5-2007

Use of Fourier Transform Infrared Spectroscopy for the Classification and Identification of Bacteria of Importance to the Food Industry

Sarah Pegram
Utah State University

USE OF FOURIER TRANSFORM INFRARED SPECTROSCOPY FOR THE IDENTIFICATION OF BACTERIA OF IMPORTANCE TO THE FOOD INDUSTRY

by

Sarah Pegram

A thesis submitted in partial fulfillment
of the requirements for the degree

of

MASTER OF SCIENCE

in

Nutrition and Food Sciences

Approved:

UTAH STATE UNIVERSITY
Logan, Utah

2007

ABSTRACT

Use of Fourier Transform Infrared Spectroscopy for the Classification and Identification of Bacteria of Importance to the Food Industry

by

Sarah Pegram, Master of Science

Utah State University, 2007

Committee Chairman: Bart Weimer
Department: Nutrition and Food Science

The aim of this work was to use Fourier Transform Infrared Spectroscopy to characterize and identify bacteria of particular significance to the food industry. FT-IR spectroscopy is a rapid technique that can be applied to all groups of bacteria. The two objectives were to determine a suitable sampling procedure to record a spectrum and to determine a suitable statistical technique to identify characteristic regions of the spectrum associated with the genus and, potentially, the species. Pure cultures of bacteria were grown in broth, suspended in saline and dried to produce a film on a halide salt crystal. These films were then used to produce FT-IR spectra. In total, 80 spectra were recorded from seven genera, seven species and four strains of bacteria. Some of the spectra were considered to be too low in intensity to be included in statistical analysis. Data points from three specific windows of the remaining spectra were used to determine spectral distances between spectra. These spectral distances were used to perform cluster analysis using Ward's method, the Complete Linkage method and the Centroid method. The

statistical analysis created successful clusters for several of the species used but was inconclusive overall in being able to distinguish between spectra at the genus, species and strain level. This may be due to inconsistent growth of bacteria and insufficient manipulation of the data. This study has shown the potential for FT-IR spectroscopy to be used to identify bacteria with significance for food but further development is needed to reproduce the consistent results demonstrated in current literature.

(104 pages)

ACKNOWLEDGMENTS

I would like to thank my committee, Drs. Bart Weimer, Marie Walsh and Donald McMahon, for their all-around help and persistence throughout the time it has taken me to complete this thesis, including supplying the bacteria samples, giving me access to laboratory facilities and seeing me through to the end of the process. Gist-brocades of Millville, Utah also supplied bacteria samples that were used in initial trials. Thank you to Dr. Joseph Irudayaraj for initiating this project.

I would also like to thank my friends and family who have supported me in every way during my time at Utah State University. I could not have done it without your love and prayers.

Sarah Pegram

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LITERATURE REVIEW

Introduction

Characterization of microorganisms, including the detection, identification, and susceptibility against antibiotics, belongs to the most frequent tasks of a wide variety of microbiological laboratories including the food industry. Identification of bacteria is of great importance when considering foodstuffs. It is important to determine the types of bacteria present, some of which are harmful and others that are useful in the production of foods.

The dairy industry relies on microorganisms for processes in production, such as maturation in cheese production. It is also concerned with spoilage bacteria that need to be identified and monitored to maximize product quality. In the meat industry there is potential for microbiological contamination from external surfaces of the animals, which is inevitably found on finished carcasses going into chill and the resulting meat products. Although organisms that cause concern for public health are sometimes present, a more persistent problem is spoilage caused by organisms and the ensuing economic losses.

As a result, microbiological testing is done at a variety of stages in the production processes including the assessment of the bacteriological quality of herd bulk raw milk and the bacteriological examination of heat-treated milk, milk products and meat products. According to one source, the average number of tests performed for food pathogens by one research scientist was 12,786 per year (Collins et al., 2004). With recent increases in food borne illness and institution of HACCP the number of pathogen tests per year are likely to be much larger.

Rapid microbe detection methods are required that can be used at different stages of production. To date, conventional techniques include traditional biochemical and serological tests and fluorescent antibody techniques. These techniques can be cumbersome and slow requiring several tests to identify bacteria even to the genus level and each taking several days. Kits have improved the efficiency of the tests but are expensive ($3-5 per sample) and are available only for select bacteria (Collins et al., 2004). Automated procedures have been developed and are available for use in research but none are used widely. Some have been developed for industrial use but each has their limitations. For example, gas chromatography and protein electrophoresis are available but these require the extraction and manipulation of bacterial constituents prior to analysis, which can be time consuming (Naumann et al., 1991). Due to the demand for faster, more reliable, specific and cost-effective techniques for microbiological analysis, new methods should meet the following criteria: 1) rapid and specific identification of all bacteria on fairly small amounts, 2) simple and uniform operating procedure, 3) differentiation strains, and 4) simple standardization and complete computerization. Although still in the developmental stage, Fourier transform infrared (FT-IR) spectroscopy promises to meet these criteria.

Infrared spectroscopy has been employed in the past as a technique to characterize microorganisms according to these criteria (Haynes et al., 1958). However, this work was severely limited by the instrumentation available and infrared spectroscopy was dismissed as a viable technique. With more recent developments in FT-IR technology, work has been carried out that has proved to be much more successful. FT-IR spectroscopy is a nondestructive technique that requires minimal sample preparation, and

allows the rapid characterization of structural features of complex, polymeric material. The infrared (IR) portion of the spectrum is rich in information regarding the vibrational and rotational motions of atoms in molecules. Specific IR absorption frequencies can be assigned to particular types of covalent bonds, and modifications of these bonds by the local electronic environment can be detected in the details of the spectrum. Bacterial specimens produce spectra that are unique "fingerprints" and can be used to identify bacteria down to the strain level.

Since its development, FT-IR spectroscopy has been used to characterize a variety of microorganisms although work has mainly been limited to clinical analysis and pathogens (Helm et al., 1991; Holt et al., 1995). In the food industry, FT-IR spectroscopy was developed for a variety of uses including protein and moisture analysis in wheat (Osborne et al., 1982). FT-IR spectroscopy has been used to identify lactobacilli in beer (Curk et al., 1994) otherwise very little work has been done on the characterization of microorganisms in the food industry using this technique. There is therefore an encouragingly large scope for work on the use of FT-IR spectroscopy for the characterization of microorganisms, particularly relating to the food industry.

Fourier Transform Infrared Spectroscopy in Food Analysis

Fourier Transform Infrared spectroscopy has been used extensively by chemists to identify compounds and used more recently in a wide variety of applied fields. Food analysis is one such field and work has been done involving quantitative and qualitative analysis of food components.

van de Voort (1992) produced a comprehensive review of the different applications of FT-IR in food analysis. In milk analysis mid-IR is routinely used for payment, dairy herd recording and quality control (van de Voort, 1992). Analytical speed is a function of the number of components analyzed such as fat, protein, lactose and other solid components. Each component requires a filter to be interposed in the IR beam to make a measurement. The analysis of milk by FT-IR does not require a filter and has shown to be as accurate as filter-based analyses (van de Voort , 1992) taking approximately 12 sec per sample.

van de Voort (1992) used FT-IR to monitor the evaporation process of sweetened condensed milk. FT-IR provides a simple means of handling such a viscous and sticky material. Using pre-analyzed sweetened condensed milk supplied by a processor, an ATR-based method was developed to directly measure fat, protein, and total solids (wet basis). The analysis correlated well with the conventional analytical methods employed by processors to monitor the evaporation process, but requires only one minute to complete the analysis. Moisture can be determined to within ±0.5% of the values obtained by the Mojonnier vacuum oven method. Mojonnier fat and Kjeldahl protein values could also be predicted adequately (van de Voort, 1992).

Belton et al. (1987) used FT-IR for the analysis of confectionery products. Confectionery products, and chocolate in particular, represent a real challenge to any optical technique. Chocolate is dark and opaque, and on melting forms a viscous liquid. The sampling methods used were ATR and photoacoustic spectroscopy (PAS). Two methods of sample preparation were used. The simplest was to break off a piece of chocolate or solidified cocoa liquor and clamp it to the ATR crystal face. The second

method involved using the samples in a molten state and then spreading them on the ATR plate or introducing them to the PAS cell. In the solid state important information regarding lipid phase changes was not destroyed, as was the case with samples in the molten state. This can be used to distinguish between fresh and stale chocolate due to the phase separation process that occurs during storage. However, quantitative lipid analysis is best determined in the molten state since the values are affected by cooling and phase separation. In a PAS spectrum a number of bands arising from fat can be seen, with the strongest band being at 1744 cm^{-1} arising from C=O (ester). Other fat related bands occur at 1477 – 1400 cm^{-1} (C-H bend), 1240 cm^{-1} and 1195 – 1129 cm^{-1} (C-O stretching). Bands arising from the other components of the chocolate such as protein (1650 and 1540 cm^{-1}) and sugars (a number in the region 1128 – 952 cm^{-1}) can also be seen. The resolution in these spectra is such that, in principle, all components could be quantified simultaneously given suitable standards (Belton et al., 1987).

van de Voort et al. (1996) used FT-IR to determine the solid fat index (SFI) of fats and oils. A method was developed that was capable of predicting the SFI profile of a sample in approximately two minutes without the need for tempering. The calibration was largely based on spectral regions that contain information related to triglyceride weight-average molecular weight and the degree and type of unsaturation of the fatty acids that make up the triglyceride. The regions used for soybean samples were the CH-cis (3135 – 2992 cm^{-1}), the high frequency side of the carbonyl ester linkage (1790 – 1759 cm^{-1}), the low frequency side of the carbonyl ester linkage (1731 – 1620 cm^{-1}), which includes the weak C=C stretching cis/trans band, and the isolated trans (992 – 926 cm^{-1}) and cis (769 – 702 cm^{-1}) bands. One of the limitations to the FT-IR approach for the

determination of SFI is that each type of oil requires a separate calibration (van de Voort et al., 1996).

Briandet et al. (1996) used FT-IR to discriminate between Arabica and Robusta coffee beans used in instant coffee. With regard to its chemical composition, coffee is one of the most complex of commonly encountered food commodities. Arabica beans are valued the most highly because they have a more pronounced and finer flavor than Robusta beans. It is therefore important that the species of the various coffee products can be identified and quantified. Green coffee beans contain a wide range of different chemical compounds, which react and interact at different stages of coffee processing to produce a final product with an even greater diversity and complexity of structure. The most important compounds in freeze-dried coffee are carbohydrates, minerals, caffeine, chlorogenic acid, proteins, and lipids. An ATR sampling method was used for compositional analysis and the discrimination was based on the different chlorogenic acid and caffeine contents of the two species. These variations were observed in the regions 1150 – 1300 cm^{-1} and 1550 – 1750 cm^{-1} and were used to successfully discriminate between the two species (Briandet et al., 1996).

Although FT-IR spectroscopy has been used widely for quantitative and qualitative analysis, its application for classification and identification of food related components are minimal. This is especially true in the area of food microbiological analysis. The following sections will review some of the current methods used for classification of bacteria and detail the previous work done in the application of FT-IR spectroscopy to microbiological analysis.

General Methods for Detection and Identification of Bacteria

The characterization of microorganisms, including detection, differentiation, identification, and antibiotic susceptibility, is one of the most frequent tasks of a wide variety of microbiological laboratories. Currently, these analyses employ three main approaches to the identification of bacteria - 'traditional', 'kit' and 'automated' (Collins et al., 2004).

Traditional Methods

Traditional methods of identification rely on biochemical and serological tests that usually involve incubating the culture in selective agar media or broth for up to several days and then performing a specific test to determine the presence of a certain species of bacteria. Several tests are required to identify bacteria by a process of elimination. For example, the esculin hydrolysis test involves innoculating esculin medium or Edwards' medium and incubating overnight. Organisms that hydrolyse esculin blacken the medium. *Serratia marcescens* and *Edwardsiella tarda* are used as positive and negative controls, respectively. A further example is the Ammonia test that involves incubating the culture in nutrient or peptone broth for 5 d. A small piece of filter paper is wetted with Nessler reagent and placed in the upper part of the culture tube. The tube is then warmed in a water-bath and the filter paper turns black if ammonia is present. The long incubation times are required to obtain a sufficient number of organisms that can be detected by the biochemical tests. The cumulative time required is therefore quite long, at least several days (Collins et al., 2004).

Fluorescent antibody techniques have also been developed to determine the morphology of the bacteria and are used with or instead of traditional serological tests. The principle of these techniques is to label proteins, including serum antibodies, with fluorescent dyes by chemical combination without alteration or interference with the biological or immunological properties of the proteins. These proteins may then be seen in microscopic preparations by fluorescence microscopy. The fluorescent dyes are detectable in much smaller concentrations than ordinary dyes. Fluorescein isothiocyanate is most commonly used and gives an apple-green fluorescence (Collins et al., 2004).

Kit Methods

The more recent development of kits has reduced the amount of time and labor required for identification of specific organisms. Most kits consist of a disposable, multichamber device, containing 15 to 30 media or substrates, designed to identify a specific group or species. They usually require 18-24h of incubation (Doyle et al., 2001). However, there are no universal kits and some organisms cannot be identified using these methods. Kits may also suggest unlikely organisms. Numeric charts or computers are often used to interpret the results by comparing the results of the tests with a library of test results. A process of deduction using the results of the tests identifies the bacterium. Kits can be expensive and have been developed for bacteria of medical importance but may not apply to organisms of industrial significance (Collins et al., 2004).

Automated Methods

To meet the demand for rapid methods that shorten the time between the receipt of a specimen and the issue of a report, automated methods have been developed that

monitor growth with various tools that change due to cellular metabolism. The instrumentation used in these methods does not require constant attention and can be allowed to proceed overnight. One example is Vitek, developed in 1976 by BioMerieux (France). This system is based on the detection of microbial growth in microwells within plastic cards. Identification cards are available for most organisms and results are interpreted automatically within eight hours (Collins et al., 2004). Another system, Biolog, includes non-clinical isolates in its initial database making it useful for environmental purposes. The principle of the system is the ability of the test organism to reduce tetrazolium violet, incorporated in the test cells, to the purple formazan. When a carbon source is not used the microwell remains colorless, as does the control well. The resulting pattern produces what is called a 'metabolic fingerprint', which is matched to the Biolog database (Collins et al., 2004).

Impedance

Impedance instrumentation has been developed to identify bacteria based on the media that is required for growth. Impedance may be defined as resistance to flow of an alternating current as it passes through a conductor and is a function of resistance, capacitance and applied frequency. Electrodes are placed in the fluid culture medium and the conductance and/or capacitance is measured. Microbial metabolism usually results in an increase in both conductance and capacitance, causing a decrease in impedance. Tests are monitored continually and when the rate of change in conductivity exceeds the user-defined pre-set criterion the system reports growth (Collins et al., 2004).

Flow Cytometry

Flow cytometry allows cell by cell analysis of the test samples and, coupled with fluorescent labels, provides a rapid and automated method for detecting microbial flora and examining its metabolic state. The sample is injected into a 'sheath fluid', which passes under the objective lens via a hydrodynamic focusing flow cell. The sheath fluid passes continuously through the flow cell, thereby focusing the sample stream into a narrow, linear flow. The sample then passes through a light beam, which causes suitably labeled cells to emit fluorescent pulses. Each pulse is detected and subsequent analysis allows pulses to be recognized as separate counts and graded in terms of fluorescence intensities. Flow cytometry analyses and sorts cells into defined populations on the basis of cell size, density and discriminatory labelling. However, it is primarily designed for bacterial counts of specific species rather than classification or identification (Collins et al., 2004).

Pyrolysis Mass Spectrometry

Pyrolysis mass spectrometry has considerable potential for the identification, classification and typing of bacteria. Samples are spread on V-shaped Ni-Fe pyrolysis foils held in pyrolysis tubes and heated in a vacuum using Curie-point techniques. This causes pyrolysis in a controlled and reproducible manner. The gas produced is passed through a molecular beam and analyzed by a rapid scanning quadrupole mass spectrometer to produce a fingerprint of the original sample. The system is entirely automated, with sample loading, extraction, indexing to next sample and data collection controlled by computer. Routine analysis takes 90 s per sample with a batch of 300 samples. PMS is limited by requiring the organisms to be originally cultured on media

that do not impose stress and consequent alteration of phenotype expression, which may obscure the relatedness of isolates. It has also had limited industrial application and is currently used only for pathogens (Collins et al., 2004).

DNA Probes

DNA probes have been developed as a universal tool for identification of bacteria, fungi, or protozoa and their application in diagnosis and epidemiology of infectious diseases as well as in industrial and environmental microbiology. A DNA probe is a piece of single-stranded DNA which can recognize and consequently hybridize with a complementary DNA sequence. The probe also carries a label which 'lights up' after hybridization. Probe specificity depends on whether it is made from total cellular DNA, short-chain oligonucleotides or by cloning DNA fragments which gives greater specificity. There are many assay formats, although all follow the same principle whereby the DNA double helix of the target organism is separated, usually by heating, and immobilized on a membrane. Immobilization prevents the complementary strands re-hybridizing and allows access for the probe. After challenge with the probe the membrane is washed to remove any unbound probe and the resulting hybridization is visualized using dot-blot or colony-blot assays (Collins et al., 2004)

DNA fingerprinting, also known as restriction fragment length polymorphism (RFLP) analysis, combines probe and restriction enzyme technologies. The assay format is generally known as Southern blotting. Restriction enzymes cut DNA into particular fragments that can be used to produce a DNA fingerprint characteristic of the source DNA (Collins et al., 2004).

To aid the use of DNA probes, Polymerase Chain Reaction (PCR) technology can be used for DNA amplication. This is based on the repetitive cycling of three simple reactions: denaturation of double-stranded DNA, annealing of single-stranded complementary oligonucleotides and extension of oligonucleotides to form a DNA copy. The conditions for these reactions vary only in incubation temperatures, all occurring in the same tube in a cascade manner. Twenty cycles can take as little as 2h and increase the amount of target DNA by a millionfold (Collins et al., 2004).

Recent development of highly efficient thermocyclers has allowed PCR technology to be used in the rapid identification of bacterial species. Fohlman et al. (2004) used 16sRNA-PCR to identify *Haemophilus influenzae* and *Neisseria meningitides*. Using superconvection with ultracentrifugation high-speed PCR, results can be obtained within 10 minutes and the amplificate can then be analyzed by DNA-sequencing to achieve species identification. However, use of this technique for routine analysis depends on the development of simple and fast procedures for nucleic acid extraction (Fohlman et al., 2004). "Rapid-cycle real-time PCR" is now available in several commercially available instruments for clinical microbiology (Collins et al., 2004).

Fourier Transform Infrared Spectroscopy Applications in Microbiology

After the initial attempts of using traditional infrared spectroscopy for microbiological analysis, more extensive and successful work has been carried out using FT-IR spectroscopy. Haynes et al. (1958) carried out work using infrared spectroscopy to analyze microorganisms. Their work distinguished between two groups of bacteria based

on whether the spectra displayed a particular peak (5.7 μ) strongly or not at all. The majority of spectra that displayed the peak belonged to the genus *Bacillus*. All strains of *Bacillus cereus* produce a spectrum containing a peak but some strains of *Bacillus megaterium* and *Bacillus licheniformis* did not produce spectra containing the peak. Further investigation revealed that the peak was not produced by any of the aerobic sporeformers, if grown in the absence of glucose, galactose, glycerol, fructose or other satisfactory precursor. Two subtypes of *Bacillus megaterium* were also observed based on the effect of culture age upon the spectrum. In one subtype, the peak reached maximum absorption with a 12 hour culture but with the other subtype, the peak reached maximum absorption after seven days (Haynes et al., 1958).

The work was limited to basic observations until the development of FT-IR. Naumann et al. (1982) used FT-IR as a tool for probing bacterial peptidoglycan. The peptidoglycan (murein) of bacterial cell walls surrounds nearly all bacteria as a protecting network. This network is probably mono-layered in gram-negative bacteria but in gram-positive bacteria it is probably multi-layered and thus constitutes up to 60% of the cell wall weight. Because of its unique structure and function in protecting bacteria against osmotic lysis, pepetidoglycan has been proved to be an ideal target for chemotherapeutic attack.

Its primary structure consists basically of disaccharide-pentapeptide subunits with unusual features such as the occurrence of alternating D and L-amino acids and a γ-bonded D-glutamic acid residue. In contrast, there is no accepted model for the three-dimensional model of peptidoglycan except for evidence for and against it having a structure similar to that of chitin or cellulose. Naumann et al. (1982) demonstrated that

infrared spectroscopy could contribute to a more detailed insight into the architecture of peptidoglycan.

The results showed that peptidoglycan samples of all different chemotypes reveal common principles of their three-dimensional architecture in spite of the multiplicity of primary structures. This was demonstrated by the conformation-sensitive amide bands being constantly centered on approximately 3300 cm^{-1}, 3080 cm^{-1}, 1657 cm^{-1} and 1534 cm^{-1} (Naumann et al., 1982). The deviating features in the infrared spectra of peptidoglycans isolated from different bacterial types seem to be useful for identification purposes by means of a fingerprint method. Some well-defined infrared bands could be employed for semi-quantitative cell wall analysis, such as determination of cross-linking indices. The following characteristic variations could be detected: (a) the presence or absence of bands near 1730 cm^{-1}, 1600 cm^{-1} and 1400 cm^{-1}, (b) difference in relative absorbance values of amide bands at 1800 – 1500 cm^{-1}, and (c) drastic variations of absorbance near 1730 cm^{-1}. The 1730 cm^{-1} band in *Bacillus subtilis*, *Micrococcus luteus* and *Corynebacterium* spectra was ascribed to the carboxylic acid groups. The 1600 cm^{-1} and 1400 cm^{-1} bands in *Escherichia coli* spectra were ascribed to carboxylate ions. This was demonstrated by titrating with hydroxide and acetic acid, respectively, which removed the bands from the spectrum. This information was used to estimate the amount of crosslinking. The distinct and characteristic differences between the infrared spectra of chitin and peptidoglycan allowed the rejection of all hypotheses based on a nearly crystalline chitin-like three-dimensional structure of murein (Naumann et al., 1982).

Nichols et al. (1985) used FT-IR to analyze bacteria, bacteria-polymer mixtures and biofilms. Analytical biochemical methods involving the identification of specific

"signature" or "biological marker" lipids have been used to determine microbial biomass and community structure. Such biochemical techniques provide qualitative and quantitative insights into environmentally relevant problems but are destructive and time consuming. FT-IR analysis can complement or supplement existing analytical procedures for the characterization of microbial samples, including biofilms. However, use of FT-IR as the sole identification tool is limited.

