Chapter 14
Analysis of Fluorinated Proteins by Mass Spectrometry

Linda A. Luck

Abstract ^{19}F NMR has been used as a probe for investigating bioorganic and biological systems for three decades. Recent reviews have touted this nucleus for its unique characteristics that allow probing in vivo biological systems without endogenous signals. ^{19}F nucleus is exceptionally sensitive to molecular and microenvironmental changes and thus can be exploited to explore structure, dynamics, and changes in a protein or molecule in the cellular environment. We show how mass spectrometry can be used to assess and characterize the incorporation of fluorine into proteins. This methodology can be applied to a number of systems where ^{19}F NMR is used.

14.1 Background

^{19}F NMR has become a useful tool in the study of protein structure and dynamics for the last three decades [1, 2]. Although high-resolution X-ray and multidimensional NMR analysis provide the best structural information, all proteins are not amenable to these techniques. ^{19}F NMR provides data for these proteins and compliments the 3D structural data by monitoring specific locations within a known structure. Fluorine labels can be incorporated into proteins biosynthetically or chemically into specific sites. New frontiers in the area of ^{19}F NMR have been progressing in the area of pharmaceuticals to reveal drug uptake, biodistribution, and metabolism in vivo [3]. The fact that there are not endogenous biological fluorines allows good detection of the probes. The nucleus also has much promise for application in molecular biology in preclinical trial in animals that can be ultimately applied to humans.

L.A. Luck, Ph.D. (✉)
Department of Chemistry, State University of New York
at Plattsburgh, 101 Broad Street, Plattsburgh, NY 12901, USA
e-mail: luckla@plattsburgh.edu

A.G. Woods and C.C. Darie (eds.), *Advancements of Mass Spectrometry in Biomedical Research*, Advances in Experimental Medicine and Biology 806,
DOI 10.1007/978-3-319-06068-2_14,

^{19}F nucleus has a spin of ½ and occurs at 100 % natural abundance which gives nearly the same sensitivity of the ^{1}H nucleus. The most notable of its features is the fact that there is a lack of fluorine in biological samples, thus there are no background signals to contend with as there are in ^{1}H NMR samples which may be nearly 90 % water. The spectra are generally simple in terms of acquisition times and usually require lower protein concentrations than those needed for high resolution multidimensional NMR studies. An additional bonus for the ^{19}F NMR studies is that the chemical shift range for the ^{19}F nucleus is 30 times larger than ^{1}H, which also adds to the simplicity of the spectra since most resonances are not overlapped. An aromatic fluorine nucleus such as one replacing a ^{1}H on a tryptophan, phenylalanine, or tyrosine residue is particularly susceptible to electrostatic effects and shows the highest dispersion of the ^{19}F NMR shifts. In addition the ^{19}F nucleus has lone pair electrons which allow exquisite sensitivity to local microenvironment. Paramagnetic probes have large affects on the fluorine nucleus and have been used for solvent accessibility studies [4–6]. These studies provide insight not obtained by the other structural techniques.

The ^{19}F nucleus is small and can be exchanged for a proton or a hydroxyl on an amino acid. A host of fluorinated amino acids (e.g., F-tryptophan, F-tyrosine, difluoromethionine, trifluoroleucine, pentafluoroleucine, F-proline, and F-serine) are commercially available. Fluorinated amino acids are usually biosynthetically incorporated into proteins by the expression of the cloned gene with the ^{19}F-labeled amino acid in the media. In addition, new methodology has surfaced in the past year where fluoroindole can be utilized in the media for biosynthesis of fluorotryptophans [7]. In this case and in other biosynthetic methods the extent of incorporation is not 100 % because unlabeled amino acids are in the media or are required during growth of the bacteria. Although the percent of incorporation of fluorine may be less than 100, all sites incorporate equally well, thus the NMR spectra are useful [5, 8]. A number of methodologies have been used to increase the percent incorporation of the ^{19}F label in the protein. These include using an auxotrophic bacterial strain that limits the synthesis of amino acid endogenously, adding label after induction of protein synthesis, and addition of glyphosate which inhibits aromatic protein synthesis. Knowing the percentage of fluorine labeling is crucial to many experiments (e.g., 2D ^{19}F Heteronuclear and Homonuclear NMR) and the effects of ^{19}F labeling on protein structure and activity must be studied in order to draw conclusions regarding the native structure when using ^{19}F NMR studies.

