mind, needed to be filled by technology with much better sensitivity, combined with better separations to address the complexity of biological mixtures, and more effective ionization (ideally independent of compound molecular weight, volatility, polarity, etc.).

In the early 1980s, in a collaboration with Milton Lee at Brigham Young University, who was then developing capillary supercritical fluid chromatography (SFC), I developed the combination of SFC with mass spectrometry (SFC-MS) [5,6]. Capillary SFC provided much faster and higher resolution separations than were feasible with liquid chromatography and appeared to offer a route for solvating intractable thermally labile and nonvolatile biological compounds. With colleague Harold Udseth, we subsequently refined the direct fluid injection SFC-MS interface that allowed compounds solvated by the supercritical fluid to be transferred to the gas phase by a rapid decompression and expansion of the fluid, where conventional ionization and mass spectrometric analysis were feasible [7]. My group's efforts to refine the SFC-MS interface and improve methods for SFC analyses continued through 1988 [8], leading to several spin-offs (e.g., the use of the same rapid fluid expansion process used in the SFC-MS interface for the production of thin films and fine powders of essentially any compound that could be dissolved in a supercritical fluid) [9–11].

It was evident from the start of my work with supercritical fluids that the major constraint for their applicability to biological research was the extremely low solubility of biopolymers in the widely used fluids (e.g., CO_2). Thus, one emphasis of our research was the exploration of alternate more polar fluid systems, such as those based on ammonia [12,13]. However, after much effort, the combination of the higher temperatures required and still insufficient solubility for most biopolymers convinced me that success was unlikely. As a final shot, I conceived an approach involving the creation of reverse micelles

and microemulsions in supercritical fluids [14] and my Group demonstrated the capability to create and apply these systems for separations and reaction processes [15–18]. This new class of solvent systems allowed, for the first time, the solvation of any substance that could be dissolved by ordinary liquids and aqueous systems while simultaneously preserving nearly all the attractive properties of supercritical fluids. (As an aside, this class of solvents has recently been adopted for applications as prosaic as dry cleaning.) Unfortunately, these reverse micelle solutions required the use of surfactants and were incompatible with interfaces for mass spectrometry. After a period in which we tried unsuccessfully to create surfactantless micelle supercritical fluid solvent systems, I turned my focus to other areas.

My work with capillary SFC had convinced me of the power and promise of microcolumn (i.e., capillary) separations, and the utility of high-resolution separations in conjunction with MS had already driven a major growth of capillary GC-MS applications (e.g., for fossil fuel analyses, another area where sample complexity is very high). It also made me a student of research in the area of capillary separations.

In the early 1980s, Professor Jim Jorgensen of the University of North Carolina demonstrated remarkable high-resolution separations of peptides and proteins by the new technique of capillary electrophoresis (CE), a method in which a high-voltage difference is provided across a capillary immersed in liquid reservoirs at both ends [19]. The electrophoretic separation process allowed biopolymer samples having microliter to nanoliter volumes to be quickly separated. I followed this early work and pondered how it might be combined effectively with mass spectrometry. Two primary problems were evident: first, the physical arrangement appeared to be incompatible with MS interfacing (because of the immersion of the ends of the capillary in buffer reservoirs that also provided the electrical contact), and second, the low flow rates, which effectively precluded use of a thermospray type interface. Beyond this, the small sample sizes (generally picomole and lower) used in CE demanded a detector enormously more sensitive than any that existed.

In 1984, Professor John Fenn of Yale University showed that ESI could be coupled with mass spectrometry to enable the analysis of small thermally labile and nonvolatile compounds [20]. In 1985, I initiated another line of research aimed at using ESI to combine capillary electrophoresis and mass spectrometry (CE-MS) and recruited postdoctoral associate Jose Olivares to work on this problem. Our first publication on CE-MS in 1987 demonstrated the achievement of subfemtomole ($<10^{-15}$ moles) detection for quaternary ammonium salts [21].

While not generally recognized, our initial work with ESI for CE-MS involved the use of a sheathless interface in which the ESI was generated using a metalized capillary terminus. The use of 100 μm i.d. capillaries provided liquid flow rates of $\sim$1 μL/min under the conditions used. Subsequent improvements included the development of the liquid sheath ESI interface in 1988 [22], which broadened the application of ESI by enabling a much wider range of solutions to be electrosprayed. This approach was later adopted in early work by others for the coupling of liquid chromatography with MS. While used less for LC-MS coupling today, the liquid sheath interface remains the most widely used CE-MS interface design. This is because it does not require special modification to the terminus of the CE capillary, because of its inherent mechanism for providing the CE electrical contact, because of the effectively zero dead volume coupling to ESI, and because of the flexibility provided by the ability to manipulate liquid composition (e.g., pH) just before ESI.

A key attraction of ESI-MS is its capability for analyzing extremely small volumes of liquids. Almost all the early ESI-MS work was done at liquid flow rates of 1–10 μL/min. Studies of electrosprays predating its use in mass spectrometry had well demonstrated that beyond a certain point the ESI current increased only slightly as the liquid flow rate was further increased. Thus, while the ESI process could be assisted to allow spraying at higher flow rates (e.g., ionspray) for interfacing of conventional LCs, from a mass spectrometric perspective all this accomplished was to consume more sample (i.e., lower sensitivity).

Our early CE-MS studies convinced me that sig-

nificant additional gains in sensitivity could be achieved using ESI, simply by delivering a solution more slowly to the electrospray emitter. Indeed, the small sample sizes and high resolution of CE separations, along with the extremely high efficiency of ESI at the low flow rates provided by CE, have generally provided the highest sensitivity measurements of biopolymers achieved using mass spectrometry, including the first attomole range measurements for proteins, in 1992 [23,24]. In 1993, we demonstrated the reduced elution speed concept for slowing CE-MS separations by controlling the CE electric field strength, thereby exploiting this understanding of the ESI phenomena [25]. We showed that the CE elution could be slowed by nearly an order of magnitude without loss of separation quality or significant reduction of MS signal intensity (i.e., with a significant gain for integrated peak intensity) and allowing more complex samples to be characterized. In 1993, we also exploited the same fundamental phenomena in a different way; postdoctoral associate Jon Wahl used metalized and contoured fused silica 10-μm i.d. capillaries for highly effective CE-MS [26,27]. Using ~10 nL/min flow rates, this work demonstrated that significantly improved sensitivity could be realized and actually constituted the first reported ESI at what is now considered nanospray flow rates.

4. Electrospray ionization–mass spectrometry and its biological applications

As a graduate student from 1971 to 1975 at the University of Utah, my thesis research under the direction of Jean Futrell involved fundamental research into ion–molecule reactions using ion cyclotron resonance (ICR) mass spectrometry, the predecessor technology of today's much more powerful FTICR technology. In 1988, I became involved in developing the research/technology basis for what would ultimately become the W.R. Wiley Environmental Molecular Sciences Laboratory (EMSL) at PNNL and, particularly, for a division including advanced mass spectrometric instrumentation. In moving away from studies of supercritical fluids, my research increasingly emphasized the development of ESI and new capabilities for characterizing biomolecules, and it was stimulated significantly by the 1988 report from Fenn and coworkers on the observation of extensive multiple charging for large polymers and proteins [28,29]. Thus, in early 1988 I began to expand my group's efforts in the area of biopolymer analysis by ESI-MS. One immediate result of this was the demonstration by postdoctoral associate Joe Loo of the now widely employed method of collisional activation in the ESI interface for providing considerable additional information in ESI mass spectra [30]. In a series of publications beginning in 1989, our research was the first to show that extensive structural or sequence-related information can be directly obtained from the tandem mass spectrometry of large multiply charged polypeptides and proteins [31–38], opening the way for the much more rapid and sensitive characterization of biopolymers. This early work also showed that extensive sequence could be readily obtained for modestly sized polypeptides [32]. It also demonstrated the use of protein dissociation in the interface with subsequent MS/MS of its fragments to obtain sequence information from intact proteins [35,37,38]. During this period, other studies examined Coulombic effects in the dissociation of the highly charged ions produced by ESI [39], developed the technique of thermally induced dissociation [40], first explored and provided insights into the role of charge state on the activation energy for gas-phase dissociation [41], and produced the first ion–ion studies of the reaction of multiply charged ions from ESI studied using a merging stream Y-tube reactor [42,43].

In recent years, the scientific community has increasingly recognized that ESI-MS can be used to study aspects of the higher-order structure of biopolymers in addition to their sequence and modifications. In 1988, our studies revealed that solvent composition affected the ESI mass spectra for proteins [44]. In 1990, we showed that the higher-order structure of proteins having intact disulfide bonds, and thus much different three-dimensional structural constraints (both in solution and in the gas phase), were reflected in their mass spectra; the more compact constrained structures had lower charge states (i.e., were shifted to

higher *m/z*) and narrower charge state distributions [45]. In 1991, Ganem and Henion showed that ESI allowed the study of intact noncovalent associations (there was actually evidence of this in earlier published work showing intact myoglobin; i.e., retaining the noncovalently associate heme moiety under some solution/ESI conditions) [46]. We subsequently showed that intact multimeric protein complexes, DNA duplexes, and other noncovalent complexes were also subject to study. Our initial studies included such tetrameric proteins as hemoglobin, avidin, streptavidin, the chaperone protein SecB, and concanavalin A. The spectra generally revealed only a few peaks corresponding to different charge states of the multimeric species and proved to be consistent with the known solution stoichiometry, e.g.; peaks indicative of trimeric or pentameric species were not observed. One of the strongest protein–ligand interactions ($K_D \sim 10^{-15}$ M) found in nature is that of the 64-kDa glycoprotein avidin (or streptavidin) with biotin, which has wide applicability in biochemistry and molecular biology. Avidin is composed of four identical subunits that associate into the active form. Each of the four subunits can accommodate one biotin molecule. My group performed the first studies of the tetrameric complexes of avidin and streptavidin with biotin, yielding results consistent with conventional understanding [47,48]. The streptavidin complex with iminobiotin (a biotin analog with $K_D \sim 10^{-7}$ M) was observed to be weaker than the biotin complex, again consistent with solution behavior [48].

The methods developed in this early work were refined and applied to a series of increasingly challenging biological applications, including the first studies of DNA- and RNA-protein complexes. It is now well established, for example, that ESI-produced ions of double-stranded (ds) DNA of $\gtrsim$10 base pair (bp) size can be stable in the gas phase if not excessively activated; stability increases with size, and larger duplex ds DNA ions will undergo dissociation of covalent bonds (e.g., base loss) in preference to dissociation of the two strands. Our work was the first to demonstrate that the noncovalent complexes formed between various drug molecules (both intercalators and minor groove binders) were also subject to study [49]. In the mid-1990s, postdoctoral associate Xueheng Cheng used ESI-MS to examine the binding of the Gene V protein (a dimeric protein) with a variety of small single-stranded (ss) oligonucleotides, providing new insights into its stoichiometry [50]. These studies showed that carefully conducted ESI-MS results reflected solution interactions and indicated that the observed complexes in the gas phase did not arise from nonspecific electrostatic interactions. We also demonstrated that collisional dissociation methods could be used to break up the complex and then retain the Gene V product in an FTICR ion trap, after which collisional dissociation of the Gene V protein ions (derived from the complex) could then be studied to provide structural information on the protein. Another early example involved complexes of the eukaryotic transcription factor, PU.1, with ds DNA, which showed that the selective formation of such complexes could be studied and that the results correctly reflected independent solution studies [51].

Although initially quite controversial, it is now broadly accepted that ESI has significant utility for the study of noncovalent complexes, and applications are increasingly common. Initially it was thought that because protein ions were generated by ESI with substantial net charge, their more compact structure in solution would be stretched out in the absence of solvent because of Coulombic repulsion between charge sites. It gradually became evident that essentially the inverse is closer to the truth; electrosprayed ions are more highly charged because they are often denatured and in a form readily extended before ESI (or sometimes when greater heating is applied during ESI [52]). As noted above, our work first showed that lower charge state ions are produced from more compact structures [53,54]. Although more closely spaced charge sites give rise to greater long-range repulsive contributions, the energy differences occurring over distances typical of chemical bonds, and available to drive dissociation, are quite small. Thus, although the activation barrier for breaking a weak bond may be lowered because of the increased Coulombic contributions, noncovalent associations based on multiple weak electrostatic, hydrogen bonding and

dispersive interactions can persist for indefinite periods in the absence of excessive thermal activation to the molecular ion. A corollary of this is that once the gas phase structure is disrupted and charge sites are moved apart, it is implausible that molecules will refold in the gas phase (unless charging is reduced), and even then it is extremely unlikely that they would refold to the same native structure as in solution. As charging increases for a given molecular structure, the effective proton affinities of charge sites for positively charged proteins (typically the basic amino acid residues), for example, decrease to a level at which proton transfer to volatilized solvent in the gas phase can occur [55]. Thus, this simple model implied that the charge state distribution would depend on both the detailed three-dimensional molecular structure and the location and nature of charge sites for the multiply charged macromolecular ions [53].

Early in our work with ESI [53,56], I hypothesized that the mechanism of ion formation summarized in Fig. 1 would be consistent with key experimental observations. In this mechanism, which has its roots in Dole's initial ESI studies [1,2], asymmetric droplet fission occurs as droplets shrink to the point where surface tension is inadequate to maintain droplet stability, generating offspring droplets that have significantly smaller diameters. The asymmetric fission process facilitates the rapid formation of smaller charged droplets that are close to their respective maximum charge, and only a small amount of additional evaporation can cause a second asymmetric fission event. The larger progeny droplets can undergo additional asymmetric fission steps as solvent evaporation again drives them to instability, but the process will be faster and require less energy input for the smaller droplets. Through one or more such steps it is feasible to isolate single molecules given the concentrations relevant to most electrospray applications. For example, a 10^{-5} M solution initially contains $\sim$3 analyte molecules per 100-nm droplet; a 10^{-6} M solution contains on average only 0.3 analyte molecules (before any solvent loss by evaporation). Asymmetric fission steps might yield droplet sizes approaching the dimensions of larger proteins ($\sim$10 nm), and further evaporation then produces the molecular ions ultimately detected. The key experimental variable in this view of ESI is the extent and the rate for which the charged droplets are heated and smaller charged species are activated by collisions. For example, sufficient energy may be imparted to just barely cause desolvation, or at the other extreme, induce extensive dissociation in the gas phase. The desolvation of electrosprayed ions is typically completed in about a millisecond, and the resulting mass spectra are also strongly influenced by the extent of collisional activation and dissociation in the ESI-MS interface.

The observation of intact noncovalent complexes by ESI-MS thus conventionally requires a balancing act between providing sufficient heating/activation for ion desolvation and keeping interface conditions mild enough to preserve the complex (Fig. 1). Insufficient desolvation will cause a loss of sensitivity. However, increasing ESI source heating or collisional activation to increase sensitivity can induce dissociation of complexes and provide a distorted view of solution associations. Increasing the biopolymer concentration is generally not a satisfactory solution; it will consume more sample and typically increase nonspecific aggregation. Trapping methods (as in FTICR) allow the use of gentler electrospray source conditions than generally applied, as the ion population initially trapped can be (on average) significantly solvated. As we have pointed out [57], the ion population can then be desolvated on a much longer time scale than possible with conventional instrumentation ($>$100 msec for ion traps vs. $<$100 μsec for conventional instrumentation). This approach allows ions to be vacuum dried in the gentlest fashion feasible, minimizes the tuning of source and interface parameters, and provides higher sensitivity. As illustrated simplistically in Fig. 2A, the need for greater activation for desolvation of some ions results in excessive activation and dissociation of others. Slower desolvation (Fig. 2B) allows a lower energy input and avoids overshoot dissociation. Qualitatively, this relationship is rationalized by the dissociation rate versus 1/(temperature) behavior being much steeper (i.e., caused by a larger energy requirement). Thus, desolvation dominates at low temperatures and long times, while

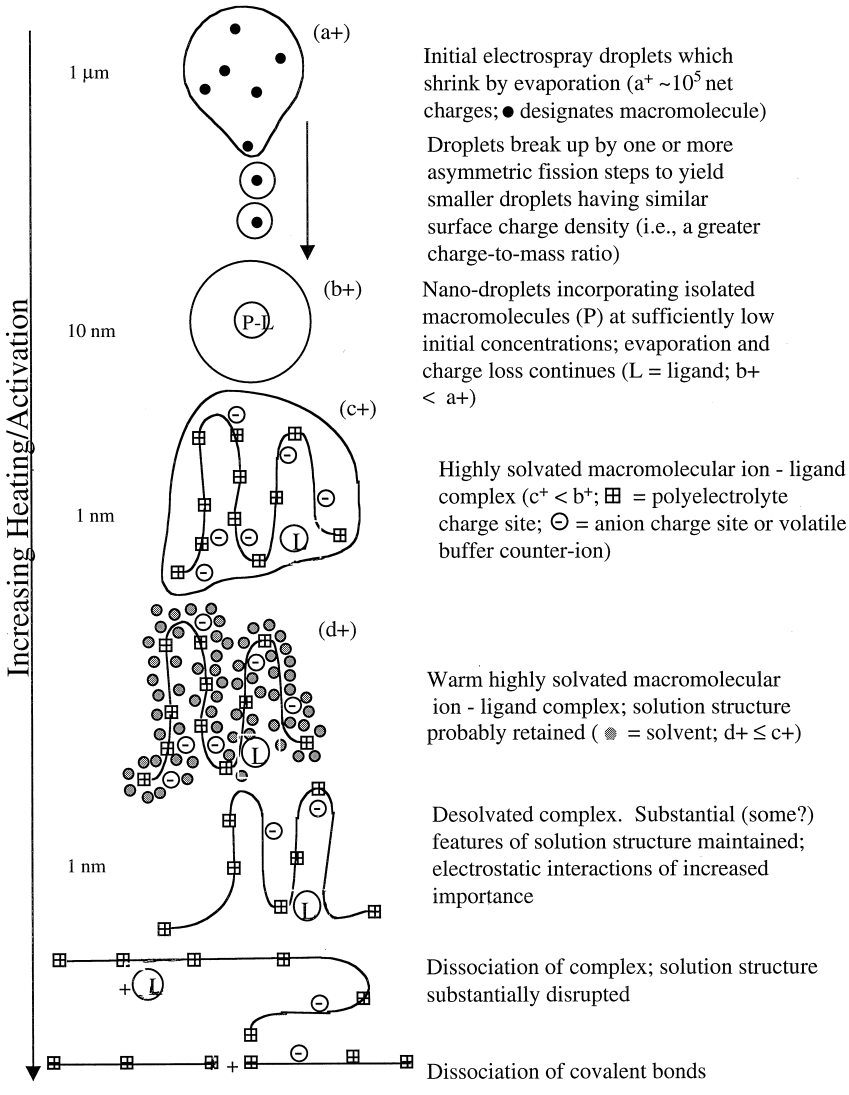

Fig. 1. Early model for formation of macromolecular ions during electrospray ionization based on asymmetric droplet fission and the resulting formation of small nanometer-diameter droplets [52,53,56]. With increasing activation, residual solvent and charge are shed. The loss of higher-order structure and noncovalent associations (e.g., ligand, L) is primarily driven by the extent of activation and repulsive Coulombic forces (after removal of the solvent).

dissociation becomes dominant at the higher temperatures needed to drive rapid desolvation.