Two groups of bacteria were analyzed using diffuse reflectance FT-IR (DRIFT) spectroscopy. Group I contained *Pseudomonas fluorescens, Desulfovibrio gigas, Staphylococcus aureus, Clostridium perfringens, Escherichia coli, Methylobacterium organophilium* (grown on methane) and *Methylosinus trichosporium* together with other sulphate-reducing bacteria and the spectra were found to be similar. Strong amide I (between 1690 and 1650 cm^{-1}) and amide II (1550 cm^{-1}) bands were present in all these organisms. Other prominent bands were present at 3300 cm^{-1} (OH stretch), 2950 cm^{-1} (CH stretch), 1450 cm^{-1}, 1250 cm^{-1} and 1090 cm^{-1} (Nichols et al., 1985)

Group II contained *Bacillus subtilis*, *Methylobacterium organophilium*, and *Nitrobacter winogradskyi* which when analyzed showed spectral features distinguishing them from the organisms listed above. In addition to the spectral bands noted above, the FT-IR spectrum of *Bacillus subtilis* contained an enlarged OH stretch and a large carbonyl peak centered at approximately 1740 cm^{-1}. When grown on methanol, the FT-IR spectrum of *Methylobacterium organophilium* contained a carbonyl stretch at 1740 cm^{-1} in addition to the strong amide I and II bands present when grown on methane. The carbonyl stretch was the largest peak detected in the spectrum of *Nitrobacter*

winogradskyi and additional spectral features were observed which were not present in the spectra of other microbes studied.

A difference spectrum obtained by the subtraction of a representative spectrum of a group I bacterium from a group II bacterium spectrum showed potential for FT-IR to enhance the differences between two organisms. The data suggested that the potential also exists for these differences to be quantified and that such techniques may be applicable to the recognition of differences in microbial community structure (Nichols et al., 1985).

Nichols et al. (1985) also studied corrosion processes to determine the role of microorganisms using DRIFT spectroscopy. IR spectra of the bacterial cellular and extra-cellular material could be easily obtained. This indicates that DRIFT spectroscopy of the metal disc surfaces may represent one technique to study changes occurring at the metal-water interface and to detect the presence of bacteria and biofilms.

Hedrick et al. (1991) used FT-IR for the rapid differentiation of archaebacteria from eubacteria in lipid preparations. Increasing interest in biotechnological applications of archaebacteria and their enzymes has led to the isolation of thousands of new strains from, for example, deep-sea hydrothermal vents. The number of isolates and archaebacterial species that do not fit into known groups, such as alkaphiles and sulfate reducers, suggested a need for rapid routine methods to accurately classify isolates as archaebacteria or eubacteria. The isoprenoid ether lipids of archaebacteria and the ester-linked long-chain fatty acids of eubacteria reliably distinguish these groups with the exception of a eubacterium which contains both ester and ether linkages.

The method was reliable using as little as 1 mg dry wt cell material. The lipid extraction and degradation procedures were simplified enabling twenty samples to be prepared in one workday. This compares with ten samples in four days for routine gas chromatographic analysis. IR spectra were obtained by DRIFT spectroscopy. Strong acid methanolysis of total lipid extracts was used to reduce the interfering peaks in the IR spectra. The ratio of the peak height of ester carbonyl stretch at 1743 cm^{-1} to the largest peak in the spectrum (the methyl stretch at 2924 cm^{-1}) was chosen to distinguish archaebacteria from eubacteria. The method reliably distinguished three species of archaebacteria and four species of eubacteria, which had ratios of approximately 10% and 60%, respectively. The method used was good for isolates but not appropriate for mixed cultures or environmental samples (Hedrick et al., 1991).

Fourier Transform Infrared Spectroscopy in Characterizing Bacteria

Initially, FT-IR spectroscopy was used to produce crude classification schemes involving whole groups of genera and component analysis. Further work was carried out to investigate the ability of FT-IR spectroscopy to distinguish bacteria down to the strain level and to use the spectra produced as a method of classification and identification.

Naumann (1984) used FT-IR to study intact, living bacteria cells and related cell-wall fragments. They used conventional absorbance/transmittance liquid cells equipped with 6-12 μm Sn spacers and BaF_2 or CaF_2 windows. Cell suspensions were injected into the IR liquid cell. Normalization of spectra to equal absorbance of amide I (spectral region where signal/noise ratio is high due to strong characteristic absorbance of water) resulted in a reproducibility of spectral features better than ± 5%. IR spectra of living

bacterial cells strongly depend on growth medium and time of growth. Consequently, extremely precise metabolic control and strict standardized handling of all samples was necessary to yield sufficient reproducibility and to avoid pitfalls when interpreting and comparing IR spectra of various bacterial cells.

Some bands were tentatively assigned: carbonyl stretching of ester groups (1745 cm^{-1}), amide I (1645 cm^{-1}), amide II (1547 cm^{-1}), asymmetric stretching of phosphate groups (1240 cm^{-1}) and complex vibrational modes of the various polysaccharides (1200-1000 cm^{-1}). Naumann (1984) concluded that spectral variations accompanying cell growth may readily be pursued, that it is possible to characterize metabolic influence of different media of growth and that chemical and antibiotic treatment of cells may be studied spectroscopically via FT-IR (Naumann, 1984).

In another study, Naumann et al. (1990) used FT-IR spectroscopy for the classification and identification of bacteria. The technique involved using 10-60 μg dry weight of late-exponential-phase cells suspended in 30 μl distilled water. An aliquot was transferred to zinc-selenite (ZnSe) optical plate and dried under a moderate vacuum to a transparent film suitable for absorbance/transmission FT-IR measurements. The spectral range was 4000 cm^{-1} to 500 cm^{-1} and was broken down into five spectral windows containing distinctive features:

Window 1 (W1)	Fatty acid region I, 3000 – 2800cm^{-1}
Window 2 (W2)	Amide region, 1800 – 1500cm^{-1}
	(Mixed region, 1500 – 1200cm^{-1})
Window 3 (W3)	Fatty acid region II, 1500 – 1400cm^{-1}

Window 4 (W4) Polysaccharide region, 1200 – 900cm^{-1}

Window 5 (W5) True Fingerprint region, 900 – 700cm^{-1}

The "fingerprint region" contained specific spectral patterns unique to each strain of bacteria although the peaks are as yet unassigned to cellular components or to functional groups.

The resulting spectra were parameterized and then analyzed by estimating and using the Pearson's product moment correlation coefficient as a measure of similarity between spectra. These values were then used in clustering procedures using Ward's algorithm. The resulting classification was in good agreement with conventional classification schemes. For example, all of the staphylococci tested formed a dense cluster, which was subdivided into two subclusters containing the coagulase-positive and the coagulase-negative strains, respectively. They concluded that FT-IR spectroscopy can classify bacteria at levels of taxonomic discrimination without any pre-selection of strains by other taxonomic criteria. In contrast to polyacrylamide gel electophoresis (PAGE) and DNA-DNA hybridization, FT-IR is useful at the serogroup, species and genus levels (Helm et al., 1991). Data acquisition is also faster than PAGE and other techniques with 100 spectra, transfer of data to the computer, data evaluation and cluster analysis being performed within 2 days. Preparation is simple and quick with only 10-60 μg of biomass required and no breakage and purification steps involved. In contrast to classical tests and chemotaxonomy, FT-IR can be applied to all groups of bacteria (Naumann et al., 1990).

Naumann et al. (1991) also used the techniques developed to establish FT-IR spectral libraries for the identification of bacteria. Carefully selected spectra representing

97 different strains out of 42 species and 17 genera were used to create a spectra library. Randomly, 72 out of 97 strains were selected to serve as "unknowns". Independently and repetitively measured spectra (n = 5) of the unknowns were recorded. Each spectrum of every unknown was chosen for the library challenge. The best fitting library spectra were used to identify unknown suspensions at the genus, species and strain level and the quotas of correct matches were calculated.

Since there is variability between spectra even when one strain is repetitively measured it was necessary to determine the reproducibility of bacterial FT-IR spectra for a selected strain under different culture conditions. These conditions were: different aliquots of identical bacterial suspensions; independent preparations of bacteria grown on agar plates from the same batch; spectra recorded over a long period of time (two years) for different preparations from bacteria grown on the same agar produced from different batches. The first condition yielded excellent reproducibility, however the third condition is more practical since it is common to use agar from the same manufacturer but different batches over a period of time.

Each of the five spectral windows was also tested singly to determine the window dependent quotas of correct matches. To enable comparison of different windows the spectral distances were suitably normalized. The entire spectrum yielded only poor quotas of correct matches: 31.9, 43.1 and 55.6% of the unknowns were identified as to strain, species and genus, respectively. This is probably due to the parts of the spectrum that provide no specific information. The quotas of correct matches for individual windows ranged from 29.2% (amide region) to 80.6% (polysaccharide region) for identity at the strain level. The most strain-specific information seemed to be contained in

the fatty acid region I, in the polysaccharide region and in the fingerprint region. Windows were then combined to get the optimal identification quotas. The data suggested the following combinations, where C1 – C4 are combinations of the spectral windows shown:

(1) $C1 \equiv \{W1 \cap W2 \cap W3 \cap W4 \cap W5\}$

(2) $C2 \equiv \{W1 \cap W3 \cap W4 \cap W5\}$

(3) $C3 \equiv \{W1 \cap W4 \cap W5\}$

(4) $C4 \equiv \{W1 \cap W4\}$

The combinations C3 and C4 obtained best results (Naumann et al., 1991).

Holt et al. (1994) used FT-IR to discriminate between different species of *Listeria*, an important food-borne pathogen. The method used was similar to that of Naumann et al. (1990) where the bacterial cells were suspended in water and then an aliquot was dried on a ZnSe crystal. For all the type species present in the set, spectra were recorded on three separate occasions using different batch cultures. At least six replicate spectra of each of the type species were recorded and in all, 59 spectra were selected for analysis. Six replicate spectra of each type species, recorded on three separate occasions using different batch cultures were considered to adequately represent the variability present in repeated measurements of the spectra of the type cultures, and hence they provide a test of the power of discrimination of the infrared method. The results were analyzed using Principal Component Analysis to reduce the data set and then Canonical Variate Analysis was used for classification (Holt et al., 1994).

Curk et al. (1994) used FT-IR to identify *Lactobacillus* species. The method again modeled Naumann et al. (1990). The reproducibility of FT-IR spectroscopy and the influence of culture temperature and time on the results were first estimated. The spectral distance, a measure of dissimilarity calculated from the Pearson's product moment correlation coefficient, varied with the strain and spectral window used. However, the minimal spectral distance between two spectra given by two different species was always greater than the spectral distance within the same strain.

The influence of the culture temperature on the spectra was estimated by comparing spectra for a strain grown at the optimal temperature (28 or 45°C) and at 37°C. The spectra for a strain grown at these two temperatures give more differences than two spectra of the same strain grown under the same conditions. Nevertheless, the distinction between species was often greater than the difference between a single strain grown at different temperatures. Time in culture had a similar effect and so, although the differences still allow discrimination between species, it is easier to keep all conditions the same to facilitate comparisons and increase reproducibility (Curk et al., 1994).

Kirschner et al. (2001) studied the use of FT-IR for the identification of enterococci at the species level. They compared the discriminatory power of vibrational spectroscopic techniques, including FT-IR, with phenotypic and genotypic methods. D. Naumann was a member of the group conducting this study and the method used to record the FT-IR spectra was the same as in his previous work (Naumann et al., 1990). Seven repetitive measurements over a period of six months from independent sample preparations of 18 strains of enterococci were performed, resulting in 126 spectra.

Statistical analysis of the spectra was carried out using OPUS software (version 3.0; Bruker). First and second derivatives of the spectra were calculated using a 9-point Savitzky-Golay filter to enhance the resolution of superimposed bands and to minimize problems from unavoidable baseline shifts. The first derivatives of the original spectra were used as input for cluster analysis and spectral distances were used for hierarchical clustering analysis.

The results indicated that FT-IR is highly reproducible and specific at the strain level. Thus, it allows accurate differentiation of closely related bacterial species such as enterococci. In comparison to conventional genotypic and phenotypic methods, FT-IR was found to be accurate for a wider range of *Enterococcus* species and more economical for routine analysis (Kirschner et al., 2001).

Oust et al. (2004) investigated the robustness of FT-IR spectroscopy to distinguish strains of bacteria under varying growth conditions. Resulting spectra were robust under small changes in growth temperature (28-32°C), growth medium (MRS and APT), growth time (42, 48 and 54 hours) and atmospheric conditions (aerobic and anaerobic) with differentiation between species and strains being unaffected. Larger variations in the growth medium only affected the differentiation of strains. However, it was also noted that any changes in conditions resulted in larger variations in the data collected. Therefore, to ensure reproducible results it is necessary to standardize the cultivation conditions as much as possible (Oust et al., 2004).

Statistical Analysis of Spectra

The statistical analyses of the spectra recorded from the bacterial samples used in the literature all involved the same basic principles. The spectra were prepared for statistical analysis using parameterization. This included filtering of the spectra (using first or second derivatives or deconvolution) to enhance the resolution and selecting a combination of certain spectral windows rather than using the whole spectra for analysis. The spectra were also normalized and scaled, usually using the strongest amide peak at 1656cm^{-1} (Naumann et al., 1990)

The statistical analysis of the spectra was performed in a variety of ways. However, each method utilized multivariate techniques such as Principal Component Analysis (PCA), Canonical Variate Analysis (CVA), Cluster Analysis (CA), and Factor Analysis (FA) (Naumann et al., 1990, Curk et al., 1994, Holt et al., 1995).

Principal Component Analysis

Principal component analysis was used as an initial treatment for the data to reduce the data set to a more manageable size suitable for further analysis (Naumann et al., 1990). The data set reduction is achieved by treating each spectrum as a point in a hyperdimensional space with as many dimensions as variables (absorbances) recorded from the spectrum. The number of variables is usually one per wavenumber over the entire spectrum or from selected regions of the spectrum. The total set of n spectra form a "cloud" of n points in this hyperdimensional space that can be described by a set of p principal components, where p is less than n. The principal components consist of eigenvectors and eigenvalues with each successive eigenvector representing a decreasing

amount of interspectral variance. Thus, approximately, only the initial two to eight eigenvectors are required to describe the bulk of interspectral variance with further eigenvectors corresponding merely to noise in the spectra. This significant reduction in the data set allows further analysis by other techniques (Johnson and Wichern, 1992).

Canonical Correlation Analysis

Canonical Correlation Analysis can classify the spectra by using linear combinations of the variables (Holt et al., 1994). The classification process considers the correlation between the linear combination of variables in one set and those of another set. It determines the pair of linear combinations with the largest correlation and then the next pair from the remaining set that is uncorrelated with the first pair, and so on. The pairs of linear combinations are called canonical variables and their correlations are called canonical correlations. The aim is to maximize the ratio of between-groups to within-group variance and hence provides discrimination between the groups (Johnson and Wichern, 1992).

Factor Analysis

Factor Analysis groups the variables according to their correlation so that all variables within a particular group are highly correlated among themselves but have relatively small correlations with variables in a different group. For example, when considering the intelligence of different individuals by comparing class scores, one may expect scores from chemistry, physics and math to be highly correlated but uncorrelated with English and history. This correlation can represent an underlying, but unmeasurable,

factor. These factors can be used to indicate common bacterial origin (Johnson et al., 1992).

Cluster Analysis

Cluster analysis uses the Euclidean (straight-line) distance between data points to determine similarity between spectra and assign them to clusters. No assumptions are made concerning the number of groups or the group structure. Hierarchical clustering methods are either agglomerative (successive mergers) or divisive (successive divisions). The results of both methods may be displayed in the form of a dendrogram (two-dimensional diagram). The following are steps in the agglomerative hierarchical algorithm for grouping N objects:

1) Start with N clusters, each containing a single entity and an NxN matrix of distances.
2) Search the distance matrix for the nearest (most similar) pair of clusters.
3) Merge these clusters to form one cluster and update the distance matrix replacing the original distances between the remaining clusters and the separate nearest clusters, with new distances between the remaining clusters and the merged cluster.
4) Repeat 2 and 3 a total of N-1 times. All objects will be in a single cluster at termination of the algorithm. Record the identity of the clusters that are merged and the levels (distances) at which the mergers take place.

(Johnson and Wichern, 1992)

This algorithm is carried out by a statistics computer package. The algorithm can vary according to how the distance between clusters is determined. An example is Ward's algorithm, which clusters groups by determining the least increase in variance or error sum of squares in the new cluster. This is calculated as the sum of the squared distances from each member of the cluster to the center of the cluster (Johnson and Wichern, 1992).

Application of Statistics for Classification

Holt et al. (1994) used canonical variate analysis (CVA) to distinguish species in the genus *Listeria.* The spectra were stored as normalized, scaled sets of data in the range of 2000 to 750 cm^{-1} at 1cm^{-1} sampling intervals. The variables were standardized by subtracting their means to give a 59-by-1250 matrix, X. PCA was used for data set reduction, producing a 59-by-r orthogonal score matrix, A, and a 1250-by-r orthonormal matrix of loadings, B, where A = XB. It can be shown that the best rank r approximation to X is AB' and so a larger r will retain more of the variability of the original data at the cost of a higher dimensionality. In this case, r was chosen to be 20. The score matrix was then used as the data set matrix for CVA. This analysis produced an r-by-m matrix of loadings, C, leading to a 59-by-m matrix of scores, D, where D = AC which was then used to distinguish species by observing the clusters in increasing dimensions. The species were discriminated in four dimensions (m = 4). The loadings were defined in terms of the original variables and used to indicate regions of the spectra that are important for discrimination. These correspond to large, positive or negative, loadings and occurred around 947-985cm^{-1} and, to a lesser extent, 1236 and 1750cm^{-1} (Holt et al., 1994)

The loadings were also used to find the canonical variance (CV) scores of new observations and place them in the group they are closest to in CV space. Assessment of the performance of the classification was obtained using the "leave-one-out" estimate. This involves taking each observation out of the data set in turn, calculating the CV's and classifying the omitted observation on the basis of these new CV's. Every observation was correctly classified by using this method (Holt et al., 1994).

Naumann et al. (1991) used spectral windows as described previously. These were analyzed using factor analysis to reduce the data set and then cluster analysis to determine the bacterial classification at the genus, species and strain level. The appropriate grouping of bacterial spectra into clusters was achieved by systematically varying some parameters prior to cluster analysis. These parameters concerned the filtering of the spectra (first or second derivative), the selection and combination of certain spectral windows and their weighting. The weighting was intended to account for the specific contributions of some cellular compounds, e.g. fatty acids of the membrane or polysaccharides of the cell wall.

The similarity between the spectra was determined using Pearson's product moment correlation coefficient. This is a measure of the linear association between two variables and is also known as the sample correlation coefficient. These coefficients were converted to *d*-values or spectral distances (a measure of dissimilarity) using the equation:

$$d_{y1y2} = (1.0 - r_{y1y2}) \cdot 1000.0$$

These could theoretically adopt values between 0 and 2000 (Naumann, 1985). A value of between 0 and 10 indicates indistinguishable spectra (Naumann et al., 1990). Since the spectra were subdivided into several spectral windows, the similarity between two spectra was calculated as a linear weighted combination of the single similarities giving the mean or overall spectral similarity. The single similarities were transformed to a uniform scaling level, multiplied by the weights, co-added and divided by the sum of the weights. The clustering was performed using Ward's algorithm (Naumann et al., 1991).

Curk et al. (1994) adopted the method and statistical analysis established by Naumann et al. (1991) to identify *Lactobacillus* species including the use of spectral windows and clustering using Ward's algorithm. Curk confirmed his identification results using DNA-DNA hybridization techniques (Curk et al., 1994).

More recently, Oust et al. (2004) used FT-IR spectroscopy to analyse 56 strains from four closely related species of Lactobacillus. Initially Hierarchical Cluster Analysis was used to study the clusters in the data but the dendrogram failed to successfully differentiate the four species. When the data was analysed with Partial Least Squares Regression, the strains were differentiated into four clusters according to species (Oust et al., 2004).

Alternative Methods for Analysis of Spectra

Hedrick et al. (1991) distinguished between the spectra of eubacteria and archaebacteria using the ratio of the peak height of ester carbonyl stretch at 1743 cm^{-1} to the largest peak in the spectrum (the methyl stretch at 2924 cm^{-1}). The archaebacteria contain isoprenoid ether lipids and the eubacteria contain ester-linked long-chain fatty acids. This method successfully distinguished between the two groups with the exception of a eubacterium that contains both ester and ether linkages. The typical value of the height of the ester carbonyl stretch peak as a percentage of the major methyl peak was found to be approximately 10% for archaebacteria and 60% for eubacteria (Hedrick et al., 1991).

Nichols et al. (1985) also used peak ratios to analyze bacteria. The analysis included dividing the bacteria studied into two groups. The first group was characterized

by a dominant amide I band and the second group of organisms displayed an additional carbonyl stretch at ~1740 cm^{-1}. The key band ratios correlate with compositions of the material and provide useful information for the application of FT-IR spectroscopy to environmental biofilm samples and for distinguishing bacteria grown under differing nutrient conditions (Nichols et al., 1985).

Given the development in technology and computerized statistical analysis techniques, FT-IR spectroscopy promises to be a very powerful technique for the classification and identification of bacteria. This could be of great significance in the food industry where there is a high demand for microbiological analysis in production processes, for safety reasons and for controlling spoilage.

HYPOTHESIS AND OBJECTIVES

Hypothesis

FT-IR spectroscopy is a reproducible method to identify bacteria.

Objectives

1. Design a sampling procedure to provide reproducible spectra for each organism.
2. Define the unique components of the spectrum to use as an identification tool.

METHODS

Obtaining Bacteria

The bacteria were obtained as cultures from Gist-brocades, Millville, Utah and Dr. Bart Weimer. The Gist-brocades bacteria samples were used in initial trials and planning stages of the project. The samples provided by Dr. Bart Weimer were used in the final study. The list of bacteria used for the study is found in Table 1.

