Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 11451–11455, December 1995
Chemistry

Sequence verification of human creatine kinase (43 kDa) isozymes by high-resolution tandem mass spectrometry

(electrospray ionization/Fourier-transform mass spectrometry/protein sequencing)

TROY D. WOOD*, LORENZO H. CHEN†, CAMILLE B. WHITE†, PATRICIA C. BABBITT†, GEORGE L. KENYON†‡, AND FRED W. MCLAFFERTY*‡

*Department of Chemistry, Baker Laboratory, Cornell University, Ithaca, NY 14853-1301; and †Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143

Contributed by Fred W. McLafferty, September 5, 1995

ABSTRACT **Amino acid sequencing by recombinant DNA technology, although dramatically useful, is subject to base reading errors, is indirect, and is insensitive to posttranslational processing. Mass spectrometry techniques can provide molecular weight data from even relatively large proteins for such cDNA sequences and can serve as a check of an enzyme's purity and sequence integrity. Multiply-charged ions from electrospray ionization can be dissociated to yield structural information by tandem mass spectrometry, providing a second method for gaining additional confidence in primary sequence confirmation. Here, accurate (±1 Da) molecular weight and molecular ion dissociation information for human muscle and brain creatine kinases has been obtained by electrospray ionization coupled with Fourier-transform mass spectrometry to help distinguish which of several published amino acid sequences for both enzymes are correct. The results herein are consistent with one published sequence for each isozyme, and the heterogeneity indicated by isoelectric focusing due to 1-Da deamidation changes. This approach appears generally useful for detailed sequence verification of recombinant proteins.**

Recombinant DNA technology now provides a standard method for protein sequencing by identifying genes that encode the protein (1). However, errors in nucleotide deletion, insertion, or misidentification do occur, and posttranslational processing of the protein cannot be identified. For verification of a cDNA sequence, molecular weight (MW) determination by mass spectrometry (MS) is increasingly valuable. Soft ionization methods such as matrix-assisted laser desorption/ionization (2) and electrospray ionization (ESI) (3) can generate molecular ions for enzymes as large as several hundred kilodaltons. However, most mass analyzers used for these techniques have a resolving power of 10^2–10^3; although ±0.01% agreement between expected and found values is often reported, failure to resolve cationized or noncovalent adducts of the molecular ion can lead to far larger, and unanticipated, errors (4, 5). Promising for this problem, Fourier-transform mass spectrometry (FTMS) (6) coupled with ESI exhibits a resolving power of 10^5–10^6 for proteins as large as albumin (7); this not only identifies adduct peaks with ease but also provides isotopically resolved $(M + nH)^{n+}$ peaks that yield directly the number of charges (z) of the measured m/z value (8). Dissociation of the electrosprayed $(M + nH)^{n+}$ ions of an enzyme by tandem mass spectrometry (9, 10) provides fragment masses to check isomeric integrity and restrict location of sequence errors (11). Tandem ESI/FTMS has been used for structural characterization of proteins as large as 67 kDa (7, 12, 13), including heterogeneity and active site location >40 kDa (11, 14). Here, tandem ESI/FTMS is applied to isozymes of creatine kinase (CK; EC 2.7.3.2), for which conflicting DNA sequences have been proposed. Based on these sequences, predicted MW values of 42,915 (15) and 42,970 (16) are reported for human muscle CK and of 42,512 (17), 42,622 (18), and 42,669 (19) for human brain CK. Further, isoelectric focusing (IEF) shows both isozymes to be mixtures and enzymatically active.

EXPERIMENTAL PROCEDURES

Materials and Sample Preparation. Overexpression of human muscle and brain CK isozymes in *Escherichia coli* has been described elsewhere (ref. 14; L.H.C., C.B.W., P.C.B., and G.L.K., unpublished data). Purification and isolation of recombinant human CKs were similar to that described for rabbit muscle CK (21) except that cells were not subjected to 50–75% ethanol fractionation after lysis.

