Determination of Soybean Oil, Protein and Amino Acid Residues in Soybean Seeds by High Resolution Nuclear Magnetic Resonance (NMRS) and Near Infrared (NIRS).

I.C. Baianu[1,2], T. You[1,2], D.M. Costescu[1,3], P.R. Lozano[1,2], V. Prisecaru[1,2], and R.L. Nelson[1,4]

[1]FSHN Dept., [2]AFC-Micro-Spectroscopy Facility, [3]Nuclear, Plasma and Radiological Engineering Dept., and [4]National Soybean Laboratory, Crop Sciences Dept., [1]University of Illinois at Urbana, IL, 61801, USA.

Email address for correspondence: **i-baianu@uiuc.edu**

Abstract

A detailed account is here presented of our high resolution nuclear magnetic resonance (HR-NMR) and near infrared (NIR) calibration models, methodologies and validation procedures, together with a large number of composition analyses for soybean seeds. NIR calibrations were developed based on both HR-NMR and analytical chemistry reference data for oil and twelve amino acid residues in mature soybeans and soybean embryos. This is our first report of HR-NMR determinations of amino acid profiles of proteins from whole soybean seeds, *without protein extraction from the seed*. It was found that the best results for both oil and protein calibrations were obtained with a Partial Least Squares Regression (PLS-1) analysis of our extensive NIR spectral data, acquired with either a DA7000 Dual Diode Array (Si and InGaAs detectors) instrument or with several Fourier Transform NIR (FT-NIR) spectrometers equipped with an integrating sphere/InGaAs detector accessory. In order to extend the bulk soybean samples calibration models to the analysis of single soybean seeds, we have analized in detail the component NIR spectra of all major soybean constituents through spectral deconvolutions for bulk, single and powdered soybean seeds. Baseline variations and light scattering effects in the NIR spectra were corrected, respectively, by calculating the first-order derivatives of the spectra and the Multiplicative Scattering Correction (MSC). The single soybean seed NIR spectra are broadly similar to those of bulk whole soybeans, with the exception of minor peaks in single soybean NIR spectra in the region from 950 to 1,000 nm. Based on previous experience with bulk soybean NIR calibrations, the PLS-1 calibration model was selected for protein, oil and moisture calibrations that we developed for single soybean seed analysis. In order to improve the reliability and robustness of our calibrations with the PLS-1 model we employed standard samples with a wide range of soybean constituent compositions: from 34% to 55% for protein, from 11% to 22% for oil and from 2% to 16% for moisture. Such calibrations are characterized by low standard errors and high degrees of correlation for all major soybean constituents. Morever, we obtained highly resolved NIR chemical images for selected regions of mature soybean embryos that allow for the quantitation of oil and protein components. Recent developments in high-resolution FT-NIR microspectroscopy extend the NIR sensitivity range to the *picogram* level, with submicron spatial resolution in the component distribution throughout intact soybean seeds and embryos. Such developments are potentially important for biotechnology applications that require rapid and ultra-sensitive analyses, such as those concerned with high-content microarrays in Genomics and Proteomics research. Other important applications of FT-NIR microspectroscopy are envisaged in biomedical research aimed at cancer prevention, the early detection of tumors by NIR-fluorescence, and identification of single cancer cells, or single virus particles *in vivo* by super-resolution microscopy/ microspectroscopy.

KEYWORDS for Book Index:

Nuclear Magnetic Resonance (NMR) Principles and Analysis of soybeans and other oil seeds; Near-Infrared (NIR) Principles, Instrumentation, Calibration and Analysis of Oil Seeds; Nondestructive Composition analysis of Soybeans Seeds, Corn and Wheat Grains; High Resolution Nuclear Magnetic Resonance Determination of Soybean Oil, Fatty Acid Composition, Isoflavones, Protein and Amino Acid Residues; HR NMR; High Resolution NMR of Proteins and Amino Acids; NMR of Oil and Protein Determinations in Mature Soybeans; High Resolution Near-Infrared Determination of Soybean Oil, Protein and Amino Acid Residues; amino acid profiles; Partial Least Square Regression-1 (PLS-1), PCR, PCA, and MLR Algorithms; Fourier Transform NIR (FT-NIR); spectral deconvolutions; Multiplicative Scattering Correction (MSC); protein, oil and moisture calibrations for Single Soybean Seed analysis; NIR chemical images for selected regions of Mature Soybean Embryos; NIR-fluorescence; dual Diode Array instruments; bulk and single soybean seed composition analysis; high-resolution ^{13}C NMR; National Soybean Research Laboratory; Near Infrared Calibrations for Oil, Protein and Amino Acid Composition Determinations in Soybeans; Nuclear Magnetic Resonance Calibrations for Oil; NIR Spectroscopy and Instrumentation; moving diffraction grating NIR/IR instruments; interference filter-based NIR instruments; NIR Spectra Pre-processing; specular reflection; light scattering corrections; baseline shifts in NIR Spectra and first derivative corrections; Savitzky-Golay algorithm for NIR data preprocessing; Stepwise Multiple Linear Regression (SMLR) algorithm; TQ Analysis software; Principal Component Regression model (PCR); Partial Least Squares models (PLS) for NIR calibrations ; Standard Methods of Soybean Composition Analysis: Protein Analysis, Kjeldahl Method, Biuret and Lowry Methods; Folin-Ciocalteau Phenol Reagent; Modified Biuret-Lowry method; High Performance Liquid Chromatography (HPLC) Analysis of Derivatized Amino Acids from Hydrolyzed Proteins; Solvent extraction methods; Moisture Determination methods; Oven drying methods; Low-resolution NMR; Karl Fischer Titration of Water; plant polysaccharide; NIR Calibrations for Protein and Oil Measurements in Mature Soybean Seeds; Bulk and Single Seed Calibrations; High Resolution Nuclear Magnetic Resonance Spectroscopy Applications to Soybean Analysis; HR-NMR Analysis of Oil Content in Seeds; 1-Pulse (1PULSE) HR-NMR Sequence for Oil Determination in Soybean seeds; Solid State ^{13}C NMR Techniques for Analysis of Soy Flour samples; 1PDNA ^{13}C SS-NMR technique for Oil Content Determination in Soybean Flours; superconducting magnetic fields; FT-NMR spectrometers; One-Pulse Decoupler; ^{13}C SS-NMR; The VACP ^{13}C SS-NMR technique for Measurements of Protein Content in Soybean Flours; Liquid-State ^{13}C NMR Measurements of Protein Content and Amino Acid Residues; Highly Hydrated Soy Flour Gels; Protein Content and Amino Acid Profile Determinations without protein extraction from the oil seeds; WALTZ-16; ^{1}H Decoupling Sequence for ^{13}C Liquid-State NMR; WALTZ-16 decoupled ^{13}C Liquid-State NMR of a soy flour gels; Soy Protein Isolate gels; Standard Calibration Plot for Soy Protein Isolates; Amino Acid Contents of Soybeans Determined by ^{13}C Liquid-State High Resolution NMR; Ion Exchange Chromatography (IEC) of Amino Acids and Soybean protein hydrolizates; Picogram sensitivity, submicron resolution FT-NIR Microspectroscopy; *in vivo* Single Molecule Detection for Ultra-sensitive Analysis in Proteomics, Functional Genomics and Agricultural/Medical Biotechnology applications.

High Resolution Nuclear Magnetic Resonance and Near Infrared Determination of Soybean Oil, Protein and Amino Acid Residues in Soybean Seeds

TABLE OF CONTENTS

ABSTRACT……...1
KEYWORDS……2
1. INTRODUCTION……..……...3
2. PRINCIPLES……..……...5
2.1. Principles of FT-NIR Spectroscopy
2.2. Principles of Nuclear Magnetic Resonance Spectroscopy
3. EXPERIMENTAL……..……..8
3.1. FT-NIR Instrumentation
3.2. FT-NIR Spectra Pre-processing
3.3. FT-NIR Calibration Models
3.4. NMR Techniques for Oil Determination in Soybean
3.4.1. One Pulse High-Resolution NMR
3.4.2. Low Resolution NMR for Oil Determination in Seeds
3.4.3. 1PDNA ^{13}C SS-NMR for Oil Content Determination in Soybean Flours
3.4.4. VACP ^{13}C SS-NMR for Measurement of Protein in Soybean Flours
3.4.5. Liquid-State ^{13}C NMR Measurements of Protein Content and Amino Acid Residues in Hydrated Soy Flour Gels
Protein Content and Amino Acid Profile Determination with WALTZ-16, ^{1}H Decoupling Sequence for ^{13}C Liquid –State NMR of Highly Hydrated Soy Flour Gels and Doughs
3.5. Standards Methods for Soybean Composition Analysis
3.5.1. Oil Determination: Solvent Extraction methods
3.6. Protein Analysis
3.7. High Performance Liquid Chromatography Analysis of Derivatized Amino Acids from Hydrolyzed Proteins
3.8. Moisture determination methods
4. RESULTS……..… 25
4.1. Validation of the NIR Calibrations for Protein and Oil Measurements in Mature Soybean Seeds
4.2. Oil and Protein Determination in Mature Soybeans Using NMR Techniques
4.2.1. Decoupling Sequence for ^{13}C Liquid-State NMR of Highly Hydrated Soy Flour Gels and Doughs
4.2.2. Amino Acid Contents of Soybeans Determined by ^{13}C Liquid-State High-Resolution NMR
4.3. Oil Determination in Soybeans by One Pulse NMR Sequence
4.4. Oil Determination in Soybean Flour by 1 PDNA Pulse Sequence NMR
4.5. Oil Determination in Soybean Flour by VACP ^{13}C SS-NMR
5. LIMITATIONS AND ADVANTAGES OF THE TECHNIQUES…… 46
6. OTHER APPLICATIONS: Selection of Soybeans Standards for NIR Calibrations of Fatty Acid Composition ……...47
7. PRACTICAL IMPLICATIONS OF OUR HIGH-RESOLUTION NIR AND NMR ANALYSES OF SOYBEAN SEEDS……...……..50
8. CONCLUSIONS AND DISCUSSION……..…..51
9. REFERENCES……..52

1. Introduction

Soybeans are the major source of plant protein and oil in the world. Commercial soybean varieties usually contain ~40% protein and ~20% oil (on a dry weight % basis). Although there remains a strong economic incentive to develop cultivars with high protein and oil contents while maintaining a competitive yield, progress has been slow. Effective breeding techniques require accurate, inexpensive and reliable soybean composition analysis. Certain areas of breeding and selection research would also benefit from single soybean seed analysis (Silvela *et al.*, 1989). Conventional composition analysis methods such as the Kjeldahl method for protein measurement and the ether extraction method for oil fraction measurements are time-consuming, expensive and impractical for measurements on large numbers of soybean samples required for molecular genetic mapping and other selection and breeding studies. In addition to problems such as low speed and high cost, wet-chemistry methods are destructive and rather inaccurate for single seed analysis, with the notable exception of the extracted protein determination by the Lowry method (1958).

Emerging practical solutions to these problems are based on Near Infrared Reflectance Spectroscopy (NIRS). When adequately calibrated with reliable primary data, NIRS generates accurate results and is less expensive than conventional or wet chemistry, composition measurement methods such as those currently adopted by the American Oil Chemists' Society (AOCS). A wide range of grains and oil seeds has been analyzed by NIRS techniques with varying degrees of success. For soybeans, early reports showed that dispersive/filter-based Near Infrared (NIR) instruments can be utilized for the determination of protein, oil (Williams, 1975) and moisture (Ben-Gera and Norris, 1968). However, in recent years significant improvements in NIR instrument performance were achieved through novel designs. A recent improvement in the design of dispersive instruments allows for high spectral acquisition speeds through the utilization of Dual Diode Array NIR detectors, such as those commercially available from Perten Instruments, Inc., (Springfield, IL). The DA-7000 NIR spectrometer model (made by Perten Instruments, Inc.) employs a dual Diode (Si/InGaAs) Array Detector, as well as a *stationary* diffraction grating, and is capable of spectral collection speeds up to 600 spectra per second (Shadow, 1998) in the range from 400 to 1,700 nm. Besides the recent development of Diode Array techniques for dispersive instruments, Fourier Transform (FT) technology is currently employed in NIR instruments to overcome most of the disadvantages of classical dispersive NIR instruments that employ moving gratings and have low acquisition speed and limited NIR resolution. Commercial FT-NIR instruments are available from manufacturers such as Thermo Nicolet, Inc. (Madison,WI), PerkinElmer Co. (Shelton,CT) and Bruker, Inc. (Madison,WI). The major advantages of FT-NIR and dual Diode Array instruments over moving grating dispersive instruments are their higher spectral resolution, higher and uniform wavelength accuracy, and also high speed of spectral acquisition/data collection. High spectral resolution is important because it facilitates long-term calibration robustness and improved separation of the sample constituents; it may also reduce the total number of samples required for calibration development because of the higher spectral information content in comparison with the other NIR instrument designs. High wavelength accuracy is critical when a calibration developed on a specific NIR instrument needs to be transferred to another instrument, and when separation of minor component constituents is sought for. Wavelength accuracy is also important for signal

averaging, which is essential for samples with low signal-to-noise ratio (S/N), as is the case of single seeds.

Although most NIRS applications are currently focused on bulk sample analysis, some recent studies on transmission instruments attempted preliminary estimates of single seed composition, such as the moisture measurement of single soybean seeds with a Shimadzu W-160 dual-beam spectrometer (Lamb and Hurburgh, 1991), and the oil measurement of single corn kernels with an Infratec model 1255 spectrometer (Orman and Schumann, 1992). These preliminary reports have indicated the potential of NIRS for single seed analysis. In addition to transmission instruments, NIR reflectance instruments have also been applied recently to single seed analysis, such as an attempt to generate color classifications (Wang et al, 1999) and an attempt to perform computational averaging of single wheat kernel spectra for composition analysis (Shadow et al, 2000). Although some progress with single seed analysis by NIR has already been reported, the potential advantages of novel NIR instrument designs such as the dual Diode Array and FT techniques have not yet been fully exploited. To take advantage of novel instrument designs, both a dual Diode Array instrument (DA-7000 by Perten Instruments, Inc., Springfield, IL) and an FT instrument (Spectrum One NTS, manufactured by PerkinElmer, Shelton, CT) were calibrated for both bulk and single soybean seed composition analysis. In recent studies we developed accurate, reliable and robust NIR calibrations for both bulk and single seed composition analyses that facilitate novel breeding/selection techniques and improve breeding efficiency.

On the other hand, previous NIRS attempts at calibrations for amino acid residues of soybean proteins in bulk soybean seeds and powdered soybean seeds have suffered until recently from two major drawbacks: the employment of primary methods involving extensive extraction and acid hydrolysis of soybean proteins from soybean seeds, and the low spectral resolution of the NIR spectra of soybean proteins and their amino acid residues. A radically different approach that circumvents such problems is afforded by high-resolution carbon-13 (^{13}C) NMR quantitative analysis of soybean protein peaks corresponding to specific ^{13}C sites of selected amino acid residues of unhydrolyzed and unmodified soybean proteins in either powdered or intact soybean seeds. Both the advantages and limitations of our novel approach to amino acid profiling and protein composition analysis of soybean seeds will be discussed, and the possible extension of this approach to developing NIRS calibrations based on the high-resolution NMR primary data will be briefly outlined. A comparison will also be presented between the results obtained with our novel NMR approach for amino acid profiles of soybean seed proteins and the corresponding data obtained through soybean protein extraction, derivatization and acid hydrolysis, followed by ion exchange chromatography and high-performance liquid chromatography (HPLC).

An attempt will be made here to present a concise overview of our recent NIR and NMR methodologies and composition measurements for a wide range of selected soybean accessions, including over 2,000 exotic soybean germplasm accessions from the USDA Soybean Germplasm Collection at the National Soybean Research Laboratory at UIUC (http:// www.nsrl.uiuc.edu).

2. Principles of Spectroscopic Quantitative Analyses

In order to achieve a successful quantitative composition analysis by spectroscopic techniques one requires a clear understanding of the underlying spectroscopic principles. A purely statistical approach --without such a basic understanding-- is more likely to result in spurious numerical data sets, that do not correspond to physical reality.

2.1. Principles of NIR Spectroscopy

IR/NIR absorption spectra occur because chemical bonds within molecules can vibrate and rotate thus generating series of different energy levels among which rapid, IR (or NIR)-induced transitions can occur. According to standard Quantum Mechanics, the vibro-rotational energy levels of a molecule can be approximately calculated with the following equations:

$$E_{NIR} = E_{rot} + E_{vib} + E_{anh} = j(j+1)Bhc + [1 - x(n+1/2)]h\nu \qquad (2.1.1)$$

where:

j: rotation quantum number: 0,1,2,3..;

n: vibration quantum number: 0,1,2,3..;

E: energy eigenvalues, and

x: anharmonic constant ≈ 0.01

The mid- and far- IR induced transitions occur mainly between neighboring energy levels ($\Delta n = \pm 1$). Such transitions are normally referred to as fundamental transitions. Absorptions caused by fundamental transitions of most molecules occur in the mid- and far- IR range of wavelengths (> 2500 nm). In addition to the fundamental transitions, molecules can also be excited from the zero energy level to energy levels beyond the first energy level ($\Delta n = \pm 2, \pm 3..$) with lower probabilities, following Boltzmann statistics. Such transitions are referred to as overtones. Absorptions caused by overtones of chemical bonds with low reduced mass (such as the O-H, N-H or C-H bond) take place in the NIR region (typical wavelengths are between 700 and 2500 nm). Therefore, the resulting NIR spectra of liquids or solids appear fairly broad and have quite low resolution by comparison with mid-IR spectra, but have higher band separation than visible absorption, or fluorescence spectra that correspond to electronic transitions in molecules. In addition to overtones, NIR transitions corresponding to (or localized at) different chemical bonds can couple and produce a combination band of such fundamental transitions. NIR absorption corresponding to combination bands of specific chemical bonds with low reduced mass (such as, O-H, N-H and C-H) also take place in the NIR region (Raghavachari, 2001; Barton, 2002). When the sample to be measured is exposed to a beam of NIR light, the beam interacts with the sample in a variety of modes, such as absorption, reflection, transmission, scattering, refraction and diffraction. From an analytical standpoint, the light

absorption is the important process, as it is directly related to constituent concentrations, as described by the **Lambert-Beer's Law**:

$$\mathbf{A = \varepsilon * L * C} \qquad \textbf{(2.1.2)}$$

where: **A** is the 'True' Absorbance, **ε** is the Extinction coefficient of the Analyte that absorbs, **L** is Path length of light through the analyzed sample, and **C** is the Analyte Concentration.

The 'true' absorbance of a sample, however, is often quite difficult to directly measure- without applying first appropriate corrections for the other light interactions that occur within the sample, especially in inhomogeneous solid or turbid liquid samples. In practice, the absorption is often calculated indirectly from the measurement of the reflectance (R), (as **A = Log 1/R**) because reflectance can be readily measured even for thick samples; the exceptions are those complex samples that possess a composite structure, such as thick, multiple layers of different composition. The calculated absorbance is usually referred to as the 'apparent absorbance,' and it can be significantly affected by specular reflection and light scattering even in the case of thin samples. Because of light scattering and specular reflection, spectral pre-processing and corrections are always required in order to obtain reliable NIR quantitative determinations of composition for samples as complex as whole seeds or intact embryos.

2.2. Principles of Nuclear Magnetic Resonance Spectroscopy

High Resolution Nuclear Magnetic Resonance (HR-NMR) spectroscopy is a powerful tool for both qualitative and quantitative analysis of foods and biological systems (see Baianu and Kumosinski, 1993 for an in-depth review of such recent applications). NMR is based upon the resonant absorption of radio-frequency (rf) waves/quanta by the nuclear spins present in a macroscopic sample when the latter is placed in a strong and uniform/constant magnetic field, $\mathbf{H_o}$. The magnetic moments **μ** of the nuclei present in the sample interact with such a strong, external magnetic field, and the magnetic interaction energy is simply:

$$E_M = -\,\boldsymbol{\mu} \cdot \mathbf{H_0} \qquad \textbf{(Eq. 2.2.1)}$$

The magnetic moments of the nuclei were shown to be able to take only certain discrete values- they are quantized- and are proportional to the total angular moments, **J :**

$$\boldsymbol{\mu} = \gamma \mathbf{J}, \text{ with } \mathbf{J} = (h/2\pi)\mathbf{I}, \qquad \textbf{(Eqs.2.2.2)}$$

where γ is the *giromagnetic ratio* characteristic of each type of nucleus, and **I** is a dimensionless angular momentum operator whose eigenvalues are called "spin number", or simply '*spin*'— an intrinsic quantum mechanical property of a nucleus that is observed only when there is an

external magnetic field present . The **I**-operator component along the NMR probe coil axis, x, is $\mathbf{I_x}$ and has m allowed values are called its eigenvalues, or spin values; such allowed m values have the form I, (I-1),…0,…(-I). Therefore, the nuclear spin energy levels derived from Eqs. 2.2.1 and 2.2.2 are:

$$E_m = -m\,\gamma(h/2\pi)\mathbf{H_0} \qquad \textbf{(Eq.2.2.3)}$$

, or in frequency (ν) units:

$$h\nu = \gamma(h/2\pi)\,H_0 \qquad \textbf{(Eq.2.2.4)}$$

where m = I, (I-1), … (-I). Allowed NMR transitions induced by resonant rf irradiation in the presence of a constant external magnetic field $\mathbf{H_0}$ will occur only for**:**

$$\mathbf{\Delta m = \pm 1.} \qquad \textbf{(Eq.2.2.5)}$$

The external magnetic field $\mathbf{H_0}$ *polarizes* the nuclear spins so that at thermal equilibrium there is an excess of nuclear magnetic moments precessing, or rotating at a constant rate, around the direction of the external magnetic field. The net result is a small, macroscopic magnetization of the sample that precesses around the magnetic field direction, z . A resonant rf pulse will tilt this precession axis and will also induce transitions between the energy levels that satisfy eq.2.2.4 (i.e., single quantum transitions). Such transitions can be observed as NMR absorption peaks in the corresponding NMR spectrum. The pulsed NMR signal- which is acquired in the time domain- has been called the *Free Induction Decay* (FID) because it is the result of a voltage induced by the nuclear spin magnetization of the sample in the coil of the NMR probe as a result of the fact that the precessing magnetization produces a variable magnetic flux through the NMR probe coil which alternates in phase with the precessing magnetization (Bloch,1956). The FID signal decays with time as the nuclear spins loose phase coherence during their precession around the external magnetic field axis (along the z-direction). The FID is then digitized at a series of points in time arranged at regular, small intervals, and it is stored in digital form in dedicated computer memory. Increasing the number of digitization points proportionally increases the spectral resolution of the NMR absorption spectrum when the computer transforms the digitized FID signal by Fast Fourier Transformation (FFT).