If a protein is 100 % labeled with fluorine then all sites contain the specific fluorinated amino acid. If the fluorine labeling is less than 100 % the proteins are a heterogeneous mixture. If for example the protein is being labeled with F-serine and there are six serine residues; one protein may have six sites labeled with fluorine and another protein may have only one site labeled. Assessment of the detrimental effects of the fluorine labeling may be misleading.

The immediate consequences for a switch of a proton atom to the more electronegative fluorine atom in proteins are the C–F bond is polarized in the opposite direction to the C–H bond and the C–F bond is more stable. In addition the C–F bond is less polarizable than the C–H bond. In an aliphatic chain the C–F bond may be a weak hydrogen bond acceptor but in a ring structure the C–F is less likely to have

this characteristic. The switch from an H to F is considered isosteric with F slightly larger than H. Fluorine substitutions have been made to increase the hydrophobicity of pharmaceuticals and improve their bioavailability [9]. Thus changes with the fluorine substitution have some bearing on the immediate environment of the ^{19}F nucleus and possibly the structure and dynamics of the protein. Each protein has its own unique character and the different nature of the fluorine label employed in the study can result in perturbations or fluorinated proteins without large deleterious effects.

There have been a number of ways to analyze the incorporation of labeled amino acids into proteins. In the case of tryptophan analogues spectroscopic methods based on the absorption spectrum and chromatographic separations can give insight into the percentage of analogue incorporation [10]. Another means to quantitate the extent of fluorine incorporation into a protein is to compare integrals of a well-resolved peak in the ^{19}F NMR spectrum to the integral of an internal standard [8]. This methodology can be used with the NMR standard in the sample or in an insert in the NMR tube during the experiment. We show a study where the internal standard actually bound to the protein of interest (vide supra). Thus this methodology is not always useful. Mass spectrum analysis has been used for analysis of fluorinated tryptophan-labeled proteins and presents itself as a method of choice for other fluorinated amino acid analogues not amenable to fluorescence studies. In light of the recent interest in designing highly fluorinated hydrophobic proteins with the specific goal of enhancing protein stability and other chemical properties, a variety of fluorinated analogues of leucine, valine, isoleucine, glycine, and other aliphatic amino acids have been utilized [11, 12]. Mass spectrum analysis hails as the best methodology by providing a general strategy for quantitative data for analogue incorporation for any of the amino acids. Substitution of fluorine for proton which is a change in mass of 18 Da is easily recognizable in the mass spectrum of proteins <70 kDa. The purpose of this chapter is to report how we have used mass spectral analysis to determine fluorine analogue incorporation into proteins and show how mass spectrometry was used to identify specific mutations in the proteins that identified peaks in the ^{19}F NMR spectrum.

14.2 Proteins

Two proteins will be discussed in this chapter to illustrate the utility of mass spectrometry as a means of analyzing fluorine incorporation into proteins and mutants produced to identify peaks in the ^{19}F NMR spectrum. The *Escherichia coli* D-galactose and D-glucose receptor (GGR) is an aqueous sugar-binding protein and the first component in the chemosensory and transport pathways for these sugars. The protein which is very stable, well characterized by structural studies with mutations easily made and tolerated by the protein structure [8]. Glucose measurements especially for blood sugar in diabetes are crucial, thus GGR has been extensively investigated as a platform for a variety of electrochemical biosensors for the detection of blood glucose levels [13–17]. This utilitarian protein has been used as our model for investigation of a number of fluorinated analogues including 4, 5, and 6 fluorotryptophan 2, 3 and 4 fluorophenylalinine, tetradeutero-5-fluorotryptophan and recently

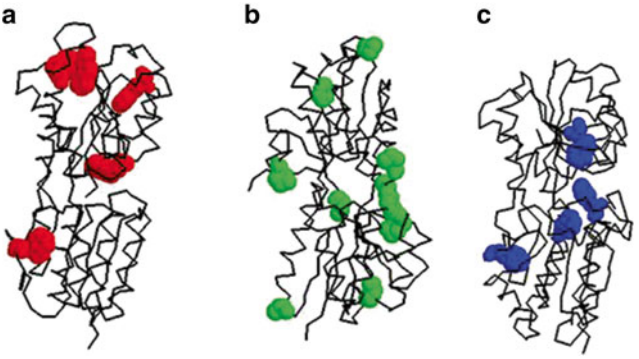

Fig. 14.1 Structure of the *E. coli* D-galactose and D-glucose receptor with the Trp residues highlighted (**a**) and the Pro residues highlighted (**b**). Structure of the *E. coli* L-leucine-specific receptor with the Trp residues highlighted (**c**) [23, 25]

