THE ANALYSIS OF COMPLEX CARBOHYDRATES BY FOURIER

TRANSFORM-INFRARED MICROSPECTROMETRY AND SINGLE-BOUNCE

ATTENUATED TOTAL REFLECTION SPECTROMETRY

by

RICHARD BRIAN MELKOWITS

(Under the direction of James A. de Haseth)

ABSTRACT

A strong need exists for the development of methodologies for compositional and structural analyses of complex carbohydrates. Vibrational spectrometry is a valuable and powerful tool for the interrogation of many chemical systems, both qualitatively and quantitatively and is applicable to the analyses of complex carbohydrates. Carbohydrates play numerous critical roles in biochemical systems, and they are ubiquitous components of living organisms. Biologically, their role as a source of energy for physiological processes is essential. More profoundly, however, they play crucial roles in the maintenance of cellular structural integrity and biosynthesis, and oligosaccharides are required for the direction of these processes and also behave as chemical messengers. Oligosaccharides are difficult and expensive to extract from biological sources, and, typically, only very small quantities are available for analysis. Therefore, a sensitive analytical method that can interrogate microgram quantities of sample is necessary.

A methodology to determine the composition of N-linked mammalian oligosaccharides was developed with the use of Fourier transform infrared microspectrometry combined with chemometrics. This initial methodology was later modified for compositional and structural elucidations by single-bounce attenuated total reflection spectrometry. These results demonstrated predictions with less than four percent error in both validation and double blind studies.

Various sources of (1-3)-β-D-glucans have been found to have tumor-necrotizing effects in mammals. Many of these anti-tumor glucans contain this structure as a backbone with O-6-linked β-glucosyl branches with a degree of branching of 1:3. A method of the structural analysis of intact polysaccharides was investigated by Fourier transform infrared spectrometry/attenuated total reflection spectrometry. The system was comprised of maltose and cellulose standards, as the sole monosaccharide subunit in these polysaccharides is D-glucose. The results indicate very high predictability of the relative extent of alpha and beta linkage.

INDEX WORDS: FT-IR, Infrared spectrometry, Infrared spectroscopy, Mid-IR, Mid-infrared, ATR, Attenuated total reflection, Attenuated total reflectance, Single-bounce, Single-reflection, Carbohydrates, Complex carbohydrates, Monosaccharide, Oligosaccharide, Polysaccharide, PLS, Partial least squares, Chemometrics

THE ANALYSIS OF COMPLEX CARBOHYDRATES BY FOURIER

TRANSFORM-INFRARED MICROSPECTROMETRY AND SINGLE-BOUNCE

ATTENUATED TOTAL REFLECTION SPECTROMETRY

by

RICHARD BRIAN MELKOWITS

B.A., Bucknell University, 1996

A Dissertation Submitted to the Graduate Faculty

of The University of Georgia in Partial Fulfillment

of the

Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2002

THE ANALYSIS OF COMPLEX CARBOHYDRATES BY FOURIER

TRANSFORM-INFRARED MICROSPECTROMETRY AND SINGLE-BOUNCE

ATTENUATED TOTAL REFLECTION SPECTROMETRY

by

RICHARD BRIAN MELKOWITS

Approved:

Major Professor: James A. de Haseth

Committee: Thomas E. Johnson
                     Ron Orlando
                     Robert S. Phillips
                     John L. Stickney

Electronic Version Approved:

Gordhan L. Patel
Dean of the Graduate School
The University of Georgia
August 2002

DEDICATION

This dissertation is dedicated to Amanda Knapp and to my family for their unwavering love and support.

ACKNOWLEDGMENTS

There are several people that deserve thanks for their patience and tenacity during my graduate career. Primarily, I would like to thank Dr. Jim de Haseth for his wisdom, guidance, and faith in me. His enthusiasm for teaching and learning has been an inspiration to me during the course of my graduate career. I would like to thank Dr. Thomas Johnson, Dr. Ron Orlando, Dr. Robert Phillips, and Dr. John Stickney for serving on my doctoral advisory committee and Dr. Carl Bergmann, Dr. Alan Darvill, Dr. Steve Levery, Jeremi Johnson, and Dan King for offering me assistance when I needed it. An overwhelming thanks goes out to the current and past research group members who have gracefully dealt with my antics over the years. I would like to thank Randy Bishop, Yu Cang, Tracey Cash, Jessica Jarman, Ushiri Kulatunga, Andrew Thomas, and Andy Todebush. I would also like to thank the former post doctorates and undergraduates who have worked in our lab: Steve Altomari, Guy David, Jennifer Gurley, Lin-Tao HE, Scott Pannell, Cora Vette Reddick, Shelley Seerly, and Charlotte Swanson. I wish all of you success and happiness in your future endeavors.

I would also like to express my gratitude towards Mai Yin Tsoi and Kris Varazo for their friendship and faith in me, and towards Andrew Thomas and Lindell Ward for proofreading this dissertation. I would like to thank my parents, Joan and Rich, my sisters, Heather and Jennifer, and my grandmother, Rosemarie, for their unwavering support over the years. Finally, I would like to express my overwhelming gratitude towards Amanda Knapp whom I love and is very special to me, and who unconditionally and continually gives me light.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 2

CHAPTER 3

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

The key to the existence of all biological species is found in the role that carbohydrates play in nature. Carbohydrates are essential for energy production and storage as well as for the provision of mechanical support for cellular structures.[1-4] The oxidation of carbohydrates is the principle metabolic mechanism for energy production throughout the biosphere.[5] Cellulose, the most abundant biopolymer, the exoskeletons of various insects, and the cell walls of bacteria, fall under the umbrella of carbohydrates.[6]

Carbohydrates also are in demand for a variety of industrial applications. Most notably, the food industry relies upon carbohydrates, particularly starch, for the production of sweeteners, bakery goods, beverages, gums, etc. Despite the introduction of synthetic materials, the textile industry largely still depends on cellulose for manufactured goods. The pharmaceutical industry has flourished from its development of synthetic vitamins and antibodies, both of which rely heavily upon carbohydrates for their synthesis. Finally, the chemical industry has benefited from the generation of monosaccharides, polysaccharides, and other related compounds in pure form.[7-10]

Complex carbohydrates, in particular, have been shown to be valuably bioactive via their participation in biochemical and physiological interactions within biological systems.[11-13] Proteins that contain carbohydrates, which are covalently bonded, are

called glycoproteins. Carbohydrate structures associated with glycoproteins, commonly referred to as glycans, can be in the form of monosaccharides, disaccharides, polysaccharides, and oligosaccharides, and affect the physiochemical and biological functions of the glycoprotein.[14,15] This is achieved by changing the properties of the glycoprotein by varying the structure of the glycan. They can also direct the protein folding and subunit assembly,[16] and are significant in their effect on immunological properties,[17-19] modification of the transmission signals for cellular response,[20] and their influence on the activity of hormones and enzymes.[21,22] For example, the recruitment of leukocytes to injured tissue was found to occur through its interaction with its cell surface carbohydrate, Sialyl-Lewis [Neu5Acα2-3Galβ1-4(Fucα1-3)GlcNAc], and endothelial cell selectins. The selectins are expressed in reaction to cytokines released during the inflammatory response.[23-25] As a result of these findings, an increased interest in the development of saccharide-based anti-inflammatory and anti-cancer drugs has commenced. Furthermore, variations in protein glycosylations have been demonstrated to be valuable molecular markers in the diagnosis of a number of human diseases.[25-29] Finally, complex carbohydrates compose the backbone of deoxyribonucleic acid, consisting of four repeating nucleotides that contain 2-deoxy-D-erythro-pentofuranose, which encodes information with regard to transcription and replication for the development of new cells in biological systems.[2]

   With increasing awareness of the biological activity of complex carbohydrates, a growing demand has incurred for a simple, rapid, accurate, and inexpensive method to characterize both the composition and structure of complex carbohydrates.[17,30,31] The

isolation of glycans from glycoproteins is difficult and expensive, and therefore, the sample quantity typically available for analysis is considerably limited.[31-33]

The full characterization of glycoproteins requires several determinations. The primary structure as well as the conformation of both the protein and the carbohydrate side chains, which are attached, is to be elucidated. In addition, the pattern of carbohydrate heterogeneity at each glycosylation site, the location of each glycosylation site, and the anomeric specificity of the linkages, are typically desired. Whereas the determination of the protein structures is relatively easy and well established, the complete characterization of the carbohydrate structures continues to be problematic.

Part of the challenge in the elucidation of carbohydrate structures is attributed to the excessive variability of potential structures. A large number of saccharide chain variations can be brought about by a small number of monosaccharide units. Carbohydrate moieties can be linked to the protein via any of four hydroxyl groups per monosaccharide, and they can exist in either form of two anomeric forms and involve either pyranose or furanose rings. As a result, two monosaccharides can form as many as thirty-two disaccharides, whereas, two amino acids can only form two dipeptides. The number of carbohydrate structures increases geometrically as the number of monosaccharide units increase because branching becomes a possibility when there are more than two units present.[33]

In addition, complex carbohydrates can be divided into two groups assigned by its linkage site on the protein. N-Linked oligosaccharides are attached to the amide group of an asparagine side chain in an Asn-X-Ser(Thr) sequence with N-acetyl-D-glucosamine as the reducing terminal monosaccharide. O-Linked

oligosaccharides are linked to the hydroxyl group of a Ser or Thr residue in the polypeptide backbone, and the reducing terminal monosaccharide is usually N-acetyl-D-galactosamine.[34] N-Linked oligosaccharides contain a common pentasaccharide core structure, Manα1-6(Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAc. N-Linked oligosaccharides are typically larger than O-linked oligosaccharides and are divided into three categories: high mannose-type sugar chains, complex-type sugar chains, and hybrid-type sugar chains.[35] Examples are shown in Figures 1.1 through 1.3.

The isolation of the glycan from the glycoprotein is a crucial step for the characterization of the carbohydrate. This isolation results in the release of the glycan from the protein to form a pool of intact oligosaccharides, which, ideally, can be subsequently separated from each other. Both of these tasks, particularly the latter, are very challenging. The former task can be achieved either enzymatically or chemically.

Two types of enzymes are known and are now commercially available for release of asparagine-linked oligosaccharides from the glycoprotein. The first, peptide-N-(N-acetyl-β-glucosaminyl)-asparagine amidase, severs the GlcNAc-Asn linkage to provide an intact oligosaccharide that possesses a reducing terminal. The second, *endo*-β-N-acetyl-D-glucosaminidase, hydrolyzes the bond between the two glucosamine residues, and therefore, one of the N-acetyl-D-glucosamine units remains attached to the protein.[36,37]

The release of oligosaccharides from glycoproteins can be achieved chemically for both N-linked and O-linked oligosaccharides with either hydrazine or trifluoromethanesulfonic acid. The former is able to differentiate between N-linked and O-linked oligosaccharides and provides the oligosaccharides with reducing termini.

Figure 1.1   Example structures of high mannose-type oligosaccharides.

Manβ(1-4)-GlcNAcβ(1-4)-GlcNAc-N

Manα(1-6)

Manα(1-3)

Manα(1-6)

Manα(1-3)

Manβ(1-4)-GlcNAcβ(1-4)-GlcNAc-N

Manα(1-6)

Manα(1-3)

Manα(1-2) —— Manα(1-6)

Manα(1-3)

Manα(1-2) —— Manα(1-2) —— Manα(1-3)
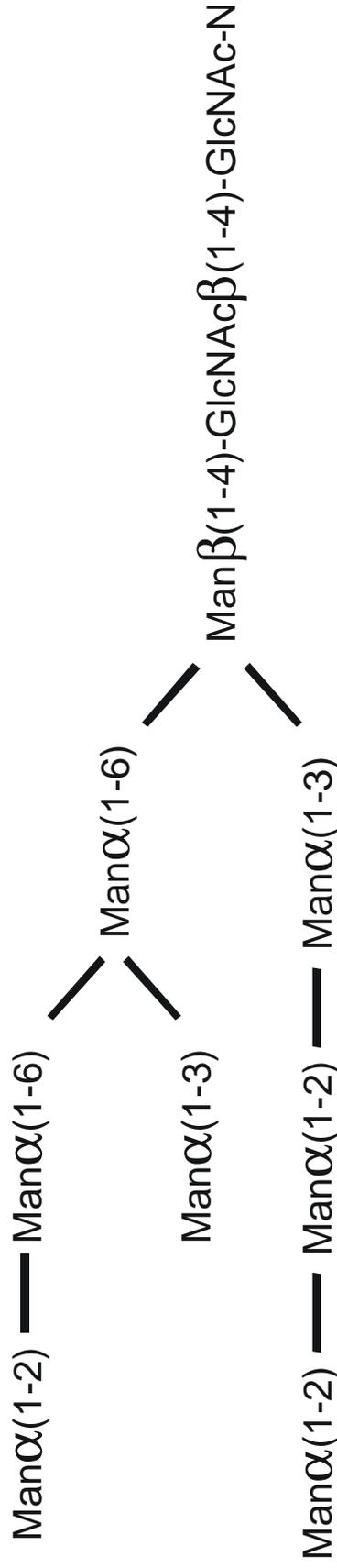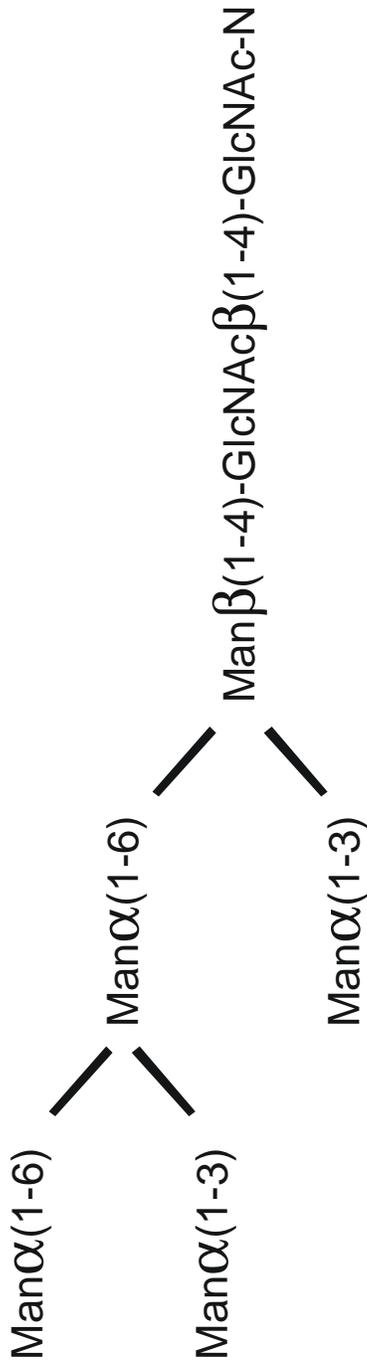
Figure 1.2   Example structures of complex-type oligosaccharides.

Galβ(1-4)-GlcNAcβ(1-2)-Manα(1-6)

Galβ(1-4)-GlcNAcβ(1-2)-Manα(1-3)

Manβ(1-4)-GlcNAcβ(1-4)-GlcNAc-N

Galβ(1-4)-GlcNAcβ(1-6)

Galβ(1-4)-GlcNAcβ(1-3)

Manα(1-6)

Galβ(1-4)-GlcNAcβ(1-2)

Manα(1-3)

Manβ(1-4)-GlcNAcβ(1-4)-GlcNAc-N
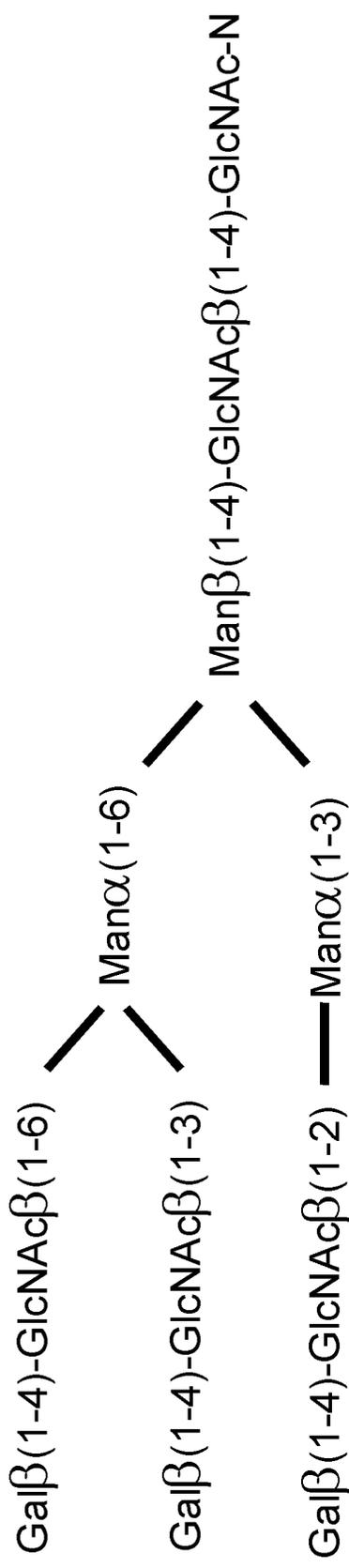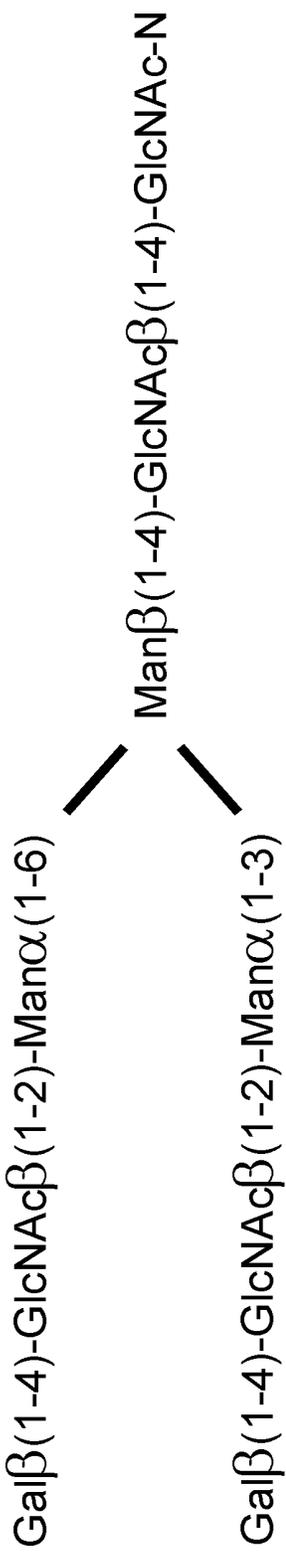
Figure 1.3   Example structures of a hybrid-type oligosaccharide and a bisected

oligosaccharide.

Manα(1-6)

Manα(1-6)

Manα(1-3)
Galβ(1-4)-GlcNAcβ(1-6)

Galβ(1-4)-GlcNAcβ(1-3)

Manβ(1-4)-GlcNAcβ(1-4)-GlcNAc-N

Manα(1-6)

Manα(1-3)

Manα(1-6)

Manα(1-6)

Manα(1-3)
Galβ(1-4)-GlcNAcβ(1-6)

Galβ(1-4)-GlcNAcβ(1-3)

GlcNAcβ(1-2) - Manβ(1-4)-GlcNAcβ(1-4)-GlcNAc-N

Manα(1-3)
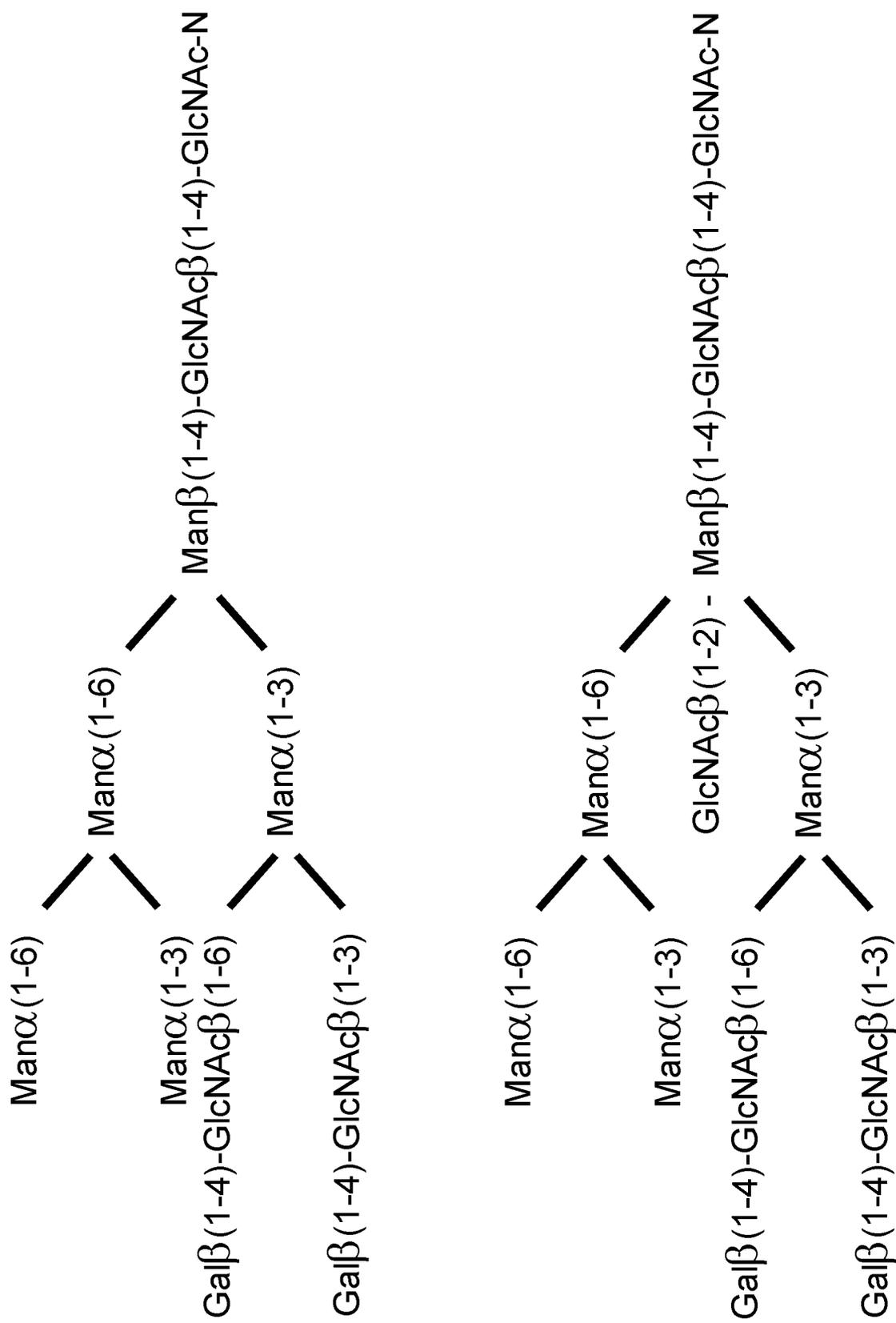
Trifluoromethanesulfonic acid, however, will preserve the intact protein whereas hydrazine will inevitably degrade it.[38]

Because the monosaccharide constituents of complex carbohydrates have similar structures, it is very difficult to separate heterogeneous glycoforms into homogeneous oligosaccharides. Frequently, the oligosaccharide composition only differs by one monosaccharide, and sometimes only the anomeric linkages vary in cases where the composition is identical.

The separation of complex carbohydrates has been attempted by a number of different methods. Among these are several common chromatographic methods such as gas chromatography, thin-layer chromatography, high-performance liquid chromatography, and supercritical fluid chromatography.[31,33,39-45] Also, methods have been applied which depend upon both size and charge. Although high-pH anion-exchange chromatography with pulsed amperometric detection is an obvious choice for charged species,[15,46,47] for uncharged oligosaccharides, gel-permeation chromatography and ion-exchange chromatography in a borate buffer, only if the sugars have *cis*-hydroxyl groups, which allow for the formation of charged borate complexes, are sometimes successful. For the elucidation of any carbohydrate structure, it is essential to analyze its monosaccharide composition. Two major restrictions obstruct the separation and detection of the monosaccharide subunits. First is their weak absorption in the ultraviolet region, which is attributed to the presence of only a small fraction of an aqueous saccharide solution in the carbonyl form, which renders the solution insufficiently chromophoric.[48] Furthermore, there is the inconvenience imparted by their lack of fluorescence since the transition states are too high in energy

to be observed at wavelengths longer than 190 nm.[49] In some cases, these problems can be circumvented by the derivatization of the carbohydrates prior to separation and detection.[30]

Mass spectrometry and gas chromatography/mass spectrometry have been applied to monosaccharide compositional analysis of oligosaccharides that are depolymerized after being subjected to hydrolysis.[50,51] Unfortunately, mass spectrometry is not inherently capable of distinguishing monosaccharides that are isobaric, such as glucose and mannose, which it is often that monosaccharides only differ in their stereochemistry. Therefore, it is usually necessary to apply a separation technique, such as gas chromatography, prior to mass-spectrometric analysis.[32,52,53] In addition, ionization of pure monosaccharides is not easily achieved and, therefore, it usually requires the derivatization of the sample to assist in ionization and detection by mass spectrometry.[16,54,55]

Nuclear magnetic resonance spectrometry has gained prominence as a method for compositional and structural elucidation of complex carbohydrates.[14,33,53,56-59] A large amount of starting material, however, is required and is often unfeasible to obtain from biological samples. Many complex carbohydrates have over 500,000 distinct ^{1}H environments, and to achieve a resolution of 0.5 part per million, a 10^{12} Hertz nuclear magnetic resonance spectrometer, which currently does not exist, would be required.[60] Finally, the rate at which the data from a single sample could be interrogated could cause the duration of an experiment to exceed a day for the acquisition of useful information. In addition to the inconvenience imposed by the excessive time consumption, this long an acquisition period poses a great challenge for any instrument

in the maintenance of a constant homogeneous magnetic field, which would be necessary to preserve background stability. Both nuclear magnetic resonance spectrometry and mass spectrometry are time-consuming, involve expensive instrumentation, and require highly specialized expertise for both operation and data interpretation.

Obviously, the introduction of a rapid, sensitive, accurate, inexpensive, and simple technique would be quite welcome to address the critical need for carbohydrate analysis. Fourier transform infrared spectrometry has been used in the past to investigate carbohydrates in the food and beverage industries.[8,9,61,62] It has not been used routinely, until recently, to analyze biological complex carbohydrates. By using Fourier transform infrared microspectrometry, small amounts of sample can be interrogated with high sensitivity. This technique offers an advantage over previous methods in several respects. It does not suffer from the inability to discern isomeric and isobaric molecules that mass spectrometry does, nor does it suffer from the slow rate of data acquisition or insensitivity from which nuclear magnetic resonance possesses. In addition, it does not require the level of expertise needed for either of the former techniques, and the instrumentation can be acquired and maintained for a fraction of the cost. Visual qualitative analysis of monosaccharides and polysaccharides by infrared spectrometry is nearly impossible since the vibrational spectra they produce in the infrared region contain no unique bands to distinguish individual saccharide species from each other. Compositional analysis of oligosaccharides is further hindered because the monosaccharides from which they are composed appear very similar to one another and contain overlapping bands. The spectra of monosaccharide mixtures with

varying concentration do contain minor variations even though they are indiscernible to visual interpretation. Recent advances in statistics and computer algorithms circumvent this restriction by offering an alternative for the interpretation the data and reliable extraction of useful compositional and structural information.

This dissertation discusses a new approach for the analysis of complex carbohydrates with vibrational spectrometry in the mid-infrared region. The following chapter describes the design of a novel methodology to determine the composition of mammalian N-linked oligosaccharides quantitatively by Fourier transform infrared microspectrometry. Chapter three takes this approach a step further through the application of another infrared spectrometric technique, attenuated total reflection spectrometry, which, for this application, proved to be superior to, and ultimately simpler than, the microspectrometric technique. A discussion of the theory and instrumentation of attenuated total reflection is provided in the introduction of the third chapter. Finally, chapter four investigates the potential of infrared spectrometry to analyze structural differences among several intact glucans. Again, the approach is quantitative and further exploits the advantages of attenuated total reflection spectrometric technology. Glucans are chosen because not only are they a critical initial step toward the analysis of intact oligosaccharides, but also glucans, themselves, are important for cancer research and are being investigated for their anti-tumor properties.[18,19,29] A method to analyze them accurately and efficiently would be invaluable to the medical community and the overall quality of life.

Fourier Transform Infrared Microspectrometry

Fourier transform infrared microspectrometry unites the areas of vibrational spectrometry and microscopy. In the scientific community, it has become a favorable method for the analysis of biological samples, as it is capable of measuring small areas of large samples. In fact, improvements in sample synthesis and purification have allowed it to be used for the measurement of small sample quantities approaching the picogram level. Infrared microspectrometric techniques are useful for a wide range of sample types especially those encountered in the forensic, textile, polymeric, and biological fields.[27,63-65]

Infrared microscope systems usually have a computer-interfaced video cameras incorporated into them, which allow an image of the sample to be visually displayed and stored in a database. Often, the stage of the microscope, where the sample resides, is a computer-interfaced translation stage, which allows spectral profiling of a desired area of a sample at about 20-μm resolution and provides for applications that involve automated stage movement. Infrared microscopes differ from conventional microscopes in that they contain two beam paths. One beam path is for infrared radiation, and the other is for visual observation. The two beam paths, however, share the same condenser, and they intersect at a remote aperture. The aperture is located at an intermediate focus rather than the focus at the sample, as the focus of the beam is smaller at the latter location. As a result, the radiation is restricted to a much smaller area thereby reducing the scattering. When a sample is in focus, its conjugate image is focused at the remote aperture, its visible image can be seen through the optical microscope, and its infrared interferogram is sent to the detector, which is usually a

small element (100 x 100 µm to 200 x 200 µm) mercury cadmium telluride detector to maximize sensitivity.[66] The microscope allows the operator to switch between the two modes without repositioning or refocusing the sample since the same cassegrain objective is used for the visible and infrared radiation. When the microscope is in the viewing mode, the helium-neon laser and infrared radiation from the spectrometer are blocked from reaching the sample. Likewise, when the microscope is in the infrared spectral collection mode, visible radiation from the illuminator does not reach the sample.[67]

Infrared microscope accessories are capable of the measurement of both transmission and reflection spectra. In Figure 1.4, the optical path for viewing the sample in transmission mode is shown. Mirror 1 redirects light from the illuminator up through cassegrain 1, which condenses the beam to a suitable size for the sample and focuses it at the sample location. The light from the sample is collected by cassegrain 2 and directed through the remote aperture after which the beam has an unobstructed path to the optical microscope. A transmission spectrum can be collected by the system as shown in Figure 1.5. Mirror 3, instead of receiving visible radiation from the illuminator, receives the infrared radiation from the spectrometer directed by mirror 2, a toroidal flipping mirror. Mirror 1, the detector mirror, is positioned into the beam above the remote aperture. Cassegrain 3, the mercury cadmium telluride detector cassegrain, condenses the beam onto the detector where the interferogram is collected.[67]

In order to view the sample through the optical microscope when the system is in reflection mode, which is shown in Figure 1.6, mirror 4 (a relay mirror) directs light received from the illuminator down one side of cassegrain 2 to the sample.

Figure 1.4   The path of the optical beam in an infrared microscope when viewing in

transmission mode.

Visible Objective

Remote Aperture

Cassegrain 2

Sample Position

Cassegrain 1

Beam Path
from Illuminator

Mirror 1

Base of Microscope
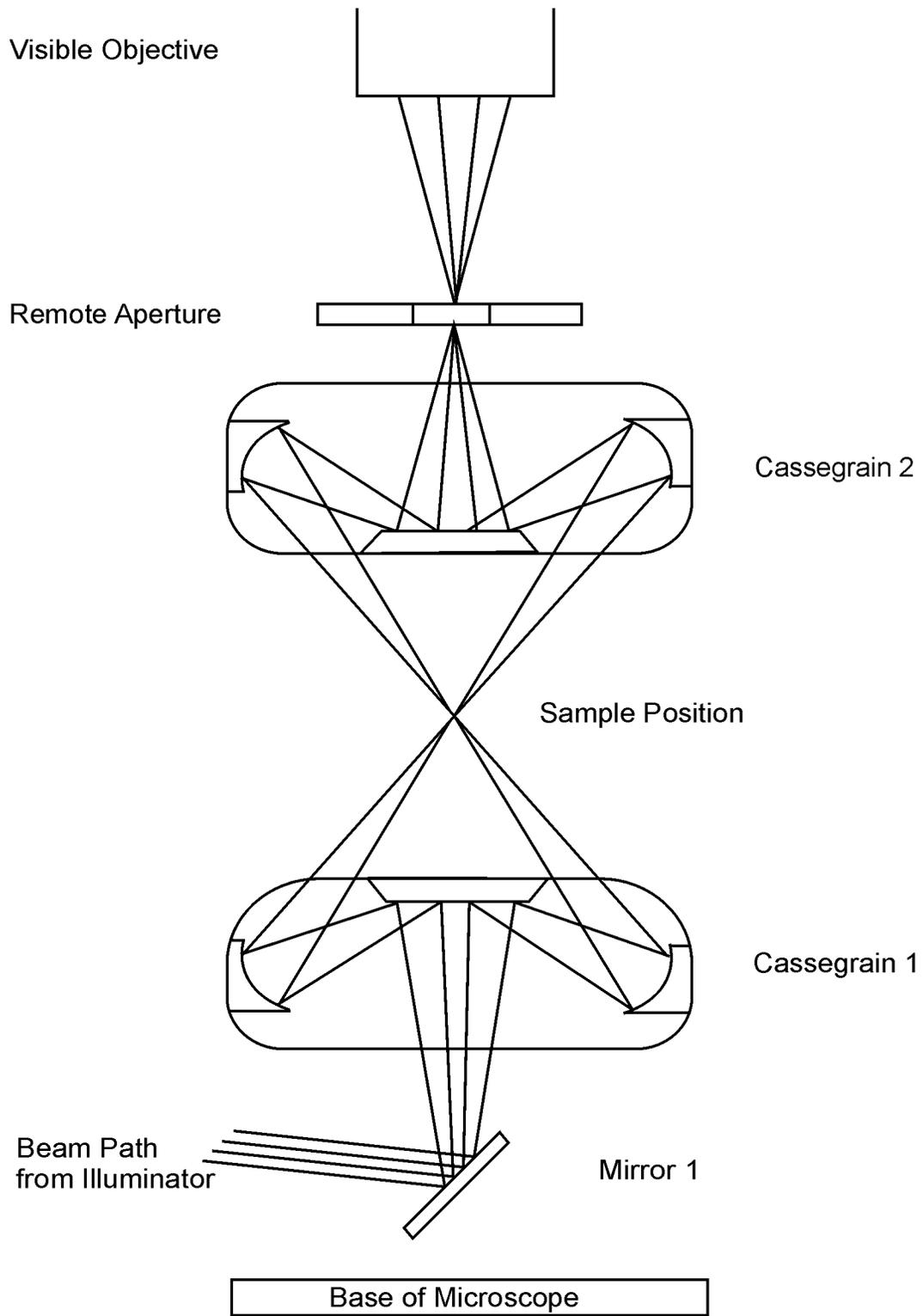
Figure 1.5   The path of the infrared beam in an infrared microscope when collecting

a spectrum in transmission mode.

Cassegrain 3

Visible Objective

Mirror 3

Detector

Remote
Aperture

Cassegrain 2

Sample Position

Cassegrain 1

Mirror 2

Mirror 1

Beam Path from
Spectrometer

Base of Microscope
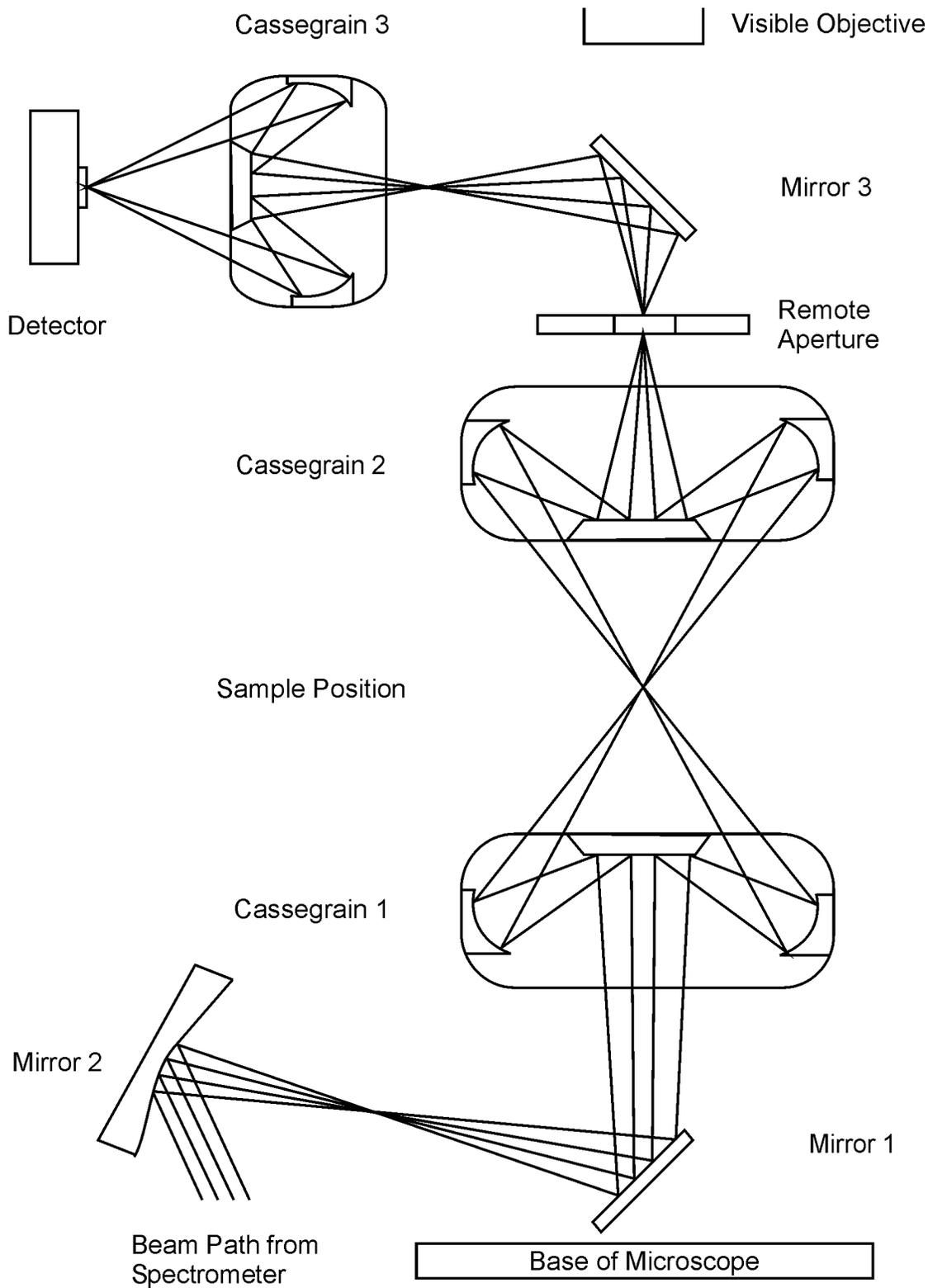
Figure 1.6   The path of the optical beam in an infrared microscope when viewing in

reflection mode.

Visible Objective

Remote Aperture

Mirror 4

Beam Path
from Illuminator

Cassegrain 2

Sample Position

Cassegrain 1

Mirror 1

Base of Microscope
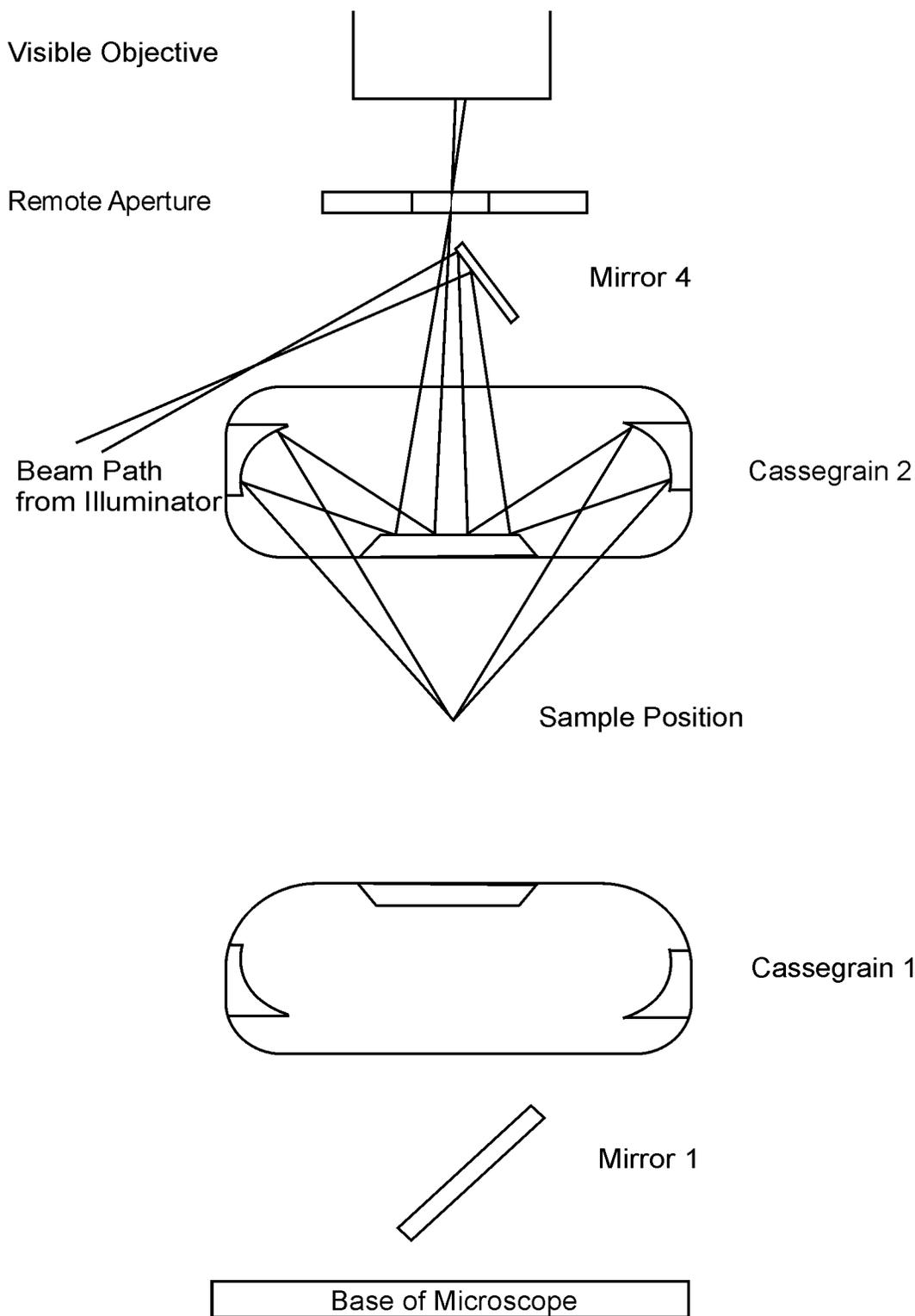
Cassegrain 2 collects the radiation from the sample and sends it through the remote aperture to the optical microscope. When the system is in this mode, the sample is shielded from the radiation of both the helium-neon laser and the infrared source of the spectrometer. To collect a spectrum in reflection mode, as shown in Figure 1.7, mirror 2 directs the infrared radiation from the spectrometer to mirror 3, rather than directing to mirror 4, and the sample receives no visible radiation from the illuminator. Mirror 4 directs the beam down through one side of cassegrain 2 to reflect off the sample, and cassegrain 2 collects the reflected radiation and directs it through the remote aperture. Mirror 1 is positioned above the remote aperture, and the beam is directed through cassegrain 3, which focuses it onto the detector where the interferogram is attained.[67]

The size of a sample image is determined by an aperture, usually called a remote aperture or targeting aperture. In the absence of the aperture, the infrared microscope would allow a large percentage of the radiation to pass through the sample and produce spectra from regions that are not of interest. That is, the beam may be too large in diameter to examine a limited region of the sample. A second aperture under the stage also rejects a large portion of unwanted infrared radiation to reduce scattering effects and consequently reduce spectral impurities. The minimum sample size that can be examined is between 10 and 15 μm in diameter, which is limited by diffraction effects rather than by the signal-to-noise ratio, which is often an important performance criterion for macroscopic measurements.[66]

Figure 1.7   The path of the infrared beam in an infrared microscope when collecting

a spectrum in reflection mode.