5. ESI-FTICR mass spectrometry

In the last decade, the combination of ESI with FTICR has greatly advanced the capabilities of mass spectrometry for biological applications [58]. The multiple charging phenomena of ESI and FTICR's better performance at lower *m/z* values makes ESI-FTICR superior to MALDI-FTICR for the study of larger biopolymers. Ions are radially confined in the FTICR trap during measurements by an applied magnetic field and axially confined by a trapping potential

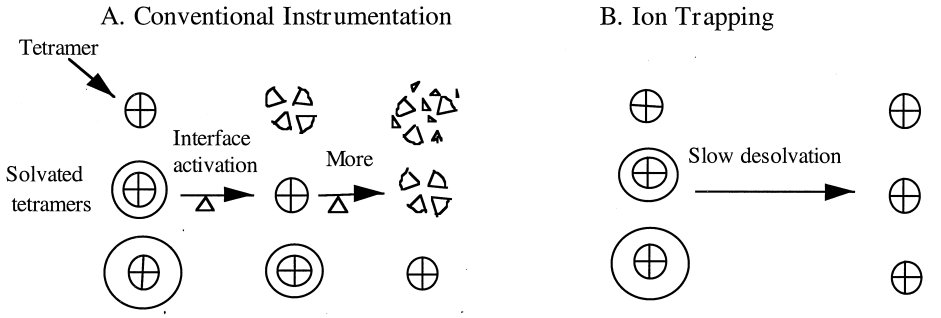

Fig. 2. Comparison of the different results obtained for desolvation because of differences in the characteristic time scales and necessary heating with conventional and ion trapping mass spectrometers (e.g., FTICR). A tetrameric complex is considered to have an initially broad variation in the extent of solvation. A slower desolvation avoids excessive activation in the ESI interface but results in greatly decreased sensitivity if incomplete at the time of analysis.

applied orthogonal to the magnetic field. The frequency of the cyclotron gyration is inversely proportional to the mass-to-charge ratio (*m/z*) and directly proportional to the strength of the applied magnetic field [59]. The orbiting ion clouds induce charges often referred to as image currents on two or more detection electrodes. When Fourier transformed, these oscillating signals provide an extremely precise measurement of the effective cyclotron frequencies (and thus the *m/z*). The resulting spectra can simultaneously yield high resolution, sensitivity, and mass measurement accuracy.

While many FTICR instruments were already in operation in 1990, most were dedicated to fundamental studies of either ion–molecule reactions or ICR-related phenomena. At that point, FTICR was widely recognized as providing measurements having resolutions and accuracies much greater than any other form of mass spectrometry, but poorly understood performance issues had made it almost completely ineffective for biological applications, and particularly for larger molecules. The reality was that there were a few demonstrations of potential utility, but essentially no real biological application. The first combination of ESI with FTICR by Hunt and McLafferty and their respective coworkers [60], like most initial demonstrations, provided no convincing evidence that either high resolution or high sensitivity could be achieved for larger biopolymers.

In my early planning for the mass spectrometry facility of the Environmental Molecular Sciences Laboratory at PNNL in 1988, I relied heavily on the unrealized promise of ESI-FTICR mass spectrometry for biological research. In 1990, my group and I organized a workshop on the future of high-performance FTICR, attended by many of the leading researchers in this area, and began planning our lab's seven-tesla FTICR instrumentation, including several new features such as beam shutters and in-magnet cryo pumping. In 1991, McLafferty and coworkers first demonstrated that high resolution could be achieved for small proteins by ESI-FTICR [61], the major breakthrough being the use of long ion cooling times (many minutes in their early work) to presumably allow a sufficient period for the evolution of a more compact ion cloud before excitation of the trapped ions to a large cyclotron radius for detection.

Beginning in mid-1992, three postdoctoral associates in my group (Brian Winger, Steven Hofstadler, and James Bruce) quickly put our FTICR instrumentation into operation and produced results demonstrating the high resolution and sensitivity achievable for the study of large biopolymers using ESI-FTICR mass spectrometry [62]. From that point I focused my group's efforts in directions aimed at both advancing the technology (along with what I hoped to be supportive fundamental studies) and moving it toward its biological application.

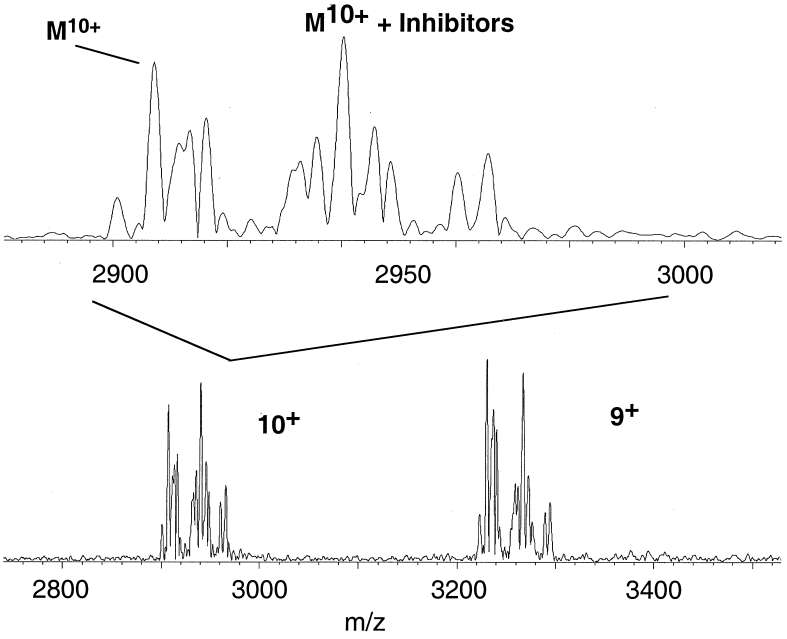

Fig. 3. Positive-ion ESI-FTICR spectrum obtained for a set of 17 inhibitors (0.05 μM each) with BCAII (1.0 μM in 10 mM ammonium acetate at pH 7). The bottom shows the narrow charge state distribution observed, typical of cases where noncovalent associations and a (presumably) compact three-dimensional structure is maintained using gentle ESI conditions. The top spectrum shows detail of the 10+ charge state region and shows a series of peaks corresponding to the enzyme-inhibitor complexes and the remaining BCAII [57].

6. ESI-FTICR mass spectrometry and its application to noncovalent complexes

In 1994, I initiated research efforts by my group aimed both at more fundamental studies of noncovalent interactions of electrosprayed complexes and at their application to protein complexes with ligands from complex combinatorial libraries. This work benefited greatly from a collaboration with Professor George Whitesides of Harvard University, which provided a model system consisting of the 29-kDa protein carbonic anhydrase (CA) and its complexes with various substituted benzenesulfonamide inhibitors [63–66]. CA is a roughly spherical Zn(II) metalloenzyme having a conical binding pocket, which catalyzes the hydration of CO_2 to bicarbonate and is an attractive model system because of the stability of CA and its well-characterized structure and ligand complexes. A large body of data correlating the structures of sulfonamide ligands with their binding constants to CA provided a basis for inferences regarding the protein's structure and its ligand interactions in the gas phase.

Under conventional conditions used for ESI (where the native structure of the protein is rapidly denatured), mass spectra for a mixture of bovine CAII (BCAII) and a set of inhibitors in acidic solution show a range of charge states extending to $>20+$ and loss of the Zn, even under very gentle ESI interface conditions. The higher charge states observed indicate a disrupted solution structure (attributed to acid catalyzed denaturation) and is consistent with the loss of Zn. A dramatic difference is evident for a similar mixture of BCAII and inhibitors in 10 mM ammonium acetate at pH 7 (Fig. 3). Two lower-charge-state species (10+ and 9+) dominated the spectrum, indicating that the structure was much more compact. A series of peaks is observed corresponding to complexes of BCAII with the various inhibitors, and Zn is fully retained in these gas phase complexes.

It was not clear from these results whether the inhibitor remained in the binding pocket after transfer

into the gas phase or whether the binding pocket structure was maintained (i.e., although the association of BCAII and inhibitor is maintained, the structure might be altered greatly). We then conducted experiments that compared the relative stabilities of BCAII complexes having both one and two identical inhibitors, and showed that the first inhibitor binds more strongly in the gas phase, consistent with its retention in the binding pocket. If structural features are lost, no special gas phase stability would necessarily be expected for the first inhibitor compared to the second. When two inhibitors are associated with the enzyme, at least one of them must be nonspecifically bound outside the binding pocket. A difference in the complex stability in the gas phase might be expected if a binding pocket were substantially preserved in the gas phase and the inhibitor specifically bound in solution remained associated; it is reasonable to expect that an inhibitor in the enzyme-binding pocket will be more strongly bound in the gas phase than an identical inhibitor randomly associated on the protein surface. The relative gas phase stabilities of the desolvated bovine carbonic anhydrase II $(BCAII)^{10+}$ complexes with both one inhibitor and two inhibitor molecules (formed using large excesses of inhibitor) were then investigated in the FTICR ion trap using sustained off-resonance irradiation (SORI) for selective collisional heating. By fixing the frequency difference between the SORI and the complexes' cyclotron frequencies, the gas pressure, and the length of the SORI event, the amplitude of the SORI irradiation was used to vary the extent of collisional activation. Thus, the dissociation efficiency as a function of the SORI peak-to-peak amplitude (Vpp) provided a measure of the gas phase stability of the complex. These experiments clearly showed that $(BCAII + 2Inh)^{10+}$ complexes were less stable in the gas phase, with loss of the second inhibitor being significantly more facile, and suggested that the enzyme-inhibitor complexes in the gas phase retained significant features from solution and that the dehydrated binding pocket likely retained the specifically bound inhibitor from solution into the gas phase. Studies also showed that stability was greatly reduced in the gas phase for inhibitor complexes with BCAII-Zn (i.e., the apo-protein) at pH 7, where the structure is known to be similar to the intact protein.

Results from other experiments also established that the para-NO_2-benzenesulfonamide-human CAII (HCAII) complex is more stable than its ortho-substituted counterpart in the gas phase [57]. The bulky NO_2 group in the ortho position has a sterically hindered interaction of the inhibitor with the binding pocket and displays reduced binding affinity in solution. In these experiments, the 10+ charge state ions of the protein-inhibitor complexes were selectively accumulated in the FTICR and then subjected to SORI collisional activation. The HCAII complexes with the para-NO_2 inhibitor showed significantly greater gas phase stability (Fig. 4). As it was known that the solvent does not directly contribute to the steric effect, a similar difference in complex stability in the gas phase would also be expected for these two isomeric inhibitors if the active site structures of the noncovalent complexes were preserved. However, randomly associated inhibitors (i.e., where either the binding pocket no longer exists or the inhibitors do not remain in the pocket after transfer to the gas phase) would not be expected to result in any difference in the gas phase stability. These results also indicated that distinctive features of the binding pocket were retained in the gas phase. These studies showed that carbonic anhydrase can maintain a relatively compact structure in the gas phase and that the binding pocket retains features that result in selective inhibitor interaction in the gas phase.

We then asked whether measurements of relative stabilities in the gas phase will generally mirror those in solution. To probe this issue, we conducted studies of BCAII-inhibitor complexes and showed dramatic differences in the relative gas phase stabilities compared with either solution binding constants or solution dissociation rates (off-rates) for an array of complexes [65]; no direct correlation with the complex stabilities in solution was observed (Fig. 5a). However, a plot of the calculated polar surface area for the inhibitors (the portion of the total molecular surface that is charged, has a large dipole, or contributes to hydrogen bonding) versus SORI amplitude revealed good (but different) correlations for sets of

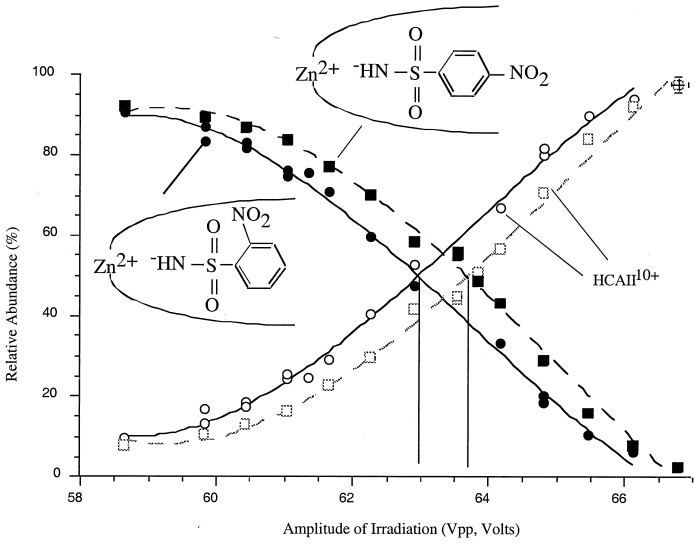

Fig. 4. A plot of the normalized abundances of the para- and ortho-NO_2-benzenesulfonamide-HCAII complex ions (filled symbols) and the dissociated HCAII product ions (open symbols) versus the extent of SORI activation. The amplitudes of irradiation at the crossing point of the two curves provide measures of the relative gas phase stabilities of the complexes in the gas phase. These results indicate that a similar steric effect exists between the protein and the ligand in the gas phase and in solution, and suggest that features of the binding pocket are retained after desolvation in the gas phase.

inhibitors both with and without aromatic amino acid residues (Fig. 5b). The stabilities of the complexes were also found to increase monotonically with the number of inhibitor amino acid residues (or the polar surface area) [65]. This suggested that the major attractive forces for noncovalent protein-ligand binding in the gas phase are caused by interactions between the polar surfaces through electrostatic, dipole–dipole, or hydrogen bonding interactions, and that the inhibitor tail has collapsed to the protein surface in the gas phase and contributes significantly to complex stability. For inhibitors having the same polar surface area, an aromatic amino acid side chain results in a stronger binding interaction with the protein in the gas phase than does an aliphatic side chain. Thus, whereas the off-rates of BCAII-inhibitor complexes in solution are mainly affected by hydrophobic interactions between the inhibitor and the enzyme, their corresponding gas phase stabilities were primarily determined by polar interactions that reflect the extent of contact area between complex constituents.

7. Bioaffinity characterization mass spectrometry

The selection of pharmacologically active molecules from large combinatorial mixtures or libraries of compounds has become a major tool in drug discovery. The ability to screen very large libraries to identify candidates for more exhaustive study allows a broad range of molecular diversity to be explored. For example, in the case of tripeptides, a library containing all variations of three amino acid residues will consist of 8000 peptides (20 possible for each of three residues, i.e., 20^3). The number of possible hexapeptides is 6.4×10^7 (i.e., 20^6). In the combinatorial approach, many compounds are synthesized (often deliberately producing a complex mixture) and then subsequently examined for their affinity to the targeted biopolymer (or other property), often by partitioning subsets of the library to facilitate the identification of the most active components.