H-*cis*-4-fluoro-proline (4F-*cis*-Pro) and H-*trans*-4-fluoro-proline (4F-*trans*-Pro) [18]. Figure 14.1a identifies the positions of the five tryptophan residues in GGR. Figure 14.1b illustrates the position of the ten proline residue sites in GGR.

The *Escherichia coli* leucine-specific protein (LBP) is a component of the transport pathway for hydrophobic amino acids. This protein as with the GGR is stable and easily manipulated for a number of biophysical studies. The LBP protein has been by ^{19}F NMR and fluorescence along with its counterpart the L-leucine, isoleucine, valine protein (LIV) [19–21]. LBP and LIV are models for the ionotropic glutamate receptors that control a wide variety of normal neuronal processes including learning and memory. Activation of these neurotransmitter receptors is involved in a number of neurodegenerative diseases, notably stroke and epilepsy. Study of the model proteins will allow us to understand the basic structural changes that contributing to the pathological effects of stroke and epilepsy, and shed light on some factors that will be beneficial in increasing cognition in our aging population. Thus we are pursuing studies on these proteins as platforms for biosensors to detect a variety of substances that may be toxic to our neurosystem [22]. Our laboratory was the first to report the affinity of LBP for phenylalanine and solved the closed structure by X-ray with phenylalanine in the binding site [23, 24]. The discovery of the phenylalanine affinity for LBP was made when 4-F-phenylalanine (4F-Phe) was used as an internal standard in the ^{19}F NMR spectrum to determine incorporation of 5-F-tryptophan (5F-Trp) in LBP (vide supra). Figure 14.1c illustrates the four positions of the tryptophan residues in closed form of LBP with leucine in the binding pocket.

14.3 Production of the Fluorine-Labeled Proteins

The production of the 4F-Trp-labeled GGR was accomplished in the strain *E. coli* NM303, a non auxotroph strain with 50 μg/mL 4F-Trp. The production of 4F-c*is*-Pro-labeled GGR utilized the compound H-*cis*-4-fluoro-proline, purchased from

BACHEM. Fifty micrograms per milliliter 4F-*cis*-Pro was used in the media and isolation of both of the fluorinated proteins was performed in the same manner as the wild-type protein [8]. The techniques and materials to generate the mutant LBP proteins and 5F-Trp-labeled LBP mutants have been described previously [19, 20, 23, 24].

14.4 NMR Measurements

NMR sample contained 0.6 mL of 1.9 mM LBP in 10 mM phosphate pH 6.9 with 10 % D_2O as lock. Spectra were obtained on a 500 MHz NMR with at 25 °C. Processing used 25 Hz linebroadening.

14.5 Mass Spectrometry Measurements

Samples were analyzed using an AB Sciex 4000 QTrap (AB Sciex, Framingham, MA) hybrid triple quadrupole/linear ion trap liquid chromatography–mass spectrometer (LCMS). Positive electrospray ionization (ESI) was used as the ionization source. Source temperature was maintained at 300 °C. Sheath gas (GS1) flow was set at 30, auxiliary gas flow (GS2) at 40, curtain gas flow at 25, and the declustering potential was set to 80. The mass spectrometer was operated in single quadrupole mode, scanning from *m*/*z* 700–1400. AB Sciex Biotools Beyesian Protein Reconstruct software was used for spectral deconvolution.

Analytes were separated using a Shimadzu Prominence high performance liquid chromatography (HPLC) system (Shimadzu Scientific Instruments, Columbia, MD) across water to acetonitrile (ACN) gradient, using 0.1 % formic acid and 0.025 % trifluoroacetic acid as ion pairing reagents. At the beginning of each run, the mobile phase was held at 30 % ACN for 3 min, increased to 70 % over 20 min, increased to 95 % over 2 min, and held at 95 % for 5 min. The LC column was then requilibrated to the initial conditions for 15 min. Flow was maintained at 80 μL/min. One microliter of each sample was injected onto a Phenomenex (Torrance, CA) Jupiter C4 300A reversed phase HPLC column (150 mm × 1 mm × 5 μm).