Cassegrain 3

Visible Objective

Mirror 3

Detector

Remote Aperture

Mirror 4

Cassegrain 2

Beam Path from
Spectrometer

Sample Position

Cassegrain 1

Mirror 1

Base of Microscope
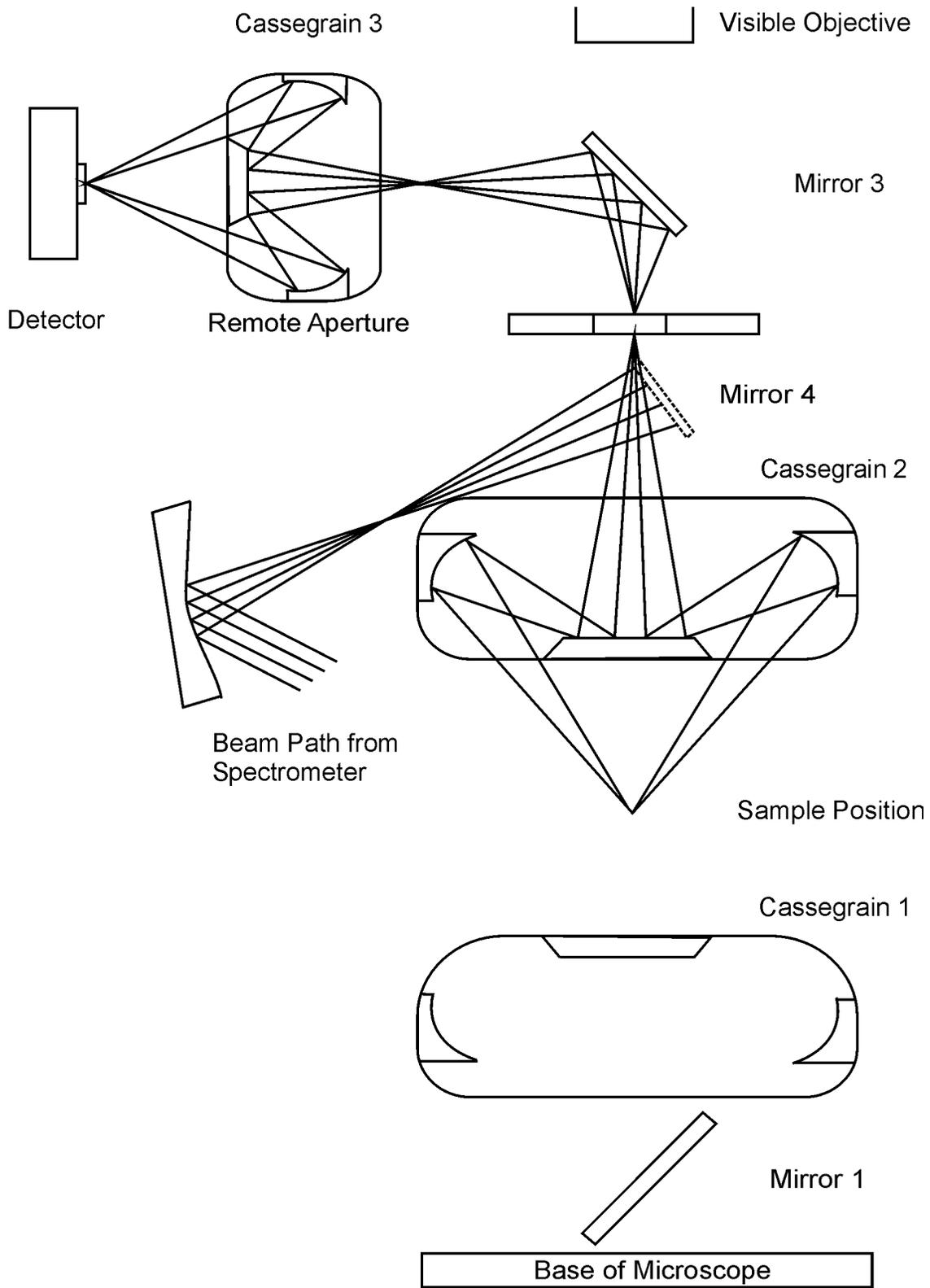
Chemometrics

For analytical chemistry, the field of relating measurements made on a chemical system via application of mathematical or statistical methods has evolved into a discipline of its own known as chemometrics. Infrared spectra of mixtures often contain numerous overlapping bands, especially when the mixtures are of closely related compounds, such as monosaccharides. Fortunately, developments in statistics and computer algorithms have generated an abundance of calibration methods. Many of these methods have expanded the capabilities of quantitative analytical techniques to address increasingly difficult problems, such as spectrometric carbohydrate analysis.[68,69]

Calibration, as described herein, is a procedure that relates instrumental measurements to analytes of interest and requires a known relationship between these measurements and the quantity of the analytes present.[70,71] Throughout this work, the Bouguer-Beer-Lambert Law, which states that the concentration, C, of an analyte is directly proportional to the measured absorbance, A, will provide this required relationship unless otherwise stated. That is,

$$A = \varepsilon bC$$

where ε is the absorptivity coefficient of the analyte (constituent), and b is the optical pathlength through the sample.[70-72]

The ultimate goal of spectrometric quantitative analysis is to create a calibration equation or series of equations with a set of standard spectra of known composition. The spectra of such a set and the calculated equation or series of equations is called a calibration set or training set. Once the calibration equations have been determined, they can be used to predict the same quantities in an unknown sample, provided that the

unknown sample is measured under the exact same conditions on the same instrument as the calibration set. Furthermore, the set of standard mixtures should reflect the composition of the unknown as closely as possible and span the range of expected concentrations.

Throughout this work, a calibration method known as Partial Least Squares is used almost exclusively. It is necessary to develop a practical conceptualization of the method to appreciate the information it provides and the advantages it offers in comparison to an abundance of other statistical methods. It is hoped that in the following discussion, the basic principles behind Partial Least Squares are elucidated sufficiently such that any misconceptions that the mathematics are mystically rooted are dispelled. In addition, several its predecessors are explored, by necessity, not only to achieve this end but also to underscore its strengths and weaknesses that differentiate it from other widely accepted methods. If there is further interest or a desire for a more mathematically rigorous discussion of any or all of these techniques, the works cited throughout the final section of the chapter are highly recommended.[73-79]

Classical Quantitation: Least Squares Regression. The most straightforward and readily understood statistical method is the linear regression. The area, or height, of a selected spectral region is assumed to be related to the concentration of the constituents of the sample. This relationship is described by a polynomial equation such as:

$$C = B_1(\text{Area}) + B_0$$

or

$$C = B_2(\text{Height})^2 + B_1(\text{Height}) + B_0$$

where C is the concentration of a constituent of interest, and B is a calibration coefficient. There is only one measurement per sample used to solve the equations, which makes this an example of a univariate model.[80]

Least Squares Regression is a technique in which model equations are solved via peak area, or height, of standards with a known constituent concentration. The coefficients are calculated such that the residuals between the known spectral responses and the predicted responses are minimized. Predicted spectral responses are extrapolated from the values on the calibration curve at known concentrations. It is possible to obtain more than one calibration coefficient, and the minimum number of calibration samples needed to solve the equation is equal to the number of calibration coefficients in the model. The peak areas, or heights, in the equation can be used to predict the concentration of a constituent in unknown samples that contain the same constituent.[80]

The main advantages of this technique are its relative simplicity and the ability to calculate very rapidly with current computer technology. This method is useful primarily for samples that consist of only a few pure compounds. There are, however, several limitations to this method.

Spectral bands that are specifically characteristic of the constituents of interest are requisite. Also, sample purity is crucial; impurities or other constituents that produce overlapping bands with the constituent of interest will produce large errors. Finally, a difficulty arises in the selection of the appropriate polynomial degree, as not all spectrometric systems adhere to linear relationships.[81]

Multivariate Statistical Methods: Classical Least Squares and Multiple Linear Regression. A technique that allows a much larger portion of the spectrum to be exploited than the Least Squares Regression model is Classical Least Squares. As with the former method, Classical Least Squares is founded on the principle of the Bouguer-Beer-Lambert Law. Reexamination of the Bouguer-Beer-Lambert Law equation indicates that, if the pathlength is held constant, which is common in spectrometric measurements, it can be incorporated into a single term with the absorptivity coefficient. This substitution yields the following equation:

$$A_\lambda = K_\lambda\, C$$

And, after rearrangement, prediction of an unknown constituent concentration is trivial:

$$C = A_\lambda/K_\lambda$$

where A_λ is the spectral response at a given wavelength, C is the concentration of the constituent of interest, and the absorptivity coefficient and pathlength are represented as a single absorptivity constant at the same wavelength, K_λ. As with the previous technique, a series of measurements may be taken at different concentrations, and a best-fit line through all of the data points may be calculated. The situation is compounded when more than one constituent is present. In cases where there are two non-interfering constituents, it is necessary to account for the absorbance of each constituent independently, which may be expressed by the equations that follow:

$$A_a = K_a\, C_a$$

$$A_b = K_b\, C_b$$

where C_a and C_b are the concentrations of constituents *a* and *b* in the mixtures, K_a and K_b are the absorptivity constants, and A_a and A_b are the absorbances. The sum of

A_a and A_b is the total absorbance for the mixture. For each additional constituent in the mixture, there must an additional equation introduced, since, as with any algebraic solution, there must be as many equations as unknowns. The above equations, however, reflect neither the dependence of the absorptivity constants on wavelength nor the absorbances that result. The equations that follow demonstrate the dependence of the total absorbance of a mixture upon wavelength:

$$A_{l1} = K_{a,l1} C_a + K_{b,l1} C_b$$

$$A_{l2} = K_{a,l2} C_a + K_{b,l2} C_b$$

where $K_{a,l1}$ and $K_{b,l1}$ are the absorptivity constants for constituents a and b at wavelength l_1, $K_{a,l2}$ and $K_{b,l2}$ are the absorptivity constants for the respective constituents at wavelength l_2, and A_{l1} and A_{l2} are the absorbances at the respective wavelengths. The same mixture may yield variations in absorbance at different wavelengths due to the incongruence of absorptivity constants for each constituent and to their dependence on wavelength.[69,70,72,80,82,83]

In addition, it is necessary to introduce another variable to compensate for errors that are always present in real measurements. Electronic noise, instrumental error, and sample handling error are among a host of other possible variations. The equations that follow take these into consideration:

$$A_{l1} = K_{a,l1} C_a + K_{b,l1} C_b + E_{l1}$$

$$A_{l2} = K_{a,l2} C_a + K_{b,l2} C_b + E_{l2}$$

where E_{l1} and E_{l2} are the residuals between the actual absorbances and the least squares fit values for each of the measured wavelengths. E_{l1} and E_{l2} essentially serve the same function as the offset coefficients in a Least Squares Regression model. If there are

more than two components present or more than two wavelengths used, then the system

of equations that result can be solved via linear algebra. In this manner, the preceding

equations can be represented in matrix terms as:

$$\mathbf{A}_{(l,n)} = \mathbf{C}_{(m,n)} \, \mathbf{K}_{(m,l)} + \mathbf{E}_{(l,n)}$$

where l is the number of calibration spectra, m is the number of constituents, and n is

the number of selected wavelengths. It follows that $\mathbf{A}$ is an $l \times n$ matrix of calibration

spectra, $\mathbf{C}$ is an $m \times n$ matrix of constituent concentrations, $\mathbf{K}$ is an $m \times l$ matrix of

absorptivity constants, and $\mathbf{E}$ is an $l \times n$ matrix of absorbance offset. The $\mathbf{K}$ matrix is

computed from the following equation:

$$\mathbf{K} = \mathbf{A}\mathbf{C}^{-1}$$

where $\mathbf{C}^{-1}$ is the inverse of the constituent concentration matrix. Determination of the

inverse of an $m \times n$ matrix requires m and n to be equal, which is not the case unless the

number of constituents is the same as the number of samples. In the above equation,

however, when overdetermination of the data forces $m \neq n$,[66] the inverse of $\mathbf{C}$ may be

substituted by its pseudo-inverse in the manner below:

$$K = \mathbf{A}(\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}'$$

where K is the least squares estimate of $\mathbf{K}$ in which the residuals are minimized, and $\mathbf{C}'$

is the transpose of $\mathbf{C}$. Once K is known, even is it is not square, it can be used to

determine the concentration of an unknown constituent concentration, $\mathbf{u}$, by:

$$\mathbf{u} = (K'K)^{-1}K\mathbf{x}$$

where $\mathbf{x}$ is the spectrum of an unknown sample.[69,80,82,83]

Classical Least Squares, also known as K matrix, has a number of advantages

over the Least Squares Regression method. One is that wavelength selection is not

necessary so long as the number of wavelengths exceeds the number of constituents in the mixtures. In addition, inclusion of a large number of wavelengths produces an averaging effect in the model, which causes it to be less susceptible to noise present in the data. Furthermore, moderately complex mixtures can be modeled, and, as in Least Squares Regression, the calculations are fast. Finally, if the Bouguer-Beer-Lambert Law is non-linear, it can be approximated as being linear over a finite range of concentrations. A nonzero intercept at zero concentration occurs in the linear region of the curve. This can be incorporated into the model by placing an extra row of ones into the **C** matrix and an extra column of constants for the intercept into the **K** matrix while the **A** matrix remains unaffected.[66]

The overwhelming disadvantage from which Classical Least Squares suffers is that the equations must be calibrated for every constituent present in the mixtures. This constraint arises from the dependence of the absorbance at a particular wavelength on the sum of all of the constituent absorbances. As stated previously, the absorbance of a constituent is the product of its concentration and absorptivity constant. If the concentration of any constituent is omitted, then the predictions may be in error. As the mixtures become increasingly complex, the predictions of the model will further suffer. If contaminants reside in the mixtures measured for analysis that either are not present in the calibration mixtures or are present in the calibration mixtures but not considered, then the ability of the model to make predictions may be critically diminished. Furthermore, this problem may also occur if there are chemical interactions among the constituents, as even slight amounts of reaction byproducts of unknown concentrations may adversely affect the calibration. Classical Least Squares performs best when

applied to both systems that have little or no inter-constituent reactions and systems that are not exceedingly complex and of which the composition is completely known.[77,84-87]

In real world systems, it is unlikely to possess complete knowledge of the composition of a sample mixture. In addition, the analysis of only a few of all of the constituents that exist in a highly complex mixture is often desired. In these cases, Classical Least Squares will not function well. One alternative arises through the rearrangement of the Bouguer-Beer-Lambert Law and expression of it as follows:

$$C = A_\lambda / \varepsilon_\lambda b$$

Combination of the absorptivity coefficient and pathlength into one constant and re-expression in matrix terms yields:

$$\mathbf{C} = \mathbf{A}_\lambda \mathbf{P} + \mathbf{E}$$

where $\mathbf{C}$ is the matrix of constituent concentrations, $\mathbf{A}_\lambda$ is the matrix of absorbances at a specific wavelength λ, $\mathbf{P}$ is the matrix of unknown calibration coefficients that relate the constituent concentrations to the absorbances, and $\mathbf{E}$ is the matrix of concentration errors. This expression of the Bouguer-Beer-Lambert Law allows the concentration to be a function of absorbance at a series of wavelengths. This is unlike Classical Least Squares, where absorbance at a particular wavelength is calculated as an additive function of concentrations. In the system of equations below, it is apparent that the matrix of coefficients can be properly calculated even if the concentrations of all of the other constituents are unknown.

$$C_a = A_{11}P_{a,11} + A_{12}P_{a,12} + E_a$$

$$C_b = A_{11}P_{b,11} + A_{12}P_{b,12} + E_b$$

The P matrix is calculated via linear algebra in manner analogous to the determination

of the K matrix in the former method, and, if the A matrix is not square, then the

calculation must be made by the substitution of A^{-1} with the pseudo-inverse of A. This

method is known as Multiple Linear Regression, Inverse Least Squares, or P matrix. It

appears to be panacea for all quantitative analysis schemes because the sample

composition with respect to only the constituents of interest is sufficient for calibration,

provided that the appropriate wavelengths, which correspond to the absorbances of the

desired constituents, are selected. Multiple Linear Regression is a multivariate method,

that is, the dependent variable is solved from a calculated solution from multiple

independent variables.[77,88]

In univariate models, a wavelength must be selected where all but one of the

constituents have absorptivity constants equal to zero, and hence, the selected

wavelengths can be used to solve separate equations for each constituent. In real

measurements, wavelength selection, even when it is possible, is seldom

straightforward, as many sample systems contain overlapping bands, and therefore, the

equations for all constituents must be solved simultaneously. This restriction is an

inherent weakness of univariate methods. Multivariate methods, with which a series of

equations are solved by use of many measurements per sample for a single calibration,

have a distinct advantage in that they permit the inclusion of spectral absorbances over a

broader range of wavelengths. Therefore, the solutions are more robust due to the

effects of averaging. The main advantage of multivariate methods, however, is that it is

possible to calibrate for a desired constituent without the need to account for any

interferences that occur in the spectra. As a result, they are more useful than univariate methods for systems that consist of highly complex mixtures.[89]

Unfortunately, there is a stipulation sufficiently profound to impose a considerable obstacle for the application of Multiple Linear Regression. The number of wavelengths selected cannot be greater than the number of training samples. Although the number of wavelengths selected can be easily augmented by inclusion of a greater number of calibration mixtures, this apparent remedy will lead to the problem of collinearity. In other words, as more samples are added to the calibration set, near-linear relationships between absorbances at multiple wavelengths begin to occur, which will induce a mathematical solution that is unstable with respect to each constituent.[78] Furthermore, overfitting arises when too many wavelengths are included, which causes the calibration to model noise that is unique to the training set, hence, a deficiency in the predictive accuracy for unknown samples results.[83]

Continuum Regression Statistical Methods: Principle Components Regression and Partial Least Squares. Another approach, which may combine some of the separate advantages of Classical Least Squares and Multiple Linear Regression, is founded on a principle known as spectral decomposition. The concept relies upon the assumption that the spectra of real samples are comprised of many different variations, and that there is only a finite number of independent variations that contribute to the spectral data. It is expected that the largest variations in the spectra of the training set may be changes that are attributed to different concentrations of the constituents of the mixtures. If it were possible to calculate a set of variation spectra, they could be used,

by multiplication by different scaling factors and subsequent co-addition, to reconstruct a spectrum that closely resembles the spectrum of an unknown mixture.[90]

These variation spectra will be referred to throughout this work as loadings, but they are also commonly known as eigenvectors, loading vectors, principle components, factors, or spectral loadings. The scaling factors that are used to reconstruct unknown spectra are called scores or eigenvalues, but the former term will be used herein exclusively. As the calculated loadings come from the original training set, they must be related to the concentrations of the constituents in the mixtures. If the same loadings can be used to predict unknown samples, then the only difference among the spectra of different mixtures would be the scores.

The scores take the place of the absorbances in either of the previous two statistical methods discussed. Because the representation of the spectra is reduced to a few scores from a multitude of wavelengths, the implementation of the Multiple Linear Regression modification to the Bouguer-Beer-Lambert Law may provide the ability to calculate concentrations among the presence of interfering constituents. In addition, the advantages of Classical Least Squares are retained in that the entire wavelength domain may be included in the calculation. All continuum regression methods share this premise, and the differences among these models, including the two discussed below, lay in the manner in which the loadings are calculated.[91] Note in particular that these models are based on variations in relative absorbances and not the absolute absorbances.

One method that uses this model of spectral variation to create the calibration equations is Principle Components Regression. The first step of this method, that is, the

calculation of all of the possible variations, is Principle Components Analysis. Before

Principle Components Analysis may be applied to the data, the spectra are typically

mean-centered. That is, the average spectrum is determined and subtracted from each

of calibration spectra. This allows greater emphasis to be given to differences among

the spectra and less attention to be given to common spectral features. Removal of the

mean simply eliminates the most common variations before the data undergoes

Principle Components Analysis.[90]

Principle Components Analysis is essentially an iterative elimination of each

independent variation from the calibration data in series. In this manner, it is possible

to produce a set of loadings that represent the variations of absorbances that are

common throughout the entire set. The matrices that result from the fully processed

calibration are shown in the simplified matrix expression of the model equation that

follows:

$$\mathbf{A}_{(i,j)} = \mathbf{S}_{(i,k)} \mathbf{L}_{(k,j)} + \mathbf{E}_{(i,j)}$$

where i is the number of calibration spectra, j is the number of data points used for the

calibration, and k is the number of loadings. It follows that $\mathbf{A}$ is an $i \times j$ matrix of

spectral absorbances, $\mathbf{S}$ is an $i \times k$ matrix of scores, $\mathbf{L}$ is a $k \times j$ matrix of loadings, and $\mathbf{E}$

is an $i \times j$ matrix of residual spectra, that is, the error in the ability of the model to

predict the calibration absorbances. Several algorithms have been developed to

calculate the loadings from a set of data. Two common ones are Decomposition of

Covariance and Non-Linear Iterative Partial Least Squares. Detailed descriptions of

these and other algorithms, such as Single Value Decomposition and Successive

Average Orthogonalization, may be found elsewhere in the literature.[73,92] In general,

the Decomposition of Covariance algorithm is the fastest to calculate, but often produces numerical errors, and, therefore, the Non-Linear Iterative Partial Least Squares algorithm, which is more robust, is widely utilized and is typically the one that is chosen for most commercial software packages that support continuum regression statistical methods.

The above equation is reminiscent of the Classical Least Squares model equation in that the entire spectrum may be included in the model. The scores and loadings, however, are used in place of the concentration and absorptivity constant matrices. Since the concentration matrix has not played a role in the model calculation, Principle Components Analysis cannot be used unaccompanied for prediction. The loadings, which represent spectral variations common to all of the calibration spectra, are employed to calculate a regression model from which constituent concentrations may be predicted. The loadings in the **L** matrix cannot be used to represent the original data without the scores matrix, **S**. The scores in the **S** matrix are unique to each spectrum, and they correlate to a given set of loadings. Therefore, it is possible to perform a regression of the concentration matrix **C** directly against the scores matrix, **S,** by:

$$\mathbf{C}_{(h,i)} = \mathbf{B}_{(h,k)}\,\mathbf{S}_{(k,i)} + \mathbf{E}_{(h,i)}$$

where i is the number of calibration spectra, h is the number of constituents used for the calibration, and k is the number of loadings. It follows that **C** is an $h \times i$ matrix of constituent concentrations, **B** is an $h \times k$ matrix of the regression coefficients, **S** is a $k \times i$ matrix of scores from the Principle Components Analysis model, and **E** is an $h \times i$

matrix of residuals. As in the Multiple Linear Regression method, the coefficient
matrix can be solved by:

$$\mathbf{B}_{(h,k)} = \mathbf{C}_{(h,i)}(\mathbf{S}'_{(k,i)}\mathbf{S}_{(k,i)})^{-1}\mathbf{S}'_{(k,i)}$$

The name, Principle Components Regression, comes from these two steps, which
combine Principle Components Analysis and Multiple Linear Regression, to calculate
the calibration equations. A single unified equation to represent the Principle
Components Regression can be produced by rearrangement of the matrix model
equation so that the scores are a function of the spectral absorbances and loadings:

$$\mathbf{S}_{(k,i)} = \mathbf{A}_{(i,j)}\mathbf{F}'_{(k,j)}$$

$\mathbf{F}$ is an orthonormal matrix in that the product of itself and its transpose is the identity
matrix. Therefore, it is not necessary to use the inverse of $\mathbf{F}$ to solve this equation. The
final equation emerges when the concentration equation and scores equation are
combined:

$$\mathbf{C}_{(h,i)} = \mathbf{B}_{(h,k)}\mathbf{A}_{(i,j)}\mathbf{F}'_{(k,j)} + \mathbf{E}_{(h,i)}$$

where $\mathbf{C}$ is an $h \times i$ matrix of constituent concentrations, $\mathbf{B}$ is an $h \times k$ matrix of the
regression coefficients, $\mathbf{A}$ is a $i \times j$ matrix of spectral absorbances, and $\mathbf{F}$ is a $k \times j$ matrix
of loadings. In addition, it is typical in the second step to add an extra unit vector
column to the scores matrix to accommodate the inclusion of an offset coefficient in the
regression.[66]

There are, however, several drawbacks to this method. Although it is presumed
that the variations in the spectral data are the result of variations in the constituents of
interest, there is no guarantee of a direct correspondence. Furthermore, the predictive
ability of this model will be adversely affected if there are collinear constituent

concentrations. This means that if a relatively large number of calibration samples must be incorporated into the model, then precautions must be taken to ensure that the constituent concentrations in the training samples are adequately randomized.[87]

Another spectral decomposition method, and the one that is used primarily in this study, is Partial Least Squares. It is similar to Principle Components Regression except that, rather than decomposition of the spectra into loadings and scores and then, as a separate step, regression against the concentration information, the concentrations are introduced into the decomposition process. As a result, spectra that contain higher constituent concentrations are more heavily weighted than those of low concentrations are weighted. The incentive is to incorporate more concentration information into the first few loadings.

This inclusion brings about two separate sets of scores and loadings. One set is for the spectral data scores $\mathbf{S}$ and for the loadings $\mathbf{L_S}$ which represent the common variations in the spectra. The other set is for the concentration data scores $\mathbf{T}$ and for the loadings $\mathbf{L_T}$, which represent the variations in the spectra that correspond to the regression components. A calibration model is constructed by relation of the two sets of scores to each other. This is done as a single step, which is unlike the two-step process used in the Principle Components Regression method. Partial Least Squares performs the spectral decomposition and the concentration data decomposition simultaneously, and, as each loading is calculated, the scores are exchanged before the contribution of the loading is removed from the data. Following this, the reduced matrices are used to calculate the next loading, and the entire process is repeated iteratively until the desired number of loadings is produced. The model equations

which result are considerably complex mathematically, and, as they are beyond the scope of this discussion, the algorithms used for the calculation of the Partial Least Squares model may be found elsewhere in the literature.[74,86,93-98]

The primary advantage of Partial Least Squares is that the spectral vectors are directly related to the concentration of the constituents of interest. Two forms of Partial Least Squares exist: PLS-1 and PLS-2. The difference between these is that, while PLS-1 calculates a separate set of scores and loadings for each constituent, PLS-2 calibrates for all of the constituents simultaneously. PLS-1, for the most part, is more robust than PLS-2 because a model generated by PLS-2 cannot be optimized for each constituent individually. The advantage of PLS-1 is more pronounced when the ranges of constituent concentrations are disparate. The calculation time requirement with respect to PLS-1 relative to PLS-2, however, increases dramatically as the number of explicitly modeled constituents increases.

Although Partial Least Squares has generally been shown to outperform Principle Components Regression, particular care must be taken to avoid collinear constituent concentrations. Furthermore, it is essential that the calibration set reflect the range of concentration variability expected in unknown samples. Both, however, provide the advantage of permitting overdetermination of the data as with Classical Least Squares by including full spectral coverage; they also produce the advantage of partial composition regression provided by Multiple Linear Regression. Partial Least Squares may be used for highly complex mixtures, as not only must solely the constituents of interest be known to apply the calibration, but also, in some cases, predictions on samples that include contaminants not found in the training samples may

be undertaken. As no definitive guidelines exist with regard to selection of the

calibration technique best suited for a particular system, reasonable intuition and

conventional wisdom often may be the only tools available for that determination.

References

1.   Kennedy, J.; Pagliuca, G. *Carbohydrate Analysis: A Practical Approach*; Chaplin, M.; Kennedy, J., Eds.; Oxford: England, 1994, pp 43-68.

2.   El Khadem, H. *Carbohydrate Chemistry*; Academic Press, Inc.: New York, 1988.

3.   Boddy, M.; Martin, P. *Journal of the Institute of Brewing* **1971**, *77* (3), 245.

4.   Shen, L.; Ginsburg, V. *Federation Proceedings* **1967**, *26* (2), 281.

5.   Lee, K.; Loganathan, D.; Merchant, Z.; Linhardt, R. *Applied Biochemistry and Biotechnology* **1990**, *23* (1), 53-80.

6.   Schmiedl, D.; Bauerlein, M.; Bengs, H.; Jacobasch, G. *Carbohydrate Polymers* **2000**, *43* (2), 183-193.

7.   Rampon, V.; Robert, R.; Nicolas, N.; Dufour, E. *Journal of Food Science* **1999**, *64* (2), 313-316.

8.   Sivakesava, S.; Irudayaraj, J. *Applied Engineering in Agriculture* **2000**, *16* (5), 543-550.

9.   Williams, P.; Cordeiro, H.; Harnden, M. *Cereal Foods World* **1991**, *36* (7), 571-574.

10.   Barker, S. *Journal of the Association of Public Analysts* **1978**, *16* (JUN), 45-48.

11.   Liebmann, L.; Robinson, J.; Mann, K. *Review of Scientific Instruments* **1991**, *62* (9), 2083-2092.

12.   Hirata, A.; Komoda, M.; Itoh, W.; Tabata, K.; Itoyama, S.; Sugawara, I. *Biological & Pharmaceutical Bulletin* **1994**, *17* (11), 1437-1440.

13.   Mattu, M.; Small, G.; Arnold, M. *Appl. Spectrosc.* **1997**, *51* (9), 1369-1376.

14.   Zhang, J.; Tian, G. *Progress in Biochemistry and Biophysics* **1998**, *25* (2), 114-119.

15.   Huang, Y.; Mechref, Y.; Novotny, M. *Carbohydrate Research* **2000**, *323* (1-4), 111-125.

16.   Orberger, G.; Geyer, R.; Stirm, S.; Tauber, R. *European Journal of Biochemistry* **1992**, *205* (1), 257-267.

17.   Talaga, P.; Bellamy, L.; Moreau, M. *Vaccine* **2001**, *19* (20-22), 2987-2994.

18.   Ooi, V.; Liu, F. *Current Medicinal Chemistry* **2000**, *7* (7), 715-729.

19.   Tsuzuki, A.; Tateishi, T.; Ohno, N.; Adachi, Y.; Yadomae, T. *Bioscience Biotechnology and Biochemistry* **1999**, *63* (1), 104-110.

20.   Hashimoto, T.; Ohno, N.; Yadomae, T. *Drug Development Research* **1997**, *42* (1), 40-48.

21.   Bousfield, G.; Baker, V.; Gotschall, R.; Butnev, V.; Butnev, V. *Methods-A Companion to Methods in Enzymology* **2000**, *21* (1), 15-39.

22.   Fujii, T.; Miyahara, Y. *Appl. Spectrosc.* **1998**, *52* (1), 128-133.

23.   Stoolman, L. *Cell* **1989**, *56* (6), 907-910.

24.   Brandley, B.; Swiedler, S.; Robbins, P. *Cell* **1990**, *63* (5), 861-863.

25.   Watson, S.; Fennie, C.; Lasky, L. *Nature* **1991**, *349* (6305), 164-166.

26.   Cohenford, M.; Qiao, L.; Rigas, B. *Gastroenterology* **1995**, *108* (4), 457A-457A.

27. Yano, K.; Ohoshima, S.; Gotou, Y.; Kumaido, K.; Moriguchi, T.; Katayama, H. *Anal. Biochem.* **2000**, *287* (2), 218-225.

28. Riebe, D.; Thorn, W. *Electrophoresis* **1991**, *12* (4), 287-293.

29. Sakagami, H.; Takeda, M. *International Journal of Oncology* **1992**, *1* (3), 283-287.

30. Paulus, A.; Klockow, A. *J. Chromatogr. A* **1996**, *720* (1-2), 353-376.

31. Starr, C. M.; Masada, R. I.; Hague, C.; Skop, E.; Klock, J. C. *J. Chromatogr. A* **1996**, *720* (1-2), 295-321.

32. Packer, N.; Lawson, M.; Jardine, D.; Redmond, J. *Glycoconjugate Journal* **1998**, *15* (8), 737-747.

33. Laine, R. A. *Glycobiology* **1994**, *4* (6), 759-767.

34. Kornfeld, R.; Kornfeld, S. *Annual Review of Biochemistry* **1985**, *54*, 631-664.

35. Kobata, A. *European Journal of Biochemistry* **1992**, *209* (2), 483-501.

36. Tarentino, A.; Gomez, C.; Plummer, T. *Biochemistry* **1985**, *24* (17), 4665-4671.

37. Mizuochi, T.; Amano, J.; Kobata, A. *Journal of Biochemistry* **1984**, *95* (4), 1209-1213.

38. Plummer, T.; Elder, J.; Alexander, S.; Phelan, A.; Tarentino, A. *J. Biol. Chem.* **1984**, *259* (17), 700-704.

39. Pretoriu, V.; Hopkins, B.; Schieke, J. *J. Chromatogr.* **1974**, *99* (NOV6), 23-30.

40. Rassi, Z.; Nashabeh, W. In *Journal of Chromatography Library Vol 58*; Rassi, Z., Ed.; Elsevier: New York, 1995, pp 267-355.

41. Roche, M.; Oda, R.; Landers, J. *Biotechnology Progress* **1997**, *13* (5), 659-668.

42. Simms, P.; Haines, R.; Hicks, K. *J. Chromatogr.* **1993**, *648* (1), 131-137.

43. Oberdorfer, F.; Kemper, K.; Kaleja, M.; Reusch, J.; Gottschall, K. *J. Chromatogr.* **1991**, *552* (1-2), 483-487.

44. Mirouze, F.; Boulou, J.; Dupuy, N.; Meurens, M.; Huvenne, J.; Legrand, P. *Appl. Spectrosc.* **1993**, *47* (8), 1187-1191.

45. Suzuki, M.; Sakamoto, R.; Aoyagi, T. *Tappi Journal* **1995**, *78* (7), 174-177.

46.   Solbriglebuhn, H. *Zuckerindustrie* **1992**, *117* (12), 979-983.

47.   Deruiter, G.; Schols, H.; Voragen, A.; Rombouts, F. *Anal. Biochem.* **1992**, *207* (1), 176-185.

48.   Vorndran, A.; Oefner, P.; Scherz, H.; Bonn, G. *Chromatographia* **1992**, *33* (3-4), 163-168.

49.   Garner, T.; Yeung, E. *J. Chromatogr.* **1990**, *515*, 639-644.

50.   Oefner, P.; Chiesa, C. *Glycobiology* **1994**, *4* (4), 397-412.

51.   Delaney, J.; Vouros, P. *Rapid Communications in Mass Spectrometry* **2001**, *15* (5), 325-334.

52.   Pons, A.; Popa, J.; Portoukalian, J.; Bodennec, J.; Ardail, D.; Kol, O.; Martin-Martin, M. J.; Hueso, P.; Timmerman, P.; Leroy, Y.; Zanetta, J. P. *Anal. Biochem.* **2000**, *284* (2), 201-216.

53.   Bendiak, B.; Fang, T. T. *Carbohydrate Research* **2000**, *327* (4), 463-481.

54.   Leskovsek, H.; Perko, S.; Zigon, D.; Faganeli, J. *Analyst* **1994**, *119* (6), 1125-1128.

55.   Harvey, D. J. *Mass Spectrometry Reviews* **1999**, *18* (6), 349-450.

56.   Bendiak, B. *Carbohydrate Research* **1999**, *315* (3-4), 206-221.

57.   Taravel, F. R.; Mazeau, K.; Tvaroska, I. *Brazilian Journal of Medical and Biological Research* **1995**, *28* (7), 723-732.

58.   van Kuik, J.; Hard, K.; Vliegenthart, J. *Carbohydrate Research* **1992**, *235*, 53-68.

59.   van Halbeek, H. *Current Opinion in Structural Biology* **1994**, *4* (5), 697-709.

60.   Laine, R.;. *Glycosciences, Status and Perspectives*; Gabius, H.-J., Ed.; Chapman and Hall: London, 1997, pp 1-11.

61.   Gallignani, M.; Garrigues, S.; Delaguardia, M. *Anal. Chim. Acta* **1994**, *287* (3), 275-283.

62.   Kameoka, T.; Okuda, T.; Hashimoto, A.; Noro, A.; Shiinoki, Y.; Ito, K. *Journal of the Japanese Society for Food Science and Technology-Nippon Shokuhin Kagaku Kogaku Kaishi* **1998**, *45* (3), 192-198.

63.   Ferrer, N. *Mikrochim. Acta* **1997**, *14* (Supplement), 329-332.

64.   Nakano, T.; Kawata, S. *Scanning* **1994**, *16* (6), 368-371.

65.   Haaland, D.; Jones, H.; Thomas, E. *Appl. Spectrosc.* **1997**, *51* (3), 340-345.

66.   Griffiths, P. R.; de Haseth, J. A. "Fourier Transform Infrared Spectrometry"; 18;
        John Wiley & Sons: New York, 1986.

67.   Perkin-Elmer. *I-Series FT-IR Microscopy User's Guide*, 1994.

68.   Ramos, L.; Beebe, K.; Carey, W.; Sanchez, E.; Erickson, B.; Wilson, B.;
        Wangen, L.; Kowalski, B. *Anal. Chem.* **1986**, *58* (5), R294-R315.

69.   Beebe, K.; Kowalski, B. *Anal. Chem.* **1987**, *59* (17), A1007-A1017.

70.   Haaland, D.; Easterling, R. *Appl. Spectrosc.* **1982**, *36* (6), 665-673.

71.   Haaland, D.; Easterling, R. *Appl. Spectrosc.* **1980**, *34* (5), 539-548.

72.   Brown, C.; Lynch, P.; Obremski, R.; Lavery, D. *Anal. Chem.* **1982**, *54* (9),
        1472-1479.

73.   Goicoechea, H.; Olivieri, A. *Chemometrics and Intelligent Laboratory Systems*
        2001, *56* (2), 73-81.

74.   Spiegelman, C.; Bennett, J.; Vannucci, M.; McShane, M.; Cote, G. *Anal. Chem.* **2000**, *72* (1), 135-140.

75.   Haaland, D.; Han, L.; Niemczyk, T. *Appl. Spectrosc.* **1999**, *53* (4), 390-395.

76.   Diaz, T.; Guiberteau, A.; Burguillos, J.; Salinas, F. *Analyst* **1997**, *122* (6), 513-517.

77.   Tauler, R.; Izquierdoridorsa, A.; Casassas, E. *Anales de Quimica* **1991**, *87* (5), 571-579.

78.   Haaland, D.; Thomas, E. *Anal. Chem.* **1988**, *60* (11), 1193-1202.

79.   Haaland, D.; Thomas, E. *Anal. Chem.* **1988**, *60* (11), 1202-1208.

80.   Thomas, E. *Anal. Chem.* **1994**, *66* (15), A795-A804.

81.   Haaland, D.; Thomas, E. *Journal of Vacuum Science & Technology A-Vacuum Surfaces and Films* **1988**, *6* (3), 1042-1043.

82.   Thomas, E.; Haaland, D. *Anal. Chem.* **1990**, *62* (10), 1091-1099.

83.   Brown, S. *Chemometrics and Intelligent Laboratory Systems* **1991**, *10* (1-2), 87-105.

84.   Nelder, J. *Journal of the Royal Statistical Society Series C-Applied Statistics* **1974**, *23* (3), 323-329.

85.   Marshall, T.; Chaffin, C.; Makepeace, V.; Hoffman, R.; Hammaker, R.; Fateley, W.; Saarinen, P.; Kauppinen, J. *J. Mol. Struc.* **1994**, *324* (1-2), 19-28.

86.   Schmidt, P.; Glombitza, B. *Trac-Trends in Analytical Chemistry* **1995**, *14* (2), 45-49.

87.   Navarrovilloslada, F.; Perezarribas, L.; Leongonzalez, M.; Polodiez, L. *Anal. Chim. Acta* **1995**, *313* (1-2), 93-101.

88.   Walmsley, A. *Anal. Chim. Acta* **1997**, *354* (1-3), 225-232.

89.   Winson, M.; Goodacre, R.; Timmins, E.; Jones, A.; Alsberg, B.; Woodward, A.; Rowland, J.; Kell, D. *Anal. Chim. Acta* **1997**, *348* (1-3), 273-282.

90.   Malinowski, E.; Howery, D. *Factor Analysis in Chemistry*; John Wiley & Sons: New York, 1980.

91. Stone, M.; Brooks, R. *Journal of the Royal Statistical Society Series b Methodological* **1990**, *52* (2), 237-269.

92. Almoy, T.; Haugland, E. *Appl. Spectrosc.* **1994**, *48* (3), 327-332.

93. Haaland, D.; Melgaard, D. *Appl. Spectrosc.* **2001**, *55* (1), 1-8.

94. Spiegelman, C.; McShane, M.; Goetz, M.; Motamedi, M.; Yue, Q.; Cote, G. *Anal. Chem.* **1998**, *70* (1), 35-44.

95. Rannar, S.; Geladi, P.; Lindgren, F.; Wold, S. *Journal of Chemometrics* **1995**, *9* (6), 459-470.

96. Bak, J.; Larsen, A. *Appl. Spectrosc.* **1995**, *49* (4), 437-443.

97. Salmain, M.; Varenne, A.; Vessieres, A.; Jaouen, G. *Appl. Spectrosc.* **1998**, *52* (11), 1383-1390.

98. Messick, N.; Kalivas, J.; Lang, P. *Microchem. J.* **1997**, *55* (2), 200-207.

CHAPTER 2

ANALYSIS OF N-LINKED OLIGOSACCHARIDES BY FOURIER

TRANSFORM INFRARED MICROSPECTROMETRY[*]

Abstract

Carbohydrates are an important class of biochemical compounds. Compositional analysis of mammalian N-linked oligosaccharides is performed routinely with gas chromatography and combined gas chromatography/mass spectrometry. Infrared spectrometry has not been used frequently in the study of oligosaccharides. Nonetheless, infrared spectrometry can be applied quite effectively because each unique structure, with the exception of optical isomers, has a corresponding distinct infrared spectrum. Even large sugars that only have minor structural differences produce distinct infrared spectra.

This investigation involves the rapid and accurate determination of monosaccharide composition in complex carbohydrates. Oligosaccharides are depolymerized in order for the individual monosaccharides to be quantified by Fourier transform infrared microspectrometry. The representative monosaccharides found in mammalian N-linked oligosaccharides are D-mannose, D-galactose, L-fucose, N-acetyl-D-glucosamine, N-acetyl-D-neuraminic acid, and N-acetyl-D-galactosamine. A common impurity is D-glucose and is therefore included in the study. A quantitative model was developed by the application of partial least squares regression methods to the spectra of mixtures of these seven monosaccharides, which mimic the composition of the oligosaccharides to be depolymerized. The standards and samples for prediction were subjected to acid methanolysis and peracetylation. Methanolysis was chosen over simple hydrolysis to protect N-acetyl-D-neuraminic acid. In addition, the peracetylated product is highly soluble in methylene chloride, which is an ideal solvent for direct deposition onto an infrared transparent window.

A representative set of samples was prepared and analyzed. Only the region from 1850 cm^{-1} to 850 cm^{-1} was used to build the partial least squares model. In addition, the spectra were baseline-corrected and normalized. Several regions for normalization were investigated in order to minimize the dependence of a particular region on any factors other than the thickness of the deposit. Leave-one-out predictions demonstrated the validity of the data set. The model was further validated by an external set of mixtures that were not incorporated into the original calibration model. Finally, predictions of unknown samples produced excellent results that surpassed the performance of the conventional gas chromatographic and combined gas chromatographic/mass spectrometric methods for analysis.

Introduction

A strong need exists for the development of methodologies for compositional analyses of complex carbohydrates.[1] Vibrational spectrometry is a valuable and powerful tool for examination of many chemical systems for both qualitative and quantitative analyses and is applicable to the analyses of complex carbohydrates. Such a method is presented.

Carbohydrates play numerous important roles in biochemical systems, and they are ubiquitous components in living organisms. Biologically, their role as a source of energy for physiological processes is essential. More profoundly, however, in addition to playing crucial roles in maintenance of structural integrity and biosynthesis, it has been found that oligosaccharides are required for the direction of these processes and also behave as chemical messengers.[2] It is the latter of these properties that guides this study towards the compositional and structural analysis of mammalian N-linked oligosaccharides. It should be noted that oligosaccharides are necessary for immune response and are present in more than fifty percent of mammalian proteins. Oligosaccharide extractions are costly, and, typically, only very small quantities are available for analysis. Therefore, a sensitive analytical method that can interrogate microgram or smaller quantities of sample is critical.[1]

The selected methodology involves the determination of the percent composition of each of seven different monosaccharides that occur in N-linked oligosaccharides. The monosaccharides of interest are D-mannose, D-galactose, L-fucose, N-acetyl-D-glucosamine, N-acetyl-D-neuraminic acid type VI, N-acetyl-D-galactosamine, and D-glucose. These seven were chosen as they, with the

exception of D-glucose, are the most common monosaccharide subunits encountered in the analysis of N-linked mammalian oligosaccharides. Glucose is included in the domain since it is an impurity that is often encountered in oligosaccharide assays. Structural representations of these seven monosaccharides are presented in Figures 2.1 and 2.2.

Several methods have been developed for compositional and structural analysis of complex carbohydrates, including gas chromatography/mass spectrometry[3-8] and nuclear magnetic resonance.[9-17] Although both of these methodologies are valuable tools for the determination of structure and composition for a wide range of compounds, each has inherent drawbacks. One disadvantage that both of these techniques have in common is the considerable expense of the instrumentation and the high level of difficulty in the development of automated routine analyses.[3]

Gas chromatography/mass spectrometry has been used routinely for analysis of oligosaccharides.[4-9,18] It is not, however, an ideal tool for these compositional analyses for several reasons. First, mass spectrometry cannot be used effectively without the separation of the components of interest prior to analysis.[3] For example, galactose and mannose, although structurally incongruent, are isobaric and, therefore, are not well-suited for distinction by mass spectrometry.[19] Gas chromatography requires derivatization of the analytes, which introduces absolute errors in these quantitative analyses as great as twenty percent. The errors may be attributed to the inability to control derivatization rates.

Nuclear magnetic resonance possesses several advantages in the areas where mass spectrometry is insufficient, such as its ability to distinguish compounds of

Figure 2.1   The structures of four of the seven monosaccharides.