In the mid-1990s we developed and demonstrated a new approach for the screening of combinatorial libraries based on the capabilities of ESI-MS and the

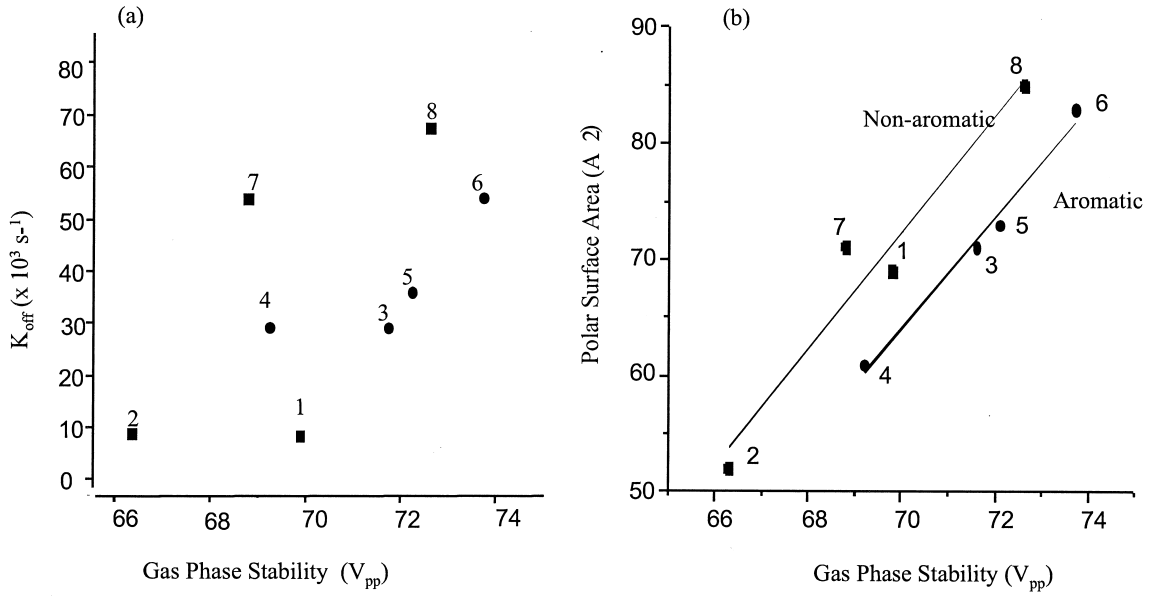

Fig. 5. (a) Plot of liquid phase dissociation constant versus gas phase stability for a set of para-substituted benzenesulfonamide inhibitor complexes with BCAII-measured SORI-induced dissociation, showing stability increases with the size of the inhibitor tail [57,65]. (b) Plot of the polar surface areas of the same inhibitors versus gas-phase complex stability. The inhibitors having aromatic amino acid residues (circles) show stronger binding with BCAII than inhibitors with aliphatic side chains in the gas phase having the same polar surface area.

use of multistage FTICR ion manipulation methods [63,67]. The bioaffinity characterization mass spectrometry approach eliminated the need for distinct separation and/or purification steps involving the original ligand mixture and the problems associated with linking the affinity ligand to a surface, as the complex can be formed in free solution at physiologically relevant pH values. The potential advantages of this approach originate from the ability to first form and study the complex in solution (in which it may have only a very low concentration), to separate it in the gas phase from all other ions (including other complexes), and finally to provide structural information on the complex and/or the affinity selected biomolecule. Thus, the separation/affinity selection step and analysis step are combined in one experiment. A representation of one implementation of our approach is given in Fig. 6. The complexes to be investigated are electrosprayed directly from a solution containing both the affinity target and the ligand library and, using appropriate ESI source conditions, transferred to the gas phase and accumulated in the FTICR ion trap. Typically, the complexes of interest are first identified in the mass spectrum and then isolated in a separate step using selected-ion accumulation (SIA); subsequently, we showed the selective accumulation of even very low concentration species [68], which fills the trap to useful capacity of the population of selected ions. The SIA step is followed by SORI dissociation of all the complexes and the retention of the corresponding higher binding affinity ligand species in the trap at lower *m/z*, after which a high-resolution mass spectrum can be obtained. In the same experiment, the inhibitors can be further dissociated to obtain structural information and for definitive identification [63,64].

The study of noncovalent complexes often requires that the concentration of the affinity target should be limited to avoid nonspecific aggregation during ESI, so the complexes of interest may be present only at very low concentrations in solution. Very large libraries would further increase the need for sensitivity and benefit significantly from the FTICR capability we developed for selectively accumulating trace level

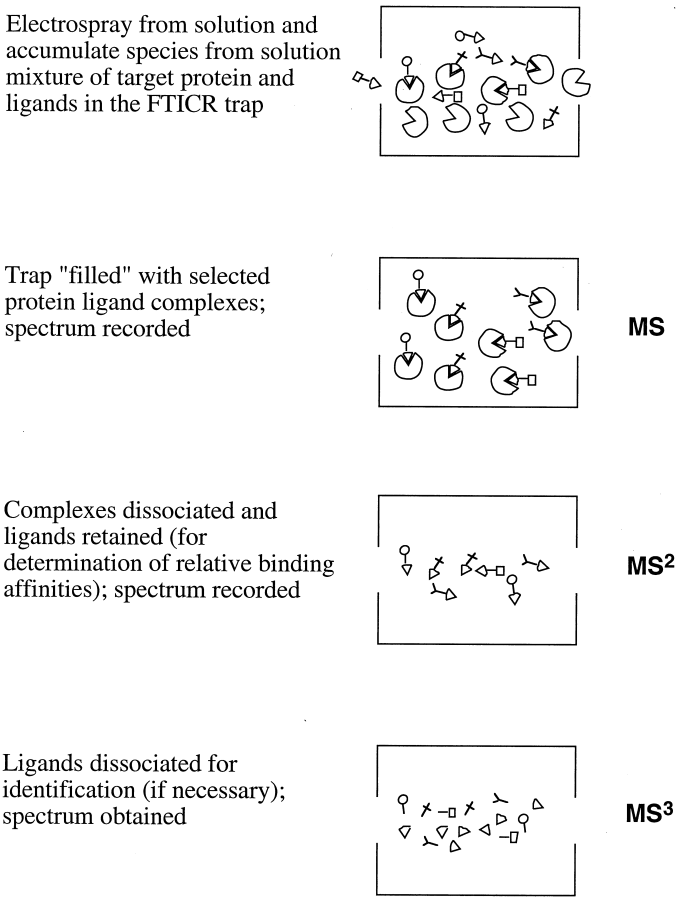

Fig. 6. A conceptual representation of the bioaffinity characterization ESI-FTICR technique for the study of mixtures of complexes. In this approach, the complexes are selectively accumulated in the trap and then dissociated with retention of the ligands that would then display relative abundances representative of their binding with the targeted biopolymer. Because the ligand ions are retained in the ion trap after a spectrum is obtained, subsequent stages of MS can be used for structural characterization of the ligands, using dissociation methods.

constituents. Importantly, the wide range of FTICR capabilities for high resolution, accurate mass measurement, and multistage MS analysis are fully available to be applied to species recovered from the dissociation of noncovalent complexes.

We initially studied two inhibitor mixtures of seven and 18 components, that had well characterized BCA-inhibitor dissociation constants spanning three orders of magnitude (10^{-6} to 10^{-9} M) [63]. One group of inhibitors consisted of compounds having dipeptide substituents with one amino acid invariant and the other position incorporated amino acids chosen to be representative of diverse size, shape, hydrophobicity, and acid/base properties. The relative abundance ratios of the various complexes observed by FTICR for two series of inhibitors were consistent with their relative affinities towards BCAII in solution. In this work, we also demonstrated the use of multistage MS methods for dissociation of the inhibitors to assist their identification (although high-resolution analysis was adequate for small libraries). To demonstrate the extension to larger combinatorial libraries, two mixtures of para-substituted benzene sulfonamide inhibitors were synthesized using solid-phase chemistries in which all combinations of 17 amino acid residues (AA_1 and AA_2) were incorpo-

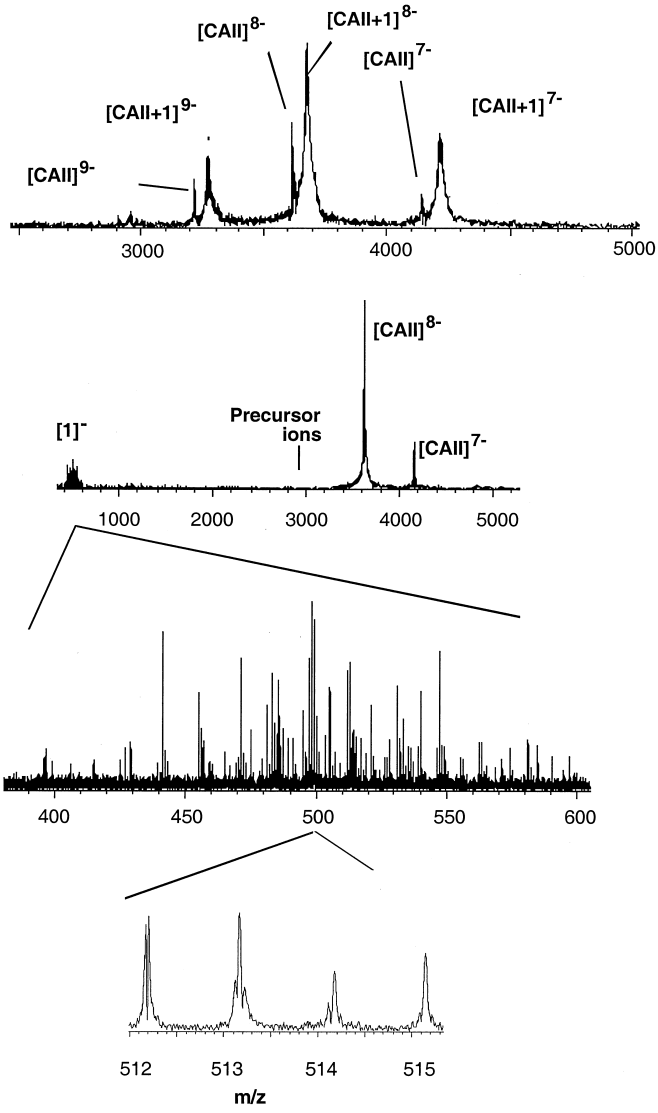

Fig. 7. (Top) FTICR mass spectrum showing the intact complexes from ESI of a mixture of the 289-component L-library (0.5 μM each) and CAII (2.5 μM) in 10 mM NH_4OAc (pH 7.0) [64]. (Bottom) Views for the mass spectrum obtained by SORI dissociation of the isolated complex ions of $[CAII+\mathbf{1}]^{9-}$. The dissociation conditions were set such that the complex ions were completely dissociated. The bottom two show expanded views of the singly charged inhibitors, $[\mathbf{1}]^{-1}$ originating from dissociation of the complexes.

rated into two positions ($17 \times 17 = 289$ components) [64]. Collision-induced dissociation of the selectively accumulated nine-charge state complexes primarily produced the eight-charge state of the protein and singly charged negative ions for nearly all the inhibitors ($[\mathbf{1}]^{1-}$; Fig. 7). The resulting mass spectra allowed the dissociated ligands to be identified based on their molecular weights; their ion intensities provide a measure of the relative binding affinities in one experiment. Fig. 8 gives a representation of the relative abundances for the dissociated ligands in terms of the composition of the amino acid residues for the two libraries. Seven individual inhibitors were then synthesized from the two libraries to determine

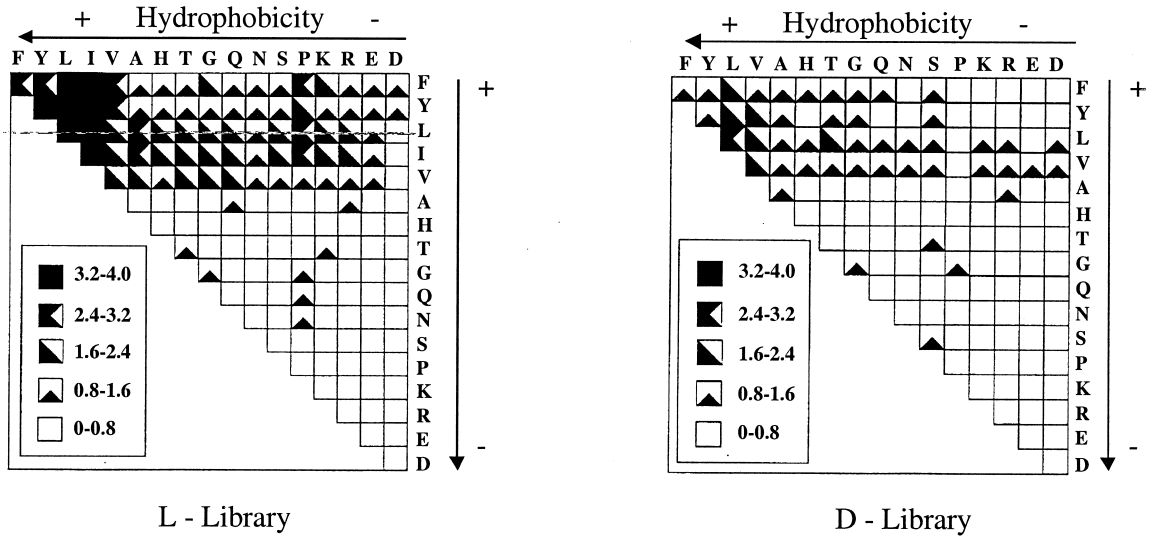

Fig. 8. The dependence of relative ion intensities on the composition of two amino acid residues in the library [57,64] used for the spectrum in Fig. 7. The amino acids (AA_1 and AA_2) for both L- and D-libraries are arranged in such a way that their hydrophobicity decreases from left to right and from top to bottom. The relative ion intensity for each ligand was obtained by comparing the ion intensity of this ligand with that of the Gly-Gly compound (present in both libraries).

their binding affinities to BCAII in solution using a fluorescence binding assay [64]. The binding constants of the seven inhibitors in solution correlated well with the relative intensities of the ions dissociated from the BCAII-inhibitor complexes (Fig. 9). The K_b for the tightest binding inhibitor identified by ESI-MS ($AA_1 = AA_2$ L-Leu; see Fig. 8) was 1.4×10^8 M^{-1}. The K_b of the compound bound most weakly ($AA_1 = AA_2$ = Gly) was 4.9×10^6 M^{-1}. The data also indicated that the addition of hydrophobic groups at the para position of benzenesulfonamide resulted in increased binding constants. Interestingly,

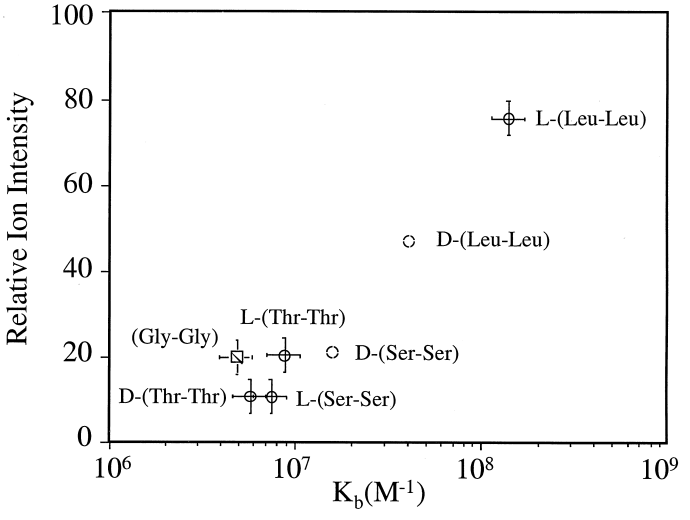

Fig. 9. Correlation of relative ion intensities versus K_b in solution for seven peptide inhibitors from L- and D-library [64].

the chirality of the amino acids also appeared to influence the binding affinities of the tripeptide inhibitors: the side chains of L-amino acids interact more effectively with the active site of CAII than did D-amino acids.

These studies demonstrated new approaches that exploited a feature of FTICR that has, in actuality, rarely been utilized: the ability to reuse ion populations. The attractions of this approach include speed and the immediate characterization of the most relevant components from complex and sometimes poorly characterized mixtures. Less obvious advantages include the potential for establishing differences in binding properties for biopolymer variants and modified biopolymers; in such cases, the biopolymers could also be subjected to characterization by an additional stage of mass analysis after dissociation of the complexes.

8. Increasing the sensitivity of ESI–mass spectrometry

A long-term aim of our research has been to increase the sensitivity of ESI–mass spectrometry for use in biological research [34,69]. The ability to address biological problems is determined in many cases by the sensitivity that can be achieved; gains in sensitivity enable fundamentally different types of measurements to be made (e.g., at the single-cell level) or studies that would be otherwise impractical (e.g., samples from small tumors, microdissection, or where samples from excessive numbers of animals would otherwise have to be pooled).