Because the various types of chemical bonds or chemical groups present in a material sample correspond to different electron density distributions surrounding the nuclear spins of the atoms involved, such nuclear spins experience different degrees of shielding from the external magnetic field, which is caused by the specific electron densities involved in chemical bonds or groups. As a result, the nuclear spins from distinct chemical groups resonate at different radio frequencies corresponding to the different degrees of shielding of such nuclear spins from the external magnetic field by the surrounding electron orbitals. Therefore, a number of such distinct nuclear magnetic resonance absorption peaks is observed which differ through their specific resonance frequencies by a value defined as the 'Chemical Shift'- proportional to the amount of electron orbital shielding surrounding each nuclear spin present. Various chemical groups will

thus exhibit a number of characteristic resonance peaks with chemical shifts specific to those groups. For convenient comparison of HR-NMR spectra obtained with different instruments utilizing magnets of different strengths, the *chemical shift* is defined as the ratio of the local magnetic field present at the observed nucleus to the full strength of the external, uniform and constant magnetic field. As the NMR measurements are usually expressed in frequency units, this definition of the chemical shift, $\boldsymbol{\delta}$, can be also expressed as:

$$\delta = (\nu_{Loc} - \nu_{ST})/\nu_{ST} \qquad \textbf{(Eq.2.2.6)}$$

, where ν_{Loc} is the nuclear spin resonance frequency of the nucleus in the sample and

ν_{ST} is the resonance frequency for a known standard chosen as a reference, such as, for example, tetra-methylsilane, $(CH_3)_4$- Si, which is the selected standard for both ^1H and ^{13}C NMR. This definition makes the chemical shift independent of the strength of the external magnetic field utilized by the HR-NMR instrument and allows for a direct comparison between spectra obtained with very different HR-NMR instruments. Very detailed, precise theoretical treatments of the NMR absorption and related processes are available in 'standard' textbooks (Abragam, 1968; Slichter, 1969). Simplified, instrument- or application- oriented textbooks (Farrar and Becker, 1971; Becker, 1980) and reviews (Baianu and Kumosinski,1993) are also available that facilitate the effective use of a wide variety of such chemically selective (and sophisticated) HR-NMR techniques by the interested analytical chemists, physical chemists, organic chemists, biochemists, or research scientists in other applied fields. As in the case of NIR spectroscopy, quantitative analyses can be performed nondestructively, quickly and routinely. The most widely employed HR-NMR techniques for quantitative analyses are based on the fact that the areas under the NMR absorption peaks corresponding to a specific component are directly proportional to the concentration of that component in the sample. Two of the most widely detected nuclei in NMR experiments are ^1H and ^{13}C. ^{13}C is a nuclear isotope of carbon that is naturally present (but with a relatively low abundance of ~ 1%) in fatty acids, lipids, and amino acids in soybean seeds. Compared to the NMR of the naturally abundant ^1H, the ^{13}C NMR has relatively low sensitivity both because of its 1% natural abundance and because of its lower resonance frequency (one quarter of the ^1H resonance frequency). Furthermore, in static solids there is a substantial line broadening caused by the chemical shift anisotropy (CSA) and by magnetic dipolar interactions. In liquids, very rapid molecular tumbling averages the chemical shift anisotropies, resulting in HR-NMR spectra with very sharp and well-resolved peaks. In static solids, chemical shift anisotropies remain as 'chemically intrinsic' features that can disguise valuable composition information which could otherwise be extracted from the isotropic chemical shifts. As a result, the ^{13}C NMR spectra of **static** solid powders are both broad and unresolved. Consequently, for the investigation of soybean solid samples, one needs to employ high-resolution NMR techniques specially designed for solids that overcome the low sensitivity and line-broadening problems. These methods, jointly labeled as 'Solid-State' NMR (SS-NMR) techniques, are employed in order to minimize first-order anisotropic nuclear interactions and to increase the S/N either by rapid sample spinning in the external magnetic field, and/or by employing special rf pulse sequences that considerably reduce magnetic dipolar interactions. Some of the more 'popular' techniques in this SS-NMR group among biochemists, analytical/organic chemists and physical chemists are the following:

- *The Magic Angle Spinning* (MAS) technique in which the whole sample is spun at an angle of $54^0 44'$ with respect to the external magnetic field, and at a rate equal to or greater than the dipolar linewidth expressed in frequency units;

- *Multiple-Pulse Sequences* (MPS) employed as composite pulse sequences that achieve homonuclear and/or heteronuclear decoupling;

- *Cross-Polarization* (CP), achieves a transfer of spin-polarization from the abundant nuclear spin population (for example, ^1H) to the rare and lower gyromagnetic ratio (e.g., ^{13}C) nuclear spin population, thus enhancing the signal to noise (S/N) for the rare nucleus.

3. EXPERIMENTATION

3.1. NIR Instrumentation

Because sample absorption data are difficult to measure directly, they are measured indirectly through reflection or transmission. NIR can, however, be employed in either the reflectance mode or the transmission mode. NIR reflectance instruments measure the amount of NIR radiation reflected from the sample at different wavelengths. NIR transmission (NIT) instruments, on the other hand, measure the amount of NIR radiation transmitted through the sample at different wavelengths. Based on the mechanism of collecting optical data at different wavelengths, NIR instruments can also be categorized as: interference filters instruments, moving diffraction grating instruments, fixed grating instruments, acousto-optical tunable filters (AOTF) instruments, Diode Array NIR (DA-NIR) instruments and Interferometer-based instruments such as FT-NIR. Filters-based NIR instruments are usually the most economical ones. The number and position of filters are designed and optimized for certain specific types of samples, and it is generally not easy to expand such instruments to other sample types. Interference filters-based NIR instruments work mostly in the transmission mode, such as the Zeltex, ZX800 and the ZX50 model instruments (manufactured by Zeltex Inc., Hagerstown, MD, http://www.zeltex.com). The major limitation of such interference filter-based instruments is that spectra are collected at only a few pre-selected wavelengths that are designed and optimized only for the major component analysis of bulk grain and oil seed samples. For the analysis of minor components like isoflavones, more flexible and powerful NIR instruments such as the DA-NIR or the Fourier Transform NIR (FT-NIR) instruments are required.

In order to collect spectral data for a large set of different wavelengths, NIR radiation can be dispersed through diffraction gratings so that signals with different wavelengths are separated, and the detector can detect signals at an individual wavelength. In the conventional configuration where a single detector is used, the diffraction grating system has to be gradually rotated, in order to project onto the detector signals of different wavelengths. Such systems are usually referred to as moving grating systems. A major limitation of such moving grating systems is due to the fact that the diffraction grating contains a moving part, which makes it difficult to obtain reproducible scans and also negatively affects the wavelength accuracy. Novel dispersive NIR instruments solve this problem by employing multiple detectors, such as diode array detectors, to detect NIR signals at different wavelengths simultaneously. In such an instrument, the NIR

radiation can still be dispersed through diffraction gratings. However, signals at different wavelengths are projected onto a stationary array of detectors, and the signals are detected simultaneously for different wavelengths. For this reason, it is no longer necessary to move the diffraction grating system. Such instruments are referred to as stationary grating systems. Since no moving grating is involved, reproducibility and wavelength accuracy/uniformity throughout the spectral range are markedly improved. Furthermore, the spectral acquisition speed is also improved dramatically because spectral data at different wavelengths are collected in parallel by such stationary grating systems, as opposed to the sequential data collection by instruments operating with moving gratings/monochromators. Typically a moving grating system takes about 30 seconds to scan an NIR spectrum at moderate resolution (i.e., 3 nm), whereas a diode-array stationary grating instrument is capable of acquiring hundreds of NIR spectra in just one second (Baianu et al., 2002[b]) at comparable resolution throughout the entire NIR spectrum.

3.2. NIR Spectra Pre-processing

NIR quantitation using the Lambert-Beer's law (**eq. 2.1.2)** requires absorbance data to be used for the concentration calculation. However, most NIR instruments do not measure absorbance directly. Instead, they measure NIR reflectance from, or transmittance through, the sample. The measured reflectance or transmittance data are then converted to absorbance data, which are normally referred to as apparent absorbance, to be differentiated from the 'true' absorbance. The apparent absorbance can be significantly affected by a variety of effects, such as specular reflection, light scattering, baseline shifts, etc. In order to improve the accuracy and reliability of NIR calibrations, NIR spectra usually have to be corrected for such effects prior to calibration model development. In fact, it has been reported that light scattering and baseline shifts may introduce more spectral variations than do the constituent contents (Williams et al, 1987). Since a calibration is the mapping between the spectral data and the constituent contents, the regression and calculations involved in the calibration development will be dominated by light scattering and specular reflection effects, instead of constituent content variations, if light scattering and specular reflection effects are not corrected first. As a result, any calibration obtained without spectral pre-processing is likely to be inaccurate, unreliable, or both.

Specular reflection effects can appear as a nonlinear baseline shift across the entire NIR spectrum. A semi-empirical approach for correcting the baseline shifts caused by specular reflection involves the definition of a set of user-selected baseline points. A baseline curve is then defined by such selected points through fitting a spline function to the points. The procedure is readily implemented with the Perkin-Elmer "SpectrumONE" program in a user-interactive mode that also allows for the subtraction of the fitted spline function/baseline curve from the NIR raw spectrum of the sample. In addition to specular reflection, the baseline shifts of NIR spectra may also be caused by electronic noise or detector response variations. In such cases, the baseline variation can appear to be constant over the entire spectral range, or may increase linearly with wavelength. First and second order derivatives of NIR spectra can then be employed effectively to remove baseline variations. The derivatives of a spectrum can be calculated by a finite difference method, which is just the difference of spectral values between two adjacent points. It is relatively easy and simple to calculate, but the S/N of the derivative spectrum will decrease. To solve this problem, Savitzky and Golay (Williams, 1987) proposed

an improved algorithm for derivative calculations, which begins with a least-squares linear regression of a polynomial of degree k over at least (k+1) data points. The derivatives of an NIR spectrum are then calculated as the derivatives of a best-fitted polynomial. The Savitzky-Golay algorithm has been proven to be very effective and the S/N is preserved in the calculated derivative spectrum.

In addition to baseline shift effects caused by the specular reflection, the electronic noise and the detector variations, light scattering is another important source of spectral variation. According to modern Quantum Electrodynamics theory (Feynman, 1963), as well as Rayleigh's simplified theory of light scattering (Kortum, 1969), when a beam of light interacts with molecules in a material, the incident light beam is partially scattered by such molecules in addition to being partially absorbed. The absorbance is linearly related to the concentrations of various components in the sample, according to eq. 2.1.2. On the other hand, light scattering is mainly caused by sample inhomogeneities, (e.g., the difference of scattering coefficients between different parts of the same sample), such as those caused by pores, a distribution of particle sizes and matrix 'texture'. The scattering coefficient is inversely proportional to the particle size of the sample, and can also be affected by variations in the packing density from sample to sample (Mie, 1908; Thiessing, 1950). According to the Kubelka-Munk theory (Kortum, 1969), light scattering affects the apparent absorbance in a multiplicative manner. Therefore, light scattering effects cannot be effectively corrected through simple, linear correction algorithms. To correct for multiplicative light scattering effects, Geladi and co-workers (Geladi *et al,* 1985) proposed a semi-empirical approach called the Multiplicative Scattering Correction (MSC), that is currently the most popular method for pre-processing NIR spectra (Isaksson, 1990). MSC begins by calculating the average spectrum of the whole set of standard samples, and then attempts to determine the multiplicative parameter (scale factor) as well as the additive parameter (shift factor) for each spectrum through a linear regression of the sample spectrum against the mean spectrum. In some applications the MSC approach was found to be very effective for correcting spectral variations caused by light scattering; as a result of MSC both the accuracy and reliability of NIR analysis were significantly improved in comparison with calibrations based on 'raw' (uncorrected) spectra. The effects of MSC applied to raw NIR spectra of single soybeans are illustrated in Fig. 3.2.1 and Fig. 3.2.2, and are quite substantial for both Dual Diode Array (Fig. 3.2.1 B) and FT-NIR spectra of soybeans (Fig. 3.2.2 B).

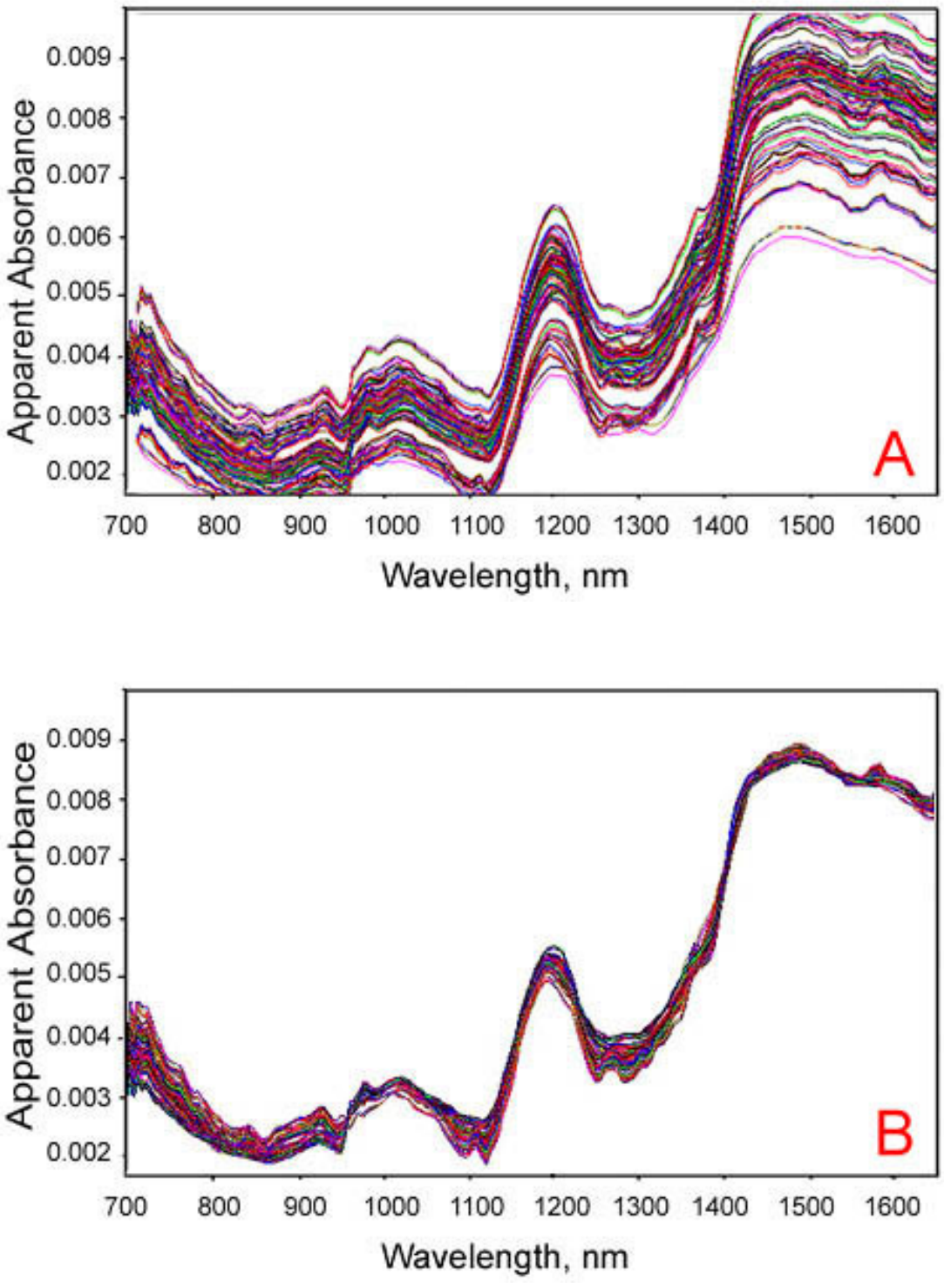

Fig. 3.2.1. An overlay plot of DA-NIR spectra of single soybean seeds obtained with the Perten DA-7000 instrument. A: Before MSC. B: After MSC.

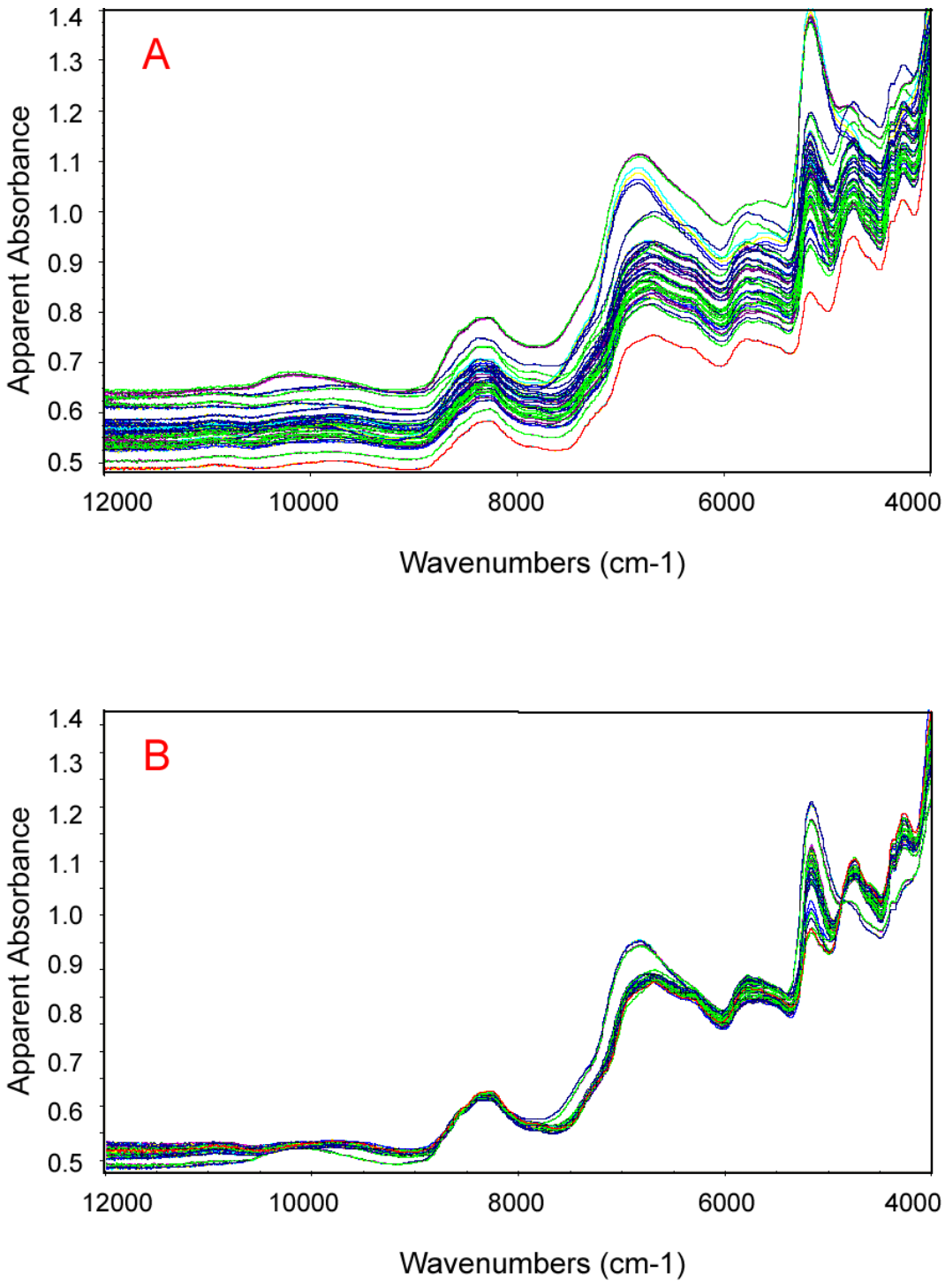

Fig. 3.2.2. An overlay plot of FT-NIR spectra of single soybean seeds obtained with the Perkin-Elmer Spectrum ONE instrument. A: Before MSC. B: After MSC.

3.3 NIR Calibration Models

After careful selection of the standard samples and accurate measurements of the composition of the standard samples for reference values, NIR spectra can be collected for such standard samples with state-of-the-art NIR instruments. With proper spectral pre-processing to correct for specular reflection and light scattering effects, the corrected NIR spectra of the standard samples can then be employed for calibration development to predict unknown samples. Calibrations are developed through regressions of the NIR spectral data against the reference values of constituent concentrations, in practice mostly through regressions of apparent absorbance data against the sample concentration data.

NIR instruments measure optical data such as reflectance from, or transmittance through, samples. The reflectance and transmittance data are usually converted into apparent absorbance. In order to predict the contents of components to be measured from the optical data, a calibration needs to be developed first. After adequate spectral data pre-processing, the calibration can be developed through regression of the corrected NIR spectral data against the reference constituent contents. As shown in the previous section on the principles of NIR, most "optical" spectroscopy quantitative analysis methods, including NIR, are based on Lambert-Beer's law which is re-cast here into a form that specifies explicitely the quantities that are wavelength-dependent :

$$a_\lambda = \varepsilon_\lambda \cdot l \cdot c \qquad (3.3.1)$$

, where a_λ is the absorbance at wavelength λ, c is the concentration of the component (analyte) to be measured, ε_λ is the absorptivity of the component at the specific wavelength λ and l is the path length. Utilizing eq.3.3.1, a direct approach to soybean NIR protein calibrations might attempt a univariate (linear) regression of the measured absorbance at an appropriately selected wavelength against the protein content of the standard soybean samples. However, because the NIR spectra of soybeans are very complex and each absorbance band often contains peaks from several different components, it remains difficult--if not impossible-- to select any specific wavelength that would be 'sufficiently' free of interference from other components to allow a reliable calibration development. One can solve this problem by taking advantage of another part of Lambert-Beer's law which simply states that the absorbance values of multiple components at are additive at any given wavelength. Consequently, an improved calibration model can be specified as:

$$a_\lambda = \varepsilon_{i\lambda} \cdot l \cdot c_i + \varepsilon_{j\lambda} \cdot l \cdot c_j + \ldots \qquad (3.3.2)$$

where a_λ and and l have the same meaning as in the previous equation, $\varepsilon_{i\lambda}$ is the absorptivity, c_i is the concentration of component i, $\varepsilon_{j\lambda}$ and c_j are defined as before for component j, and so on, for all the components present in the sample. With this model, one has to measure the absorbance for at least two different wavelengths if there are two interfering components to be measured, and a multivariate regression procedures would have to be employed. Unfortunately, even such a multivariate model is of little practical use for NIR calibration development. The major drawback of such a model is that it would require a knowledge of the complete

composition (concentration of every component) in the calibration samples, whereas in most situations one may only be interested in certain components.

One solution to this problem is obtained either by rearranging the Lambert- Beer's equation as follows:

$$c = \frac{a_\lambda}{\varepsilon_\lambda \cdot l} \qquad (3.3.3)$$

, or by combining the absorptivity coefficient (ε) and the path length (l) into a single constant , so that it takes the simpler form:

$$c = p_\lambda \cdot a_\lambda \qquad (3.3.4)$$

For complex samples such as soybean seeds, where most major components do interfere with each other, absorbance data obtained at more than one wavelength are often utilized in practice, and the above model is extended to all such selected wavelengths:

$$c = a_{\lambda_1} p_{\lambda_1} + a_{\lambda_2} p_{\lambda_2} + \ldots + a_{\lambda_m} p_{\lambda_m} \qquad (3.3.5)$$

The above model is known as the Inverse Least Squares (ILS), or the Multiple Linear Regression (MLR) model, and is widely applied in filter-based NIR instruments that collect spectral data at only a few pre-selected wavelengths. For either DA-NIR or FT-NIR instruments, that collect spectral data for hundreds of different wavelengths, it is impractical to apply such an MLR model directly to all of the acquired data points throughout the entire spectral range because such a procedure would require the calculation of a total of *m* regression parameters (usually several hundreds or thousands) with such an MLR model; this would, therefore, require that a minimum set of *m* standard samples (several hundreds to thousands) to be available for the calibration training set. One solution to this potentially severe problem would be to apply the MLR model to only a small number of spectral data at pre-selected wavelengths, but such number must not exceed the number of standard samples employed for calibration because otherwise there would be some undetermined variables. The pre-selection of such wavelengths is critical to building an accurate and robust calibration, but it is also quite difficult to do. One may know which wavelength regions should be included from the corresponding spectra of the pure components. The selection of the exact wavelengths for calibration from such regions can still be difficult, because most modern instruments have high, or very high, resolution, and therefore, even in a narrow spectral region there will be a large number of points present.

Another approach is to specify the most important region(s), based either on the pure-component spectra or the deconvolved spectra of the standard samples. Then, one could utilize a computer algorithm to select the rest of the wavelengths for the calculation, such as in the case of the Stepwise Multiple Linear Regression (SMLR) procedure provided by the TQ Analysis

software package (copyright by Nicolet Co). Even with the SMLR approach, if the number of data points included in the model is not carefully selected, over-fitting may readily occur (that is, the calibration model would have utilized too many factors); as a result, the calibration may fit the standard samples perfectly, but it will fail to predict samples that are not in the calibration training set. An improved, advanced approach utilizes a *statistical factor analysis* method, that leads to two other directly related NIR calibration models: the Principal Component Regression model (PCR), and the Partial Least Squares model (PLS). Both the PCA/PCR and the PLS model are based on ***factor analysis***, which was developed to solve problems that have many factors; such factors may also happen to be highly co-linear when the MLR is over-fitting. The principle on which both PCR and PLS are based stems from the observation that, although there are usually many different variations that make up a spectrum (such as: inter-constituent interactions, instrument variations, differences in sample handling, etc.), after proper data pre-treatments (such as, baseline corrections, light scattering corrections, e.g., MSC, etc.) the largest variations remaining in the calibration set would be only due to the chemical composition variations of the standard samples. The main purpose of both PCR and PLS is then to calculate a set of 'variation spectra' that represents only the variations caused by composition. Such calculated 'variation spectra' are sometimes called loading vectors, principal components, or more frequently, ***factors***. The calculation of such spectra usually involves an iterative process that manipulates n-samples of proper numerical values called 'eigenvectors', and for this reason PCR and PLS algorithms are also called 'eigenvector methods'. Once the factors are calculated, they are utilized instead of the raw spectra for building the calibration model; therefore, the possibility of over-fitting can be minimized by choosing the correct number of factors. Although the concepts of PLS and PCR are similar, the approaches to the calculation of the factors (loading vectors) are quite different. The PCR algorithm calculates the factors independent of the concentration information, whereas the PLS algorithm utilizes *both* the concentration and spectral information of the calibration set to calculate the factors.