CH=O
——OH
HO——
HO——
——OH
CH₂OH

D-Galactose

CH=O
HO——
HO——
——OH
——OH
CH₂OH

D-Mannose

CH=O
HO——
——OH
——OH
HO——
CH₃

L-Fucose

CH=O
——OH
HO——
——OH
——OH
CH₂OH
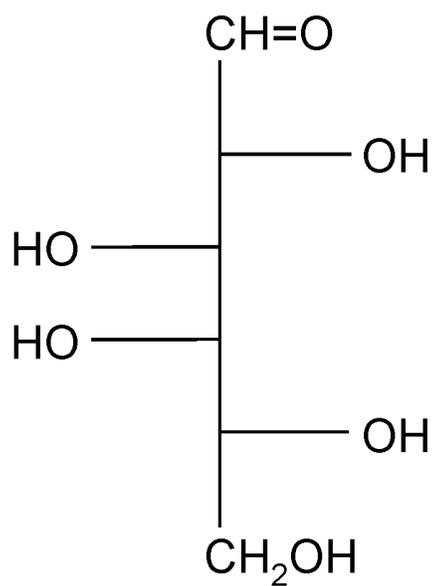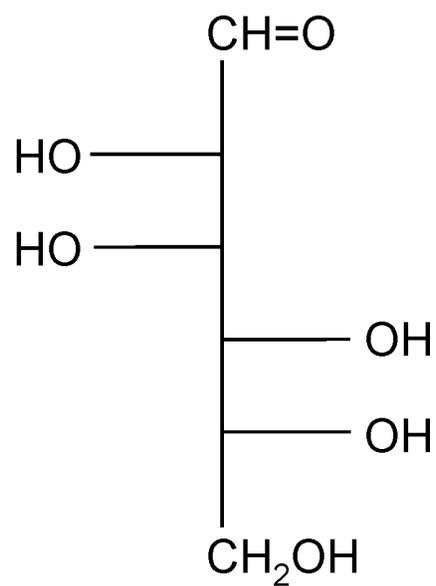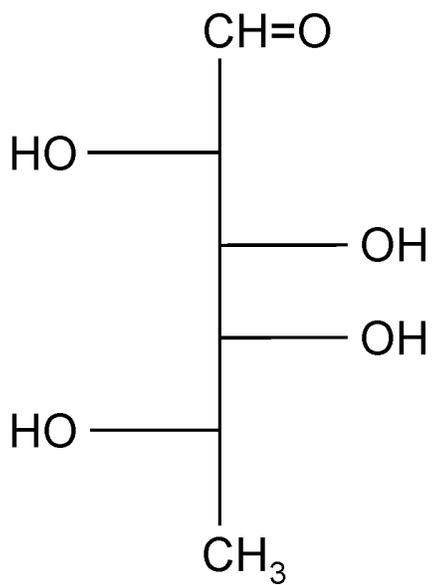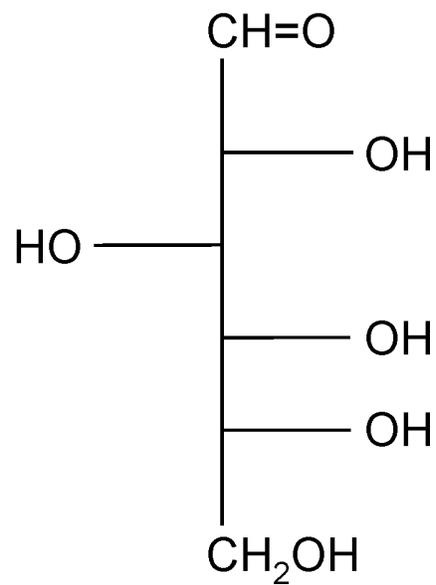
D-Glucose

Figure 2.2   The structures of the remaining three of the seven monosaccharides.

CH=O

NH(C=OCH₃)

HO

HO

OH

CH₂OH

N-Acetyl-D-Galactosamine

CH=O

NH(C=OCH₃)

HO

OH

OH

CH₂OH

N-Acetyl-D-Glucosamine

COOH

CH=O

HO

(CH₃C=O)HN

HO

OH

OH

CH₂OH
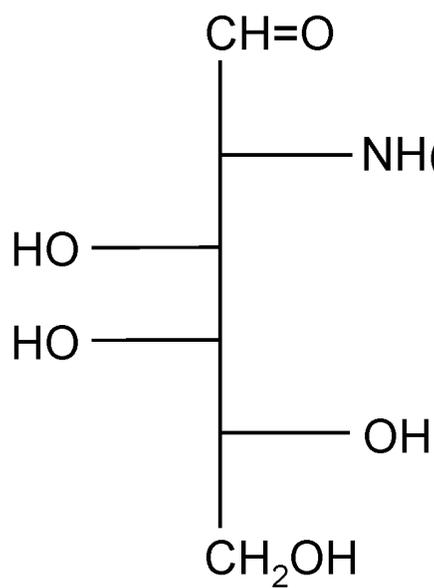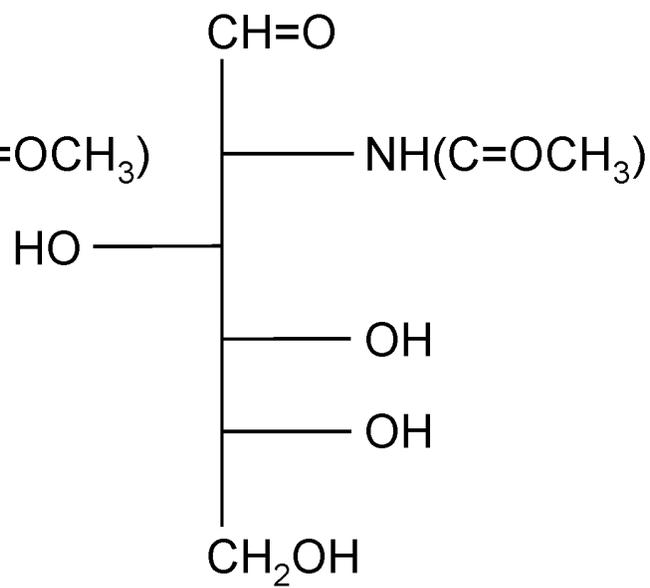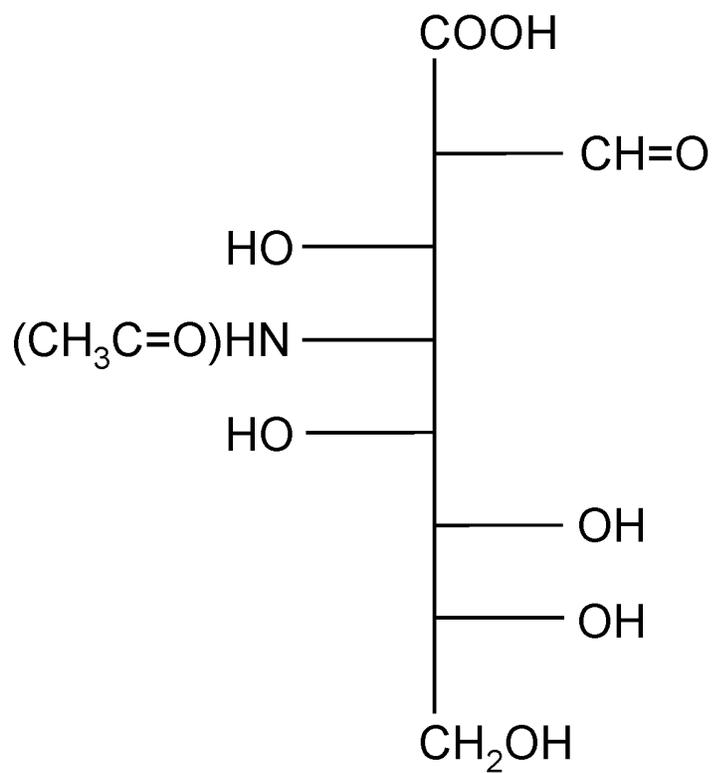
N-Acetyl-D-Neuraminic Acid

identical molecular weights; however, it poses other inherent challenges. First, and foremost, is its intrinsic insensitivity as a detector. Nuclear magnetic resonance requires greater sample mass than is economically feasible in order to obtain an adequate signal intensity. In addition, quantitative analysis, which requires reproducible spectral interrogation, is thwarted by the relative inability to maintain a homogeneous magnetic field in nuclear magnetic resonance. It is nearly impossible to obtain identical spectra from the same sample upon multiple acquisitions. Although nuclear magnetic resonance seldom poses a problem for qualitative analysis, reproducible data are a prerequisite for statistical analysis.[20,21]

In contrast, Fourier transform infrared spectrometry is a sensitive analytical tool with a wide variety of applications. Fourier transform infrared spectrometry can frequently identify chemical species unequivocally where other analytical techniques cannot provide definitive structural and compositional information. Furthermore, it is a rapid and simple technique, which makes it a valuable tool for process analysis and quality control. The instrumentation is relatively inexpensive and is stable enough to permit automated analyses.

Infrared spectra are highly reproducible and absorbance correlates to sample concentration. Therefore, Fourier transform infrared spectrometry is well-suited for use in quantitative analysis. Least squares regression is only useful for systems that involve mixtures of compounds that have few components and whose infrared spectra do not have overlapping absorption bands. Multivariate analysis, particularly partial least squares, works very well in conjunction with vibrational spectrometric analysis.

Fourier transform infrared spectrometry has received little attention for analysis of carbohydrates with the possible exception of sugar concentrations in food products and beverages.[22,23] Near-infrared spectrometry combined with multivariate quantitative analysis has been the recipient of considerable development in the last decade.[24-31] Only relatively recently has mid-infrared spectrometric analysis, in conjunction with partial least squares, begun to see greater acceptance for a wider range of applications in the literature.[32-37]

Complex carbohydrates have sufficiently similar spectra that visual interpretation of the spectra of a mixture of sugars is not possible. Minor variations in the spectra, however, are adequate in order for partial least squares to distinguish varying concentrations of constituents in a mixture reliably and with high precision. Unlike mass spectrometry, infrared spectrometry can differentiate among structural isomers, and, although it shares the inability to distinguish optical isomers, determination of optical isomers is not necessary for carbohydrate analysis in mammalian systems. In addition, as there is no interaction among the saccharide constituents of the system being analyzed, the quantitative methodology is reduced in complexity. As a result, the requirement with respect to the size of the standardization set necessary to build the multivariate calibration model is relaxed in relation to the size of a set in which the components would interact.

For the analysis of oligosaccharide composition, it is only requisite to determine the relative concentration of each of the seven common monosaccharide structures that are contained in the overall structure. If the oligosaccharides are depolymerized, then the mixture that results is a simple mixture of monosaccharides, which may be

considered as a simple mixture of seven separate constituents despite there being five potential forms of each monosaccharide. These five different forms are shown in Figure 2.3. This permits the design of a standardization set that can be produced from mixtures of commercially available monosaccharides. Such a sample set is required for partial least squares analysis to serve as the calibration model.

The standardization set is an unbiased representation of monosaccharide compositions that spans the expected range of relative monosaccharide content present in typical oligosaccharide structures. The spectra of all of the mixtures in the standardization set and the known concentrations of all of the constituents present in the mixtures serve as the basis for the partial least squares calibration model. Once the calibration model is built, it is tested first by a subset of the standardization set, then, it is tested for validation by an external sample set. If the oligosaccharide samples are depolymerized prior to analysis, then the model subsequently can be used for composition prediction.

An important aspect of method development for routine analysis is cost efficiency. Isolation of oligosaccharide samples is considerably expensive. Therefore, it is essential that the technique only require small sample quantities. To attain this goal, infrared microspectrometry is employed. This method permits analysis of microgram quantities of analytes; however, in order to use an infrared microscope system, several accommodations must be made.

Direct deposition is a suitable technique for infrared microspectrometric analysis; direct deposition of aqueous solutions of hygroscopic sugars, however, is not feasible. To circumvent this restriction, a modest adjustment in sample preparation can

Figure 2.3   The structures of the five possible forms in which monosaccharides

exist. Mannose is used in this example.
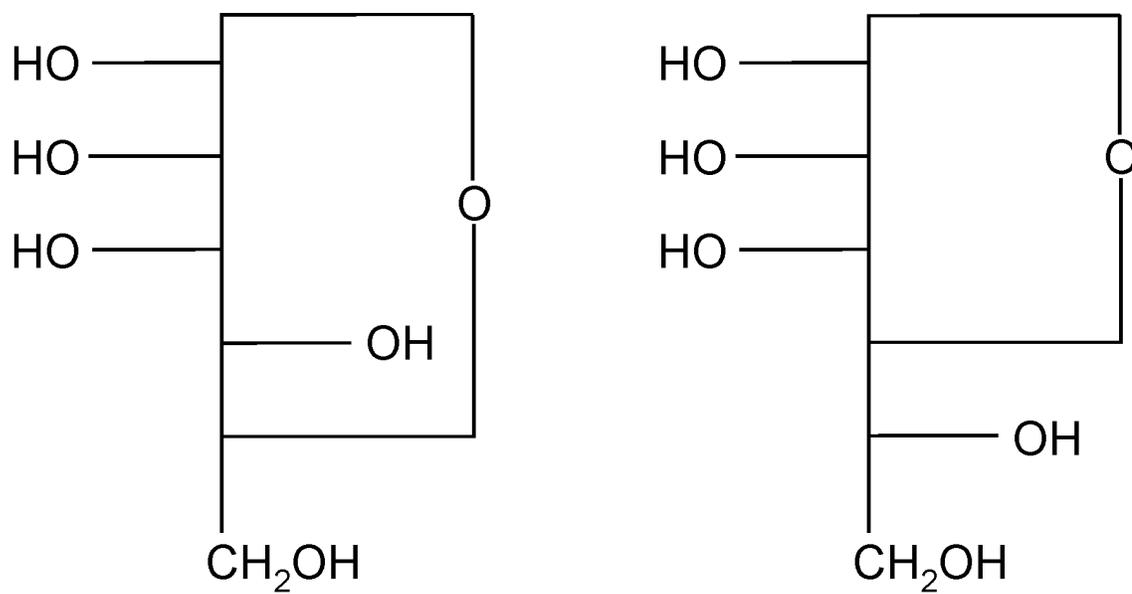
Open Chain Form of Mannose
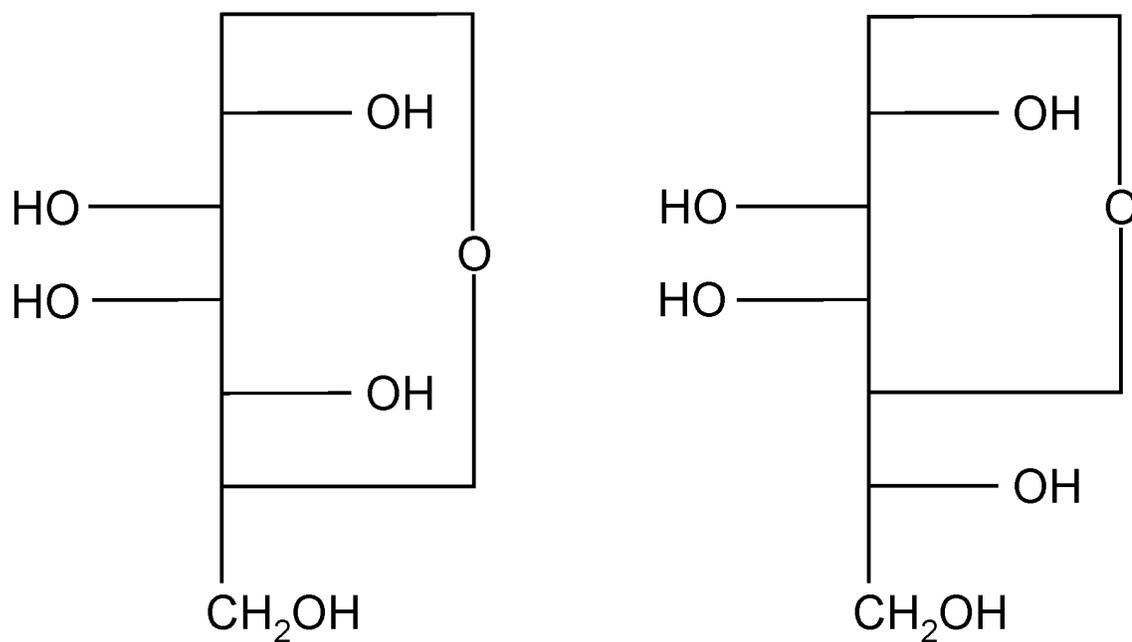
be implemented. Specifically, use of a highly volatile solvent system sufficiently addresses the problem of solvent elimination during sample deposition.

The preparation of the standards entails acid methanolysis and peracetylation to yield methyl glycosides. Note that in all cases where N-acetyl-D-neuraminic acid is concerned, the methanolysis reaction always produces a methyl ester of the C-1 carboxyl group, and the resulting methyl ester methyl glycoside is unaffected by the subsequent per-N,O-acetylation. In the interest of brevity, however, the products of all of the standards discussed in the present chapter as well as subsequent chapters will simply be referred to as "methyl glycosides" throughout. The structure of a methyl glycoside is shown in Figure 2.4. The objective of these processes is two-fold. Methanolysis will depolymerize the polysaccharides into their monosaccharide subunits. Furthermore, peracetylation effectively substitutes acetyl groups in the place of every hydroxyl group on the saccharides. The resulting methyl glycosides are considerably more hydrophobic, and thus, they are soluble in nonpolar organic solvents, which is preferable for direct deposition.

Experimental

Sample Preparation. Eighty-eight monosaccharide mixtures, which contain two to all seven constituents, were prepared as the standardization set, as well as an additional twenty monosaccharide mixtures for the validation set. The monosaccharide constituents are D-mannose, D-galactose, L-fucose, N-acetyl-D-glucosamine, N-acetyl-D-neuraminic acid type VI, N-acetyl-D-galactosamine, and D-glucose. The sugars and all of the other reagents and solvents were purchased from

Figure 2.4   The structures of the products of methanolysis and peracetylation.

Mannose is also used here as an example.
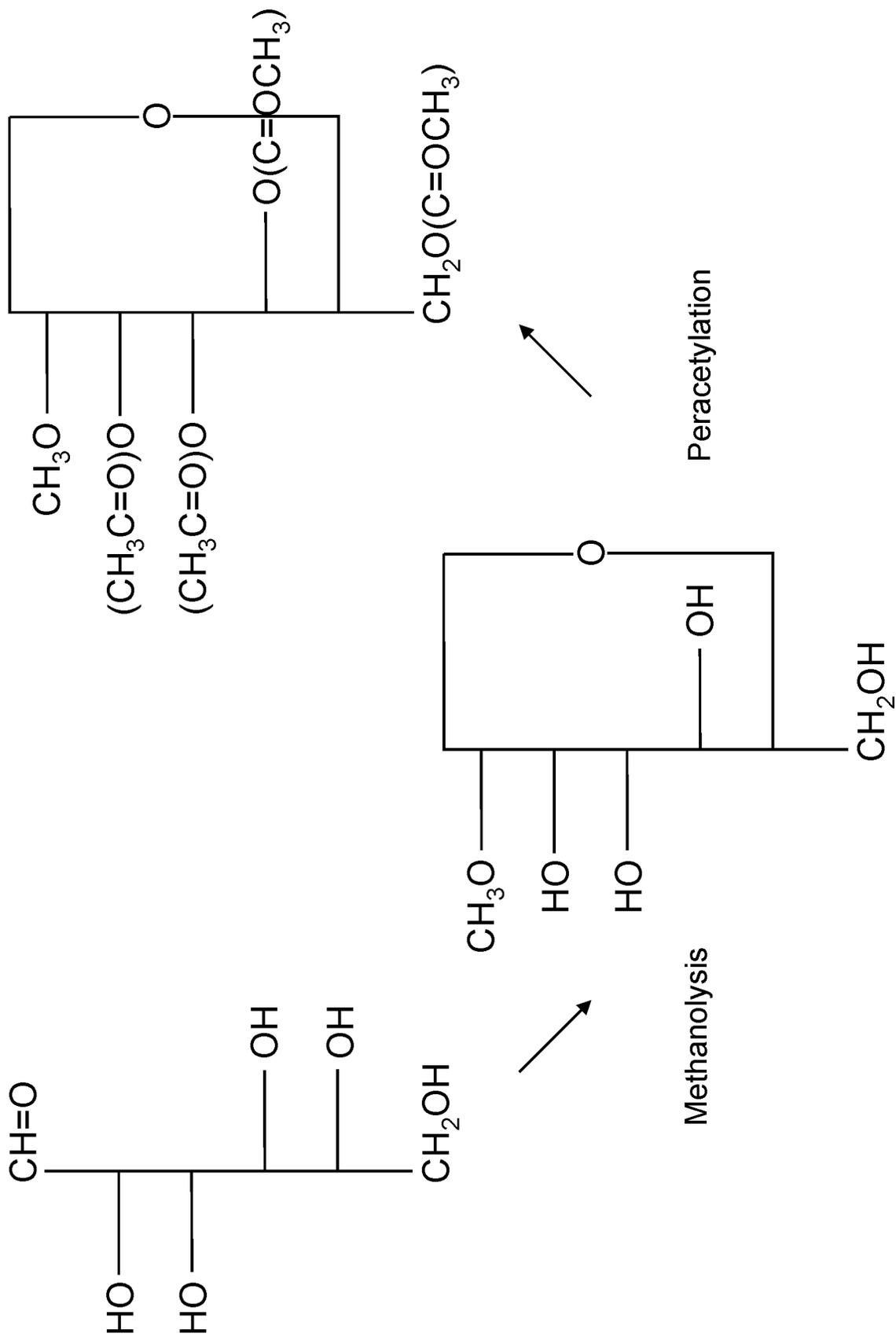
CH=O

HO

HO

OH

OH

CH₂OH

Methanolysis

CH₃O

HO

HO

OH

CH₂OH

Peracetylation

CH₃O

(CH₃C=O)O

(CH₃C=O)O

O(C=OCH₃)

CH₂O(C=OCH₃)

Sigma-Aldrich Co., St. Louis, MO, with the exception of the 3 N methanolic

hydrochloric acid which was purchased from Supelco, Bellefonte, PA. The weight

percentage ranges were selected to mimic the composition of mammalian N-linked

oligosaccharides that may be encountered in biological systems. Some of the ranges

reach beyond what is typically encountered to increase the robustness of the model,

however. Tables 2.1 through 2.4 list the specific composition for each of the

eighty-eight training set samples. The weights of individual sugars were assessed to

±0.05 mg precision on a high precision analytical balance (Ohaus Corporation, Florham

Park, NJ) and diluted with 18 MΩ water such that the resulting concentration is

4 mg/mL. The stocks were pipetted in the appropriate volumes to achieve the

prescribed composition for each of the standard mixtures. This was accomplished via a

series of Eppendorf™ Autoclavable Pipettes (Brinkman Instruments, Inc., Westbury,

New York), with which volumes of 10 – 100 μL, 20 – 200 μL, or 100 – 1000 μL can be

delivered. In all cases, each sample consisted of two milligrams of each mixture in

500 μL of water. The mixtures were contained in 17 mm x 63 mm borosilicate glass

vials with polytetrafluoroethylene-lined caps (Fisher Scientific, Pittsburgh, PA). The

vials were transported to and placed in a low-temperature freezer (–80 ºC). After

freezing, the samples were placed in a lyophilizer until the solvent was completely

removed.

The methanolysis and peracetylation procedures are as follows: Into each of the

samples, 1 mL of 1 N anhydrous methanolic hydrochloric acid was pipetted. Particular

care was taken to ensure that the vial caps were barely finger-tight immediately after

Table 2.1   The composition of calibration samples one through twenty-two.

Table 2.1

Sample	Gal	Man	Fuc	GlcNAc	Neu5	GalNAc	Glu
1	10%	15%	45%	4%	10%	6%	10%
2	20%	35%	25%	5%	0%	7%	8%
3	30%	55%	0%	6%	6%	0%	3%
4	40%	16%	0%	7%	20%	5%	12%
5	5%	36%	44%	8%	5%	2%	0%
6	0%	56%	24%	9%	7%	4%	0%
7	50%	17%	0%	10%	0%	20%	3%
8	8%	37%	0%	9%	12%	0%	34%
9	18%	57%	0%	8%	8%	4%	5%
10	28%	18%	23%	7%	4%	14%	6%
11	35%	38%	0%	6%	14%	0%	7%
12	7%	58%	0%	5%	17%	0%	13%
13	12%	19%	42%	4%	11%	12%	0%
14	14%	39%	22%	11%	0%	0%	14%
15	13%	59%	0%	12%	0%	16%	0%
16	45%	20%	0%	13%	18%	0%	4%
17	0%	40%	41%	6%	9%	0%	4%
18	0%	60%	21%	6%	2%	11%	0%
19	3%	21%	0%	16%	3%	34%	23%
20	0%	41%	0%	17%	23%	9%	10%
21	0%	61%	0%	18%	0%	0%	21%
22	0%	22%	20%	19%	2%	10%	27%

Table 2.2   The composition of samples twenty-three through forty-four.

Table 2.2

Sample	Gal	Man	Fuc	GlcNAc	Neu5	GalNAc	Glu
23	9%	42%	0%	20%	16%	13%	0%
24	2%	62%	0%	21%	0%	0%	15%
25	1%	23%	39%	22%	0%	15%	0%
26	0%	43%	19%	23%	15%	0%	0%
27	0%	63%	0%	24%	13%	0%	0%
28	6%	24%	0%	25%	0%	0%	45%
29	0%	44%	38%	5%	5%	6%	2%
30	0%	64%	18%	5%	0%	2%	11%
31	16%	25%	0%	28%	31%	0%	0%
32	5%	45%	0%	30%	0%	0%	20%
33	0%	65%	0%	5%	0%	5%	25%
34	17%	26%	17%	15%	25%	0%	0%
35	4%	46%	0%	14%	10%	10%	16%
36	0%	66%	0%	13%	15%	0%	6%
37	0%	27%	36%	12%	22%	0%	3%
38	0%	47%	16%	11%	0%	26%	0%
39	0%	67%	0%	10%	23%	0%	0%
40	23%	28%	0%	9%	0%	7%	33%
41	0%	48%	35%	8%	3%	0%	6%
42	0%	68%	15%	7%	0%	0%	10%
43	27%	29%	0%	6%	0%	35%	3%
44	0%	49%	0%	5%	29%	0%	17%

Table 2.3   The composition of samples forty-five through sixty-six.

Table 2.3

Sample	Gal	Man	Fuc	GlcNAc	Neu5	GalNAc	Glu
45	0%	69%	10%	4%	11%	6%	0%
46	14%	30%	14%	5%	25%	0%	12%
47	7%	50%	0%	6%	6%	9%	22%
48	9%	70%	0%	7%	7%	3%	4%
49	8%	31%	33%	8%	0%	6%	14%
50	0%	51%	13%	9%	16%	0%	11%
51	0%	71%	0%	10%	9%	6%	4%
52	21%	32%	0%	9%	10%	4%	24%
53	0%	52%	32%	8%	0%	3%	5%
54	0%	72%	12%	7%	0%	0%	9%
55	32%	33%	0%	6%	15%	14%	0%
56	11%	53%	0%	5%	11%	11%	9%
57	2%	73%	0%	4%	0%	5%	16%
58	8%	34%	31%	4%	23%	0%	0%
59	3%	54%	11%	4%	0%	28%	0%
60	0%	74%	0%	5%	0%	0%	21%
61	0%	75%	0%	5%	10%	5%	5%
62	0%	85%	0%	5%	0%	10%	0%
63	0%	90%	0%	5%	5%	0%	0%
64	0%	80%	8%	5%	0%	0%	7%
65	0%	70%	10%	6%	4%	0%	10%
66	0%	60%	0%	6%	0%	0%	34%

Table 2.4   The composition of samples sixty-seven through eighty-eight.

Table 2.4

Sample	Gal	Man	Fuc	GlcNAc	Neu5	GalNAc	Glu
67	36%	50%	0%	6%	8%	0%	0%
68	7%	40%	27%	7%	19%	0%	0%
69	19%	30%	11%	7%	0%	33%	0%
70	37%	20%	0%	7%	0%	26%	10%
71	9%	10%	0%	8%	12%	11%	50%
72	0%	15%	10%	8%	0%	50%	17%
73	22%	25%	5%	8%	37%	0%	3%
74	0%	35%	0%	8%	0%	0%	57%
75	0%	45%	0%	9%	46%	0%	0%
76	11%	55%	9%	9%	5%	0%	11%
77	0%	65%	4%	9%	0%	22%	0%
78	0%	75%	0%	10%	15%	0%	0%
79	0%	86%	0%	5%	0%	0%	9%
80	8%	17%	8%	11%	50%	0%	6%
81	1%	27%	3%	12%	0%	41%	16%
82	5%	37%	0%	13%	45%	0%	0%
83	3%	47%	0%	14%	0%	36%	0%
84	0%	57%	7%	16%	0%	0%	20%
85	0%	67%	2%	15%	0%	16%	0%
86	0%	77%	0%	5%	18%	0%	0%
87	0%	87%	0%	5%	0%	0%	8%
88	24%	14%	11%	12%	39%	0%	0%

addition of the reagent to maintain an anhydrous system by the minimization of its

exposure to atmosphere. The vials were heated for twenty minutes in order to soften the

polytetrafluoroethylene lining, after which the caps were further tightened. The

reaction was maintained in a thermostatic heat block at 80 ºC for 24 hours.

Both the acid elimination with nitrogen gas and the rinse of the remaining

mixtures with anhydrous methanol were executed in an N-evaporator (Organomation

Associates, Inc., Berlin, MA). Equal amounts of pyridine and acetic anhydride were

dispensed to each sample. The caps were retightened, and the reaction was permitted to

continue overnight at room temperature and in the absence of light. After the reaction

was complete, the reagents were eliminated, and the products were rinsed with toluene

in a manner congruent to the acid elimination and methanolic rinse. Finally, the excess

toluene was eliminated under low vacuum (viscous flow region), and the products were

stored in a low-temperature freezer (-16 ºC).

Data Acquisition and Computation. The methyl glycoside mixtures were

dissolved in 150 µL of methylene chloride. Calcium fluoride (CaF_2) windows (Spectral

Systems, Inc., Hopewell Junction, NY) were utilized as the infrared transparent

windows for the deposition method discussed in the introduction. For each deposit, a

window was heated by a two hundred watt incandescent lamp positioned approximately

20 cm above the surface. Additionally, a vacuum airflow was supplied about 10 mm

above the face of the window by a disposable pipette affixed to a rough pump by means

of Teflon tubing as shown in Figure 2.5. A 1-µL aliquot of the diluted product was

Figure 2.5   A drawing of the arrangement of the apparatuses used in direct

deposition.

Direct Deposition

Vacuum Pump

Incandescent Source

1 μL Micropipette

Sample Deposit on Infrared Transparent Window
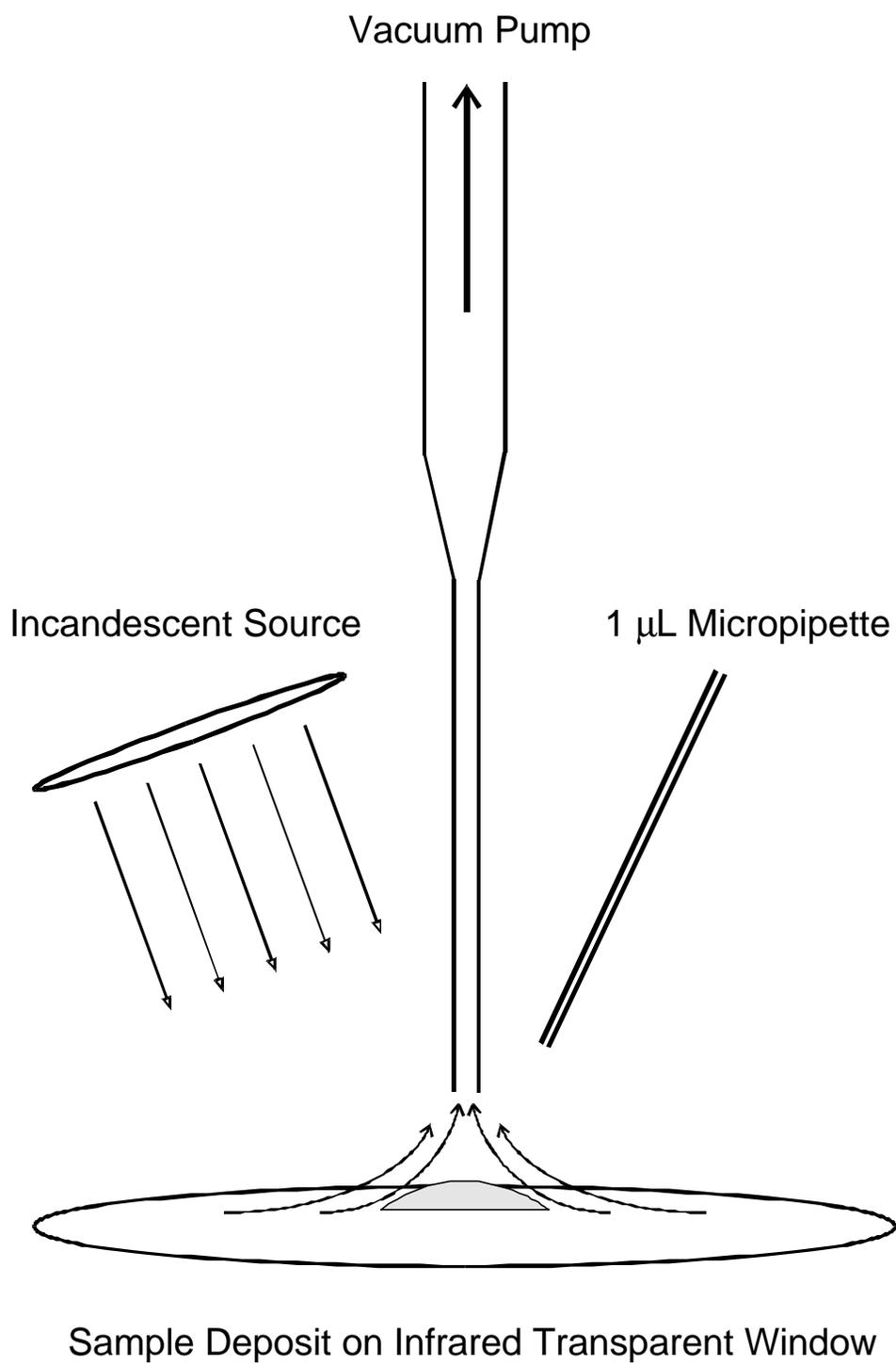
deposited, with the use of a micro capillary (Drummond Scientific Co., Broomall, PA), onto the window directly below the disposable pipette. The window, after deposition, was positioned onto the stage of a Perkin-Elmer *i*-Series infrared microscope that is coupled to a Spectrum 2000 Fourier transform infrared spectrometer (Perkin-Elmer, Norwalk, CT). The window was repositioned with the microscope stage control such that an appropriate 50 µm x 50 µm region of the deposit intersected the optical path. The aperture was set to contain a 50 µm x 50 µm area as well. A mercury cadmium telluride detector was chosen for detection, and the instrument collected transmission spectra from 4000 cm^{-1} to 700 cm^{-1} at 4 cm^{-1} resolution. Each 100-scan data set was Medium Norton-Beer apodized, Fourier transformed, and converted into absorbance in the respective sequence. This was realized for each of the samples, and the resulting spectra were truncated at 1850 cm^{-1} and 850 cm^{-1}. A two-point linear baseline-correction, incorporating both leveling and zeroing functions, was applied to each spectrum with 1850 cm^{-1} and 850 cm^{-1} designated as the reference points. Each spectrum was unit area normalized over the entire domain from 1850 cm^{-1} and 850 cm^{-1} before being placed into the training set. This was accomplished by integrating the area under the spectrum over the entire domain and rescaling such that the resulting area is equal to unity since only the relative concentrations are of interest.

A partial least squares algorithm, specifically PLS-1, was used to create a model from the eighty-eight training spectra. GRAMS PLSIQ™ version 5.2 (Galactic Industries Corporation, Salem, NH) was used for the statistical computations. The algorithm calculated iterations up to and including the possible eighty-five factors, utilized mean-centering, and performed a cross-validation adjusted to leave one

spectrum out at a time for each of the iterations. Selection of the appropriate number of factors for each of the constituents further optimized the final calibration model. Once the model was constructed, predictions were made on each of the internal spectra, the validation spectra, and, later, on a set of unknown spectra. The results, including predicted concentration, root-mean-square deviations, coefficients of determination, F-ratios, and prediction residual error sum of squares, were imported into a spreadsheet where error analysis was tabulated.

Results and Discussion

Although solvent elimination can be adequately achieved, merely ensuring complete solvent elimination is not the only challenge presented by direct deposition. For spectral quantitative analysis, it is necessary to suppress any deviations or artifacts in the spectra attributed to anything other than chemical composition. According to the Bougeur-Beer-Lambert Law, pathlength and concentration also affect absorption. In sample deposits, relative concentration variability is eliminated since only the pure analytes of interest are present. Pathlength, however, is governed by the deposit thickness as the spectra are measured by transmission. This variable is partially resolved by creating deposits that contain a pathlength gradient. This results in a non-uniform thickness of the deposit. A non-uniform thickness permits selection of an area of the deposit with the desired approximate thickness for spectral acquisition. Sample area selection, however, does not entirely provide pathlength reproducibility, as it is inaccurate and difficult to undertake. Correction for pathlength inconsistencies is essential and is provided by spectral unit area normalization over the region of interest.

Figures 2.6 and 2.7 show the infrared spectrum of a sample before and after
normalization.

A substantial parameter to be considered, when a multivariate calibration model
is constructed, is the spectral range. It is necessary to select the range such that the
highest correlation occurs between the absorbances in the spectra and the concentrations
of the constituents. In addition, it is not advisable to include regions that have weak or
no absorbance as this will result in a pronounced deficiency in the signal-to-noise ratio.
Another consideration for the selection of the spectral regions is the desire to avoid
selection of regions where the absorbance may be non-linear. In particular, although
PLS-1 is robust enough to correct for some non-linearity, selection of non-linear regions
is detrimental to the model. Monitoring the spectra in the training set during
interrogation, however, alleviates the problem of non-linearity via judicious selection of
the location within the sample deposit for analysis. The selected area must supply a
great enough absorbance signal to produce a maximal signal-to-noise ratio, yet the
signal should not be excessively intense where non-linearity is introduced into the
absorbance-to-concentration relationship. The carbonyl band is the strongest and is
within the included wavelength region. Therefore, particular care must be exercised to
ensure that too strong an absorption, i.e., greater than 0.5 absorbance units, does not
occur in the spectrum at 1650 cm^{-1} prior to normalization.[38-40]

For the set of training spectra, the region that exhibited the greatest correlation
between absorbance and concentration is the region between 1850 cm^{-1} to 850 cm^{-1}. A
two-point linear baseline-correction provides the best results. There are no bands that
absorb at either 1850 cm^{-1} or 850 cm^{-1}. Thus, these two points are most attractive to

Figure 2.6   A baseline-corrected and truncated spectrum of a peracetylated

monosaccharide mixture prior to normalization.
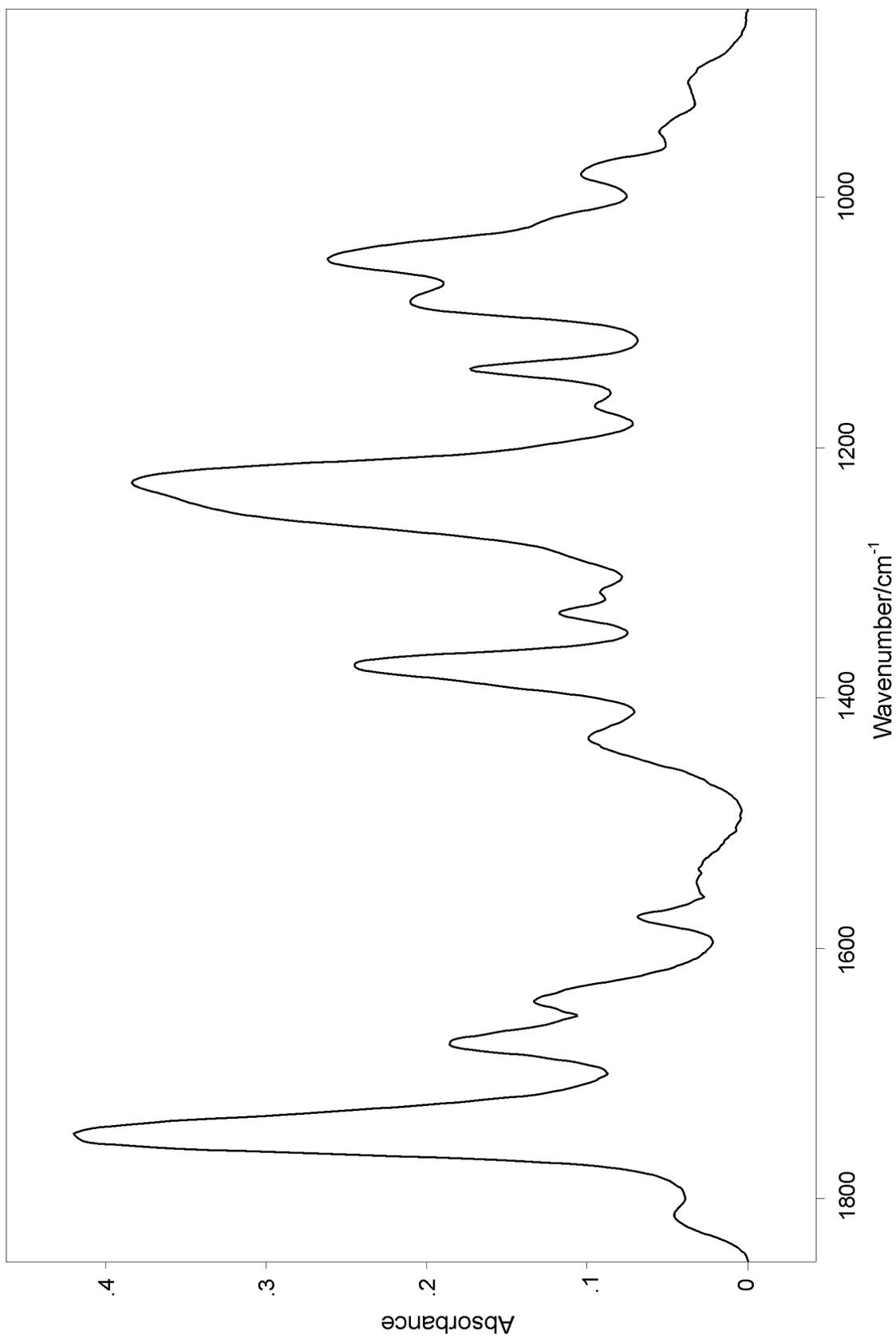
Figure 2.7   A baseline-corrected and truncated spectrum of a peracetylated

monosaccharide mixture after normalization.
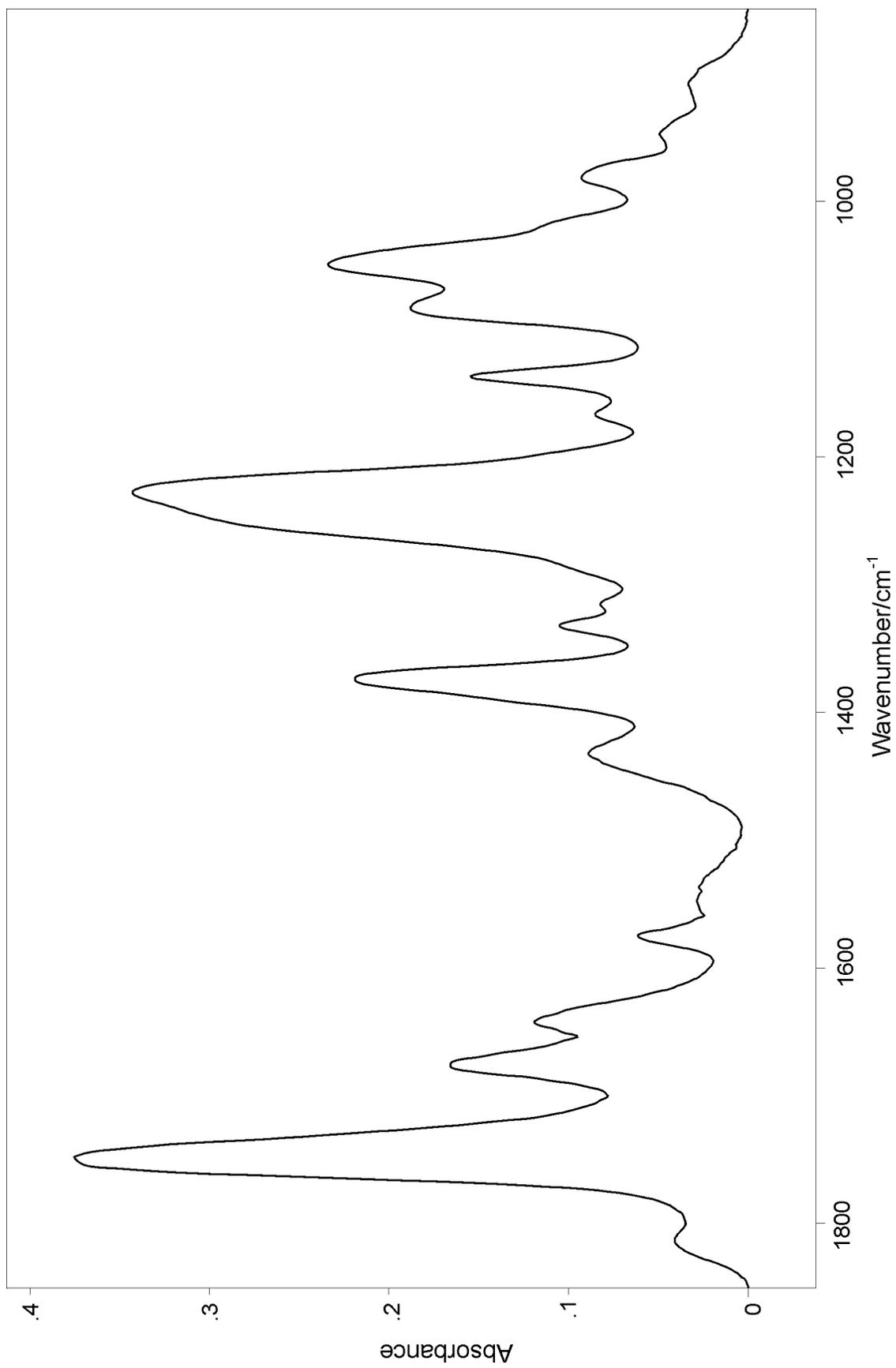
select for baseline-correction considering, as they both are the bandpass extrema, they

not only allow for the entire domain to be spanned, but also supply no chemical

information into their absorbance values. In addition to baseline-correction,

normalization demonstrates improvement in the model considerably. Although the

method of interrogation allows the operator to locate an appropriate area of the sample

that has a suitable thickness, neither this method nor the method of direct deposition is

sufficiently precise to ensure a consistent pathlength among the standard mixtures. The

improvement normalization provides is ascribed to the correction of the inconsistencies

of the pathlengths within the training set due to the inability to create reproducibly thick

deposits.