Most of our early studies, while in the context of the ESI interface for CE-MS, were generally applicable to all ESI-MS studies. In 1993, postdoctoral associate Dave Gale first applied the use of very low flow rates for general direct infusion, using small-diameter contoured (etched) capillaries and flow rates of as small as 50 nL/min to reduce sample consumption and enhance sensitivity [70]. As for our CE-MS work with very low flow rates [26,27], sensitivity and other aspects of performance were significantly enhanced.

Increasing the sensitivity achieved in conjunction with other features of the overall analysis has also been of interest. In many cases where MS^n studies can potentially provide useful structural information, the species of interest are of too low relative abundance or are trapped along with many other species. As molecular weight increases, these situations present significant experimental challenges, particularly when sample is limited, as is generally the case in biological research. In 1993, we developed FTICR methods for selectively accumulating specific species in the ion trap using quadrupole excitation (QE) methods [68]. This development exploited a unique in magnet cryopanel feature of our instrumentation that allowed us to obtain high transient pressures for effective quadrupole axialization in the seven-tesla magnetic field. (Quadrupole cooling had been ineffective in other similar magnetic field instrumentation because of lack of this capability.) The selective ion accumulation (SIA) methodology greatly improved the analysis sensitivity by allowing the FTICR cell to be filled with specific species from complex mixtures. SIA can enhance sensitivity for trace species because longer ion accumulation periods are possible without the normal limitations imposed by the expansion of magnetron radii from the presence of the buffer gas added to facilitate trapping. The greatly extended dynamic range is also important. Elimination of any high-abundance species allows more ions in the desired *m/z* range to be trapped and detected. In collaboration with Andy Ewing (Penn. State University) in 1995, we showed that the combination of capillary electrophoresis with ESI-FTICR allowed high-resolution mass spectra to be obtained for hemoglobin from a single red blood cell [71,72]. We demonstrated broadband QE methods for improved trapping of ion populations and SORI methods for collisional dissociation of ions in the FTICR ion trap, and showed that partial amino acid sequence information could be obtained with online CE-FTICR [73]. Later developments have since greatly improved on the sensitivity of this approach.

Key to many applications is the achievable dynamic range of measurements, and this becomes particularly important when sample size is limited and

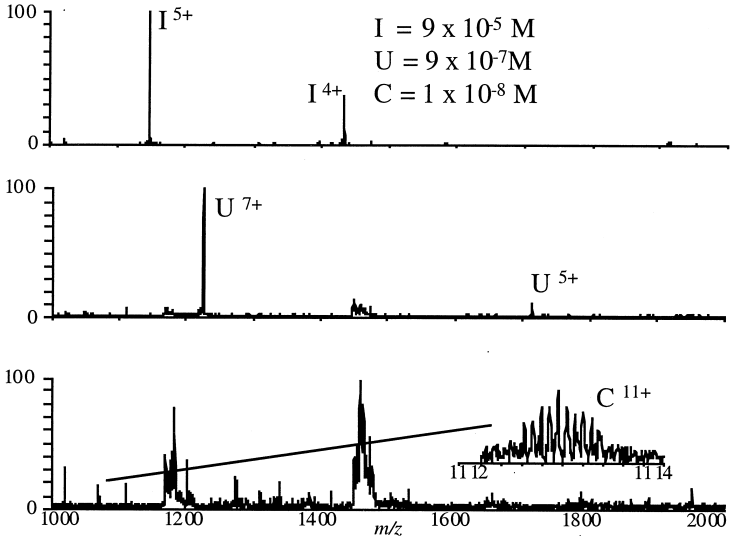

Fig. 10. Demonstration of the use of colored-noise waveforms to obtain a large effective increase in dynamic range and sensitivity with FTICR. The solution contained three proteins, insulin (I), ubiquitin (U), and cytochrome *c* (C) having relative concentrations differing by nearly four orders of magnitude. (A) shows that the conventional spectrum was dominated by the major insulin species (I). The information in this spectrum was then used to rapidly and automatically generate a colored-noise waveform designed to remove the major species and produce the middle spectrum (B), where the intermediate concentration species (U) now dominates. The same approach was then automatically followed by using the new peaks to create additional notches in the colored-noise waveform to remove the additional major species in the middle spectrum and generating the bottom spectrum (C). The bottom spectrum shows many trace level impurities as well as peaks caused by cytochrome *c* (C^{11+}, clearly identifiable because of its isotopic envelope structure).

sensitivity is also a concern. For this reason, in 1996 we extended and generalized the SIA approach by using rapidly synthesized colored noise excitation waveforms to enable automated dynamic range expansion [74]. The FTICR trap has a useful charge capacity that is three to four orders of magnitude larger than that used in quadrupole ion traps (depending on trap size, magnetic field strength, and other operational details), and that provides the basis for an inherently greater dynamic range within any given spectrum. Our earlier studies with SIA had shown that lower concentration species could be effectively accumulated in the FTICR trapped ion cell, while more abundant species (not subjected to QE) were continually removed by expansion of the magnetron orbit radius. The result is that more space is made in the trap for the selected lower-abundance species accumulated over greatly extended periods. Our approach used colored (or filtered) noise waveforms to apply QE to any desired parts of the mass spectrum and performed this in an essentially real-time and data-dependent mode of operation [74]. This approach was demonstrated by the example shown in Fig. 10 for ESI-FTICR of a mixture of insulin, ubiquitin, and cytochrome *c* at various concentrations ranging from $\sim 10^{-4}$ to 10^{-8} M. Under normal in trap ion accumulation conditions, the dynamic range obtained was ~100, and the lower abundance species were not evident in the mass spectrum (Fig. 10, top). The application of selected-ion accumulation using a colored noise waveform synthesized from the initial spectrum (i.e., where notches were applied at the cyclotron frequencies of the major detected species in the initial spectrum) gave a greater than two orders of magnitude gain in dynamic range and allowed the detection of cytochrome *c* present at 10^{-8} M concentration in the same solution. The speed of this approach makes it viable for use with on-line separations of proteins or peptides and should be of particular utility for the proteomic strategies discussed later. Indeed, we are presently implementing a variation upon this approach in which the data dependent

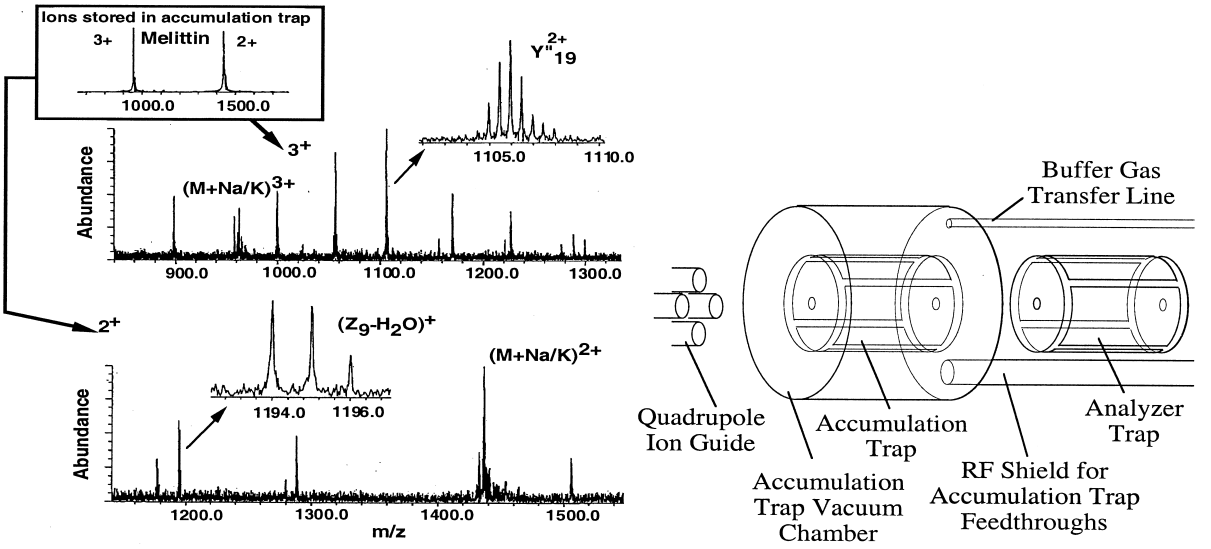

Fig. 11. (Right) Schematic representation of a dual cell arrangement originally developed for our seven tesla instrumentation [75]. The source cell is enclosed to produce higher-buffer gas pressures without sacrificing vacuum quality in other parts of the system. Shielding of cell wires reduced cross talk, allowing the simultaneous and independent operation of the two cells. (Left) Demonstration of the selective and sequential transfer of two different charge states of melittin parent ions (inset shows initial spectrum) between the two traps. This exploited the selective axialization of the desired species while the remaining species have their magnetron motion excited to larger radii and cannot escape through the lower trapping barrier that applies for ions that are axialized. After transfer, each charge state was individually selected for dissociation, producing the spectra shown.

dynamic range enhancement is conducted external to the FTICR cell, in an RF-only quadrupole.

An important aspect of optimizing sensitivity by MS involves increasing the information that can be obtained from a given sample and ultimately wasting as few ions as possible. The nondestructive nature of FTICR detection offers the potential for remeasurements of the same ion population and their use in tandem or multistage (i.e., MS^n) dissociation experiments from the introduction of a single group of ions that can be dissociated and remeasured multiple times, providing extensive structural information with minimal sample consumption. While the basis for such measurements was evident, a series of conceptual and practical issues had, in reality, long prevented such applications. As a step toward circumventing these issues, we developed and demonstrated a new dual trap arrangement shown in Fig. 11 (right) [75]. One of our goals with the dual cell was to provide improved isolation of the two cells and provide greater flexibility, more effective multistage studies, and greater overall sensitivity. High-efficiency ion accumulation was accomplished in the first higher-pressure cell contained in an enclosed region having its own gas inlet. Fig. 11 (left) shows a demonstration of sequential ion transfer between the cells in combination with the CAD of selected ions. First, different charge states of melittin were trapped in the high-pressure cell and their mass spectrum was obtained. Using single-frequency QE, ions of different charge states were sequentially transferred into the analyzer cell, where they were subjected to SORI CAD. Thus, the entire procedure with CAD of different species was performed with the same ion population initially trapped in the higher pressure cell. These studies showed that with proper QE control of the magnetron expansion of ions in the higher-pressure cell and the careful choice of trapping potentials during the transfer process, the ions not transferred from the higher-pressure cell could be reaxialized and subsequently transferred to the analyzer cell. The advantage of this approach is the large gain in effective duty cycle and

achievable sensitivity when there are multiple species of interest initially present.

The overall effectiveness in utilizing ions produced by the ESI source is the key to optimizing sensitivity. In most single trap FTICR instruments, nearly all of the sample consumed is actually wasted, as ions are accumulated in the trap for only a small fraction of the experimental sequence (i.e., the ionization duty cycle is typically $<10\%$). External accumulation of ions alone does not address this issue unless ion selection during or before accumulation is also accomplished. Thus, while external accumulation will certainly increase the duty cycle, it does so generally only because either the transfer to the region or the accumulation itself is inefficient; if this were not the case the ion trap would generally be almost immediately overwhelmed by the number of ions that can be produced by ESI. Present efforts in our laboratory are now refining external ion selection methods that provide a somewhat analogous capability to that demonstrated for dynamic range expansion (see Fig. 10) [74] but for which the overall duty cycle can approach 100% [76].

Further increases in the sensitivity of FTICR mass spectrometers depend significantly on increasing the overall ion transmission from solution to the FTICR cell, as well as the efficiency of trapping the ions in the cell. While ESI at atmospheric pressure can be very efficient for dilute samples delivered to the electrospray emitter at low flow rates [69], the reduction or elimination of losses during ion transport from the atmospheric pressure region of the ESI source to the second vacuum stage at a pressure of few hundred microtorr, has been inefficient.

After considering the issues involved for quite some time, I developed the concept for the electrodynamic ion funnel [77,78]. Our earlier work had shown that in a well-designed ESI source and interface the largest ion losses occur in the higher-pressure regions. Previous work had shown that RF ring electrode devices can be used to effectively confine ions at lower pressures and in the presence of collisions [79,80] and that RF quadrupole, hexapole, or octapole devices (i.e., multipole devices in general) can be used at pressures in the range of 1–100 mtorr for ion focusing because of collisional damping [81]. The advantages of these devices are substantial because both spatial focusing (which gives a higher ion current) and energy focusing (which gives a narrower, more effectively trapped kinetic energy distribution) are obtained [82]. However, such devices do not work well at higher pressures and they do not have the ability to refocus ions dispersed over large volumes (i.e., the acceptance area of multipole devices is closely related to the emmitance area of the device). The ion funnel uses RF ring electrodes of progressively smaller diameter and with the opposing RF phases applied to adjacent elements. The result of this arrangement is to confine ions within the funnel. A separated DC field is applied to drive ions down the throat of the funnel. The ion funnel concepts were initially implemented by postdoctoral associates Keqi Tang and Scott Shaffer and have been extensively evaluated and used in a number of our mass spectrometers [77,78,82–85]. Our initial evaluation of performance with a 3.5-tesla ESI-FTICR mass spectrometer demonstrated the anticipated high-sensitivity performance for proteins and peptides [76] and enabled detection limits in low zeptomole range. Further improvement in the instrument sensitivity will require increasing the probability of ion trapping—a probability of trapping close to unity—and should enable detection limits to be extended to the subzeptomole range. If achieved, such a level of sensitivity would enable, for example, the characterization of many proteins from a single cell.

9. ESI-FTICR studies of individual ions

The ultimate in instrumental sensitivity will be achieved when single molecules can be selected from complex mixtures and effectively analyzed. An early goal of my interest in FTICR was to develop methods that would enable measurements of single ions. Although physicists had previously achieved the detection of small ions in Penning traps [86], this had been done with narrow-band (e.g., SQUID) detectors that, in effect, required that the mass detected be known in advance.

My interest in single-ion detection by FTICR began shortly after learning the extent of multiple charging that could be realized for the ESI of large molecules. From the viewpoint of sensitivity, it appeared that FTICR detection methodology would be more than adequate for detection of individual large ions. At that time, the Human Genome Project was just beginning, and I conceived of an approach for DNA sequencing based on the detection of large individual ions [87]. The concept failed, not because of an inability to measure the masses of individual ions but, rather, because of the failure to establish a gas-phase process that would sequentially cleave nucleotide bases from a DNA strand.

Our initial publication showed that we could not only detect single large ions [88] but that the signal arising from single (i.e., individual) ions was detectable for ions having as few as 30 charges using our seven-tesla instrument and a conventional FTICR cell [89]. For larger molecules that carried more charges, detection of a single ion trapped in the FTICR cell became progressively easier, and we have reported the detection of ds DNA ions of $>$100 MDa size with $\sim$30,000 negative charges [90]. These large ions were shown also to produce cyclotron signals extending for much longer times after dipolar excitation than for smaller ions. This observation was because of the large mass of these ions and the fact that dephasing or loss of coherence of an ion packet, a major factor in the conventional damping of the transient signal after dipolar excitation, obviously cannot occur for a single ion that is stable.

Our initial publication also showed that we could determine the mass of individual ions on the basis of directly observed step-wise charge-state changes [88]. Because mass spectrometers measure the *m/z* of ions and ESI often produces a series of peaks corresponding to different charge states for which neither *m* nor *z* is initially known, obtaining an MW measurement requires that the charge of the detected ions be determined in some manner. In conventional ESI-MS measurements this is accomplished by deconvolution of the spectrum, exploiting the distribution of charge states observed for a given species. FTICR typically makes use of the 1-Da isotopic spacing. However, neither of these approaches was directly applicable for the case of single ions, as a simple measurement yields a single *m/z* value for which neither *m* nor *z* is known (actually, we showed that the approximate charge could be determined from the signal amplitude, but this approach is unlikely to yield measurements with better than 5%–10% uncertainty [91]). We conceived and demonstrated a method in which charge state changes are induced by ion–molecule reactions during the detection event; the *m/z* of multiple charge states of the same ion are measured allowing unambiguous charge state determination. This time resolved ion correlation (TRIC) method, applied using software developed by Gordon Anderson and Jim Bruce, allowed for accurate charge state determination [88,89]. As shown in Fig. 12 for a plasmid DNA sample, simultaneous and independent measurements of several ions (depending on the charge capacity of the cell) are feasible with this approach (0.1% accuracy was demonstrated in measurements for 2 MDa plasmid DNA [92]). The potential advantages of this approach include unmatched sensitivity and the capability for making accurate mass measurements of molecules that are much larger than feasible with any other approach yet known.

Our initial studies provided mass measurements based on a limited number of individual ions. As such, the molecular weights obtained were quite good for the set of ions studied, but could vary substantially from the sample average molecular weight because of inherent polydispersity (because of isotopic content, adduction, etc.) of high molecular weight compounds and statistical limitations caused by the number of ions studied.