ESI/FTMS. Human muscle and brain CKs were prepared as 20 μM solutions in methanol/water/acetic acid, 76:22:2 (vol/vol), and electrosprayed through a syringe needle biased at 2.6–2.9 kV (flow rates, 0.5–1.0 μl/min). The resulting ions were transported by three radio-frequency-only quadrupoles through five stages of differential pumping to the trapped ion cell in a 6.1-Tesla magnetic field of a modified Extrel Waters FTMS-2000 (22). For tandem mass spectrometry, ions were dissociated by nozzle/skimmer (NS) (13, 23), by infrared multiphoton dissociation (IRMPD) with a Synrad (Bothell, WA) model 48-2 25-W continuous wave CO_2 laser (24), or by sustained off-resonance irradiation–collisionally activated dissociation (25, 26) of charge states isolated by stored waveform inverse Fourier transform (27).

RESULTS AND DISCUSSION

Human Muscle CK. The mass spectrum (Fig. 1*B*) of heterogeneous human muscle CK under gentle source conditions chiefly shows intact $(M + nH)^{n+}$ peaks and little evidence of fragmentation, similar to results for rabbit muscle CK and its Cys-282 → Ser mutant (14). Geometric deconvolution of the resolved isotopic peak abundances (Fig. 1*C*, n = 20+ to 48+) gives a MW value of 42,970, consistent with one (16) of the published sequences ($^{13}C_{27}$, 42,970) but markedly different from that expected ($^{13}C_{27}$, 42,915) by another (15). Obtaining accurate MW information would be difficult for low-resolution MS because of Na adduction (replacing H^+ with Na^+) and water loss, which would make the $(M + nH)^{n+}$ peak give an unresolved mass of 42,981 (Fig. 1*A*). A small abundance of the α_2 dimer can also be observed at odd charge states (e.g., 43+,

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: CK, creatine kinase; ESI, electrospray ionization; FTMS, Fourier-transform mass spectrometry; IRMPD, infrared multiphoton dissociation; IEF, isoelectric focusing; NS, nozzle/skimmer; MW, molecular weight.

‡To whom reprint requests should be addressed.