After the partial least squares model was built, a leave-one-out internal

cross-validation was performed. Table 2.5 presents the root-mean-square deviations for

each constituent. The prediction was derived from the model where the optimal number

of factors was assigned. Plots of the actual concentration versus predicted

concentration for each constituent are presented in Figures 2.8 through 2.14. The

optimal number of factors was determined by examining the F-ratio for every number

of factors, that is, value of f, as well as consideration of where the first local minima for

the values of the predicted residual error sums of squares occur. Figures 2.15

through 2.21 are plots of the number of factors versus the predicted residual error sum

of the squares for each constituent. The coefficients of determination, R^2, for each of

the constituents, D-galactose, D-mannose, L-fucose, N-acetyl-D-glucosamine,

N-acetyl-D-neuraminic acid, N-acetyl-D-galactosamine, and D-glucose, are 0.9514,

Table 2.5   The root-mean-squared deviations, coefficients of determination, and

number of factors for the seven monosaccharide constituents.

Table 2.5

Constituent	Root Mean Squared Deviation	Coefficient of Determination	Number of Factors
Gal	2.84%	0.9514	14
Man	3.39%	0.9567	13
Fuc	2.94%	0.9553	12
GlcNAc	1.75%	0.9400	17
Neu5	2.80%	0.9394	14
GalNAc	1.88%	0.9704	15
Glu	2.02%	0.9624	12

93

Figure 2.8   A plot of actual concentration versus predicted concentration for

D-galactose.

D-Galactose
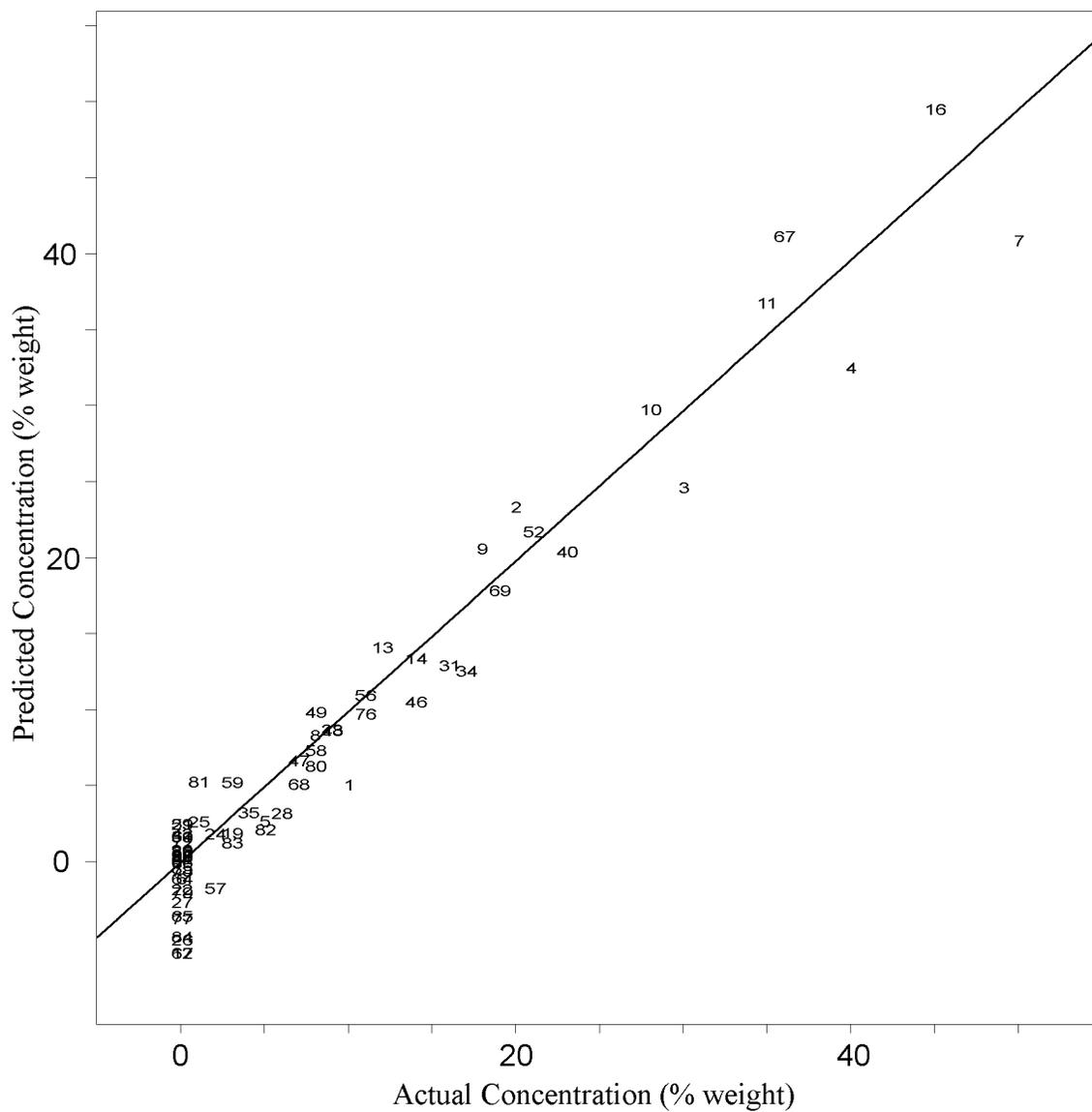
Figure 2.9   A plot of actual concentration versus predicted concentration for

D-mannose.

D-Mannose
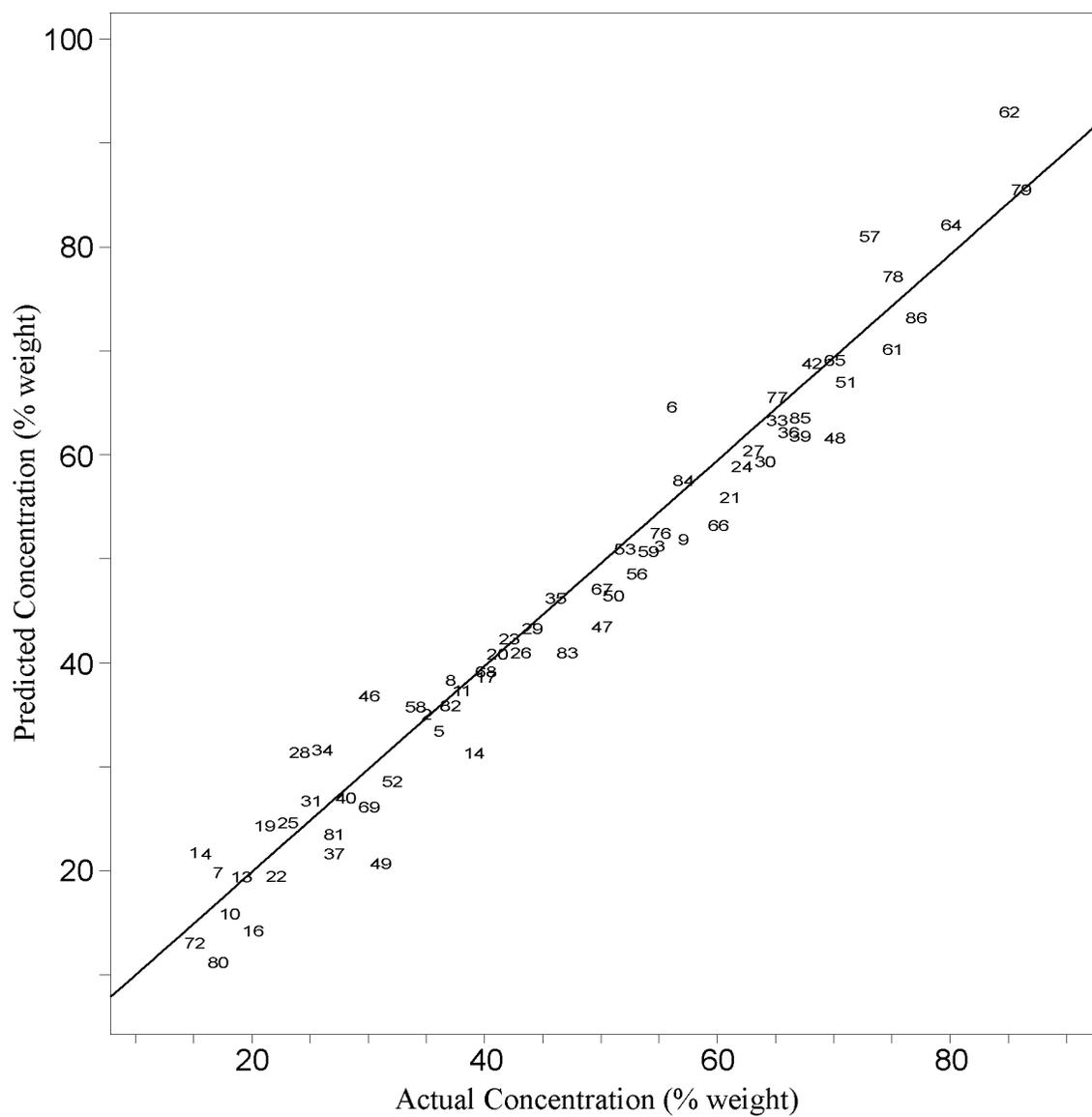
Figure 2.10   A plot of actual concentration versus predicted concentration for

L-fucose.

L-Fucose
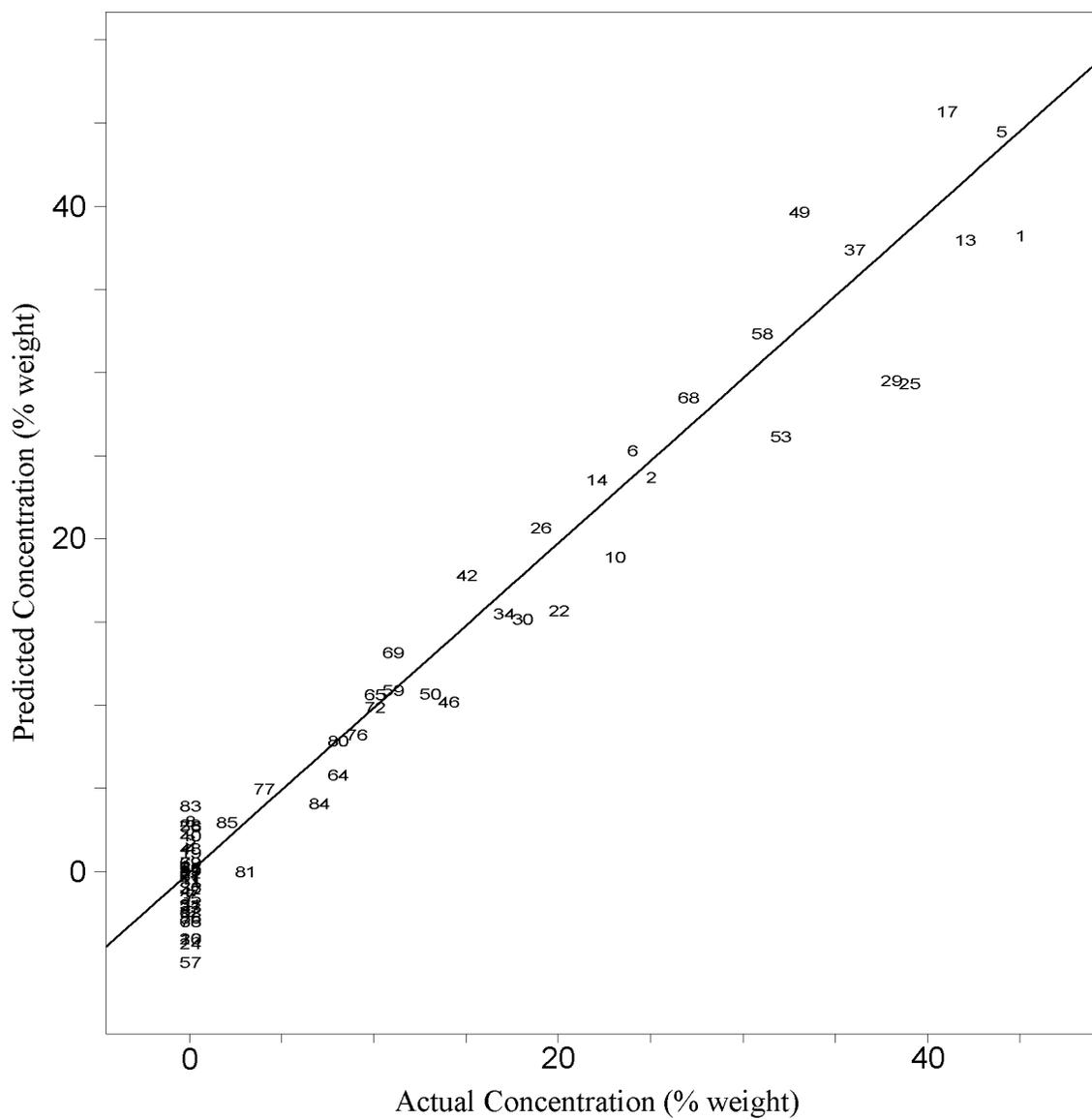
Figure 2.11   A plot of actual concentration versus predicted concentration for

N-acetyl-D-glucosamine.

N-Acetyl-D-Glucosamine
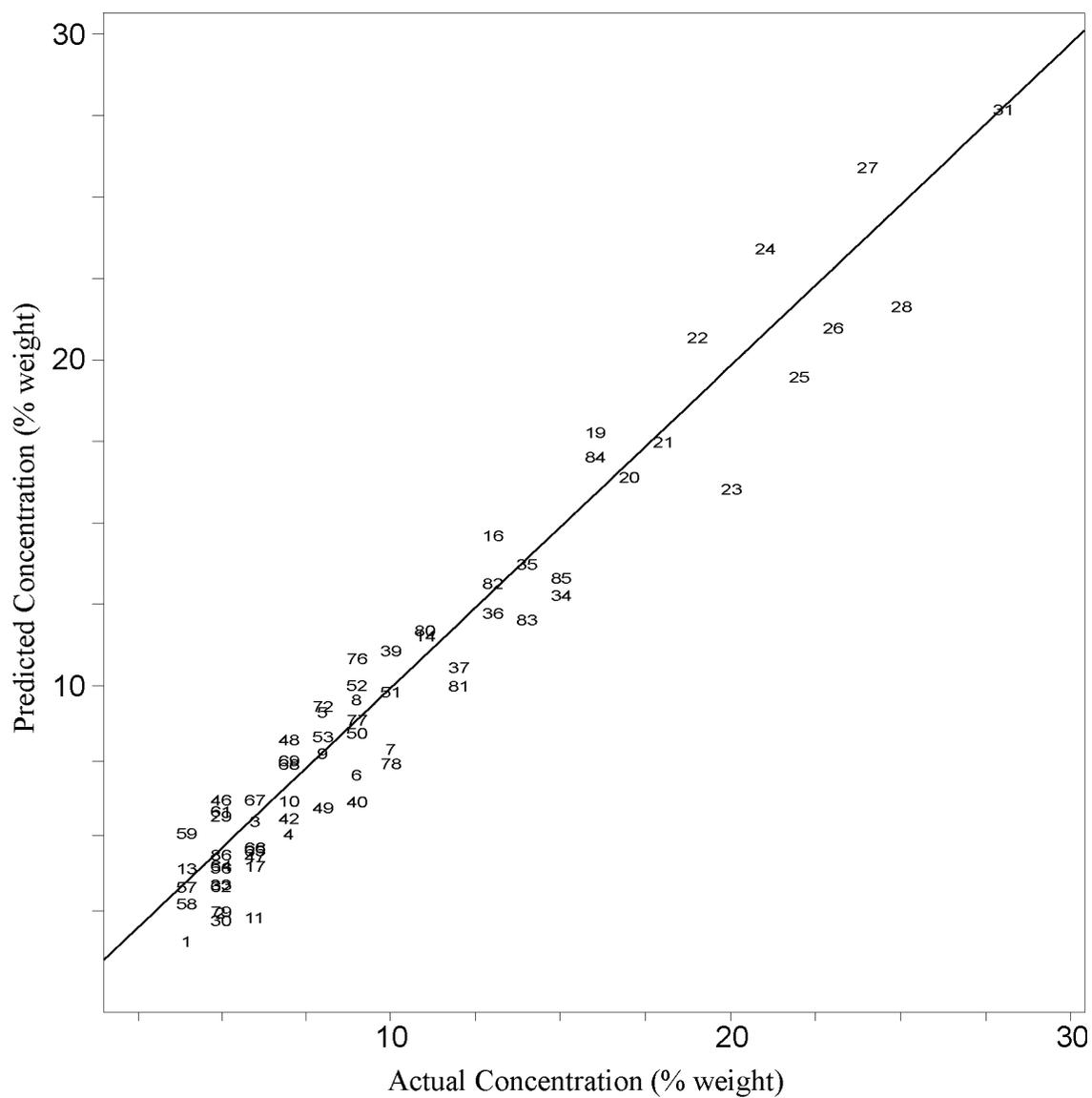
Figure 2.12   A plot of actual concentration versus predicted concentration for

N-acetyl-D-neuraminic acid.

N-Acetyl-D-Neuraminic Acid
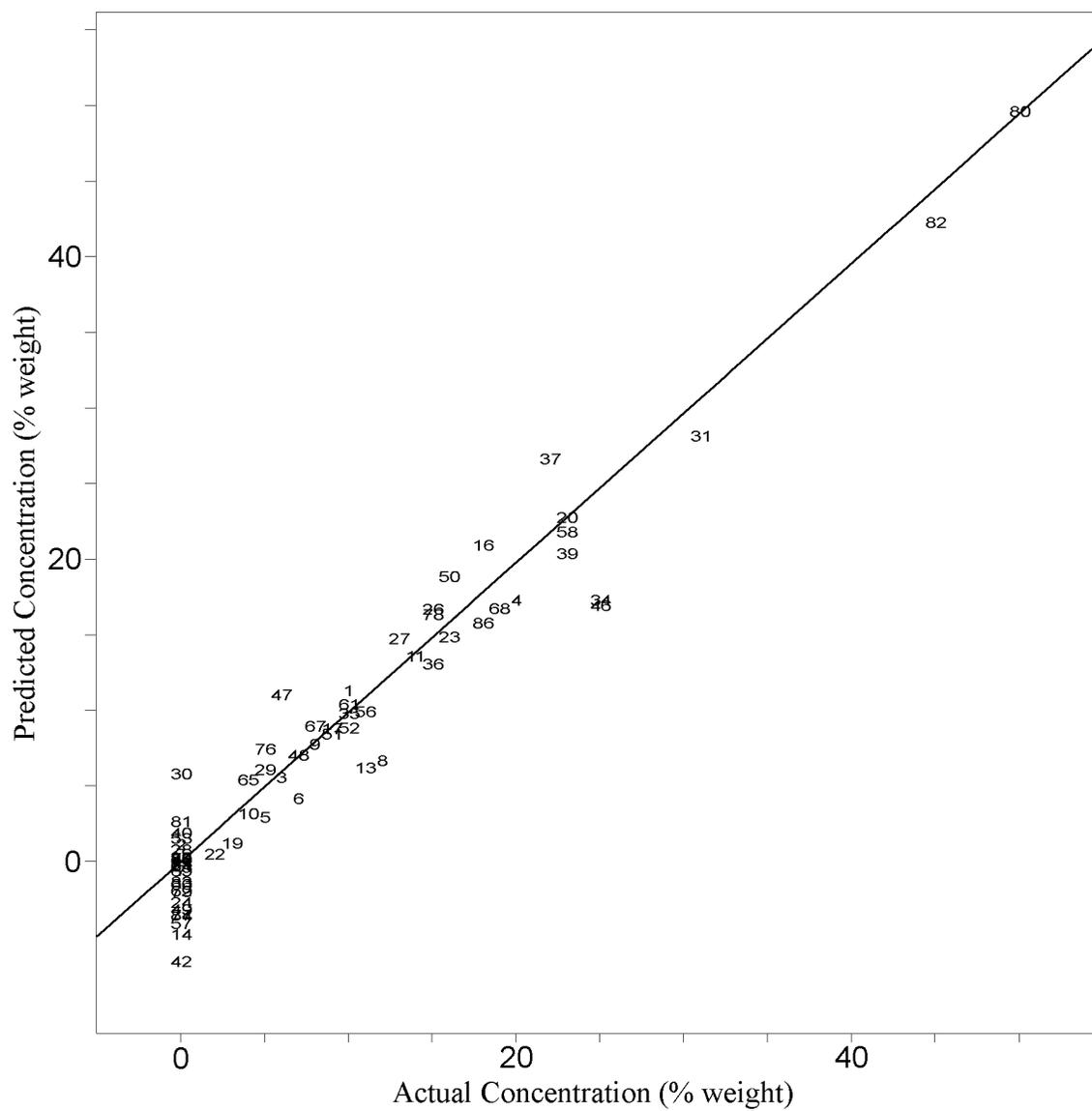
Figure 2.13   A plot of actual concentration versus predicted concentration for

N-acetyl-D-galactosamine.

N-Acetyl-D-Galactosamine
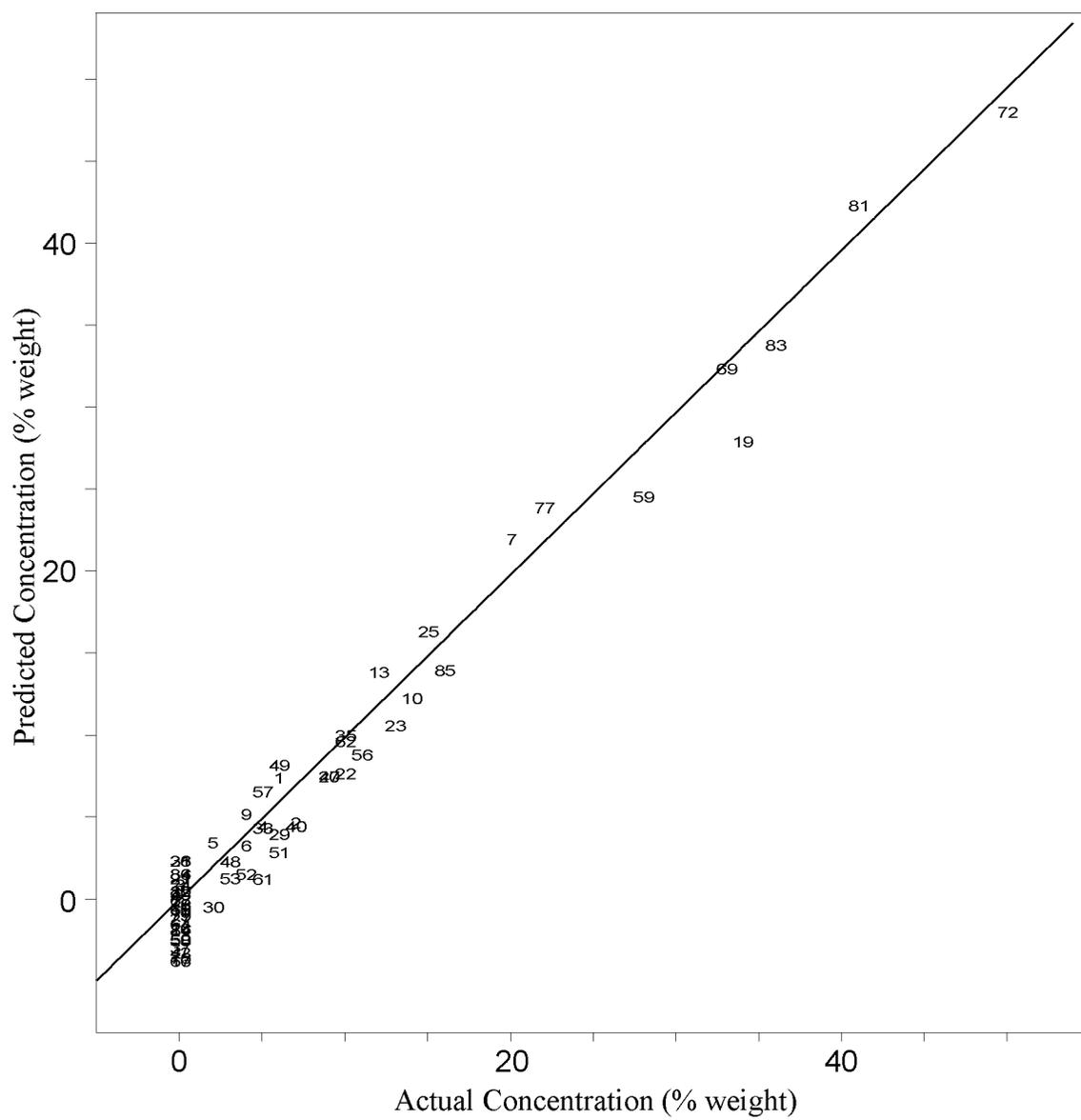
Figure 2.14   A plot of actual concentration versus predicted concentration for

D-glucose.

D-Glucose
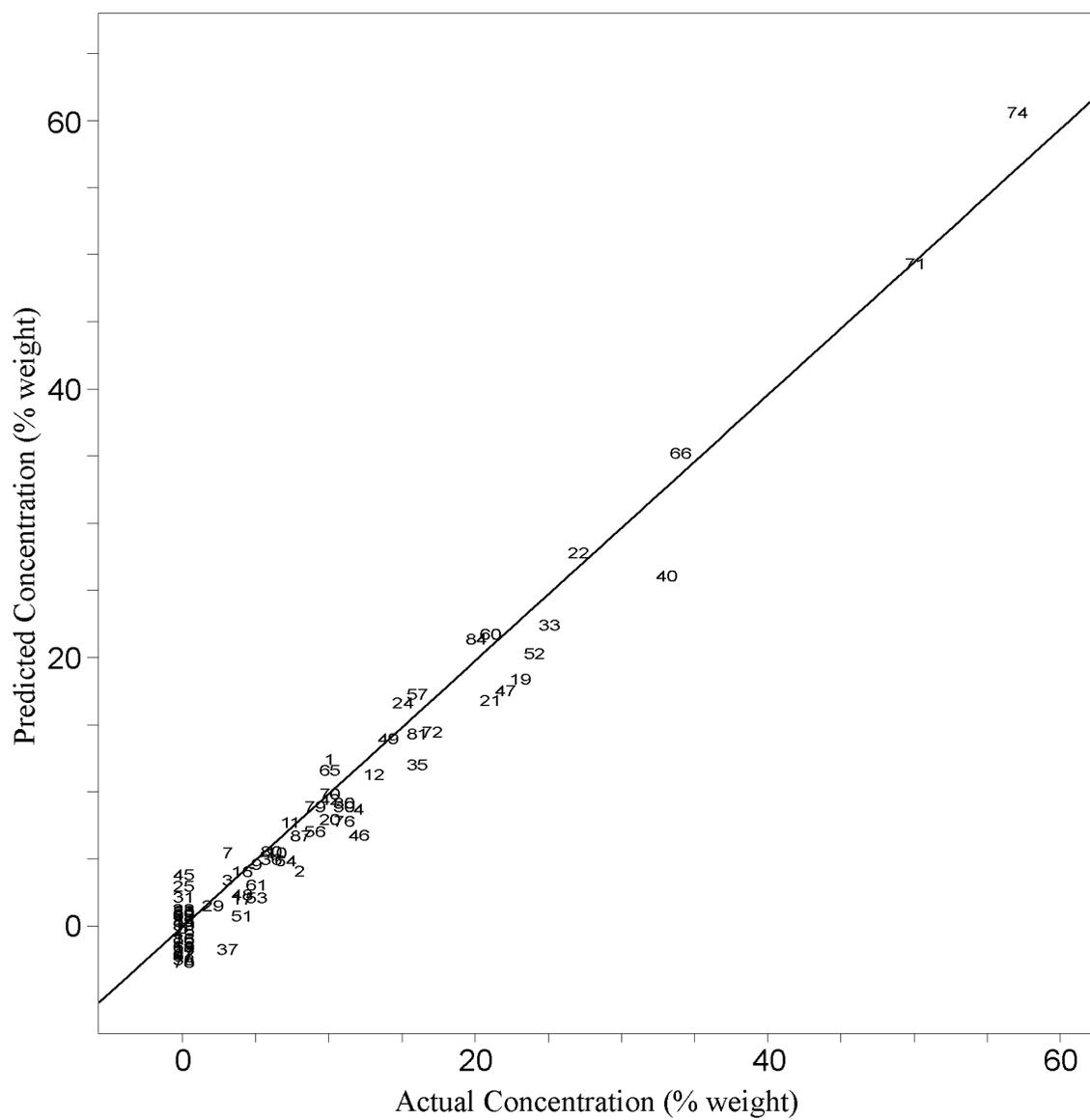
Figure 2.15   A plot of predicted residual error sum of squares versus the number of

factors for D‑galactose.

D-Galactose
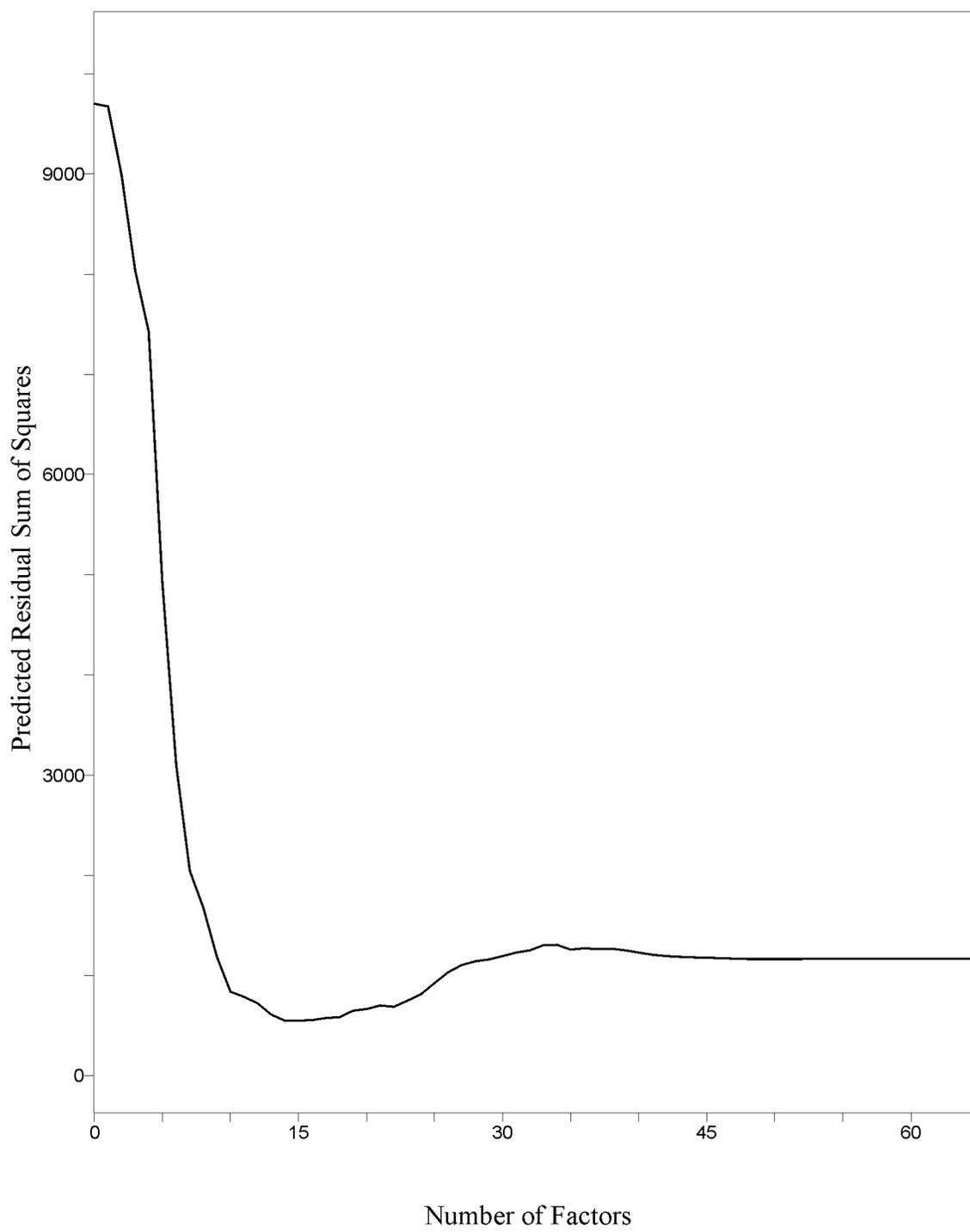
Figure 2.16   A plot of predicted residual error sum of squares versus the number of

factors for D-mannose.

D-Mannose
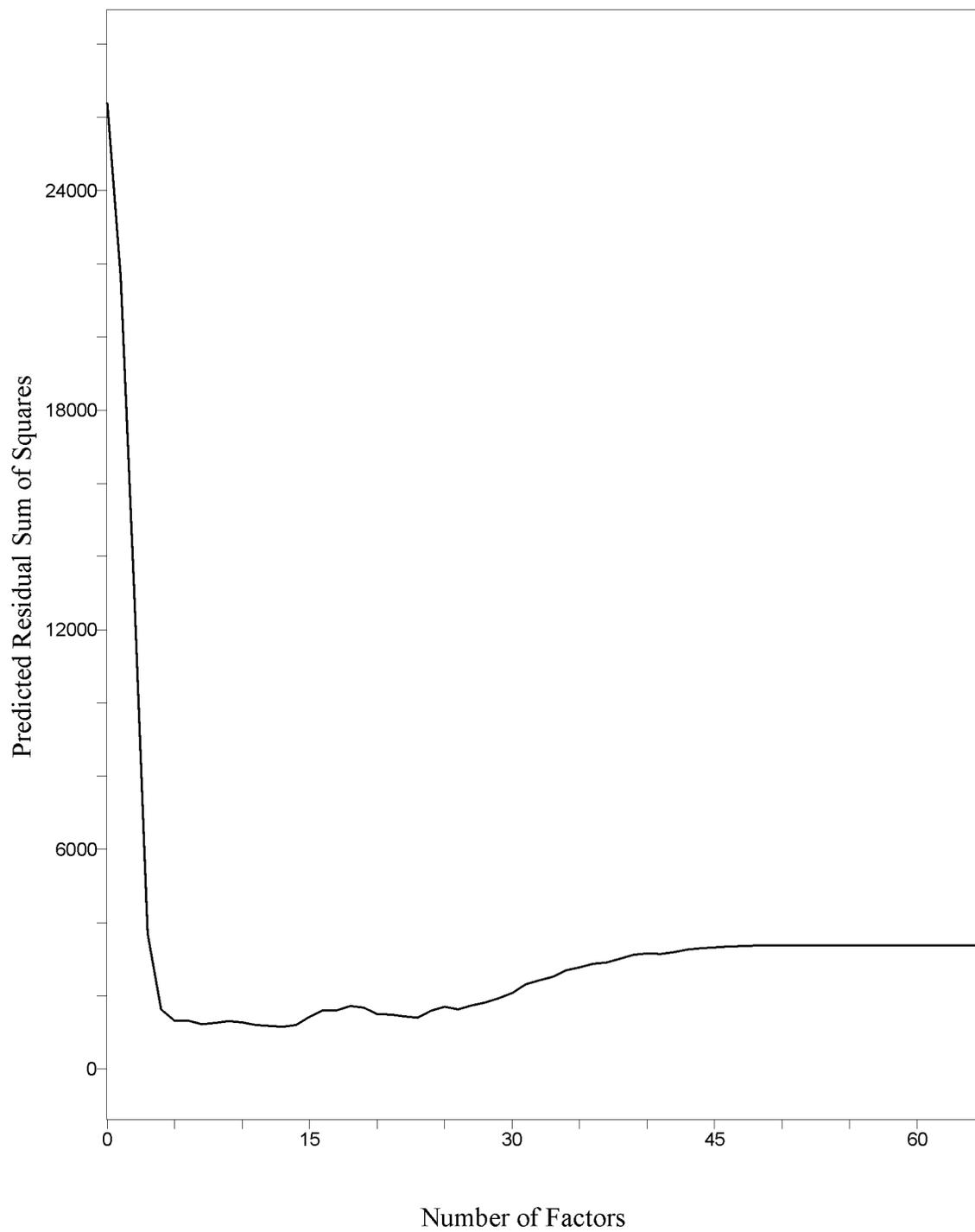
Figure 2.17   A plot of predicted residual error sum of squares versus the number of

factors for L-fucose.

L-Fucose

Predicted Residual Sum of Squares
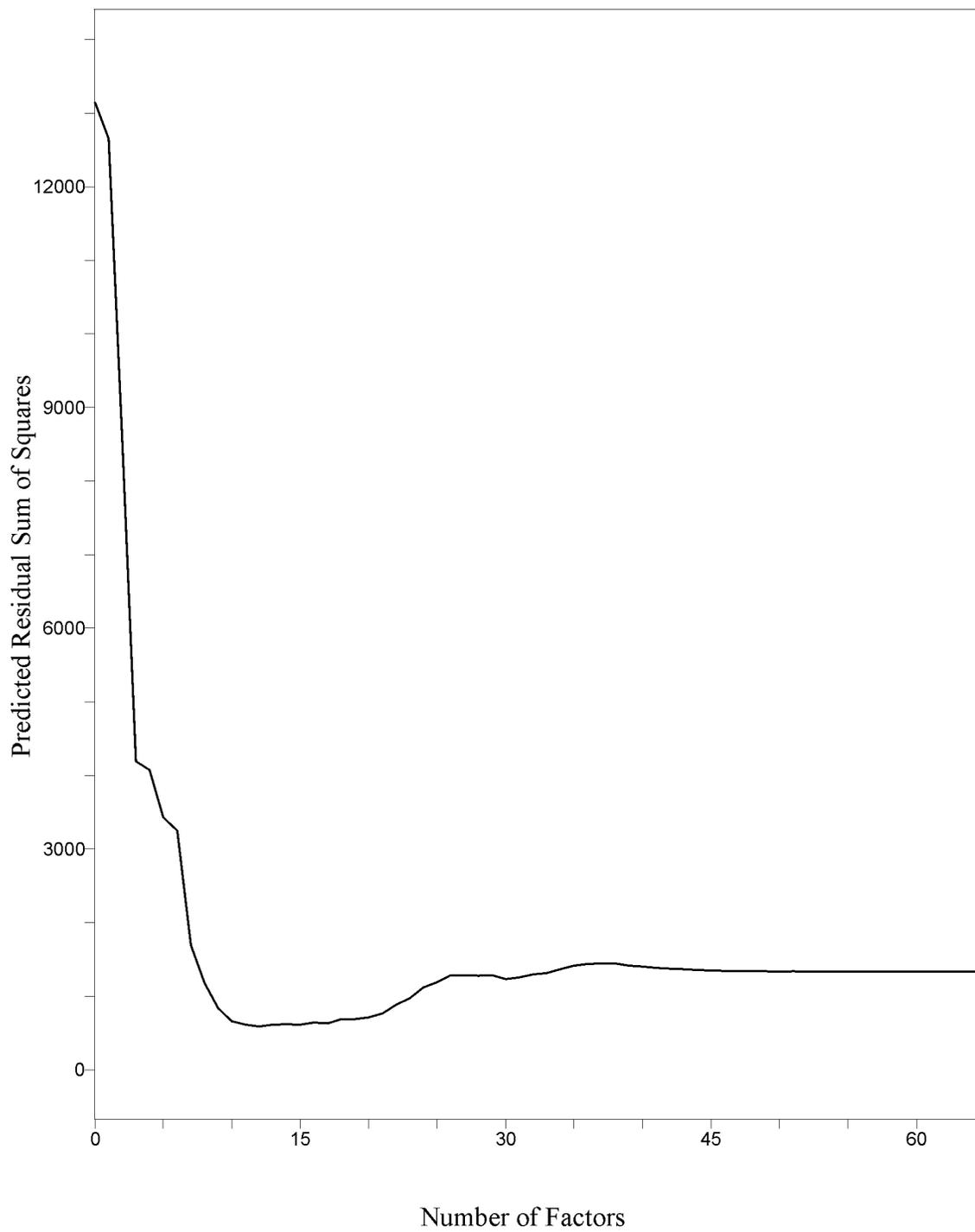
Number of Factors

Figure 2.18   A plot of predicted residual error sum of squares versus the number of

factors for N-acetyl-D-glucosamine.

N-Acetyl-D-Glucosamine
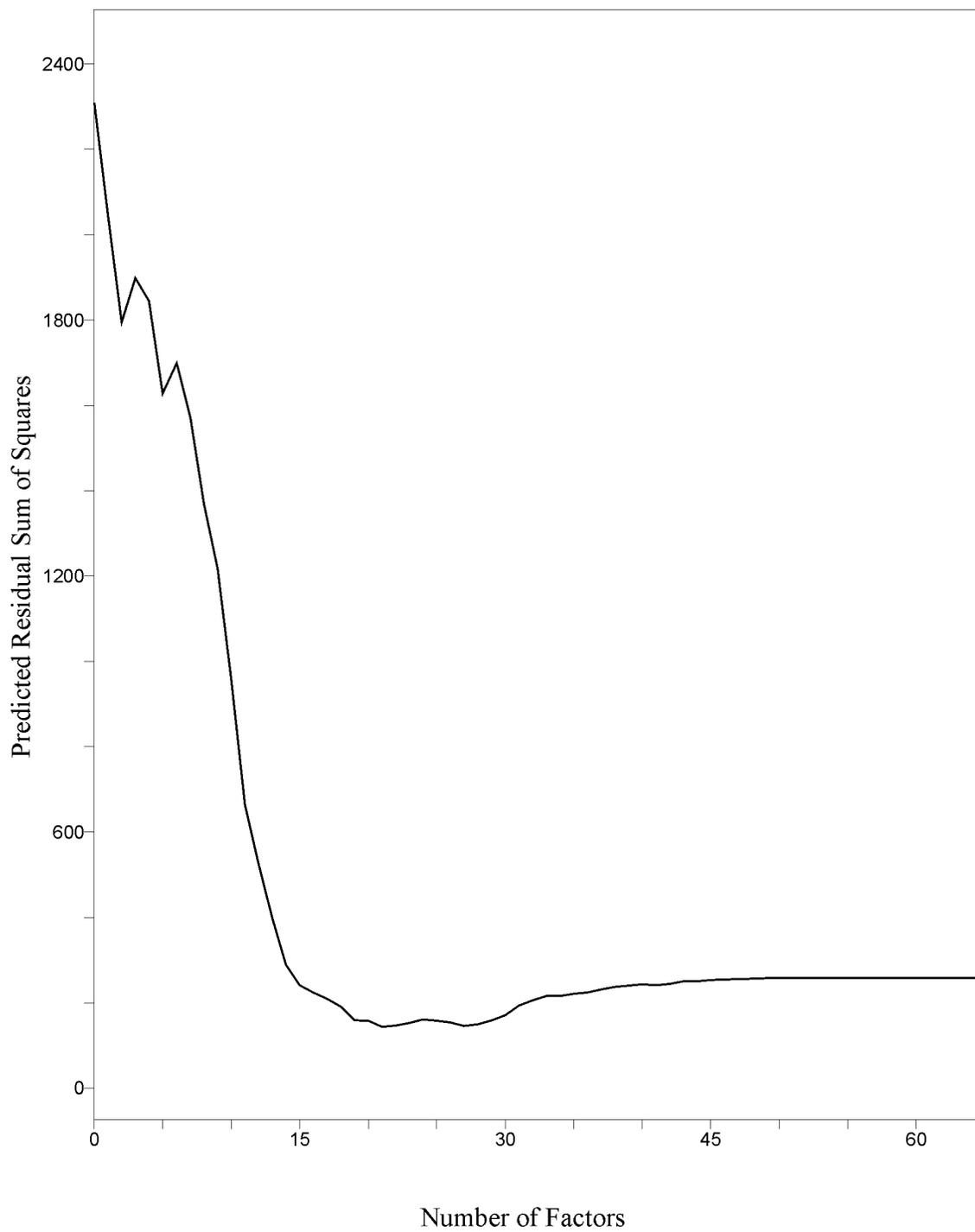
Figure 2.19   A plot of predicted residual error sum of squares versus the number of

factors for N-acetyl-D-neuraminic acid.