Sampling Techniques

The freezer cultures were thawed and used to inoculate the respective broth. The inoculated broth was incubated for 16-24h at the appropriate temperature (Table 1). The broth was centrifuged (5000 rpm, 10 min) to collect the cells which were then washed three times with saline (0.85% NaCl) and then re-suspended in 350 µl of saline. A 25 µl aliquot was placed on a crystal and dried to a transparent film in a desiccator containing calcium sulfate, under a moderate vacuum (approximately 25 Torr). The crystals used were either barium fluoride or calcium fluoride and these were used interchangeably as they are themselves transparent to the infrared radiation. The resulting film was used to obtain a spectrum. A background spectrum was obtained using a blank saline film on the crystal.

All spectra were recorded on an FTS-7 FT-IR spectrometer (BIORAD) using 32 scans. The spectral resolution was 8 cm^{-1}. The data point resolution was one point per three wavenumbers, from 4000cm^{-1} to 500cm^{-1}.

Table 1 Bacteria Used in Study

Genus	Species	Subspecies	Strain	Growth Temperature °C	Broth	Conditions	Number of Batches
Lactococcus	*lactis*	*cremoris*	SK11	30	Elliker	Facultative Anaerobic	4
		lactis	S3				4
Leuconostoc				20-30	Elliker	Anaerobic	1
Pediococcus	*acidiladici*		33314 ATCC	25-40 (37)	MRS	Microaerophilic	3
	pentosaceus		25744				3
Lactobacillus	*helveticus*		212	30-40	Elliker	Facultative Anaerobic	2
	casei		201,202				2 (per strain)
Bacillus	*subtilis*			30	Nutrient	Aerobic	1
Pseudomonas	*fluorescens*		AFT29, B52	30	Nutrient	Aerobic	2 (per strain)
Escherichia	*coli*		PTRKL2	37	Nutrient	Facultative Anaerobic	1

Recording of Spectra

The first set of spectra was collected from one replicate of each bacterium used in the study. The first set of replicates of bacteria consisted of *Bacillus subtilis*, *E. coli, Leuconostoc, Pediococcus pentosaceus, Pediococcus acidiladici, Lactococcus lactis ssp. cremoris*, *Lactococcus lactis ssp. lactis, Lactobacillus casei* 201, *Lactobacillus casei* 202, *Lactobacillus helveticus*, and *Pseudomonas fluorescens* B 52 and *P. fluorescens* AFT 29. The bacteria were grown and analyzed as described above. Four spectra were recorded for each bacterium.

A second set of spectra was collected using another sample of each of the bacteria. These bacteria were chosen from those in the first set of replicates but were grown from a second sample under the same growth conditions. The bacteria chosen were those that successfully produced spectra from the first set of replicates. These were *P. pentosaceus, P.acidiladici, L. lactis ssp. lactis, L. casei* 201, *L. casei* 202, *L. helveticus, P.fluorescens* B 52 and *P. fluorescens* AFT 29. These bacteria were considered successful because they grew sufficiently to produce spectra with clearly distinguishable peaks of reasonable intensity rather than background noise. Four spectra were recorded for each bacterium.

Statistical Analysis of the Spectra

As well as visual comparison of the spectra, statistical analysis was carried out to determine if the spectra could be used successfully to distinguish between different organisms. The total number of variables for each spectrum was equivalent to the number of data points and the spectrometer used recorded one data point per three wavenumbers.

Hence, for each spectrum the raw data contained over 1100 data points. To prepare the raw data for analysis, the number of variables needed to be reduced while conserving as much of the spectral information as possible. To achieve the needed data reduction while preserving the most information the spectrum was divided into windows as defined by Naumann et al. (1990). The windows used for further analysis corresponded to the wavenumbers 3000 – 2800 cm-1, 1500 – 1400 cm-1 and 900 – 700 cm-1. The similarity between spectra was then determined by comparing the absorption values of these wavenumbers using the Pearson product coefficient to obtain correlation values between all the spectra obtained. These correlations were converted into spectral distances, a measure of dissimilarity, using the equation:

$$d_{y1y2} = (1.0 - r_{y1y2}) \cdot 1000.0$$

where d is the spectral distance between two spectra and r is the correlation between those spectra. The spectral distances were then compared to see which spectra are most similar to each other. This was done statistically using cluster analysis performed using Ward's method, Centroid method and the Complete Linkage method (Johnson et al. 1992). Each method produced the same dendrogram.

RESULTS AND DISCUSSION

Results

In total 112 spectra were recorded from seven genera, nine species and two strains for each of two species. Each sample had at least four sub-sample spectra. The spectra were stored as sets of data in the range of 4000cm^{-1} to 500cm^{-1} at 3cm^{-1} sampling intervals. All of the spectra produced can be found in Appendix B.

The spectra produced are the result of the infrared light being absorbed by the chemical bonds in the bacterium as each wavelength causes resonance. The strongest absorbance peaks occur due to the following components in the sample: water (3500 cm^{-1}, broad), amide bonds (1700 – 1500 cm^{-1}), fatty acids (1500 – 1400 cm^{-1}) and polysaccharides (1200 – 900 cm^{-1}). The region below 900cm^{-1} is referred to as the "fingerprint region" because it consists of several unassigned peaks, which are specific to each chemical compound that may be related to a specific bacterium (Appendix A). The tentative assignment of the chemical bond for other peaks frequently found in a bacterial spectrum can be found in Appendix A.

Reproducibility of Spectra

The effort of this work was to confirm that a reproducible spectrum could be collected. This was achieved to some degree by producing a spectrum with distinct peaks, which could then be reproduced from the same replicate of the each bacterium. However, the fingerprint region did vary between spectra. One example of this is the set of four spectra collected from the first replicate of *Pediococcus pentosaceus* (Fig. 1).

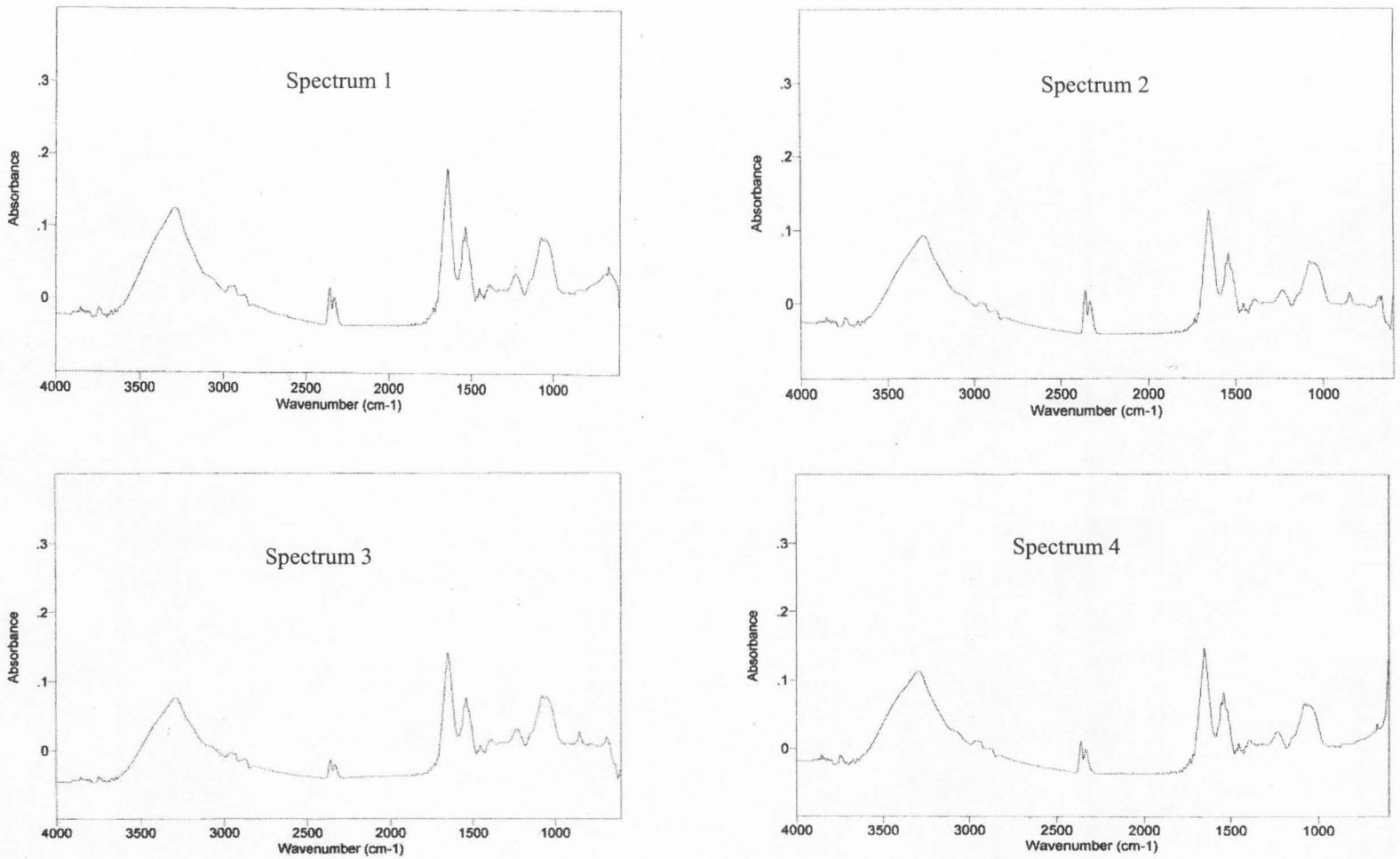


Figure 1. Spectra of sub-samples of *Pediococcus pentosaceus*, replicate 1.

Four separate aliquots were used and four very similar spectra with distinct peaks were recorded. Reproducible spectra with distinct peaks were collected for *P. pentosaceus, P.acidiladici, L. lactis ssp. lactis, L. casei* 201, *L. casei* 202, *L. helveticus, P.fluorescens* B 52 and *P. fluorescens* AFT 29. The spectra recorded for *Bacillus subtilis* and *Escherichia coli* had indistinguishable peaks because they did not grow as expected (see Fig. 2).

In order to confirm the reproducibility of the spectra it was necessary to compare the spectra of the first replicate with spectra of a second replicate of the same bacterium. For example, the spectra of the first replicate of *P. pentosaceus* can be compared with the set of spectra collected from the second replicate of *P. pentosaceus* (Fig. 3). Although the intensity of the peaks is approximately 0.1 absorbance units higher in the second set of spectra the actual shape of the peaks is the same in both sets. Some distinct features observed in both spectra are the multiple splitting of the amide peak at 1550 cm^{-1} and the splitting of the peak at 1100 cm^{-1}. This reproducibility was also seen in the spectra for the second replicate of the other bacteria. Once the spectra were collected with sufficient reproducibility, the data were examined for specific differences between the bacteria.

Specificity of Spectra

The second objective was to determine if a spectrum contained specific features that allowed determination of each bacterium down to the strain level and hence, could be used to distinguish between closely related bacteria. The first level of distinction we

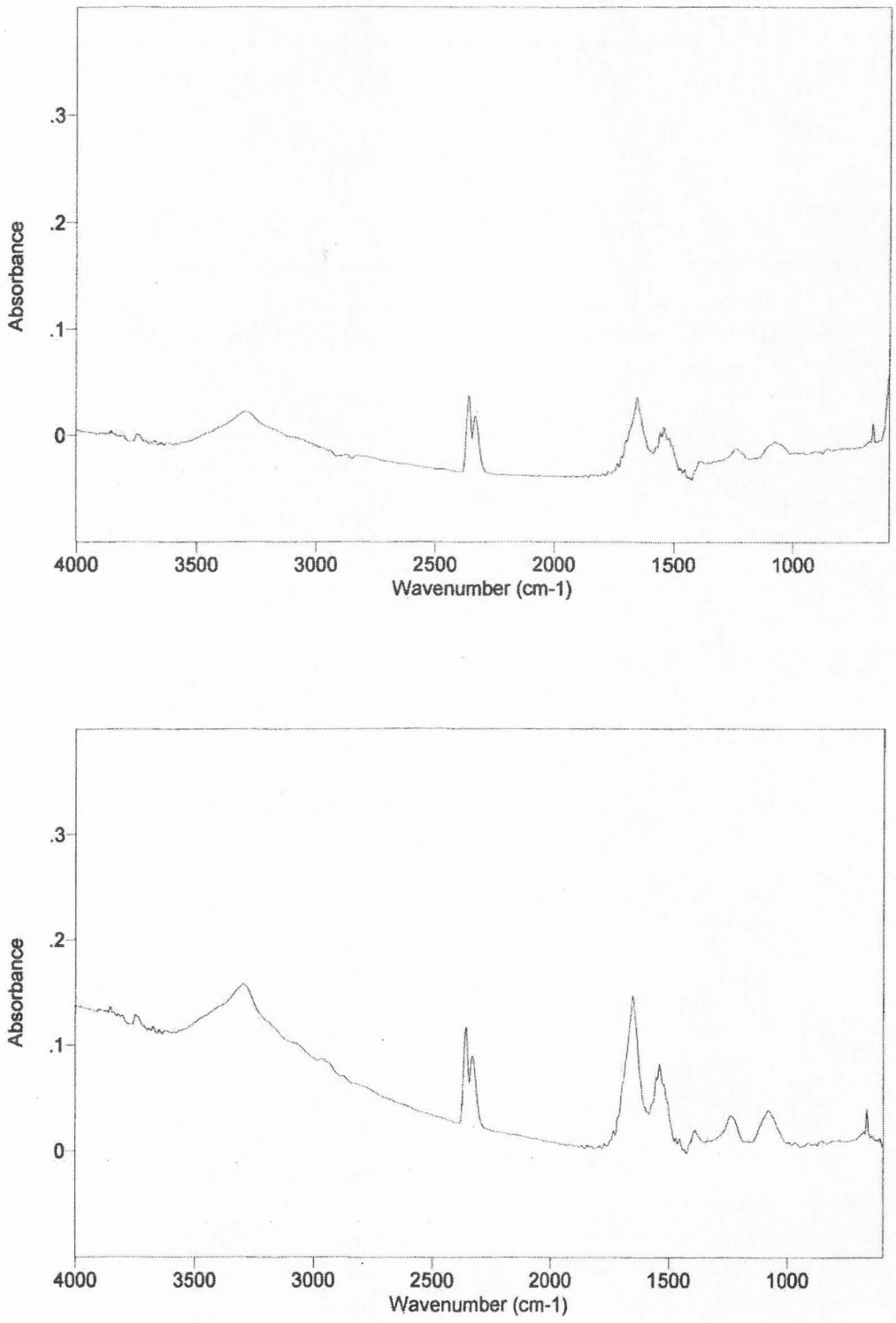


Figure 2. Spectra of sub-samples 1 (top) and 2 (bottom) of Esherichia coli, replicate 1

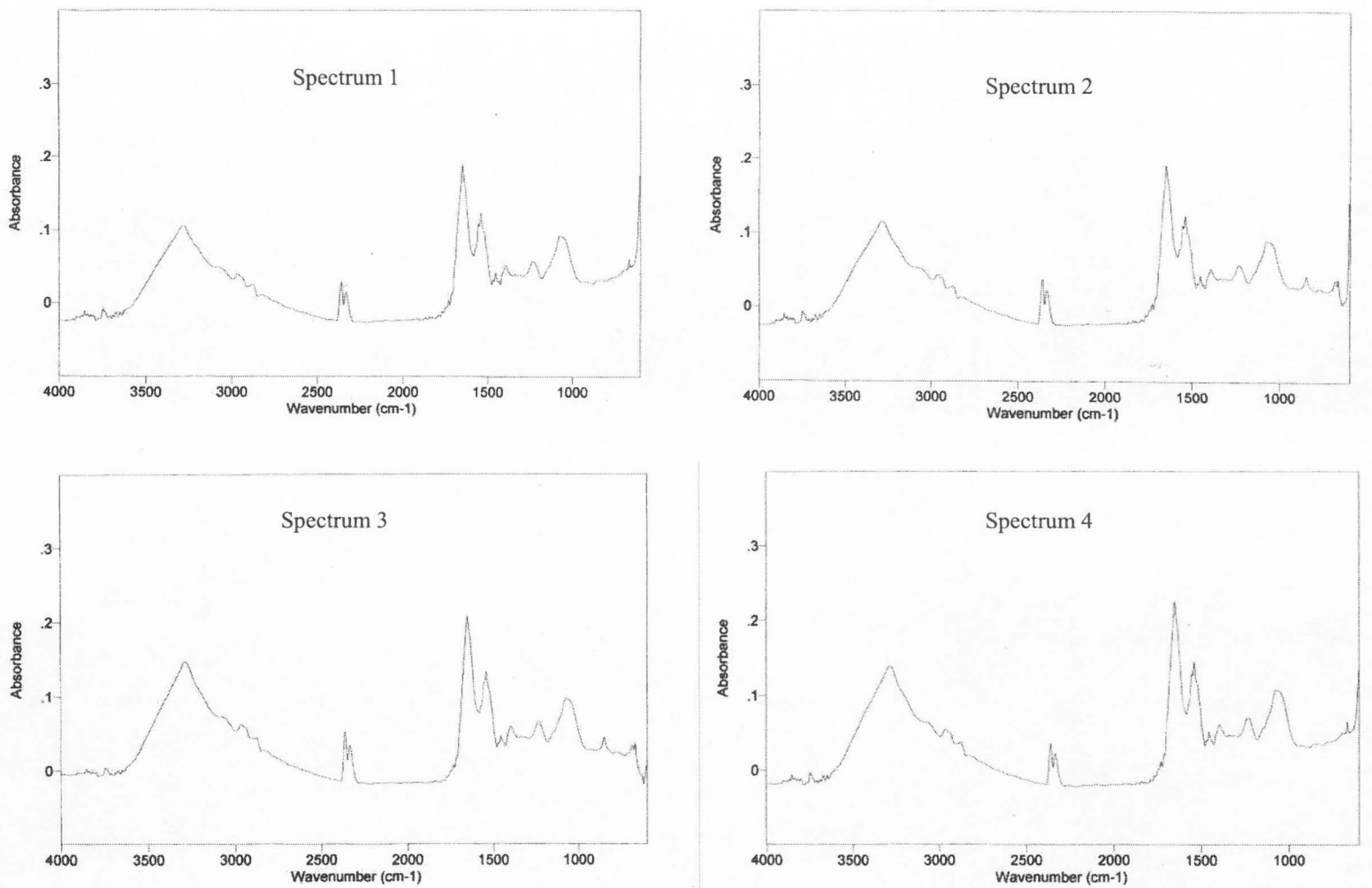


Figure 3. Spectra of sub-samples of *Pediococcus pentosaceus*, replicate 2.

examined was between the genera. This was observed using the spectra collected from *P. fluorescens* AFT 29 and *P. pentosaceus* (Fig. 4). Although the absolute intensity of absorbance was different, some specific peaks were observed that may distinguish between the two genera. The peak at 1100cm^{-1} was split into two peaks of approximately equal intensity in the *pediococcal* spectrum. In the *Pseudomonas* spectrum although the same peak was split, one of the resulting peaks was more intense than the others. The peak at 1550cm^{-1} showed multiple splitting in the *pediococcal* spectrum and appears to be a distinguishing feature. However, it was hard to determine if this multiple splitting appears because of the lower intensity of the absorbance or if it is due to specific bacterial components.

The second level of distinction tested was between each species. This was observed using the spectra collected from *Lactobacillus helveticus* and *Lactobacillus casei* (Fig. 5). One distinguishing feature between these two spectra may be the relative height of the peak at 1200 cm^{-1} (1) to the peak at 1100 cm^{-1} (2). In the *L. helveticus* spectrum peak 1 was consistently higher than peak 2. In the *L. casei* spectrum peak 1 was consistently lower than peak 2.

The third level of distinction was between strains. This can be observed in the spectra collected from *Pseudomonas fluorescens* AFT 29 and *Pseudomonas fluorescens* B52 (Fig. 6). There are some differences between these spectra, such as a small difference in splitting of the peak at 1100 cm^{-1}, but overall the spectra appear to be very similar. It is hard to distinguish between differences in the spectra that are specific to the strains and differences that are due to other factors such variations in the intensity. It is

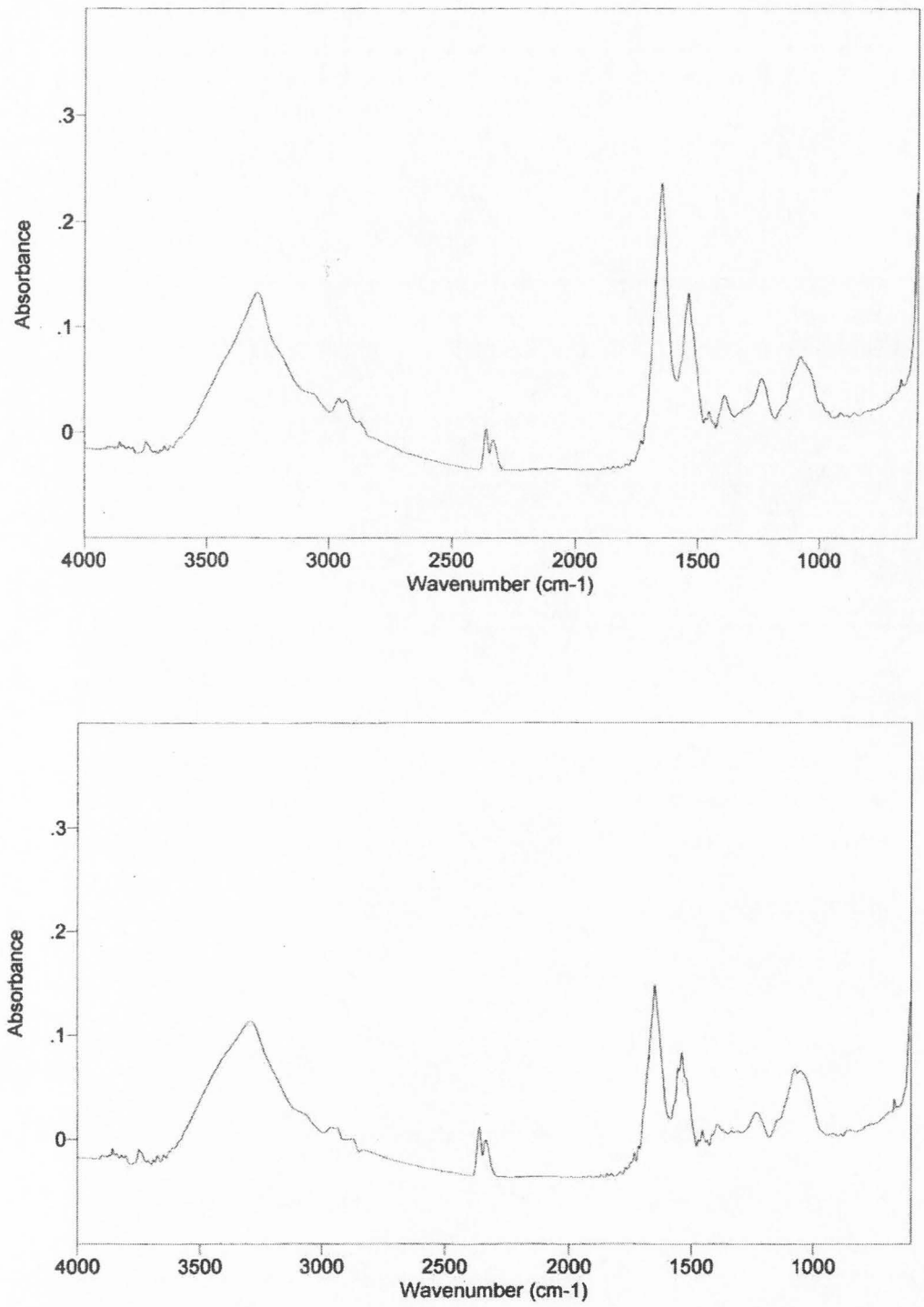


Figure 4. Spectra of *Pseudomonas fluorescens* AFT 29 (top) and *Pediococcus pentosaceus* (bottom).