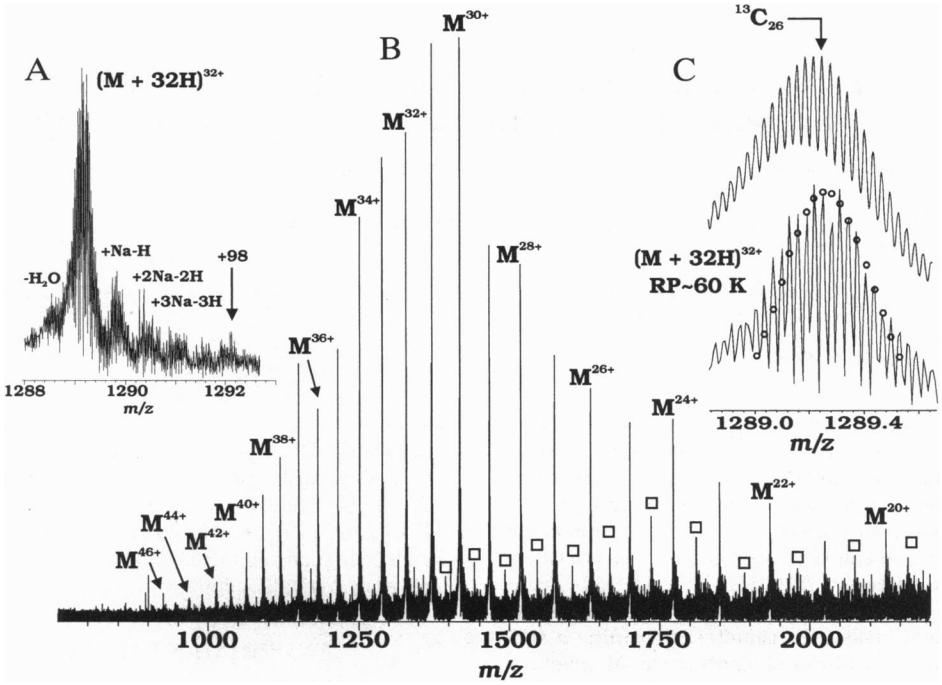

FIG. 1. Human muscle CK. (*A*) Expanded m/z scale for $(M + 35H)^{35+}$, revealing contaminants due to salt adduction and water loss. (*B*) ESI/FTMS mass spectrum, 100-V NS, and sum of 25 scans (□, dimer ions). (*C*) Expanded m/z scale, $(M + nH)^{n+}$ (60,000 resolving power). Lower curve, $n = 35$, isotopic distribution predicted from cDNA sequence of Trask *et al.* (16) (○); upper curve, geometric deconvolution of $n = 20–48$.

45+, 47+, etc.; Fig. 1*A*; the dimer isotopic peaks of even charge states lower the apparent resolving power).

The two published human muscle CK gene sequences (15, 16) have eight base-pair differences in the coding region; six of these are nonconservative and lead to differences in the expressed sequence at amino acid residues 46 [Thr → Ile; Trask *et al.* (16) → Perryman *et al.* (15)], 129 (Arg → Pro), 192 (Leu → Gln), 209 (Asp → His), 214 (Arg → Pro), and 323 (Gly → Ala). Though the predicted MW of one published human muscle CK sequence matches the observed MW (16), this alone does not ensure that the sequence is correct, as shown in a previous ESI/FTMS study of thiaminase I (11). Confidence in the sequence assignment can be enhanced further from ion fragment masses. Increasing the NS potential (140 V) extensively dissociates the $(M + nH)^{n+}$ ions, and the generated fragment masses are tabulated in Fig. 2. The mass sum of one pair, 24,094 ($^{13}C_{15}$) + 18,878 ($^{13}C_{12}$) = 42,972, agrees with the measured MS ±2 Da. These correspond to b_{212} (N-terminal) and y_{168} fragments, respectively, of the Trask sequence (16), a fragmentation also found for NS of rabbit muscle CK (≈95% sequence homology with human muscle CK) (14). C-terminal ion fragments as large as y_{48} cannot distinguish between the two sequences since both are identical in this region; however, the y_{98} mass (Fig. 2) is consistent with Gly-323, not Ala-323; based on this and the mass of y_{168} (Fig. 2), the only reported sequence difference between residues 213 and 282 correlates with Arg-214, not Pro-214. Fragment masses for b_{23}, b_{24}, and b_{43} are in agreement with both sequences as they are identical