N-Acetyl-D-Neuraminic Acid
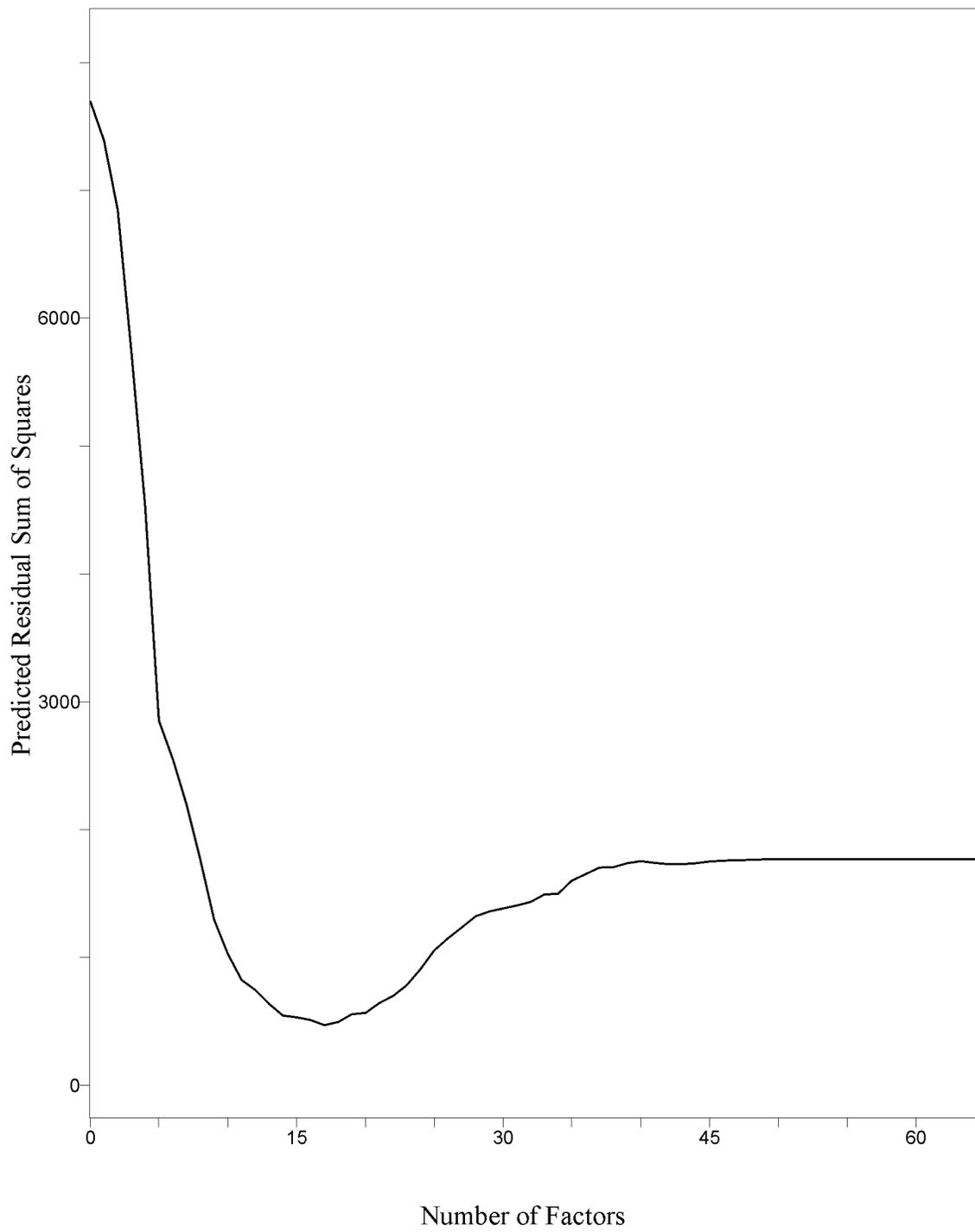
Figure 2.20   A plot of predicted residual error sum of squares versus the number of

factors for N-acetyl-D-galactosamine.

N-Acetyl-D-Galactosamine
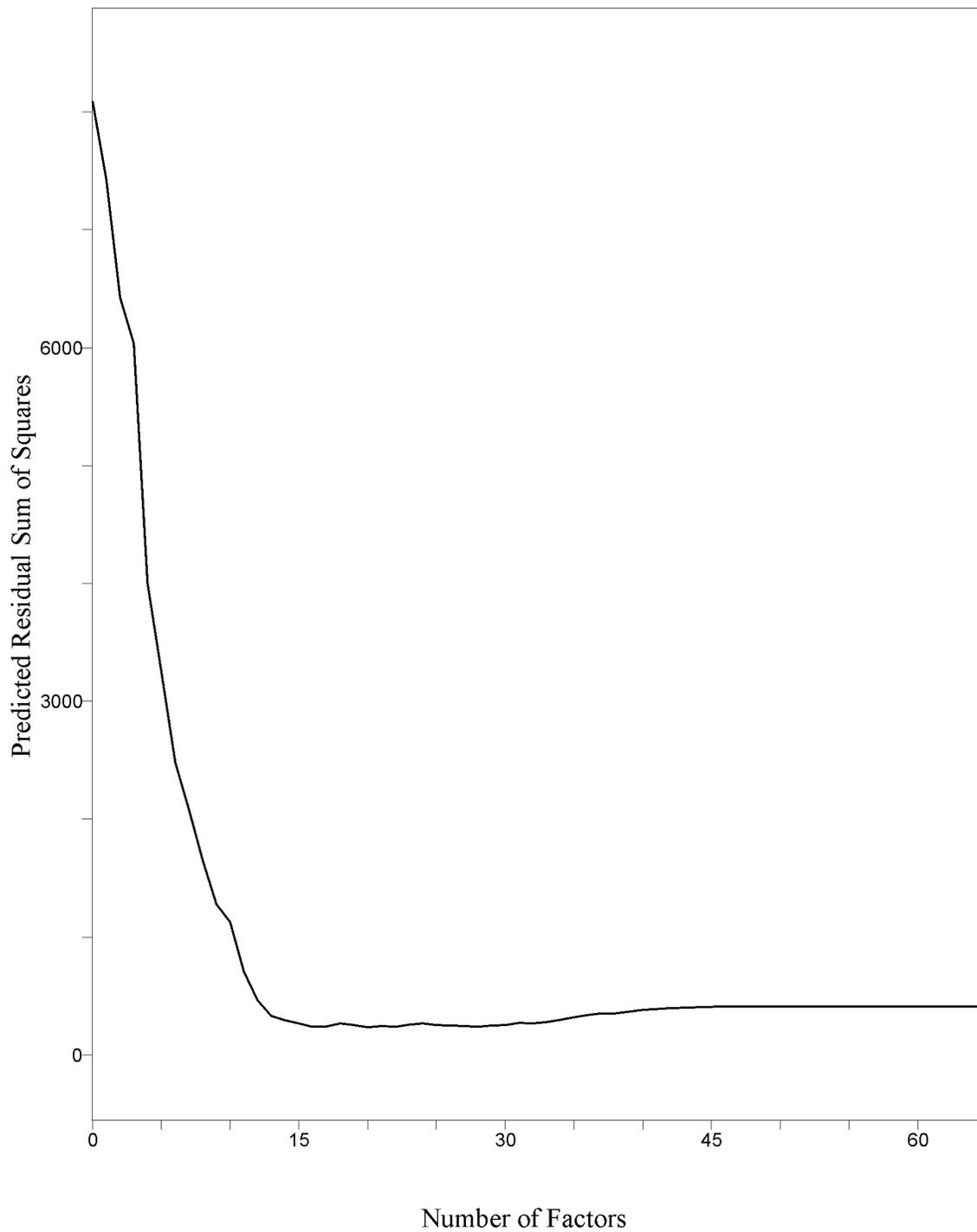
Figure 2.21   A plot of predicted residual error sum of squares versus the number of

factors for D-glucose.

D-Glucose
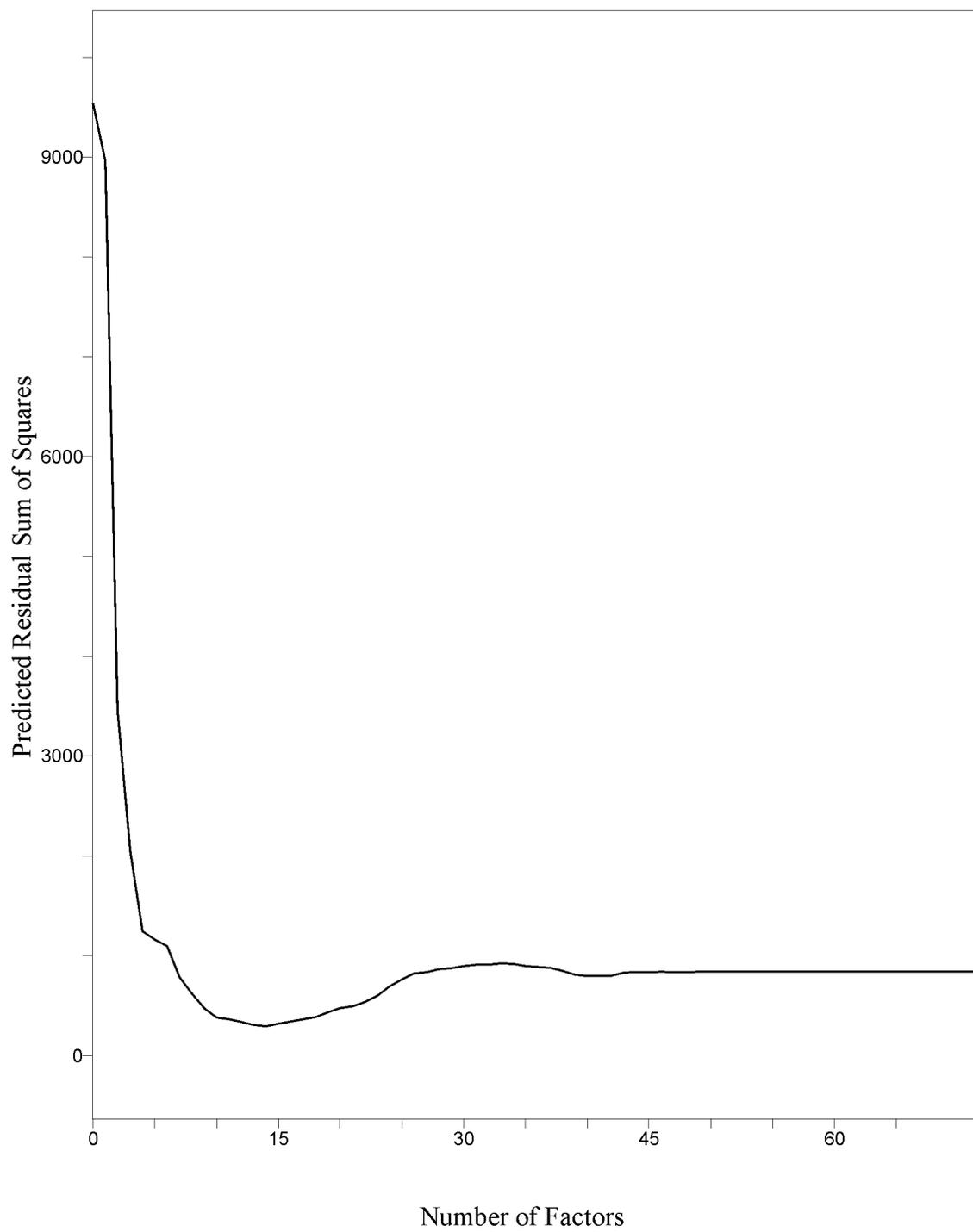
0.9567, 0.9553, 0.9400, 0.9394, 0.9704, and 0.9624, respectively. The values of f

selected for each of the above constituents are 14, 13, 12, 17, 14, 15 and 12,

respectively. These results were verified with a validation set.

The average error of prediction for each of the seven constituents among the

twenty samples in the validation set are 5.36%, 5.56%, 5.48%, 2.36%, 3.41%, 3.75%,

and 3.55%, respectively. Table 2.6 presents the average error of prediction among all

constituents for each sample. None of these results yields an absolute error greater than

six percent. Although the errors are much larger than the errors in the internal

predictions, they are still significantly smaller than the expected errors in gas

chromatography/mass spectrometry compositional analysis. After passing validation, a

set of ten unknowns was predicted with the model. These samples were prepared at the

Complex Carbohydrate Research Center at the University of Georgia to provide a blind

study; their preparation adhered to the procedure applied to the training and validation

sets, and their composition was undisclosed to the analyst. The results from the blind

study, of which the prediction range was comparable to that of the validation set, are

presented in Table 2.7.

Several methods to enhance the robustness of the model were investigated.

Alternate baseline-correction methods, such as polynomial fits of varying orders,

superchop methods, e.g., ultrasmoothing, and alternative linear interpolations,

underwent exploration for their viability. None of these alternatives, however, provided

as robust a model as the linear two-point correction applied at the extrema of the

selected calibration region of the spectra.

Table 2.6   The average percent error of prediction for all seven constituents for

each of the twenty validation samples.

Table 2.6

Validation Sample Number	Error of Prediction	Validation Sample Number	Error of Prediction
1	3.06%	11	6.38%
2	4.74%	12	2.68%
3	2.23%	13	5.73%
4	5.20%	14	1.92%
5	3.41%	15	3.27%
6	3.22%	16	3.12%
7	7.68%	17	2.96%
8	7.71%	18	3.21%
9	6.93%	19	3.13%
10	3.58%	20	4.13%

Table 2.7   Errors of prediction in percent for each constituent in each of the ten

unknown samples.

Table 2.7

#	Gal	Man	Fuc	GlcNAc	Neu5	GalNAc	Glu
1	0.73%	1.66%	1.82%	1.87%	2.67%	4.31%	3.48%
2	2.37%	5.02%	1.94%	0.25%	0.00%	3.89%	4.83%
3	0.24%	5.19%	2.76%	3.72%	0.19%	5.32%	3.26%
4	2.37%	2.32%	3.88%	1.12%	2.86%	2.87%	1.91%
5	5.87%	13.23%	4.91%	0.72%	5.94%	6.83%	1.63%
6	1.28%	1.74%	6.08%	2.27%	4.48%	3.35%	0.00%
7	1.63%	4.27%	3.19%	2.91%	1.46%	0.52%	0.30%
8	10.18%	11.04%	6.80%	0.00%	1.51%	7.28%	0.00%
9	11.12%	8.80%	9.78%	1.58%	10.07%	4.95%	4.57%
10	6.65%	7.18%	9.10%	7.64%	8.78%	0.31%	5.50%

In addition to alternate baseline-correction methods, other data pretreatment methods to compensate for anomalies in the spectra were investigated. A common method is to use the first or second derivative spectra in the training set. This essentially allows instrumental drift to be ignored or at least to be de-emphasized. This method was unsuccessful as an order of magnitude of noise is introduced into the spectrum for every second derivative computed. One preprocessing method, however, which demonstrates success, and therefore, was employed in the final model, is mean-centering. Mean-centering builds the model from the differences between each spectrum and the average of all of the spectra in the set. It reduces the prevalence of the common features among the spectra while placing emphasis on the unique characteristics of each spectrum. When mean-centering is not included in the calibration method, the coefficients of determination, R^2, for each of the constituents, D-galactose, D-mannose, L-fucose, N-acetyl-D-glucosamine, N-acetyl-D-neuraminic acid, N-acetyl-D-galactosamine, and D-galactose, decline to 0.9415, 0.9420, 0.9453, 0.9315, 0.9248, 0.9634, and 0.9573, respectively. In addition, the standard errors of prediction for all of the constituents increase. These findings support the contention that mean-centering enhances the method.

Finally, alternative methods of normalization were investigated to correct for pathlength differences among the sample deposits. One option was to normalize the spectra over the area of the carbonyl band (~ 1650 cm^{-1}). This approach seemed promising because it was postulated that the characteristics of the carbonyl bands are unaffected by the isomeric differences among the monosaccharides. No significant improvement was observed, however, by the utilization of this technique, and it was

therefore abandoned. In addition, normalization of the spectra by the intensity of a single point rather than the area of a region was explored. It was found, however, that the use of this procedure was detrimental to the calibration. Therefore, only unit normalization over the entire domain is employed in the final model.

Conclusions

The predictions made for all of the constituents in this study were reasonably accurate; however, one possible source of error may be attributed to the purity of the stock. N-acetyl-D-neuraminic acid, for example, is not commercially synthesized in high purity and is therefore isolated from either sheep submaxillary glands or E. coli. N-acetyl-D-neuraminic acid produced from E. coli was used in this investigation. The manufacturer reports that the purity of this monosaccharide is only ninety-eight percent.

There are several other sources of error in the analysis other than the impurities present in the starting materials. The greatest source of error is associated with the sample preparation itself. The model cannot be more accurate than the accuracy of the known composition of the training set from which it is built. Although pathlength discrepancies present during infrared analysis may be a profound source of error, the sample preparation elicits further investigation. The preparation is suspect as the solid monosaccharides are hygroscopic and, during mass determination, are exposed to atmospheric conditions for no less than several minutes. Another consideration, in addition to the inherent problem of exposure to water vapor, is the precision of the analytical balance. The balance provides three confident significant figures; however, due to a considerable amount of deviation in the last significant digit (one hundredth of

a milligram), this digit is highly questionable. Furthermore, the use of three different pipettes, while necessary, introduces a bias with regard to the accuracy of each. The systematic error of the individual pipettes results in the further propagation of error.

The next logical source of error is the discrepancy in pathlength attributed to variations of the deposit thickness. Although the technique allows for a reasonable thickness gradient within a single deposit, selecting the same thickness for every sample acquisition is nearly impossible. A good approximation is made by monitoring the infrared response during relocation of the stage control. Ideally, a spot is deemed appropriate if the maximum absorbance is no more than 0.5 absorbance units. Unfortunately, the local variation in thickness within the 50 µm x 50 µm data acquisition area is essentially indeterminate by both the infrared spectra and the display through the visible objective. Although it is desirable to collect full spectra at the highest signal possible, thus maximizing the signal-to-noise ratio, exceeding the prescribed limit, i.e., entering the range of overabsorbance, however, will result in spectral response where the apparent absorbance and actual absorbance do not adhere to the linear relationship expressed by the Bougeur-Beer-Lambert law. Non-compliance with the Bougeur-Beer-Lambert Law may have catastrophic effects on the calibration model. Although reasonable provisions are made to safeguard against this scenario, it is difficult to guarantee the nonoccurrence of local maxima in deposit thickness that exceed the range of linear response.

Normalization, to some extent, removes the error attributed to pathlength differences and improves the calibration, but it is not a panacea for non-uniform deposits. Normalization does not correct for the distortions that arise from the

non-uniformity of the deposit surface. Distortions due to surface effects require further investigation for remedy. Finally, the infrared microscope has many off-axis beam paths that further compromise pathlength consistency.

PLS-1 is the most suitable algorithm for use with this system. Although PLS-2 is a faster calculation, especially as the number of constituents increase, it is only as robust as PLS-1 when the domain is identical among all of the constituents. The precision of this method does exceed both gas chromatography/mass spectrometry and nuclear magnetic resonance in compositional analysis. Other statistical procedures were investigated to optimize the results, but none surpassed the performance of PLS-1.

In future work, it is proposed that both the methods of analysis and the sample preparation be subjected to reexamination. One important point is to assess the maximum storage time of the standards and reagents before their decomposition is initiated. It has been determined that the 3 N methanolic hydrochloric acid stock is the first of all of the chemical supplies to be compromised under long term storage despite refrigeration and separate dilutions to 1 N for individual experiments. As a precaution, this reagent is replaced every six months, and all other reagents and standards are replenished annually.

Perhaps of greater significance is that the Fourier transform infrared analysis itself will be subjected to revision to the extent that it is proposed that the microscope setup be abandoned entirely in favor of an alternate methodology. A separate investigation, as will be discussed in the following chapter, is underway in an attempt to eliminate some of the difficulties in using infrared microspectrometry. It will utilize the measurements, made by a Harrick Split-Pea™ single-bounce attenuated total reflection

accessory interfaced to the spectrometer, to construct the statistical models. The new method may eliminate both the need to baseline-correct and to normalize spectra from the analyzed samples. As a result, this may increase the reproducibility of the spectra and minimize the influence of environmental fluctuations on the system.

References

1.   Hu, G. F. *J. Chromatogr. A* **1995**, *705* (1), 89-103.

2.   Albersheim, P.; Darvill, A.; Davis, K.; Doares, S.; Gollin, D.; Mcneil, M.;
        Oneill, R.; York, W. *Federation Proceedings* **1985**, *44* (4), 1048-1049.

3.   Torto, N.; Cohen, A.; Gorton, L.; Laurell, T.; van der Hoeven, R. *Laboratory
        Robotics and Automation* **1998**, *10* (6), 361-367.

4.   Hachey, D. L.; Parsons, W. R.; McKay, S.; Haymond, M. W. *Anal. Chem.* **1999**,
        71 (20), 4734-4739.

5.   Karlsson, N. G.; Nordman, H.; Karlsson, H.; Carlstedt, I.; Hansson, G. C.
        Biochemical Journal **1997**, *326*, 911-917.

6.   Huyghuesdespointes, A.; Yaylayan, V. A.; Keyhani, A. *Journal of Agricultural
        and Food Chemistry* **1994**, *42* (11), 2519-2524.

7.   Leskovsek, H.; Perko, S.; Zigon, D.; Faganeli, J. *Analyst* **1994**, *119* (6),
        1125-1128.

8.   Tevesz, M. J. S.; Schwelgien, S. F.; Smith, B. A.; Hehemann, D. G.; Binkley, R. W.; Carter, J. G. *Veliger* **1992**, *35* (4), 381-383.

9.   Pons, A.; Popa, J.; Portoukalian, J.; Bodennec, J.; Ardail, D.; Kol, O.; Martin-Martin, M. J.; Hueso, P.; Timmerman, P.; Leroy, Y.; Zanetta, J. P. *Anal. Biochem.* **2000**, *284* (2), 201-216.

10.   Bendiak, B.; Fang, T. T. *Carbohydrate Research* **2000**, *327* (4), 463-481.

11.   Duus, J. O.; St Hilaire, P. M.; Meldal, M.; Bock, K. *Pure and Applied Chemistry* **1999**, *71* (5), 755-765.

12.   Bendiak, B. *Carbohydrate Research* **1999**, *315* (3-4), 206-221.

13.   Mazzoni, V.; Bradesi, P.; Tomi, F.; Casanova, J. *Magnetic Resonance in Chemistry* **1997**, *35*, S81-S90.

14.   Taravel, F. R.; Mazeau, K.; Tvaroska, I. *Brazilian Journal of Medical and Biological Research* **1995**, *28* (7), 723-732.

15.   van Halbeek, H. *Current Opinion in Structural Biology* **1994**, *4* (5), 697-709.

16.   van Kuik, J.; Hard, K.; Vliegenthart, J. *Carbohydrate Research* **1992**, *235*, 53-68.

17.   Allerhand, A.; Berman, E. *J. Am. Chem. Soc.* **1984**, *106* (8), 2400-2412.

18.   Tsunesada, T.; Takehara, A.; Mitsuishi, K.; Ihara, T. *Bunseki Kagaku* **2000**, *49* (6), 437-442.

19.   Laine, R. A. *Glycobiology* **1994**, *4* (6), 759-767.

20.   MacNamara, E.; Hou, T.; Fisher, G.; Williams, S.; Raftery, D. *Anal. Chim. Acta* **1999**, *397* (1-3), 9-16.

21.   Peyron, M.; Pierens, G. K.; Lucas, A. J.; Hall, L. D.; Potter, G. F.; Stewart, R. C.; Phelps, D. W. *Magnetic Resonance Imaging* **1994**, *12* (2), 295-298.

22.   Sivakesava, S.; Irudayaraj, J. *Applied Engineering in Agriculture* **2000**, *16* (5), 543-550.

23.   Gallignani, M.; Garrigues, S.; Delaguardia, M. *Anal. Chim. Acta* **1994**, *287* (3), 275-283.

24.   Lewis, C. B.; McNichols, R. J.; Gowda, A.; Cote, G. L. *Appl. Spectrosc.* **2000**, *54* (10), 1453-1457.

25.   Smola, N.; Urleb, U. *Anal. Chim. Acta* **2000**, *410* (1-2), 203-210.

26.   Hall, J. W.; Pollard, A. *Clinical Biochemistry* **1993**, *26* (6), 483-490.

27.   Marbach, R.; Koschinsky, T.; Gries, F. A.; Heise, H. M. *Appl. Spectrosc.* **1993**, *47* (7), 875-881.

28.   Miller, C. E. *Spectrochimica Acta Part A-Molecular and Biomolecular Spectroscopy* **1993**, *49* (5-6), 621-632.

29.   Lee, M. H.; Cavinato, A. G.; Mayes, D. M.; Rasco, B. A. *Journal of Agricultural and Food Chemistry* **1992**, *40* (11), 2176-2181.

30.   Miller, C. E.; Eichinger, B. E. *Journal of Applied Polymer Science* **1991**, *42* (8), 2169-2190.

31.   Rasco, B. A.; Miller, C. E.; King, T. L. *Journal of Agricultural and Food Chemistry* **1991**, *39* (1), 67-72.

32. Salmain, M.; Varenne, A.; Vessieres, A.; Jaouen, G. *Appl. Spectrosc.* **1998**, *52* (11), 1383-1390.

33. Ahmed, M. K.; Daun, J.; DeClercq, D. *Appl. Spectrosc.* **1998**, *52* (7), 990-993.

34. Lopez-Anreus, E.; Garrigues, S.; de la Guardia, M. *Analyst* **1998**, *123* (6), 1247-1252.

35. Woodhead, A.; Harrigan, F.; Church, J. *Vibrational Spectroscopy* **1997**, *15* (2), 179-189.

36. Cooper, J. B.; Wise, K. L.; Welch, W. T.; Sumner, M. B.; Wilt, B. K.; Bledsoe, R. R. *Appl. Spectrosc.* **1997**, *51* (11), 1613-1620.

37. Andrade, J. M.; Muniategui, S.; Prada, D. *Fuel* **1997**, *76* (11), 1035-1042.

38. Zhu, C. J.; Griffiths, P. R. *Appl. Spectrosc.* **1998**, *52* (11), 1403-1408.

39. Zhu, C. J.; Griffiths, P. R. *Appl. Spectrosc.* **1998**, *52* (11), 1409-1413.

40. Ramsay, D. A. *J. Am. Chem. Soc.* **1952**, *74* (1), 72-80.

CHAPTER 3

ANALYSIS OF N-LINKED OLIGOSACCHARIDES BY FOURIER

TRANSFORM INFRARED SPECTROMETRY/ATTENUATED TOTAL

REFLECTION SPECTROMETRY

In the preceding chapter, a method for the determination of the monosaccharide composition of N-linked mammalian oligosaccharides was investigated and developed. It is an entirely novel approach to complex carbohydrate analysis, and although it has demonstrated the feasibility of infrared microspectrometry, which is an attractive tool for reasons discussed previously, as a methodology for carbohydrate analysis, considerably more investigation is necessary to improve its efficacy. This chapter explores a second approach to the same question, undertaken to improve the previous method and is intended to resolve several areas of difficulty that were encountered during the prior work. One area of concern involves the deficiencies associated with the predictive performance that are related to the preparation of standard mixtures of monosaccharides.

Initially, monosaccharide mixtures were prepared by weighing the predetermined quantity of each of the pure, dry monosaccharides directly into the reaction vial for that mixture. This procedure has since been characterized as inappropriate for mainly two reasons. The first regards the precision of the analytical balance. Even if the balance assesses reliably a given mass within a tolerance of

$\pm$ 0.05 mg, which is characteristic for a high precision analytical balance, for a two-milligram sample, the error with respect to percent composition may be no less than 2.5%. The second problem is that the dry, powdered monosaccharides are hygroscopic, which introduces additional error. Because it takes no less than ten minutes to prepare a given mixture, the effect of water absorption may be pronounced and easily increase the expected error by several percent. This introduction of water into the monosaccharide mixtures is also a concern as it may adversely affect the peracetylation reaction. After the mixtures of monosaccharides are dissolved into water, however, it was determined by gravimetric analysis that lyophilization nearly removes all of the moisture. This creates a sufficiently anhydrous condition for the peracetylation to take place unencumbered.

The inclusion of lyophilization into the sample procedure allowed the use of stock solutions, which alleviated some uncertainty with regard to the mass determination of the individual components during preparation. Dilutions of the individual sugars to prepare the stock solutions could be assessed with higher precision because a greater mass is used in the preparation, which yields a smaller relative error in mass determination by an analytical balance. Also, the use of micropipettes for dilutions reduces the error by an order of magnitude relative to the error incurred by an analytical balance operating in the milligram region. The use of aqueous stocks resulted in the period of exposure of the monosaccharides to atmosphere during mass determination to diminish to a fraction of the time encountered during the previous method. Finally, it became apparent that implementation of the mass determination in a dry box is neither practical nor necessary.

Another area of concern is in relation to the infrared microspectrometry itself. For this investigation, a method that utilized attenuated total reflection spectrometry in lieu of infrared microspectrometry was developed. The use of a single-reflection attenuated total reflection accessory may actually reduce the sample size requirement. In addition, attenuated total reflection spectrometry permits the use of a deuterated tri-glycine sulfate detector, which responds more linearly to radiational intensity than does a mercury cadmium telluride detector.[1,2] Finally, the necessity for normalization is eliminated since the pathlength is held constant in attenuated total reflection spectrometry if the condition is met that the depth of penetration is exceeded by the thickness of the sample and is discussed in greater detail in the subsequent section.

Attenuated Total Reflection Spectrometry

Total internal reflection is a phenomenon that can be observed during ordinary experiences. In a glass of water, for example, if one looks through the surface of the water at a small angle from normal incidence, then the surface of the glass under the surface of the water appears to be completely silvered, and it is not possible to see objects on the other side of the glass.[3,4] The radiation striking the glass is completely reflected, and the objects behind the glass do not receive the radiation from this optical pathway. The phenomenon ceases to occur, however, when an object is either brought into contact with the outside reflecting surface of the glass or brought into very close proximity. The destruction of this phenomenon results from the establishment of a standing wave normal to the reflecting surface in the denser medium, and an evanescent nonpropagating field is generated in the less dense medium of which the amplitude

decays exponentially with the distance from the surface. A criterion that must exist for total internal reflection to take place is that the angle of incident radiation θ must exceed the critical angle θ_C. The critical angle varies as a function of the ratio of refractive indices of the two media in the following manner:

$$\theta_C = \arcsin(\eta_2/\eta_1)$$

where η_1 is the refractive index of the denser medium and η_2 is the refractive index of the scarcer medium.[3] Figure 3.1 is a graphical representation of the conditions needed for total internal reflection to take place.

There are many applications of total internal reflection, such as the ability to greatly enhance the image contrast in fingerprinting techniques;[5] however, the specific application relevant to the present work is in regard to the application of total internal reflection to vibrational spectrometry, known as attenuated total reflection spectrometry. This name arises from the occurrence of attenuation in the evanescent wave at frequencies that an object absorbs the radiation when it is brought into contact with or in the proximity of the reflecting surface. This technique has found a considerable amount of use in the infrared wavelength region.[6-15] Since infrared spectrometry has become one of the most powerful analytical techniques, the increased need for the ability make surface and bulk analysis has grown significantly. There has been tremendous development in attenuated total reflection accessories to accompany infrared spectrometry.[9,16-32] Unlike other accessories designed for infrared spectrometry, these accessories offer a number of important advantages. Mainly, attenuated total reflection spectrometry is not hindered by many of the sampling problems from which transmission experiments suffer. Quite often, samples are not

Figure 3.1   A representation of total internal reflection and the criteria required for

the phenomenon to occur.

Beam below critical angle obeys Snell's Law

Low refractive index medium, η_1

High refractive index medium, η_2

θ_c

Beam above critical angle reflects internally
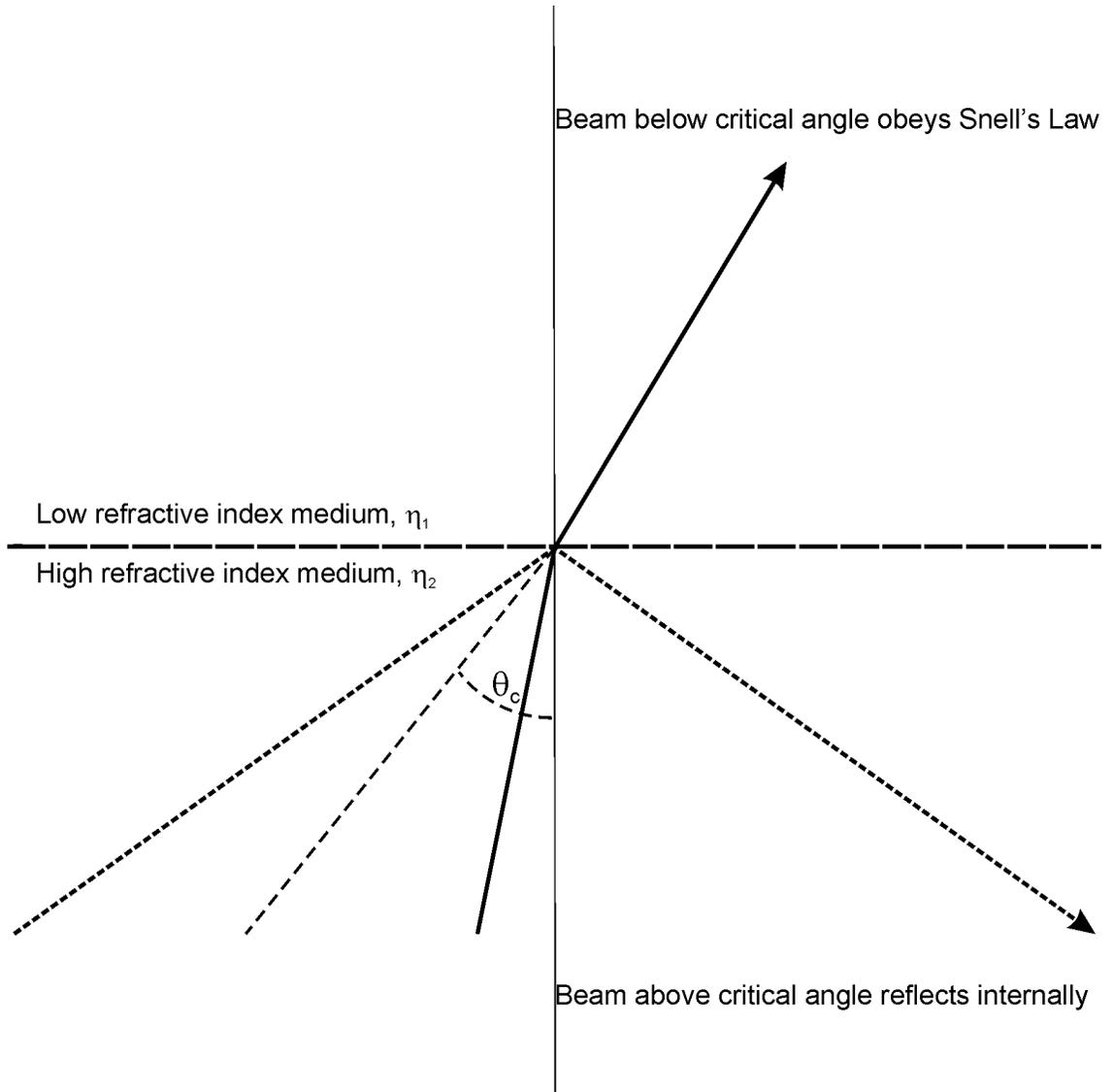
sufficiently thin optically for transmission measurements. The preparation of the samples to make them suitable for transmission experiments usually compromises the sample, or at least alters the morphology of the samples, through dilution or compression. Figures 3.2 and 3.3 represent the orientation of a sample measured by transmission spectrometry and attenuated total reflection spectrometry, respectively.

Little preparation is needed for the implementation of total attenuated total reflection experiments, as it is only necessary to place a sample into contact with the accessory. Typically, only minor pressure application is required for intimate contact with the sample, thereby preserving the integrity of the sample. As a result, this technique can be applied to minute samples in the forms of liquid, irregularly shaped solids, powders, etc., so long as sufficient contact between the accessory element and the sample is maintained. The distance from the surface of the element where the evanescent wave is generated is on the order of micrometers, which makes attenuated total reflection insensitive to the bulk thickness of a sample, and therefore the analysis of thick or strongly absorbing materials becomes possible.

An important parameter in attenuated total reflection spectrometry is the depth of penetration d_p, which is the distance that the evanescent wave extends into the sample. The depth of penetration, which is also wavelength dependent, is defined as the distance from the element/sample interface where the intensity of the evanescent wave decays to $1/e$ of its initial value at the interface and is determined as follows:

$$d_p = \lambda/(2\pi\eta_1(\sin^2\theta - (\eta_2/\eta_1)^2)^{1/2})$$

where λ is the wavelength of the incident radiation, and θ is the angle of incidence.[3,33,34] Figure 3.4 shows an approximation of the relationship of the intensity of an evanescent

Figure 3.2   The orientation of sample measured by a transmission experiment.

Transmission

Infrared Source

Sample

Detector
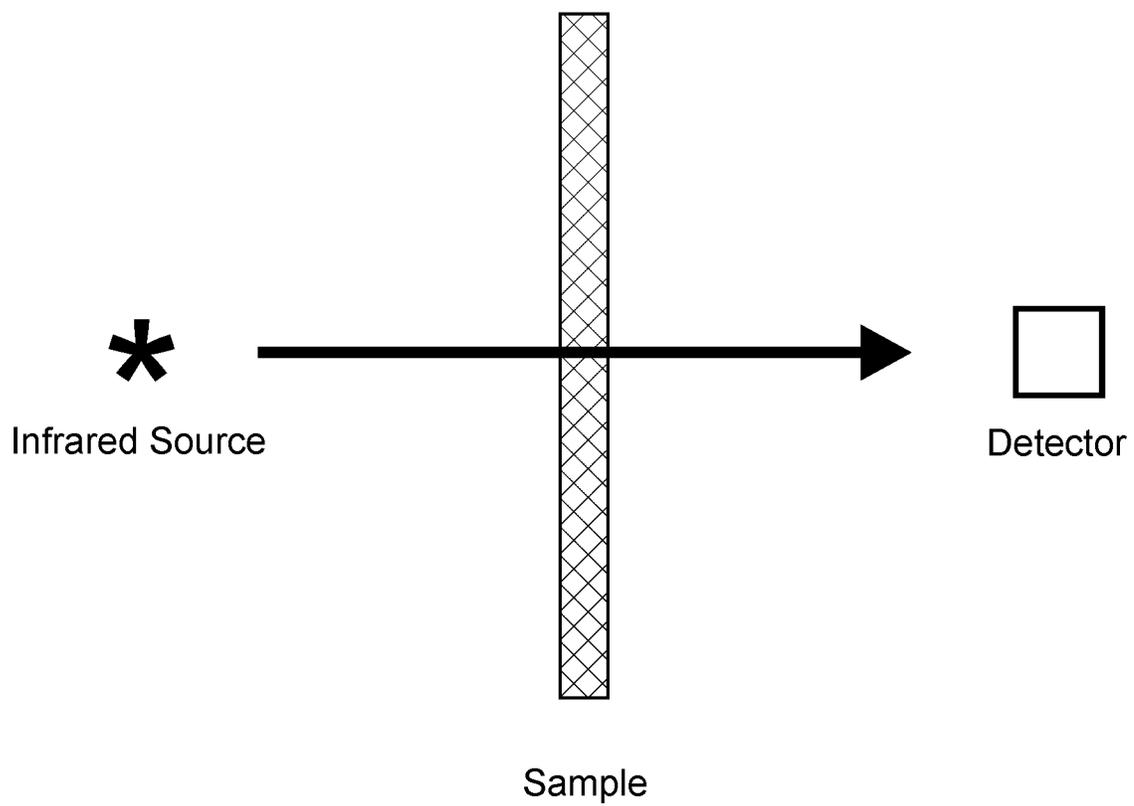
Figure 3.3   The orientation of a sample measured by an attenuated total reflection

experiment.

Internal Reflection

Sample

Internal Reflection Element

✻

Infrared Source

Detector
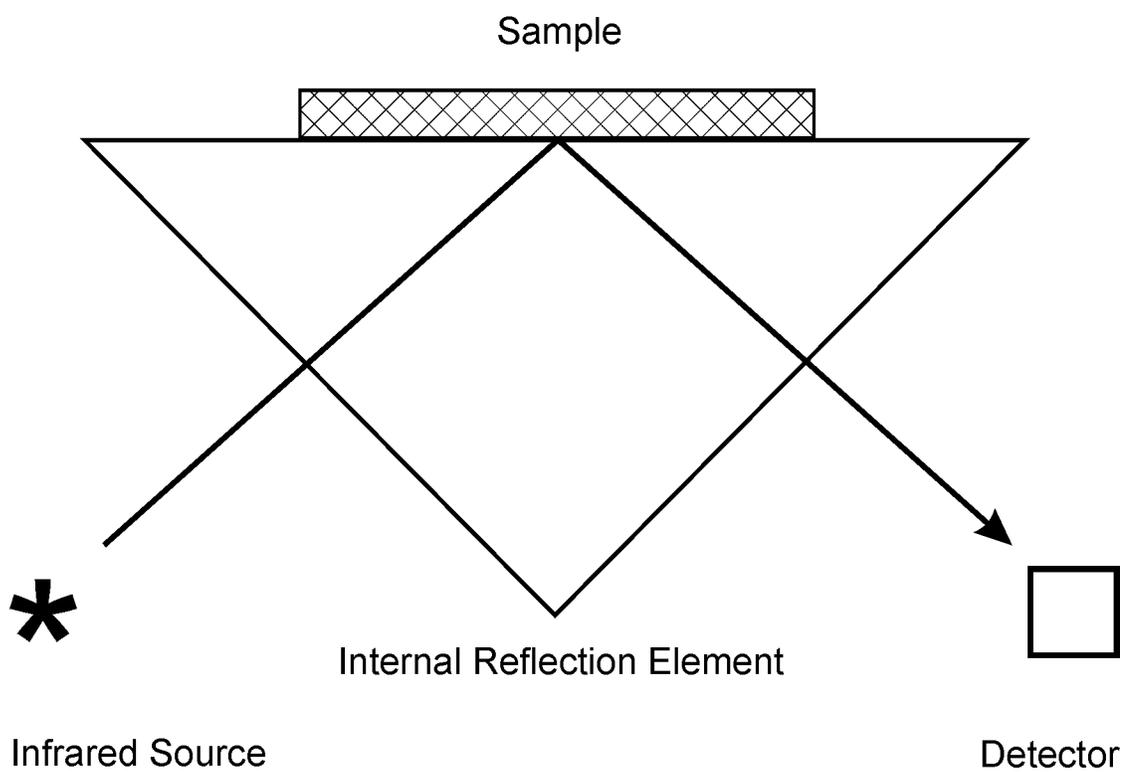
Figure 3.4   The depth of penetration of an evanescent wave into a sample. The

darkness of the shading represents the intensity of the evanescent wave.

Depth of Penetration of Evanescent Wave

Low Refractive Index Medium

High Refractive Index Medium

wave into a sample with the distance from the sample/element interface. The effective pathlength is the depth of penetration multiplied by the number of reflections. As with transmission spectra, where the spectral intensity is directly proportional to the thickness of the sample through which the radiation passes, in attenuated total reflection spectrometry the effective pathlength is directly proportional to the spectral intensity. This dissertation is only concerned with single-reflection (or single-bounce) total attenuated reflection; therefore, for the purpose of discussion herein, the effective pathlength and depth of penetration are rendered equivalent.

The choice of material for the composition of the internal reflection element has several consequences. The refractive index of the element is inherent to its composition and has two main effects on the spectrometry. The first is that as the refractive index of the internal reflection element material increases the critical angle increases as well. This is of particular importance when measuring samples with high refractive indices, since it is advisable that the angle of incidence greatly exceeds the critical angle to avoid distortions in the spectra.[29] Secondly, as can be seen from the equation above, a higher refractive index of the internal reflection element will decrease the depth of penetration thereby decreasing the intensity of the spectrum. Additionally, the choice of material for the internal reflection element determines the spectral range that can be acquired.[4] Finally, suitable elements must not be composed of materials that interact chemically with the samples. For mid-infrared attenuated total reflection spectrometry, germanium and silicon are common internal reflection element materials, and they have refractive indices of 4.0 and 3.5, respectively.[6,9]

Finally, the sample contact efficiency and area of the sample contact have an impact on quantitative analysis. It is imperative that both of these values be as close to 100% as possible. For liquid samples, the contact efficiency is very high because the liquid will adhere uniformly to the surface of the internal reflection element. The sample must also completely cover the active area of the element in order for the spectra intensities to be maximized, and more importantly, in order for the spectra to be reproducible. Not only must the area be completely covered, but also the thickness of the sample over the entire area must exceed the depth of penetration to ensure that the effective pathlength remains constant. The reproducibility of the effective pathlength is of paramount importance for the performance of the statistical analysis on the spectral data.