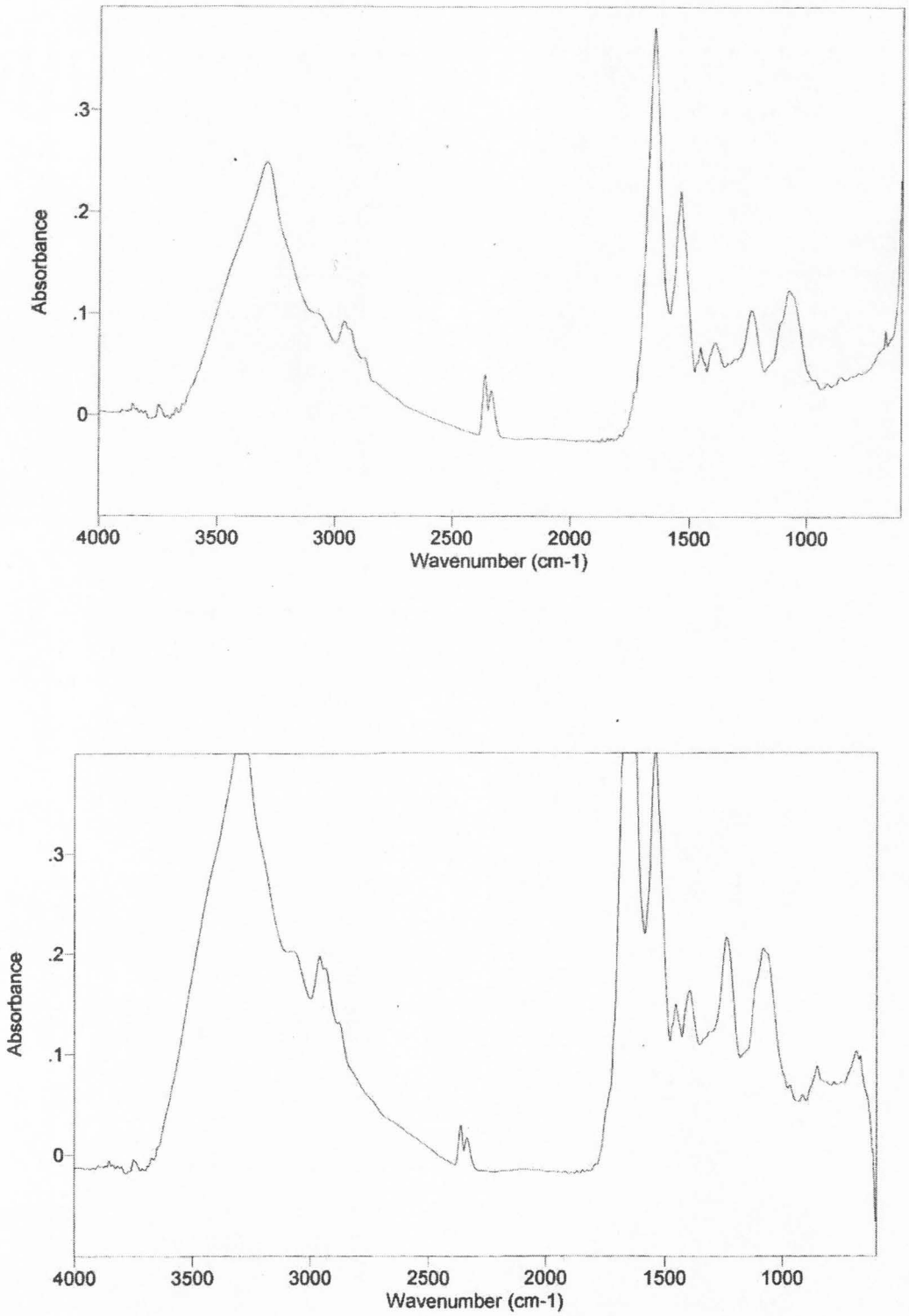


Figure 5. Spectra of *Lactobacillus casei* (top) and *Lactobacillus helveticus* (bottom).

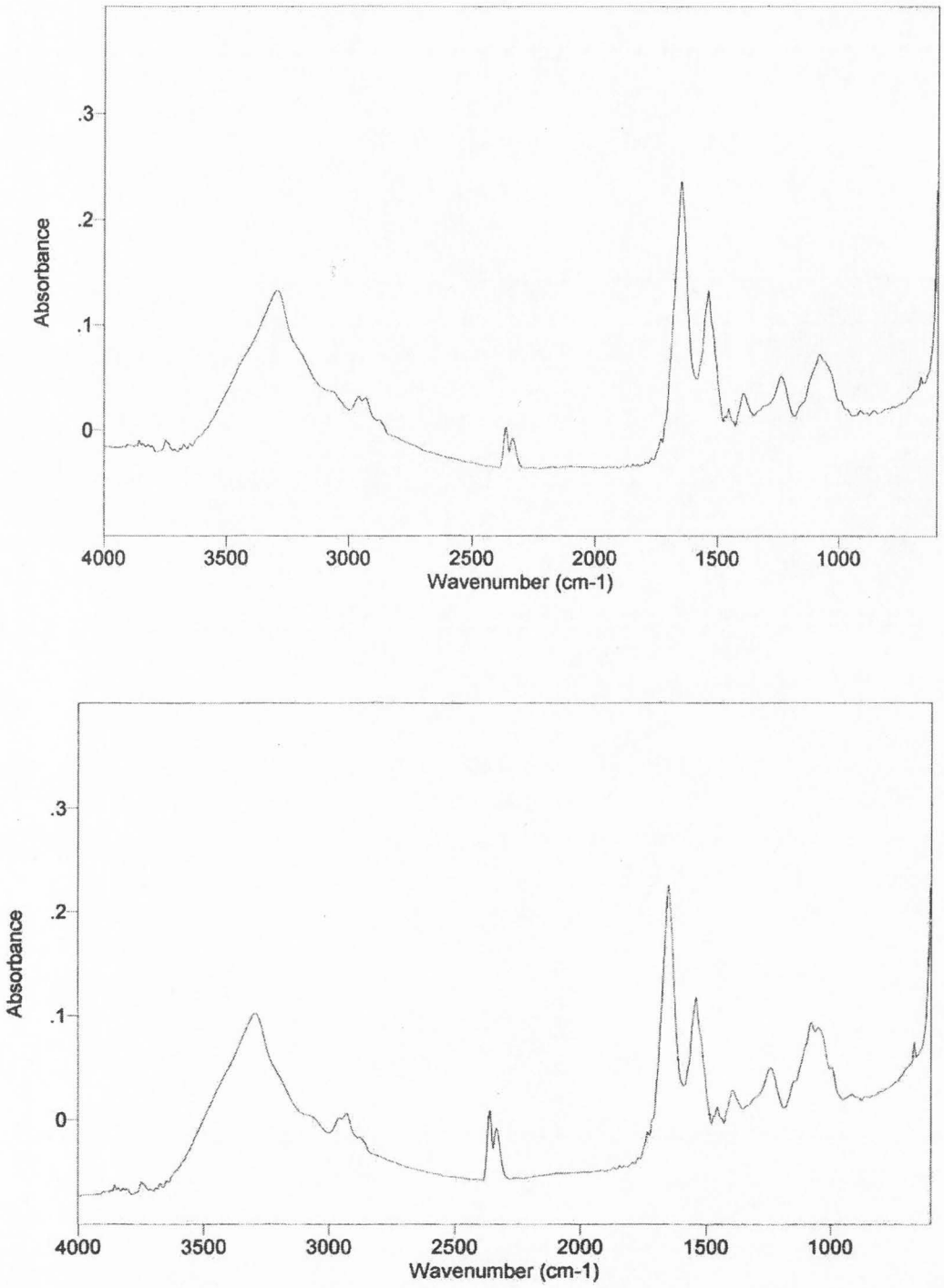


Figure 6. Spectra of *Pseudomonas fluorescens* AFT 29 (top) and *Pseudomonas fluorescens* B52 (bottom).

very difficult to determine the differences between the spectra from different organisms by visual inspection. Therefore, a more complete analysis was done using statistical procedures and spectral analysis of specific regions.

Statistical Analysis of the Spectra

The total number of variables for each spectrum was equivalent to the number of data points and the spectrometer used recorded one data point per three wavenumbers. Hence, for each spectrum the raw data contained over 1100 data points. To prepare the raw data for analysis, the number of variables needed to be reduced while conserving as much of the spectral information as possible. To achieve the needed data reduction while preserving the most information the spectrum was divided into windows as defined by Naumann et al. (1990). The windows used for further analysis corresponded to the wavenumbers 3000 – 2800 cm-1, 1500 – 1400 cm-1 and 900 – 700 cm-1. The similarity between spectra was then determined by comparing the absorption values of these wavenumbers using the Pearson product coefficient to obtain correlation values between all the spectra obtained. These correlations were converted into spectral distances, a measure of dissimilarity, using the equation:

$$d_{y1y2} = (1.0 - r_{y1y2}) \cdot 1000.0$$

where d is the spectral distance between two spectra and r is the correlation between those spectra. The spectral distance values can be seen in Appendix C, Tables 1a-e. The spectral distances were then compared to see which spectra are most similar to each

other. This was done statistically using cluster analysis performed using Ward's method, the Centroid method and the Complete Linkage method (Johnson and Wichern, 1992). Ward's method and the Centroid method produced the same dendrogram (Figs. 7 and 8) and the Complete Linkage method produced a dendrogram (Fig. 9) that was almost identical except for the switching of two clusters (number 7, *P. fluorescens* and number 14, *Leuconostoc*).

The dendrograms created five initial clusters represented by the five different colors, which were further subdivided into smaller clusters. The different colors might be expected to be homogeneous with regard to the type of bacterium they contain. The smaller the distance required to connect two organisms or clusters, the greater the similarity between them. The graph alongside below the dendrograms indicates the most appropriate number of clusters by measuring the level of similarity at each fusion of two clusters. Thus, a steep rise in similarity distance, indicating a significant increase in dissimilarity between two clusters, would indicate the most appropriate number of clusters. It would be expected that this would occur after ten or twelve clusters relating to the number of species or strains present in the data. However, the graph has significant increases in dissimilarity after 20, 24 or 27 clusters indicating that the spectra for different genera, species or strains of bacteria were not sufficiently similar to be clustered in 10 or 12 clusters.

Taking genera, species and strain each in turn, it can be seen that of the genus *Pediococcus*, three of the replicates (17, 20 and 21) are clustered quite closely together but that the others are in different clusters. For the genus *Lactobacillus*, all but one of the replicates are contained within two clusters although the similarity distance between the

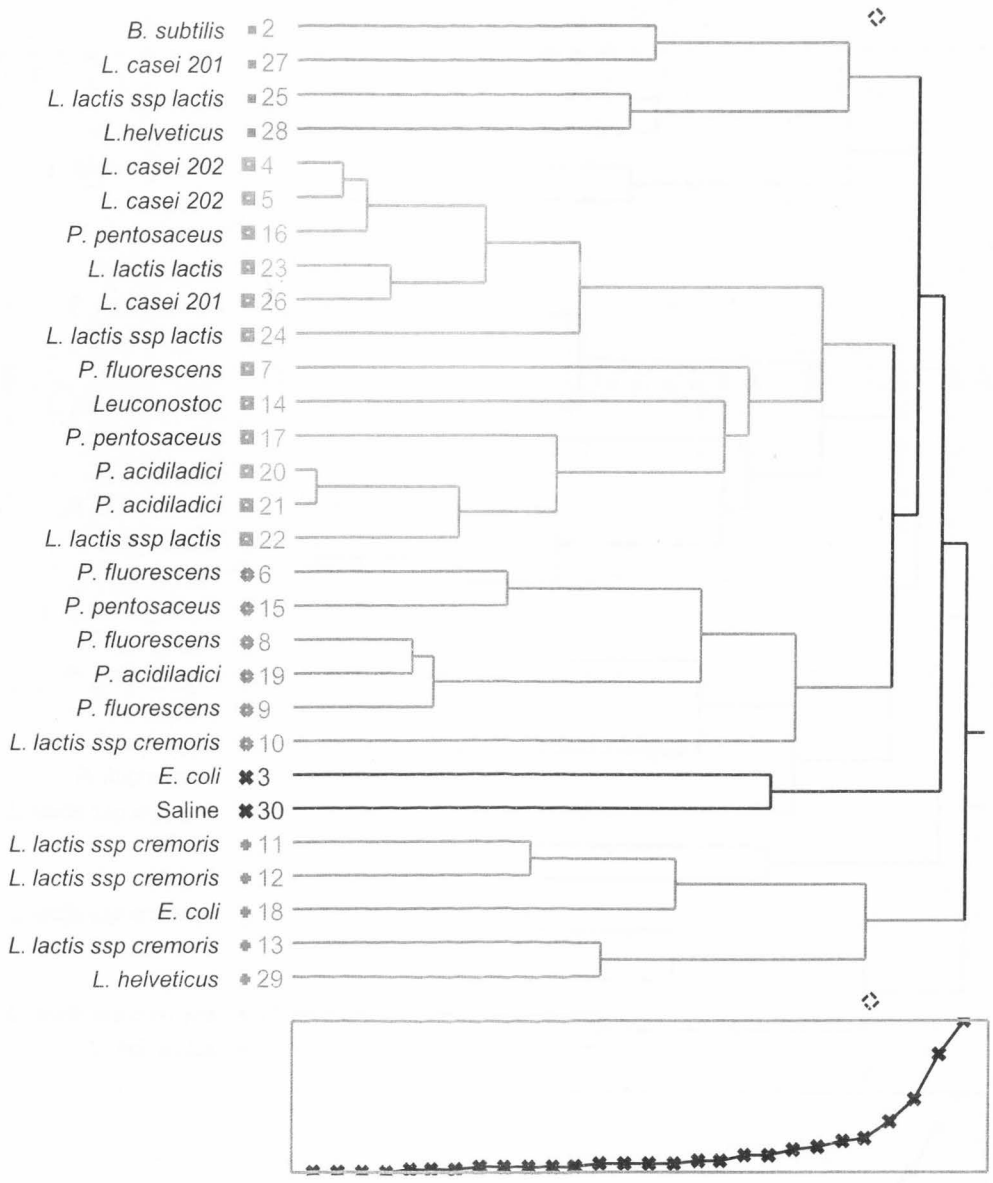


Figure 7. Dendrogram produced by Ward's Method.

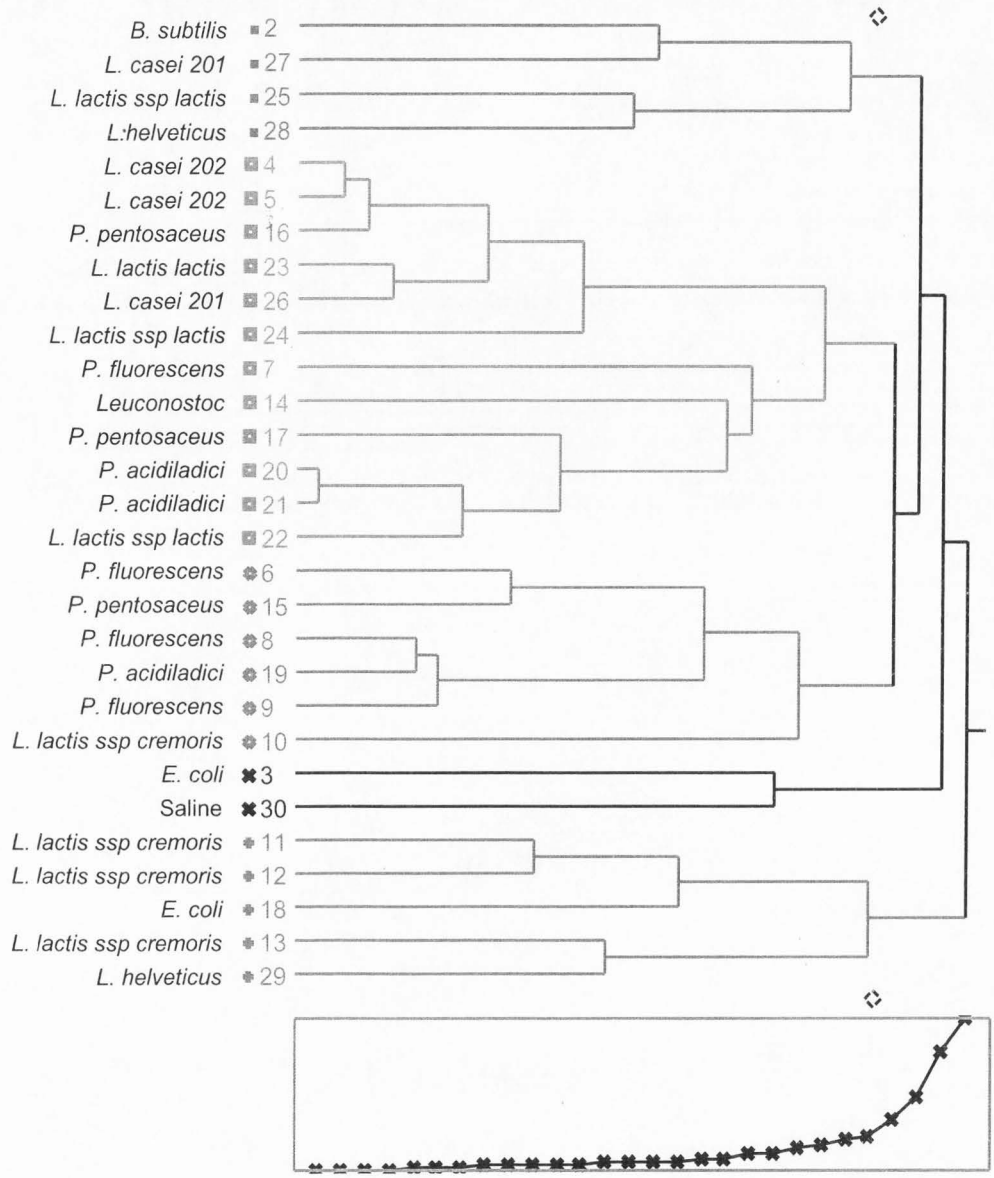


Figure 8. Dendrogram produced by the Centroid Method.

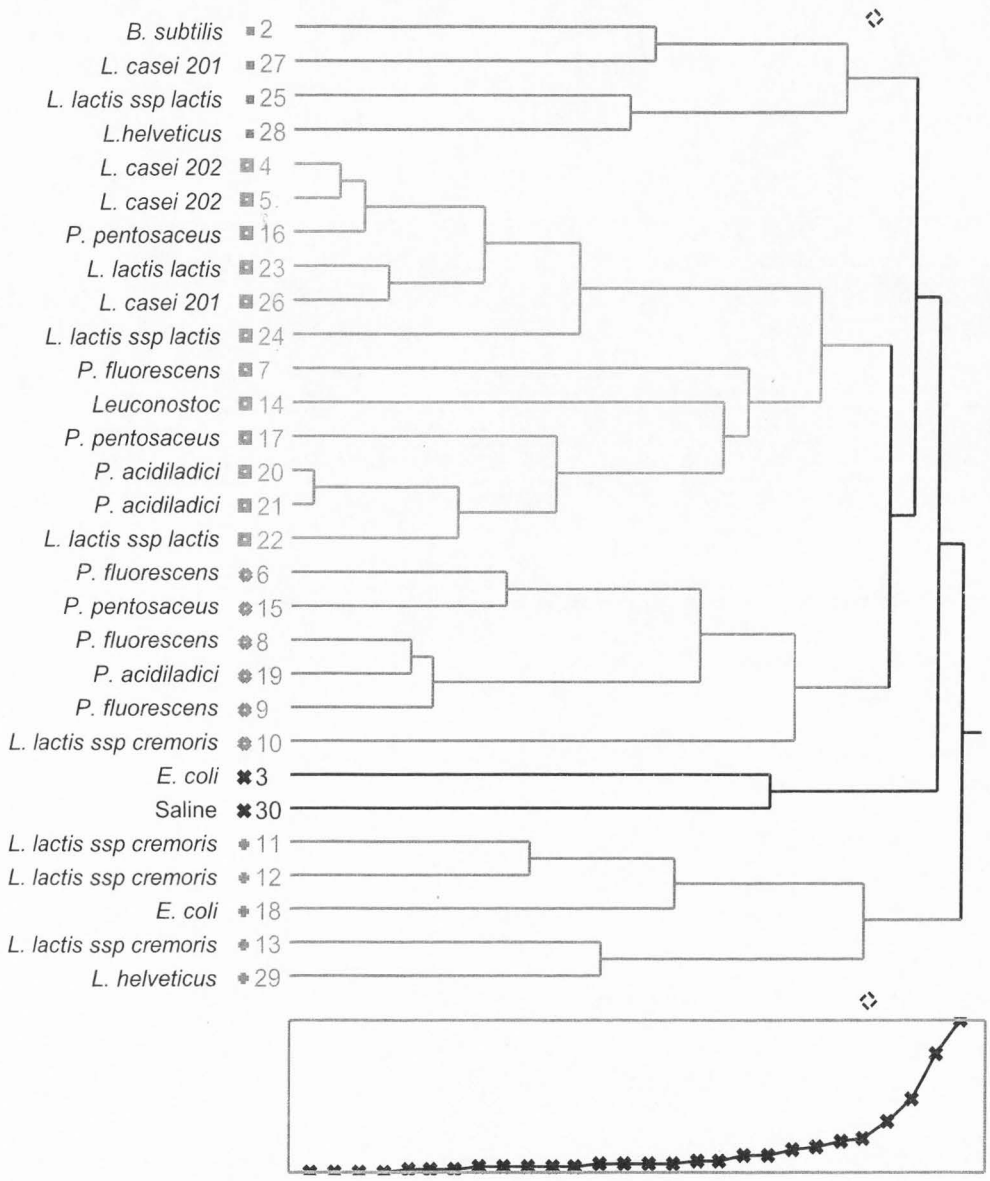


Figure9. Dendrogram produced by the Complete Linkage Method.

two clusters is large. The genus *Pseudomonas* is the most successful with three out of four of the replicates (6, 8 and 9) being found in the same cluster. The genus Lactococcus is the least successful with the replicates being spread out over all the clusters.