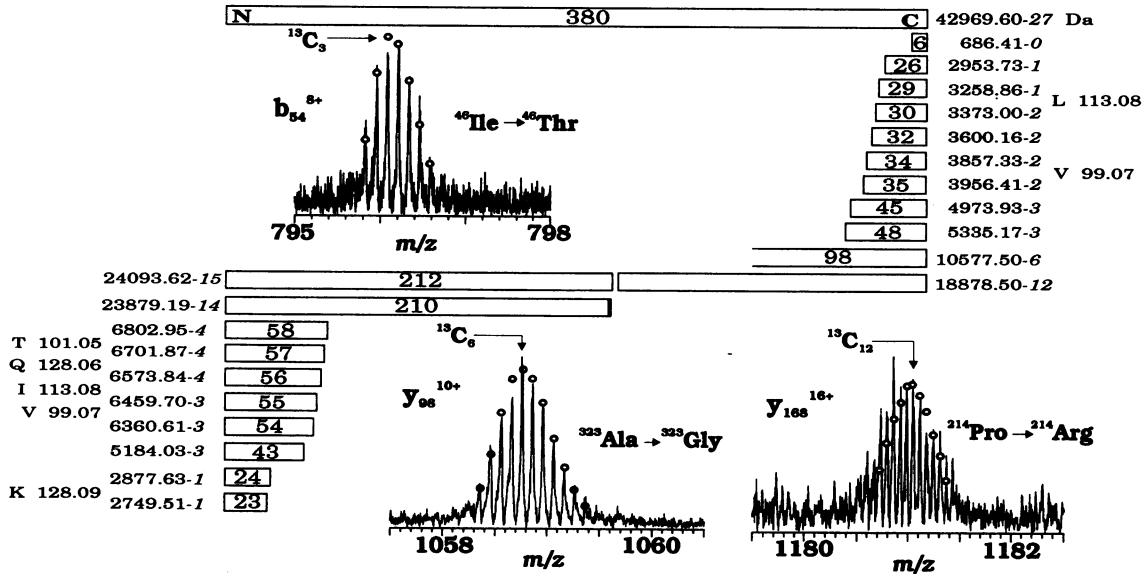

FIG. 2. Composite of NS spectral data of human muscle CK. External columns: expected mass differences for indicated amino acids. (*Insets*) Fragment isotopic distribution as described in Fig. 1.

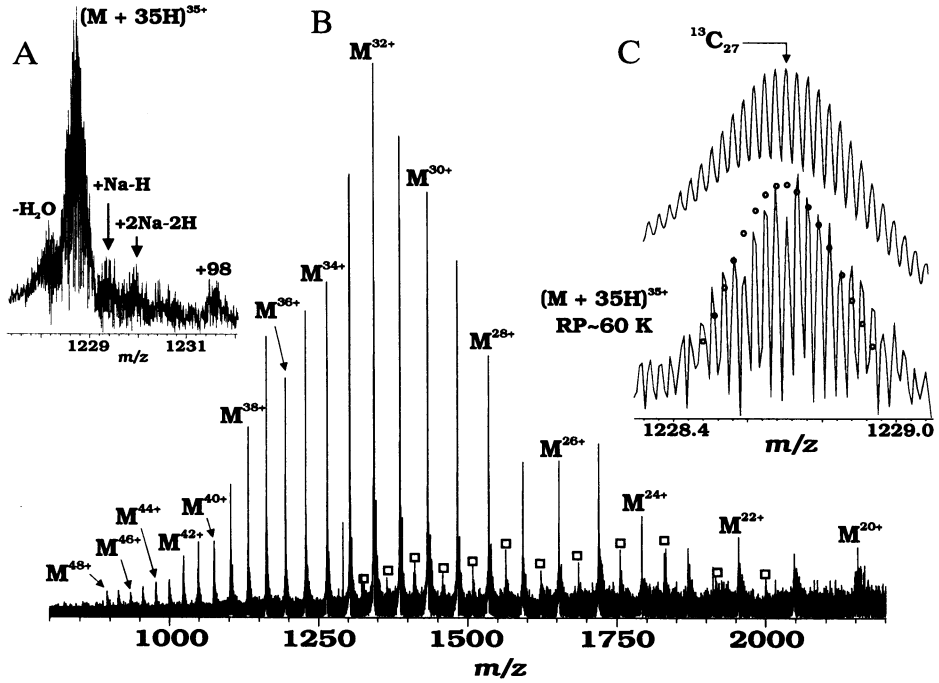

FIG. 3. Human brain CK. (*A*) Expanded m/z scale for $(M + 32H)^{32+}$, revealing contaminants due to salt adduction and water loss. (*B*) ESI/FTMS mass spectrum, 100-V NS, and sum of 25 scans (□, dimer ions). (*C*) Expanded m/z scale, $(M + nH)^{n+}$ (60,000 resolving power). Lower curve, $n = 32$, isotopic distribution predicted from cDNA sequence of Mariman *et al.* (17) (○); upper curve, geometric deconvolution of $n = 20–47$.

in this region; however, masses corresponding to fragments b_{54}–b_{58} indicate that the correct assignment at residue 46 is Thr-46, not Ile-46. Though no amide bond dissociations are observed in the region between b_{58} and b_{210}, the mass of the latter is consistent with the reported amino acid differences at residues 129 (Arg → Pro), 192 (Leu → Gln), and 209 (Asp → His) (16).