The PLS method is considered, in general, to be more reliable than PCR. Besides the numerical calculation of regression parameters for the calibration, the PLS algorithm also provides qualitative information for model validation, through the first loading vector, which is usually a first-order approximation to the pure-component spectrum (Haaland et al, 1988; Sorvaniemi et al, 1993). Although PLS is an advanced multivariate regression algorithm, and has been widely applied for NIR calibration development, care still needs to be taken when applying PLS to NIR data of complex samples such as soybeans. Unlike MLR- which usually requires manually selecting the wavelengths or spectral regions for the calculation- PLS has the intrinsic ability to automatically build calibration models over the entire spectral range, thus eliminating the requirements of either manual selection of wavelengths or spectral regions. Whereas this feature might be an advantage for most types of samples, it may lead to a severe limitation of the results obtained with the PLS in the special case of samples which happen to have a very high degree of correlation between two or more component concentrations. In such special cases, the first-order loading vectors of the two correlated components may look similar, and the calibration would remain unreliable regardless of the algorithm(s), models or method(s) employed for calibration. In special cases, one might be able to minimize this problem by manually selecting for the PLS calculation those spectral regions where the pure-component absorption dominates (an approach reminiscent of MLR).

The computations of PLS and PCR are usually carried out with professional chemometrics software. There are currently several chemometrics software programs available for calibration development with PLS and PCR, such as the ThermoGalactic Graphic Relation Array Management System (GRAMS/32) (Salem, NH, www.galactic.com), ThermoNicolet TQ Analyst (Madison, WI, www.nicolet.com), Perkin-Elmer Quant+ (www.perkin-elmer.com), and Bruker OPUS (www.bruker.com). The GRAMS/32 software package is a professional spectroscopic analysis software package that supports light scattering corrections as well as PLS and PCR regression algorithms. The calibration results, including correlation plots, loading spectra, SECV plots, etc., can be exported to Microsoft Office subprograms such as Excel. It can also be expanded by allowing the user to write special programs in the Array Basic programming language. The TQ Analysis software package, on the other hand, provides several calibration features that are user-friendly. It supports light scattering corrections (MSC), as well as spectral smoothing, and also includes the options of CLS, MLR, PCR and PLS regression analyses. Even though the TQ program is not as expandable as GRAMS/32, it is specifically designed and optimized for FT-NIR instruments. In our NIRS and FT-NIR studies, both the GRAMS/32 and the TQ Analyst were routinely employed.

3.4. NMR Techniques for Oil Determination in Soybean

3.4.1 Simple One-Pulse (1PULSE) High-Resolution NMR

The simple, 1PULSE ^{1}H NMR method provides a direct means for measuring the oil content in somatic soybean embryos and soybean oil samples. This method uses just one radio frequency (rf) pulse during each acquisition cycle (Fig 3.4.1.1). The rf pulse excites all ^{1}H nuclei in a sample, and a characteristic ^{1}H NMR time-domain signal is observed. The single pulse employed by this method has a defined width that maximizes the initial amplitude of the NMR signal; this pulse width is the time interval during which the resonant rf pulse of average power *pw* is applied to the sample, resulting in a 90 degree flip of the nuclear spin magnetization from the direction of the constant, external magnetic field.

The hydrogen nucleus (^{1}H), with a spin of $\frac{1}{2}$, is usually selected for NMR measurements because it is the most abundant isotope present in natural biomaterials. The rf pulse selected for HR-NMR has a characteristic, resonance frequency which is proportional to the magnetic field strength employed by the instrument. In our measurements, a Varian U-400 spectrometer model was employed, and the applied radio frequency pulse was at the ^{1}H resonance frequency of 400 MHz, in an external magnetic field of 9.4 T.

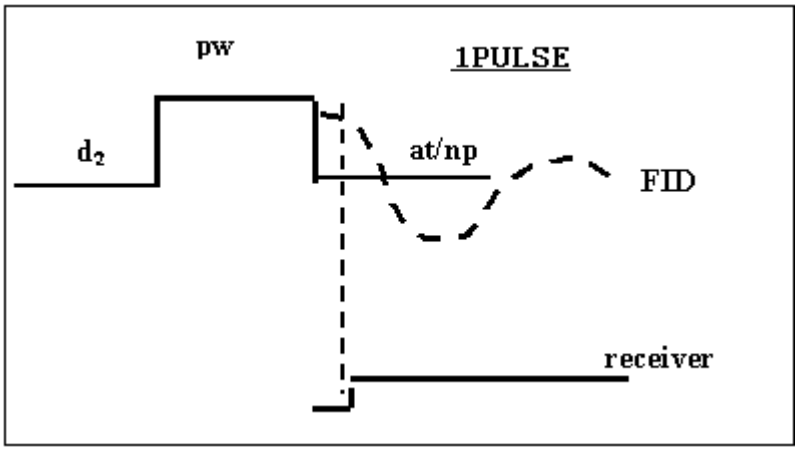

Fig. 3.4.1.1. Simple 1PULSE Sequence for High-Resolution NMR Analysis of Oil in seeds.

In the case of our high-resolution NMR studies of oil in mature soybean seeds and embryos, the number of selected points was np = 65,536. The FFT of an FID produces an HR-NMR spectrum which represents the variation of the NMR absorption intensity with the nuclear spin resonance frequency. To avoid the possibility of rf saturation, nuclear spins must be allowed to relax (that is, without any additional rf excitation being applied) for a significant interval of time called delay time, or d_2, until the next 90^0 rf pulse is applied. For a low viscosity liquid that does not contain either paramagnetic or ferromagnetic species, the length of time required for the nuclear spin relaxation to occur is at least on the order of the reciprocal of the half-height linewidth for the sharpest observed absorption peak in the HR-NMR spectrum of the liquid. For typical HR-NMR studies the line broadening (lb) is selected to be less than ~0.2 Hz, and therefore the delay time, d_2 , required for nuclear spin relaxation is typically on the order of 5 s or longer. To compensate for the very weak NMR absorption signal of oil from the soybean seed or embryo samples, S/N in the oil spectra was improved more than twenty-fold through the accumulation of at least 400 transients, while the gain parameter of the rf pre-amplifier and receiver was held constant during all HR-NMR acquisitions.

3.4.2 Low-resolution NMR for Oil Determination in Seeds: AOCS recommended method Ai 3-75 for Oil Content

The time-domain pulsed NMR method is an AOCS recommended standard method (AOCS Recommended Practice AK4-95) for rapid and simultaneous determinations of oil and moisture contents of oil seeds. This method can accurately measure oil seed samples with less than 10% moisture. Drying is stated to be necessary for the higher moisture samples. The method usually involves the following steps:

1. Place the test sample into the magnetic field of the NMR spectrometer;

2. Apply an intense 90^o radio frequency (rf) pulse to excite all the hydrogen nuclear spins;

3. Record the free induction decay (FID) following the 90^{o} rf pulse. The maximum amplitude of the FID signal is proportional to the total number of protons from the water and oil phases of the sample;

4. Apply a second, 180^{o} , refocusing, rf pulse to produce a spin-echo signal when only the signal from the oil phase contributes to the FID;

5. Calculate the difference between the two component signal amplitudes, one of which is proportional to the oil, while the other is proportional to the moisture content. Then convert the measured signal intensity from water and oil into percentages of oil or moisture content with an established calibration.

This method has been applied to soybean and sunflower seed analysis and was reported to have only 0.6% error for oil determination. The calibrations employed to relate the FID signal to oil and moisture percentages are critical for the accuracy and reliability of this method. For best performance, the calibration samples should be homogenous, free from impurities, and of the same type as the test samples; this is so because different types of oil seeds may have different fatty acid profiles which would result in different time dependences for the FID amplitude. It is recommended that the oil content of calibration standards should be determined with the reference method described in AOCS Ai 3-75.

3.4.3 1PDNA ^{13}C SS-NMR technique for Oil Content Determination in Soybean Flours

Soybean flours can be directly measured for oil content determination by employing a composite, 1PDNA pulse sequence. Solid-state ^{13}C NMR spectra were recorded with a General Electric, GN300WB model, FT-NMR instrument, operating with a 7.05 T, wide bore superconducting magnet. The pencil-shaped CP-MAS probe allowed for the insertion of a 7.5 mm diameter rotor made of zirconium. The NMR pencil probe components are as shown in Fig. 3.4.3.1. The same NMR probe is employed for experiments which require spinning the rotor at high-speed rates, with the rotor axis at the magic angle (54deg,44min) with respect to the external manetic field (z) direction. The maximum spinning rate of the rotor was ~6 kHz with all our samples and was simply achieved with nitrogen gas from the building supply. The active volume in the coil could be filled with ~300 mg of sample. Considering the fact that the gyromagnetic ratio for ^{13}C is just one quarter of that for ^{1}H, the center frequency for the ^{13}C NMR spectrum in the 7.05 T superconducting magnetic field of the GN300WB spectrometer was near 75 MHz.

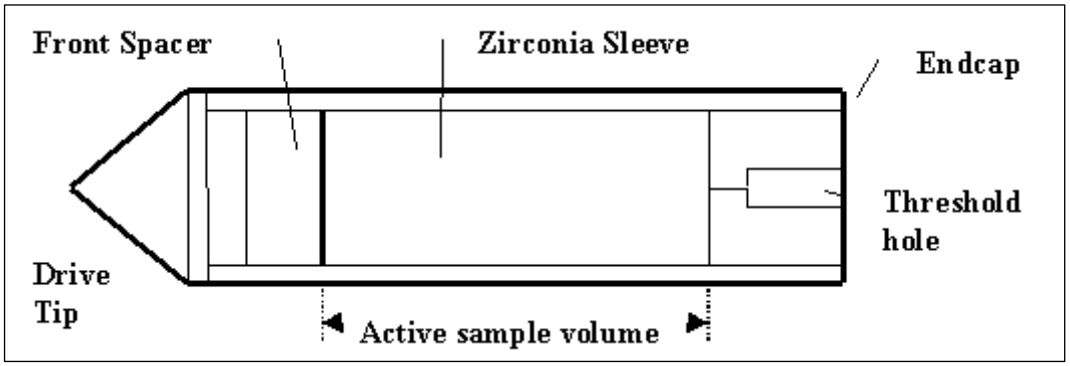

Fig. 3.4.3.1. Diagram of the pencil probe employed in a General Electric, GN300WB model, FT-NMR spectrometer, with a zirconium rotor sleeve, Kel-f drive tip, Teflon front spacer and end cap.

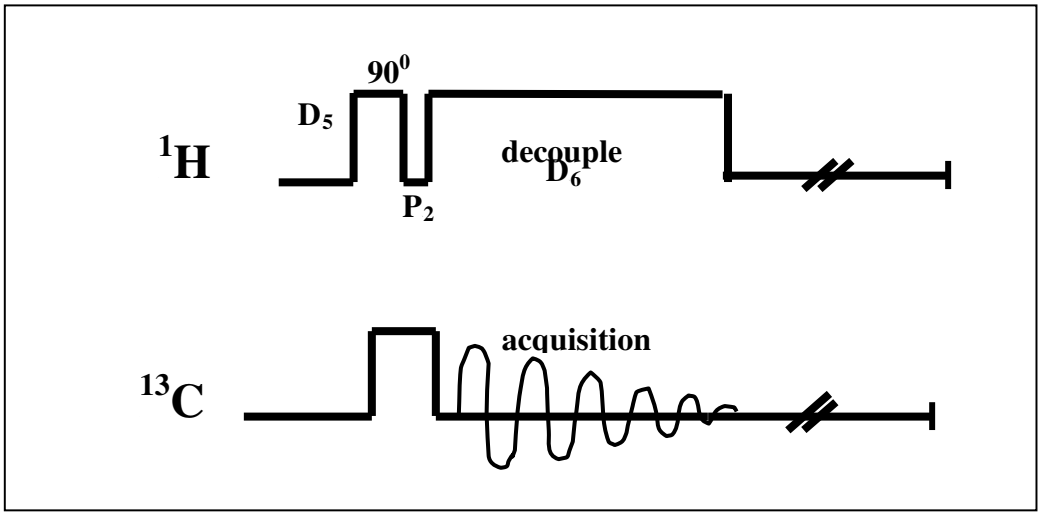

Fig. 3.4.3.2. The 1PDNA NMR pulse sequence employed in our ^{13}C SS-NMR measurements of oil content in soybean flours.

3.4.4. The VACP ^{13}C SS-NMR technique for Measurements of Protein Content in Soybean Flours

The Variable Amplitude Cross-Polarization (VACP) experiment is performed by applying a pulse sequence that transfers polarization from the ^{1}H to ^{13}C nuclear spins, in the presence of sample spinning at the magic angle with respect to the external magnetic field. The artificially imposed, fast sample spinning averages out the ^{13}C chemical shift anisotropy. The purpose of the

VACP NMR pulse sequence is to enhance the ^{13}C NMR signal through cross-polarization from ^{1}H to the neighboring ^{13}C nuclear spins. The pencil probe for solids was employed in the General Electric GN300WB (7.04 T) spectrometer to measure 300 mg samples of soybean flours without any additional sample preparation. The number of transients selected in this case was 1,600 for each soybean flour sample, thus allowing for a 40-fold improvement in S/N.

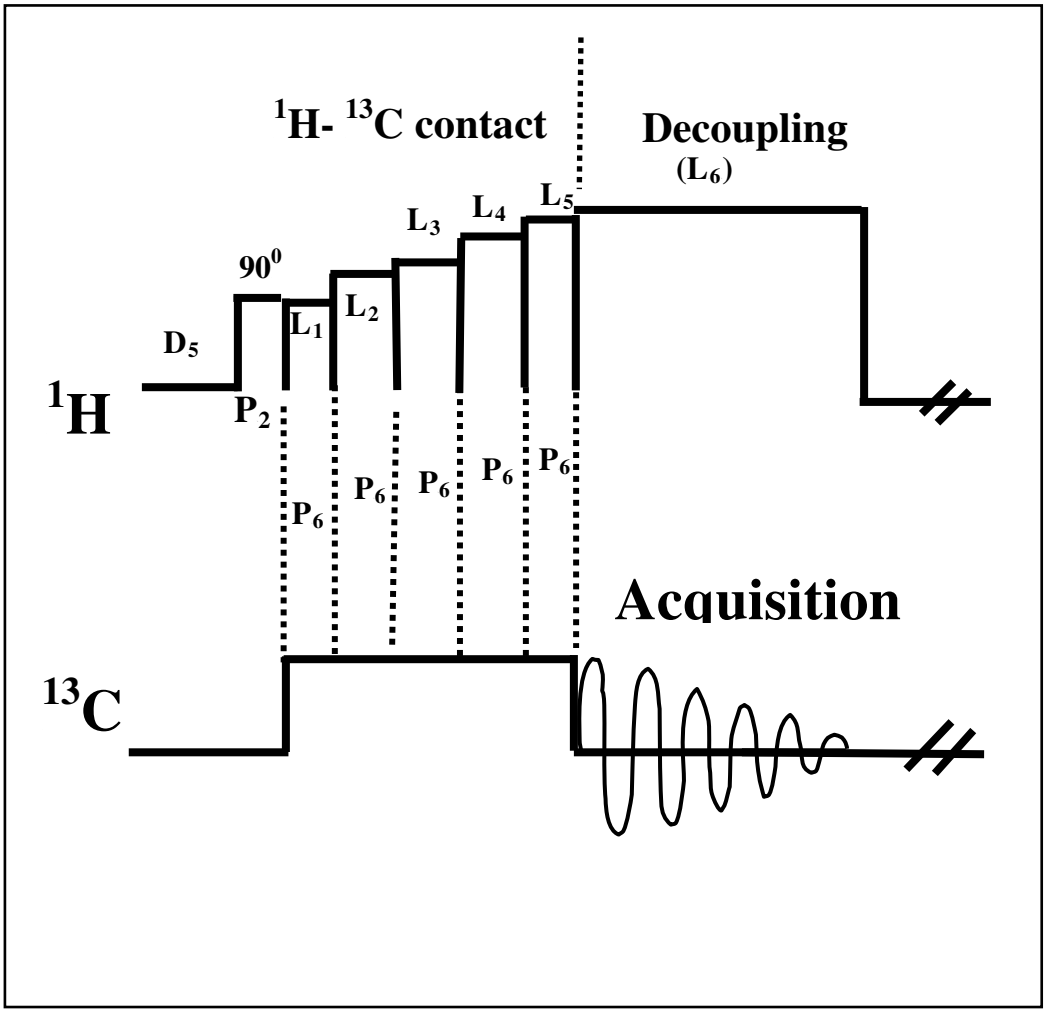

Fig. 3.4.4.1 The VACP NMR pulse sequence employed in our ^{13}C SS-NMR measurements of Protein Content in Soybean Seed Flours**.**

3.4.5 Liquid-State ^{13}C NMR Measurements of Protein Content and Amino Acid Residues in Hydrated Soy Flour Gels

Solid sample composition information that could be provided by the averaged, Isotropic Chemical Shift Isotropy (CSI) is hidden by the very broad bands present in static and rigid solids that possess large Chemical Shift Anisotropy (CSA). In liquids, rapid molecular tumbling averages out anisotropies and, therefore, NMR spectroscopists often employ dilute solutions to

acquire high-resolution NMR spectra. Nevertheless, it is often the case that highly hydrated concentrated samples, such as hydrated gels, still exhibit higher resolution ^{13}C NMR spectra than those obtained with the help of various SS-NMR techniques, by virtue of the segmental mobility in high molecular weight biopolymers, in those sample regions that are highly hydrated as in soft gels of various hydrated biopolymers (Baianu et al, 1989).

3.4.6 Protein Content and Amino Acid Profile Determination with the WALTZ- 16, ^{1}H Decoupling Sequence for ^{13}C Liquid-State NMR of Highly Hydrated Soy Flour Gels and Doughs

The WALTZ-16 ^{1}H decoupling pulse sequence for ^{13}C NMR, is a composite pulse sequence that employs ^{1}H broadband decoupling, as well as refocusing of the heteronuclear interactions, by applying a refocusing 180 deg. pulse to the ^{13}C nuclear spins as shown in Fig 3.4.6.1. In order to determine the protein content and amino acid profiles of soybean seeds we employed a Varian UI-600 spectrometer that operates at 150 MHz resonance frequency for ^{13}C NMR in a 14.1 T external magnetic field. Samples of soy flour gels of various dilutions in D_2O at pH ~11.2 were carefully placed in a 10 mm probe for solutions. Spectra were recorded with 10,000 transients, with a ^{13}C pulse width of 8.0 μs; the recycle delay employed was 4.0 s and the acquisition time was 0.62 s. The selected spectral width was 52.8 kHz (~350 ppm).

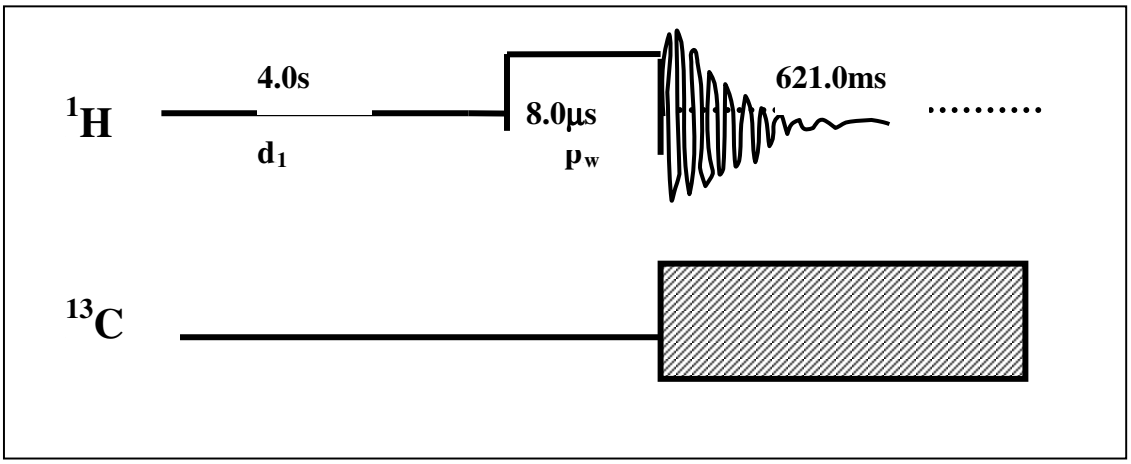

Fig. 3.4.6.1. The WALTZ-16 Decoupling pulse sequence for Liquid-State ^{13}C NMR.

3.5 Standard Methods for Soybean Composition Analysis

Understanding the limitations and assumptions involved in standard methods is essential for generating high quality calibrations; any large and unexplained variations in the content of any of the components in the standard samples can result in large errors of prediction for the constituents of interest. Therefore, the analytical methods for oil, protein and moisture determination will be briefly discussed as they have been employed for the purpose of NIR calibrations for these major soybean seed components.

3.5.1 Oil determination

Compared to protein determination methods, the oil determination method most commonly employed is relatively straightforward. Both oil and fats belong to the class of lipids, which by definition is a group of substances generally soluble in organic solvent and insoluble in water. Oil refers to liquid lipids at room temperature whereas 'fat' refers to the solid lipids at room temperature. Since oil consists of a mixture of hydrophobic molecules that are soluble in organic solvent and insoluble in water, the total oil content of a sample can be determined by organic solvent extraction.

3.5.1.1. Solvent extraction methods (AOCS official method Ac 3-44)

Based on the extraction operation, the organic solvent extraction method can be categorized as a continuous solvent extraction method, a semicontinuous solvent extraction method, or a discontinuous solvent extraction method. The semicontinuous extraction method is most widely employed in analytical laboratories and it normally utilizes a Soxhlet distiller or similar devices. The AOCS official method (Ac 3-44) for oil determination of soybean samples is the semi-continuous method.

The AOCS official method specifies petroleum ether as the solvent to extract oil from ground soybean meal in a Butt-type extraction apparatus such as Soxhlet distiller. The basic operation involves the following steps: First, weigh 2 g of ground sample and enclose the sample in filter paper. Then place the sample in the Butt tube device and extract the sample with petroleum ether for 5 hrs. Next, evaporate the petroleum ether on a steam bath or in a water bath. Finally, weigh the mass of the extracted oil. The oil content of the sample can be calculated as the percentage of extracted oil over the total mass of the sample. To get accurate and reliable results, it is important that the powder sample is fine enough as it has been found that particle size of the ground soybean affects the extraction. In addition, the moisture content of the sample is also important. If the moisture in the sample is too high (>10%), the sample may also need a drying pretreatment.

3.6. Protein Analysis

Various techniques were utilized to determine the protein content in soybeans. However, each one has its advantages or drawbacks, and they should be therefore considered as complementary to each other.

The Kjeldahl method is one of the widely employed methods for measuring organic nitrogen content in grains, and it is also the official method for protein analysis recommended by the AOCS (Ac 4-91). The total organic nitrogen of the sample is calculated and converted into the percentage of protein by multiplying by a predefined constant. However, the digestion process requires some catalysts to increase speed and it is affected by changes in temperature.

The Biuret method is also employed to determine protein content for relatively large samples. It is considered by many researchers to be more accurate than the Kjeldahl method for protein measurements because it utilizes the reaction between the peptide bond and copper ions; on the other hand, Kjeldahl qunatitates only the total nitrogen, and cannot distinguish between protein and non-protein nitrogen. The Biuret method does have relatively low sensitivity, and it requires calibration with known protein concentration standards. A related method to Biuret is the Lowry method, which is perhaps the most widely applied method for determination of protein content in solutions. It combines the Biuret reaction with the reduction of the Folin-Ciocalteau phenol reagent (phosphomolybdic-phosphotungstic acid) by aromatic amino acids tyrosine and tryptophan residues in the proteins. The Lowry method has very high sensitivity; however, the color reaction may vary with different proteins to a greater extent than with the Biuret method. Ohnishi and Barr made a modification of the Lowry method in their procedure, thus combining the advantages of the Biuret method with those of the Lowry method, and also resolving the limitations of the latter (Ohnishi and Barr, 1978). Their procedure is the basis for the current Sigma Chemical Co. (St. Louis, Missouri) micro-protein determination procedure No.690. This procedure has also been employed in our laboratory for protein determination and was calibrated with soybean protein standards of known purity and composition.