At the species level, *P. pentosaceus* was spread out between clusters but two out three of the *P. acidiladici* replicates (20 and 21) are in one cluster with a very small similarity distance. The third replicate (19) is in another cluster with a large similarity distance with the other two replicates. The two L. helveticus replicates (28 and 29) were not very similar at all being found at the top and bottom of the dendrogram with a very large similarity distance. The most successfully clustered species is *L. casei* where three out of four of the replicates (4,5 and 26) are found in one cluster with a small similarity distance. The fourth replicate (27) is again found in another cluster with a large similarity distance with the other three. Of the three *L. casei* replicates in one cluster, two (4 and 5) are of the same strain, *L. casei* 202, and have a very small similarity distance, making a cluster of their own. However, *L. casei* 201 have not clustered together and have a large similarity distance.

Overall, it would be expected that each genera of bacteria would have a cluster and within that cluster there would be clusters for each species, and then again for each strain, with the similarity distance decreasing for genera, species and strain respectively. It would also be expected that spectra of different replicates of the same bacterium would be in the same cluster, and ideally next to each other on the dendrograms, indicating the greatest similarity. However this has not been accomplished consistently even though it has been successful to a very limited extent in a few cases.

Interpretation of Results

Naumann et al. (1990) achieved complete differentiation of the clusters of different genra of bacteria. The clustering was also conclusive for most species and strains of bacteria. This was not consistently achieved in this study. One reason for this is the difficulty in recording spectra of a sufficient quality as in the case of *Esherichia coli.* The spectra showed the multiple splitting at 1550cm^{-1} that can also be seen in the spectra for *Bacillus subtilis*, *Leuconostoc* and *Lactococcus lactis ssp. cremoris*. This splitting pattern was probably due to the over all low intensity recorded on the spectra rather than being caused by a specific bacterial component. The peak pattern was also more variable between spectra. The peak patterns in the spectra recorded from these bacteria were probably caused by a weak concentration of the bacterial sample and had relatively little information about the specific bacteria that were used. This trouble is an indication of the potential problems in using the method consistently. A large number of cells and the operate skill play a crucial role in determining cellular identity.

These concerns could be overcome in part by using a minimum bacterial concentration to record the spectra. The bacterial sample could also be used for plate growth to ensure that there was no contamination in the sample. A minimum absorbance could be required for analysis and at the analysis stage, the spectra could also be normalized using the total absorbance or the strong amide peak to remove differences in absorbance due to bacterial concentration.

The statistical analysis could also be further developed to achieve more conclusive results. This could include using first or second derivatives of the spectra, using different windows of the spectra, and weighting the different windows that are used

for the analysis. For example, Naumann et al. (1990) used the first derivatives of the spectral ranges 3000 - 2800 cm^{-1}, 1200 - 1900 cm^{-1} and 900 - 700 cm^{-1}, with weightings of 1.0, 3.0 and 1.0, respectively, for the classification of the genera *Staphylococcus, Streptococcus* and *Clostridium*. But in the same paper, the second derivatives of the spectral ranges 1200 – 900 cm^{-1} and 1500 – 1200 cm^{-1}, with weightings of 2.0 and 1.0 respectively, were used for the cluster analysis of different species of *Clostridium*. The paper also indicates that the parameters were systematically varied until the resulting classification agreed, by and large, with the desired classification, according to the principles of the means-ends analysis (Newell and Simon, 1972). However, this empirical approach also limits the amount of predictive information that can be derived from a spectrum.

In conclusion, the results presented have demonstrated that FT-IR spectroscopy can be used to produce reproducible spectra from bacteria. The statistical analysis of these created successful clusters for several of the species used but was inconclusive overall in being able to distinguish between spectra at the genus, species and strain level. This may be due to the operate skill required to grow a sufficient number of bacterial cells needed to produce spectra of a consistently good quality. A second factor may be the statistical analysis required to obtain more conclusive results including normalizing the spectra, the use of first and second derivatives and weighting the various spectral windows according to their significance in distinguishing bacteria.

CONCLUSION

The results of this study have shown that it was difficult to produce highly reproducible spectra for bacteria using this sampling method. This may be due to the operate skill required to maintain consistent growth conditions in order to achieve a less variable final concentration of bacteria in saline solution. In future work, this could be improved upon by using plate growth to determine the density and possible contamination of the cells. This would confirm the bacterial concentration of the sample used to produce spectra and could also be used to determine a minimum concentration required for reproducible spectra. This would minimize the variability in absorbance that is observed in the spectra.

Furthermore, the regions of the spectra used for analysis were taken from the literature where they had proved successful (Naumann et al., 1990). However, this study failed to conclusively distinguish between the bacteria using these windows, even though there were some sub-samples that were successfully clustered. Future work could include further statistical manipulation of the data including normalizing the spectra on total absorbance or an amide peak, using first or second derivatives of the data and weighting the different windows until a more successful cluster analysis was achieved. Other windows of the spectra could also be used for analysis.

So overall, this study has not been able to demonstrate that FT-IR spectroscopy is a reproducible method to identify bacteria. Spectra were successfully recorded using the sampling technique but future work would need to include determination of the bacterial concentration, and further statistical manipulation of the data.

REFERENCES

Belton, P. S., Saffa, A. M., Wilson, R. H., 1987. The potential of Fourier Transform Infrared Spectroscopy for the analysis of confectionary products. Food Chem. 53-61.

Briandet, R., Kemsley, E. K., Wilson, R. H., 1996. Discrimination of Arabica and Robusta in Instant Coffee by Fourier Transform Infrared Spectroscopy and Chemometrics. J. Agric. Food Chem. 44, 170 – 174.

Collins, C. H., Lyne, P. M., Grange, J. M., 2004. Microbiological Methods, eighth edition. Arnold, London.

Curk, M. C., Peladan, F., Hubert, J. C., 1994. Fourier Transform infrared (FTIR) spectroscopy for identifying *Lactobacillus* species. FEMS Microbiology Letters 123, 241-248.

Doyle, M. P., Beuchat, L. R., Montville, T. J., 2001. Food Microbiology Fundamentals and Frontiers, second edition. ASM Press, Washington, DC.

Fohlman, J., Blomberg, J., Froman, G., Engstrand, L., Johansson, A., Friman, G., 2004. Microbial diagnosis with PCR will become clinically beneficial with a faster analysis. Lakartidningen 101, 1488-92.

Haynes, W. C., Melvin, E. H., Locke, J. M., Glass, C. A., Senti, F. R., 1958. Certain factors affecting the infrared spectra of selected microorganisms. Appl. Microbiol. 6, 298-304

Hedrick, D. B., Nivens, D. E., Stafford, C., White, D. C., 1991. Rapid differentiation of archaebacteria from eubacteria by diffuse reflectance Fourier-transform IR spectroscopic analysis of lipid preparations. J. Microbiol. Methods 13, 67-73.

Helm, D., Labischinski, H., Naumann, D., 1991. Elaboration of a procedure for identification of bacteria using Fourier-Transform IR spectral libraries: a stepwise correlation approach. J. Microbiol. Methods 14, 127-142.

Helm, D., Labischinski, H., Schallehn, G., Naumann, D., 1991. Classification and identification of bacteria by Fourier-transform infrared specroscopy. J. Gen. Microbiol. 137, 69-79.

Holt, C., Hirst, D., Sutherland, A., McDonald, F., 1995. Discrimination of species in the genus Listeria by Fourier Transform infrared spectroscopy and canonical variate analysis. Appl. Environ. Microbiol. Jan, 377-378.

Holt, J.G., Krieg, N.R., Sneath, P.H.A., Staley, J.T., Williams, S.T. (Eds.), 1994. Bergey's Manual of Determinative Bacteriology, ninth edition. Williams and Wilkins, Baltimore.

Johnson, R. A., Wichern, D. W., 1992. Applied Multivariate Statistical Analysis, third edition. Prentice Hall, Upper Saddle River, NJ.

Kirschner, C., Maquelin, K., Pina, P., Ngo Thi, N. A., Choo-Smith, L.-P., Sockalingum, G. D., Sandt, C., Ami, D., Orsini, F., Doglia, S. M., Allouch, P., Mainfait, M., Puppels, G. J. and Naumann, D., 2001. Classification and identification of Enterococci: a comparative phenotypic, genotypic, and vibrational spectroscopic study. J. Clinical Microbiol. May, 1763-1770

Naumann, D., Helm, D., Labischinki, H., Giesbrecht, P., 1990. The characterization of microorganisms by Fourier-Transform Infrared Spectroscopy (FT-IR). In: Nelson, W.H. (Ed.), Modern Techniques for Rapid Microbiological Analysis. VCH, New York, pp. 43-96.

Naumann, D., Helm, D., Labischinski, H., 1991. Microbiological characterizations by FT-IR spectroscopy. Nature. 351, 81-83.

Naumann, D., Barnickel, G., Bradaczek, H., Labischinski, H., Giesbrecht, P., 1982. Infrared spectroscopy, a tool for probing bacterial peptidoglycan. Eur. J. Biochem. 125, 505-515.

Naumann, D., 1984. Some ultrastructural information on intact, living bacterial cells and related cell-wall fragments as given by Fourier Transform infrared spectroscopy. Infrared Phys. 24(2/3), 233-238.

Nichols, P. D., Henson, J. M., Guckert, J. B., Nivens, D. E., White, D. C., 1985. Fourier Transform infrared spectroscopic methods for microbial ecology: analysis of bacteria, bacteria-polymer mixtures and biofilms. J. Microbiol. Meth. 4, 79-94.

Osborne, B. G., Douglas, S., Fearn, T., Willis, K. H., 1982. The development of universal calibrations for measurement of protein and moisture in UK home-grown wheat by Near-Infrared Reflectance Analysis. J. Sci. Food Agric. 33, 736 – 740.

Oust A., Moretro T., Kirschner C., Narvhus JA., Kohler A., 2004. Fourier Transform infrared spectroscopy for identification of closely related lactobacilli. J. Microbiol. Methods 59, 149-162

van de Voort, F. R., 1992. Fourier Transform infrared spectroscopy applied to food analysis. Food Res. International 25, 397-403.

van de Voort, F. R., Memon, K. P., Sedman, J., Ismail, A. A., 1996. Determination of Solid Fat Index by Fourier Transform Infrared Spectroscopy. JAOCS 73(4), 411-416.

APPENDICES

A: TENTATIVE ASSIGNMENT OF BANDS IN BACTERIAL SPECTRA

TENTATIVE ASSIGNMENTS OF SIGNIFICANT BANDS IN BACTERIAL SPECTRA

Table A1 Tentative Assignments of Significant Bands in Bacterial Spectra (Adapted from Naumann et al., The Characterization of Microorganisms by Fourier-Transform Infrared Spectroscopy (FT-IR). *Modern Techniques for Rapid Microbiological Analysis*; Nelson, W.H., Ed.; VCH: New York, 1991.)

Frequency cm^{-1}	Relative Intensity	Tentative Assignment
3500	m	OH str of OH hydroxyl
3200	m-s	NH str (amide I) of proteins.
2900 - 2800	m	Different CH str in fatty acids
1655	s	Amide I of α-helical structures
1637	s	Amide I of β-pleated sheet structures
1548	s	Amide II band
1515	m	"Tyrosine" band
1468	w-m	CH sc of CH_2 methylene
1250-1220	w-m	PO str of PO phosphodiesters
1200-900	m	COC, CO dominated by the ring vibrations of polysaccharides
900-600	w	"fingerprint region"

Abbreviations:

s	strong
w	weak
m	medium
v	very
asym	asymmetrical
sym	symmetrical
str	stretch
sc	scissoring

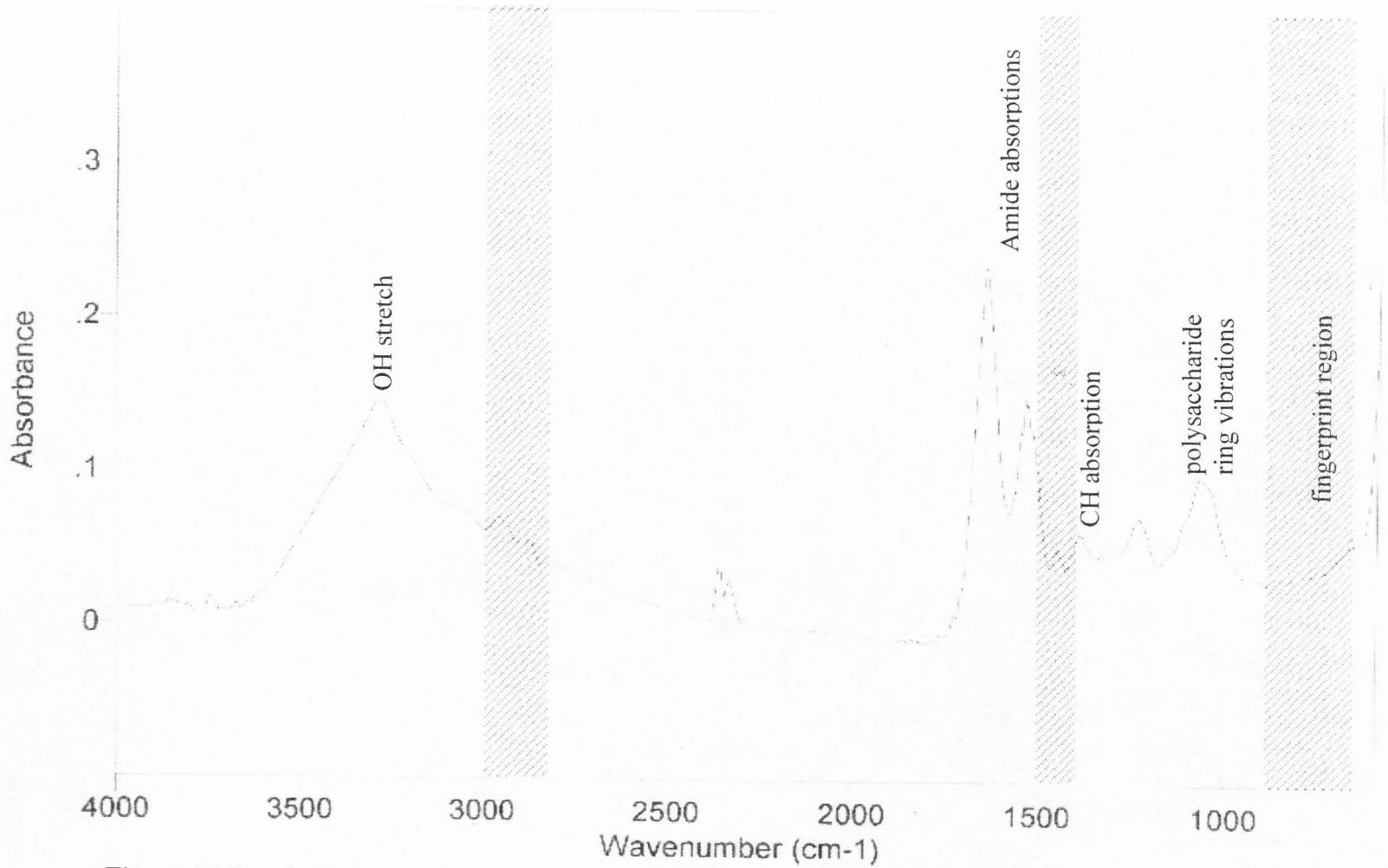


Figure A1 Sample Spectra showing tentative assignment of peaks and windows used for analysis (shaded regions)