Heterogeneity. The masses of all the fragments except for b_{212} ($^{13}C_{15}$, ±2 Da) match that expected within ±1 Da (16). However, IEF of human muscle CK shows at least three components with enzymatic activity (data not shown); not only do all three components have MW values within ±2 Da of the values here, but their fragments are also within ±1 Da. Heterogeneity from Gln/Glu and Asn/Asp ($NH_2 \rightarrow OH$; 16 → 17 Da) differences (28) was indicated from rabbit muscle CK ESI/FTMS data (14). However, the measured MW from ESI/FTMS is not higher than expected from the identified cDNA sequence for human muscle CK (16). There are stretches in the sequence between amide bond cleavages (the longest is 152 residues) where a number of amidation/deamidation steps could occur to give the predicted fragment masses and might explain the nature of the multiple IEF bands. Any structural differences present are probably subtle differences of the sequence in Trask *et al.* (16). Interestingly, this sequence has better sequence homology with muscle CK sequences of other species (29–33), than the sequence in Perryman *et al.* (15), suggesting that the latter contains gene sequencing misreadings (e.g., GC transposition) and that it is not a polymorphic CK isozyme.

Human Brain CK. As with rabbit and human muscle CK, the mass spectrum of human brain CK (Fig. 3*B*) shows intact $(M + nH)^{n+}$ peaks and little fragmentation. Water losses and sodium and phosphate (+98) adducts are distinguished (Fig. 3*A*) and the α_2 dimer can be observed at odd charge states (e.g., 47+, 49+, 51+, etc.). Geometric deconvolution (Fig. 3*C*) of the resolved isotopic peak abundances ($n = 20+$ to 47+) gives a MW value of 42,512, which matches the MW predicted ($^{13}C_{26}$, 42,512) based on one sequence (17) while differing significantly ($^{13}C_{26}$, 42,669, and $^{13}C_{26}$, 42,622, respectively) from two others (18, 19).

Interestingly, NS dissociation of human brain CK (150 V) produces limited fragmentation from the N terminus (b_{42}, b_{43}, b_{54}–b_{58}) and none containing the C terminus; under similar NS conditions, rabbit (14) and human muscle CK produced extensive fragmentation from both termini. Human brain CK is ≈80% identical to the muscle isozyme (17), so it is possible that structural differences between these lead to different fragmentation under identical conditions. The three published sequences of human brain CK have the following differences. Kaye *et al.* (19) agree completely with Mariman *et al.* (17) except at residues 77 (assigned Gly → Asp) and 129 (Gly → Arg); Villarrel-Levy *et al.* (18) agree with Mariman *et al.* (17) except at residues 40 (Asp → Glu), 41 (Val → Leu), 97 (Arg → Gly), 98 (Arg → Gly), 104 (Asp → Glu), 105 (Asp → His), 131 (Ala → Arg), 214 (Ala → Arg), 215 (Arg → Gly), and 296

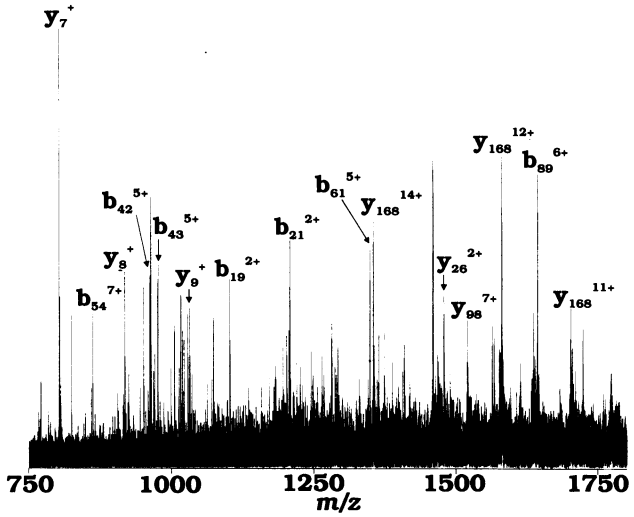

FIG. 4. IRMPD of human brain CK (sum of five scans).