A fascinating aspect of individual ion measurements using FTICR is the potential for characterization of much larger molecules than possible with conventional measurements involving large ensembles of ions. However, FTICR measurements of large populations of ions become increasingly problematic as MW increases, as large numbers of trapped charges cause greater Coulomb-induced frequency shifts and ion coalescence phenomena [93]. We showed that TRIC measurements could be effectively multiplexed, as a number of ions could be measured in each

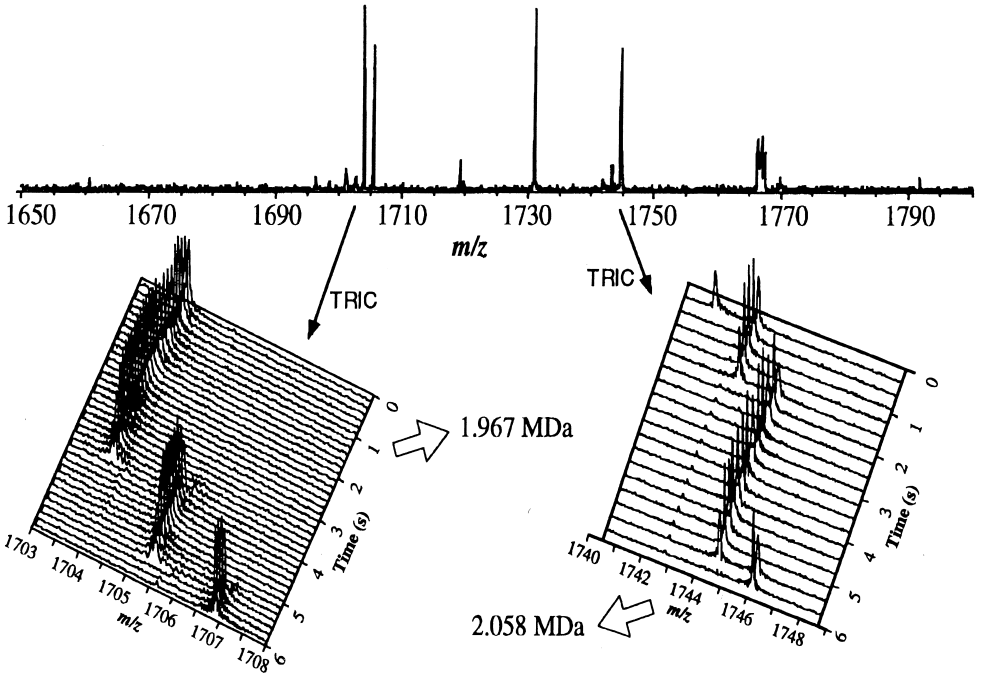

Fig. 12. Time-resolved ion correlation plots for two of a group of ~2-MDa plasmid DNA ions from a single FTICR measurement. Note that several peaks appear in the spectrum, corresponding to the different charge states of each ion that were present during the observation time. The different charge states resulted from ion–molecule reactions of the individual ion.

spectrum and several sets of spectra could be acquired to produce a compilation (i.e., histogram) of all resultant masses from the TRIC analyses. Limitations of this approach, however, were the relatively long observation time required to detect charge state shifts caused by ion–molecule reaction, the limited number of ions in each spectrum (so that the reactant and product pairs from charge state changes could be reliably identified), and the resulting significant data processing and computational requirements.

We subsequently developed an alternative approach involving the summation of sparsely to moderately populated individual ion spectra [94]. This approach allowed us to define charge states or to provide sufficient resolution to define isotopic (1 Da) spacing so as to enable the use of conventional methods of mass determination [94]. A primary advantage here was that the large number of charges that are required to define the charge state distribution, and that may under ordinary circumstances preclude FTICR measurements or lead to degraded performance because of Coulombic interactions, are spread out temporally over many independent acquisition steps (i.e., spectra). A second advantage is that this approach provides much higher resolution and mass measurement precision, as space charge related effects are minimized and time-domain signal duration is much longer (providing the basis for more precise *m/z* measurements). A third advantage is that ion cloud stability limitations (that our work had shown to significantly define the upper MW for which isotopic resolution is achievable [93]) are circumvented, opening the door to more routine measurements of much larger biopolymers. The initial demonstration of this approach for the molecular weight measurement of bovine serum albumin (BSA; ~66 kDa) is shown in Figure 13 [94]. The combination of longer time domain signals and improved quality trapping fields should allow much greater precision and accuracy in mass measurements of very large ions. This approach not only circumvents the limitations on high-resolution measurements for large molecules but also pro-

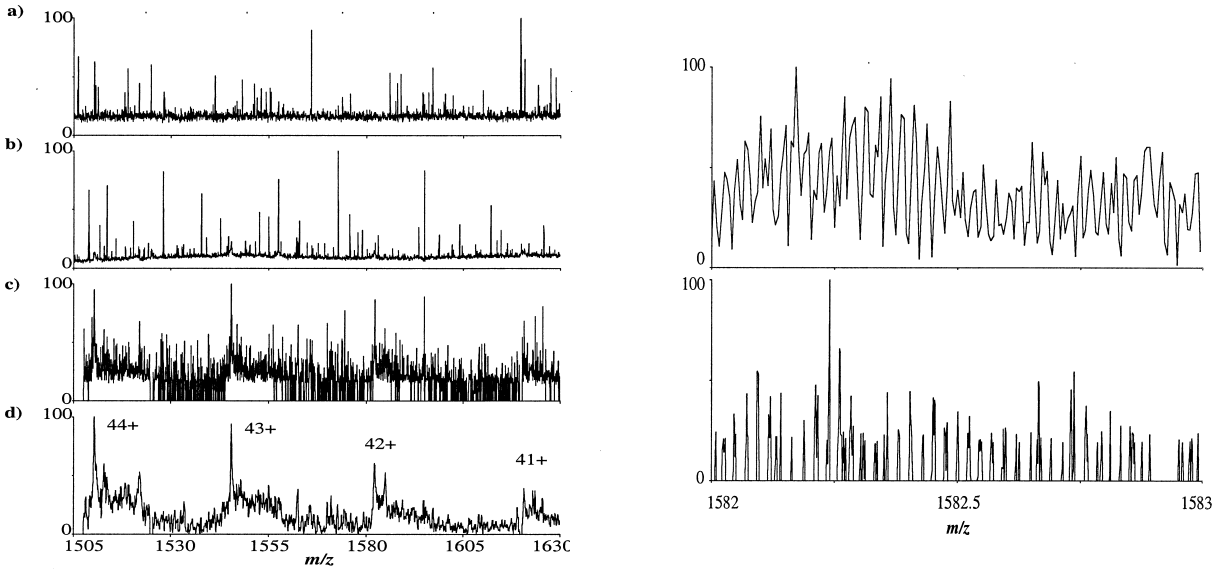

Fig. 13. (Left) Demonstration of the individual ion spectral summation approach for BSA. (a) shows a typical BSA individual ion spectrum [94]. The trapped ion population is insufficient to define the charge state or isotopic distributions but results in high-performance detection because of low space-charge conditions. (b) Direct summation of 100 spectra showing reduced S/N (produced by conventional summation without prior noise reduction because of the large number of mass-to-charge ratios possible). (c) Summation of 100 peak picked and processed individual ion spectra, demonstrating the retention of resolution for the temporally distributed large population of ions. (d) Result of box car integration over the resultant summed spectrum in (c), more clearly defining charge state peaks. (Right) Comparison of the isotopically resolved peaks resulting from the conventional SIA of a single-charge state of albumin ions (top) and summation of 100 individual ion spectra (bottom), as on the left. Isotopic peaks from the individual ion summation approach appear narrower and at slightly higher frequency, presumably because of the reduced ion population (and space-charge effects) present during detection of the individual ion spectra. These results show that better isotopic resolution and higher-precision mass measurements can be obtainable for higher–molecular weight species using this approach.

vides a foundation for achieving the ultimate in sensitivity if, for example, using methods such as those described earlier, analyte ions could be stored and delivered at desired rates to the FTICR trap.

10. ESI-FTICR mass spectrometry and proteomics

I believe some of the most important contributions of mass spectrometry over the next decade will be made in the area of proteomics. In one sense, proteomics is not a new field, and it is sometimes argued that almost any protein characterization can be classified as proteomics. Because the use of MS for the identification of spots excised from two-dimensional (2D) gels has been refined over the last decade and is now broadly applied, it can also be argued that MS's role in proteomics is not new.

The proteome may be defined as the entire complement of proteins that can be expressed by a particular cell, organism, or tissue; proteomics can be similarly defined as the study of the protein complement expressed at a given time or under a specific set of environmental conditions. One of the key differences in the role of mass spectrometry in proteomics is a matter of scale; the interest is often in many proteins and obtaining as complete a representation of the proteome as practical. Thus, to some, proteomics is simply a matter of higher throughput for protein characterization.

From another perspective, however, I believe the field of proteomics represents a real paradigm shift related to how we think about and use the information. Ignited by advances during the Human Genome Program, biological research is moving rapidly into a postgenomic era involving the explicit study of complex biological systems. In this new paradigm, cellular pro-

cesses are increasingly studied globally and new understandings are gained regarding their systems-level behavior and the emergent properties arising from their complex nature. A major aim is to understand the cellular function of proteins as well as their more conventionally delineated molecular function. This view of cellular systems can provide a basis for understanding the robustness of cellular systems, the nature of multigenic diseases and individual variability, and so forth. In addition, we gain understanding of the molecular nodes, typically proteins, that can be targeted for drug development, gene therapy, genetic manipulations, etc. The improved understanding of cellular systems will, for example, open a wide range of opportunities that can never be addressed by the present single-gene approaches that dominate conventional (reductionist) biological research. The higher-level cellular systems view provides a basis for understanding the robustness of cells, the nature of multigenic diseases, epigenic diseases, and individual variability and a starting point for the development of predictive modeling capabilities for the effects caused by perturbations such as low-level exposures.

While the availability of complete genome sequences opens the door to important biological advances, much of the real understanding of cellular systems and the roles of its constituents will necessarily be based on proteomics. The capability to precisely measure changes in the expression (and modifications) of many proteins simultaneously enables identifying/inferring and, ultimately, understanding the function of the proteins participating in the multiple pathways and will provide insights into how cellular networks are linked and how new otherwise unpredictable system properties arise. When combined with the capability to model and simulate cellular systems and, ultimately, develop a predictive capability, proteomics will provide the basis for attacking some of the most important problems in biology.

Until now, however, available analytical techniques have fallen far short of the capabilities needed to provide such high-throughput measurements of an organism's proteome. While 2D-PAGE separations have demonstrated the ability to provide a detailed view of thousands of proteins expressed by an organism or cell in a single 2D display [95], it remains a relatively slow, labor-intensive, and cumbersome technology. The sensitivity of this method is limited by the amount of a protein needed for visualization (typically on the order of ~10 femtomoles) [96] as well as the quantity of protein that may be loaded onto the gel. Indeed, there is the growing recognition that global proteome displays based on 2D gels are largely constrained to the more stable and abundant (e.g., housekeeping) proteins and that important classes involved in signal transduction and regulation of expression, for example, and other lower-abundance (and generally less stable) proteins go largely undetected. 2D-PAGE is also biased against membrane proteins, highly acidic or basic proteins, and very large or small proteins.

High sensitivity and dynamic range is of crucial practical importance to the study of cellular pathways and networks because many important protein classes, such as those involved in signal transduction, will be present only at low concentrations. Not only is high sensitivity crucial for the detection of important regulatory proteins, it also improves the quality of abundance measurements for low-level proteins. Increased sensitivity allows proteomic measurements to be extended to smaller cell populations or tissue samples (e.g., obtained from microdissection). High dynamic range is also important so that low abundance proteins can be detected in a mixture also containing higher-abundance proteins.

The approaches to proteomics being developed and applied in my group can be divided between those that are based on intact protein analysis and those that utilize enzymatic processing.

11. Proteomics based on capillary isoelectric focusing with ESI-FTICR for intact proteins

The mass spectrometry community's approach to protein characterization (and proteomics) has been dominated by methods involving the proteolytic processing of proteins to produce more analytically tractable polypeptides (generally of <3-kDa size). Our laboratory is strongly interested in the development and application of a complementary capability

for intact proteins. The driving force for this is the complementary nature of the information generated; for example, only by measuring the intact species can one be truly confident as to the origin of a polypeptide fragment. Information on the coordination of multiple modification sites is largely unobtainable unless the different species are first separated (and the separated species identified). A further pragmatic reason for this approach is that substantially less sample processing is required, and obtainable detection limits should be better than those obtained using proteolytic processing, thus, for example, providing the basis for more effective measurements at the single-cell level.

Capillary isoelectric focusing (CIEF) employs a polyampholyte mixture in free solution when an electric field is applied bridging high-pH and low-pH solutions, sets up a pH gradient in the capillary through which proteins migrate until they have zero net charge and thus focus at their respective isoelectric points (pI). CIEF can handle extremely small sample sizes and offers enhanced speed and resolution, ease of automation, and the high sensitivity arising from the natural concentration effect (a factor of ~100-fold) associated with focusing. After focusing, protein bands are mobilized to the detector by one of several methods, such as chemical mobilization (anodic or cathodic) or hydraulic mobilization (i.e., pressure-, vacuum-, or gravity-driven flow). We have shown separations of analytes having pI differences as small as 0.004 pH units [97]. The CIEF-MS combination is thus roughly analogous to a 2D-PAGE separation, in the sense that it provides information on pI and molecular mass, but with added potential advantages that include greater speed and sensitivity, ease of automation, much more accurate mass measurements, and the potential for use of advanced multidimensional MS methods. The amount of total protein loaded in the CIEF capillary in our initial work was typically 100–300 ng, whereas 2D-PAGE can commonly consume 100 μg to >10 mg of total protein. For faint spots, protein extracts from as many as 20 2D-PAGE separations are often pooled for analysis by conventional MS methods.

We initially demonstrated the effectiveness of CIEF for obtaining high-resolution separations of complex protein mixtures of cell lysates extracted from *Escherichia coli* [98,99]. *E. coli* was harvested, lysed, and dual microdialyzed, and soluble proteins were prepared for CIEF-FTICR measurements (by adding Pharmalyte 3-10 carrier ampholytes to a concentration of 0.5%) and injected to fill the capillary for CIEF. A wealth of mass measurement data can be obtained from the numerous mass spectral peaks attributed to proteins (Fig. 14). Automated data analysis software was developed that allowed visualization of the results in the form of familiar 2D-PAGE displays, as illustrated in Fig. 15. The 2D displays are produced by plotting molecular mass versus scan number, which is correlated to protein pI. The spectral intensity in this display is derived from ion abundances and is communicated through the use of variable spot sizes. The 2D display shown in Fig. 15 contained ~900 "spots," that is, unique putative protein masses from *E. coli*. It should be noted that the strategies for identification of intact proteins have limitations. Our experience has shown that only a modest fraction of intact proteins can be initially identified based solely on mass measurements. While MS/MS of smaller proteins is useful for identification, it is generally not well suited for the identification of large proteins, and this is particularly problematic when sample size is very limited or with on-line separations. We have more recently shown that specific amino acid isotopic labeling using auxotrophic organisms can significantly aid protein identification [100].

12. New approaches to proteomics based upon global proteolysis

While proteome studies analyzing intact proteins by CIEF-FTICR provide a high density of data, they also provide an incomplete view of the proteome, as insoluble and very high MW proteins remain problematic. In addition, protein identification can be difficult. The alternative, and more classical, approach is to identify proteins by measurements of the peptides obtained from an enzymatic (e.g., tryptic) digest of the proteins. While the masses of several peptides

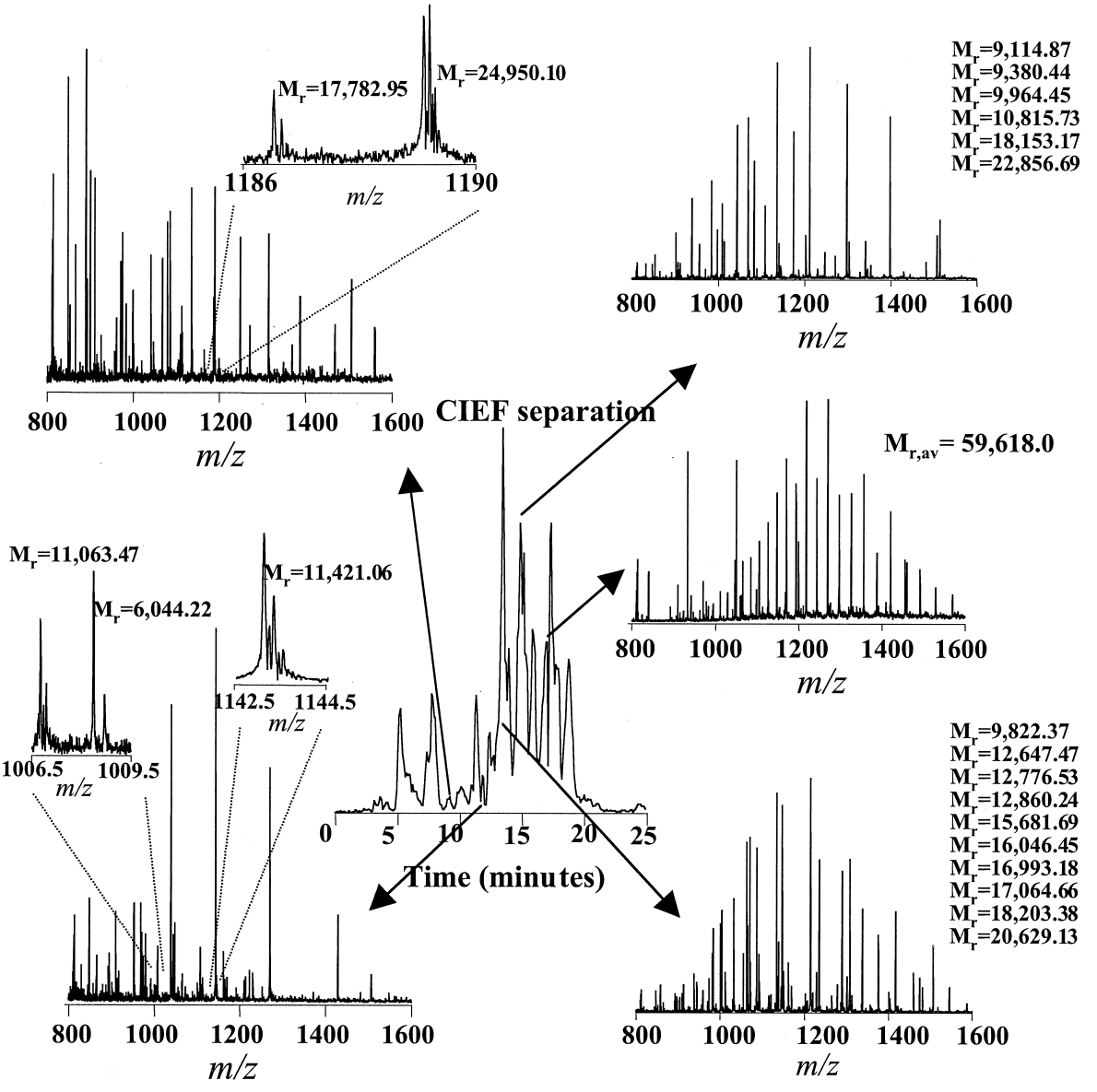

Fig. 14. Demonstration of the high density of information obtained in a single CIEF-FTICR analysis of *E. coli* lysate. The reconstructed electropherogram (center) shows several resolved peaks; however, analysis of each scan reveals that several protein masses can be present within a single scan. This analysis provided ~900 putative protein masses from a single 30-min experiment.

are generally required, MS/MS data for only a single peptide is often all that is required for protein identification.