Infrared spectra acquired by attenuated total reflection spectrometry do not exactly resemble the optical constants of the material, the index of refraction, nor the attenuation index, but rather they are a complicated composite of these factors.[3,4,18] This is not a serious problem, however, because the spectra are highly reproducible, and reference libraries of spectra exist for purposes of comparison. Figures 3.5 and 3.6 show a transmission spectrum and an attenuated total reflection spectrum, respectively, of the same material. The depth of penetration, as stated earlier, is wavelength dependent in that as the wavelength increases so does the depth of penetration. Often spectrometers have a mathematical attenuated total reflection correction routine that can be applied to the data to make the reflection spectra more closely resemble the transmission spectra. A spectrum that has undergone such a correction is shown in Figure 3.7. In the present case, although the correction is useful for the analyst, who

Figure 3.5   A transmission spectrum of a peracetylated monosaccharide mixture.
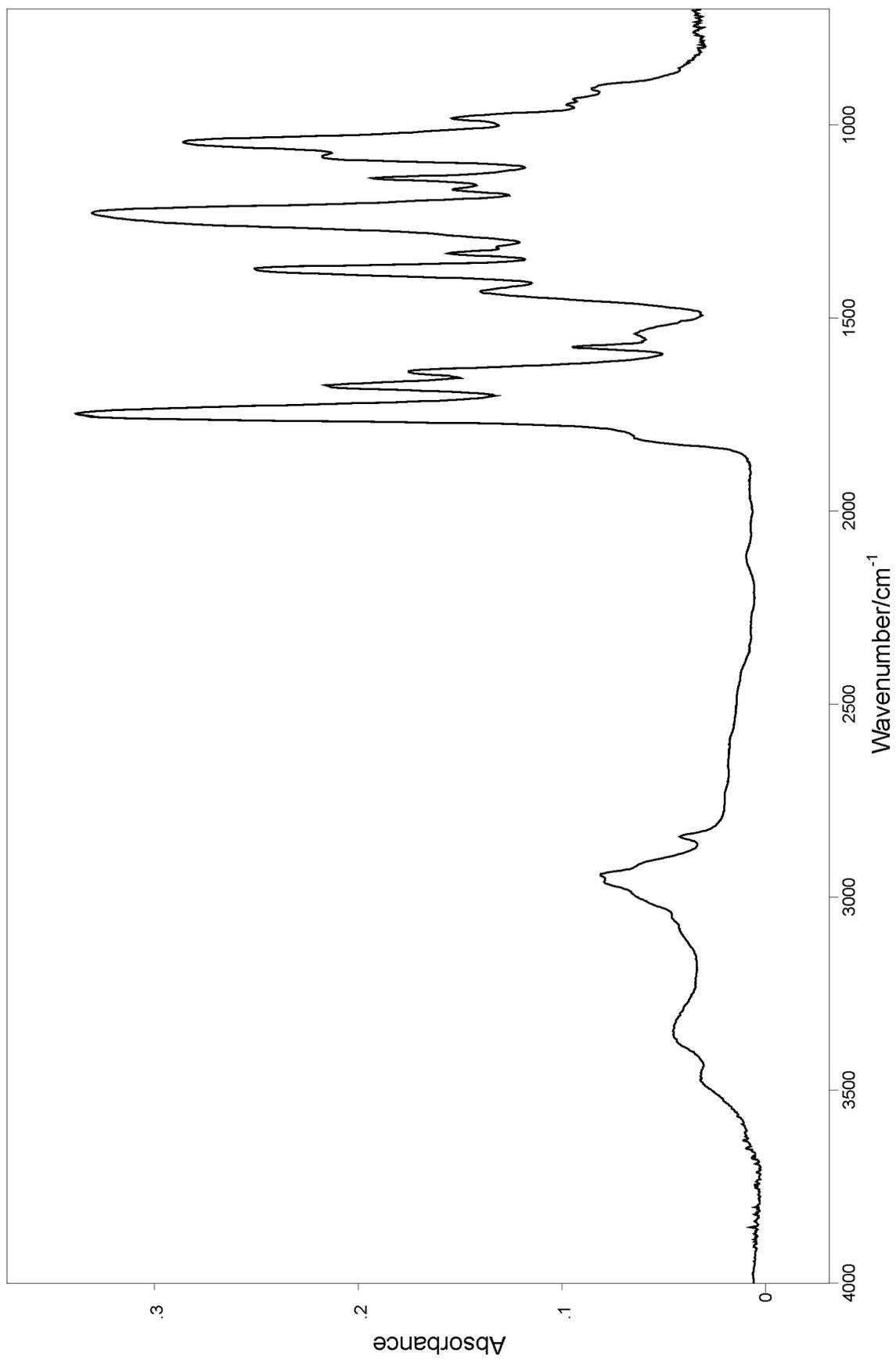
Figure 3.6   An attenuated total reflection spectrum of a peracetylated

monosaccharide mixture.
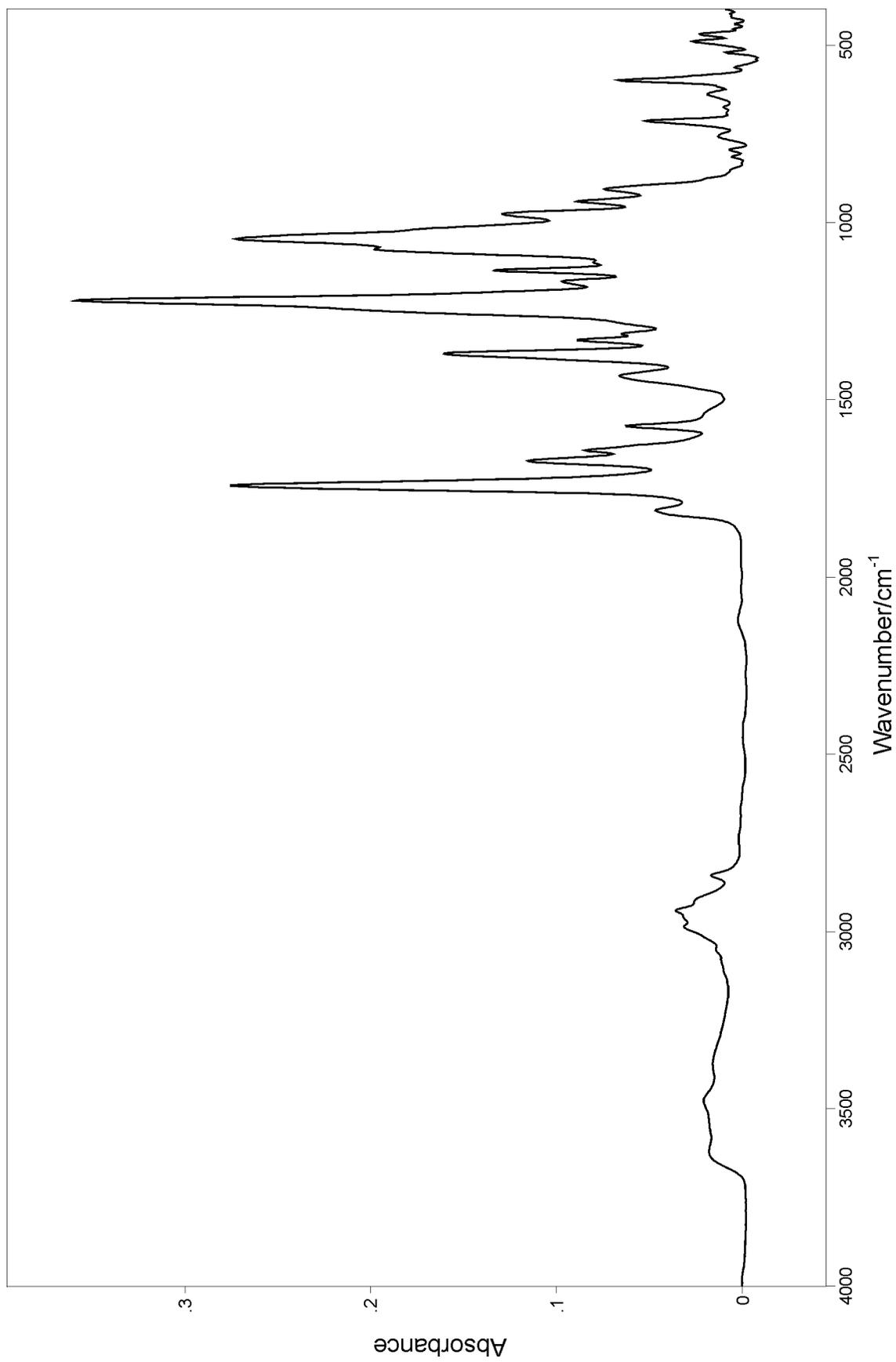
Figure 3.7   An attenuated total reflection spectrum of a peracetylated

monosaccharide mixture after undergoing a correction routine to make

the spectrum resemble a transmission spectrum.
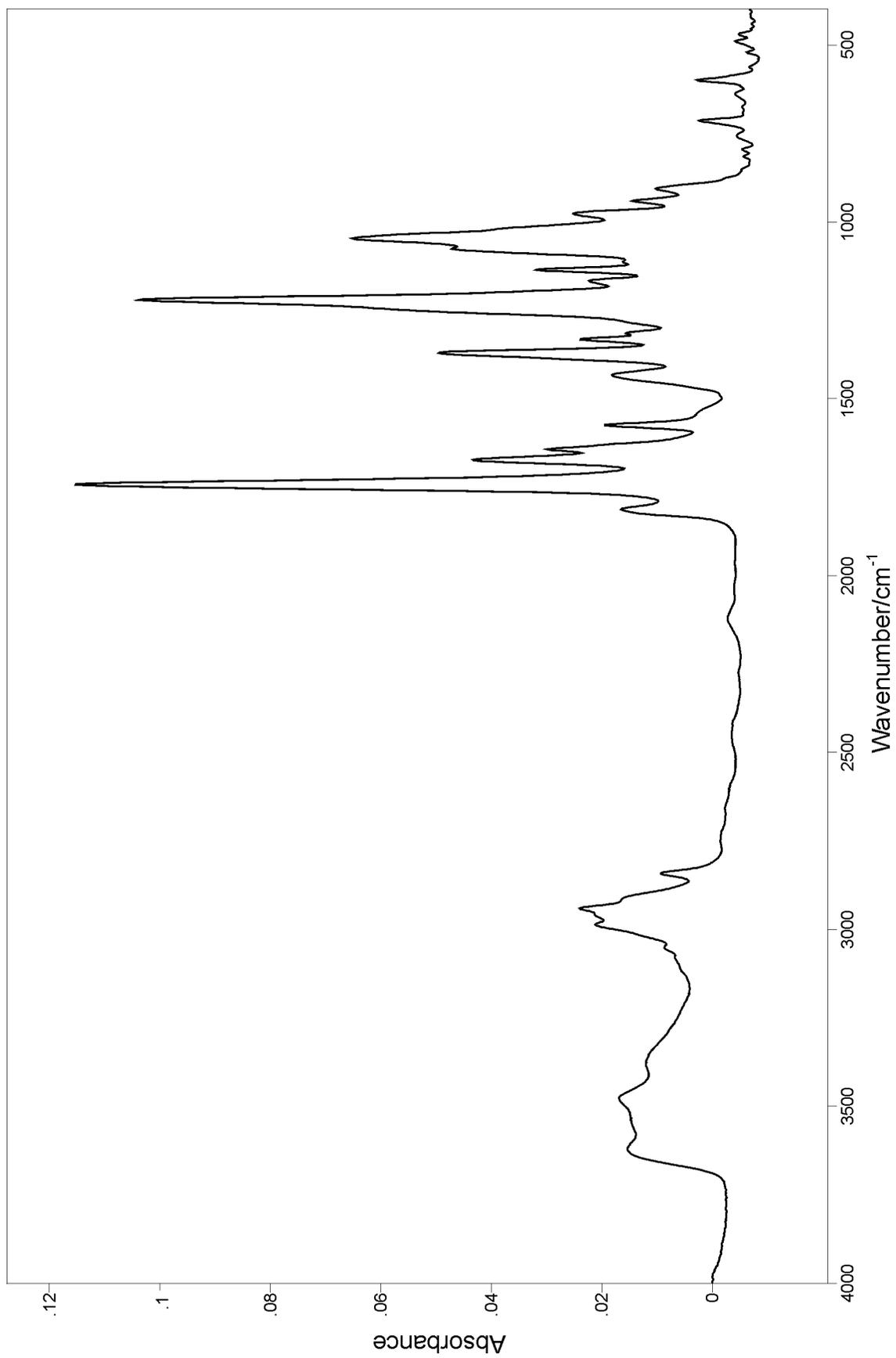
may have a preconceived expectation of the visual characteristics of the transmission spectrum, this correction has a negligible effect on the results generated through quantitative analysis.

Numerous internal reflection element designs exist. They fall into two main categories: single-reflection and multiple-reflection. The multiple-reflection elements have several types of configurations including fixed-angle plate, shown in Figure 3.8, variable-angle plate, unipoint, V-shaped, modified hemicylinder, and cylindrical rods and fibers.[19-24,35] Multiple reflection elements are not used in the present study, and therefore this discussion will be confined to single-reflection elements.

Single-reflection internal reflection elements are available in a number of different geometries. Among the varieties of single-reflection elements that have a variable-angle geometry are the hemicylindrical and hemispherical designs.[3,26,27,33] The configuration of a hemicylindrical element collimates the optical beam to a diameter equal to the radius of the element, and the beam condensation is only in one dimension at the reflecting surface. Although this may be suitable for measuring bulk materials in situations where the sample contact is easily maintained, this is not optimal for measuring small amounts of sample. To overcome this limitation, the geometric design selected for this investigation is a hemispherical configuration. This geometry is a hemisphere with a cone-shaped section at the top that plateaus to facilitate contact with the sample. In addition to this advantage, beam condensation is provided by this design thereby increasing the sensitivity in the examination of small samples (Figures 3.9 and 3.10).[26] The smallest active area commercially available in an internal reflection

Figure 3.8   A schematic of a fixed-angle plate multiple-bounce internal reflection

element.

Fixed Angle Internal Reflection Element
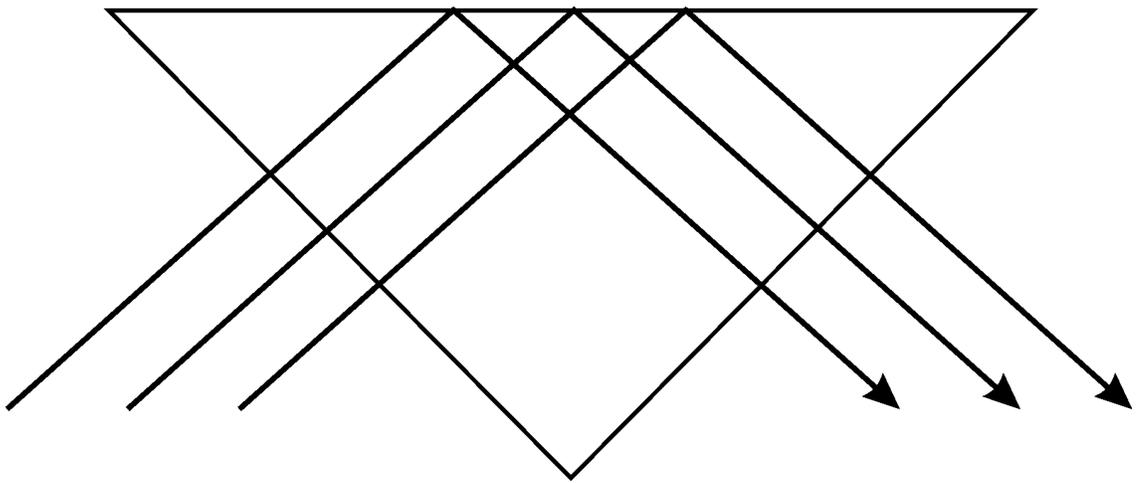
Figure 3.9   A schematic of a hemispherical internal reflection element.

Side View of Hemispherical Internal Reflection Element
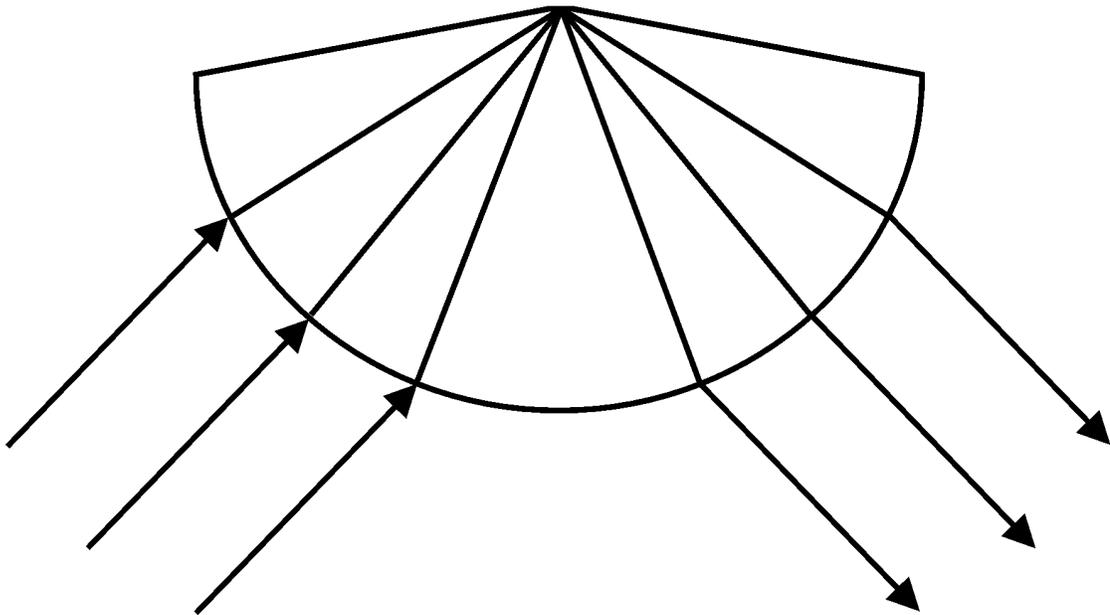
Figure 3.10   Another perspective of a hemispherical internal reflection element.

Top View of Hemispherical Internal Reflection Element
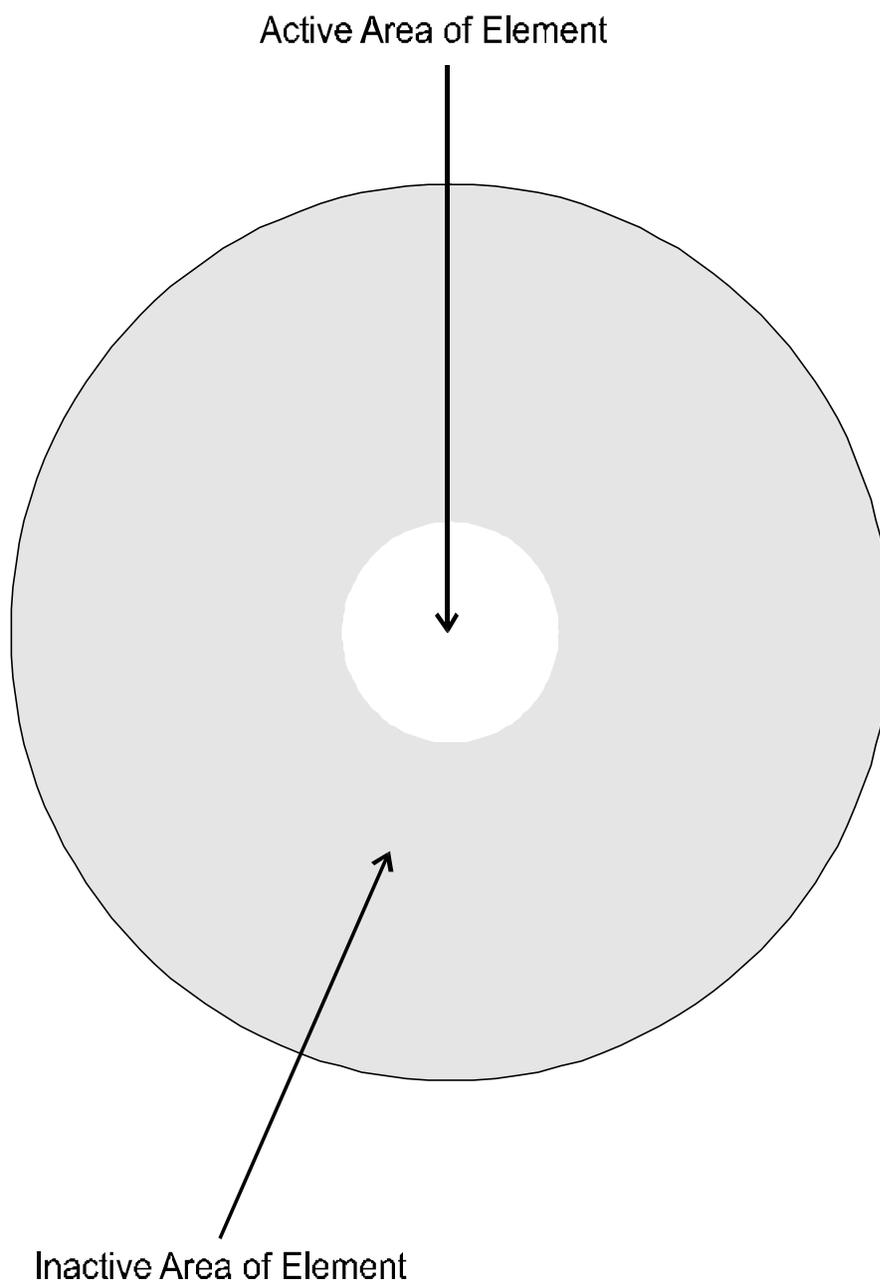
Active Area of Element

Inactive Area of Element

element is about 250 μm, which is found in the Harrick Split-Pea™ single-bounce

attenuated total reflection accessory (Harrick Scientific Corporation, Ossining, NY).

This accessory also conveniently fits directly into the sample compartment of the

spectrometer. For measuring solid samples, a pressure applicator is available to assist

in the maintenance of sample/element contact for reproducible effective pathlengths. In

direct deposition, however, the pressure applicator is not necessary, as the sample will

adhere to the surface of the element upon solvent elimination. Figure 3.11 shows how

the optical beam travels through the Split-Pea™ accessory from the infrared

spectrometer. Mirrors 1 and 2 direct the infrared beam toward mirror 3, which is an

elliptical mirror, and the beam is focused onto the internal reflection element. Then the

beam is collected by mirror 4, which is an elliptical mirror as well, and directed to

toward mirrors 5 and 6 to be sent on to the detector.

Experimental

Sample Preparation. Eighty-six monosaccharide mixtures were prepared for

the standardization set, as well as an additional twenty monosaccharide mixtures for the

validation set. The monosaccharide constituents are D-mannose, D-galactose, L-fucose,

N-acetyl-D-glucosamine, N-acetyl-D-neuraminic acid type VI,

N-acetyl-D-galactosamine, and D-glucose. The standards and all of the other reagents

and solvents were purchased from Sigma-Aldrich Co., St. Louis, MO, with the

exception of the 3 N methanolic hydrochloric acid which was purchased from Supelco,

Bellefonte, PA. The weight percentage ranges were selected to mimic the composition

Figure 3.11   A diagram showing the optical pathway through a Split-Pea™ attenuated

total reflection accessory.
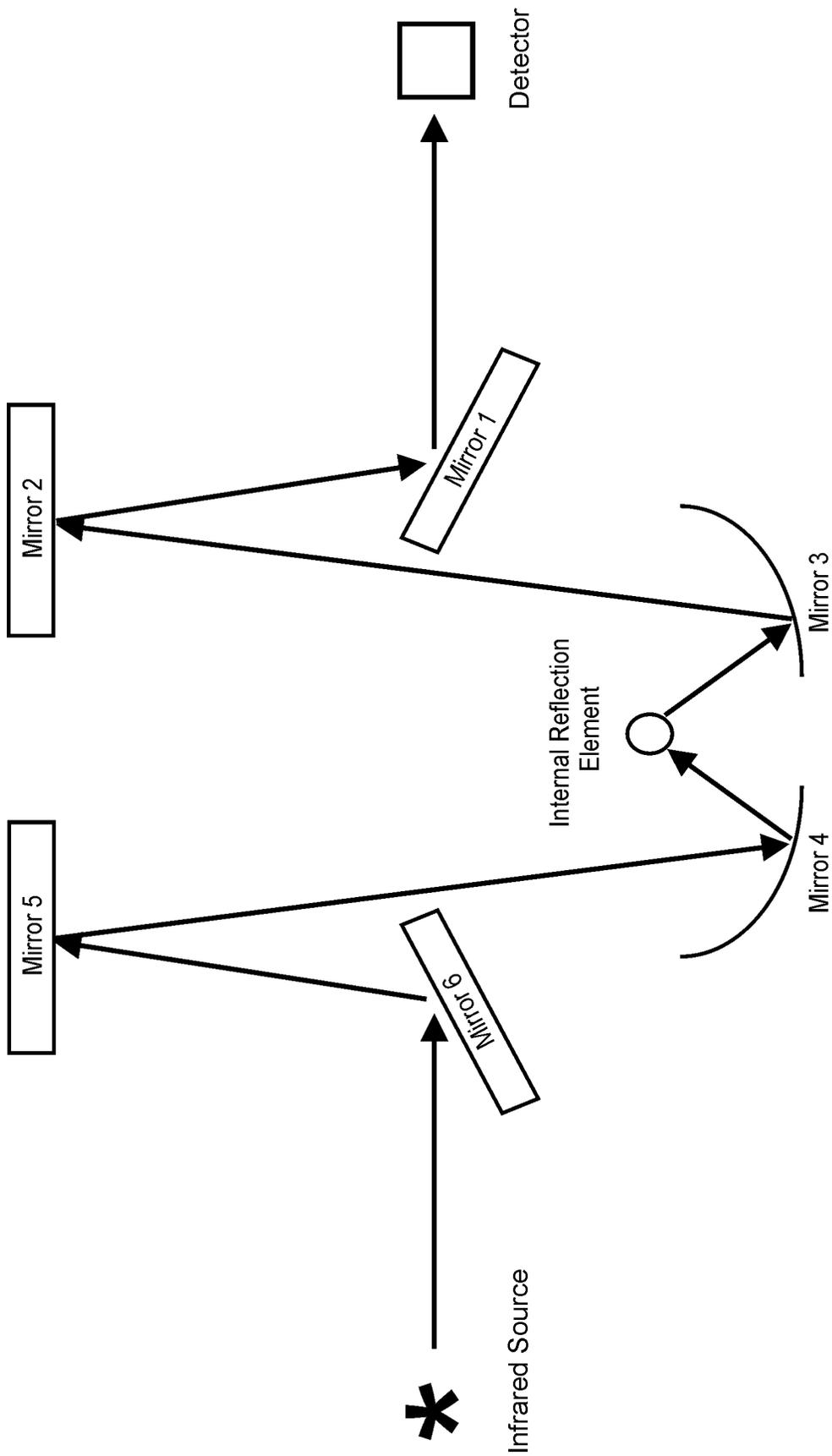
of mammalian N-linked oligosaccharides. Some of the ranges surpass what is normally encountered in oligosaccharide analysis, however, in order to increase the robustness of the model. Table 3.1 lists the compositional range for each constituent in the eighty-six training set samples. The weights of individual sugars were determined with ±0.05 mg precision on a high precision analytical balance (Ohaus Corporation, Florham Park, NJ) and subsequently diluted with 18 MΩ water to make standard solutions via a series of Eppendorf™ Autoclavable Pipettes (Brinkman Instruments, Inc., Westbury, New York) to achieve the prescribed composition for each mixture such that the resulting sugar concentration is 4 mg/mL, i.e., two milligrams of each saccharide mixture in 500 µL of water. The mixtures were contained in 17 mm x 63 mm borosilicate glass vials with polytetrafluoroethylene-lined caps (Fisher Scientific, Pittsburgh, PA). The vials were transported to and placed in a low-temperature freezer. After freezing, the solvent was completely removed from the sample via lyophilization.

The methanolysis and peracetylation procedures are as follows: Into each reaction vial, 1 mL of 1 N anhydrous methanolic hydrochloric acid, which was diluted from 3 N methanolic hydrochloric acid with high-performance liquid chromatography grade methanol, was pipetted. Particular care was taken to ensure that the vial caps were barely finger-tight immediately after addition of the reagent to minimize the exposure of the samples to atmospheric moisture. After heating for twenty minutes, the vial caps were tightened, and the reaction was maintained in a thermostatic heat block at 80 ºC for 24 hours.

Both the acid elimination with nitrogen gas and the rinse of the remaining mixtures with anhydrous methanol were executed in an N-evaporator (Organomation

Table 3.1   Range of composition for each constituent in the calibration sample set.

Table 3.1

Constituent	Minimum Composition	Maximum Composition
Gal	0%	50%
Man	5%	90%
Fuc	0%	45%
GlcNAc	5%	30%
Neu5	0%	40%
GalNAc	0%	50%
Glu	0%	50%

Associates, Inc., Berlin, MA). Equal amounts of pyridine and acetic anhydride were

dispensed into each sample. The caps were retightened, and the reaction was permitted

to continue overnight at room temperature and in the absence of light. After the

reaction was completed, the reagents were eliminated, and the products were rinsed

with toluene. Finally, the excess toluene was eliminated under vacuum, and the

products were relocated to a low-temperature freezer.

Data Acquisition and Computation. The instrumentation used to perform

single-reflection attenuated total reflection/Fourier transform-infrared spectrometry was

a DigiLab FTS 4000 Fourier transform infrared spectrometer (DigiLab Laboratories,

Randolph, MA) interfaced to a Harrick Split-Pea™ single-reflection attenuated total

reflection accessory. The accessory has a silicon hemispherical internal reflection

element with a 250 µm diameter active area and a 45° angle of incidence. The

Split-Pea™ accessory is designed to fit into the sample compartment of the spectrometer

and maintain sufficient purge for data collection. Figures 3.12 and 3.13 are schematics

of the Split-Pea™ accessory.

The methyl glycoside mixtures were dissolved in 60 µL of methylene chloride.

A 1-µL aliquot of the diluted mixture was deposited, with the use of a micro capillary,

directly onto the silicon element of a Split-Pea™ accessory taking care to ensure that the

flat portion of the internal reflection element was entirely covered. The spectrum was

monitored until the solvent completely evaporated after which the data acquisition

commenced. A deuterated tri-glycine sulfate detector was used, and the spectrometer

Figure 3.12   A Split-Pea™ accessory drawn in correct proportion. This diagram

provides another perspective on the arrangement of the optics inside.

(Copyright Harrick Scientific Corporation)
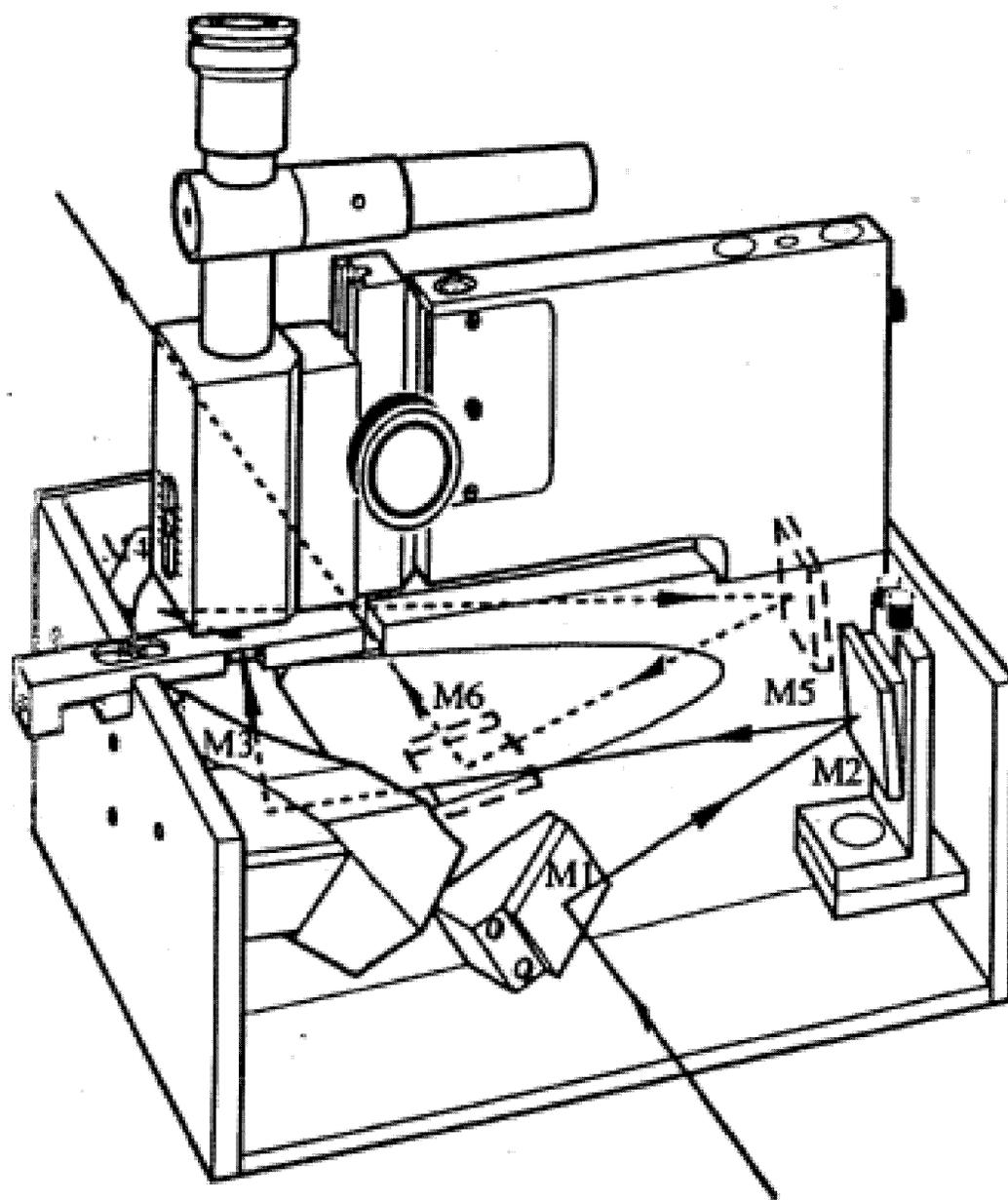
Figure 3.13   The outside of a Split-Pea™ accessory drawn in correct proportion. This

diagram provides a perspective on the placement of the accessory into an

infrared spectrometer. (Copyright Harrick Scientific Corporation)
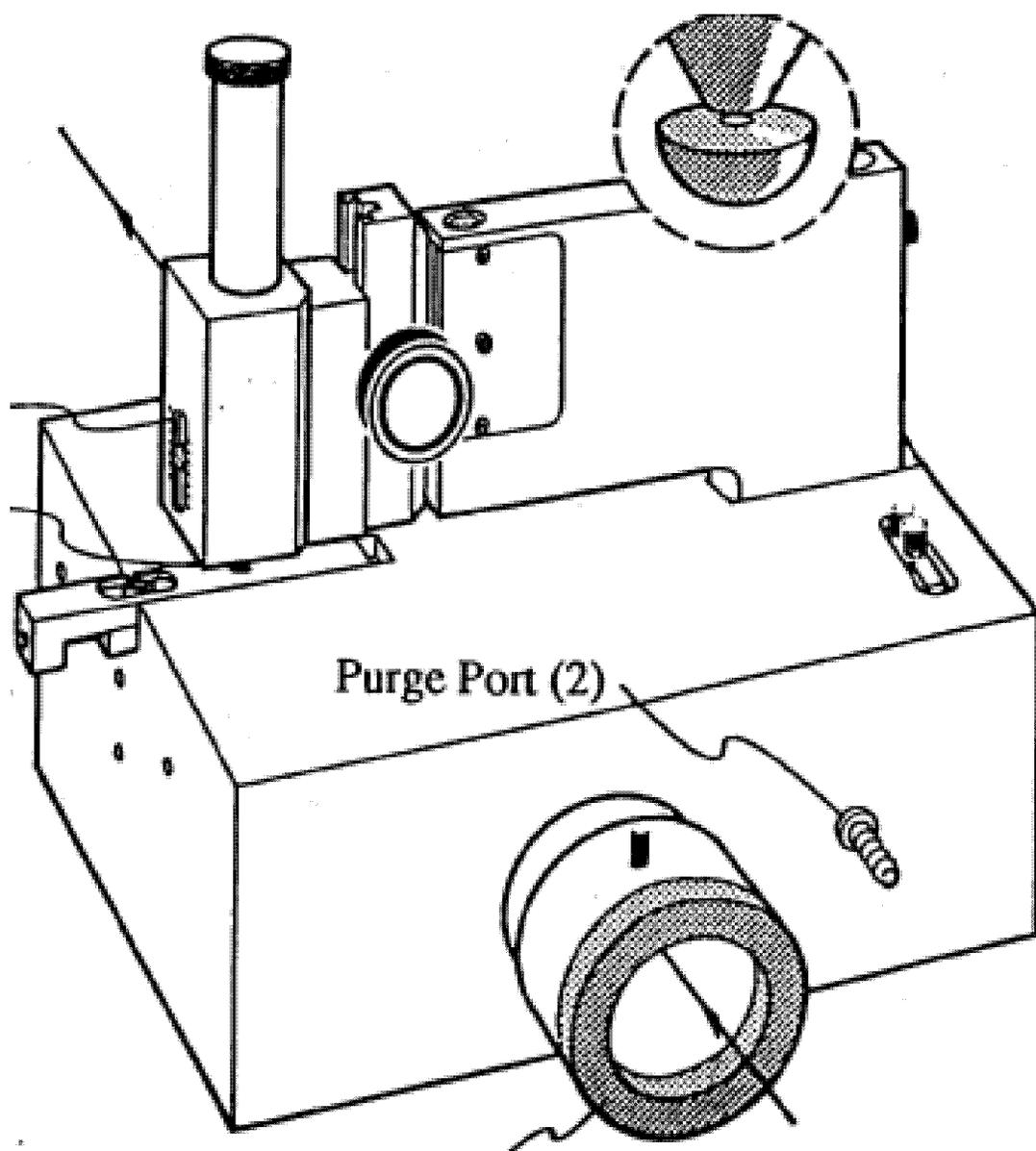
Purge Port (2)

collected reflection spectra from 4000 cm^{-1} to 400 cm^{-1} at 4 cm^{-1} resolution. Each 100-scan data set was Medium Norton-Beer apodized, Fourier transformed, and co-added. The spectrometer scanning speed, electronic filter, and undersampling ratio were set to 5 kHz, 1.2 kHz, and 2, respectively. The resulting spectra were truncated at 1850 cm^{-1} and 850 cm^{-1}. A two-point linear baseline-correction, using both leveling and zeroing functions, was applied to the spectra with 1850 cm^{-1} and 850 cm^{-1} designated as the reference points.

The Partial Least Squares algorithm, PLS-1, was used to create a model from the eighty-six training spectra. GRAMS PLSIQ™ version 6.0 was used for the statistical computations. The algorithm calculated iterations up to and including the eighty-three possible factors, utilized mean-centering, and performed a cross-validation adjusted to leave out one spectrum at a time during each of the iterations. The model was further optimized by appropriate selection of the number of factors for each of the constituents. Once the model was constructed, predictions were made on each of the internal spectra, the validation spectra, and the set of unknown spectra. The results, including predicted concentration, root-mean-square deviations, coefficients of determination, F-ratios, and prediction residual error sum of squares were imported into a spreadsheet where error analysis was tabulated.

Results and Discussion

The results provided by the single-reflection attenuated total reflection methodology demonstrated its superiority to infrared microspectrometry for this application. Table 3.2 presents the root-mean-square deviation for each of the

Table 3.2   The root-mean-squared deviations, coefficients of determination, and

number of factors for the seven monosaccharide constituents.

Table 3.2

Constituent	Root Mean Squared Deviation	Coefficient of Determination	Number of Factors
Gal	1.13%	0.9923	12
Man	0.99%	0.9976	9
Fuc	0.87%	0.9977	10
GlcNAc	0.74%	0.9859	16
Neu5	0.94%	0.9922	15
GalNAc	0.89%	0.9943	14
Glu	0.84%	0.9945	12

constituents. After being optimized, it produced correlation coefficients for each of the constituents, D‑galactose, D‑mannose, L‑fucose, N‑acetyl‑D‑glucosamine, N‑acetyl‑D‑neuraminic acid, N‑acetyl‑D‑galactosamine, and D‑glucose, equal to 0.9923, 0.9976, 0.9977, 0.9859, 0.9922, 0.9943, and 0.9945, respectively. Plots of actual concentration versus predicted concentration for each constituent are presented in Figures 3.14 through 3.20. The values of f selected for each of the above constituents are 12, 9, 10, 16, 15, 14, and 12, respectively. Figures 3.21 through 3.27 are plots of the number of factors versus the predicted residual error sum of the squares for each constituent, which assisted in the determination of the appropriate number of factors. These results were verified with a validation set.

The average error of prediction for each of the constituents among the twenty samples in the validation set are 2.83%, 2.86%, 2.62%, 1.05%, 3.33%, 2.26%, and 1.80%, respectively. Table 3.3 indicates the average error of prediction among all of the constituents for each validation sample. None of these yields an absolute error greater than four percent. The results from the blind study, of which the prediction range was comparable to that of the validation set, are presented in Table 3.4. Finally, lactose, which is a disaccharide that consists of D‑galactose and D‑glucose, was analyzed and resulted in errors of prediction of 0.83% and 2.72% for D‑galactose and D‑glucose, respectively.

Conclusions

There is no question that single-reflection attenuated total reflection spectrometry outperformed the infrared microspectrometric technique. Infrared

Figure 3.14   A plot of actual concentration versus predicted concentration for

D-galactose.

D-Galactose
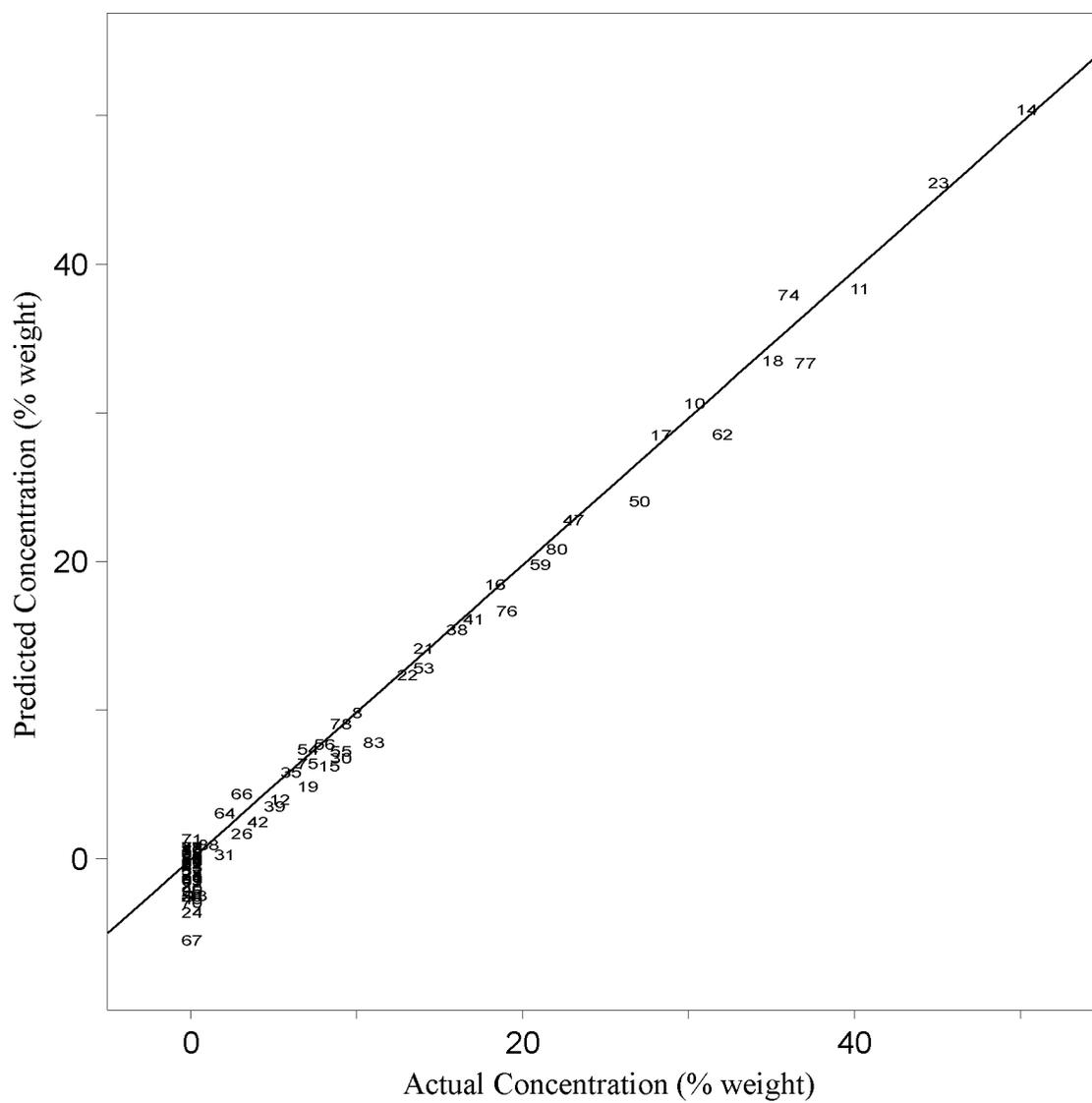
Figure 3.15   A plot of actual concentration versus predicted concentration for

D-mannose.

D-Mannose
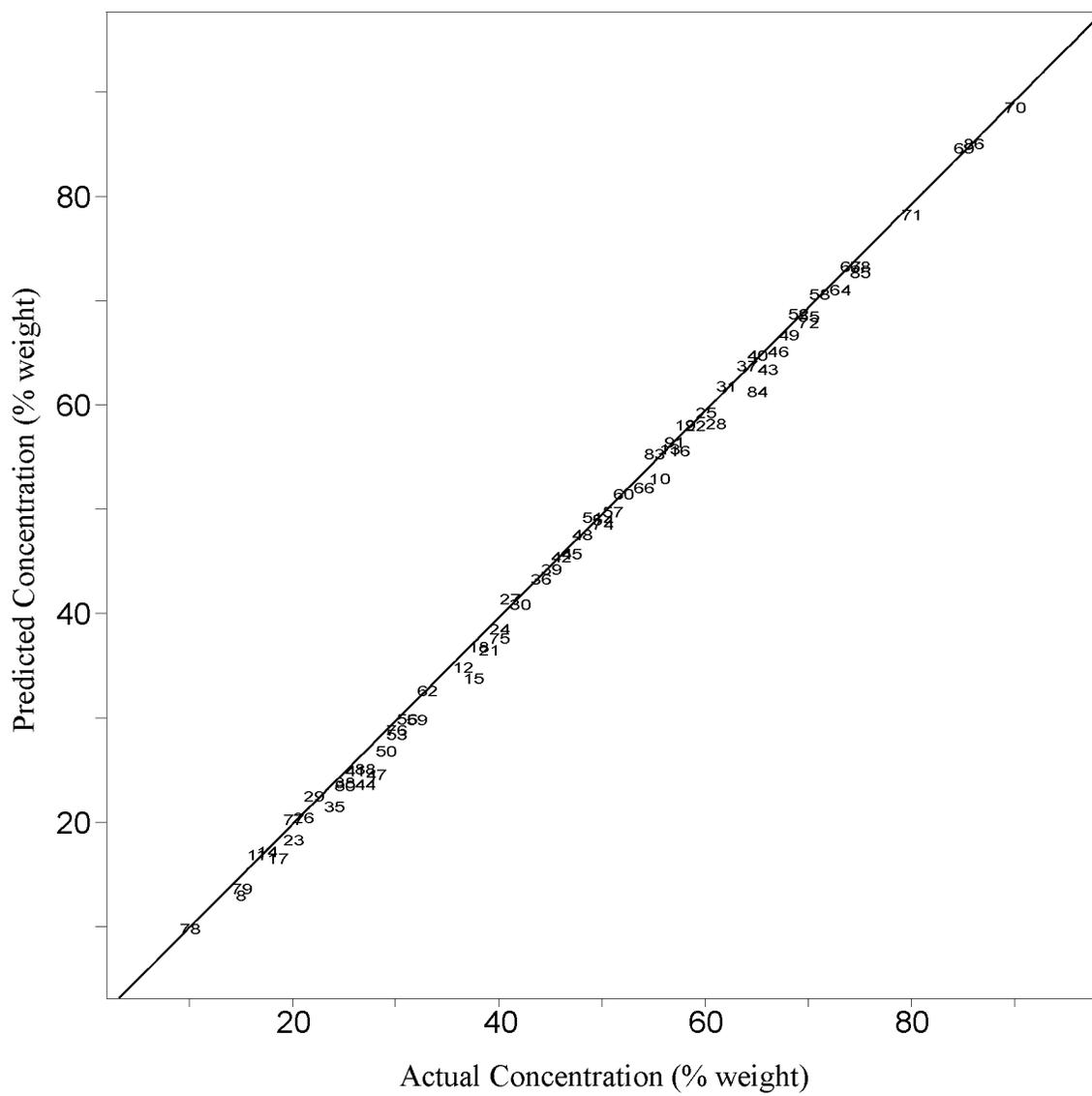
Figure 3.16   A plot of actual concentration versus predicted concentration for

L-fucose.