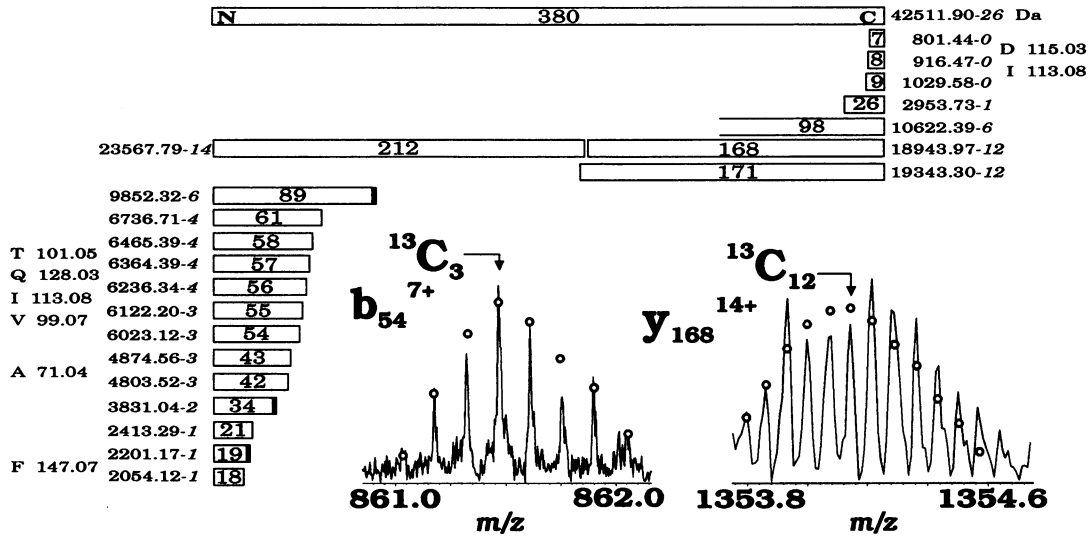

FIG. 5. Composite of NS, IRMPD, and sustained off-resonance irradiation spectral data of human brain CK, correlated with the cDNA sequence of Mariman *et al.* (17) (heavy vertical lines, cleavages N-terminal to proline). External columns and insets are as in Fig. 2.

(Asp → His). The observed fragment masses from NS rule out one published sequence (18) but cannot distinguish between the other two (17, 19), which are identical through b_{58} (see Fig. 5). Sustained off-resonance irradiation–collisionally activated dissociation of the isolated 31+ charged state generated one complementary pair of fragments, b_{212}/y_{163} (23,568 + 18,943), but with relatively low signal/noise ratio; the masses of the fragments and their sum agree with theory ±1 Da for one sequence (17) and ruled out the others based on the b_{212} mass (18, 19). Additional confidence in these conclusions was drawn from fragmentation data by using IRMPD (Fig. 4). Dissociation by IRMPD produced extensive fragmentation, including y-type fragments not observed by NS, e.g., y_{168} (Fig. 5 *Inset*). IRMPD produced b_{61} and b_{89} fragments; while two of the sequences are identical through b_{61} (17, 19), their masses for b_{89} differ by 58 Da since either a Gly or Asp at residue 77 is predicted; the observed mass of 9852 is consistent with Asp-77.