3.7. High Performance Liquid Chromatography Analysis of Derivatized Amino Acids from Hydrolyzed Proteins

A method that is often preferred by analytical laboratories in order to generate 'standard' amino acid profiles of proteins is High Performance Liquid Chromatography (HPLC) of hydrolyzed proteins. However, this method does not allow for the measurement of Tryptophan (Trp), Glutamine (Gln) and Asparagine (Asn) residues. Only values of Glx = Gln +Glu and Asx=Asp +Asn can be reported with this method as the acid hydrolysis converts all Gln into Glu (Glutamic Acid), and all Asn into Asp (Aspartic Acid). Before actual HPLC measurement, the remaining 18 amino acid residues are derivatized with special fluorochrome reagents, such as the AccQ-Fluor reagent (6-aminoquinolyl-N-hydroxysuccinimidyl carbamate) in a borate buffer (Waters Co., Milford, MA, USA). After obtaining linear HPLC standard plots for the 18 amino acid residues that are contained in acid hydrolyzates of proteins, one can proceed to attempt NIR calibrations based on such partial HPLC data for the same group of protein hydrolyzates. This approach was recently attempted with soybean samples and a brief summary of NIR calibrations was reported (Anderson, Killam and Orf, 2002) for amino acid profiles of ground soybean samples measured with the dispersive NIRS Model 6500 instrument (NIRS Systems, Silver Springs, MD) operated in the reflection mode. The only major drawback of this approach, apart from the Gln and Asn conversion to the acid forms, are the relatively large errors introduced by the acid hydrolysis for several of the more labile amino acid residues, thus limiting the usefulness of the approach to perhaps 10 of the 18 amino acid residues that are being separated by HPLC.

3.8. Moisture determination methods

Moisture is probably the most widely analyzed component for food products. There are, however, several precautions that need be observed in order to obtain accurate and reproducible moisture measurements. Water in food products and oil seeds can be dynamically distributed over at least three different types of water populations, i.e., free, adsorbed and trapped. Most moisture determination methods determine the amount of water in food products by measuring the difference of mass before and after removing water from the sample, in most cases by drying the sample for extended periods of time at temperatures close to the water boiling point. Because not all the water populations present in a food product, or an oil seed, can be readily removed by drying at a specific temperature, drying methods for moisture determination are susceptible to inconsistency. The most widely employed moisture determination method for grains and oil seeds is the oven drying method. For oven drying, the sample is heated under specified conditions and the weight loss is measured to calculate the moisture content of the sample. Drying conditions, such as the type and condition of the oven, and the time and temperature of drying, can significantly affect the results. In the ASAE standard method (ASAE S352.2) for soybean moisture determination, it is required that 15 grams of whole, ungrounded soybean seeds be dried at 103 °C for 72 hrs. To determine the moisture content of low moisture products the Karl Fischer titration method could also be applied. This chemical method is based on the fundamental reaction involving the reduction of iodine by SO_2 in the presence of water. However, its rate of success with several oil seeds, such as corn and soybean seeds has been rather low.

4. RESULTS

4.1 Validation of the NIR Calibrations for Protein and Oil Measurements in Mature Soybean Seeds: Bulk and Single Seed Calibrations

After appropriate spectral corrections for light scattering effects and baseline shifts, the DA-NIR and FT-NIR spectra of the standard samples were employed for calibration development. For both the DA-NIR and FT-NIR instruments, calibrations were developed based on the PLS-1 model and they were validated with the corresponding deconvoluted spectra. The number of factors for the PLS-1 models was optimized by cross validation; the prediction errors of the calibration models were also estimated by employing cross validation. The correlation coefficients (R) and Standard Error of Cross Validation (SECV) of the DA-NIR calibration for protein and oil measurements are presented in Figures 4.1 to 4.4 for the FT-NIR instrument, and in Figures 4.5 to 4.8 for the DA-NIR instrument. In addition, the calibration results are also presented in Tables 4.1 and Table 4.2. From Figures 4.1 to 4.4 and Table 4.1, one can see that the SECV values for protein and oil analysis for both bulk soybean samples and single seed soybean samples are fairly low. For bulk sample analysis, the SECV value is quite low, ~0.1% for both protein and oil calibrations. For the single seed analysis, the SECV value for protein analysis is 1.1% and that for oil is 0.5%. From Figures 4.1 to 4.4 and Table 4.1, one may note that very accurate results can be obtained with the FT-NIR instrument. The SECV values for protein and the oil FT-NIR analysis of bulk samples are similar to the results obtained with the DA-NIR instrument, whereas for single seed analysis, the FT-NIR instrument seems to be more accurate. This is as expected, and it is easily explained by the fact that FT-NIR instruments utilize an integrating sphere accessory and a much narrower beam, which is more appropriate for single seed analysis.

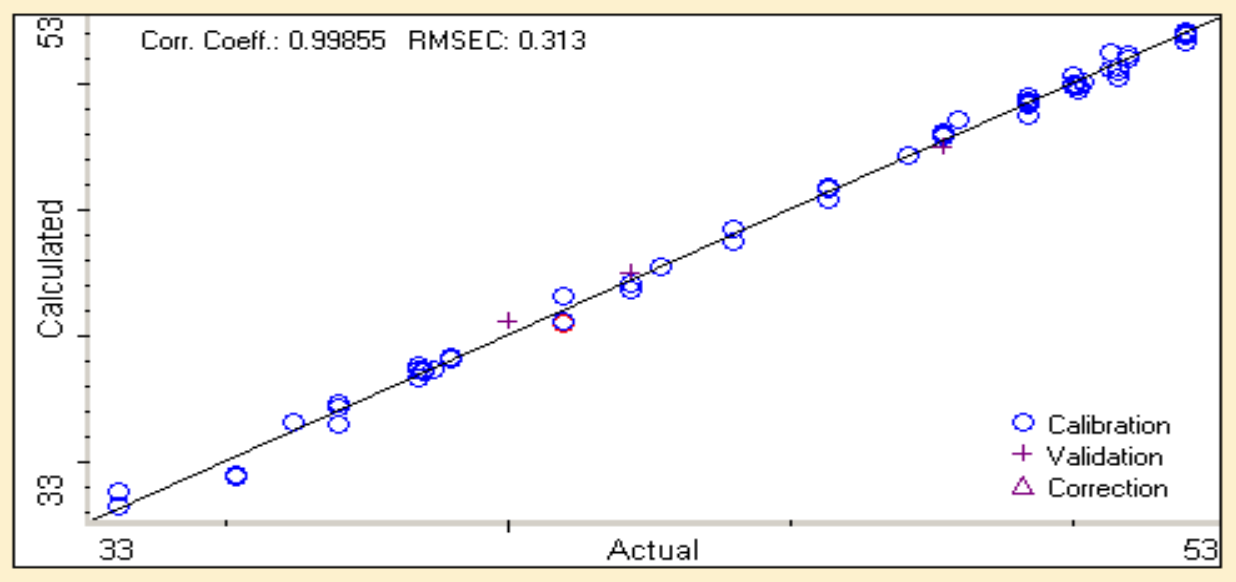

Fig. 4.1.1. Standard Protein Values vs. Calculated Values by FT-NIR Calibrations for Single Seed Soybean Analysis.

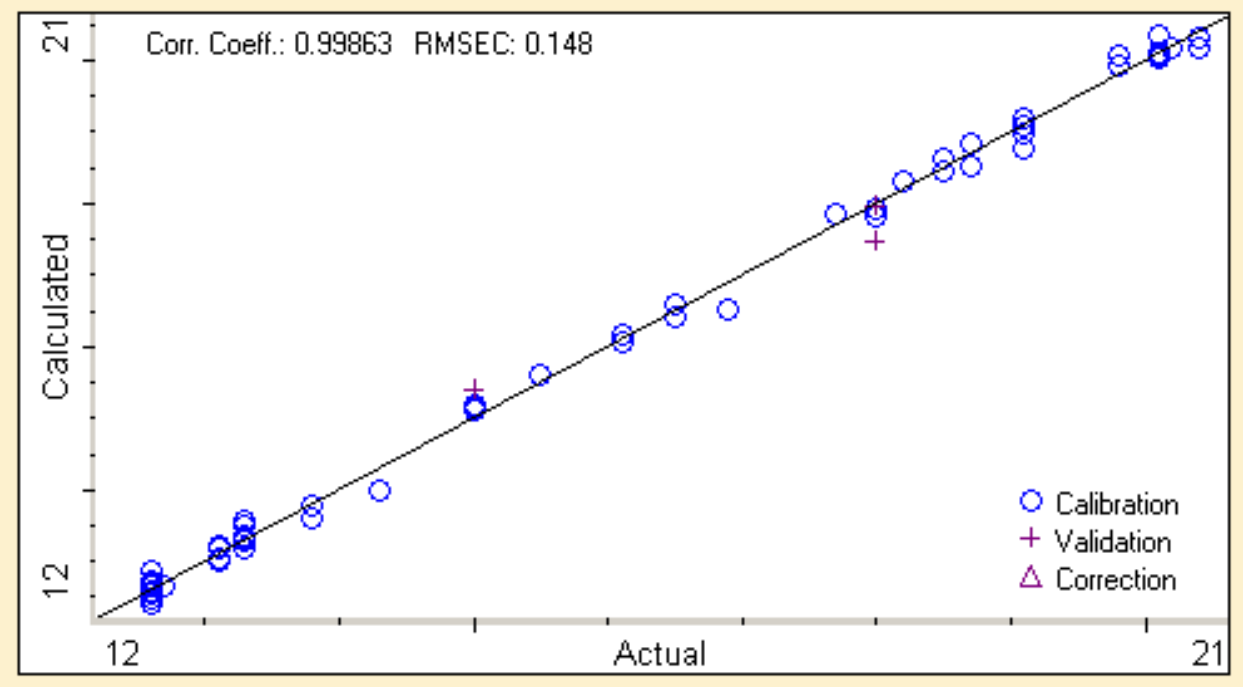

Fig. 4.1.2. Standard Oil Values vs. Calculated Values by FT-NIR Calibrations for Single Seed Soybean Analysis.

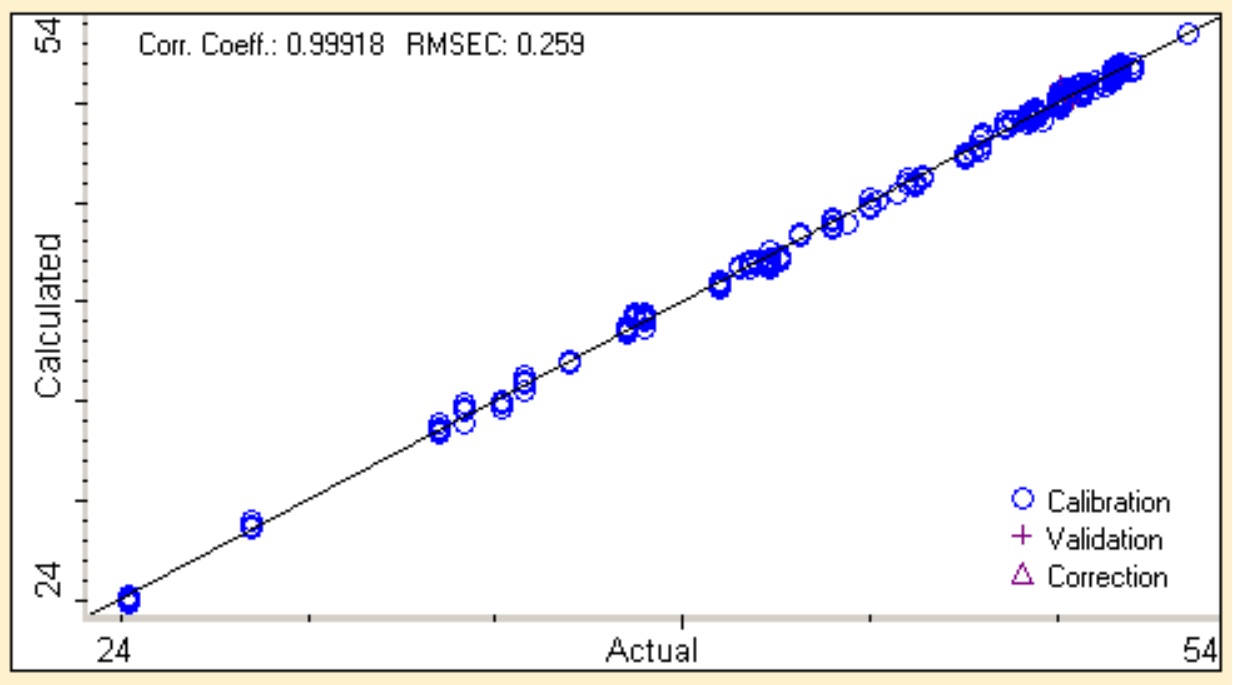

Fig. 4.1.3. Standard Protein Values vs. Calculated Values by FT-NIR Calibrations for Bulk Soybean Sample Analysis.

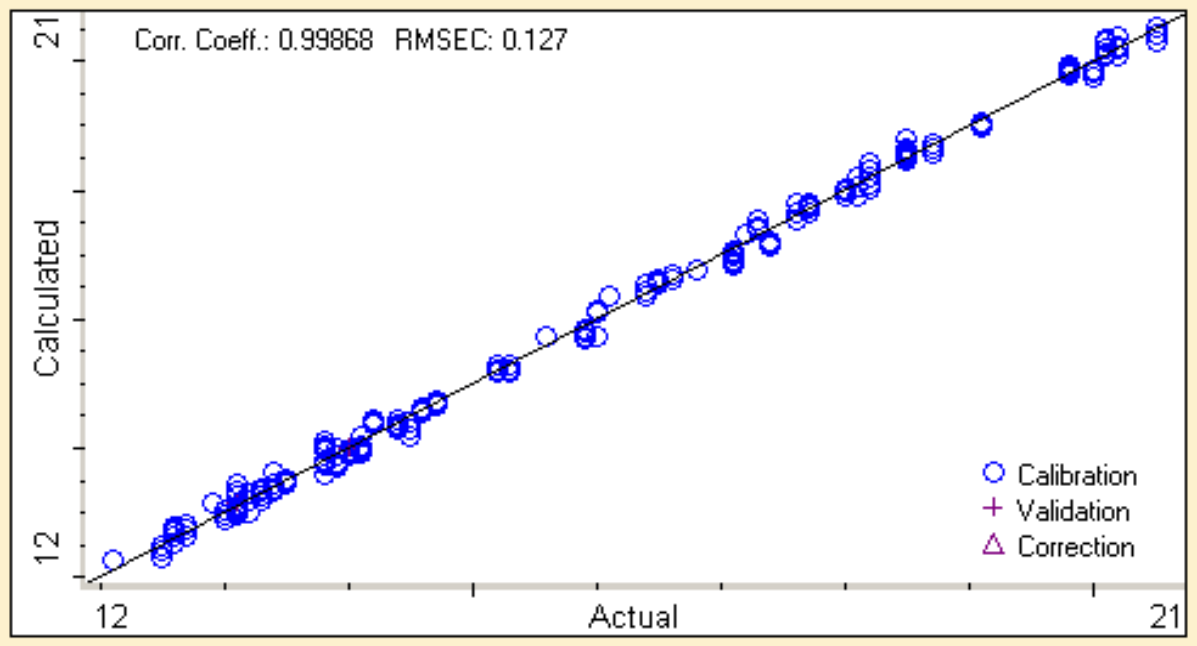

Fig. 4.1.4. Standard Oil Values vs. Calculated Values by FT-NIR Calibrations for Bulk Soybean Sample Analysis

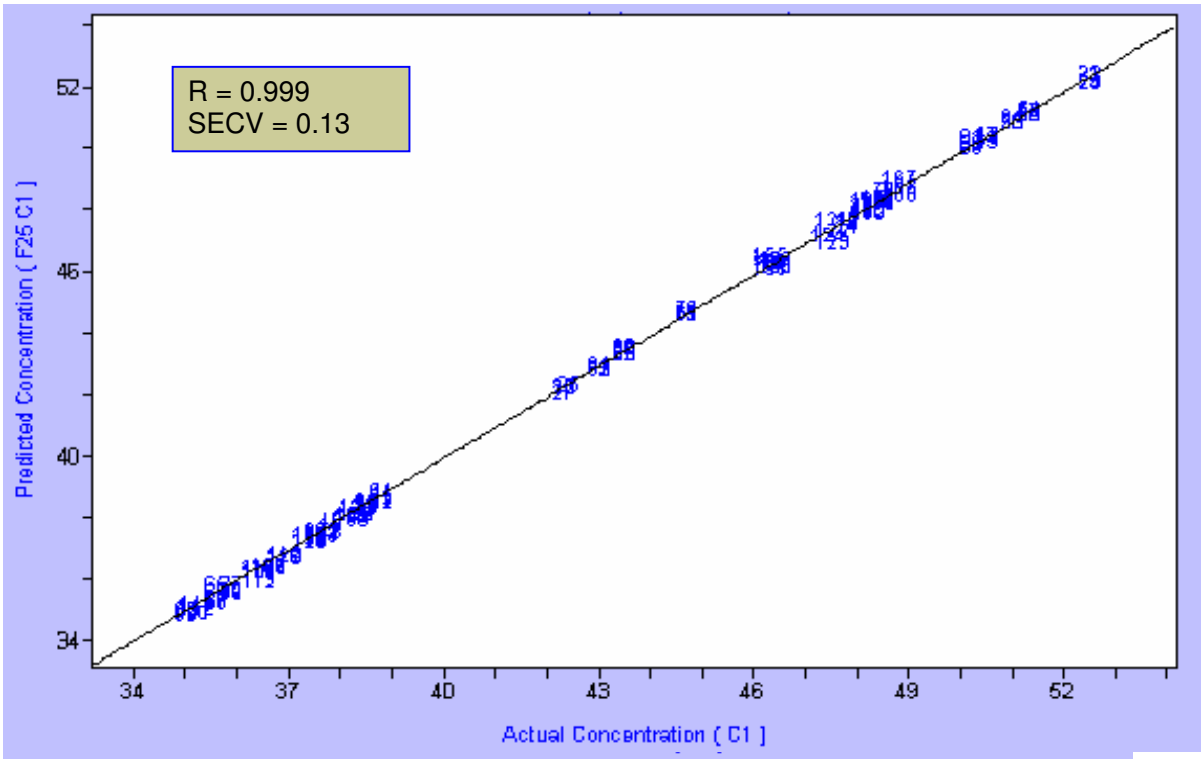

Fig. 4.1.5. Standard Protein Values vs. Calculated Values by DA-NIR Calibrations for Bulk Soybean Sample Analysis.

Table 4.1.1. Correlation coefficients (R) and Standard Error of Cross Validation (SECV) for Soybean Protein, Oil, and Moisture Analysis on the Perkin-Elmer Spectrum ONE NTS FT-NIR Instrument.

Components	**Protein**		**Oil**	
	Bulk Sample	**Single Seeds**	**Bulk Sample**	**Single Seeds**
SECV	0.3	0.3	0.1	0.2
R	99.9%	99.9%	99.9%	99.9%

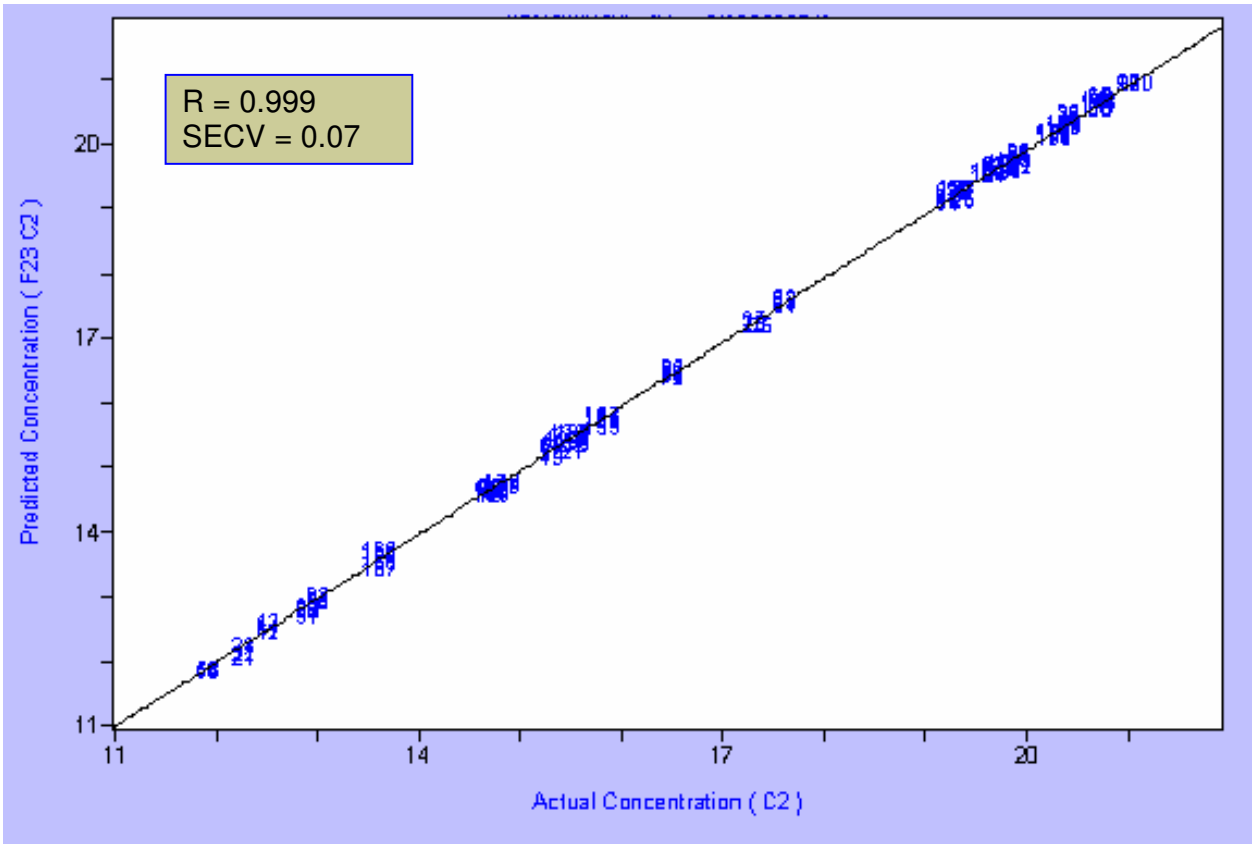

Fig. 4.1.6. Standard Oil Values vs. Calculated Values by DA-NIR Calibrations for Bulk Soybean Sample Analysis.

Table 4.1.2. Correlation coefficients (R) and Standard Error of Cross-Validation (SECV) for Soybean Protein and Oil Analysis on the Perten DA-7000, Dual Diode-Array DA-NIR Instrument.

Components	**Protein**		**Oil**	
	Bulk Sample	**Single Seeds**	**Bulk Sample**	**Single Seeds**
SECV	0.1	1.1	0.1	0.5
R	99.9%	98.5%	99.9%	98.5%

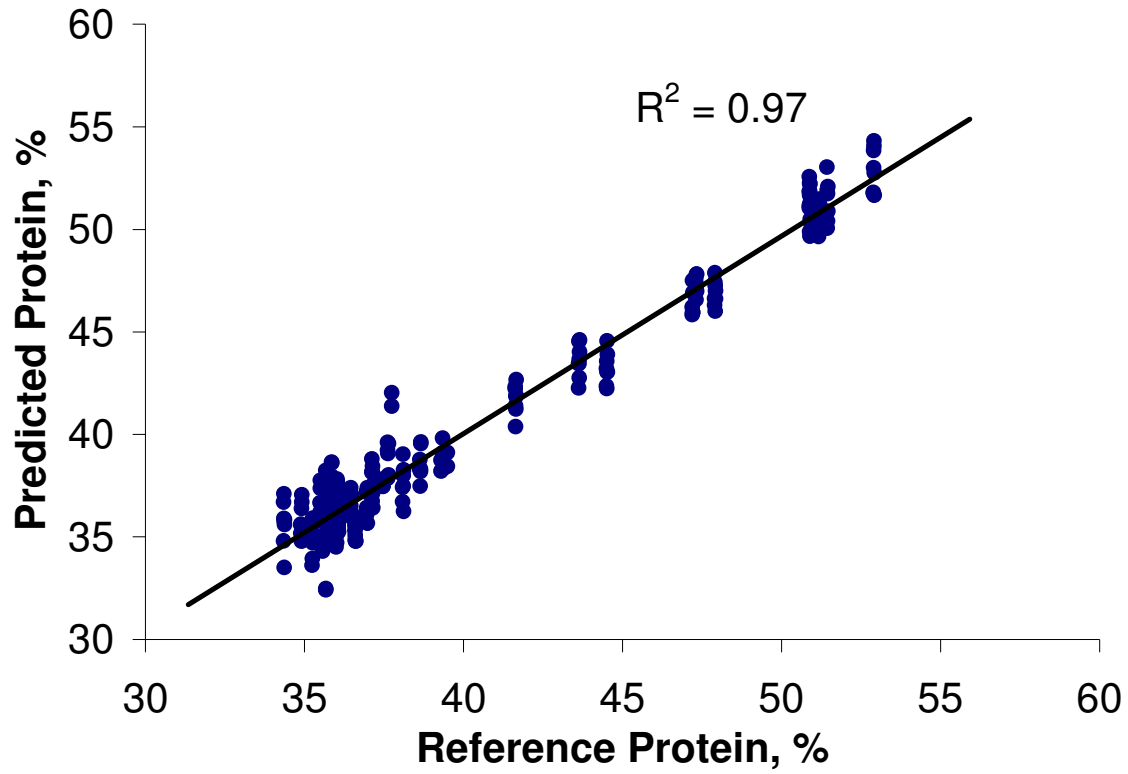

Fig. 4.1.7. Standard Protein Values vs. Calculated Values by DA-NIR Calibrations for Single Seed Soybean Analysis.