This fact led to a concept that opens the door to high-throughput approaches for broad proteomic measurements: If the molecular mass of a single peptide can be measured with high sufficient mass measurement accuracy (MMA), such that its mass is unique among all of the possible peptides predicted from a genome, it can then be used as an accurate mass tag (AMT), or an effective biomarker for unambiguous protein identification. The use of such AMTs allows

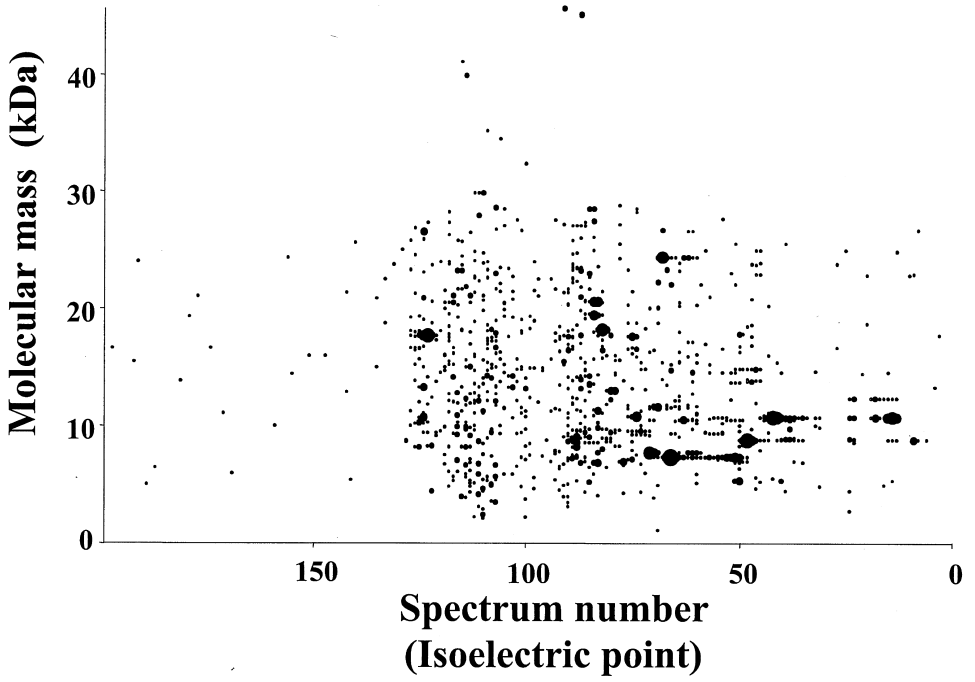

Fig. 15. Virtual 2D-gel display of the CIEF-FTICR analysis of lysates from *E. coli* grown in normal medium.

the proteolytic fragments generated from an entire proteome or complex protein mixture to be analyzed and identified with much greater speed.

In this approach, we utilize global digestion methods for the proteome mixture and analyze the data in the context of calculated digest products from of all of the proteins predicted from the subject genome (or other appropriate databases). Conventional MS/MS approaches can then be used to fill in ambiguous mass measurements and for the initial validation of AMTs. Initial validation of the protein identification enables subsequent proteome measurements to be based almost exclusively on the use of AMTs, and thus substantially increases throughput. This approach obviates the routine need for additional identification efforts (e.g., using MS/MS) and is important for subsequent quantitative studies where samples from many perturbations, sampling times, and so on, are of interest to understand the cellular function of proteins and the cellular system.

It is possible to predict, with some qualifications, the set of proteins possibly expressed by a given organism by translating all hypothetical open reading frames (ORFs) predicted from genomic sequence information. The set of potential protein molecular masses can then be calculated based on the predicted amino acid sequences. We have analyzed all of the predicted tryptic polypeptides for a number of sequenced organisms using the predicted ORFs. Fig. 16 shows the percentage of unique tryptic peptide masses predicted for yeast and *Caenorhabditis elegans* at varying levels of MMA. For predicted peptides from yeast and *C. elegans* having masses of ~2000 Da, ~65% and 60%, respectively, have unique masses at a MMA of 0.1 ppm. A MMA of 1 ppm only decreases the percentage of peptides having unique masses from 65% to 50% in the case of yeast and from 60% to 40% for *C. elegans*. At MMA levels of 10 ppm, however, <10% and 5% of the predicted peptides for yeast and *C. elegans*, respectively, have unique masses; at 100 ppm, almost none. The advantages of higher MMA are even greater as the peptide mass increases. Calculations show that for both yeast and *C. elegans* >80% of the predicted peptides heavier than 2500 D

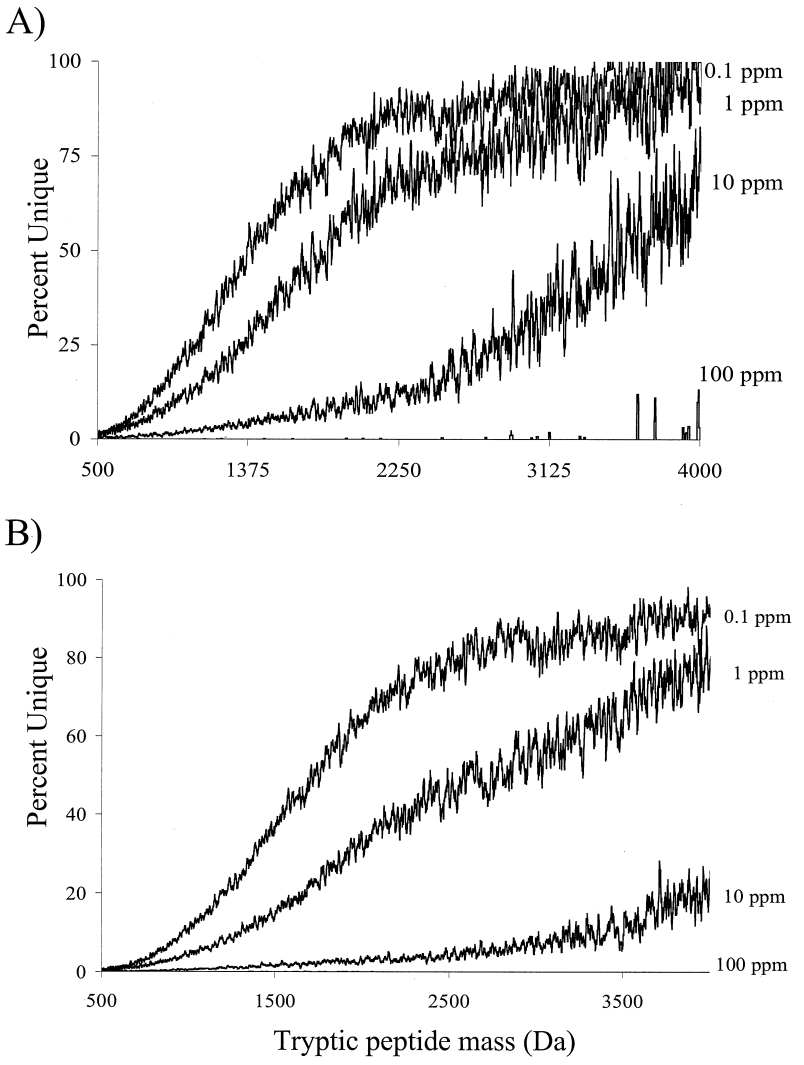

Fig. 16. Percentage unique tryptic fragments (potential accurate mass tags) as a function of tryptic fragment mass at four different levels of mass measurement accuracy for the predicted proteins of (A) yeast and (B) *C. elegans*.

have unique masses in the proteome at sub–parts per million MMA.

As already discussed in the context of intact proteins, approaches that provide additional information on amino acid composition can greatly aid protein identification. For this purpose, Aebersold and coworkers have developed isotope coded affinity tags to reduce the complexity of the proteome analysis [101]. The approach involves both affinity selection of cysteine-containing polypeptides (Cys-polypeptides) by modifying the proteins with a Cys-specific reagent containing a biotin group and affinity isolating the modified peptides using immobilized avidin. Not only does this method substantially reduce the complexity of the proteome mixture but it also significantly reduces the MMA required for the generation of AMTs.

Obtaining the greatest possible numbers of AMTs will not only require high MMA (as discussed above) but will require the ability to detect as many peptides as possible within the context of a relatively rapid overall separation strategy. As shown in Fig. 17, >22,000 peptides (from among >35,000 peaks) were

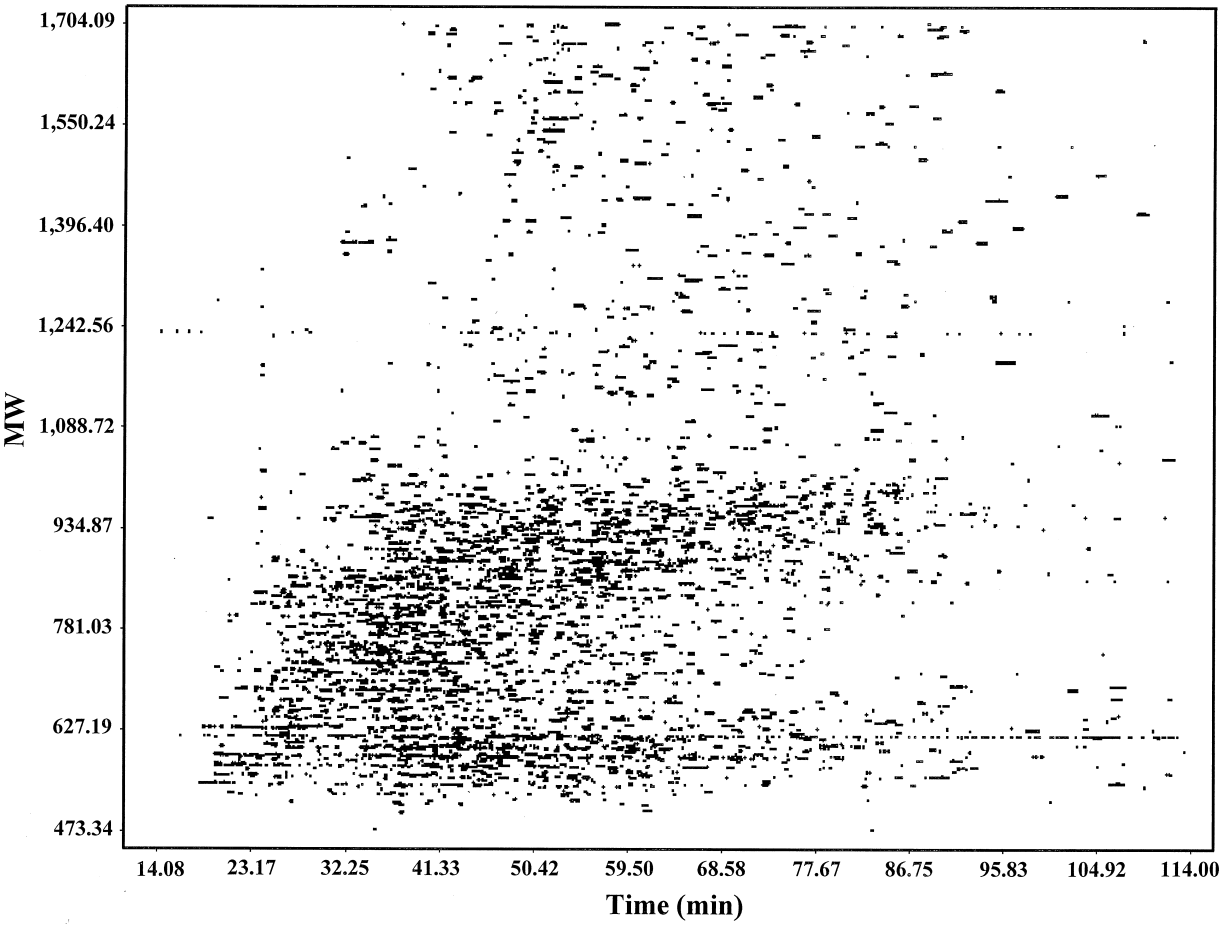

Fig. 17. A two-dimensional capillary LC-FTICR display from a tryptic digest of a soluble protein lysate from *Deinococcus radiodurans* [102]. More than 22,000 putative polypeptides were observed during this single analysis. The implementation of automated dynamic range expansion methods is expected to significantly extend proteome coverage.

initially observed in a single capillary LC-FTICR analysis of polypeptides isolated from a tryptic digestion of a soluble protein lysate from the highly radiation resistant microorganism *Deinococcus radiodurans* [102]. An ideal tryptic digest would produce about 60,000 polypeptides in the 500 to 5000 Da range for the set of proteins predicted from the DNA sequence, from which more than 99% of proteins could be identified. A single scan from one LC-FTICR experiment is shown in Fig. 18. While several high-intensity peaks are observed in the full spectrum, the advantage of FTICR for peptide detection is clearly demonstrated when regions that appear to contain little information are expanded (Fig. 18B). Further expansion of this region reveals even more lower-abundant peptide signals (Fig. 18C). The ability to detect low-abundance peptides not only allows the broadest possible proteome coverage but, more important, allows important regulatory proteins that may be present in low copy number to be studied.

More than 1000 peptide species can sometimes be detected in a single spectrum, such as shown in Fig. 18, and $>10^4$ should be measurable in a single spectrum if peptides were similar in abundance (a number that can likely be increased further by the use of the dynamic range expansion methods described earlier). As the peak capacity of the LC separation is ~500, the theoretical

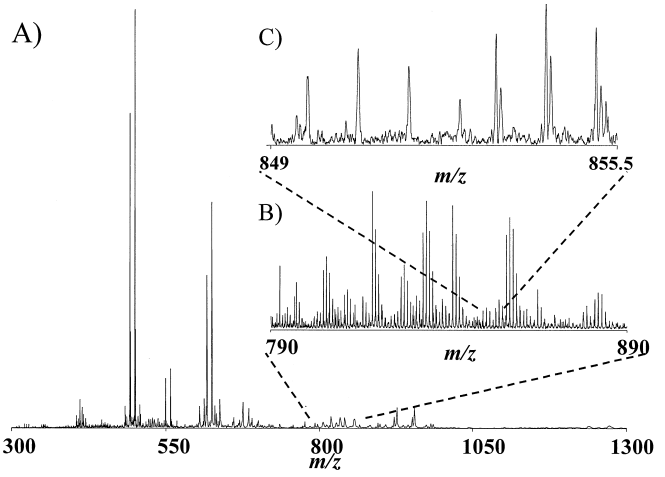

Fig. 18. Single mass spectrum from a capillary LC-FTICR analysis of a whole-proteome tryptic digest [102]. While several highly abundant peptides were observed within the entire *m/z* range, expansion of various regions shows the presence of lower-abundant species detectable because of the high sensitivity, resolution, and dynamic range achievable using FTICR. More than 2800 different polypeptides can be counted in this single spectrum.

peak capacity of the combined LC-FTICR approach is at least 10^6.