The ESI/FTMS MW and ion fragmentation data (summarized in Fig. 5) are consistent with only one published sequence of human brain CK, namely that of Mariman *et al.* (17). The excellent agreement between the observed MW and fragment masses (±1 Da) vs. that predicted from this sequence gives high confidence in its correctness. IEF of human brain CK shows evidence of unresolved isoforms (data not shown) and multiple peaks (all enzymatically active) when subjected to column chromatography. There is no evidence of heterogeneity in the ESI/FTMS results here; however, MW measurements of human brain CK 3 months after those in Fig. 3 showed an increase in MW of ≈2 Da; this suggests that chemical deamidations, which would increase the mass of 1 Da, may also occur over an extended period of time. Unfortunately, the protein degrades over time, and the only fragment mass observed (y_{168}, low signal/noise ratio) did not appear to exhibit a mass increase, through from the limited signal/noise ratio, a 1- to 2-Da mass increase in this fragment cannot be ruled out. Nonetheless, this result suggests heterogeneity in human brain CK may be due to chemical deamidations. There is better sequence alignment for this human brain CK sequence (17) vs. brain CK sequences from other species (34–36) than for the other two proposed sequences (18, 19); this suggests that the latter proposed sequences also result from nucleotide sequencing errors and are not true polymorphic isozymes.

The overall ESI/FTMS results for both human muscle and brain CKs show it is possible to distinguish between proposed sequences based on MW and ion fragmentation in the MS. Several key advantages to this approach vs. chemical degradation procedures are that MW for the intact molecule and masses of fragment ions can be obtained with high speed (1- to 5-s acquisitions) and sensitivity [<1 pmol consumed for each scan; fmol samples can now be analyzed (20)]. Because of these and the previously mentioned advantages of high-resolution ESI/FTMS, this approach should become increasingly used to distinguish among proposed sequences and for assessment of recombinant protein purity and structural integrity.

We thank P. B. O'Connor for assistance and discussions on deconvolution methods. We also thank D. Aaserud, E. Fridricksson, Z. Guan, N. L. Kelleher, D. P. Little, M. W. Senko, and G. A. Valaskovic for helpful discussions. Financial support was from the National Institutes of Health (F.W.M.), Public Health Service (G.L.K.), and an American Heart Association Postdoctoral Fellowship (to L.H.C.).

1. Hunkapiller, T., Kaiser, R. J., Koop, B. F. & Hood, L. (1991) *Science* **254,** 59–67.
2. Hillenkamp, F., Karas, M., Beavis, R. C. & Chait, B. T. (1991) *Anal. Chem.* **63,** 1193A–1203A.
3. Fenn, J. B., Mann, M., Meng, C. K. & Wong, S. F. (1990) *Mass Spectrom. Rev.* **9,** 37–70.
4. Wang, R. & Chait, B. T. (1994) *Curr. Opin. Biotechnol.* **5,** 77–84.
5. McLafferty, F. W. (1994) *Acc. Chem. Res.* **27,** 379–386.
6. Marshall, A. G. & Grosshans, P. B. (1991) *Anal. Chem.* **63,** 215A–229A.
7. Speir, J. P., Senko, M. W., Little, D. P., Loo, J. A. & McLafferty, F. W. (1995) *J. Mass Spectrom.* **30,** 39–42.
8. Henry, H. D., Quinn, J. P. & McLafferty, F. W. (1991) *J. Am. Chem. Soc.* **113,** 5447–5449.
9. Smith, R. D., Baringa, C. J. & Udseth, H. R. (1989) *J. Phys. Chem.* **93,** 5019–5022.
10. Loo, J. A., Quinn, J. P., Ryu, S. I., Henry, K. D., Senko, M. W. & McLafferty, F. W. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 286–289.
11. Kelleher, N. L., Costello, C. A., Begley, T. P. & McLafferty, F. W. (1995) *J. Am. Soc. Mass Spectrom.* **6,** 981–984.
12. Senko, M. W., Beu, S. C. & McLafferty, F. W. (1994) *Anal. Chem.* **66,** 415–417.
13. O'Connor, P. B., Speir, J. P., Senko, M. W., Little, D. P. & McLafferty, F. W. (1995) *J. Mass Spectrom.* **30,** 88–93.
14. Wood, T. D., Chen, L. H., Kelleher, N. L., Little, D. P., Kenyon, G. L. & McLafferty, F. W. (1995) *Biochemistry,* in press.
15. Perryman, M. B., Kerner, S. A., Bohlmeyer, T. J. & Roberts, R. (1986) *Biochem. Biophys. Res. Commun.* **140,** 981–989.
16. Trask, R. V., Strauss, A. W. & Billadello, J. J. (1988) *J. Biol. Chem.* **263,** 17142–17149.
17. Mariman, E. C. M., Broers, C. A. M., Claesen, C. A. A., Tesser, G. I. & Wieringa, B. (1987) *Genomics* **1,** 126–137.