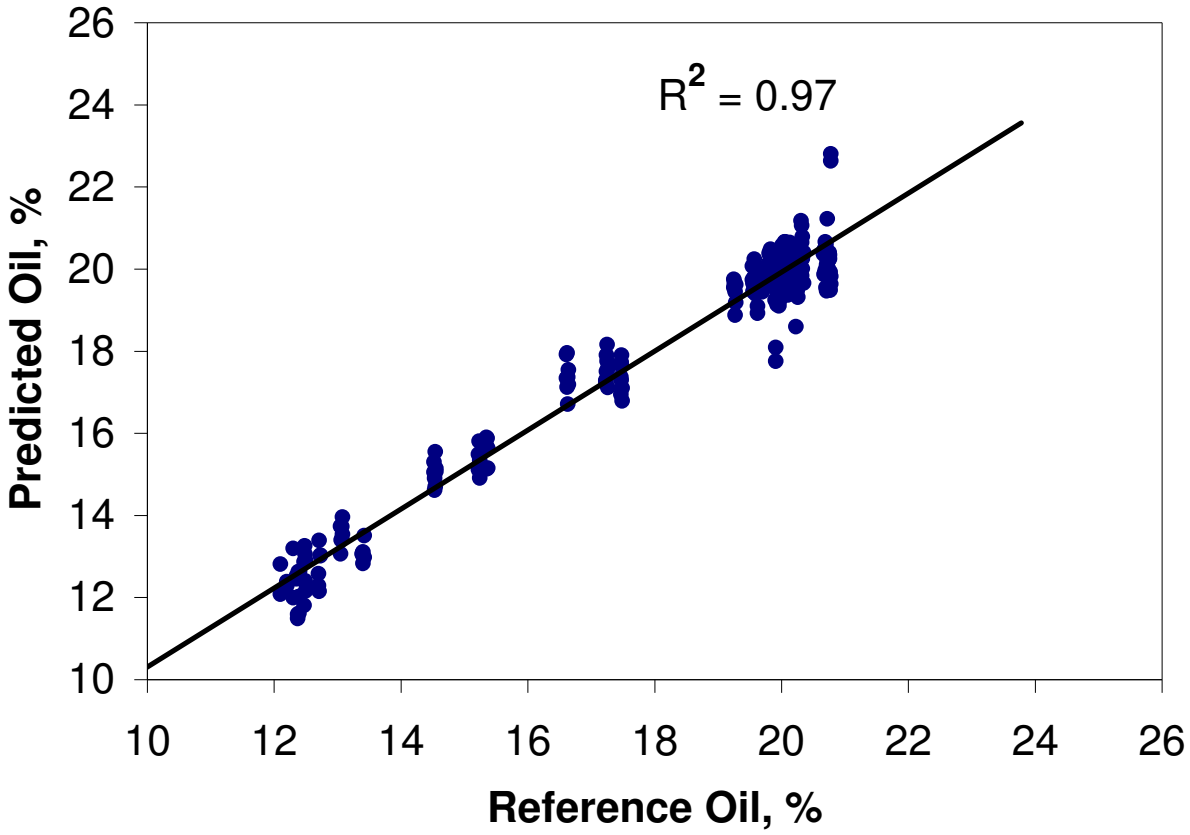

Fig. 4.1.8. Standard Oil Values vs. Calculated Values by DA-NIR Calibrations for Single Seed Soybean Analysis.

4.2Oil and Protein Determination in Mature Soybeans Using NMR Techniques

4.2.1 Decoupling Sequence for ^{13}C Liquid-State NMR of Highly Hydrated Soybean Flour Gels and Doughs

The ^{1}H decoupled ^{13}C NMR spectra of gel samples of soybean flour, protein isolate and oil that were recorded with the WALTZ-16 ^{1}H decoupling pulse sequence are presented in Figs. 4.2.1.1 to 4.2.1.3.

It was previously reported for soybean proteins (Baianu and Kumosinski, 1993; Kakalis and Baianu, 1989, 1990) that the region of interest for soybean protein content determination is located in spectral region 4, between 173 ppm and 181 ppm, as shown in Fig.4.2.1.1. Indeed, we found the ^{13}C NMR peaks of 18 amino acid residues to be present in this region. In the same figure, spectral region 3 was dominated by signals coming from the major components of the teguments, cellulose and hemicellulose. Signals from different carbons of triacylglycerols have resonances in region 2. The peak in spectral region 1 is assigned to the methyl group signal, and

the peaks in region 5 are assigned to glycoproteins. In Fig. 4.2.1.3 the peaks at 59 ppm and 66 ppm are assigned to the Cα and Cβ carbons of glycerol, whereas the peaks at 125 ppm and 127 ppm are assigned to the ethylene carbons of fatty acids.

The proposed assignments of the C_α peaks for all essential amino acids were in accordance with BioMagResBank, as follows: His (81.40 ppm), Ile (77.23 ppm), Leu (69.90 ppm), Lys (42.63 ppm), Met (47.13 ppm), Phe (73.46 ppm), Thr (51.93 ppm), Trp (47.13 ppm) and Val (51.93 ppm). Similar ^{13}C NMR assignments were made previously for wheat proteins (Baianu, 1981; Baianu et al., 1982; 1989) and corn zeins (Augustine and Baianu,1986; Baianu, 1987; Baianu and Kumosinski,1993). We also found that the carbonyl peaks of 18 of the amino acids present were close to ~172 ppm. The amino acid profiles of the soy protein and/or soybean flour were obtained from the integral values of the C_α peaks ratioed to the integral value of the carbonyl peak at ~172 ppm. The amino acid profile can then be used as a database for the protein evaluation method called the *protein digestibility corrected amino acid score (PDCAAS)*, which takes into account the amino acid profiles of specific protein groups.

In previous reports of high-resolution ^{13}C NMR studies of purified soybean protein fractions in solutions (Kakalis and Baianu, 1989, 1990; Wei, 1990), it was shown that the best resolution of the amino acid residue peaks in the NMR spectra was obtained for pH values close to 11. Therefore, the protein content of soybean flour gels was determined under such alkaline pH conditions by employing the standard plot obtained for various dilution levels of SPI gels. A standard ^{13}C NMR calibration plot for dilute SPI solutions is presented in Fig. 4.2.1.4. This was obtained by plotting the ratios of the peak integral values for the region between 173 ppm and 181 ppm in the ^{13}C NMR spectra of SPI solutions and/or gels, against the known SPI concentration. The linear regression equation for fitting the standard plot was then employed to calculate the soybean protein content of soybean flour gels.

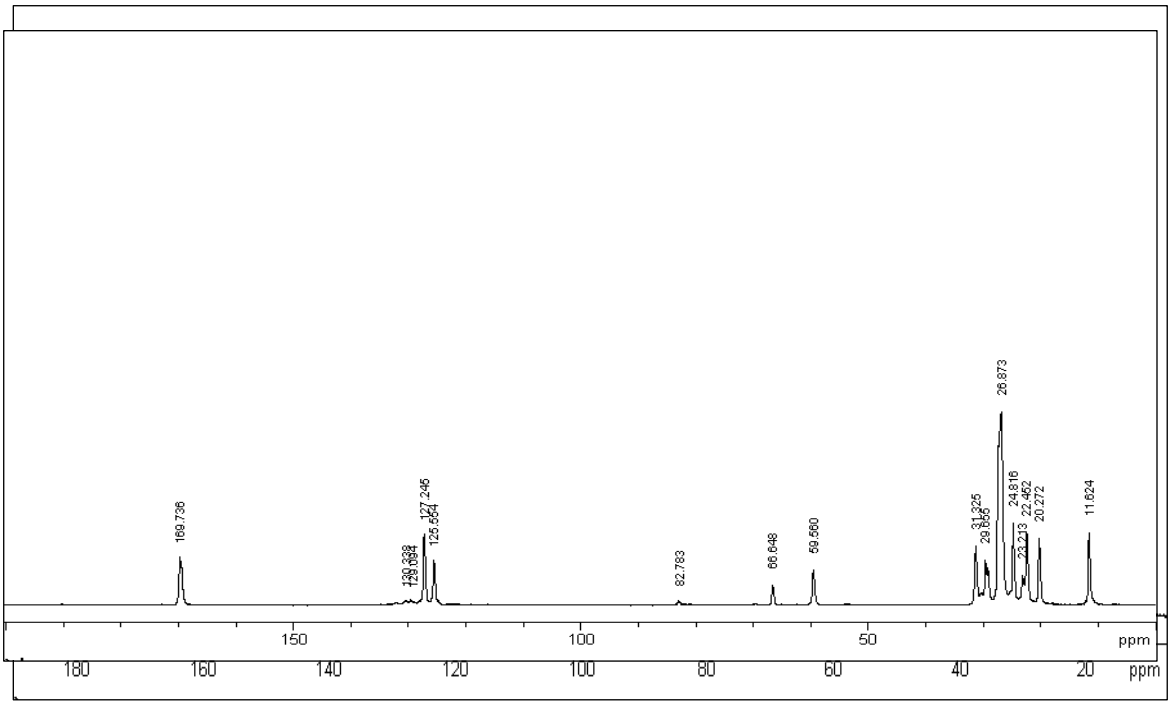

Fig. 4.2.1.1 WALTZ-16 decoupled ^{13}C Liquid-State NMR of a soybean flour gel sample with 38.7% protein content. Spectrum recorded with 10,000 transients on a Varian UI600 NMR spectrometer, in a 14.1 T external magnetic field**.**

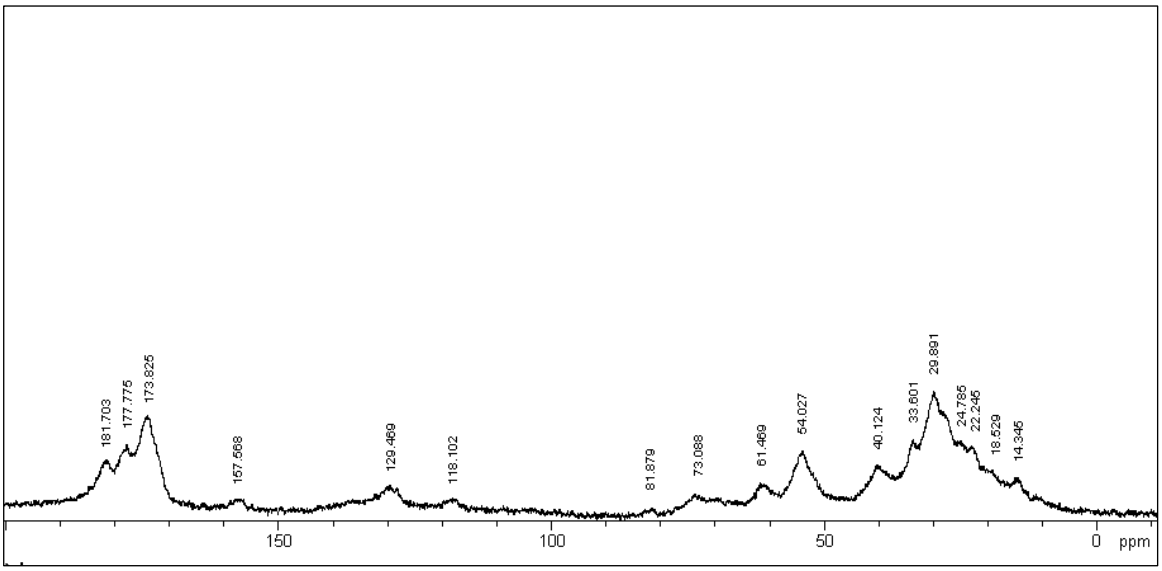

Fig. 4.2.1.2. WALTZ-16 decoupled ^{13}C Liquid-State NMR of a Soybean Protein Isolate gel sample with 50.8% protein content. Spectrum recorded with 5,000 transients, on a Varian UI600 NMR spectrometer, in a 14.1 T external magnetic field.

Fig. 4.2.1.3. WALTZ-16 ^{13}C Liquid-State NMR of a soybean oil sample. The spectrum was recorded with 112 transients on a Varian UI-600 NMR spectrometer, in a 14.1 T external field.

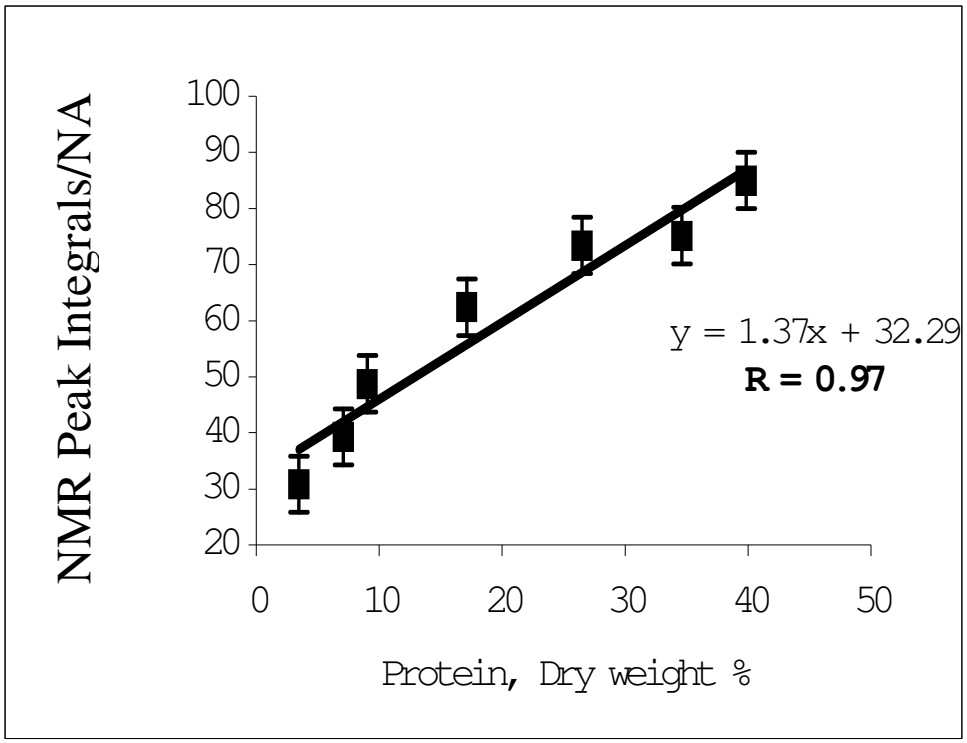

Fig. 4.2.1.4. Standard Calibration Plot for Soybean Protein Isolates.

2.3.2 Amino Acid Contents of Soybeans Determined by ^{13}C Liquid-State, High-Resolution NMR, Ion Exchange Chromatography and HPLC. Correlations between Amino Acid and Soybean Protein Contents

Amino acid profiles of a set of 100 standard soybean seed samples were determined by high performance liquid chromatography of Amino Acids. The NIR values for protein and oil of this standard sample set are highly correlated, as shown in Fig. 4.2.2.1. Furthermore, we found that the amino acid contents of this standard soybean seeds set, calculated on a dry basis as a percent of the total soybean seed sample weight (D %, AA) are highly correlated with the soybean protein content calculated on a dry basis from the total sample weight (as shown in the following Figures 4.2.2.2 to 4.2.2.5). Therefore**,** our data presented here in Figures 4.2.2.2 to 4.2.2.5 imply that NIR calibrations for amino acid contents of soybeans may encounter difficulties with most standard sets because of the close correlation (>85%) between the amino acid residue and the protein content of soybeans. Similar, high correlations were also found for other amino acid residues of acid hydrolyzed soy proteins: Asx, His, Ala, Ile, Cys and Met.

Table 4.2.2.1 presents a comparison between our amino acid composition analyses by ^{13}C Liquid-State, High-Resolution NMR of unhydrolyzed soybean seed samples and the corresponding data obtained by ion exchange chromatography of extracted soybean proteins after acid hydrolysis. As shown in a previous report (Augustine and Baianu, 1984), there is very

good agreement between the amino acid analysis of acid hydrolyzed, extracted soybean protein samples by ^{13}C Liquid-State, High-Resolution NMR and the corresponding data obtained by ion exchange chromatography of the same extracted soybean protein samples after acid hydrolysis. Therefore, the remaining differences between the results obtained by the two different approaches are most likely to be caused by acid hydrolysis. Furthermore, our NMR results include data for Trp, Gln, Glu, Asn and Asp in unhydrolyzed soybean proteins that cannot be obtained for acid hydrolyzed protein samples.

Table 4.2.2.1. Comparison between Amino Acid Contents of Soybean Proteins in Soybean Seeds Determined by ^{13}C Liquid-State, High-Resolution NMR and Ion Exchange Chromatography (IEC).

Wt % Total :	**Ala**	**Val**	**Leu**	**Ileu**	**Gly**	**Asn**	**Asp**	**Asx**	**Gln**	**Glu**	**Glx**
NMR	5	5	7	4.5	4	7	5	12	11	8	19
IEC	4	5.2	7.3	4.7	2.8	ND	ND	12.1	ND	ND	21.3

	Ser	**Thr**	**Arg**	**Lys**	**Trp**	**Tyr**	**His**	**Phe**	**Cys**	**Met**
NMR	5	4	8	7	1	3	3	6	1.5	1.0
IEC	4.6	3.6	9.5	7.8	ND	3.5	2.6	5.5	ND	0.9

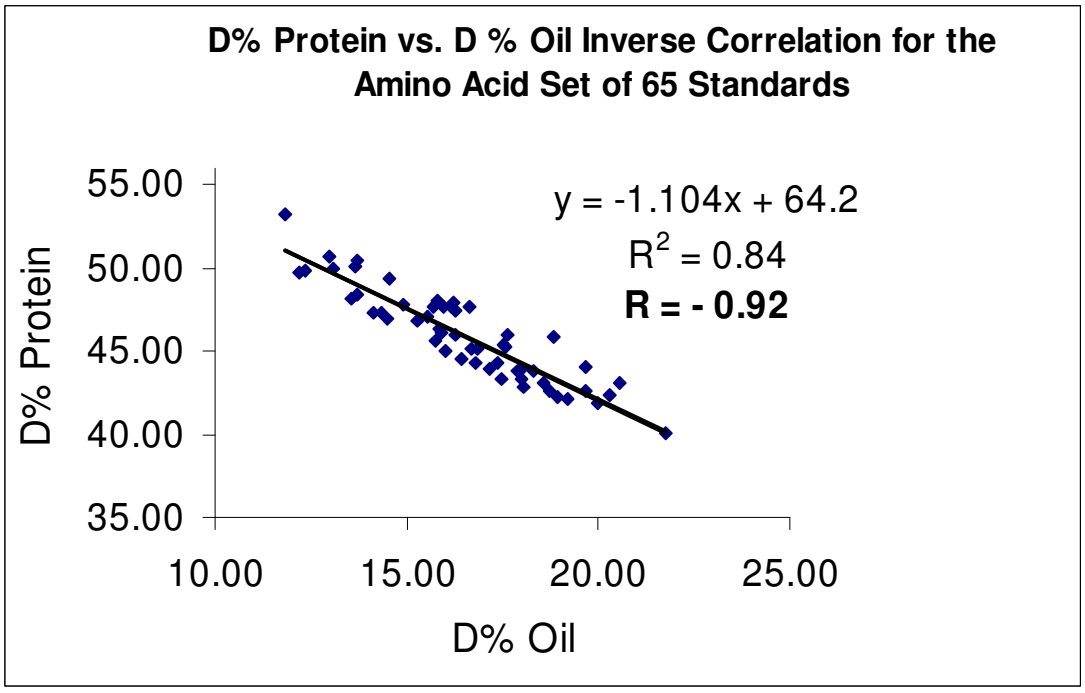

Fig. 4.2.2.1. Inverse Correlation between Protein (Dry weight %) vs. Oil (Dry weight %) for the Amino Acid Set of 65 Standards selected for the NIRS AA-Calibration (in addition to the remaining 35 standards employed for independent validation).

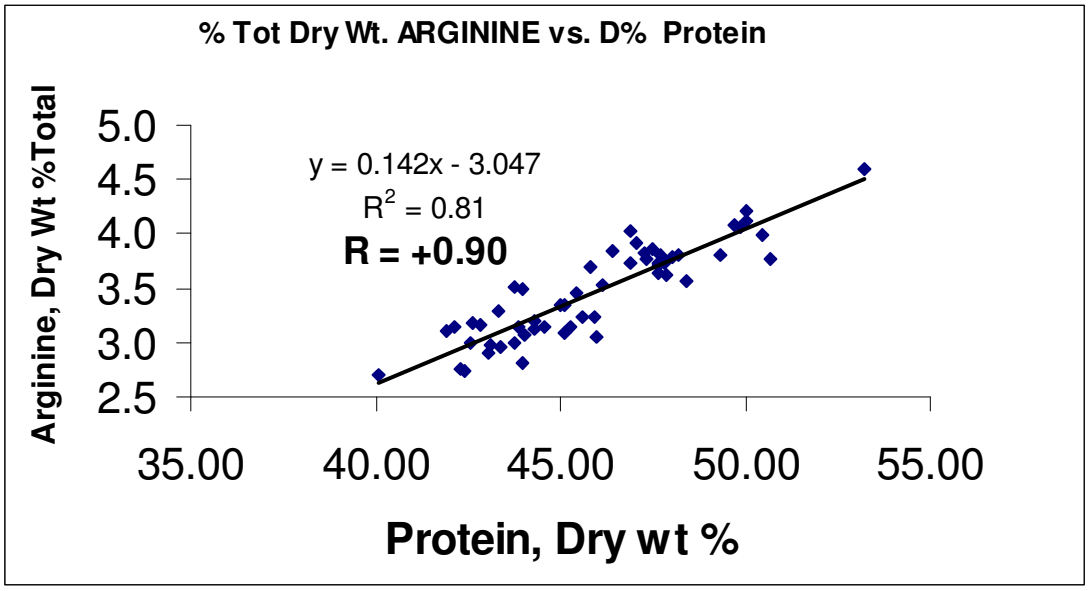

Fig. 4.2.2.2. Arginine as % Tot Dry Wt. vs. D% Protein. Note the very high degree of correlation between the Arginine content (as % of Total dry weight) and the Protein (as % of Total Dry weight) for the 65 amino acid standards measured.

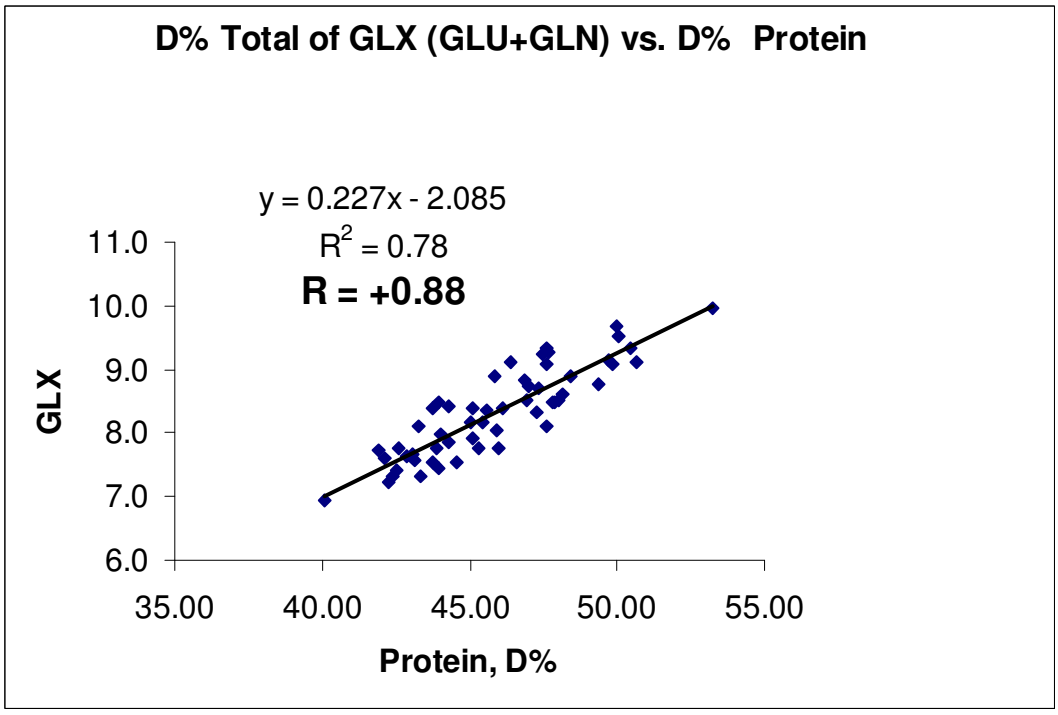

Fig. 4.2.2.3. Glutamine plus Glutamic Acid (Glx), as % Tot Dry Wt. vs. D% Soybean Protein. Note the very high degree of correlation between the Glx content (as % of Total dry weight) and the Protein (as % of Total Dry weight) for the 65 amino acid standards measured.