In our approach, high-throughput proteomic studies will broadly involve two phases, one in which the proteins are identified, and a second phase in which many measurements are made of proteome-wide changes in expression of proteins (derived using AMTs). In the initial identification phase, both useful AMTs and other ambiguous putative polypeptides are subjected to MS/MS to confirm protein identification, identify modifications, and so forth.

We have recently shown that the high MMA provided by FTICR opens the opportunity to analyze a number of peptides simultaneously; that is, multiplexed MS/MS [103]. In contrast to the conventional tandem mass spectrometric approach for protein identification, where an individual polypeptide is sequentially selected and dissociated, multiplexed MS/MS selects and dissociates several species simultaneously (Fig. 19). In the case of polypeptides, a limited set of fragmentation pathways is usually favored, and we have found that high MMA allows most of the fragment ions in a MS/MS experiment from a limited set of parent polypeptides to be attributed to a specific parent species [103]. In our initial demonstration, the most abundant species were simultaneously selected using a SWIFT (Fig. 19B), and then subjected to IRMPD (Fig. 19C). In this case, all seven species were uniquely identified from the *C. elegans* full-genome database, which is roughly equivalent in complexity to that projected for a specific differentiated mammalian cell type.

13. Quantitative proteome-wide measurements

A necessary component of effective proteomics involves quantitation of protein abundance, for example, in comparison of two cellular populations as a result of some perturbation. The predominant method for measuring changes in protein expression levels using current proteomic technology (e.g., based on 2D-PAGE separations) is to compare protein spot intensities. Not only does this strategy fail to detect subtle changes in protein expression levels and lower abundance proteins, the precision of the measurements is much less than desired, and it is also difficult to compare results between different labs. To address this need, we developed an approach based on the use of stable-isotope labeling methods to provide effective internal standards for each protein [99]. Our initial application of this approach was in the context of intact

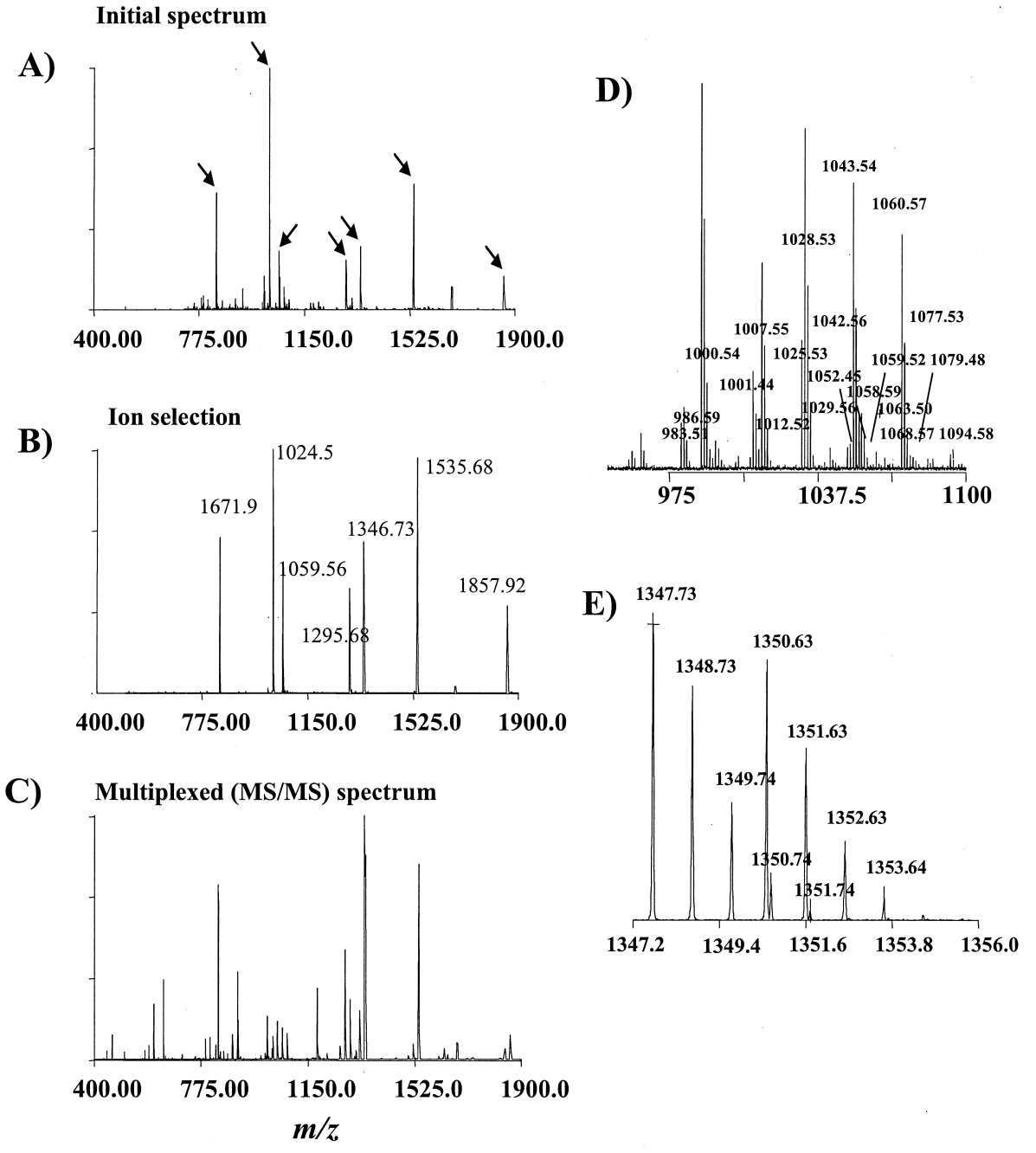

Fig. 19. Demonstration of multiplexed MS/MS using FTICR as a basis for high-throughput protein identification. (A) The seven most abundant ions from the MS spectrum of a mixture were selected using a SWIFT waveform and (B) subsequently subjected to IRMPD (C). Expanded views of the *m/z* range 975–1100 (D) show >20 detected isotopic distributions from a total of 105. The *m/z* range 1347–1356 shows two overlapping but resolved isotopic distributions (E).

protein measurements, but our laboratory and (independently) others have also demonstrated the approach in several different label strategies in conjunction with enzymatic processing [99,101,104]. In our original implementation, two isotopically distinct versions of each protein were generated and analyzed simultaneously, providing calibrants for all detected proteins, thus enabling precise proteome-wide measurement of changes in protein abundance resulting from cellular perturbations [99]. Cells are cultured in stable-isotope-labeled media (e.g., ^{15}N, ^{13}C enriched or ^{15}N, ^{13}C depleted), and the proteins extracted from this source are compared with proteins extracted from cells grown in a natural (or other) isotopic abundance medium.

Our initial demonstration of isotopic labeling strategies for whole proteomes was for the analysis of intact proteins using CIEF-FTICR. These studies examined the cadmium (Cd^{2+}) stress response in *E. coli* K-12 MG1655. In these studies, *E. coli* was grown in both normal (i.e., natural isotopic abundance) and rare-isotope-depleted media. The experimental approach involved mixing two cell populations; thus, the resulting mass spectra show two versions of each protein and the ratio of the two isotopically different and resolvable versions reflects relative protein abundance (its abundance ratio, or AR). A range of responses to Cd^{2+} stress were observed, with proteins being suppressed ($AR < 1$), induced ($AR > 1$), or showing more complex behaviors with time. The ARs for the 200 most abundant proteins detected ranged from <0.1 to >30 [99]. Replicate analyses indicate that the precision of protein abundance ratios was $\pm 6\%$, far better precision than achievable using 2D-PAGE technology and MS methods without the use of internal standards [99].

In an initial demonstration of quantitation with the peptide-AMT approach, we have cultured yeast in a medium containing the natural abundance of the isotopes of nitrogen, that is, ^{14}N (99.6%) and ^{15}N (0.4%), and a second pool (e.g., a mutant) in the same medium enriched in ^{15}N (>98%). Cells grown in normal isotopic abundance and ^{15}N-enriched media were mixed and the soluble proteins extracted. This combined protein extract was analyzed using our capillary LC-FTICR approach. The two isotopic versions of each peptide were easily distinguished because of the distinctive isotopic distribution of the ^{15}N-labeled peptide. An example of two isotopically distinct versions of a peptide is shown in Fig. 20, with the simulated isotopic distribution expected for each (Fig. 20 A, inset). The number of nitrogen atoms present in the peptide is obtained from the mass differences between unlabeled and ^{15}N-labeled versions of the peptide. This knowledge of the number of nitrogen atoms in the peptide provides an additional constraint that is useful for protein identification.

Another approach, developed by Aebersold and coworkers, incorporates an isotopic label on Cys residues after proteins have been extracted [101]. The labeling strategies that use isotopically defined media offer the highest precision (as labeling occurs at the earliest possible point), the broadest proteome coverage, and minimal handling. In contrast, post-isolation isotopic labeling is broadly applicable to proteins extracted from every conceivable source (e.g., tissue samples), can provide a reduced mixture complexity in some labeling schemes, and can aid identification by providing an additional sequence constraint. However, this is obtained at the cost of additional sample processing, somewhat decreased protein coverage, and a possible small decrease in precision (caused, in greatest part, by the variables that precede the labeling step). The use of stable isotope labels with the modified Cys-labeled polypeptides provides a basis for proteome-wide precise quantitation of expression levels and protein identification, as well as a significant reduction in the complexity of the polypeptide mixture when combined with the use of bioaffinity purification methods [101].

In collaboration with Aebersold and coworkers, we have applied this approach in conjunction with FTICR to develop a high-throughput method to obtain simultaneous proteome-wide identification and quantitation of yeast proteins. The additional Cys-constraint generally allows for unambiguous peptide identification at only low parts per million MMA. A few representative spectra showing the presence of differentially labeled Cys-polypeptides, along with their calculated abundance ratios and the yeast peptide identified, are shown in Fig. 20 (B). Of particular note is the fact that

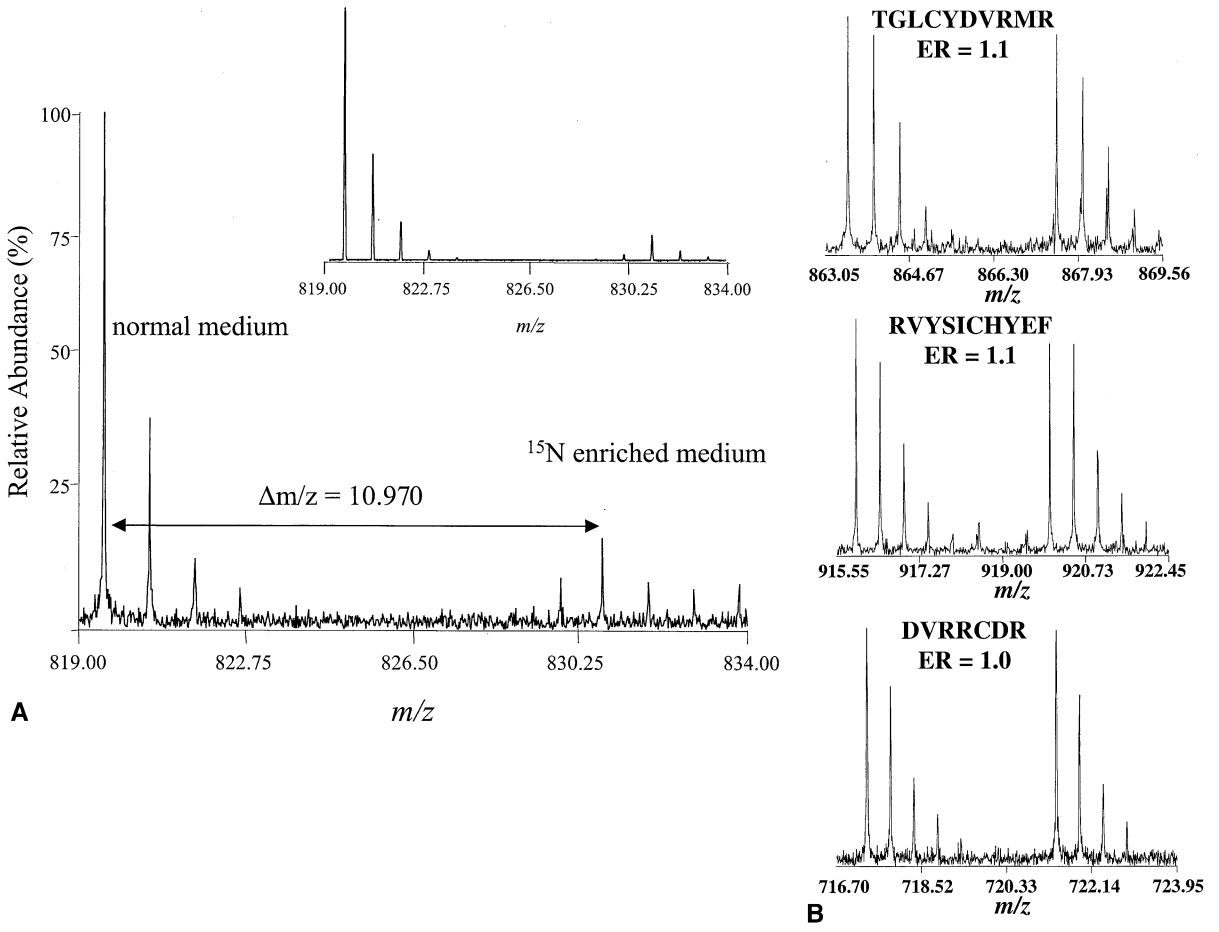

Fig. 20. (A) Mass spectrum displaying the natural isotopic abundance and ^{15}N-labeled versions of a peptide observed in the capillary LC separation of a digested CIEF fraction. (B) Examples of Cys-polypeptides observed in a capillary LC-FTICR analysis of a yeast lysate labeled with the heavy and light isotope-labeled versions of the bioaffinity tag. The presence of two isotopically distinct versions of the same peptide provides the basis for comparative displays of the relative protein expression levels and a measure of the number of nitrogen atoms. The simulated mass spectrum for the two versions of this peptide is also shown (inset).

the measured ARs are consistent with the fact that identical protein fractions were processed for this control experiment. I believe that the use of such labeling approaches, combined with the methods described above, provides a basis for the extension to the much more complex mammalian proteomes.

14. A closing comment

Mass spectrometry is positioned to play a significant role in proteomics as well as the broader spectrum of biological research. Developments in instrumentation and computational technologies, along with the large amounts of genome sequence information continuously being generated for a variety of different cell types and organisms, are leading to revolutionary advances in the use of mass spectrometry for analyzing the protein content of these systems. The possibility now exists for studying hundreds or thousands of proteins within a single experiment, instead of focusing on a single species. This new approach to protein analysis potentially allows for pathways and networks, as well as how they effect one another, to be characterized in hours instead of months or years.

To reach this exciting goal, proteomic studies must be highly sensitive to achieve the broadest possible proteome coverage and also be both quantitative and high throughput. Mass spectrometry appears destined to have an important role in this enterprise. The paradigm shifts once proteins are identified, and I envision that a major application of mass spectrometry will consist of high-throughput comparative studies of proteomes (e.g., from perturbation studies). Finally, the ongoing advances in sensitivity are certain to increasingly enable studies at the single-cell level.

This more prominent role is causing significant changes in the practice of mass spectrometry, involving the increased use of automation, increasingly sophisticated software, and the integration of bioinformatic tools. More fundamental is the increased scope of the enterprise and the highly collaborative nature of research and the different emphasis for the more basic studies in the field of mass spectrometry. While this is fundamentally changing many facets of what we do, how we interact with colleagues, and particularly the nature of the opportunities for those entering the field, there is much fun yet to be had.

Acknowledgements

I gratefully acknowledge an extensive array of collaborators, coworkers, and postdoctoral associates from whom I have learned and benefited in our interactions. Key collaborations referred to in this manuscript have included Reudi Aebersold (University of Washington), Andy Ewing (Pennsylvania State University), Jean Futrell (PNNL), Milton Lee (Brigham Young University), and George Whitesides (Harvard University). Present and past coworkers and postdoctoral associates have included Gordon Anderson, Charles Barinaga, Michael Belov, Scott Berger, Jim Bruce, Ruidan Chen, Xueheng Cheng, Tom Conrads, Charlie Edmonds, Bingbing Feng, Brenda Gale, Dave Gale, Hongying Gao, Dave Goodlett, Michael Gorshkov, Richard Harkewicz, Amy Harms, Steve Hofstadler, Pam Jensen, Taeman Kim, Mary Lipton, Chaunliang Liu, Joe Loo, Rachel Ogorlalek Loo, Suzana Martinovic, Christophe Masselon, Dale Mitchell, Dave Muddiman, Jose Olivares, Ljiljana Pasa Tolic, Alan Rockwood, Joanne Severs, Scott Shaffer, Keqi Tang, Aleksey Tolmachev, Tim Veenstra, Steve Van Orden, Karen Wahl, Jon Wahl, Brian Winger, Dave Wunschel, Quinyuan Wu, Naxing Xu, Rui Zhao, Zack Zhao, and especially my longtime friend and colleague Harold Udseth. I also thank the National Institutes of Health, through NCI (CA81654), NINDS (NS39617), and NCRR (RR12365), for their current support, and the U.S. Department of Energy Office of Biological and Environmental Research for longtime support of my research. Pacific Northwest National Laboratory is operated by Battelle Memorial Institute for the U.S. Department of Energy under contract DE-AC06-76RLO 1830.