18. Villarreal-Levy, G., Ma, T. S., Kerner, S. A., Roberts, R. & Perryman, M. B. (1987) *Biochem. Biophys. Res. Commun.* **144,** 1116–1127.
19. Kaye, F. J., McBride, O. W., Battey, J. F., Gazdar, A. F. & Sausville, E. A. (1987) *J. Clin. Invest.* **79,** 1412–1420.
20. Valaskovic, G. A., Kelleher, N. L., Little, D. P., Aaserud, D. J. & McLafferty, F. W. (1995) *Anal. Chem.* **67,** 3802–3805.
21. Chen, L. H., Babbitt, P. C., Vasquez, J. R., West, B. L. & Kenyon, G. L. (1991) *J. Biol. Chem.* **266,** 12053–12057.
22. Beu, S. C., Senko, M. W., Quinn, J. P., Wampler, F. M., III, & McLafferty, F. W. (1993) *J. Am. Soc. Mass Spectrom.* **4,** 557–565.
23. Loo, J. A., Edmonds, C. G. & Smith, R. D. (1991) *Anal. Chem.* **63,** 2488–2499.
24. Little, D. P., Speir, J. P., Senko, M. W., O'Connor, P. B. & McLafferty, F. W. (1994) *Anal. Chem.* **66,** 2809–2815.
25. Gauthier, J. W., Trautman, T. R. & Jacobson, D. B. (1991) *Anal. Chim. Acta* **246,** 211–225.
26. Senko, M. W., Speir, J. P. & McLafferty, F. W. (1994) *Anal. Chem.* **66,** 2801–2808.
27. Marshall, A. G., Wang, T.-C. L. & Ricca, T. L. (1985) *J. Am. Chem. Soc.* **107,** 7893–7897.
28. Schimmel, P. (1989) *Acc. Chem. Res.* **22,** 232–233.
29. Putney, S., Herlihy, W., Royal, N., Pang, H., Aposhian, H. V., Pickering, L., Belagaje, R., Biemann, K., Page, D., Kuby, S. & Schimmel, P. (1984) *J. Biol. Chem.* **259,** 14317–14320.
30. Ordahl, C. P., Evans, G. L., Cooper, T. A., Kunz, G. & Perriard, J.-C. (1984) *J. Mol. Biol* **259,** 15224–15227.
31. Giraudat, J., Devillers-Thierry, A., Perriard, J.-C. & Changeux, J.-P. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 7313–7317.
32. Buskin, J. N., Jaynes, J. B., Chamberlain, J. S. & Hauschka, S. D. (1985) *J. Mol. Evol.* **22,** 334–341.
33. Roman, D., Billadello, J., Gordon, J., Grace, A., Sobel, B. & Strauss, A. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 8394–8398.
34. Pickering, L., Pang, H., Biemann, K., Munro, H. & Schimmel, P. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 2310–2314.
35. Benfield, P. A., Henderson, L. & Pearson, M. L. (1985) *Gene* **39,** 263–267.
36. Hossle, J. P., Rosenberg, U. B., Schafer, B., Eppenberger, H. M. & Perriard, J.-C. (1986) *Nucleic Acids Res.* **14,** 1449–1463.