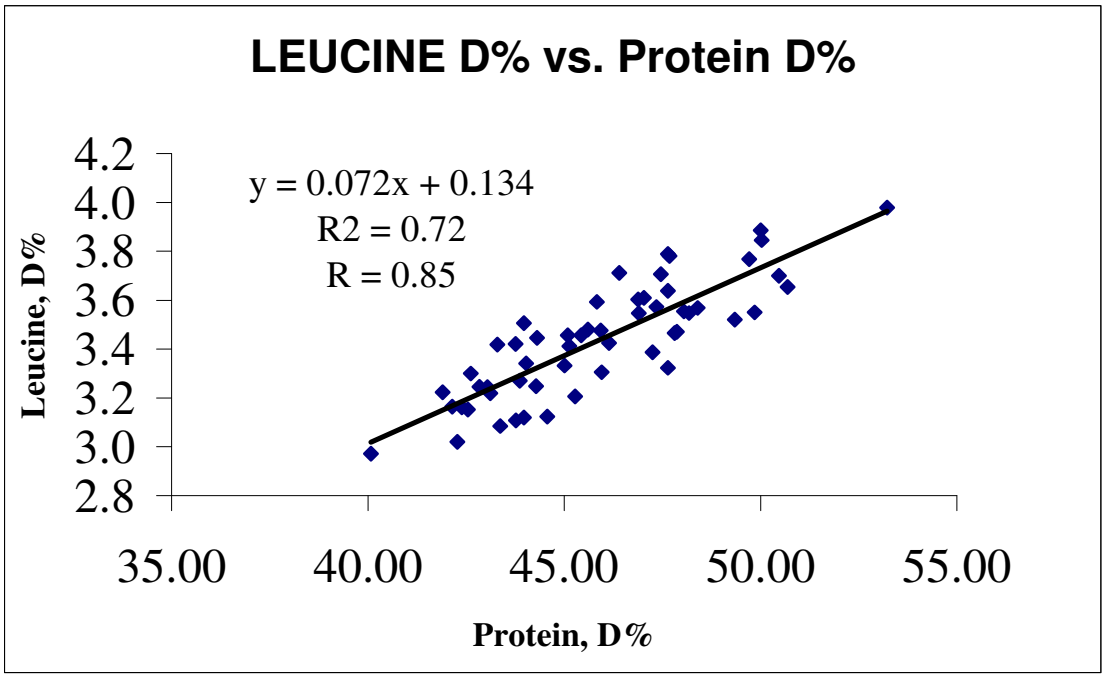

Fig. 4.2.2.4. Leucine as % Tot. Dry Wt. vs. D% Protein. Note the very high degree of correlation between the Leucine content (as % of Total dry weight) and the Protein (as % of Total Dry weight) for the 65 amino acid standards measured.

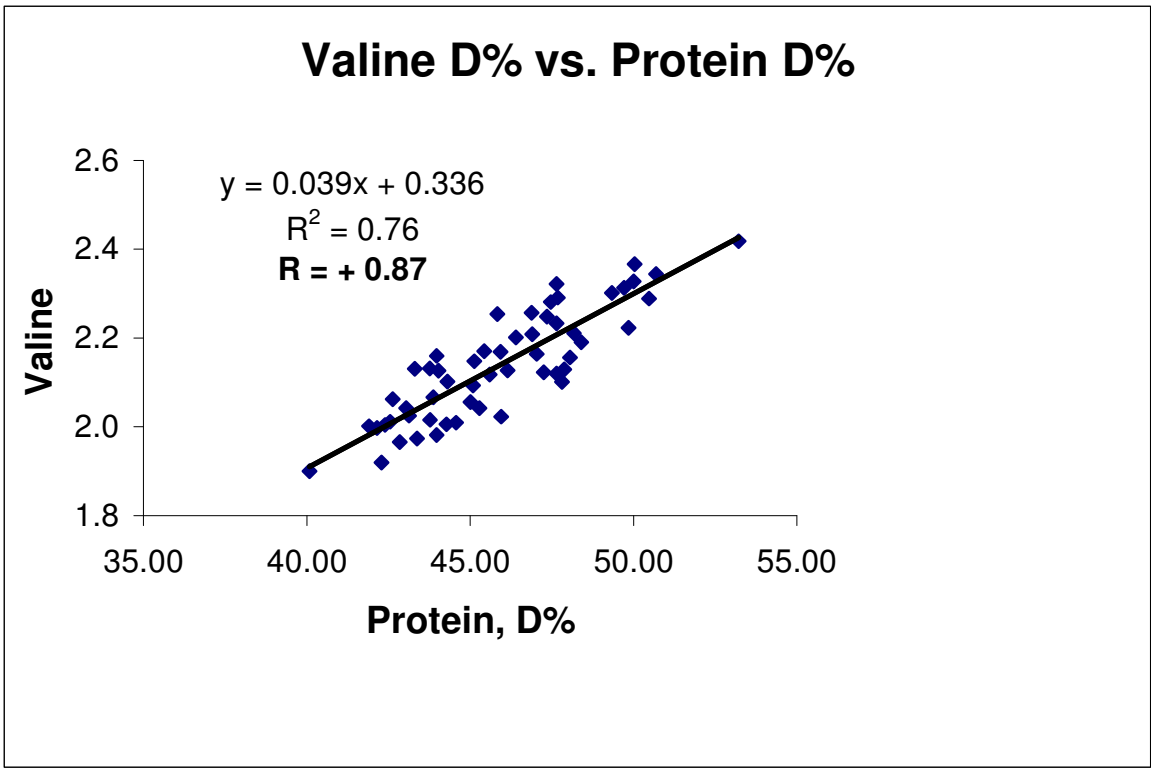

Fig. 4.2.2.5. Valine as % Total Dry Weight vs. D% Protein. Note the very high degree of correlation between the Valine content (as % of Total dry weight) and the Protein (as % of Total Dry weight) for the 65 amino acid standards measured.

Furthermore, our data indicates that the range of amino acid content variation among different soybean lines is relatively narrow, thus limiting the scope that breeders may have for selecting soybean lines with improved amino acid profiles. This interesting aspect requires further experimental studies, as well as careful analysis of the NIR data for the amino acid standard sets (e.g., deconvolution of the soybean protein spectra and their comparison with the deconvoluted NIR spectra of the individual amino acids.)

4.3 Oil Determination in Soybeans with the 1PULSE HR-NMR method.

Because the oil in plant seeds is in a liquid-like form, the oil protons are highly mobile and are detected by the 1PULSE NMR sequence. On the other hand, the much less mobile protons of carbohydrates and proteins in the solid matrix of the soybean seeds remain undetected as their FID signals decay very rapidly (within <30 μs), and are therefore within the dead-time of high-resolution NMR probes for liquids (Rutar *et al.*, 1989). The NMR signal amplitude is proportional to the mass of a sample (Abragam, 1961), and the intensity of a peak is simply the integral value of the area under the corresponding HR-NMR absorption peak. Experimental conditions and methodology were as described in Section 3, and references cited therein.
Fig. 4.3.1 illustrates the proportional increase of the NMR peak height with increasing quantity of oil for soybean seed standard samples. The standard, linear plot shown was first obtained for pure soybean oil standards. The slope of the standard regression line, expressed from the

intensity of the NMR peak as a linear function of the corresponding amount of oil in the standards, was then employed for the calculation of the quantities of oil for unknown soybean samples. Unknown sample oil contents were predicted by direct comparison of the measured NMR peaks for oil with the regression line in the oil standard plot (Fig. 4.3.2). Similar oil measurements were carried out previously for rapeseed or canola seeds, without oil extraction from the seeds.

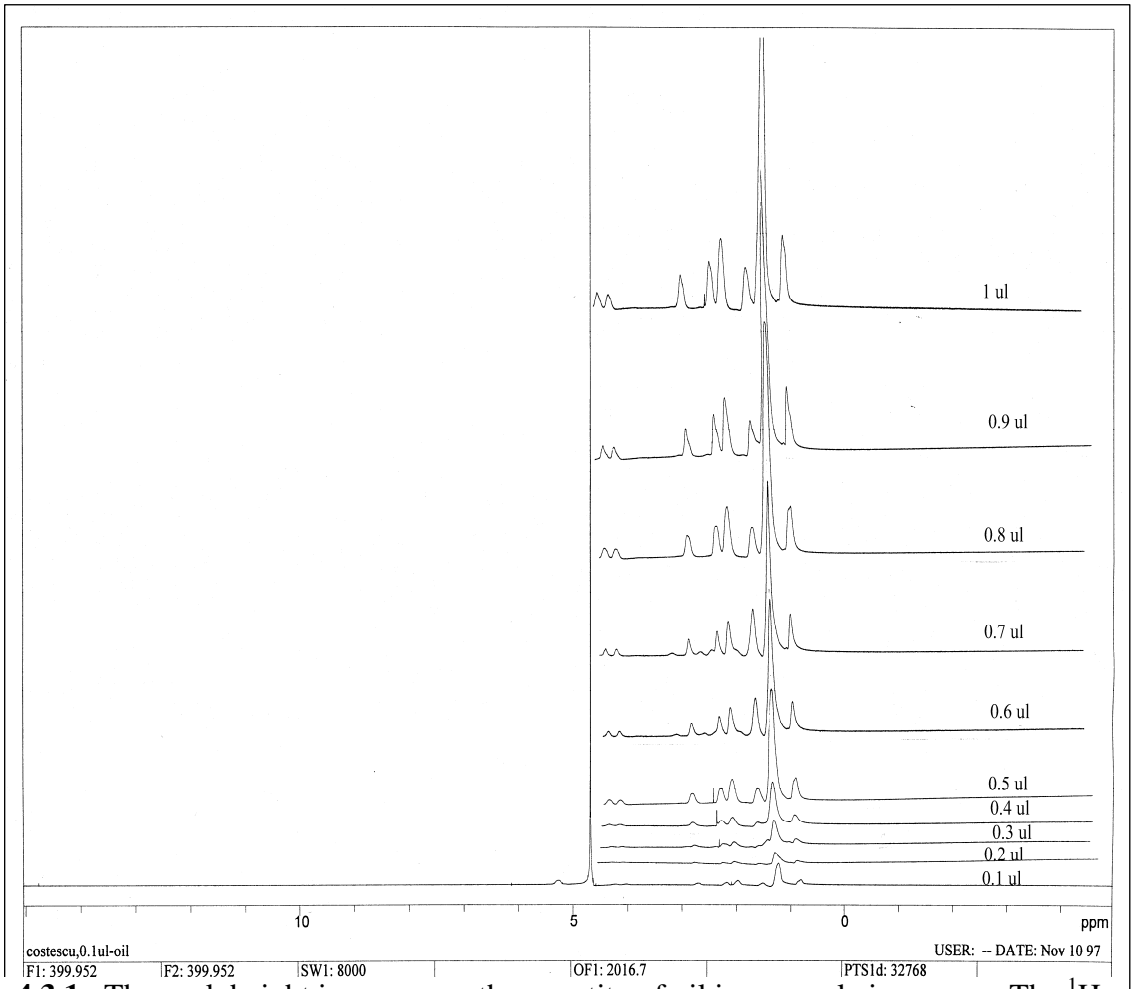

Fig. 4.3.1. The peak height increases as the quantity of oil in a sample increases. The ^{1}H spectrum was taken with an NMR spectrometer Varian U-400 and a Nalorac 5 mm ^{1}H NMR Quad probe, in an external magnetic field of 9.4 T, at a resonance frequency of 400 MHz.

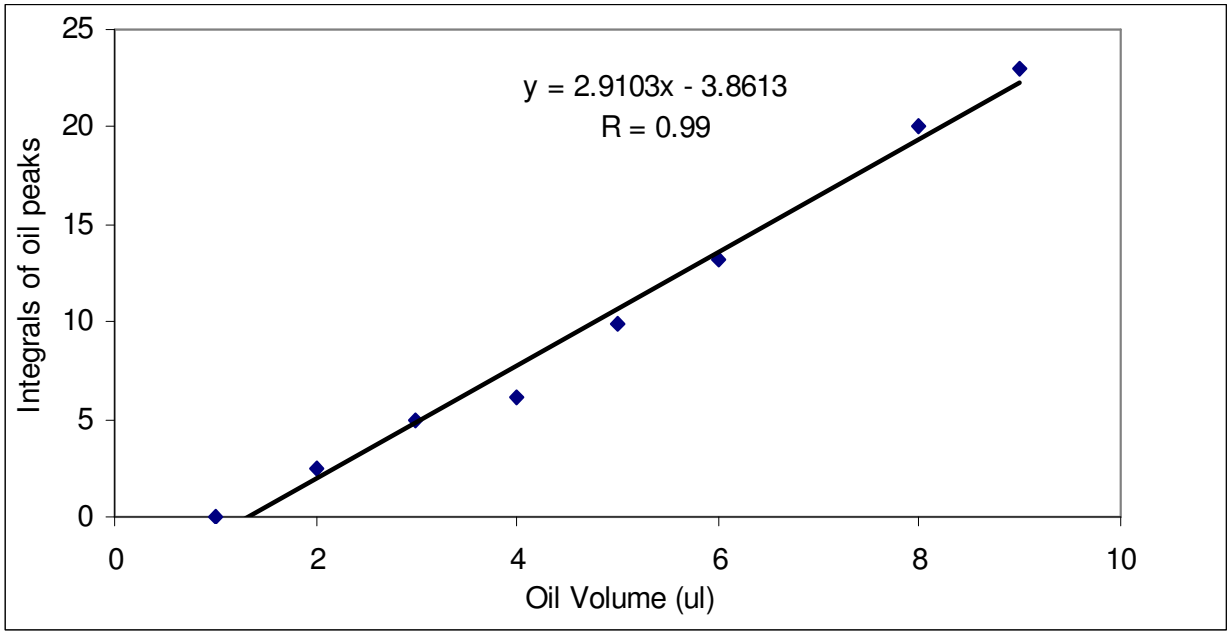

Fig. 4.3.2. The soybean oil standard plot for 400 MHz ^{1}H NMR measurements on the Varian U-400. The probe was a Nalorac 5 mm QUAD for high-resolution ^{1}H NMR.

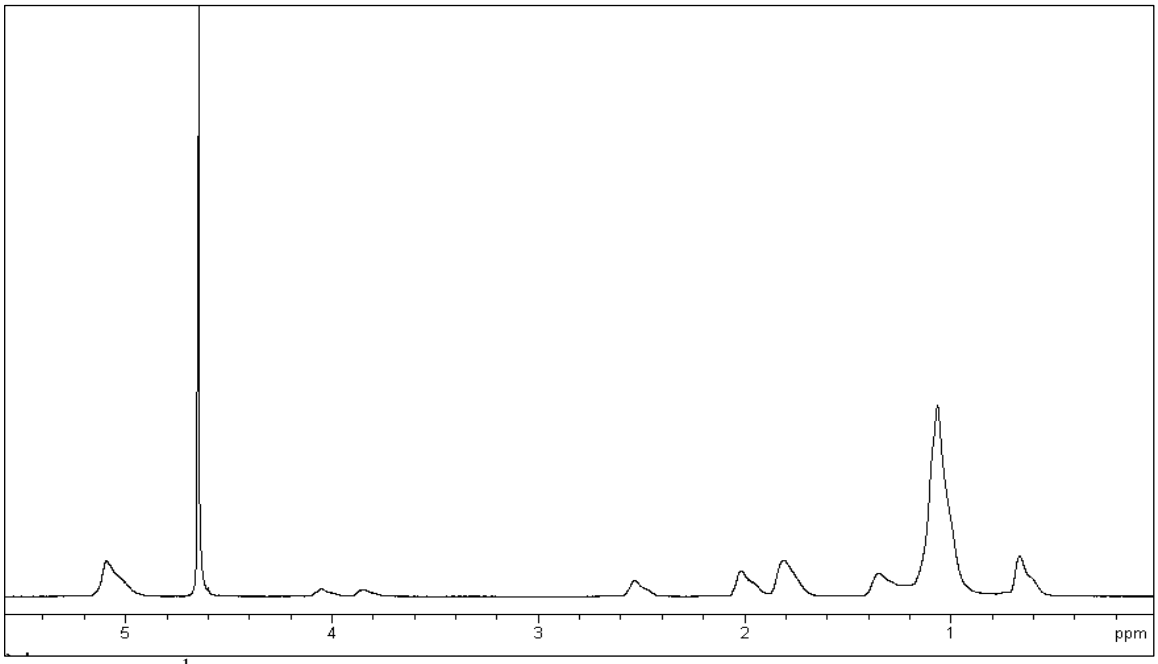

Fig. 4.3.3. The ^{1}H NMR oil peaks in a 0.9 µL soybean oil sample are in the following regions: 0 to 3 ppm, ~4 ppm and ~5 ppm. The ^{1}H NMR peak of water is at 4.67 ppm. The ^{1}H NMR spectra was acquired by averaging 500 transients with a Varian U-400 NMR spectrometer and a Nalorac 5 mm Quad probe tuned to a resonance ^{1}H center frequency of 400 MHz in an external magnetic field of 9.4 T.

4.4 Oil Determination in Soybean Flour with the 1 PDNA NMR Pulse Sequence

The 1PULSE witd Decoupler turned on During Acquisition (1PDNA) pulse sequence was employed for the carbonyl group detection of oil in soy flours. The lattice relaxation time for oil, $T_{1,}$ measured with the inverse recovery method, and was found to be T_1= 0.52 s. Therefore, the delay time, or the inter-pulse time interval, should be selected as $d_5 \geq 5T_1$, or ~2.5 s, and thus ensure that all FID signals are reproducible. However, because the FID is acquired for an interval time T_2*, which is much shorter than the actual T_1, the acquisition delay can be reduced in practice to ~3 T_2*, or even less if the selected pulse width was smaller than the 90^0 pulse. Therefore, in our oil determination measurements by NMR, d_5 was selected as 2 s, and the number of accumulations was 400, for a duration of the 90^0 rf pulse of 5.5 μs.

The chemical shift values were obtained by comparison to the glycine carbonyl chemical shift value of 176.03 ppm relative to TMS, as shown in Fig. 4.4.1. ^{13}C NMR spectra of soybean flours and soybean oil recorded with the 1PDNA experiment with 400 transients are presented in Figs. 4.4.2. and 4.4.3. Based on our published assignments (Baianu et al.,1993), the peaks in the spectral region ~130 ppm were assigned to the ethylenic carbons of the fatty acid signals.

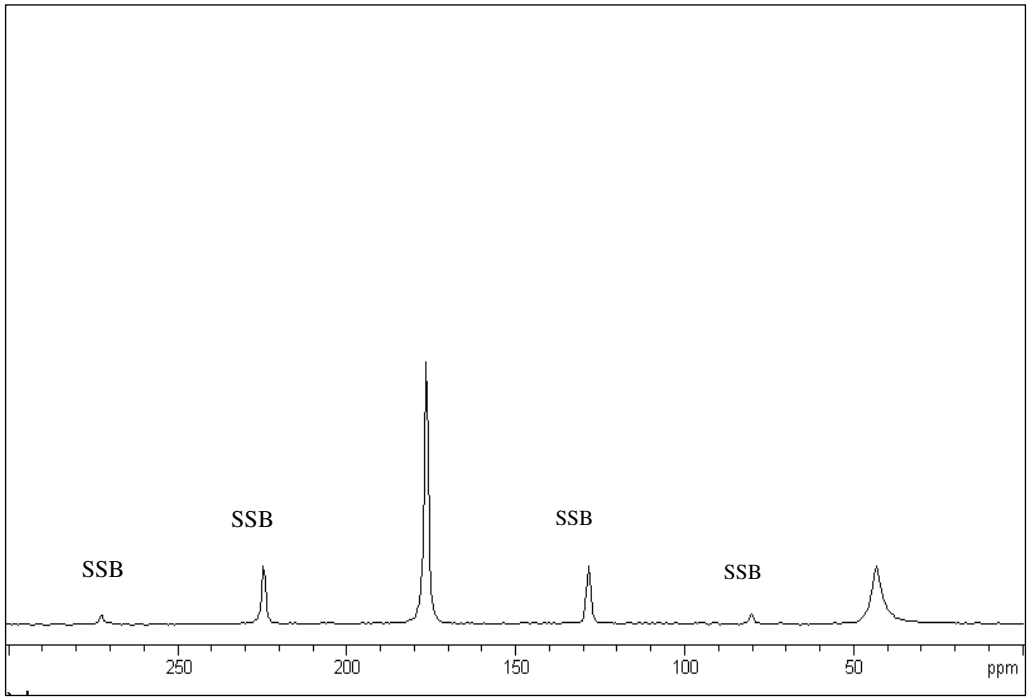

Fig. 4.4.1. Solid Glycine powder spectra with the carbonyl peak at 176.03 ppm relative to TMS. (SSB = spinning sidebands).

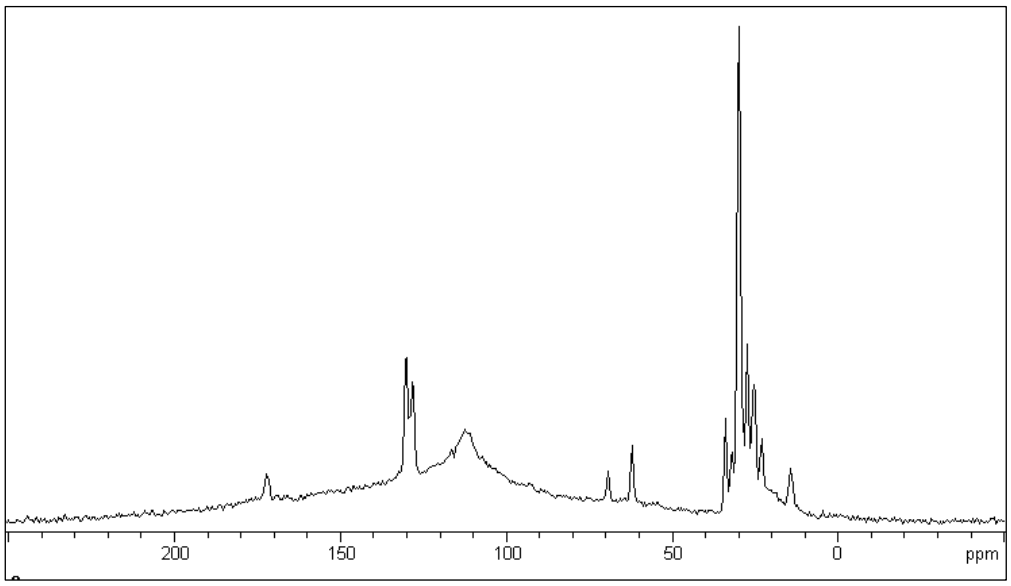

Fig 4.4.2. 1PDNA ^{13}C SS-NMR of soybean flour for a standard sample with 20% oil content (~300 mg total weight). The soybean oil signal of interest exhibits several resolved peaks close to 130 ppm.

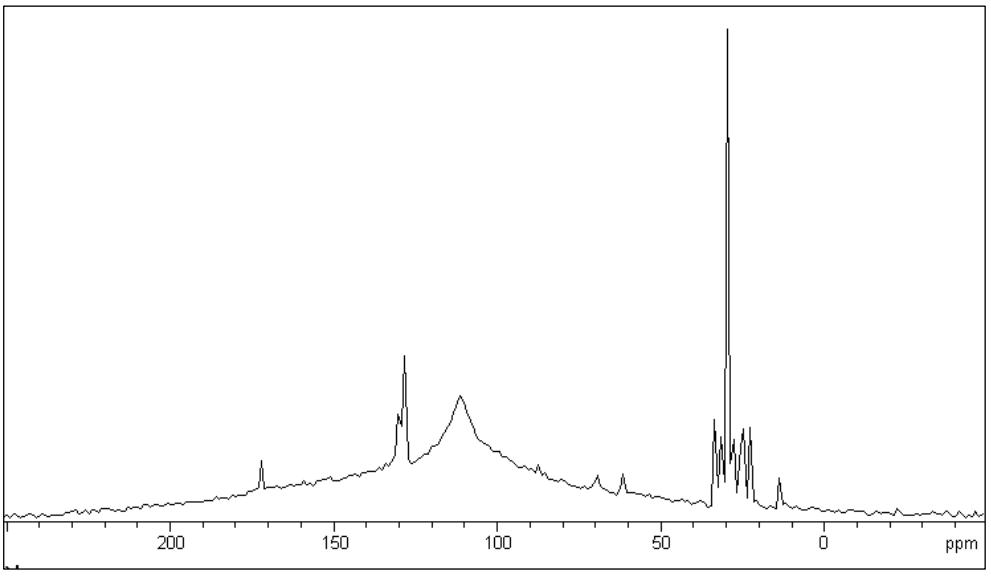

Fig. 4.4.3. 1PDNA ^{13}C SS-NMR of a standard sample of 53.5 mg of soybean oil in Al_2O_3.