14.6 Mass Spectrometry of 4-Fluorotryptophan and 4F-*cis*-Pro-Labeled GGR

The ^{19}F NMR studies of 5F-Trp-labeled GGR showed great utility for the conformational changes that occurred during ligand binding [4, 8, 26]. The protein is easily manipulated and we have further explored the incorporation of a number of fluorine and carbon-13 labels. One question that plagued us during these studies was the multiplicity of the ^{19}F NMR peaks near the calcium binding site [26]. To address

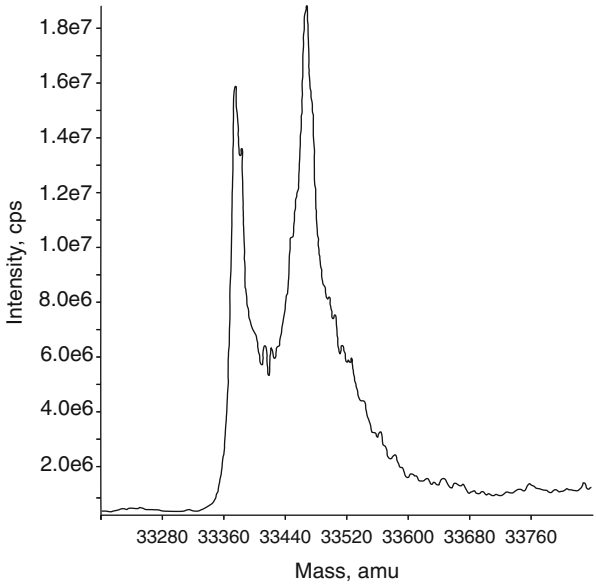

Fig. 14.2 The mass spectrometry of the 4F-Trp-labeled GGR

this question we have started experiments to determine whether these peaks are due to the 5F-Trp analogue or the neighboring proline residues that may show *cis* or *trans* orientation. To accomplish this we are exploring these questions by ^{19}F NMR studies and X-ray crystallography of the fluorinated proteins. We have commenced our investigation by incorporating 4F-Trp, 4F-*cis*-Pro, and 4F-*trans*-Pro into GGR. The 4F-Trp analogue is essentially nonfluorescent at ambient temperature; making it a "silent" analogue for absorption spectroscopy, therefore fluorine incorporation studies using an absorption technique referred to as LINCS analysis was not possible [27]. Figure 14.2 shows the initial mass spectrometry data for the incorporation of 4F-Trp into GGR in a non auxotrophic strain of *E. coli*. The spectrum shows a peak for native GGR at a mass of 33,370 and a broad peak for the incorporation of five fluorine atoms at 33,460. The percent incorporation is over 50 % which is sufficient for ^{19}F NMR studies but not for a number of physical biochemical methods such as differential scanning calorimetry for protein integrity or titrating calorimetry for determining the binding of sugar and calcium ligands to the fluorinated protein. Our efforts continue to increase the incorporation to nearly 100 % using an auxotroph strain of *E. coli*.

The 4F-*cis*-Pro incorporation into the ten sites on the GGR protein was the most complex project we have attempted to date. The mass spectral analysis of our third attempt to incorporate this fluorine label is shown in Fig. 14.3. The spectrum shows a distinct resonance for native GGR at 33,373 mass units. Since it is not fully labeled GGR shows a heterogeneous population of masses which corresponds to fluorine in any of the ten proline sites. Thus the observed masses for the fluorinated GGR at 33,388, 33,406, 33,424, 33,442, 33,460, 33,478, 33,496, 33,514, 33,532 and 33,550