References

[1] M. Dole, L.L. Mack, R.L. Hines, R.C. Mobley, L.D. Ferguson, M.B. Alice, J. Chem. Phys. 49 (1968) 2240.
[2] M. Dole, R.L. Hines, Macromol. 1 (1968) 96.
[3] R.D. Smith, A.L. Johnson, Anal. Chem. 53 (1981) 739.
[4] R.D. Smith, J.E. Burger, A.L. Johnson, Anal. Chem. 53 (1981) 1603.
[5] R.D. Smith, J.C. Fjeldsted, M.T. Lee, W.D. Felix, Anal. Chem. 54 (1982) 1883.
[6] R.D. Smith, J.C. Fjeldsted, M.L. Lee, J. Chromatogr. 247 (1982) 231.
[7] R.D. Smith, H.R. Udseth, Anal. Chem. 55 (1983) 2266.
[8] R.D. Smith, B.W. Wright, C.R. Yonker, Anal. Chem. (1988) A1323.
[9] R.C. Petersen, D.W. Matson, R.D. Smith, J. Am. Chem. Soc. 108 (1986) 2100.
[10] D.W. Matson, R.C. Petersen, R.D. Smith, Materials Lett. 4 (1986) 429.
[11] D.W. Matson, R.C. Petersen, R.D. Smith, Ceramic Powder Sci. 21 (1987) 109–120.
[12] C.R. Yonker, B.W. Wright, H.R. Udseth, R.D. Smith, Ber. Bunsen-Ges. Phys. Chem. 88 (1984) 908.
[13] R.D. Smith, H.R. Udseth, R.N. Hazlett, Fuel 64 (1985) 810.
[14] R.W. Gale, J.L. Fulton, R.D. Smith, J. Am. Chem. Soc. 109 (1987) 920.
[15] R.W. Gale, J.L. Fulton, R.D. Smith, Anal. Chem. 59 (1987) 1977.
[16] R.D. Smith, C.R. Yonker, J.L. Fulton, J.M. Tingey, J. Supercritical Fluids 1 (1988) 7.
[17] R.D. Smith, J.L. Fulton, H.K. Jones, Sep. Sci. Tech. 23 (1988) 2015.
[18] J.L. Fulton, R.D. Smith, J. Phys. Chem. 92 (1988) 2903.

[19] J.W. Jorgensen, K.D. Lukaes, Science 222 (1984) 266.
[20] M. Yamashita, J.B. Fenn, J. Phys. Chem. 88 (1984) 4671.
[21] J.A. Olivares, N.T. Nguyen, C.R. Yonker, R.D. Smith, Anal. Chem. 59 (1987) 1230.
[22] R.D. Smith, C.J. Barinaga, H.R. Udseth, Anal. Chem. 60 (1988) 1948.
[23] J.H. Wahl, D.R. Goodlett, H.R. Udseth, R.D. Smith, Anal. Chem. 64 (1992) 3194.
[24] J.H. Wahl, D.R. Goodlett, H.R. Udseth, R.D. Smith, Electrophoresis 14 (1993) 448.
[25] D.R. Goodlett, J.H. Wahl, H.R. Udseth, R.D. Smith, J. Microcolumn Sep. 5 (1993) 57.
[26] R.D. Smith, J.H. Wahl, D.R. Goodlett, S.A. Hofstadler, Anal. Chem. 65 (1993) 574A.
[27] J.H. Wahl, D.C. Gale, R.D. Smith, J. Chromatography A 659 (1994) 217.
[28] C.K. Meng, M. Mann, J.B. Fenn, Z. Phys. D Atom Mol. Clusters 10 (1988) 361.
[29] S.F. Wong, C.K. Meng, J.B. Fenn, J. Phys. Chem. 92 (1988) 546.
[30] J.A. Loo, H.R. Udseth, R.D. Smith, Rapid Commun. Mass Spectrom. 2 (1988) 207.
[31] R.D. Smith, C.J. Barinaga, H.R. Udseth, J. Phys. Chem. 93 (1989) 5019.
[32] C.J. Barinaga, C.G. Edmonds, H.R. Udseth, R.D. Smith, Rapid Commun. Mass Spectrom. 3 (1989) 160.
[33] R.D. Smith, C.J. Barinaga, Rapid Commun. Mass Spectrom. 4 (1990) 54.
[34] R.D. Smith, J.A. Loo, C.G. Edmonds, C.J. Barinaga, H.R. Udseth, Anal. Chem. 62 (1990) 882.
[35] J.A. Loo, C.G. Edmonds, R.D. Smith, Science 248 (1990) 201.
[36] J.A. Loo, C.G. Edmonds, H.R. Udseth, R.D. Smith, Anal. Chim. Acta 241 (1990) 167.
[37] J.A. Loo, C.G. Edmonds, R.D. Smith, Anal. Chem. 63 (1991) 2488.
[38] J.A. Loo, C.G. Edmonds, R.D. Smith, Anal. Chem. 65 (1993) 425.
[39] A.L. Rockwood, M. Busman, R.D. Smith, Int. J. Mass Spectrom. Ion Processes 111 (1991) 103.
[40] A.L. Rockwood, M. Busman, R.D. Smith, Rapid Commun. Mass Spectrom. 5 (1991) 582.
[41] M. Busman, A.L. Rockwood, R.D. Smith, J. Phys. Chem. 96 (1992) 2397.
[42] R.R. Ogorzalek Loo, H.R. Udseth, R.D. Smith, J. Phys. Chem. 95 (1991) 6412.
[43] R.R. Ogorzalek Loo, H.R. Udseth, R.D. Smith, J. Am. Soc. Mass Spectrom. 3 (1992) 695.
[44] J.A. Loo, H.R. Udseth, R.D. Smith, Biomed. Environ. Mass Spectrom. 17 (1988) 411.
[45] J.A. Loo, C.G. Edmonds, H.R. Udseth, R.D. Smith, Anal. Chem. 62 (1990) 693.
[46] B. Ganem, Y.-T. Li, J.D. Henion, J. Am. Chem. Soc. 113 (1991) 6294.
[47] B.L. Schwartz, K.J. Light-Wahl, R.D. Smith, J. Am. Soc. Mass Spectrom. 5 (1994) 201.
[48] B.L. Schwartz, D.C. Gale, R.D. Smith, A. Chilkoti, P.S. Stayton, J. Mass Spectrom. 30 (1995) 1095.
[49] D.C. Gale, D.R. Goodlett, K.J. Wahl, R.D. Smith, J. Am. Chem. Soc. 116 (1994) 6027.
[50] X. Cheng, A.C. Harms, P.N. Goudreau, T.C. Terwilliger, R.D. Smith, Proc. Natl. Acad. Sci. 93 (1996) 7022.
[51] X. Cheng, P.E. Morin, A.C. Harms, J.E. Bruce, Y. Ben-David, R.D. Smith, Anal. Biochem. 239 (1996) 35.
[52] B.E. Winger, K.J. Light-Wahl, R.R.O. Loo, H.R. Udseth, R.D. Smith, J. Am. Soc. Mass Spectrom. 4 (1993) 536.
[53] R.D. Smith, K.J. Light-Wahl, Biol. Mass Spectrom. 22 (1993) 493.
[54] K.J. Light-Wahl, B.L. Schwartz, R.D. Smith, J. Am. Chem. Soc. 116 (1994) 5271.
[55] E.R. Williams, J. Mass Spectrom. 31 (1996) 831.
[56] R.D. Smith, H.R. Udseth, A.L. Rockwood, B.E. Winger, S.A. Hofstadler, D.R. Goodlett, K.J. Light-Wahl, in: T. Matsuo (Ed.), Proceedings of the Kyoto '92 International Conference on Biological Mass Spectrometry, Kyoto, Japan (1992), pp. 16–17.
[57] R.D. Smith, J.E. Bruce, Q. Qu, Q.P. Lei, Chem. Soc. Rev. 26 (1997) 191.
[58] R.D. Smith, J.E. Bruce, Q. Wu, X. Cheng, S.A. Hofstadler, G.A. Anderson, R. Chen, R. Bakhtiar, S.O. Van Orden, D.C. Gale, M.G. Sherman, A.L. Rockwood, H.R. Udseth, in: A. Burlingame, S. Carr (Eds.), Mass Spectrometry in the Biological Sciences, Humana Press, Totowa, NJ, 1996, p. 25.
[59] M.B. Comisarow, A.G. Marshall, Chem. Phys. Lett. 25 (1974) 282.
[60] K.D. Henry, E.R. Williams, B.H. Wang, F.W. McLafferty, J. Shabanowitz, D.F. Hunt, Proc. Natl. Acad. Sci. 86 (1989) 9075.
[61] K.D. Henry, J.P. Quinn, F.W. McLafferty, J. Am. Chem. Soc. 113 (1991) 5447.
[62] B.E. Winger, S.A. Hofstadler, J.E. Bruce, H.R. Udseth, R.D. Smith, J. Am. Soc. Mass Spectrom. 4 (1993) 566.
[63] X. Cheng, R. Chen, J.E. Bruce, B.L. Schwartz, G.A. Anderson, S.A. Hofstadler, D.C. Gale, R.D. Smith, J. Gao, G.B. Sigal, M. Mammen, G.M. Whitesides, J. Am. Chem. Soc. 117 (1995) 8859.
[64] J. Gao, X. Cheng, R. Chen, G.B. Sigal, J.E. Bruce, B.L. Schwartz, S.A. Hofstadler, G.A. Anderson, R.D. Smith, G.M. Whitesides, J. Med. Chem. 39 (1996) 1949.
[65] Q. Wu, J. Gao, D. Joseph-McCarthy, G.B. Sigal, J.E. Bruce, G.M. Whitesides, R.D. Smith, J. Am. Chem. Soc. 119 (1997) 1157.
[66] Q. Wu, Q.P. Lei, R.D. Smith, J. Gao, J. Carbeck, G.M. Whitesides, Biophys. J. 76 (1999) 3253.
[67] J.E. Bruce, G.A. Anderson, R. Chen, X. Cheng, D.C. Gale, S.A. Hofstadler, B.L. Schwartz, R.D. Smith, Rapid Commun. Mass Spectrom. 9 (1995) 644.
[68] J.E. Bruce, G.A. Anderson, S.A. Hofstadler, S.L. Van Orden, M.S. Sherman, A.L. Rockwood, R.D. Smith, Rapid Commun. Mass Spectrom. 7 (1993) 914.
[69] R.D. Smith, J.A. Loo, C.G. Edmonds, C.J. Barinaga, H.R. Udseth, J. Chromatography 516 (1990) 157.
[70] D.C. Gale, R.D. Smith, Rapid Commun. Mass Spectrom. 7 (1993) 1017.
[71] S.A. Hofstadler, F.D. Swanek, D.C. Gale, A.G. Ewing, R.D. Smith, Anal. Chem. 67 (1995) 1477.

[72] S.A. Hofstadler, J.C. Severs, R.D. Smith, F.D. Swanek, A.G. Ewing, Rapid Commun. Mass Spectrom. 10 (1996) 919.
[73] S.A. Hofstadler, J.C. Severs, R.D. Smith, F.D. Swanek, A.G. Ewing, J. High Resol. Chromatogr. 19 (1996) 617.
[74] J.E. Bruce, G.A. Anderson, R.D. Smith, Anal. Chem. 68 (1996) 534.
[75] M.V. Gorshkov, L. Pasa-Tolic, J.E. Bruce, G.A. Anderson, R.D. Smith, Anal. Chem. 69 (1997) 1307.
[76] M.E. Belov, M.V. Gorshkov, H.R. Udseth, G.A. Anderson, R.D. Smith, Anal. Chem. 72 (2000) 2271.
[77] S.A. Shaffer, K. Tang, G.A. Anderson, D.C. Prior, H.R. Udseth, R.D. Smith, Rapid Commun. Mass Spectrom. 11 (1997) 1813.
[78] S.A. Shaffer, D.C. Prior, G.A. Anderson, H.R. Udseth, R.D. Smith, Anal. Chem. 70 (1998) 4111.
[79] W. Paul, B. Lucke, S. Schlemmer, D. Gerlich, Int. J. Mass Spectrom. Ion Processes 150 (1995) 373.
[80] D. Gerlich, State-Selected and State-to-State Ion-Molecule Reaction Dynamics. Part 1. Experim. New York, Wiley, LXXXII, 1992, p. 1.
[81] D.J. Douglas, J.B. French, J. Am. Soc. Mass Spectrom. 3 (1992) 398.
[82] A.V. Tolmachev, H.R. Udseth, R.D. Smith, Anal. Chem. 72 (2000) 970–978.
[83] S.A. Shaffer, A. Tolmachev, D.C. Prior, G.A. Anderson, H.R. Udseth, R.D. Smith, Anal. Chem. 71 (1999) 2957.
[84] M.E. Belov, M.V. Gorshkov, H.R. Udseth, G.A. Anderson, A.V. Tolmachev, D.C. Prior, R. Harkewicz, R.D. Smith, J. Am. Soc. Mass Spectrom. 11 (2000) 19.
[85] T. Kim, A.V. Tolmachev, R. Harkewicz, D.C. Prior, G.A. Anderson, H.R. Udseth, R.D. Smith, T.H. Bailey, S. Rakov, J.H. Futrell, Anal. Chem. 72 (2000) 2247.
[86] R.M. Weisskoff, G.P. Lafyatis, K.R. Boyce, E.A. Cornell, R.W. Flanagan, D.E. Pritchard, J. Appl. Phys. 63 (1988) 4599.
[87] C.G. Edmonds, R.D. Smith, L.M. Smith, Proceedings of the Workshop on The Relevance of Mass Spectrometry to DNA Sequence Determination: Research Needs for the Human Genome Program (U.S. Department of Energy Report CONF-9004251) November (1990).
[88] R.D. Smith, X. Cheng, J.E. Bruce, S.A. Hofstadler, G.A. Anderson, Nature 369 (1994) 137.
[89] J.E. Bruce, X. Cheng, R. Bakhtiar, Q. Wu, S.A. Hofstadler, G.A. Anderson, R.D. Smith, J. Am. Chem. Soc. 116 (1994) 7839.
[90] R. Chen, X. Cheng, D.W. Mitchell, S.A. Hofstadler, A.L. Rockwood, Q. Wu, M.G. Sherman, R.D. Smith, Anal. Chem. 67 (1995) 1159.
[91] R. Chen, Q. Wu, D.W. Mitchell, S.A. Hofstadler, A.L. Rockwood, R.D. Smith, Anal. Chem. 66 (1994) 3964.
[92] X. Cheng, D.G. Camp II, Q. Wu, R. Bakhtiar, D.L. Springer, B.J. Morris, J.E. Bruce, G.A. Anderson, C.G. Edmonds, R.D. Smith, Nucleic Acid Res. 24 (1996) 2183.
[93] D.W. Mitchell, R.D. Smith, J. Mass Spectrom. 31 (1996) 771.
[94] J.E. Bruce, G.A. Anderson, H.R. Udseth, R.D. Smith, Anal. Chem. 70 (1998) 519.
[95] J. Klose, U. Kobalz, Electrophoresis 16 (1995) 1034.
[96] A. Shevchenko, M. Wilm, O. Vorm, M. Mann, Anal. Chem. 68 (1996) 850.
[97] Y. Shen, F. Xiang, T.D. Veenstra, E.N. Fung, R.D. Smith, Anal. Chem. 71 (1999) 5348.
[98] P.K. Jensen, L. Pasa-Tolic, G.A. Anderson, J.A. Horner, M.S. Lipton, J.E. Bruce, R.D. Smith, Anal. Chem. 71 (1999) 2076.
[99] L. Pasa-Tolic, P.K. Jensen, G.A. Anderson, M.S. Lipton, K.K. Peden, S. Martinovic, N. Tolic, J.E. Bruce, R.D. Smith, J. Am. Chem. Soc. 121 (1999) 7949.
[100] T.D. Veenstra, S. Martinovic, G.A. Anderson, L. Pasa-Tolic, R.D. Smith, J. Am. Soc. Mass Spectrom. 11 (2000) 78.
[101] S.P. Gygi, S. Rist, S.A. Gerber, F. Turecek, M.H. Gelb, R. Aebersold, Nature Biotech. 17 (1999) 994.
[102] L. Pasa Tolic, M.S. Lipton, P.K. Jensen, G.A. Anderson, Y. Shen, T. Conrads, H.R. Udseth, R. Harkewicz, C. Masselon, T.D. Veenstra, R.D. Smith, in preparation.
[103] C. Masselon, G.A. Anderson, R. Harkewicz, J.E. Bruce, L. Pasa-Tolic, R.D. Smith, Anal. Chem. 72 (2000) 3349.
[104] Y. Oda, K. Huang, F.R. Cross, D. Cowburn, B.T. Chait, Proc. Natl. Acad. Sci. 96 (1999) 6591.