APPENDIX B: FULL SET OF SPECTRA RECORDED

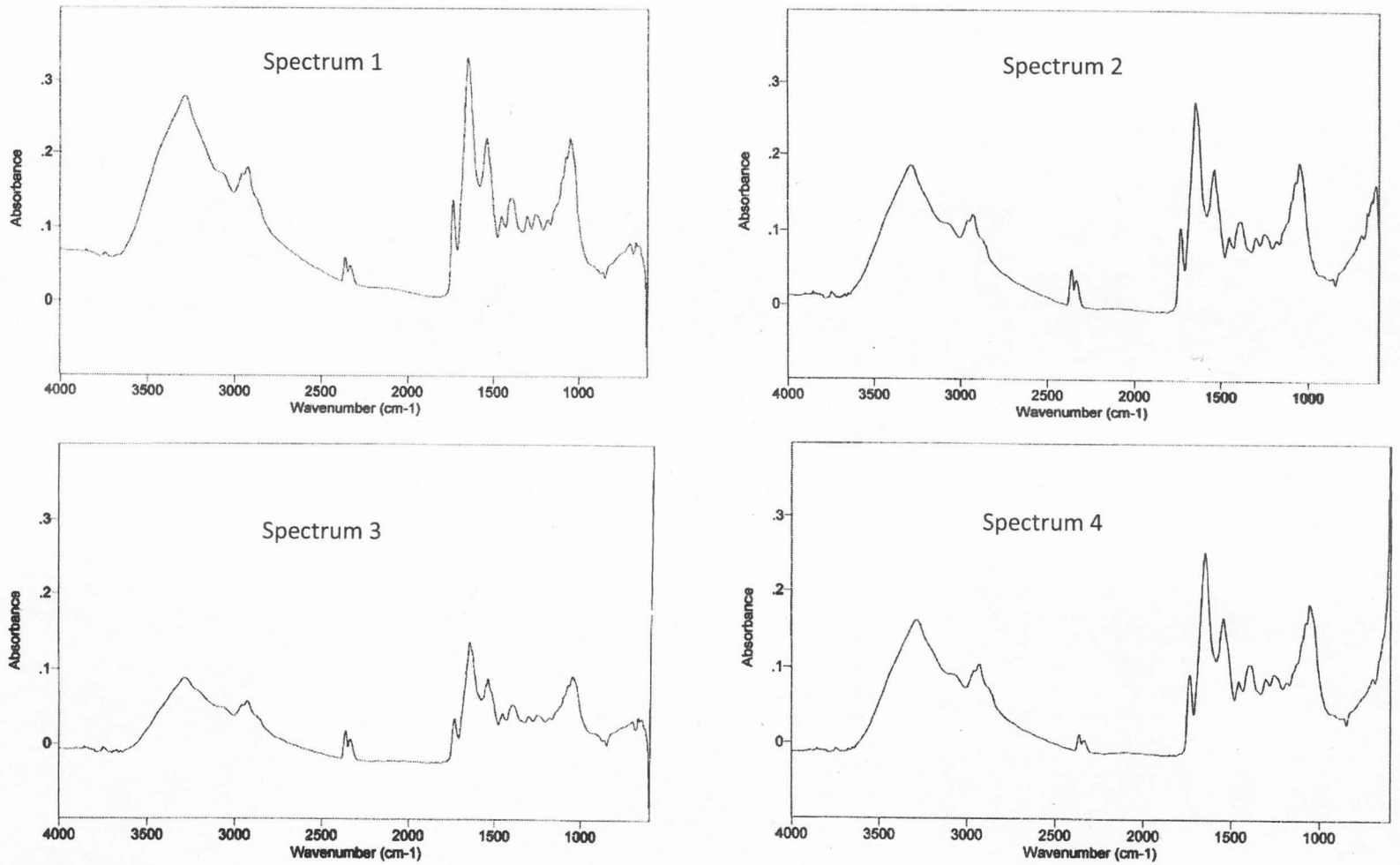


Figure B1. Spectra of sub-samples of *Bacillus subtilis*, replicate 1

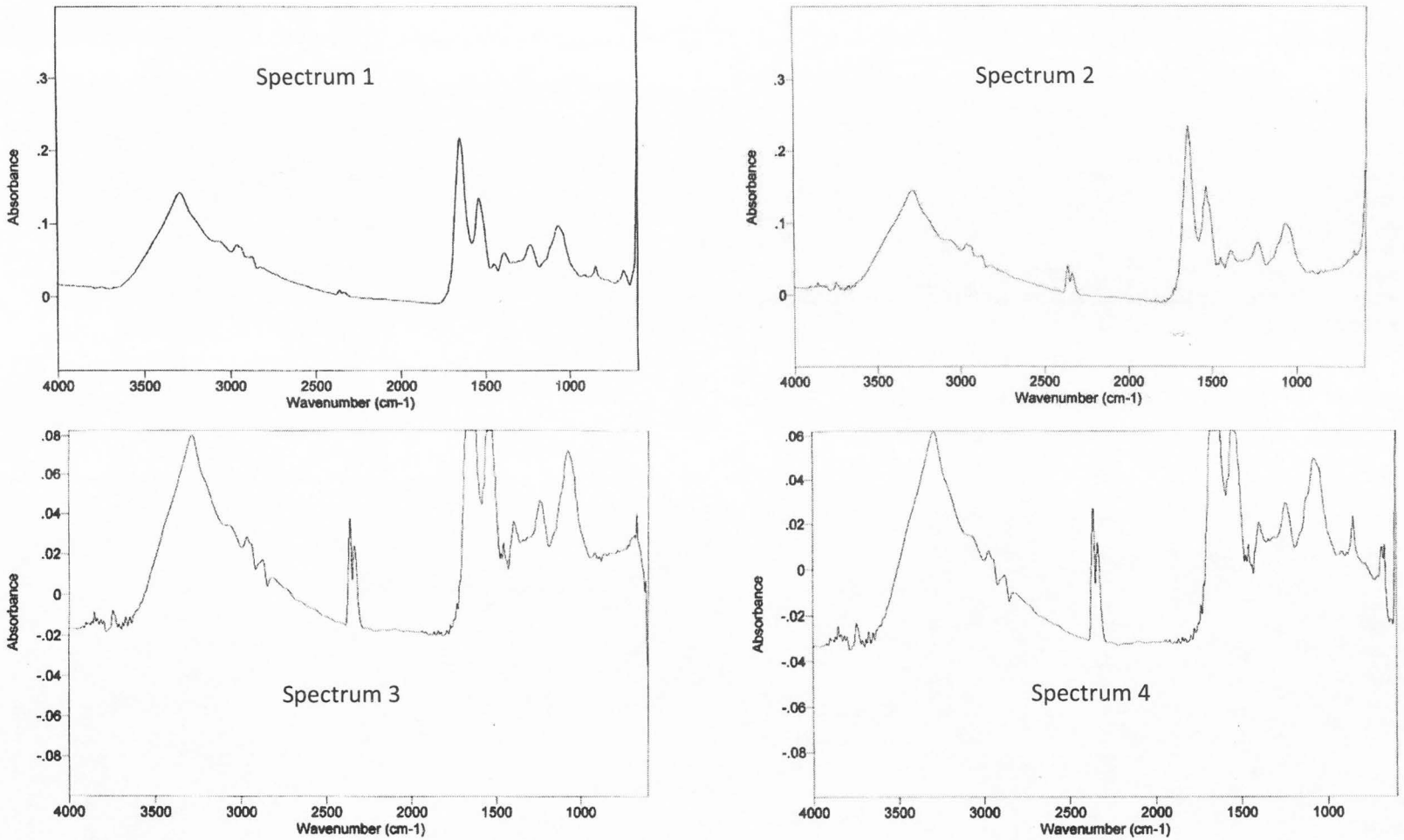


Figure B2. Spectra of sub-samples of *Lactobacillus casei* 202, replicate 2

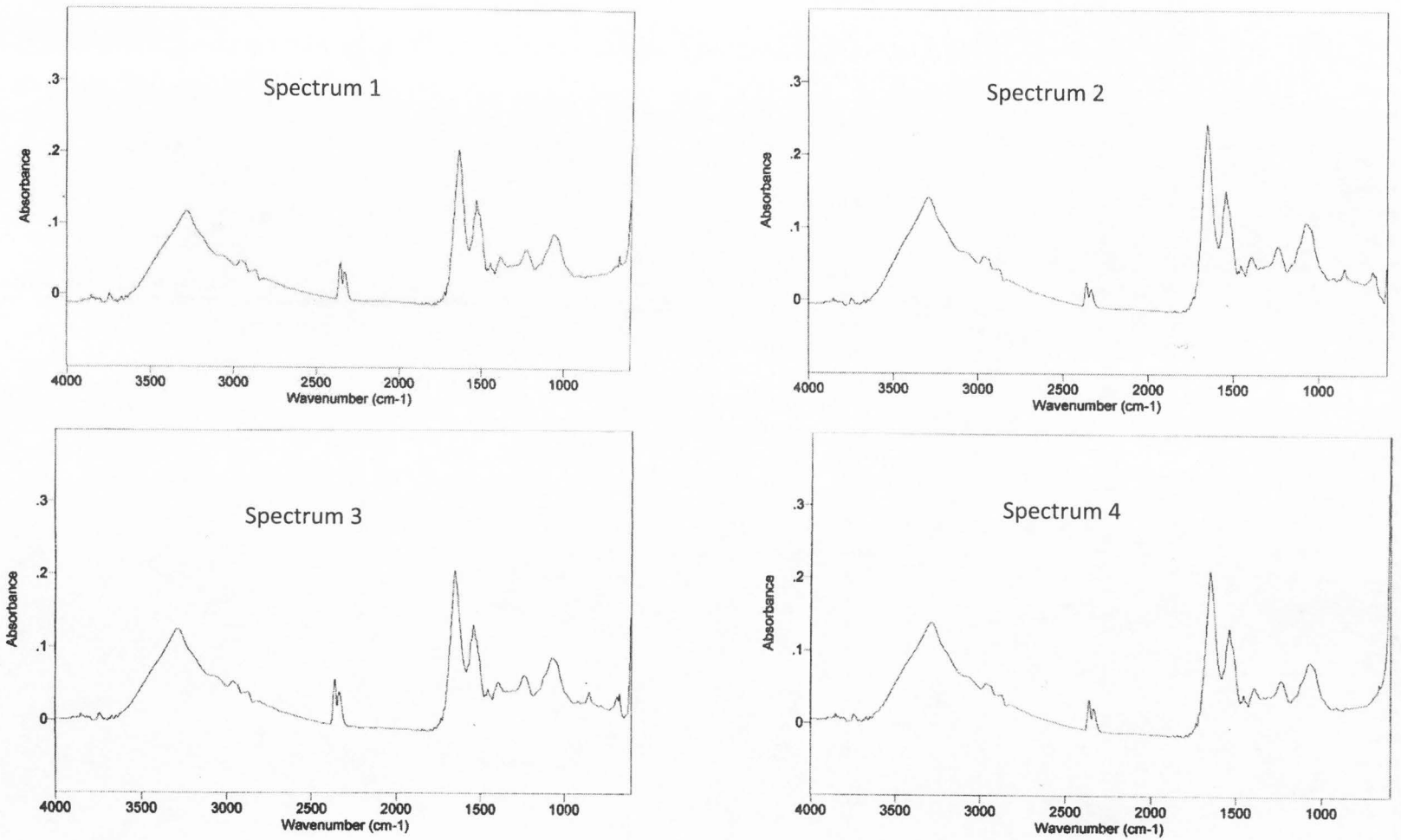


Figure B3. Spectra of sub-samples of *Lactobacillus casei* 202, replicate 3

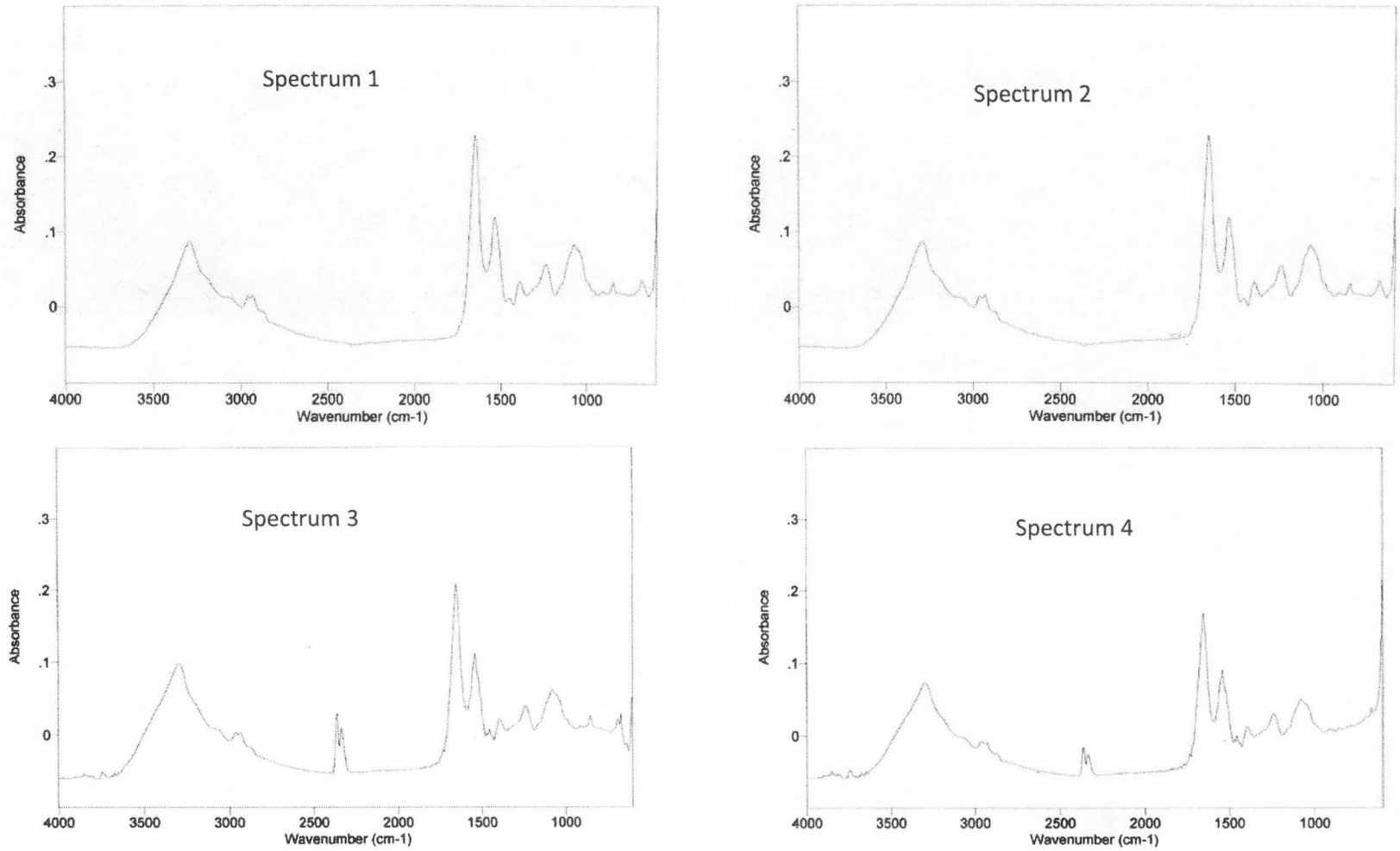


Figure B4. Spectra of sub-samples of *Pseudomonas fluorescens* AFT 29, replicate 1

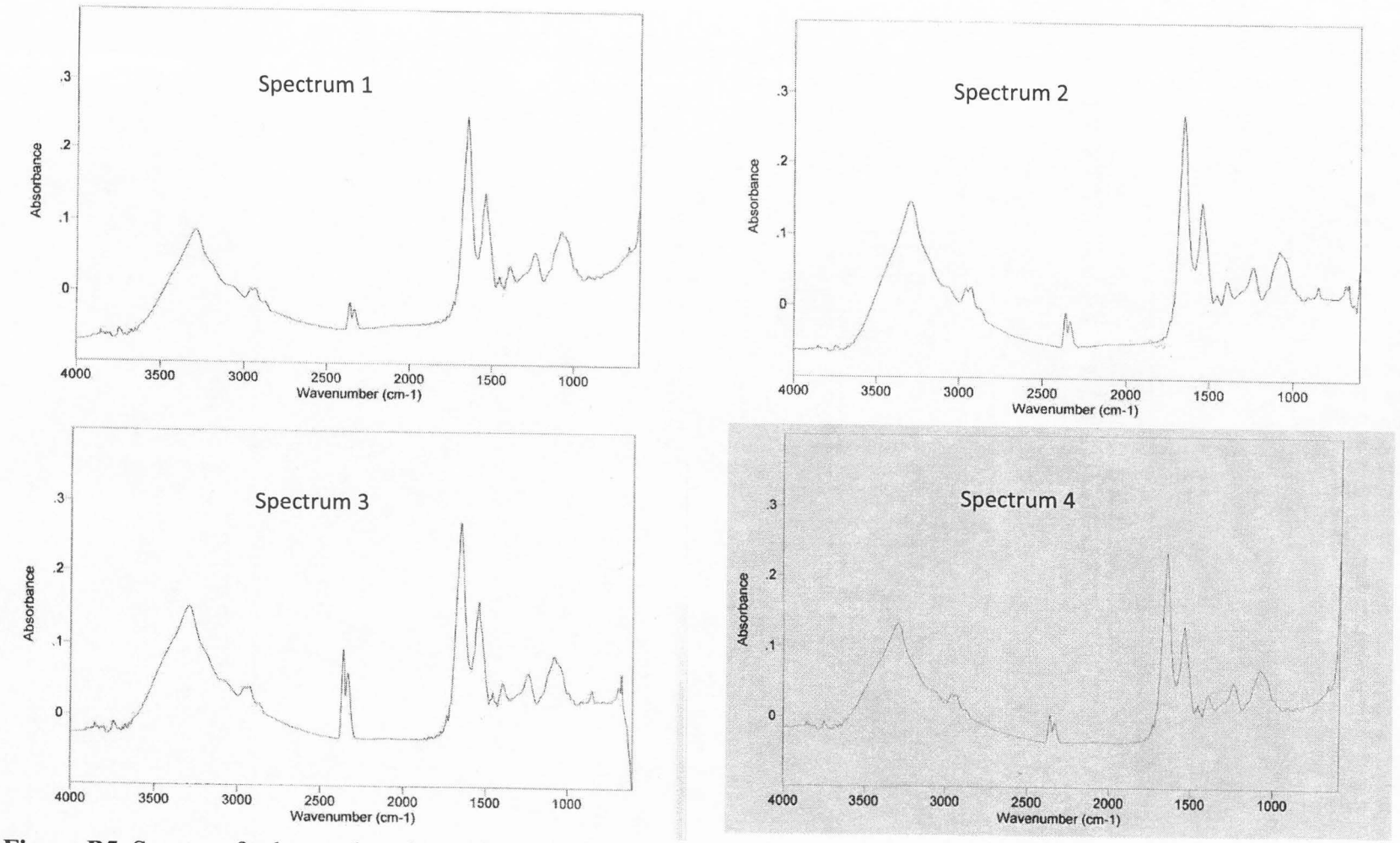


Figure B5. Spectra of sub-samples of *Pseudomonas fluorescens* AFT 29, replicate 2

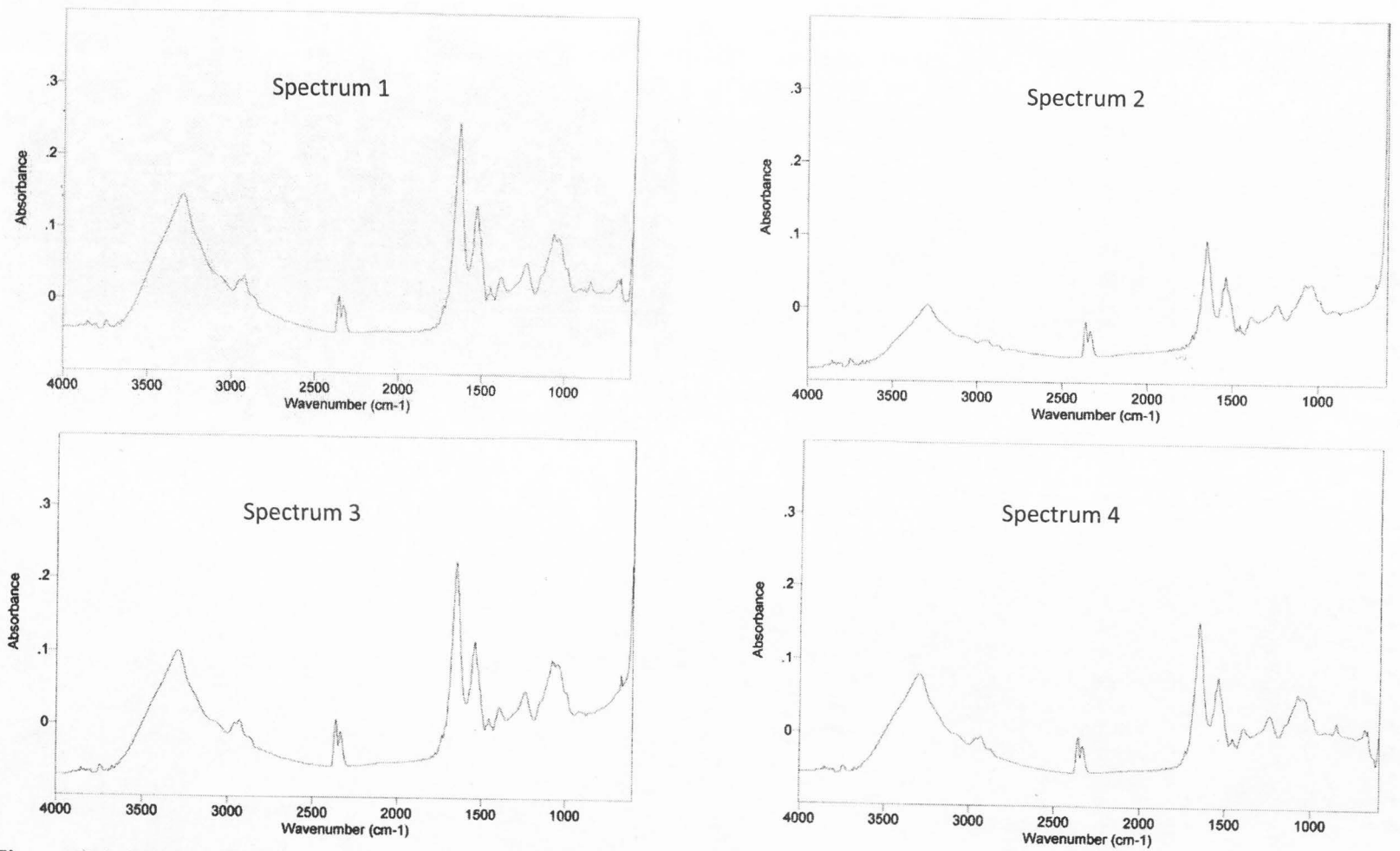


Figure B6. Spectra of sub-samples of *Pseudomonas fluorescens* B52, replicate 1

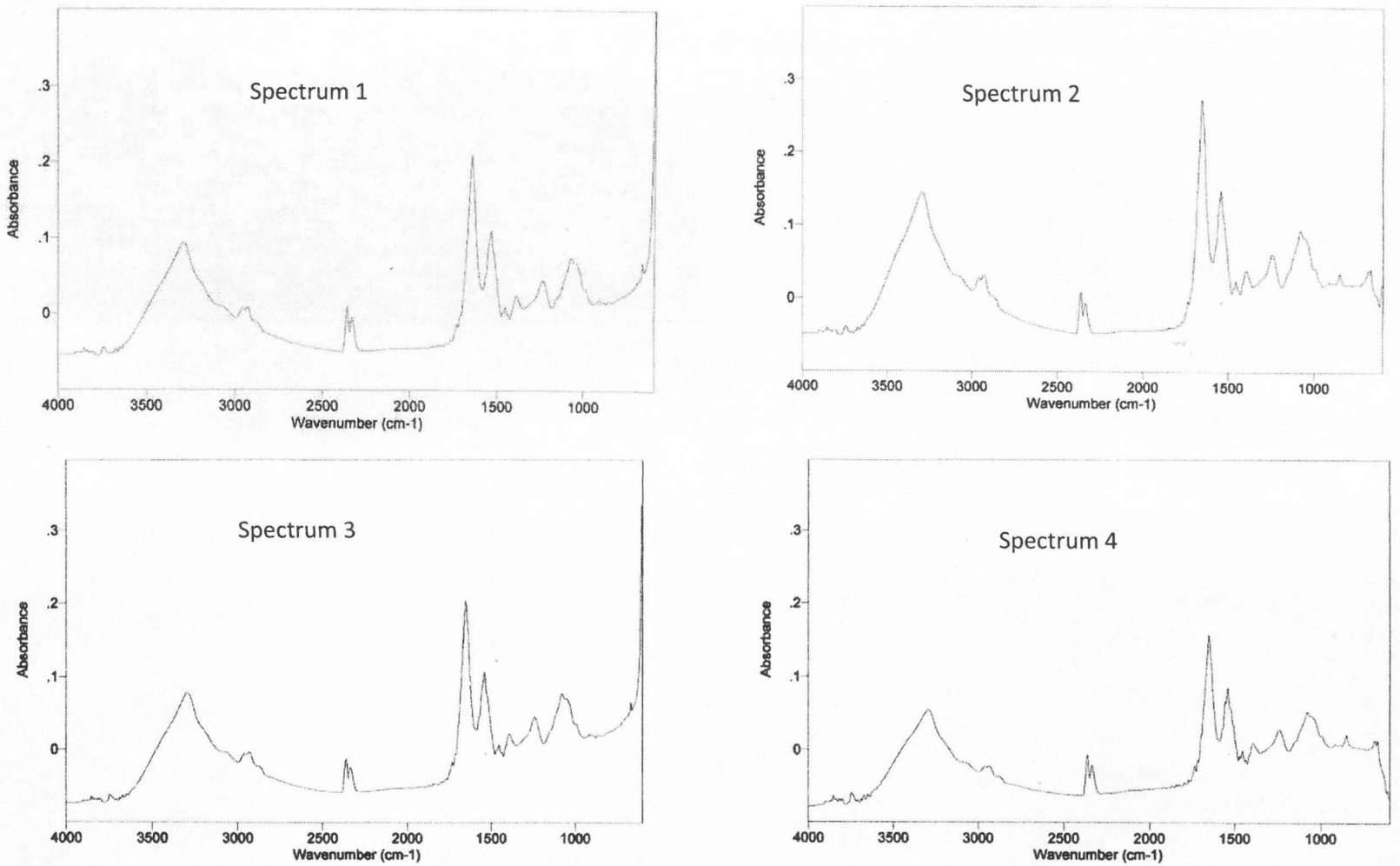


Figure B7. Spectra of sub-samples of *Pseudomonas fluorescens* B52, replicate 2

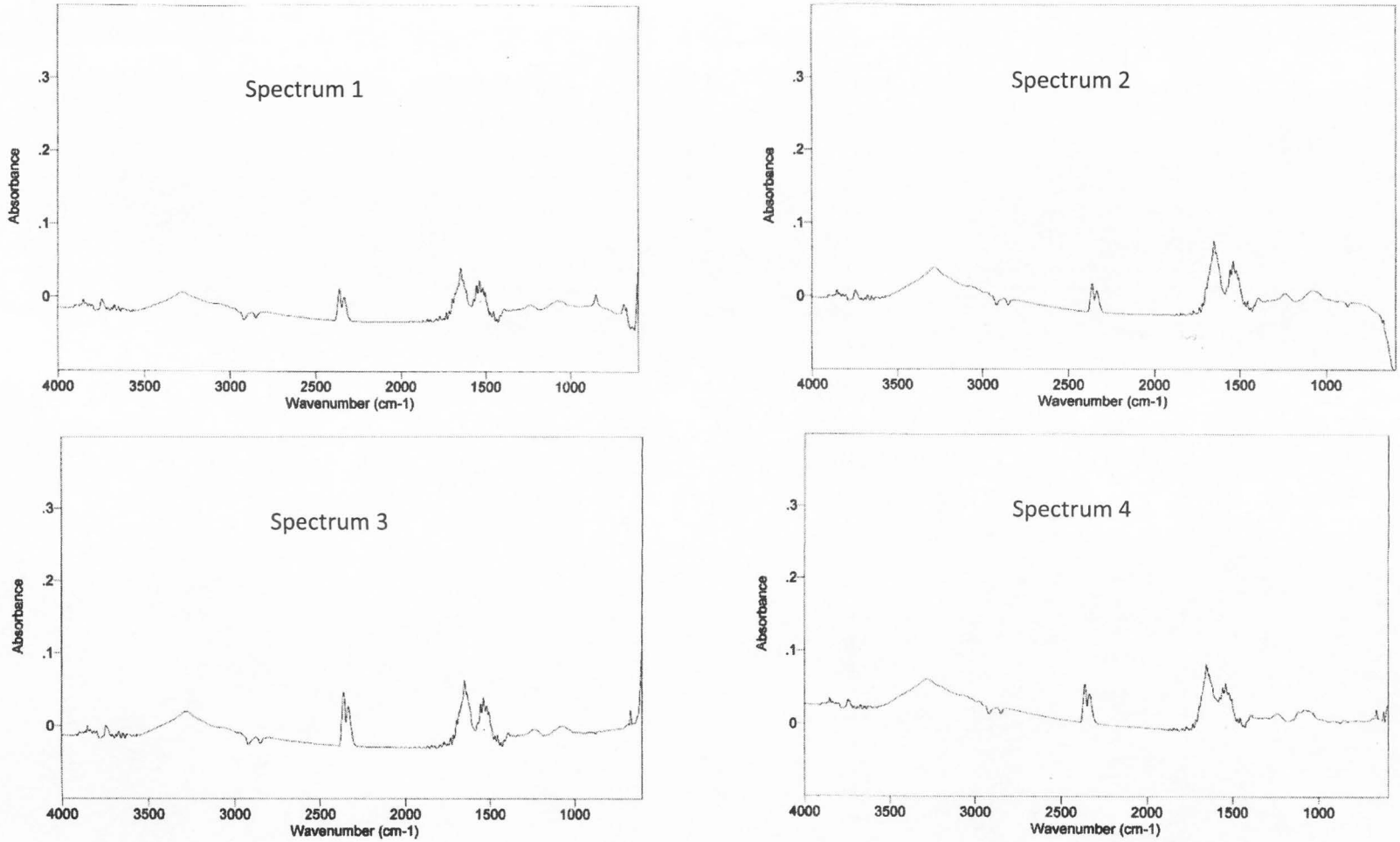


Figure B8. Spectra of sub-samples of *Lactococcus lactis* ssp. *cremoris*, replicate 1