L-Fucose
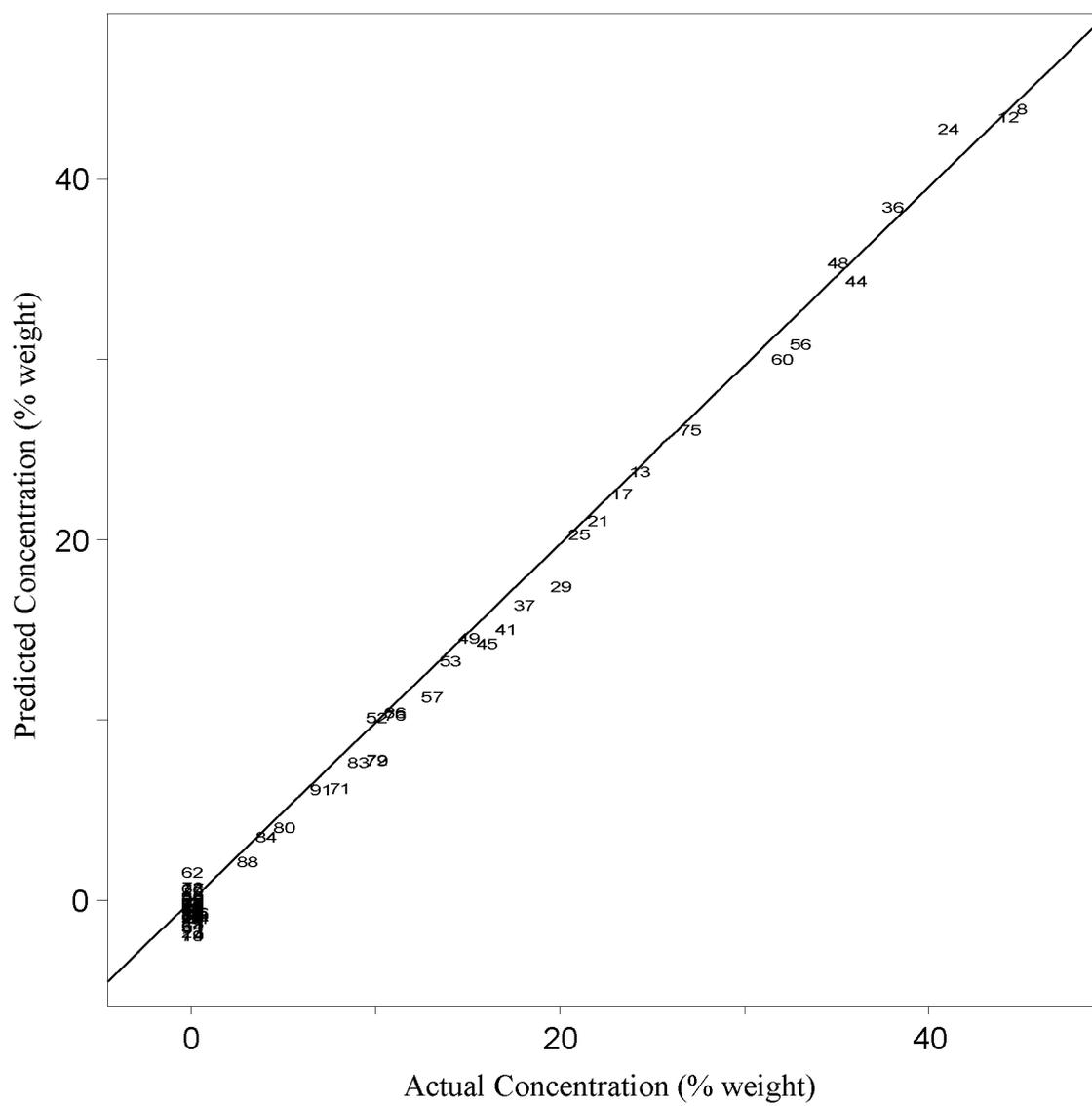
Figure 3.17   A plot of actual concentration versus predicted concentration for

N-acetyl-D- glucosamine.

N-Acetyl-D-Glucosamine
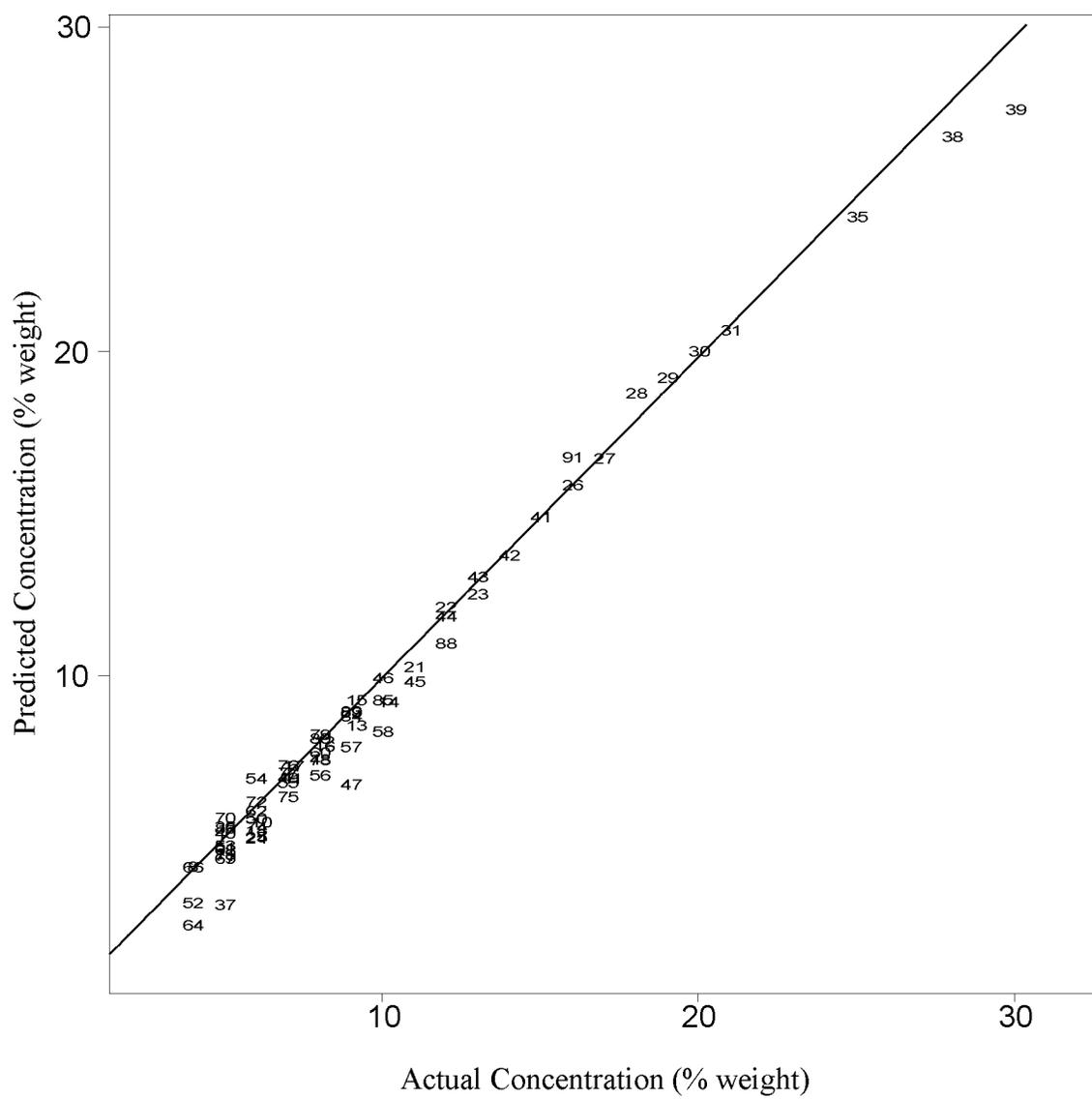
Figure 3.18   A plot of actual concentration versus predicted concentration for

N-acetyl-D-neuraminic acid.

N-Acetyl-D-Neuraminic Acid

Actual Concentration (% weight)

Figure 3.19   A plot of actual concentration versus predicted concentration for

N-acetyl-D-galactosamine.

N-Acetyl-D-Galactosamine
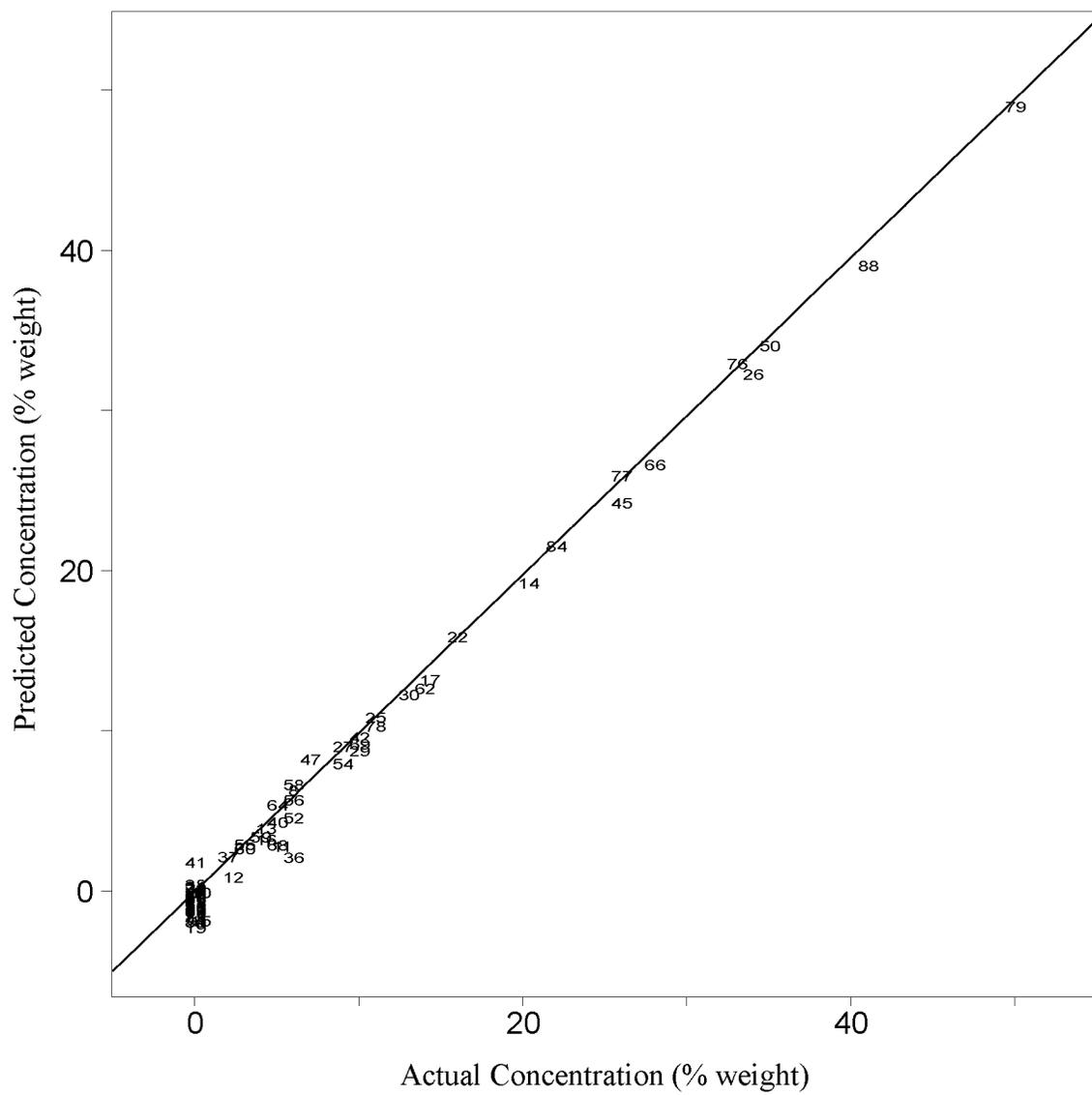
Figure 3.20   A plot of actual concentration versus predicted concentration for

D-glucose.

D-Glucose

Figure axes: Actual Concentration (% weight) versus Predicted Concentration (% weight)

Figure 3.21   A plot of predicted residual error sum of squares versus the number of

factors for D-galactose.

D-Galactose
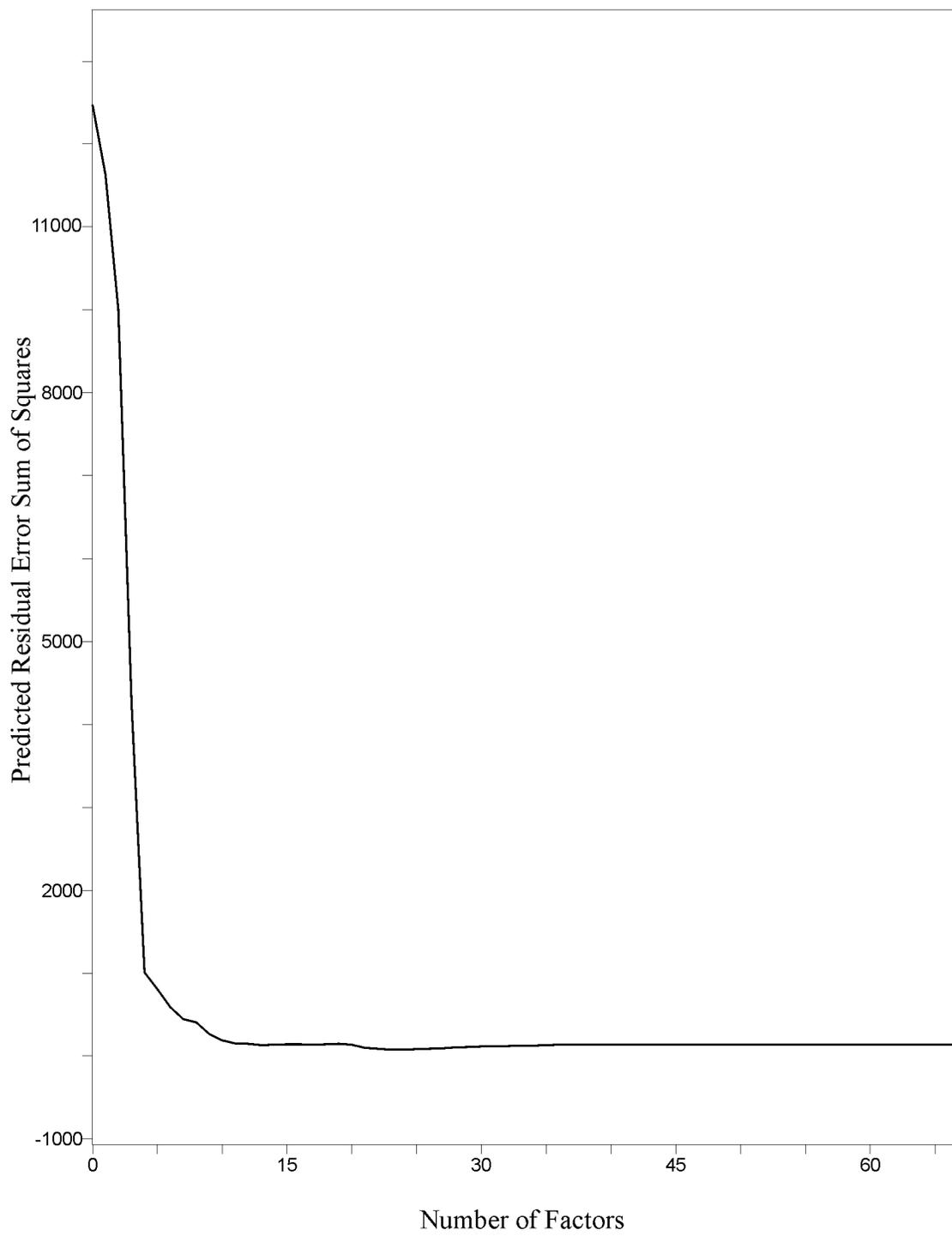
Figure 3.22   A plot of predicted residual error sum of squares versus the number of

factors for D-mannose.

D-Mannose
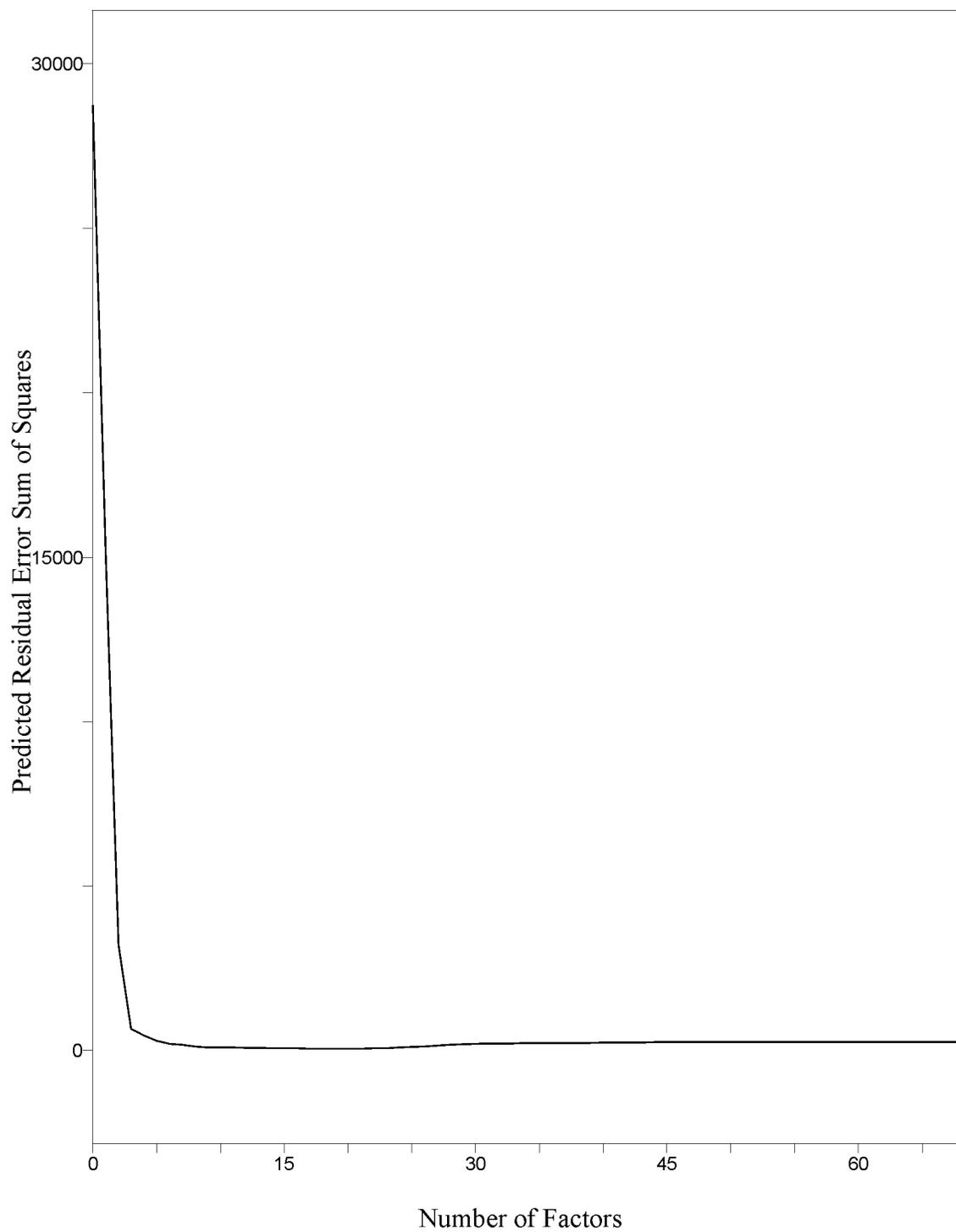
Figure 3.23   A plot of predicted residual error sum of squares versus the number of

factors for L-fucose.

L-Fucose
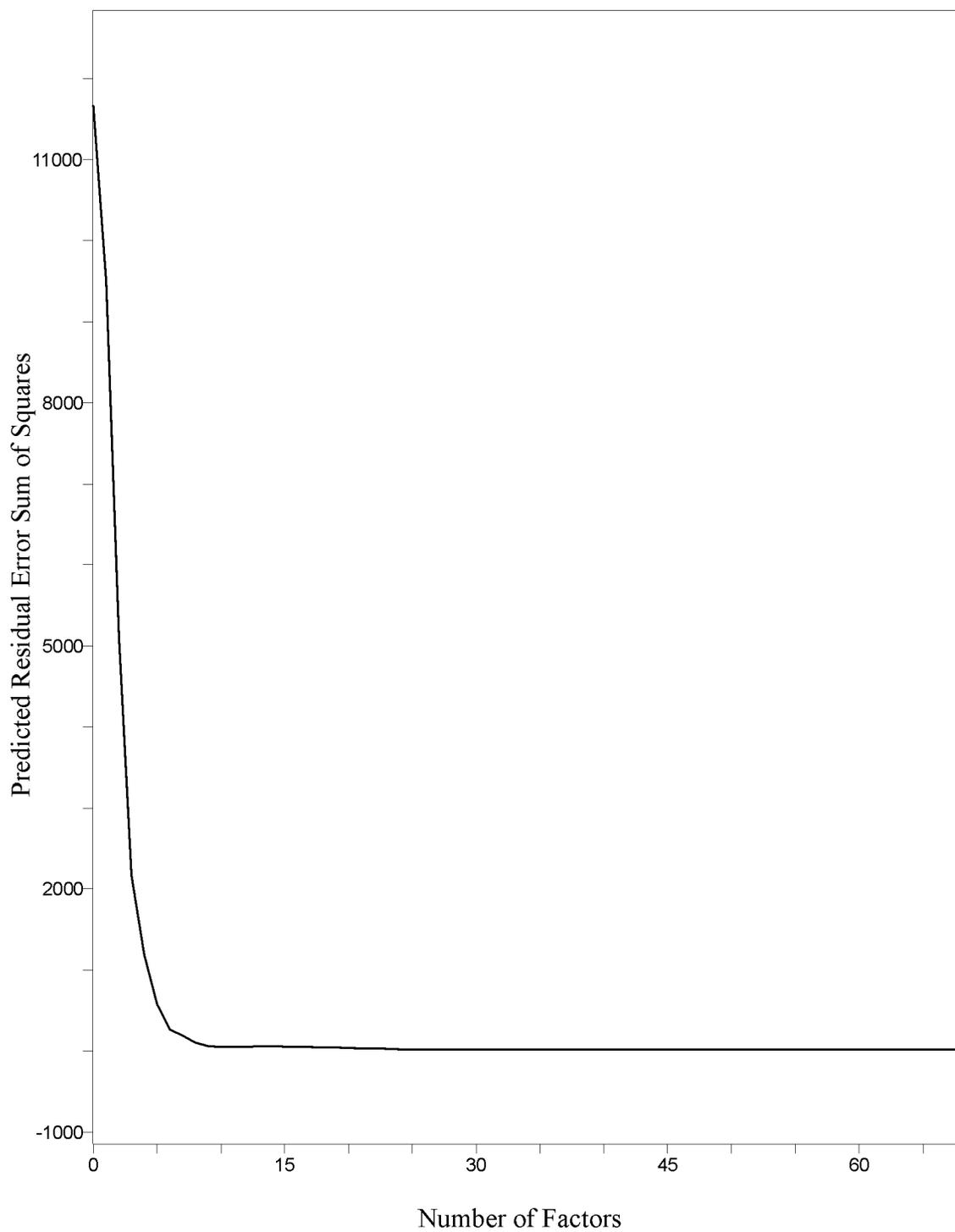
Figure 3.24   A plot of predicted residual error sum of squares versus the number of

factors for N-acetyl-D-glucosamine.

N-Acetyl-D-Glucosamine
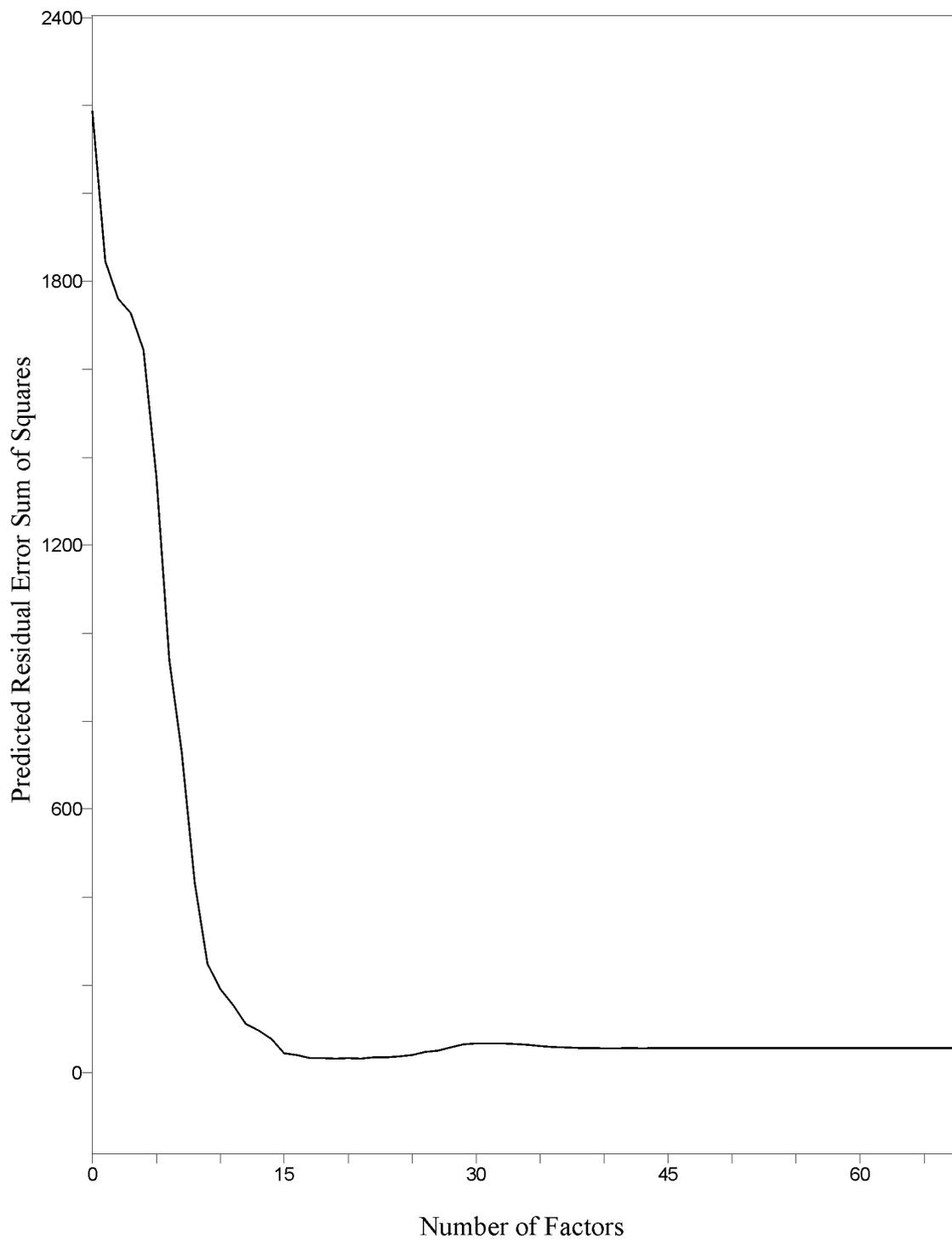
Figure 3.25   A plot of predicted residual error sum of squares versus the number of

factors for N-acetyl-D-neuraminic acid.

N-Acetyl-D-Neuraminic Acid
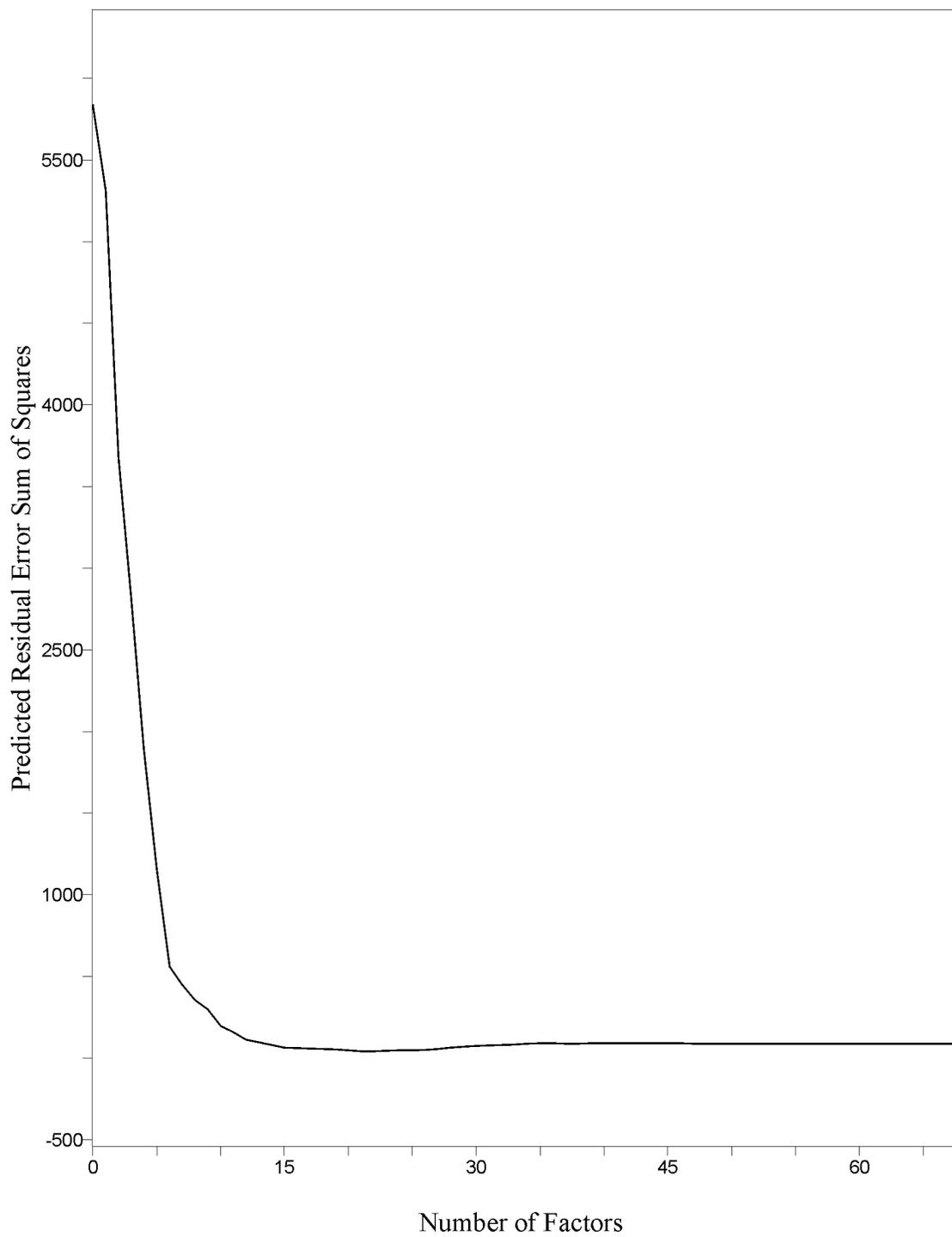
Figure 3.26   A plot of predicted residual error sum of squares versus the number of

factors for N-acetyl-D-galactosamine.

N-Acetyl-D-Galactosamine
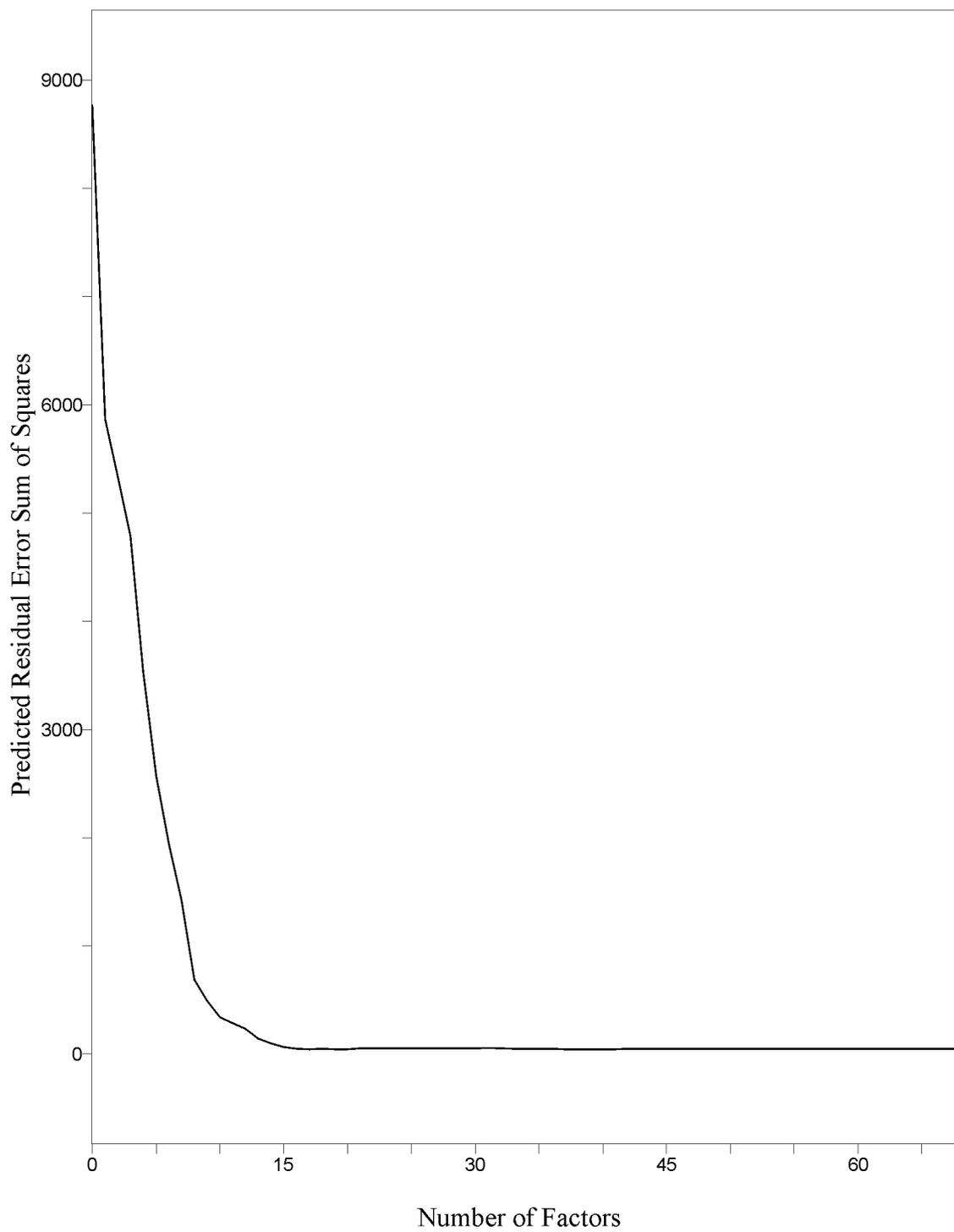
Figure 3.27 A plot of predicted residual error sum of squares versus the number of factors for D-glucose.

D-Glucose
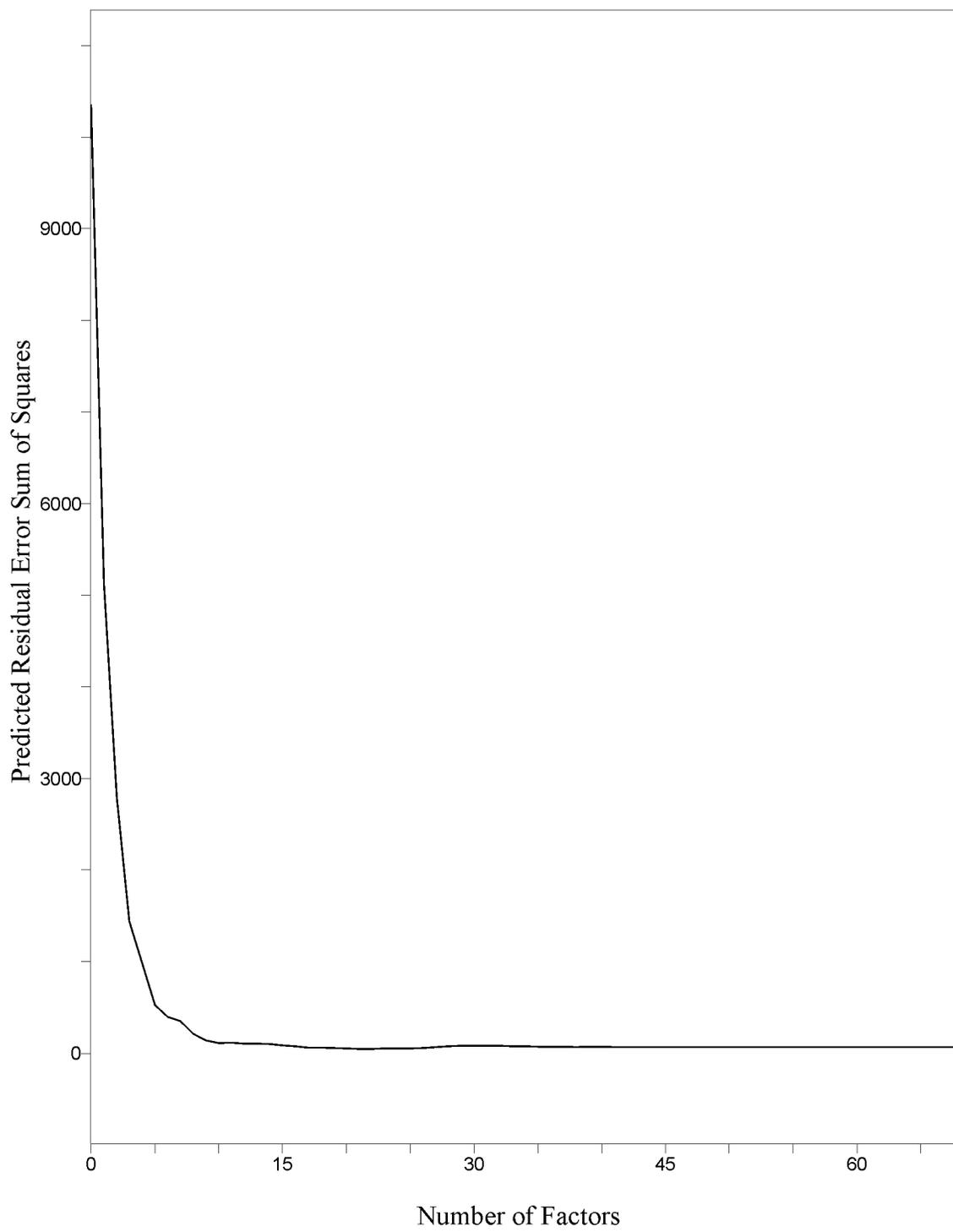
Table 3.3   The average percent error of prediction for all seven constituents for

each of the twenty validation samples.

Table 3.3

Validation Sample Number	Error of Prediction	Validation Sample Number	Error of Prediction
1	2.61%	11	3.79%
2	2.31%	12	1.47%
3	1.76%	13	2.74%
4	3.66%	14	2.04%
5	1.92%	15	3.04%
6	1.31%	16	3.86%
7	3.24%	17	2.70%
8	1.65%	18	1.36%
9	1.36%	19	3.23%
10	2.36%	20	1.44%

Table 3.4   Errors of prediction in percent for each constituent in each of the ten

unknown samples.

Table 3.4

#	Gal	Man	Fuc	GlcNAc	Neu5	GalNAc	Glu
1	4.53%	0.97%	1.83%	2.04%	2.65%	2.37%	2.66%
2	0.63%	3.03%	0.27%	0.89%	3.96%	2.33%	2.25%
3	0.86%	0.72%	1.12%	0.35%	3.45%	0.00%	1.30%
4	3.86%	1.40%	0.99%	2.17%	0.91%	0.58%	0.38%
5	5.29%	3.41%	0.43%	0.70%	5.25%	1.69%	1.79%
6	7.06%	5.58%	1.35%	1.55%	0.56%	1.47%	0.48%
7	0.37%	1.69%	3.36%	0.40%	7.73%	7.19%	0.12%
8	1.96%	0.62%	0.83%	0.79%	3.94%	0.00%	0.49%
9	4.76%	1.44%	5.09%	1.66%	4.02%	4.47%	2.05%
10	6.95%	8.39%	0.00%	0.51%	5.35%	2.48%	0.00%

microspectrometric studies in this area of research will most likely be abandoned in favor of single-reflection attenuated total reflection spectrometry. Not only is this methodology more accurate, but also it is also much easier and less time-consuming to execute. A number of reasons account for these advantages. The Split-Pea™ accessory benefits from a more reliably constant background; i.e., upon repeated background acquisitions, the spectrum drifts far less than it does in its microspectrometric counterpart. As the accessory fits into the sample compartment of the spectrometer, the system is sealed more extensively, and therefore the accessory shares a common purge with the spectrometer. Furthermore, the optical path never deviates from the purged system since the sample is dried to the surface of the reflection element and is reflected back into the accessory without ever entering the outside atmosphere. On the other hand, during operation of an infrared microscope, since the seal of the purge ring is somewhat questionable, the proximity of the operator to the instrument, and the number of other people present, substantially affects the background signal.

Another advantage from which single-reflection attenuated total reflection spectrometry benefits is the maintenance of a constant effective pathlength through the sample. This is the case so long as the thickness of the sample exceeds the depth of penetration of the infrared beam. This conclusion was ensured empirically by laying multiple deposits onto the internal reflection element until no further increase in absorption is observed in the infrared spectrum. The concentration of the mixture is constant after the solvent elimination has concluded, and only the pure sample matrix remains. The only other degree of freedom impacting the sample quantity subjected to interrogation is the area of the sample. This area is defined by the size of the internal

reflection element, which is constant, and attention is given to ensure that the active area of the element is completely covered during each interrogation. Therefore, the only remaining variable that affects spectral absorption is the chemical composition of the mixtures.

Additionally, single-bounce attenuated total reflection is much less time-consuming than measurements made by the infrared microscope system. It is not necessary to search for a suitable spot on the sample deposit for data acquisition since inhomogeneities regarding the surface of the deposit are inconsequential due to the sample/element interface being constrained to the fixed geometry of the element. In this manner, the sample/atmosphere interface is never interrogated. In addition, the heater/vacuum apparatus is rendered obsolete. The heating aspect is not needed because the internal reflection element receives enough thermal energy, provided by the source of the spectrometer, to maintain the sample/solvent system above ambient temperature and promotes solvent elimination. In addition, the vacuum function is no longer necessary to influence the profile of the upward facing surface of the sample deposit.

The throughput of the internal reflection element, when properly aligned, is approximately 15% of the total energy.[26,28] As a result, it is possible to use a deuterated tri-glycine sulfate detector. Two advantages are offered by a deuterated tri-glycine sulfate detector. A wider spectral range can be acquired in comparison with that which can be measured by a mercury cadmium telluride detector. More significantly, deuterated tri-glycine sulfate detectors respond much more linearly over a wider range of spectral intensities, and are, therefore, more viable candidates for quantitative

applications. Finally, using the hemispherical configuration of the internal reflection element, beam condensation is provided, which ultimately reduces the sample mass requirement for the successful implementation of the technique. These strengths may facilitate the development of future extensions of this application.

The amount of spectral preprocessing is reduced, in comparison to that which is required for microspectrometric application, as it is no longer necessary to normalize the spectra. In fact, normalization did not significantly improve the results. As with the infrared microscope system, several computational techniques and processes were investigated, but PLS-1 was the best performing statistical algorithm for this application, and the utilization of standard spectra that had undergone two-point linear baseline-correction with leveling and zeroing functions applied to the extrema of the spectral range included in the calibration (1850 cm^{-1} to 850 cm^{-1}) provided the optimum results. Overall, the amount of time required for analysis of an unknown specimen is reduced from over an hour, as in the case of infrared microspectrometry, to about five minutes. Attenuated total reflection spectrometry is the preferred method for further investigations on different systems.

The predictive errors in this application may be contributed by the compositional accuracy of the standardization set. The validation set and unknown set produced errors comparable to the standardization set. The use of three different pipettes, while necessary, introduces a bias with regard to the accuracy of each. Even after the pipettes are calibrated within tolerances specified by the manufacturer and reproducibly deliver the same volume within these tolerances, the use of three different pipettes creates a bias since each may be calibrated to a different point of the range of

allowable tolerances. In addition, the precision of a given pipette fluctuates across the range of volume delivery for that pipette.

In future work, it is desirable to be able to automate the direct deposition method. Work on this has already commenced in our laboratory, and it is expected to be applicable to this methodology. Other areas that are to be addressed include the design of a calibration transfer method so that the analyses can take place, using spectrometers other than the one on which the calibration model was generated. Not only is this important for minimizing the number of times a calibration model has to be generated in other laboratories with other spectrometers, but it is also important to address inevitable long-term instrumental drift. Currently, a calibration model needs to be regenerated every six months before predictive accuracy begins to significantly deteriorate. A calibration transfer method may allow for future unknown spectra to be collected and corrected in order to resemble the spectrum as though it had been collected at the time of the standardization set measurements. Finally, when a calibration method is successfully implemented, it would be considerably convenient to automate the interrogation method. This may result in an overall semi-automated method for the entire analytical process.