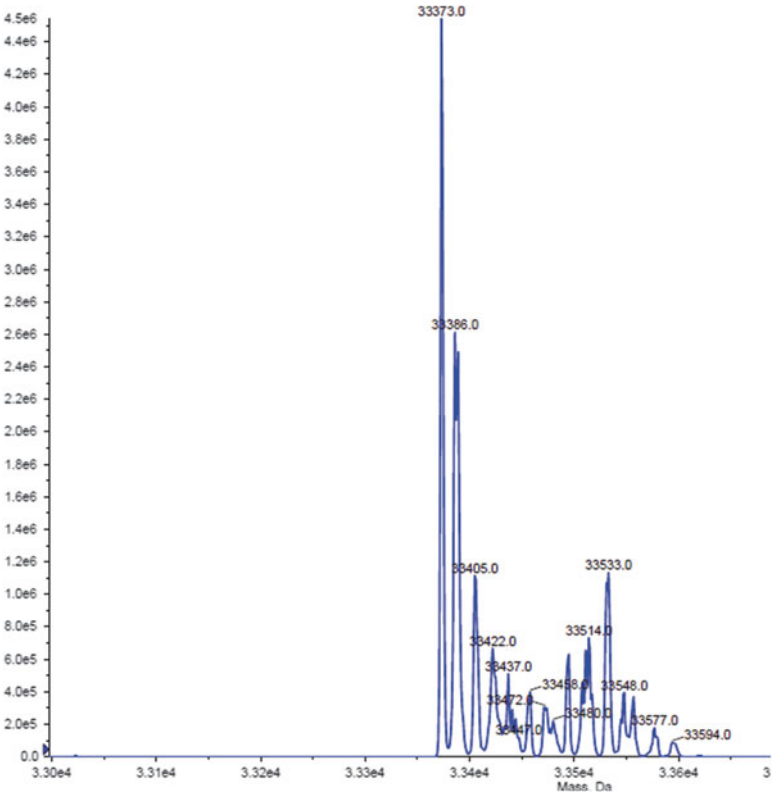

Fig. 14.3 The mass spectrum of 4F-*cis*-Proline-labeled GGR. The largest peak is the native GGR with no label

correspond to 1–10 fluorine atoms incorporated per protein. The peaks for these masses are detected in the spectrum, therefore we can conclude that the 4F-*cis*-Pro is tolerated by the biosynthetic system even though there have been reports where proteins are compromised with this analogue [28]. We will continue our work by transforming our plasmid into the proline auxotroph cell line JM83 with increased amounts of 4F-*cis*-Pro in the media to enhance the incorporation percentage of this fluorine analogue.

14.7 Mass Spectrometry of 5-Fluorotryptophan LBP and LBP Mutants

Initial studies of the 5F-Trp-labeled LBP protein used an internal standard of 4F-Phe as an internal standard for the ^{19}F NMR spectrum and for calculation of the percent incorporation of fluorine into the protein as was done for the GGR [4]. Figure 14.4

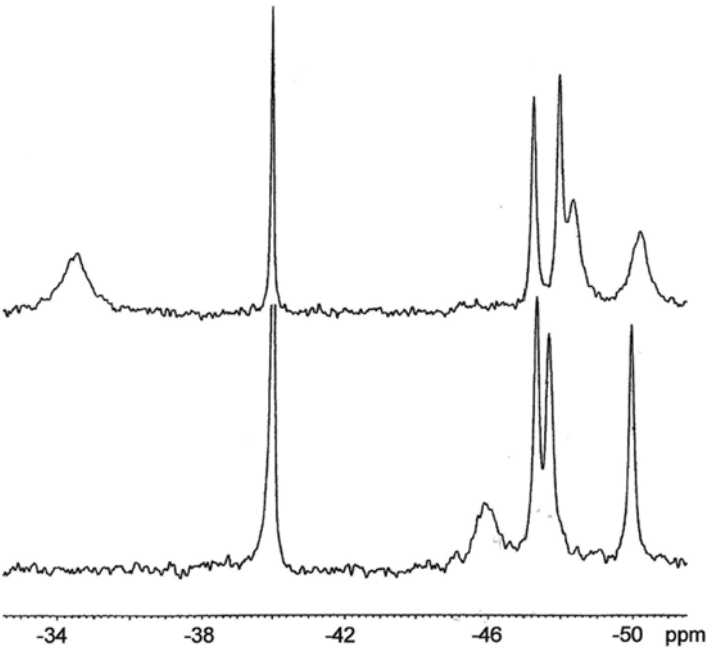

Fig. 14.4 The ^{19}F NMR spectra of 5F-Trp-labeled LBP with 4F-Phe added as an internal standard. Top spectrum is without ligand L-leucine and the bottom spectrum is with excess L-leucine. The internal standard resonance is present in the top spectrum because LBP does not bind to the D form of 4F-Phe. The standard 4F-Phe was a racemic mixture of both D and L forms