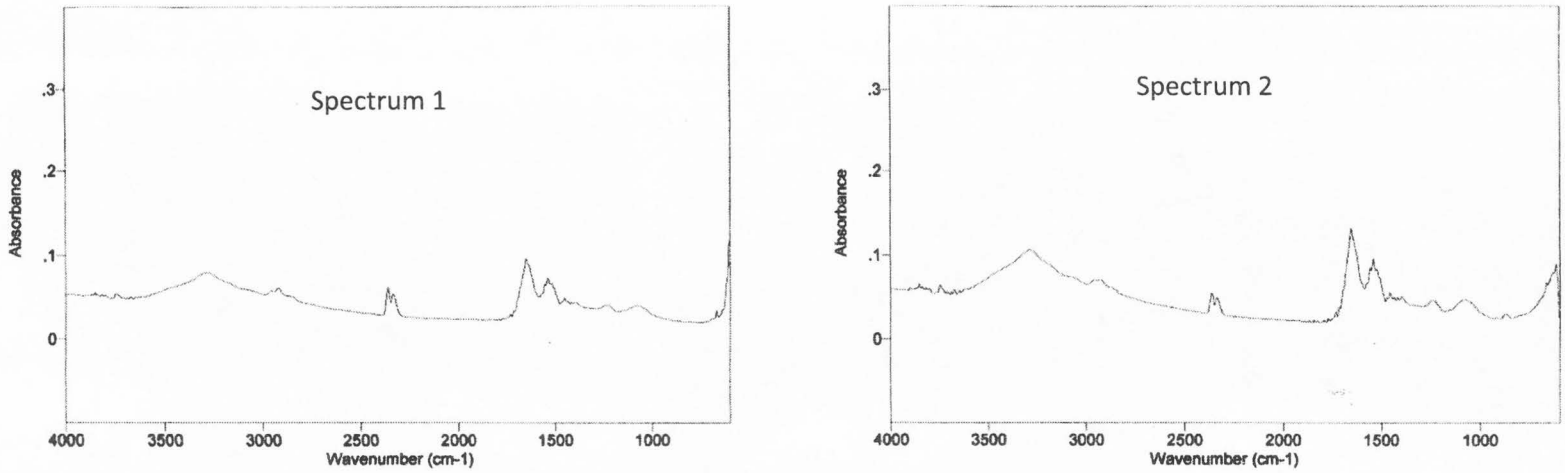


Figure B9. Spectra of sub-samples of *Lactococcus lactis* ssp *cremoris*, replicate 2

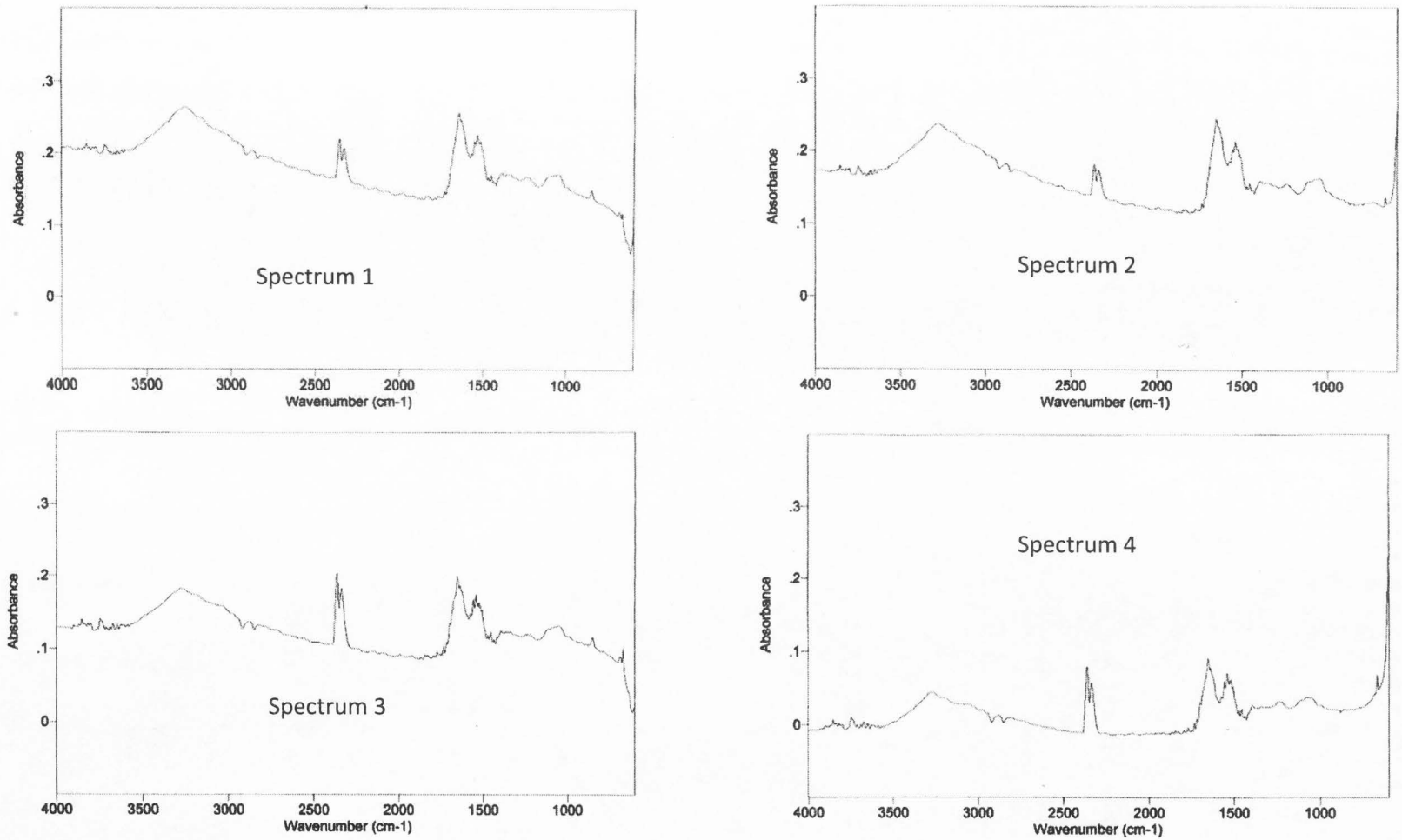


Figure B10. Spectra of sub-samples of *Lactococcus lactis* ssp. *cremoris*, replicate 3

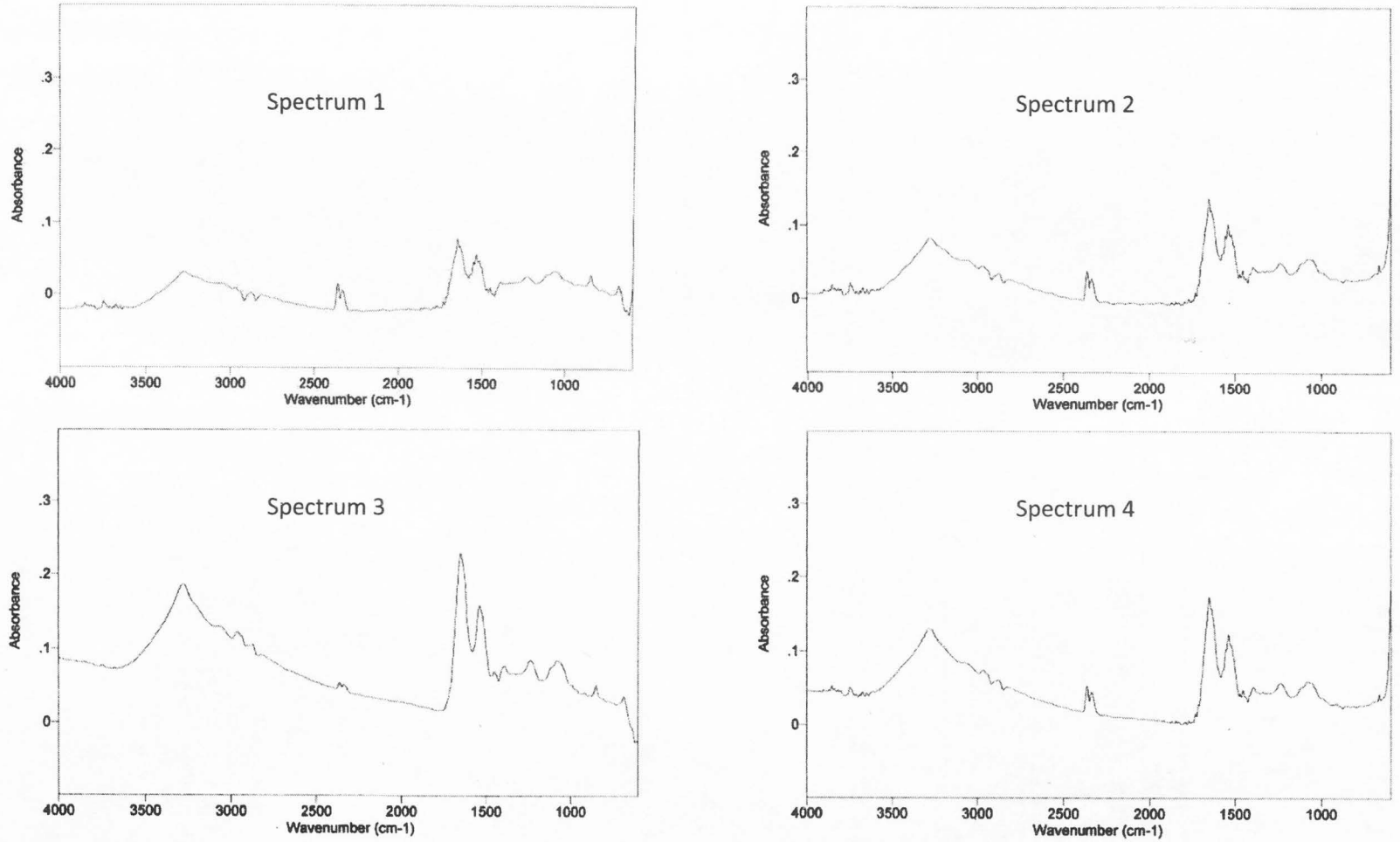
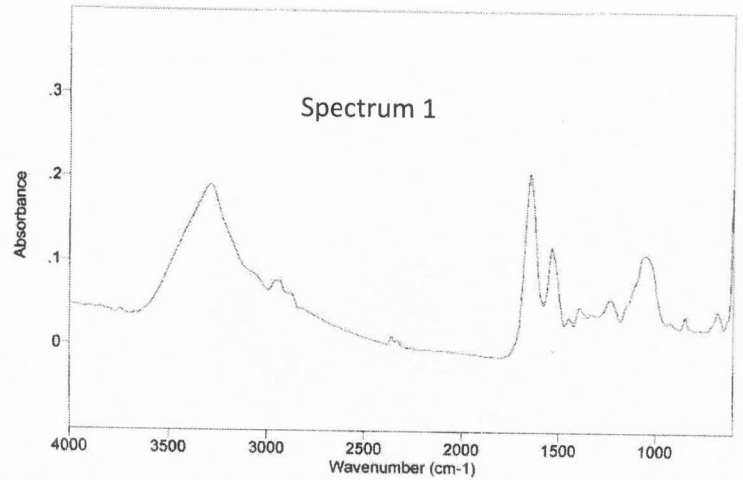
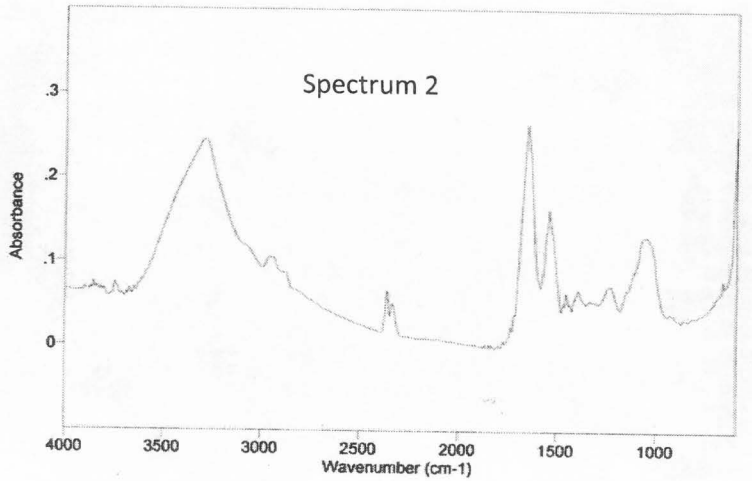
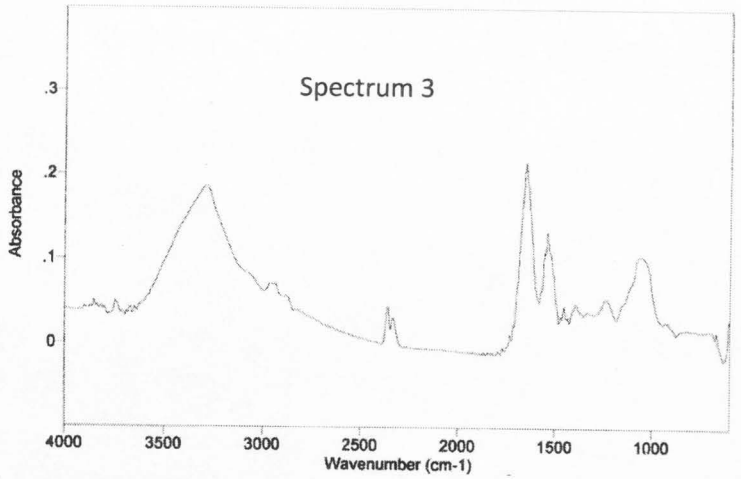


Figure B11. Spectra of sub-samples of *Lactococcus lactis* ssp. *cremoris*, replicate 4