References

1.   Richardson, R.; Griffiths, P. *Appl. Spectrosc.* **1998**, *52* (1), 143-153.

2.   Richardson, R.; Yang, H.; Griffiths, P. *Appl. Spectrosc.* **1998**, *52* (4), 565-571.

3.   Harrick, N. *Internal Reflection Spectroscopy*; Harrick Scientific Corp.: Ossining, New York, 1987.

4.   Harrick, N. *Appl. Spectrosc.* **1971**, *25* (1), 142.

5.   Winograd, N.; Harrick, N. *Appl. Spectrosc.* **1971**, *25* (1), 143.

6.   Hosono, H. *Journal of Applied Physics* **1991**, *69* (12), 8079-8082.

7.   Oberg, K.; Fink, A. *Anal. Biochem.* **1998**, *256* (1), 92-106.

8.   Poston, P.; Rivera, D.; Uibel, R.; Harris, J. *Appl. Spectrosc.* **1998**, *52* (11), 1391-1398.

9.   Sommer, A.; Hardgrove, M. *Vibrational Spectroscopy* **2000**, *24* (1), 93-100.

10.   Widayati, S.; Stephens, S.; Dluhy, R. *Mikrochim. Acta* **1997**, *14* (Supplement), 679-681.

11.   Kameoka, T.; Okuda, T.; Hashimoto, A.; Noro, A.; Shiinoki, Y.; Ito, K. *Journal of the Japanese Society for Food Science and Technology-Nippon Shokuhin Kagaku Kogaku Kaishi* **1998**, *45* (3), 192-198.

12.   Bellonmaurel, V.; Vallat, C.; Goffinet, D. *Appl. Spectrosc.* **1995**, *49* (5), 556-562.

13.   Fujii, T.; Miyahara, Y.; Watanabe, Y. *Bunseki Kagaku* **1995**, *44* (6), 465-469.

14.   Toft, J.; Kvalheim, O.; Karstang, T.; Christy, A.; Kleveland, K.; Henriksen, A. *Appl. Spectrosc.* **1992**, *46* (6), 1002-1008.

15.   Millot, J.; Allam, N.; Manfait, M. *Anal. Chim. Acta* **1994**, *295* (3), 233-241.

16.   Johnson, B.; Doblhofer, K. *Electrochimica Acta* **1993**, *38* (5), 695-701.

17.   Deblase, F.; Harrick, N.; Milosevic, M. *Journal of Applied Polymer Science* **1987**, *34* (5), 2047-2051.

18.   Harrick, N. *Annals of the New York Academy of Sciences* **1963**, *101* (3), 928-959.

19.   Harrick, N. *Anal. Chem.* **1964**, *36* (1), 188-192.

20.   Harrick, N. *Appl. Opt.* **1965**, *4* (12), 1664-1665.

21.   Harrick, N. *Anal. Chem.* **1965**, *37* (11), 765-768.

22.   Harrick, N. *Appl. Opt.* **1966**, *5* (7), 1263-1237.

23.   Harrick, N. *Appl. Opt.* **1966**, *5* (1), 1-3.

24.   Harrick, N. *Anal. Chem.* **1971**, *43* (11), 1533-1535.

25.   Harrick, N. *Appl. Spectrosc.* **1987**, *41* (1), 1-2.

26.   Harrick, N.; Milosevic, M.; Berets, S. *American Laboratory* **1992**, *24* (3), MM50-PP50.

27.   Harrick, N.; Milosevic, M.; Berets, S. *American Laboratory* **1992**, *24* (9), 29-32.

28.   Harrick, N.; Milosevic, M.; Betets, S. *Appl. Spectrosc.* **1991**, *45* (6), 944-948.

29. Levitt, R.; Harrick, N. *Journal of the Optical Society of America* **1966**, *56* (4), 557.

30. Powell, G.; Milosevic, M.; Lucania, J.; Harrick, N. *Appl. Spectrosc.* **1992**, *46* (1), 111-125.

31. Lu, Y.; Drelich, J.; Miller, J. *Journal of Colloid and Interface Science* **1998**, *202* (2), 462-476.

32. Coates, J. *Spectroscopy* **1997**, *12* (3), 16.

33. Griffiths, P. R.; de Haseth, J. A. "Fourier Transform Infrared Spectrometry"; 18; John Wiley & Sons: New York, 1986.

34. Harrick, N.; Carlson, A. *Appl. Opt.* **1971**, *10* (1), 19-23.

35. Harrick, N. *Abstracts of Papers of the American Chemical Society* **1973**.

CHAPTER 4

THE DETERMINATION OF GLUCAN STRUCTURE BY FOURIER TRANSFORM INFRARED SPECTROMETRY/ATTENUATED TOTAL REFLECTION SPECTROMETRY[*]

Abstract

Various sources of (1-3)-β-D-glucans have been found to have tumor-necrotizing affect in mammals. Many of these anti-tumor glucans contain this structure as a backbone with O-6-linked β-glucosyl branches with a degree of branching of 1:3. A preliminary investigation determined the feasibility of collecting the reflection spectrum of a powder on a Harrick Split-Pea™ attenuated total reflection accessory with a silicon internal reflection element. Mixtures of the glucans are measured in crystalline form to eliminate the effects of hydrogen bonding that may occur if in an aqueous solution. The results demonstrate a high degree of reproducibility of data on single analyte interrogations measured at the appropriately selected mechanical pressures.

This study is an investigation for the extraction of bonding information from infrared attenuated total reflection spectra of polysaccharides. A set of standards is generated by the measurement of the infrared spectra of simple polysaccharides and correlation of the binding properties through partial least squares regression. Information on the monosaccharide composition is omitted as all of the standards share a common monosaccharide subunit composition. The binding properties investigated include the ratio of alpha bonds to beta bonds.

The system which the investigation involved was maltose and cellulose. Maltose is a polysaccharide comprised solely of alpha-linked D-glucose, and cellulose is the beta version of maltose. So far, it has been determined that the following sugars are commercially available: maltobiose, maltotriose, maltotetraose, maltopentaose, maltohexaose, maltoheptaose, cellobiose, cellotriose, cellotetraose, and cellopentaose. The results provided accurate and reproducible predictions of relative number of alpha

and beta linkages in glucan polysaccharides. The results may be applied to intact

polysaccharides of biological significance.

Introduction

The lack of attention historically given to complex carbohydrates by the pharmaceutical and academic communities probably has resulted from the underestimation of the biological significance of this class of naturally occurring compounds. One of the most significant realizations in the study of carbohydrates in recent years is the recognition within the medicinal research community that they may be a tremendous source of drug discovery leads.[1-3] One such lead is that complex carbohydrates, polysaccharides in particular, have been demonstrated to possess antigenic properties against tumor development and pneumococcal and meningococcal infections, as well as other infections.[4]

The focus of the current study is the structural determination of glucan polysaccharides. Analogues of (1-3)-glucan polysaccharides are potential candidates as therapeutics for they possess a considerable level of immunostimulant activity.[1,2] Fungal (1-3)-β-D-glucans have attracted much chemical and pharmaceutical attention in the last twenty years. Other natural sources of (1-3)-β-D-glucans have been found to have tumor-necrotizing effects in mammals. Many of these anti-tumor glucans contain this structure as a backbone containing O-6-linked β-glucosyl branches with a degree of branching of 1:3.[3] That is, the ratio of branching residues to all the (1-3)-linked D-glucose residues of the backbone is equal to 1/3.

Extensive studies on a variety of fungal glucans have demonstrated the wide range of structural diversity that exists with regard to the arrangement of the O-6-linked β-glucosyl branching. For example, *Ganoderma lucidum* is a highly-branched, insoluble fungal glucan with a degree of branching equal to 4:5, whereas *Auriculalria*

auricula, another fungal glucan, has a degree of branching of only 1:16.[3] It has also been suggested that the arrangement of the O-6-linked β-glucosyl side chains plays a substantial role in the tumor growth inhibitory activities of various (1-3)-β-D-glucans.[5] Obviously, based on these examples, the structure of polysaccharides is critical to their medicinal applications, and, therefore, it would be of significant value to determine structural information of polysaccharides by a method that is both accurate and cost-effective.

A number of methods has been applied for the structural analysis of polysaccharides, including gas chromatography/mass spectrometry[6-11] and nuclear magnetic resonance spectrometry.[12-20] Although both of these methodologies are valuable tools for the determination of structure and composition for a wide range of compounds, each has inherent drawbacks. Disadvantages that both of these techniques have in common are the considerable expense of the instrumentation and the high level of expertise required for operation and interpretation of the data, as well as the difficulty for the development of automated routine analyses.[6,21]

Gas chromatography/mass spectrometry has been used routinely for analysis of complex carbohydrates.[7-12,22] It is not, however, an ideal tool for these analyses for several reasons. First, mass spectrometry typically cannot be used effectively without the separation of the components of interest prior to analysis.[6] Gas chromatography typically requires derivatization of the analytes, which introduces considerable error. The error may be attributed to the inability to control derivatization rates or extents of derivatization for multiple analyses.

Nuclear magnetic resonance spectrometry possesses several advantages in the areas where mass spectrometry is insufficient, such as its ability to distinguish compounds of identical molecular weights; however, it poses other intrinsic drawbacks. Most notably is its inherent insensitivity as a detector. Nuclear magnetic resonance spectrometry requires a large sample mass in order to achieve a sufficient signal-to-noise ratio. In addition, it suffers from the relative challenge to maintain a homogeneous magnetic field in nuclear magnetic resonance spectrometers. Therefore, it is difficult to obtain identical spectra from the same sample upon multiple acquisitions.

Fourier transform infrared spectrometry has received little attention for analysis of carbohydrates with the possible exception of sugar concentrations in food products and beverages.[23,24] It is, however, a sensitive analytical tool with a wide variety of applications. Fourier transform infrared spectrometry can frequently identify chemical species unequivocally, where other analytical techniques cannot provide definitive structural and compositional information. Furthermore, it is a rapid and simple technique, which makes it a valuable tool for process analysis and quality control. The instrumentation is relatively inexpensive and is stable enough to permit repeated automated analyses.

Complex carbohydrates have sufficiently similar spectra that visual inspection of the spectra, such as those of glucan polysaccharides to determine the structure, is nearly impossible. Minor variations in the spectra, however, are adequate for a statistical approach such as partial least squares to distinguish isomers that differ in only one linkage reliably and with high precision. In this study, it is shown that statistical

methods are not necessarily limited to the determination of concentrations of components in a mixture, and, in fact, they may be powerful tools for structural elucidation with appropriate design of the model.

An important aspect of method development for routine analysis is cost-efficiency. Isolation of biological samples is considerably expensive. Therefore, it is essential that the technique require only small sample quantities. To meet this requirement, infrared spectrometry is joined by single-reflection attenuated total reflection spectrometry. Attenuated total reflection spectrometry permits the analysis of microgram quantities of sample. Recently, there has been tremendous development in attenuated total reflection accessories to accompany infrared spectrometry, and these accessories offer a number of important advantages over conventional infrared spectrometric systems.

Mainly, attenuated total reflection spectrometry alleviates many of the sampling problems from which conventional infrared transmission experiments suffer. As many samples are not sufficiently thin optically for transmission measurements, special preparations to the sample typically have to occur. The preparation of the samples to make them suitable for transmission experiments often compromises the sample, or at least alters the morphology of the samples, by dilution or compression. Little preparation is needed for attenuated total reflection experiments, as it is only necessary to place a sample in contact with the accessory. Usually, only minor pressure application is required to ensure adequate contact with the sample, thereby preserving the integrity of the sample. As a result, this technique can be applied to minute samples

of polysaccharides, which are in powder form, so long as intimate contact between the accessory element and the sample is maintained.

The smallest active area commercially available in an internal reflection element is about 250 µm, which is found in the Harrick Split-Pea™ single-bounce attenuated total reflection accessory (Harrick Scientific Corporation, Ossining, NY) with a silicon internal reflection element. The geometry of the internal reflection element is a hemisphere with a cone-shaped section at the top that plateaus to facilitate contact with the sample. In addition to this advantage, beam condensation is provided by this configuration, thereby increasing the sensitivity in the examination of small samples. This accessory is also convenient because it fits directly into the sample compartment of the spectrometer and shares a common purge with the instrument. This enables measurements of multiple samples without breaking the system purge, which results in the preservation of a highly stable background. In addition, the sample contact area and depth of penetration are intrinsic to the geometry and composition of the internal reflection element, and as a result, remain constant throughout the study. This assists in the reproducibility of the effective pathlength, and thereby reduces the extent of spectral preprocessing, such as normalization, that is necessary for statistical analysis. As the samples to be interrogated are solids, a pressure applicator may be employed to assist in the maintenance of sample/element contact efficiency, which also contributes to the reproducibility of effective pathlengths. An advantage of interrogation of the polysaccharides in crystalline form is the elimination of the effects of hydrogen bonding on the spectra that may occur in aqueous assays.

This study is confined to the analysis of (1-4)-β-D-glucans and (1-4)-α-D-glucans, i.e., cellulose and maltose, respectively. These two classes of glucans are selected mainly due to their commercial availability. As attenuated total reflection infrared spectrometry is shown to determine the ratio of beta linkages to alpha linkages in mixtures of these two compounds, then it is also suggested that it will be able to determine other linkages, such as 1-2, 1-3, and 1-6, when mixtures of glucans containing these linkages are prepared as standards.

Experimental

Sample Preparation. Twenty-four glucan polysaccharide mixtures were prepared for the standardization set, as well as an additional five glucan polysaccharide mixtures for the validation set. The polysaccharide constituents were maltobiose, maltotriose, maltotetraose, maltopentaose, maltohexaose, maltoheptaose, cellobiose, cellotriose, cellotetraose, and cellopentaose, and were all purchased from Sigma-Aldrich Co., St. Louis, MO. The weights of individual sugars were determined with ±0.05 mg precision on a high precision analytical balance (Ohaus Corporation, Florham Park, NJ) and diluted with 18 MΩ water to make standard solutions via a series of Eppendorf™ Autoclavable Pipettes (Brinkman Instruments, Inc., Westbury, New York) to achieve the prescribed composition. The composition of the standard mixtures is shown in Tables 4.1 and 4.2. The mixtures were contained in 17 mm x 63 mm borosilicate glass vials with polytetrafluoroethylene-lined caps (Fisher Scientific, Pittsburgh, PA). The vials were transported to and placed in a

Table 4.1   The composition of the first twelve samples of the standardization set.

Table 4.1

Sample	Alpha-Linked Constituent	Beta-Linked Constituent
1	85% maltobiose	15% cellobiose
2	70% maltotriose	30% cellobiose
3	45% maltopentaose	55% cellobiose
4	15% maltoheptaose	85% cellobiose
5	70% maltotetraose	30% cellotriose
6	55% maltopentaose	45% cellotriose
7	45% maltohexaose	55% cellotriose
8	30% maltoheptaose	70% cellotriose
9	30% maltobiose	70% cellotetraose
10	15% maltotriose	85% cellotetraose
11	85% maltotetraose	15% cellotetraose
12	70% maltopentaose	30% cellotetraose

Table 4.2   The composition of the remaining twelve samples of the standardization

set.

Table 4.2

Sample	Alpha-Linked Constituent	Beta-Linked Constituent
13	45% maltobiose	55% cellopentaose
14	30% maltotriose	70% cellopentaose
15	15% maltotetraose	85% cellopentaose
16	85% maltopentaose	15% cellopentaose
17	55% maltoheptaose	45% cellopentaose
18	100% maltotriose	none
19	100% maltotetraose	none
20	100% maltopentaose	none
21	100% maltohexaose	none
22	none	100% cellobiose
23	none	100% cellotriose
24	none	100% cellotetraose

low-temperature freezer. After freezing, the water was removed from the samples via lyophilization.

Data Acquisition and Computation. The instrumentation used to perform single-reflection attenuated total reflection/Fourier transform-infrared spectrometry was a DigiLab FTS 4000 Fourier transform infrared spectrometer (DigiLab Laboratories, Randolph, MA) equipped with a Harrick Split-Pea™ single-reflection attenuated total reflection accessory. The accessory has a silicon hemispherical internal reflection element with a 250 μm diameter active area and a 45º angle of incidence. The Split-Pea™ accessory is designed to fit into the sample compartment of the spectrometer and maintain sufficient purge for data collection. A pressure applicator was used to press the mixtures directly onto the silicon element of a Split-Pea™ accessory, with care taken to ensure that the flat portion of the internal reflection element was entirely covered. The spectrum was monitored until the maximum 500 kg/cm^2 of pressure was exerted, after which the data acquisition commenced. A deuterated tri-glycine sulfate detector was used, and the spectrometer collected reflection spectra from 4000 cm^{-1} to 400 cm^{-1} at 4 cm^{-1} resolution. Each 100-scan data set was Medium Norton-Beer apodized, Fourier transformed, and co-added. The spectrometer scanning speed, electronic filter, and undersampling ratio were set to 5 kHz, 1.2 kHz, and 2, respectively.

The partial least squares algorithm, PLS-1, was used to create a model from the twenty-two training spectra. GRAMS PLSIQ™ version 6.0 was used for the statistical computations. The algorithm calculated iterations up to and including the twenty-one possible factors, utilized mean-centering and multiplicative scatter correction, and

performed a cross-validation adjusted to leave out one spectrum at a time during each of the iterations. The spectral regions selected for analysis between 3250 cm^{-1} and 2700 cm^{-1} and between 1500 cm^{-1} and 660 cm^{-1}. The model was further optimized by the appropriate selection of the number of factors. Once the model was constructed, predictions were made on each of the internal samples, the validation samples, and a set of three unknown samples. The results, including predicted composition, root-mean-square deviations, coefficients of determination, F-ratios, and prediction residual error sum of squares were imported into a spreadsheet where error analysis was tabulated.

Results and Discussion

Figure 4.1 is a spectrum of one of the calibration mixtures. Conventional baseline-correction was not used in this study, as all of the samples are solid powders. The baseline curvature in the spectra is due to light scattering effects caused by the inhomogeneous distribution of particles both in size and in orientation with respect to the incident beam of radiation. The degree of scattering is dependent on wavelength and, therefore, is not uniform throughout the spectrum; that is, its influence is generally more pronounced at longer wavelengths. To compensate for this effect, multiplicative scatter correction is utilized prior to partial least squares analysis.

This correction method is based on the supposition that the wavelength dependency of the scattering is independent of the wavelength dependency of absorption due to chemical composition.[25,26] The removal of the effects of scattering is attempted by the linearization of each spectrum to an ideal spectrum. In this case, the

Figure 4.1   A reflection spectrum of one of the calibration mixtures.
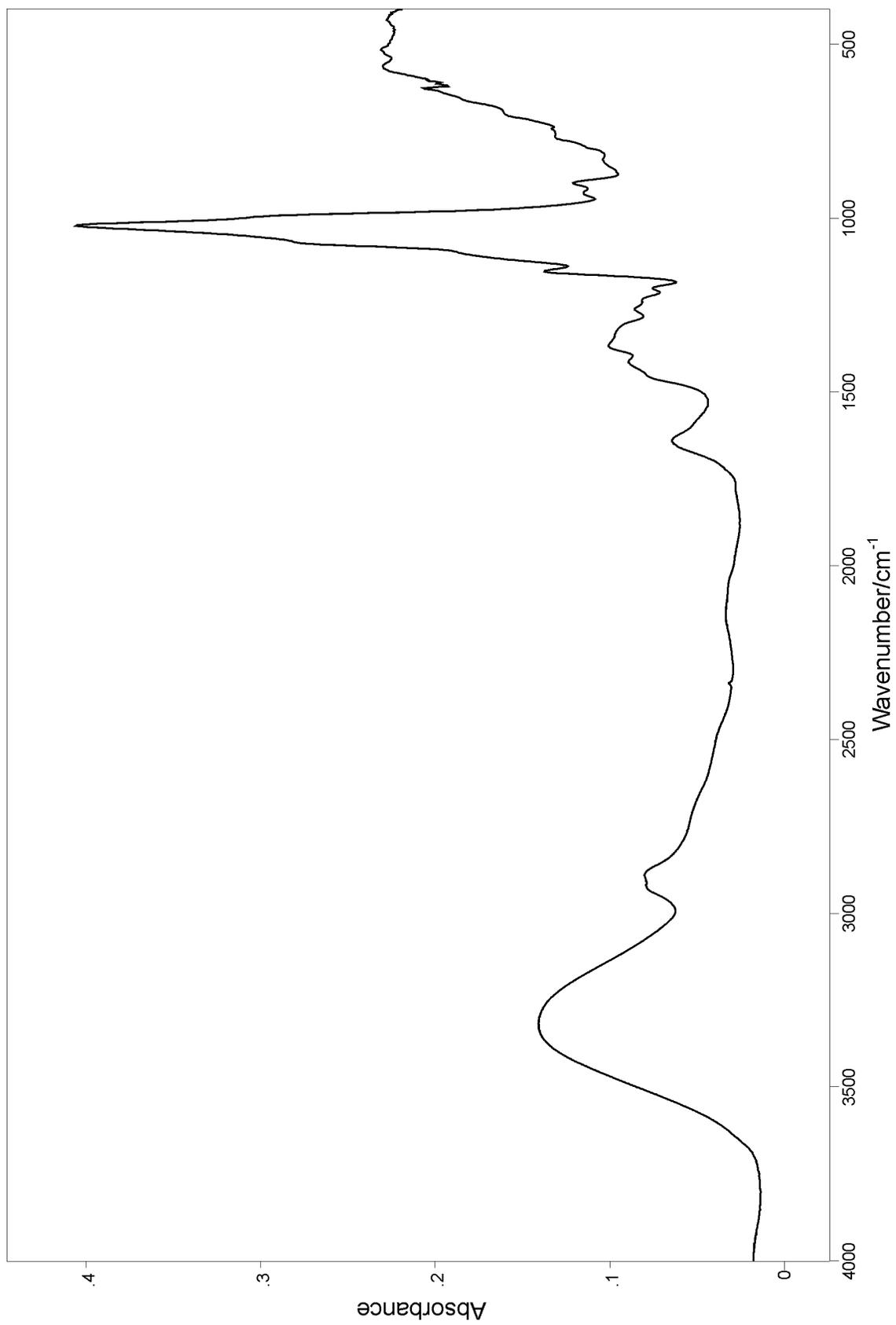
average of all of the calibration spectra serves as an estimate of an ideal spectrum. A linear regression of the spectral responses in each calibration mixture against the corresponding points in the average spectrum is calculated. Multiplicative scatter corrected spectra are subsequently produced by the subtraction of the offset value of the regression from the original spectra, then the division of the differences by the slope value from the regression. By adjusting the slope and offset of the original data to the average spectrum, the variation due to chemical composition is preserved while the major source of random variation among the spectra is minimized. In this set, two spectral regions are used for the calibration model and are multiplicative scatter corrected separately to ensure the calculation of the correct slope, as the regions are discontinuous.

The selection of the spectral regions for inclusion in the model was based upon the value of the coefficient of determination at different wavelengths. The correlation spectrum in Figure 4.2 shows the justification of the selection of the two discrete regions at 3250 cm^{-1} to 2700 cm^{-1} and 1500 cm^{-1} to 660 cm^{-1}. The ability of the model to predict the relative presence of alpha linkages and beta linkages is very high. The coefficient of determination produced by the model is 0.9985, the optimal number of factors chosen is five, the root-mean-squared deviations for the standardization set is 1.608%. The errors of prediction of each of the calibration mixtures are given in Table 4.3. A plot of predicted values versus actual values is shown in Figure 4.3, and a plot of the predicted residual error sum of squares versus the number of factors is presented in Figure 4.4. These results were verified with a validation and an unknown

Figure 4.2   The correlation spectrum for the calibration model.
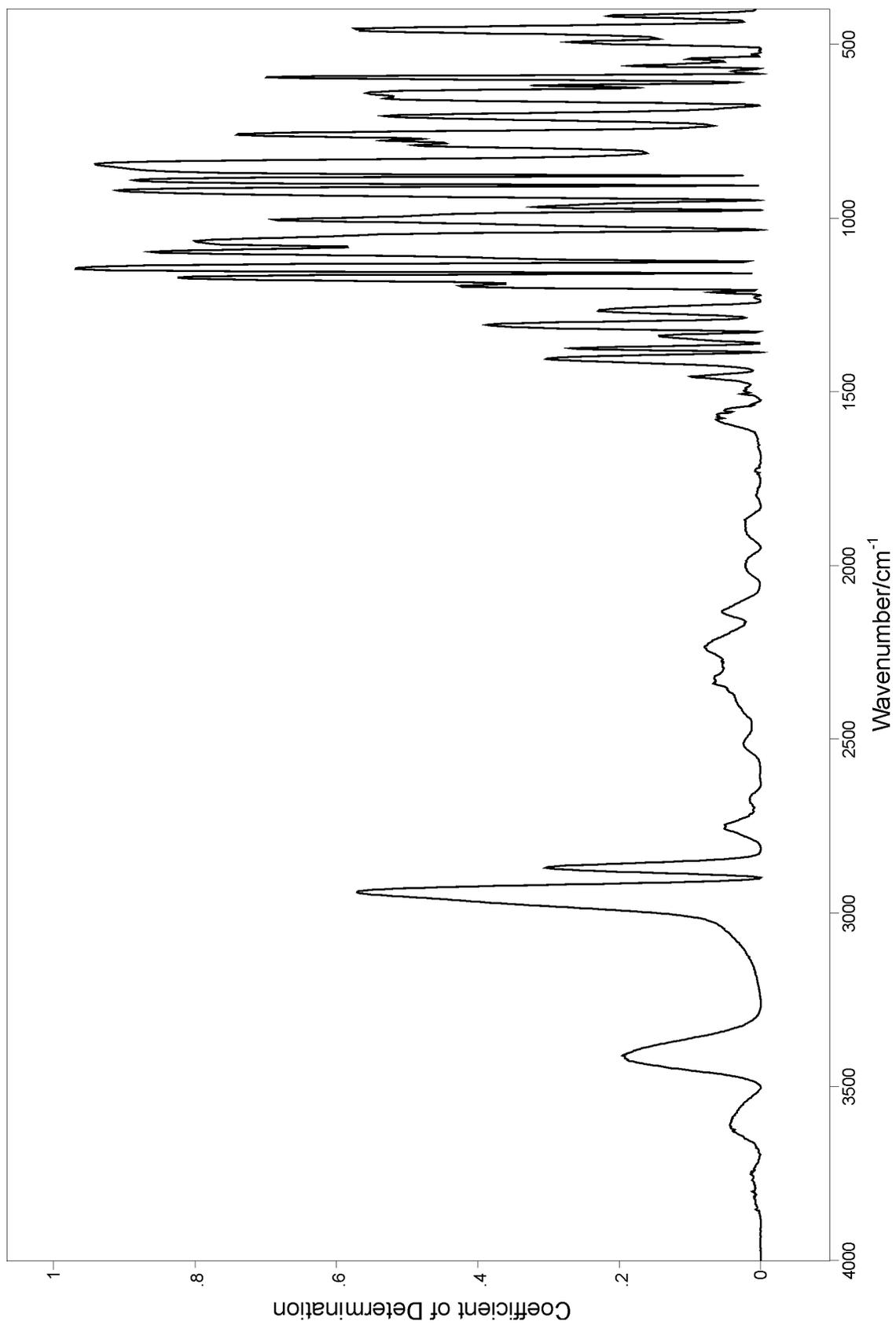
Table 4.3   The errors of prediction for the standardization set.

Table 4.3

Sample	Error of Prediction for β/(α+β)	Sample	Error of Prediction for β/(α+β)
1	2.25%	13	2.65%
2	1.67%	14	1.36%
3	1.09%	15	6.42%
4	0.11%	16	2.73%
5	0.38%	17	0.52%
6	0.08%	18	2.20%
7	0.68%	19	0.52%
8	0.74%	20	1.99%
9	2.13%	21	1.52%
10	2.91%	22	0.47%
11	0.13%	23	3.34%
12	1.50%	24	0.05%

Figure 4.3   A plot of actual β/(α+β) versus predicted β/(α+β).

$$\beta/(\alpha + \beta)$$
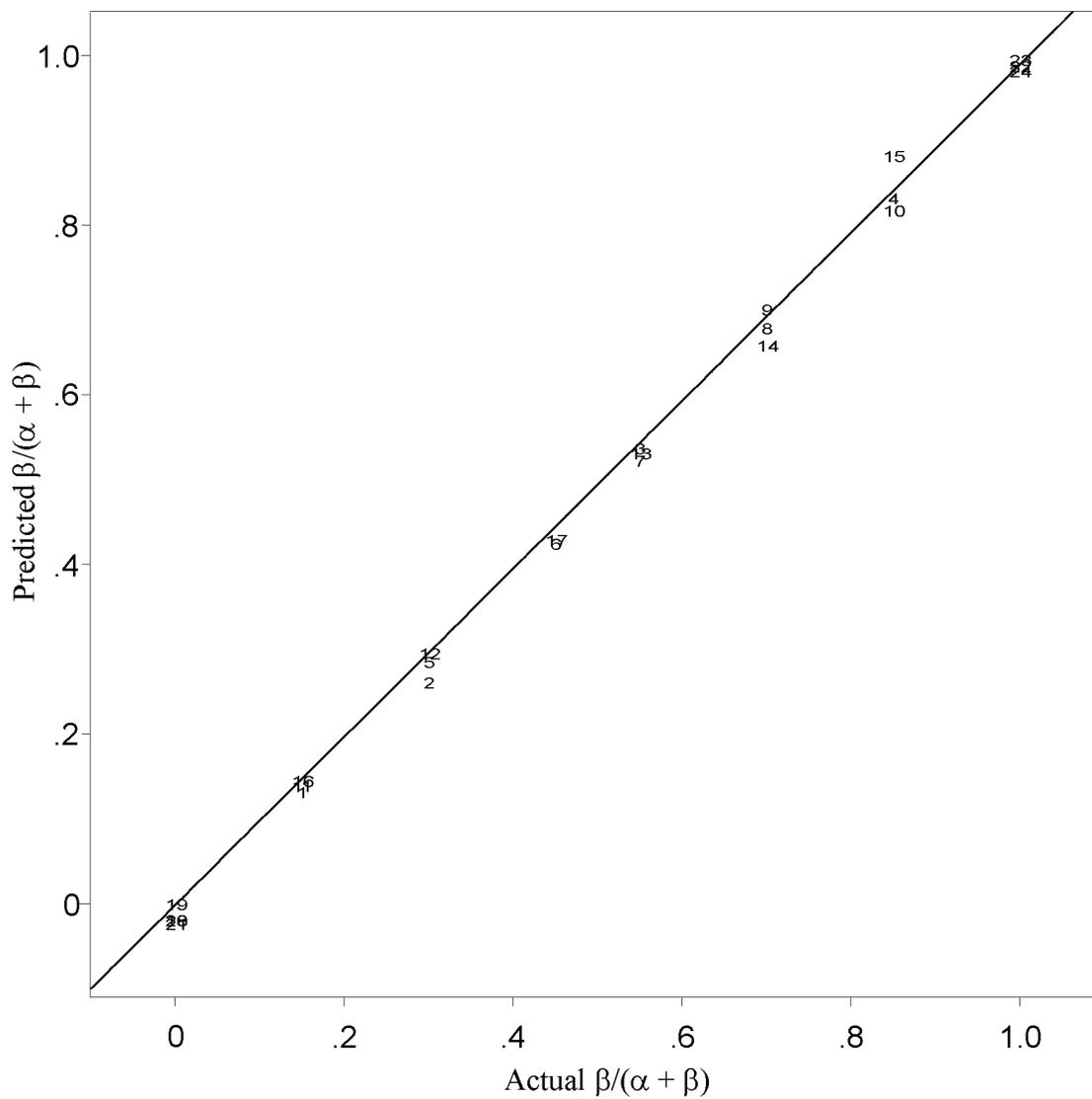
Figure 4.4   A plot of predicted residual error sum of squares versus the number of

factors for $\beta/(\alpha+\beta)$.

$$\beta/(\alpha + \beta)$$
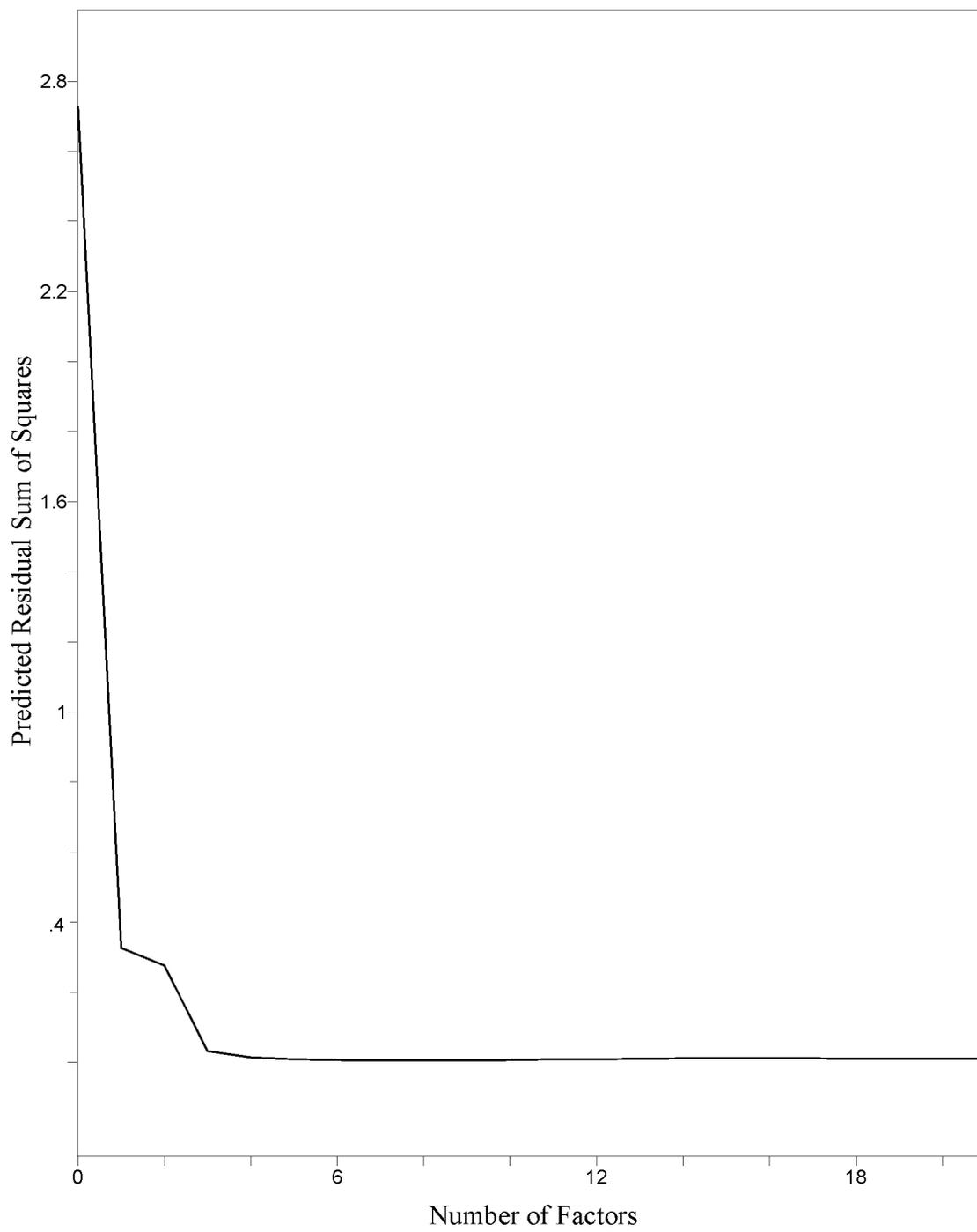
set. The results from the validation set and the unknown set are presented in Tables 4.4 and 4.5, respectively.

Conclusions

The ability of infrared spectrometry to make structural determinations of complex carbohydrates has been demonstrated unequivocally in this study. Several other statistical algorithms were investigated for their feasibility; however, partial least squares outperformed both principle component regression and multiple linear regression. As there is only one constituent determined in this system, the results from the PLS-1 and PLS-2 algorithms were identical, and neither had an advantage with respect to the time required for the completion of the calculation. In addition, mean-centering enhanced the model considerably. Other methods of scatter correction, such as standard normal variate correction, both with and without detrending, were investigated, but none surpassed multiplicative scatter correction.

The scope of this investigation was limited by the commercial availability of standards. Future studies will focus on the acquisition of other standards, which will require synthesis, to extract further structural information from polysaccharides by Fourier transform infrared spectrometry/single-bounce attenuated total reflection spectrometry. The next logical step would be an investigation of linear glucans that contain both alpha and beta linkages in the same molecule. In addition, it can be surmised from the results of this investigation that the prediction other linkages, e.g., 1-2, 1-3, and 1-6, will be a trivial matter once standards of sufficient variety are acquired. Once that proves successful, investigation of the degree of branching for this class of compounds would be quite attractive. Finally, an investigation of

Table 4.4   The average percent error of prediction for each of the five validation

samples.

Table 4.4

Validation Sample	Actual $\beta/(\alpha + \beta)$	Predicted $\beta/(\alpha + \beta)$	Error of Prediction
1	0.00%	0.00%	0.00%
2	55.00%	56.02%	1.02%
3	70.00%	70.72%	0.72%
4	15.00%	14.58%	0.42%
5	30.00%	31.13%	1.13%

Table 4.5   The average percent error of prediction for each of the three unknown

samples.

Table 4.5

Unknown Sample	Actual $\beta/(\alpha + \beta)$	Predicted $\beta/(\alpha + \beta)$	Error of Prediction
1	0.00%	1.37%	1.37%
2	45.00%	44.41%	0.59%
3	22.00%	21.80%	0.20%

polysaccharides containing different monosaccharide compositions should establish a

powerful means by which infrared spectrometry/attenuated total reflection spectrometry

may determine structures of intact oligosaccharides.

References

1. Ooi, V.; Liu, F. *Current Medicinal Chemistry* **2000**, *7* (7), 715-729.

2. Hatanaka, K. In *Polysaccharides in Medicinal Applications*; Dumitru, S., Ed.; Marcel Dekker, Inc.: New York, 1996, pp 3-105.

3. Misaki, A.; Kakuta, M. In *Carbohydrates in Drug Design*; Witczak, Z.; Nieforth, K., Eds.; Marcel Dekker, Inc.: New York, 1997, pp 655-689.

4. Varki, A. *Glycobiology* **1993**, *3* (2), 97-130.

5. Musser, J.; Kreft, A. *Journal of Medicinal Chemistry* **1992**, *35* (14), 2501-2524.

6. Torto, N.; Cohen, A.; Gorton, L.; Laurell, T.; van der Hoeven, R. *Laboratory Robotics and Automation* **1998**, *10* (6), 361-367.

7. Hachey, D. L.; Parsons, W. R.; McKay, S.; Haymond, M. W. *Anal. Chem.* **1999**, *71* (20), 4734-4739.

8. Karlsson, N. G.; Nordman, H.; Karlsson, H.; Carlstedt, I.; Hansson, G. C. *Biochemical Journal* **1997**, *326*, 911-917.

9.   Huyghuesdespointes, A.; Yaylayan, V. A.; Keyhani, A. *Journal of Agricultural and Food Chemistry* **1994**, *42* (11), 2519-2524.

10.   Leskovsek, H.; Perko, S.; Zigon, D.; Faganeli, J. *Analyst* **1994**, *119* (6), 1125-1128.

11.   Tevesz, M. J. S.; Schwelgien, S. F.; Smith, B. A.; Hehemann, D. G.; Binkley, R. W.; Carter, J. G. *Veliger* **1992**, *35* (4), 381-383.

12.   Pons, A.; Popa, J.; Portoukalian, J.; Bodennec, J.; Ardail, D.; Kol, O.; Martin-Martin, M. J.; Hueso, P.; Timmerman, P.; Leroy, Y.; Zanetta, J. P. *Anal. Biochem.* **2000**, *284* (2), 201-216.

13.   Bendiak, B.; Fang, T. T. *Carbohydrate Research* **2000**, *327* (4), 463-481.

14.   Duus, J. O.; St Hilaire, P. M.; Meldal, M.; Bock, K. *Pure and Applied Chemistry* **1999**, *71* (5), 755-765.

15.   Bendiak, B. *Carbohydrate Research* **1999**, *315* (3-4), 206-221.

16.   Mazzoni, V.; Bradesi, P.; Tomi, F.; Casanova, J. *Magnetic Resonance in Chemistry* **1997**, *35*, S81-S90.

17.   Taravel, F. R.; Mazeau, K.; Tvaroska, I. *Brazilian Journal of Medical and Biological Research* **1995**, *28* (7), 723-732.

18.   van Halbeek, H. *Current Opinion in Structural Biology* **1994**, *4* (5), 697-709.

19.   van Kuik, J.; Hard, K.; Vliegenthart, J. *Carbohydrate Research* **1992**, *235*, 53-68.

20.   Allerhand, A.; Berman, E. *J. Am. Chem. Soc.* **1984**, *106* (8), 2400-2412.

21.   Talaga, P.; Bellamy, L.; Moreau, M. *Vaccine* **2001**, *19* (20-22), 2987-2994.

22.   Tsunesada, T.; Takehara, A.; Mitsuishi, K.; Ihara, T. *Bunseki Kagaku* **2000**, *49* (6), 437-442.

23.   Sivakesava, S.; Irudayaraj, J. *Applied Engineering in Agriculture* **2000**, *16* (5), 543-550.

24.   Gallignani, M.; Garrigues, S.; Delaguardia, M. *Anal. Chim. Acta* **1994**, *287* (3), 275-283.

25.   Isaksson, T.; Naes, T. *Appl. Spectrosc.* **1988**, *42* (7), 1273-1284.

26.   Geladi, P.; Macdougall, D.; Martens, H. *Appl. Spectrosc.* **1985**, *39* (3), 491-500.

CHAPTER 5

FUTURE STUDIES

The main goal of the research described in this dissertation is the design of a methodology for the analysis of complex carbohydrates, including monosaccharides, oligosaccharides, and polysaccharides, by Fourier transform infrared spectrometry. Monosaccharide component analysis of oligosaccharides has been done routinely by gas chromatography/mass spectrometry for at least two decades.[1-6] This method requires the derivatization of the monosaccharide mixtures, which result from depolymerization of the oligosaccharides, into volatile species before undergoing separation by gas chromatography. Although high-pH anion-exchange chromatography/pulsed amperometric detection has since been used,[7-9] it is not an information-rich technique. That is, it does not possess the potential of vibrational spectrometry, as demonstrated herein, to allow for the structural elucidation of complex carbohydrates.

In this dissertation, Fourier transform infrared microspectrometry was established as a viable method for the performance of quantitative analysis of carbohydrates in conjunction with partial least squares regression. The technique was considerably enhanced, however, when single-bounce attenuated total reflection spectrometry was utilized instead of microspectrometry. The method that was developed by its application provided higher sensitivity and detection of smaller sample quantities, and there was no longer a need to normalize the spectra, as was necessary

during the microspectrometric method. This resulted in a superior analytical method, as it was more accurate and simpler to implement. Further development of Fourier transform infrared spectrometry/single-bounce attenuated total reflection spectrometry for routine analysis of carbohydrates demands the development of a calibration correction method that compensates for long-term instrumental drift. This will eliminate the need to regenerate the calibration model every several months and will permit the analysis to be transferable to other instruments and laboratories.

Perhaps the most profound findings in this research were provided by the simplest experiments. The results obtained from the measurement of neat samples of glucan polysaccharides demonstrated the power of infrared spectrometry and clearly points the direction in which future studies will progress. The accuracy in the prediction of the characterization of alpha and beta linkages in polysaccharides is extraordinary, especially when one considers that this was done completely independently of the length of the polysaccharide, which was randomized. Unfortunately, the scope of this study was limited by the variety of polysaccharide standards that are commercially available. Other polysaccharides, however, may be synthesized for future investigation. It is extremely likely that 1-2, 1-3, 1-4, and 1-6 linkages in glucan polysaccharides, with the availability of standards that contain a variety of these configurations, may be discerned with the same level of success as the characterization of α and β linkages. Finally, models built with variations in monosaccharide subunit species within the polysaccharides should establish a powerful means through which infrared spectrometry/attenuated total reflection spectrometry may determine structures of intact oligosaccharides.

References

1.   Torto, N.; Cohen, A.; Gorton, L.; Laurell, T.; van der Hoeven, R. *Laboratory Robotics and Automation* **1998**, *10* (6), 361-367.

2.   Hachey, D. L.; Parsons, W. R.; McKay, S.; Haymond, M. W. *Anal. Chem.* **1999**, *71* (20), 4734-4739.

3.   Karlsson, N. G.; Nordman, H.; Karlsson, H.; Carlstedt, I.; Hansson, G. C. *Biochemical Journal* **1997**, *326*, 911-917.

4.   Huyghuesdespointes, A.; Yaylayan, V. A.; Keyhani, A. *Journal of Agricultural and Food Chemistry* **1994**, *42* (11), 2519-2524.

5.   Leskovsek, H.; Perko, S.; Zigon, D.; Faganeli, J. *Analyst* **1994**, *119* (6), 1125-1128.

6.   Tevesz, M. J. S.; Schwelgien, S. F.; Smith, B. A.; Hehemann, D. G.; Binkley, R. W.; Carter, J. G. *Veliger* **1992**, *35* (4), 381-383.

7.   Huang, Y.; Mechref, Y.; Novotny, M. *Carbohydrate Research* **2000**, *323* (1-4), 111-125.

8.   Solbriglebuhn, H. *Zuckerindustrie* **1992**, *117* (12), 979-983.

9.   Deruiter, G.; Schols, H.; Voragen, A.; Rombouts, F. *Anal. Biochem.* **1992**, *207* (1), 176-185.