The 1PDNA sequence pulse can be employed to record the ^{13}C NMR spectra of compounds in their liquid phase (or in solutions) as it is the case with oil in soybeans. Since all oil soybean components are tumbling fast relative to the ^{13}C NMR resonance frequency employed in our measurements, the dipolar interactions were reduced and their NMR signals gave sharp, well-resolved peaks. Thus, the measured diffusion constant for oil in immature soybean embryos was in the range of 10^{-5} cm^2s^{-1}. All ^{13}C NMR spectra were baseline corrected prior to integration. The quantity of oil in a sample was assumed to be given by the value of the integral of the peaks at ~132 ppm. The standard oil curve was obtained with the integral values of the ^{13}C NMR peaks at ~132 ppm (which are assigned to the ethylenic carbons of fatty acids). The results of our ^{13}C NMR measurements were in very good agreement with our NIR results obtained for the same soybean samples, as shown in Table 4.4.1 and Fig. 4.4.4.

Table 4.4.1. Example of 1PDNA ^{13}C SS-NMR Measurements of Oil Content for Soybean Flours, and their direct, linear correlation with the corresponding NIR data.

Soybean Seed ID	M2123	W2101	W2403	M309	M285	96-960A2-2687	W1228	96-959A6-1447	LG00-13523	LG00-13251
% oil by NIR	24.7	23.9	22.1	20.4	21.6	19.8	18.2	16.4	14.2	13.3
%oil by NMR	24.9	24.0	22.7	21.0	22.0	19.0	19.0	16.9	14.3	13.9

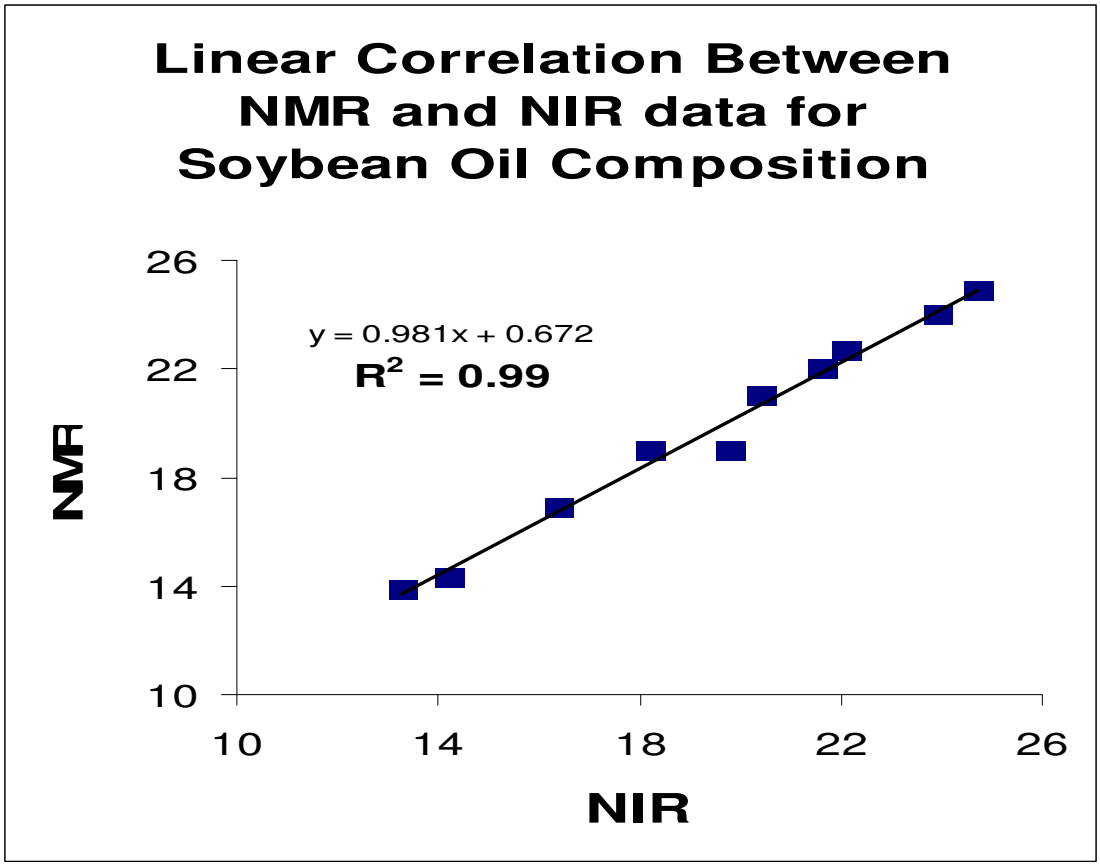

Fig. 4.4.4. 1PDNA ^{13}C SSNMR of Oil Content of Soybean Flours and their direct linear correlation with the corresponding NIR data.

4.5 Oil Determination in Soybean Flour by VACP ^{13}C SS-NMR

The protein content of soybean flours was determined from the standard plot for Soy Protein Isolates (SPI), (SPI source: 95% protein content, Archer Daniels Midland Company, IL, USA). NMR measurements of soy protein content are in agreement with the corresponding NIR results for the same samples, as shown in Table 4.5.1 and Fig. 4.5.2

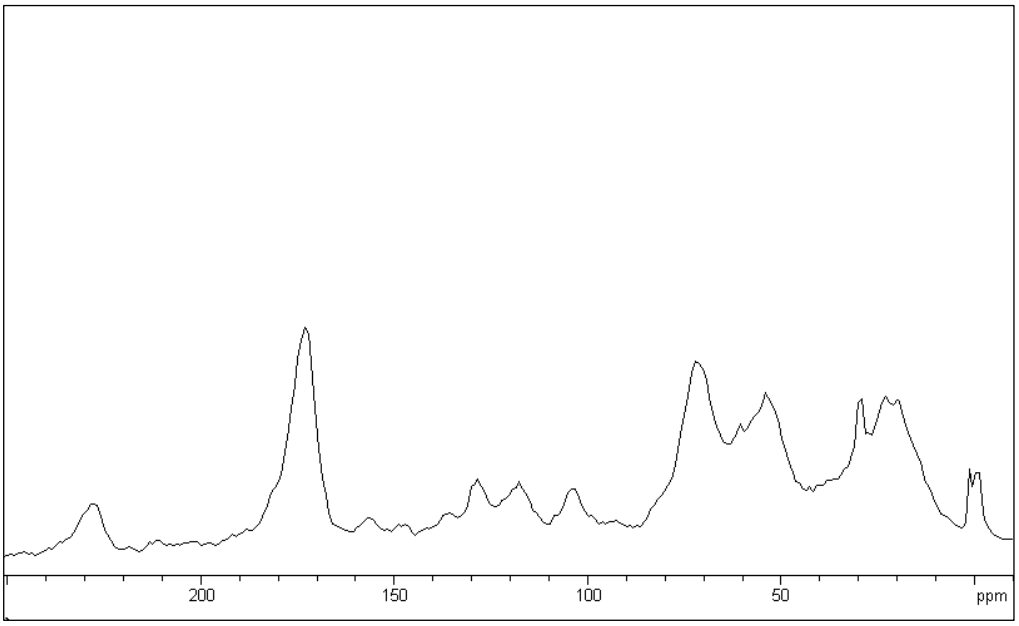

Fig. 4.5.1. VACP ^{13}C SSNMR of a Soybean Flour sample with 41% protein content (300 mg total weight). The component of interest has peaks in the spectral region around 174 ppm.

Table 4.5.1. VACP ^{13}C SS-NMR Measurements of Protein Content of Soybean Flours and Their Direct Correlation with the NIR Data (as shown in Section 2.6).

Soybean Seed ID	M2123	W2101	W2403	M309	M285	96-960A2-2687	W1228	96-959A6-1447	LG00-13523	LG00-13251
% protein by NIR	37.0	39.1	40.8	42.7	43.8	46.2	49.6	52.6	55.0	56.7
% protein by NMR	37.1	40.0	41.5	42.0	44.0	47.0	49.0	53.1	55.0	58.0

37.0

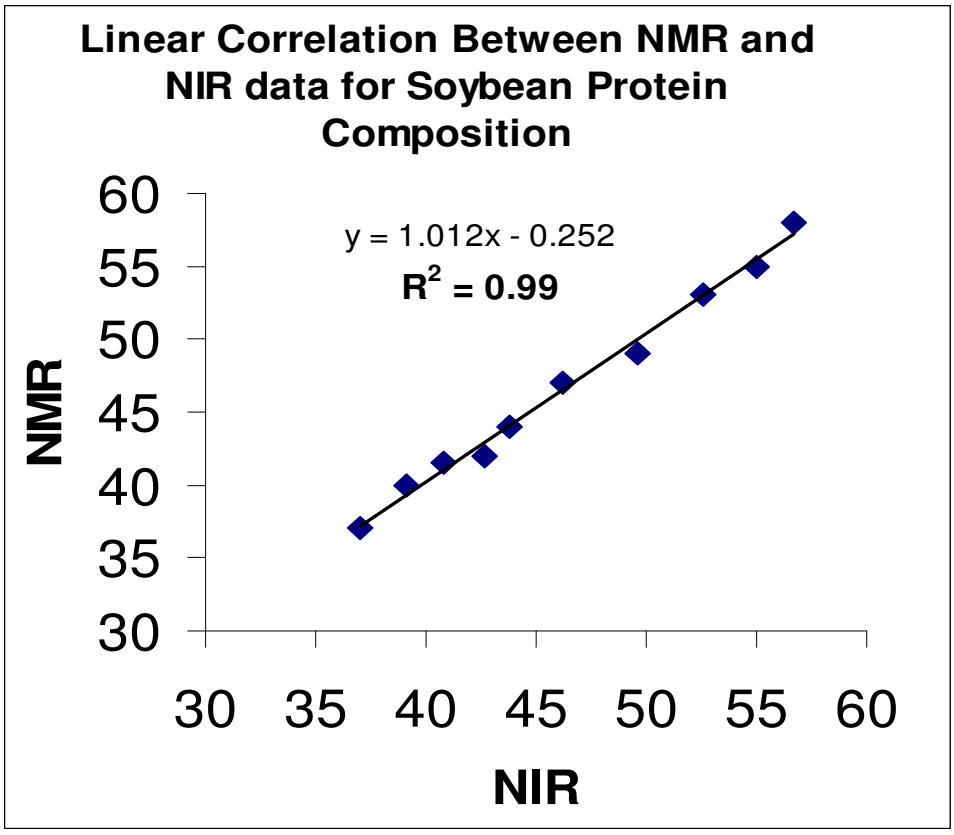

Fig. 4.5.2. VACP ^{13}C SS-NMR Measurements of Protein Content in Soybean Flours and their direct, Linear Correlation with the Corresponding NIR Data.

The 99% linear correlation between the ^{13}C SS-NMR and the corresponding Dual DA-NIRS oil and protein measurements on the same samples of well-defined soybean accessions from the USDA Soybean Germplasm Collection at UIUC, suggests that both techniques are suitable for the non-destructive, practical determination of both oil and protein content of soybean flours.

5. Limitations and Advantages of the techniques

FT-NIR instruments offer relatively high sensitivity, higher spectral resolution and considerably shorter spectral acquisition time in comparison with either filter-based or dispersive instruments that employ moving gratings.

Moisture determination errors can significantly affect the accuracy of any calibration in the NIR region because all the remaining components are specified on a wet basis in the chemometric programs that utilize, for example, the PLS1 algorithm. Furthermore, FT-NIR measurements on powdered soybeans take longer than whole seed analysis, and moisture calibrations are significantly less accurate for soybean powders in comparison with whole seeds because of the rapid moisture changes that can, and do, occur in soybean powders. Therefore, any inaccuracy in moisture determinations for the standard samples will significantly affect the predicted values for the other soybean seed/ oil seed components. The effect is even greater when smaller concentrations need to be predicted such as, for example, in the case of isoflavone or fatty acid calibrations. Errors made in moisture control of seed samples during transfers between measurement sites (i.e., between reference/analytical wet chemistry and routine NIR laboratories) have often been the cause of inaccurate and unreliable NIR calibrations that were not externally pre-validated. This is also one of the possible reasons why NIR methods for protein, oil and moisture are not yet recommended by AOCS for oil seeds.

In comparison with other methods of composition analysis, FT-NIR reflectance has not only the advantage of being convenient (with little or no sample preparation required), but it is also high speed, low cost per sample analysis and highly reproducible when calibrated correctly. These are indeed considerable advantages to be weighted against its minor disadvantages.

Last but not least, the soybean coat, and especially the darker coat colors, have been found to have a major influence on the NIR scattering and reflection properties of the soybean seed that does limit the applications of NIR for the composition analysis of dark color coat soybeans such as black, brown or green.

On the other hand, HR-NMR techniques for both liquids and solids have superior spectral resolution in comparison with FT-NIR but their sensitivity is lower than that of FT-NIR; therefore, composition analysis by HR-NMR are slower and more costly than by FT-NIR.

6. OTHER APPLICATIONS:

Selection of Soybeans Standards for Near Infrared Calibrations of Fatty Acid Composition

Unlike the resolved peaks of fatty acids in mid-IR spectra, the NIR bands are much less resolved for fatty acids thus making calibrations more difficult by NIR. NIR determinations of fatty acid contents in oil seeds have been previously reported, but the robustness of such calibrations, as well as the validation problems encountered determined suggested that the work reported should be considered as preliminary screening (Velasco, 1999; Sato, 2002).

We have therefore decided to develop new calibration procedures that would select reliable calibration standard sets so as to minimize potential validation problems that were previously encountered (Pazdernik *et al.,* 1997). Two sets of 66 soybean standards were selected for fatty acid calibrations, based on the wide range of expected fatty acid values suggested by previous analytical measurements of fatty acid contents for the same soybean lines from the collection of soybean lines at the Soybean Research Laboratory at the University of Illinois at Urbana. These selected lines were sealed in airtight containers and stored under constant temperature. In order to develop a reliable FT-NIR calibration there are several requirements that should be met. First, all factors affecting FT-NIR spectra must be represented in the calibration set. Such factors include physical and chemical characteristics of the sample, methods of sample preservation and processing, as well as instrument and sample environment (Windham, et al., 1989). Secondly, since the fatty acid composition is reported as a percentage of total oil content, the latter needs to be verified on site in order to eliminate potential errors that may occur at transfers between two different sites, for example. Furthermore, since the developments of the NIR calibration models have to be carried out with contents expressed on a wet basis it is very important to re-check the moisture contents of all sample standards that are employed for the calibration, and correct for any moisture changes that are likely to occur between different measurement sites.

Results

Standard gas chromatography (GC) methods (AOCS, Ca 5b-71) were employed to estimate the reference fatty acid content of the calibration set. The fatty acids analyzed were: 16:0, 18:0, 18:1, 18:2 and 18:3 and the results were expressed as percentage of total oil and are presented in Table 6.1 (measurements were carried out at the USDA Peoria Laboratory).

Table 6.1. **Range of constituents for 66 selected soybean samples selected as standards for Fatty Acid NIR Calibrations (data is courtesy of the USDA Peoria Laboratory). (Soybean lines are identified as: Stoneville 1999, MG V–VIII. USDA Germplasm, National Research Center, Urbana, IL.)**

Simple Statistics	Dry Protein	Dry Oil	Moisture	%16:0	% 18:0	% 18:1	% 18:2	% 18:3
Mean	47.61	17.11	5.86	11.91	3.24	21.19	55.42	8.24
Stdev	2.03	1.21	0.17	0.70	0.49	2.51	2.08	1.03
Max	53.74	19.32	6.19	14.10	4.35	27.63	60.74	10.93
Min	43.80	13.70	5.48	10.54	2.33	14.24	51.09	6.32

Our laboratory's NIT measurements with Zeltex transmission instruments of the total oil contents for the selected standards are compared in Fig. 6.1 with those provided by the Northern Region USDA Peoria Laboratory. One notes a very high degree of correlation (r = 0.98) between the dry oil values measured independently by the two laboratories. This suggests that potential problems with sample transfers between the two independent measurement sites have been thus avoided, and also that validation of the calibration results with the second set of validation sample standards processed with the same procedure is likely to avoid the validation problems reported previously (Velasco, 1999; Sato, 2002; Windham, *et al*., 1989). Furthermore, a statistical exploratory analysis of the fatty acids contents of the standards was performed using SPSS ® and SAS ® software to detect the normality of the data as well as the presence of outliers within the standards.

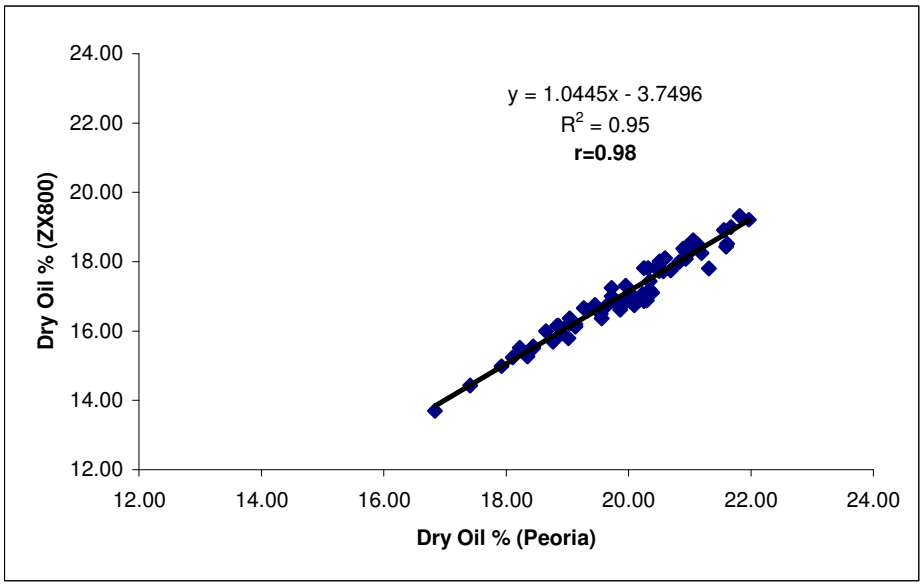

Fig. 6.1. Correlation between dry oil (USDA Peoria) vs. dry oil ZX800 of 66 fatty acids selected soybean standards.

The palmitic acid distribution of the calibration set was significantly normal (Shapiro-Wilks *P* = 0.05). This means that the sample size follows probability laws and that non extreme values were selected as standards. The palmitic distribution appeared to be leptokurtic, and also skewed towards high values, but it contained only mild outliers. In addition, because of the selection of stable lines, its coefficient of variation was smaller than 6 %.

The stearic acid distribution of the fatty acid calibration standards was also normal, with a 0.05 probability of type I error (p-value = 0.246). It was also leptokurtic as well as skewed towards the higher values, but with a coefficient of variation smaller than 15%.

In the oleic distribution, the central tendency statistics were extremely close and therefore the coefficient of variation was small (CV = 11.83%). The oleic and linoleic distributions of the calibration set were also normal, both with coefficients of variation less than 4%.

On the other hand, the linolenic acid distribution was not normal with a 0.05 probability of type I error (Shapiro-Wilks p-value= 0.011) It was skewed towards the highest values and leptokurtic, with its inter-quartile range being extremely narrow and its coefficient of variation equal to 12.49%.

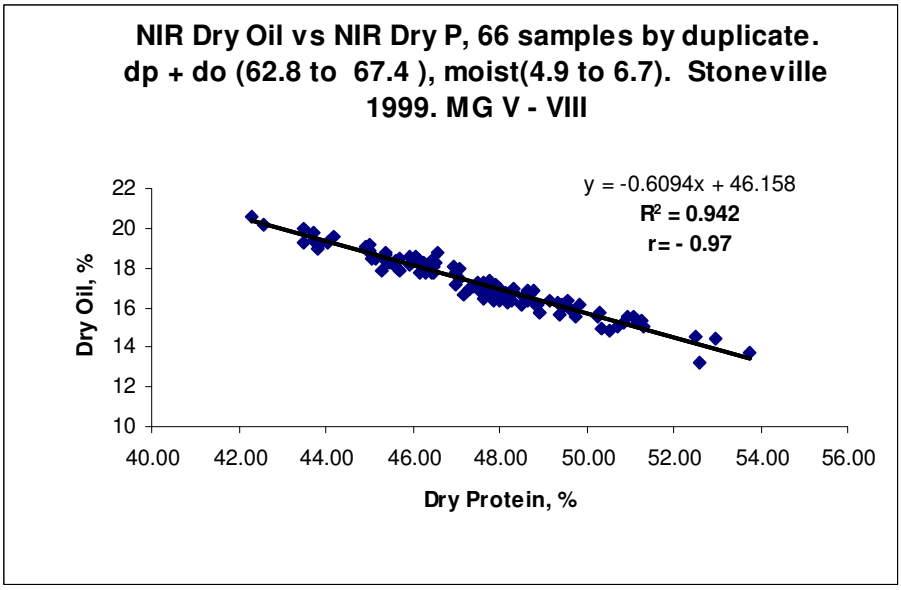

Fig. 6.2. Inverse oil--protein correlation of 66 fatty acids selected samples measured with NIT instruments.

Guo et al. (2002) suggested in a recent study of 5000 different soybean lines that there is a high degree of inverse correlation (-r>0.90) between protein and oil contents of seeds drawn from large sets of soybean lines. The high degree of inverse correlation present between the oil and protein contents of our selected standard set for fatty acid calibrations shown in Fig. 6.2 is consistent with these recent NIT and NIR studies, and it thus provides an independent, external validation for the protein values of our selected standard set of soybean seeds.

7. Practical Implications of our High-Resolution NIR and NMR Analyses of Soybean Seeds

A brief illustration of our NIR results for a large set of soybean samples is presented in Figure 7.1. A high degree of inverse correlation exists between the protein and oil contents for the 5,000 bulk soybean seed samples measured by NIR with the calibrations. This group of soybean seed samples was selected from a large set of experimental lines planted at six different locations over two years. The data presented in Figure 7.1 is therefore consistent with the robustness of our NIR calibrations for protein and oil, and also indicate that even after a small number of generations (~3) the degree of inverse correlation between protein and oil can be very high (>90%). This fact limits the soybean protein content increase that can be practically achieved for commercial soybean applications.

Furthermore, from a commercial viewpoint the high seed yield of soybeans appears to be accepted by the industry as being more important than high protein content. It has been generally found that the protein concentration and seed yield of soybeans are inversely correlated. However, in a few recent studies (Thompson and Nelson,1998; Kabelka *et al.*, 2002) it was suggested that the mean protein compositions of soybean populations could be increased to a certain extent, and without a significant loss of agronomic yield (Wilcox and Cavins, 1995). In the same study (Wilcox and Cavins, 1995), one soybean line "Pando" (498 g/kg protein) was backcrossed to another line "Cutler 71" (408 g/kg protein) to determine if the yield of Cutler 71 could be recovered in addition to the high protein from Pando. Random F4-derived lines, as well as three lines with highest protein concentration from the initial cross were evaluated for agronomic traits for 1 yr. Seeds were evaluated for protein and oil concentration using either NIR reflectance or NIR transmission (NIT). The parent line for each backcross was selected first for protein, and then for yield, similarly to Cutler 71. Random F4-derived progenies and the cultivars Pando, Cutler 71, and Hamilton were evaluated for 2 yrs. In each backcross generation, lines were identified with seed protein in excess of 470 g/kg and progressively approached the Cutler 71 yield. One line averaged 472 g/kg seed protein and was significantly ($p = 0.05$) higher in seed yield than Cutler 71, and similar in yield to the Hamilton cultivar. In each population, there were inverse relationships between yield and protein ($R^2 = 0.33$ to 0.06) and between seed protein and seed oil (R^2 0.55 to 0.94). In successive backcross populations, minimum oil values increased from 148 to 174 g/kg, indicating a trend toward recovering the oil concentration of Cutler 71 (204 g/kg). The data demonstrate that high seed protein can be backcrossed to a soybean cultivar, fully recovering the seed yield of the cultivar. *This suggests the absence of physiological barriers to combining high seed protein with high seed yield in these soybean populations.*

For the soybean lines investigated in a recent study (Kabelka et al., 2002), we have also found by NIT that the degree of protein-oil inverse correlation was as high as that of the soybean experimental lines represented in Fig.6.1. This is the case in spite of the fact that the mean protein content of the latter was substantially higher (by ~7%) than that of the former soybean group that was also characterized for agronomic yield by Kabelka et al. (2002).