shows the ^{19}F NMR spectra of the LBP protein without L-leucine (top) and with excess L-leucine (bottom). The resonance at −39 ppm is the internal standard 4F-Phe. Contrary to the studies with ^{19}F NMR GGR which used the 3F-Phe as an internal standard without compromising the experiments, the internal standard 4F-Phe bounds to the LBP. This consequently led to further studies identifying the new ligands for LBP. The resonance to the far left of the top spectrum in Fig. 14.4 corresponds to the L-4F-Phe in the bound state in the LBP protein. This broad resonance disappears when excess L-leucine is added to the NMR tube. The ^{19}F NMR of the LBP without 4F-Phe in the sample shows four distinct sharp resonances with L-leucine [19]. The spectrum in Fig. 14.4 (bottom trace) shows a protein in slow exchange between two ligands in the NMR tube. This is a good example why an internal standard should not be used directly in a sample and why mass spectrometry is a better choice for calculating the percent incorporation of a fluorine label.

LBP has an inducible promoter and thus with minimal media and enhanced amounts of fluoro amino acids added prior to induction the percentage of fluorinated label is greatly increased. Figure 14.5 shows the mass spectrum of 5F-Trp-labeled LBP with 100 % incorporation. Native LBP has a 36,983 mass and 5F-Trp-labeled protein shows a mass of 37,054.

To identify the resonances in the NMR spectrum we changed the tryptophan residues to phenylalanine or tyrosine. Mass spectrometry can also be used to quickly

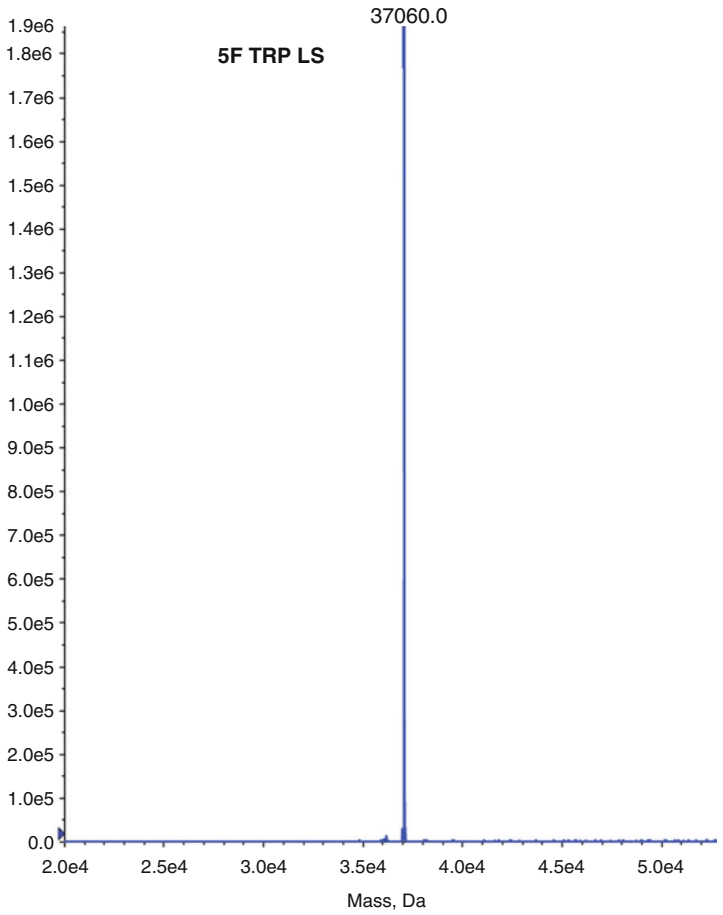

Fig. 14.5 The mass spectrum of 5F-Trp-labeled LBP

Table 14.1 The calculated masses and masses obtained by mass spectrometry of LBP proteins and mutants

Protein	Calculated weight	Mass spec weight
LBP WILD TYPE	36,983	36,986
5F-Trp LBP	37,054	37,060
LBP W336F	36,944	36,943
LBP W18, 278, 320F	36,869	36,868
LBP W18F	36,944	36,948
LBP W18Y	36,960	36,966
LBP W278F	36,944	36,942
LBP D1C	36,972	36,988

identify if the correct mutation is made. Table 14.1 shows the mass spectrometric analysis of a number of LBP mutants that were used in our NMR, biosensor, and fluorescent studies.