Our novel NIR calibrations have been extensively and successfully tested with a wide range of different soybean lines and exotic germplasm soybean accessions for both accuracy and

robustness; therefore, our calibrations can be employed equally well for the rapid and reliable NIR analysis of soybean composition throughout the industrial soybean distribution chain, from harvesting to post-harvest processing. Such improved soybean lines often have lower oil content as a trade-off for the increased protein content, but still could have an acceptable seed yield for soybean lines with significantly increased protein content above the range of commercial soybean cultivars.

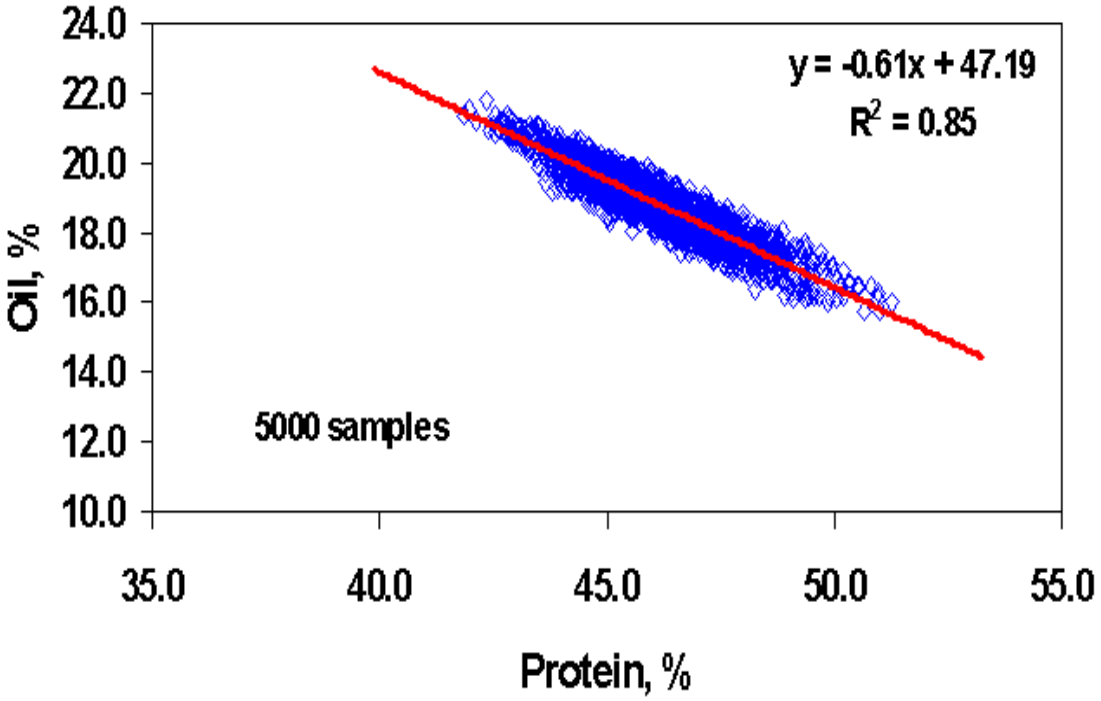

Fig. 7.1. Protein-Oil Inverse Correlation of 5,000 Soybean Samples of Experimental Lines at UIUC.

8. Conclusions and Discussion

Oil content determinations for whole soybeans seeds were carried out with either Diode Array or FT-NIR instruments; such determinations were based upon calibrations that utilized a PLS1 regression models and were extensively validated with a large number of soybean lines. Our NIR calibrations were undertaken in parallel with the higher resolution (but slower and more expensive) NMR measurements. These calibrations had very high correlation coefficients (R >0.99) between the NIR predicted values and the reference data. Both high-resolution NIR and NMR calibrations and methodologies were employed -- with HR-NMR employed to calibrate the NIR-- and, respectively, carry out a large number of protein and oil composition analyses of soybean seeds (~50,000; by NIR and NIT) for breeding and selection purposes over a period of three years. A wide range of soybean experimental lines and more than 2,000 exotic

soybean germplasm accessions were thus characterized accurately and reproducibly for selection and breeding programs at UIUC. Therefore, our results demonstrate the usefulness of this novel NIR approach for soybean selection and breeding purposes.

A high degree of inverse correlation was found between the protein and oil contents of 5,000 bulk soybean seed samples predicted by NIR with the new calibrations that we developed. This group of soybean seed samples was selected from an even larger set of experimental lines that were planted at six different locations over a time interval of three years. Improved soybean lines are often found to have lower oil content as a trade-off for the increased protein content, but still could have an acceptable seed yield for soybean lines with significantly increased protein content well-above the range of commercial soybean cultivars.

Our novel NIR calibrations may also be employed to develop procedures for the rapid and reliable analysis of soybean composition throughout the industrial soybean distribution chain, from harvesting to post-harvest processing.

9. REFERENCES

1. Abragam, A., *"The Principles of Nuclear Magnetism*", Clarendon Press, Oxford, 1961.
2. AOAC International.1995.Official Methods of Analysis, 16th ed., AOAC International, Gaithersburg, MD.
3. Anderson, R, Killam, A.S. and Orf, J.H. 2002. Amino Acid Analysis of Soybean Powders by NIR. Soy2002 Conference. (Abstr.)
4. Augustine, M. and Baianu, I.C.1984. High-resolution ^{13}C NMR and Ion Exchange Chromatography Amino Acid Profiling of Corn and Soybean Protein Hydrolizates. FASEB Proceed. 1125.
5. Augustine, M. and Baianu, I.C.1986. High Resolution Carbon-13 NMR Studies of Maize Proteins. J. Cereal Sci. 4: 371-378.
6. Baianu, I.C. and H. Förster.1980. Cross-polarization, high-field Carbon-13 NMR techniques for studying Physicochemical Properties of Wheat Grains and Wheat Proteins., J. Applied Biochemistry, 2 : 347-355.
7. Baianu, I.C.1981.Carbon-13 and Proton NMR studies of Wheat Proteins in Solution." J.Sci. of Food & Agriculture, 32 : 309-313.
8. Baianu, I.C. et al. 1982. High Resolution Carbon-13 and Nitrogen-15 NMR Studies of Flinor Wheat Gliadins: Spectral Assignments, Changes with concentration and heat treatments. J.Sci. of Food & Agriculture, 33 : 373-383.
9. Baianu, I.C. 1987. 125 MHz Carbon-13 NMR Spectra of Corn Zeins in Alkaline Solutions. Unpublished results.

10. Baianu, I.C. 1989. Ch.9 in *"NMR in Agriculture.",* P.Pfeffer and W. Gerasimowicz. Eds., Boca Raton, FL: CRC Press.

11. Baianu, I.C., et al.1990. Ch8**.** In *"NMR Applications in Biopolymers*". Edited by J.W. Finley et al., Plenum Press, New York.

12. Baianu, I.C. 1992. Ed. "*Physical Chemistry of Food Processes*. Vol. 1: *Fundamental Aspects*." Van Nostrand Reinhold. New York, NY.

13. Baianu, I.C., and Kumosinski, T.F. 1993. Ch9 in "*Physical Chemistry of Food Processes. Vol.: Advanced Techniques, Structures, and Applications*." Van Nostrand Reinhold. New York, NY.

14. Baianu, I.C., Costescu, D., You, T., 2002. Novel Techniques for Microspectroscopy and Chemical Imaging Analysis of Soybean Seeds and Embryos. Soy2002 Conference, Urbana.

15. Baianu, I. C., You, T., Guo, J., Nelson, R. L, 2002. Calibration of Dual Diode Array and Fourier Transform Near Infrared Reflectance Spectrometers for Composition Analysis of Single Soybean Seeds in Genetic Selection, Cross-Breeding Experiments. Soy2002 Conference, Urbana, IL.

16. Bardet, M., M.F. Foray, J. Bourguinon and P. Krajewski. Investigation of seeds with high-resolution solid-state ^{13}C NMR. Magnetic Resonance in Chemistry, 39:733-738, 2001.

17. Barton F.E., 2002, Spectroscopy Europe, 14(1) : 12-18.

18. Becker, E.D. 1980. High Resolution NMR. Theory and Chemical Applications, 2nd Ed. New York : Academic Press.

19. Ben-Gera, I., Norris, K.H. 1968. Isr. J. Agric. Res. 18:125-132.

20. Bloch, F. 1946. Nuclear Induction. Phys. Rev. 70:460-474.

21. Burlingame, A.L., R.K. Boyd, S.J. Gaskell, Analytical Chemistry, 66: 634R, 1994

22. Burton, J.W., and R.F. Wilson, C.A. Brim, and R.W. Rinne. Registration of soybean germplasm line with modified fatty acid composition of seed oil. Crop Science 29:1583; 1989

23. Burton, J.W., and R.F. Wilson.1994.Registration of N88-480, a soybean germplasm line with high concentration of oil in seeds. Crop Science 34: 313-314.

24. Carr, H. Y. And E. M. Purcell, Effects on diffusion on free precession in Nuclear Magnetic Resonance experiments, Physical Review, 94 (3), 1954.

25. Delwiche, S.R., 1995. Cereal Chem. 72: 11-16.

26. Derome, A.E., Modern NMR techniques for chemistry research, p. 129, Pergamon Press, New York, 1987.

27. Farrar, T.C. and E.D. Becker.1971. Pulsed and Fourier Transform NMR. New York: Academic Press.

28. Feynman, R.P., "*Lectures on Physics"* (Addison-Wesley Publishing Company, Reading, MA, 1963), Vol. 3.

29. Fraenkel-Conrat, H, Cooper, M., 1944. "The Use of Dyes for the Determination of Acid and Basic Groups in Proteins,"J. Biol. Chem., 154: 239-246 (1944).

30. Geladi, P., MacDougall, D., Martens, H. 1985. Linearization and Scatter-Correction for Near-Infrared Reflectance Spectra of Meat. Applied Spectroscopy. 39(3): 491-500.

31. Gornall, AG, CS Bardawill, and MM David. 1949.*J. Biol. Chem. 177:* 751.

32. Haaland, D.M. , Thomas, E.V. , 1988. *Analytical Chemistry*, 60, 1193.

33. Hammond, E.G., and W.R. Fehr. 1982. Progress in breeding for soybean oil with low linolenic acid. Journal of American Oil Chemist Society 59 (4): 304A.

34. Hammond, E.G., and 14. W.R. Fehr. 1983. Registration of A6 germplast line of soybean. Crop Science 23: 192-193;

35. Harris, Robin K., *"Nuclear Magnetic Resonance spectroscopy, A Physicochemical View",* Pitman, London, 1983.

36. Hofmann, E.N., Initiation and chemical mutagenesis of somatic embryogenic cultures of soybean (*Glycine max (*L.) *Merr*, M.S.Thesis, 1998. University of Illinois at Urbana-Champaign (UIUC).

37. Isaksson, T., Naes, T. 1988. The effect of Multiplicative Scatter Correction (MSC) and Linearity Improvement in NIR Spectroscopy. Applied Spectroscopy. 42(7): 1273-1284.

38. Kabelka, E. A. , Diers, B. W. , Fehr, W. R. , LeRoy, A. R. , Baianu, I. C., You, T. , Neece, D. J. , Nelson, R. L. 2003. Identification of putative yield enhancing quantitative trait loci from exotic soybean germplasm. (In Press)

39. Kakalis, L.T. and Baianu, I.C. 1989. A high-resolution ^{13}C NMR study of glycinin in D_2O solutions. J. of Agric. & Food Chem. 37:1222-1228.

40. Kakalis, F.T. and Baianu, I.C. 1990. High-resolution ^{13}C NMR study of the soybean 7S storage protein fraction in solution. J. of Agric. & Food Chem. 38:1314-1320.

41. Kortum, G. 1969. ed*:"Reflectance Spectroscopy--Principles, Methods, Applications."* Springer-Verlag, NewYork, NY.

42. Lamb, D., Hurburgh, C.R. 1991. Trans. ASAE. 34: 2123-2129.

43. Lazzeri, P.A., D.F. Hildebrand, and G.B. Collins. 1987. Soybean somatic embryogenesis: Effects of nutritional, physical and chemical factors. Plant Cell, Tissue and Organ Culturen,10: 209-220.

44. Lehninger, Albert L., *"Biochemistry: The Molecular Basis of Cell Structure and Chemistry"*, Worth Publishers, New York, 1978.

45. Levitt, Malcom H., *"Spin Dynamics: Basics of Nuclear Magnetic Resonance",* John Willey & Sons, New York, 2001.

46. Lowry, O. H., N.J. Rosebrough, A. L. Farr and R. J. Randall, 1951. Protein measurement with the Folin phenol reagent. Journal of Biological Chemistry 193: 265-275.

47. Maluszynski, M., B.S. Ahloowalia, and B. Sigurbjornsson. 1995.Application of in vivo and in vitro mutation techniques for crop improvement. Euphytica, 85: 303-315.

48. Mie, G. 1908. Ann. Physik. 25: 377.

49. Ohnishi, S.T. and Barr, J.K.1978. A simplified method of quantitating protein using the biuret and phenol reagents. *Anal. Biochem*. **86**, 193-200.

50. Orman, B.A., Schumann, R.A., 1992. JAOCS. 69: 1036-1038.

51. Pazdernik, D.L., Killam, A. S., Orf, J. H., 1997. Agronomy Journal. 89:679-685.

52. Raghavachari, R., ed. 2001. Near-Infrared Applications in Biotechnology, Marcel-Dekker, New York, NY.

53. Robertson, J. A. and F. E. Barton.1984.Oil and water analysis of sunflower seeds by Near Infrared Reflectance spectroscopy, JAOCS, 61 (3): 543-547.

54. Rutar, V. 1982. A new possibility for nondestructive protein content determination in viable seeds, Applied Spectroscopy, 36: 259.

55. Rutar, V. 1989. J. of Agric. And Food Chem. 37:67.

56. Rutar, V., M. Kovac and G. Lahajnar.1989. Nondestructive study of liquids in single seeds using Nuclear Magnetic Resonance and Magic Angle Sample Spinning, JAOCS, 66: 961-965.

57. Rutar, V.1989. Magic-Angle Spinning NMR spectroscopy of liquids as nondestructive method for studies of plant seeds, J. Agric. Food Chem., 37: 67-70.

58. Rutar, V. Ch.3. NMR Studies in Intact Seeds, in *"Nuclear Magnetic Resonance in Agriculture"* editors Philip E. Pfeffer and Walter V. Gerasimowicz, CRC Press Inc., Boca Raton, Florida, 1984.

59. Sato, 2002. New Estimation Method for Fatty Acid Composition in Oil using Near Infrared Spectroscopy. J generation systems. *Biosci. Biotechinol. Biochem*. 66: 2543-2548.

60. Shadow, W. 1998. Rapid analysis for the food industry using near-infrared spectroscopy. Perten Instruments North America, Inc.

61. Shadow,W. and A. Carrasco. 2000. Practical single-kernel NIR/visible analysis for small grains. Cereal Foods World 45(1):16-18.

62. Silvela, L., Rodgers, R., Barrera, A., Alexander, D.E. 1989.Ther. Appl. Genet. 78:298

63. Smith, P.K., Krohn, R.I., Hermanson, G.T., Mallia, A.K., Gartner, F.H., Provenzano, M.D., Fujimoto, E.K., Goeke, N.M., Olson, B.J., and Klenk, D.C. 1985. Measurement of protein using bicinchoninic acid. Anal. Biochem. 150:76-85.

64. Sorvaniemi, J., Kinnunen, A., Tsados, A., Malkki, Y. 1993. Using partial least squares regression and multiplicative scatter correction for FT-NIR data evaluation of wheat flours, Lebensmittel-Wissenschaft & Technologie, 26(3): 251-258.

65. Srinivasan, V.T., B.B. Singh, P.K. Chidambareswaran and V. Sundaram. 1985. Cottonseed oil estimation by pulsed Nuclear Magnetic Resonance technique, JAOCS, 62(6).

66. Theissing, H. H, J. Opt. Soc. Am. 40, 232 (1950).

67. Thompson, J.A. and R.L. Nelson.1998.Utilization of diverse germplasm for improving soybean yield. Crop Sci. 38: 1362-1368.

68. Thorpe, T.A. Morphogenesis and regeneration. In: *"Plant cell and tissue culture."* I. K. Vasil and T.A. Thorpe (eds) Kluwer Academic Publishers, Dordrecht, pp.17-36: 1994.

69. Tiwari, P. N. And P. N. Gambhir.1995.Seed oil determination without weighing and drying the seeds by combined Free Induction Decay and Spin-Echo Nuclear Magnetic Resonance Signals, JAOCS, 72(9).

70. You, T., Guo, J., Baianu, I. C., Nelson, R. L, 2002. Determination of Isoflavones Contents for Selected Soybean Lines by Fourier Transform Near Infrared Reflectance Spectroscopy. Soy2002 Conference, Urbana, IL.

71. You, T., Guo, J., Baianu, I. C., Nelson, R. L 2002. Rapid Determination of Protein, Oil, Moisture, and Isoflavones Contents of Single Soybean Seeds by Fourier Transform Near Infrared Reflectance Spectroscopy. China & International Soy Conference & Exhibition 2002, Beijing, China.

72. You, T., Guo, J., Baianu, I. C., Nelson, R. L, 2003. Single Soybean Seeds Calibration and Composition Analysis by Near Infrared Reflectance Spectroscopy with Diode Array and Fourier Transform Instruments. (To be published).

73. You, T., Guo, J., Baianu, I.C., Nelson, R.L, 2003. Diode-Array Near Infrared Spectroscopy for Rapid Soybean Composition Analysis: Light Scattering Corrections for Intact and Ground Soybean Seeds.(*In submission).*

74. Velasco, L., Goffman, F.D., Becker H.C., 1999. Development of Calibration Equations to Predict Oil Content and Fatty Acid Composition in *Brassicaceae* Germplasm by Near-Infrared Reflectance Spectroscopy. JAOCS, 76: 25-30.

75. Wang, D., Dowell, F.E., Lacey, R.E. 1999. Cereal Chem. 76: 30-33.

76. Wei, T.C.1988. Molecular basis of soy protein functionality: relationship between rheological properties, hydration and aggregation of soy proteins, M.S. Thesis,UIUC

77. Wei, T.C. 1990. Molecular Dynamics, Conformation and Hydration Studies of Soybean Proteins and Amino Acids by NMR. PhD Thesis, University of Illinois, Urbana.

78. Wilcox, J. R. and Cavins, J. F., July/Aug 1995. Backcrossing high seed protein to a soybean cultivar. Crop-Sci, 35(4):1036-1041.

79. Wilcox, J.R., J.F. Cavins, and N.C. Nielsen. 1984. Genetic alteration of soybean oil composition by a chemical mutagen. Journal of the American Oil Chemists Society, 61: 97-100.

80. Williams, P.C. 1975. Cereal Chem. 52: 561-576.

81. Williams, P., Norris, K. 1987. *"Near-Infrared technology in the agricultural and food industries."* American Association of Cereal Chemists, Inc. St.Paul, MN.

82. Wilson, R. F. Advances in the genetic alteration of soybean oil composition. In: R.F. Wilson (ed). *"Designing Value-Added Soybean for Markets of the Future."* American Oil Chemists Society, Champaign, Illinois p.38-53, 1991.

FIGURE CAPTIONS

Fig. 3.2.1. Overlay plot of DA-NIR spectra of single soybean seeds obtained with the Perten DA-7000 instrument. A: Before MSC; B: After MSC.

Fig. 3.2.2. Overlay plot of FT-NIR spectra of single soybean seeds obtained with the Perkin-Elmer Spectrum ONE instrument. A: Before MSC; B: After MSC

Fig. 3.4.1.1. Simple One-Pulse sequence for high-resolution NMR analysis of oil.

Fig. 3.4.3.1. A pencil probe of a General Electric GN300WB FT-NMR spectrometer, with a zirconia rotor sleeve, Kel-f drive tip and Teflon front spacer and endcap

Fig. 3.4.3.2. The 1PDNA pulse sequence employed in ^{13}C SS-NMR experiments of oil content determination in soybean flours.

Fig. 3.4.7.1. The WALTZ-16 Decoupling Pulse Sequence for Liquid- State ^{13}C NMR.

Fig. 3.5.1. The VACP pulse sequence employed in ^{13}C SS-NMR Measurements of Protein Content in Soybean Flours**.**

Fig. 4.1.1. Standard Protein Values vs. Calculated Values by FT-NIR Calibrations for Single Seed Soybean Analysis.

Fig. 4.1.2. Standard Oil Values vs. Calculated Values by FT-NIR Calibrations for Single Seed Soybean Analysis.

Fig. 4.1.3. Standard Protein Values vs. Calculated Values by FT-NIR Calibrations for Bulk Soybean Sample Analysis.

Fig. 4.1.4. Standard Oil Values vs. Calculated Values by FT-NIR Calibrations for Bulk Soybean Sample Analysis

Fig. 4.1.5. Standard Protein Values vs. Calculated Values by DA-NIR Calibrations for Bulk Soybean Sample Analysis.

Fig. 4.1.6. Standard Oil Values vs. Calculated Values by DA-NIR Calibrations for Bulk Soybean Sample Analysis.

Fig. 4.1.7. Standard Protein Values vs. Calculated Values by DA-NIR Calibrations for Single Seed Soybean Analysis.

Fig. 4.1.8. Standard Oil Values vs. Calculated Values by DA-NIR Calibrations for Single Seed Soybean Analysis.

Fig. 4.2.1.1. WALTZ-16 decoupled ^{13}C Liquid-State NMR of a soy flour gel sample with 38.7% protein content. Spectrum recorded with 10,000 transients on a Varian UI-600 NMR spectrometer, in a 14.1 T external magnetic field**.**

Fig. 4.2.1.2. WALTZ-16 decoupled ^{13}C Liquid-State NMR of a Soy Protein Isolate gel sample with 50.8% protein content. Spectrum recorded with 5,000 transients, on a Varian UI-600 NMR spectrometer, in 14.1 T external magnetic field.

Fig. 4.2.1.3. WALTZ-16 ^{13}C Liquid- State NMR of soy oil sample. The spectrum was recorded with 112 transients on a Varian UI-600 NMR spectrometer, in 14.1 T external field.

Fig. 4.2.1.4. Standard Calibration Plot for Soy Protein Isolates.

Fig. 4.2.2.1. Protein (Dry weight %) vs. Oil (Dry weight %) Inverse Correlation for the Amino Acid Set of the 65 Standards selected for the NIR AA Calibration, in addition to the remaining 35 standards employed for independent validation.

Fig. 4.2.2.2. Arginine as % Tot Dry Wt. vs. D% Protein.

Fig. 4.2.2.3. Glutamine plus Glutamic, Glx, as % Tot Dry Wt. vs. D% Soybean Protein.

Fig. 4.2.2.4. Leucine as % Tot Dry Wt. vs. D% Protein.

Fig. 4.2.2.5. Valine as % Total Dry Weight vs. D% Protein.

Fig. 4.3.1. The ^{1}H NMR spectrum taken with an NMR spectrometer Varian U 400 and a Nalorac 5mm ^{1}H Quad probe, in an external magnetic field of 9.4 T, at an ^{1}H NMR resonance frequency of 400 MHz.

Fig. 4.3.2. The soybean oil standard plot for 400 MHz ^{1}H NMR measurements on the Varian U-400. The probe was a Nalorac 5 mm QUAD for ^{1}H NMR.

Fig. 4.3.3. The ^{1}H NMR oil peaks in a 0.9 μL soybean oil sample collected by accumulating 500 transients with a Varian U400 NMR spectrometer and a Nalorac 5mm Quad probe tuned to a resonance frequency of 400 MHz in an external magnetic field of 9.4 T.

Fig. 4.4.1. Glycine spectra with its carbonyl peak at 176.03 ppm relative to TMS.

Fig. 4.4.2. 1PDNA ^{13}C SS-NMR of soybean flour for a standard sample of 20% oil content (~300 mg total wt). The oil signal of interest exhibits several resolved peaks close to 130 ppm.

Fig. 4.4.3. 1PDNA ^{13}C SS-NMR of a standard sample of 53.5 mg soybean oil in Al_2O_3.

Fig. 4.4.4. 1PDNA ^{13}C SSNMR of Oil Content of Soybean Flours and their direct correlation with the corresponding NIR data.

Fig. 4.5.1. VACP ^{13}C SSNMR of a Soybean Flour sample of 41% protein content (300 mg total weight). The component of interest has peaks in the spectral region centered at 174 ppm.

Fig. 4.5.2. VACP ^{13}C SS-NMR Measurements of Protein Content in Soybean Flours, and their direct, Linear Correlation with the corresponding NIR data.

Fig. 6.1. Protein-Oil Inverse Correlation of 5000 Soybean Samples of Developmental Lines at UIUC.