14.8 Conclusions

The studies shown in the chapter demonstrates the applicability of mass spectrometry for studying the percent incorporation of fluorine labels in proteins prepared for ^{19}F NMR studies. For 4F-Trp incorporation this method is crucial since absorption methods are not amenable. We clearly make a case that this is the method of choice for most fluorine analogues. In addition the results from the mass spectrometry strongly support the site directed mutagenesis data and confirm the mutant protein production for analysis of the resonances in the ^{19}F NMR spectra. The effectiveness of the ^{19}F NMR data is greatly facilitated by careful quantification and protein characterization by liquid chromatography electrospray mass spectrometry.

Acknowledgments The author would like to thank Bruce O'Roarke for analyzing the proteins by mass spectrometry. This work has been supported by the NSF for the LCMS instrumentation in the Chemistry Department at the University of Vermont (CHE MRI-0821501, LAL Co-PI), SUNY at Plattsburgh President's Award and Mini Grant.

References

1. Danielson MA, Falke JJ (1996) Annu Rev Biophys Biomol Struct 25:163–195
2. Yu JX, Hallac RR, Chiguru S, Mason RP (2013) Prog Nucl Magn Reson Spectrosc 70:25–49
3. Reid DG, Murphy PS (2008) Drug Discov Today 13:473–480
4. Luck LA, Falke JJ (1991) Biochemistry 30:6484–6490
5. Zemsky J, Rusinova E, Nemerson Y, Luck LA, Ross JB (1999) Proteins 37:709–716
6. Crabb JW, Carlson A, Chen Y, Goldflam S, Intres R et al (1998) Protein Sci 7:746–757
7. Crowley PB, Kyne C, Monteith WB (2012) Chem Commun (Camb) 48:10681–10683
8. Luck LA, Falke JJ (1991) Biochemistry 30:4248–4256
9. Yu J, Cui D, Zhao D, Mason RP (2008). In: GH A. Tressaud (ed) Fluorine and health. Elsevier B.V., Amsterdam, pp 198–276
10. Senear DF, Mendelson RA, Stone DB, Luck LA, Rusinova E, Ross JB (2002) Anal Biochem 300:77–86
11. Buer BC, Marsh EN (2012) Protein Sci 21:453–462
12. Salwiczek M, Nyakatura EK, Gerling UI, Ye S, Koksch B (2012) Chem Soc Rev 41:2135–2171
13. Luck LA, Moravan MJ, Garland JE, Salopek-Sondi B, Roy D (2003) Biosens Bioelectron 19:249–259
14. Carmon KS, Baltus RE, Luck LA (2004) Biochemistry 43:14249–14256
15. Andreescu S, Luck LA (2008) Anal Biochem 375:282–290
16. Tripathi A, Wang J, Luck LA, Suni II (2007) Anal Chem 79:1266–1270
17. Sokolov I, Subba-Rao V, Luck LA (2006) Biophys J 90:1055–1063
18. Luck LA, Vance JE, O'Connell TM, London RE (1996) J Biomol NMR 7:261–272
19. Salopek-Sondi B, Luck LA (2002) Protein Eng 15:855–859
20. Salopek-Sondi B, Skeels MC, Swartz D, Luck LA (2003) Proteins 53:273–281
21. Salopek-Sondi B, Swartz D, Adams PS, Luck LA (2002) J Biomol Struct Dyn 20:381–387
22. Ahmed AH, Loh AP, Jane DE, Oswald RE (2007) J Biol Chem 282:12773–12784
23. Magnusson U, Salopek-Sondi B, Luck LA, Mowbray SL (2004) J Biol Chem 279:8747–8752

24. Luck LA, Johnson C (2000) Protein Sci 9:2573–2576
25. Vyas MN, Jacobson BL, Quiocho FA (1989) J Biol Chem 264:20817–20821
26. Luck LA, Falke JJ (1991) Biochemistry 30:4257–4261
27. Waxman E, Rusinova E, Hasselbacher CA, Schwartz GP, Laws WR, Ross JB (1993) Anal Biochem 210:425–428
28. Holzberger B, Obeid S, Welte W, Diederichs K, Marx A (2012) Chem Sci 3:2924–2931