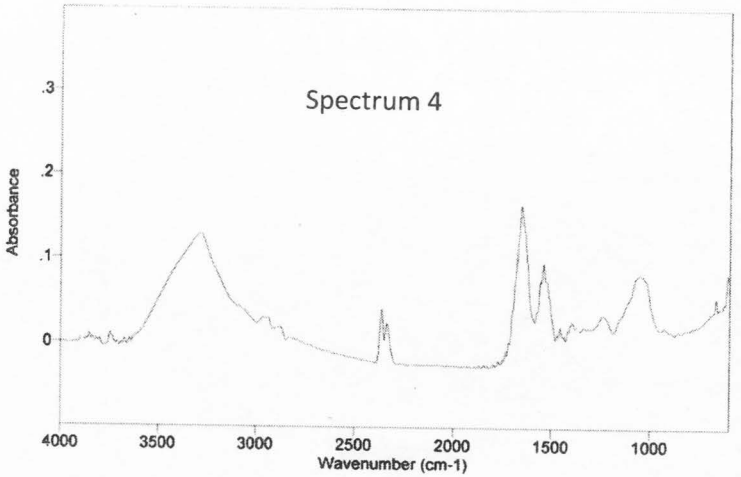


Figure B12. Spectra of sub-samples of *Leuconostoc*, replicate 1

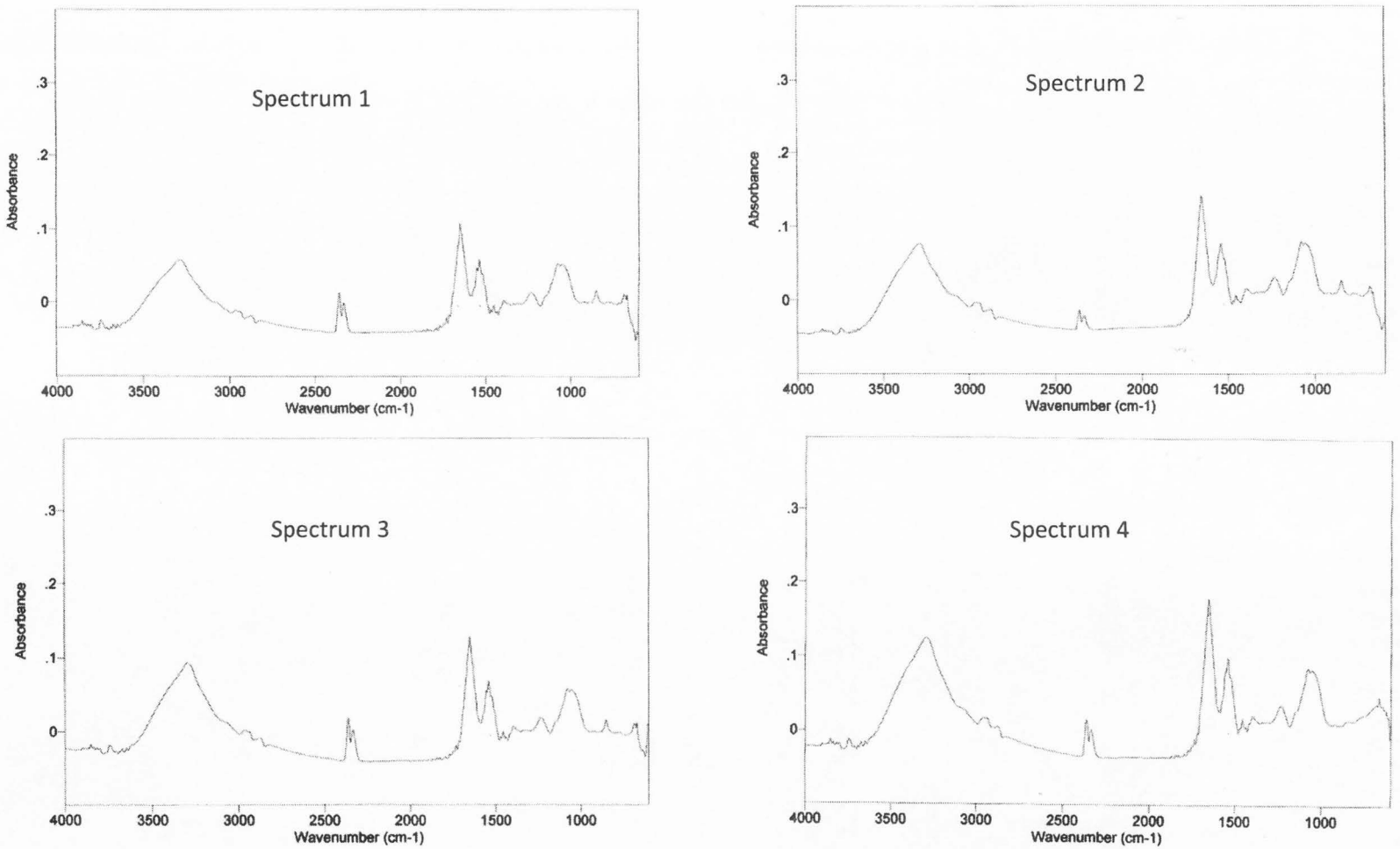


Figure B13. Spectra of sub-samples of *Pediococcus pentosaceus*, replicate 1

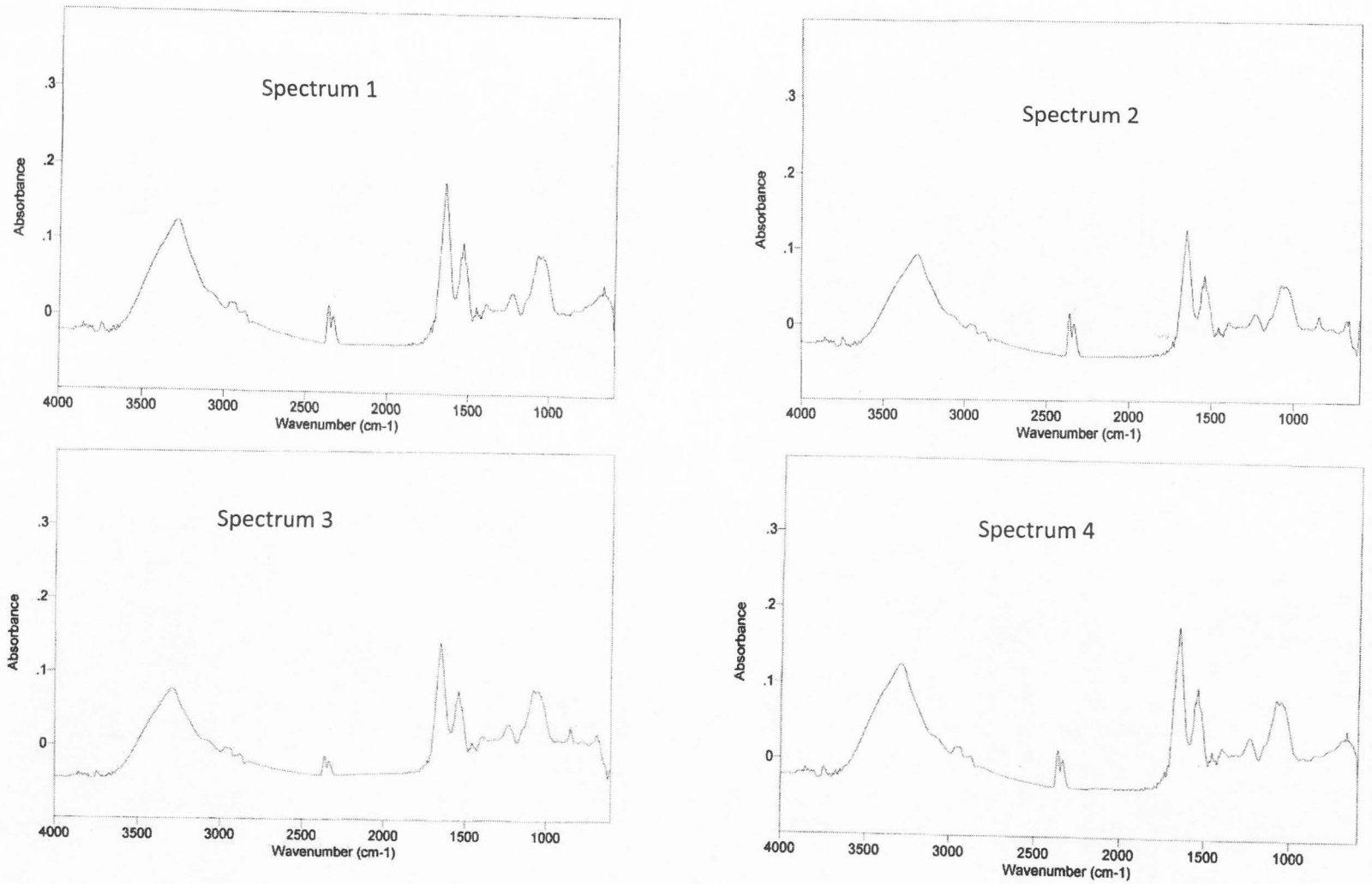


Figure B14. Spectra of sub-samples of *Pediococcus pentosaceus*, replicate 2

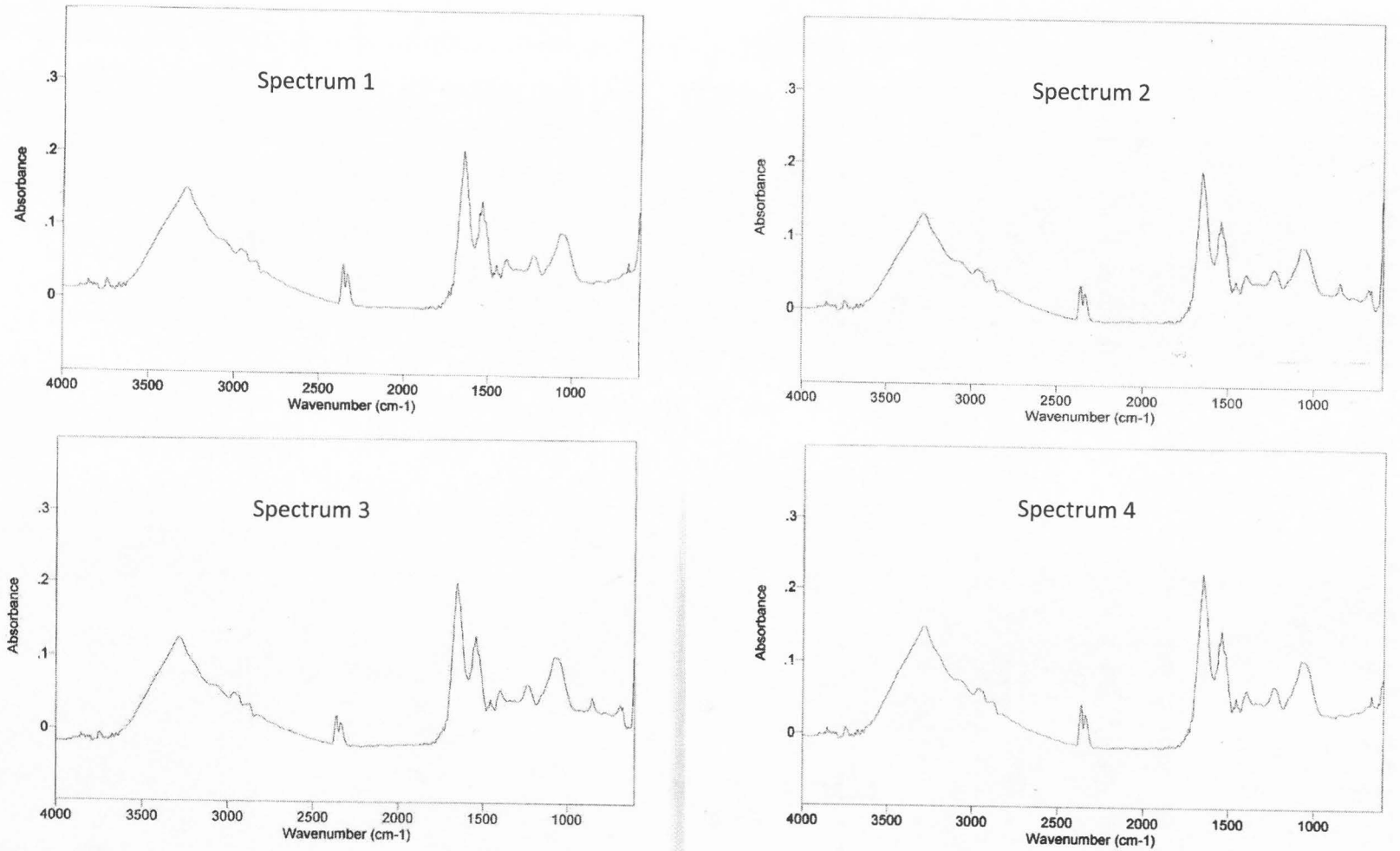


Figure B15. Spectra of sub-samples of *Pediococcus pentosaceus*, replicate 3

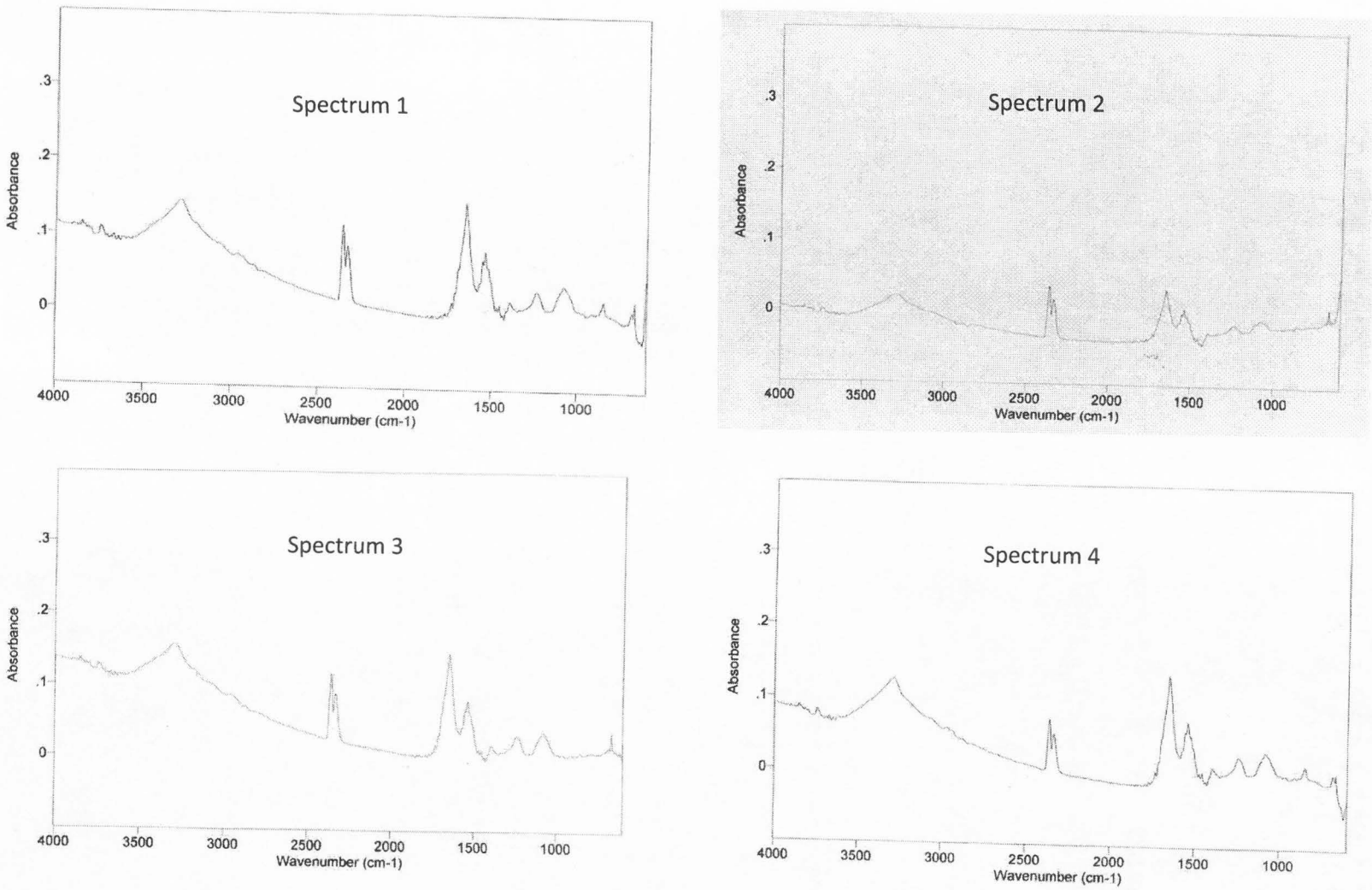
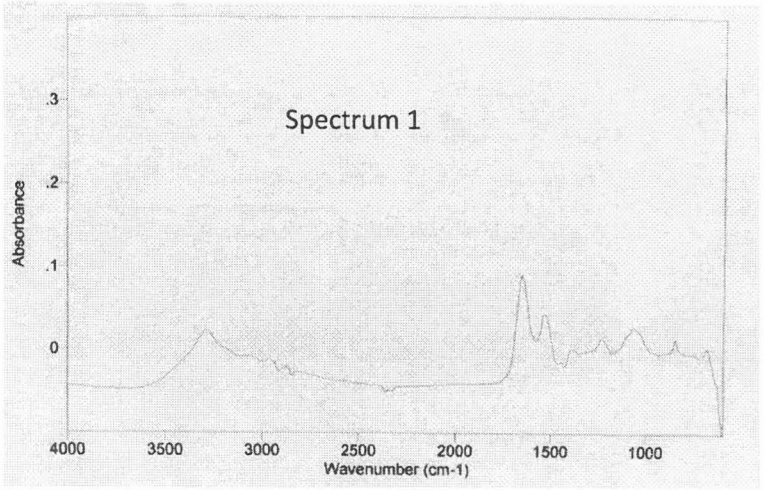
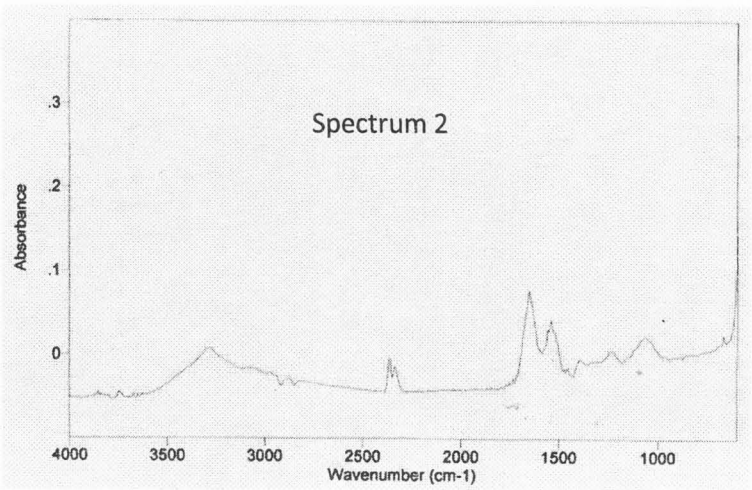


Figure B16 Spectra of sub-samples of *Escherichia coli*





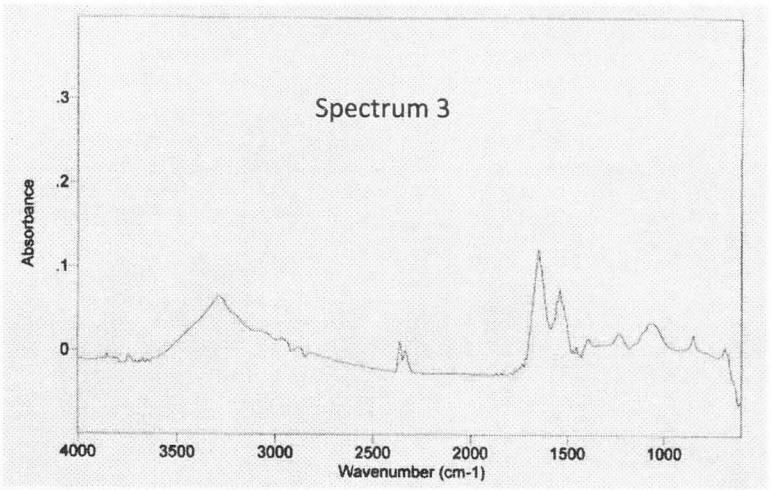


Figure B17 Spectra of sub-samples of *Pediococcus acidiladici*, replicate 1

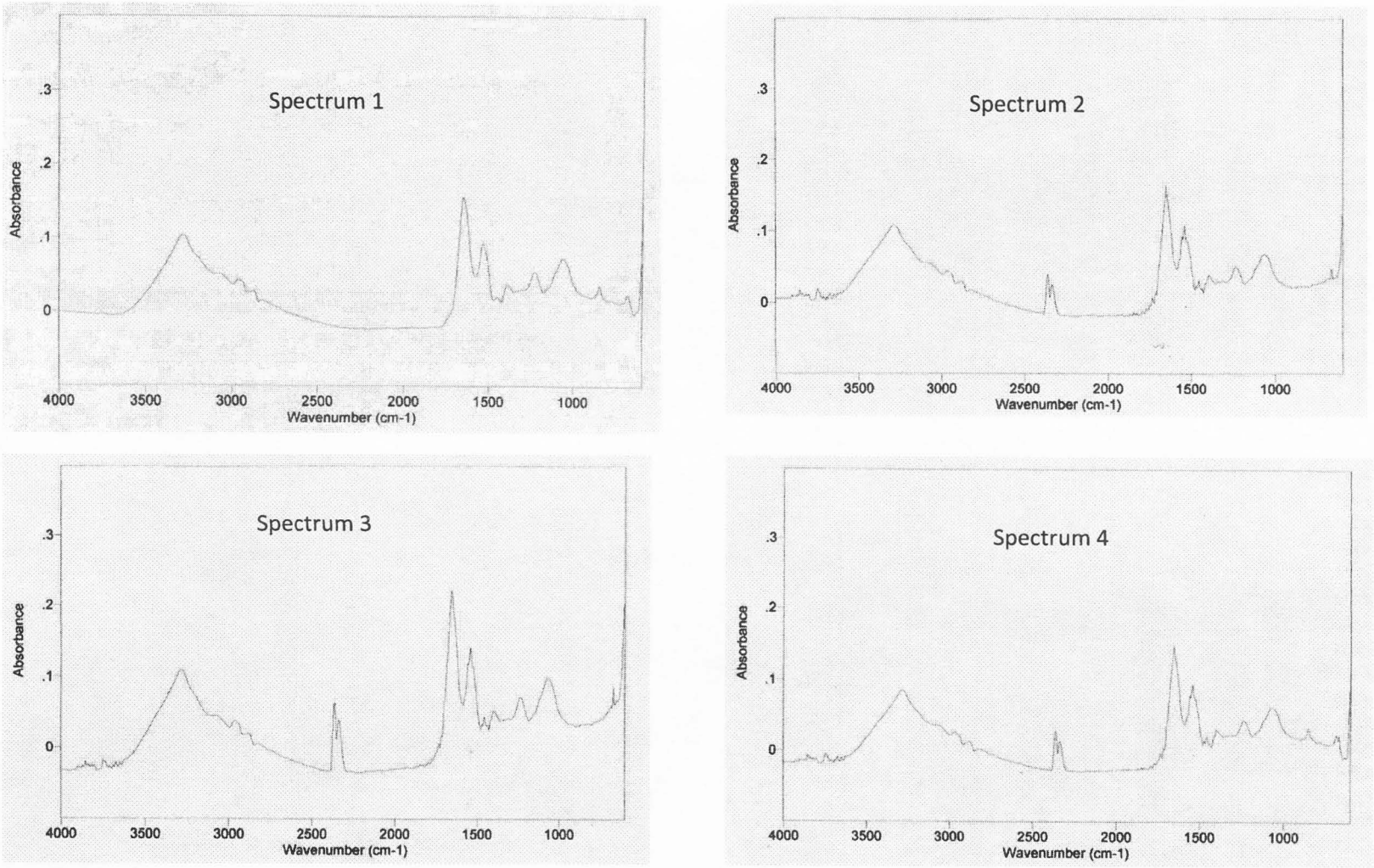


Figure B18 Spectra of sub-samples of *Pediococcus acidiladici*, replicate 2

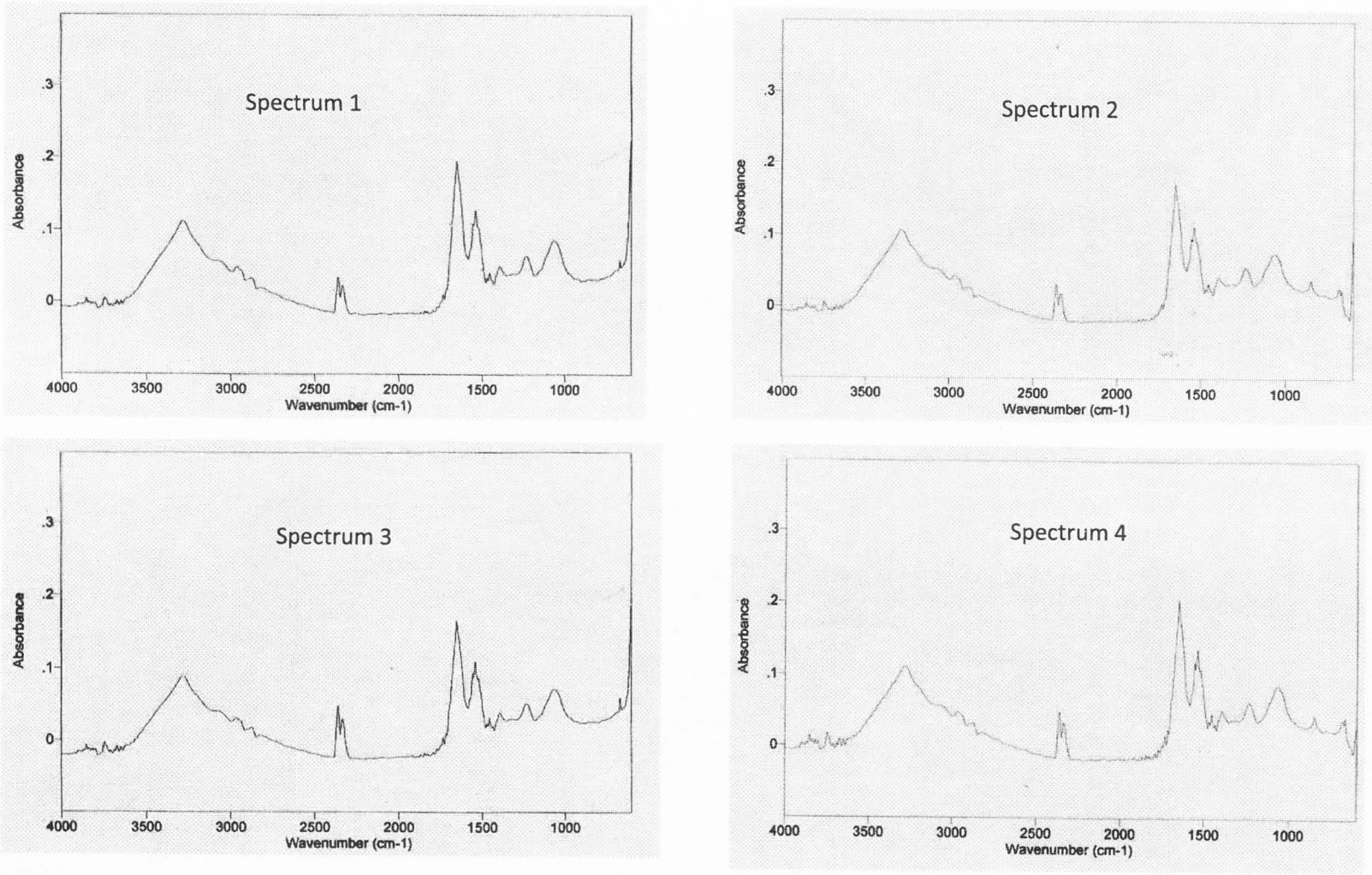


Figure B19 Spectra of sub-samples of *Pediococcus acidiladici*, replicate 3

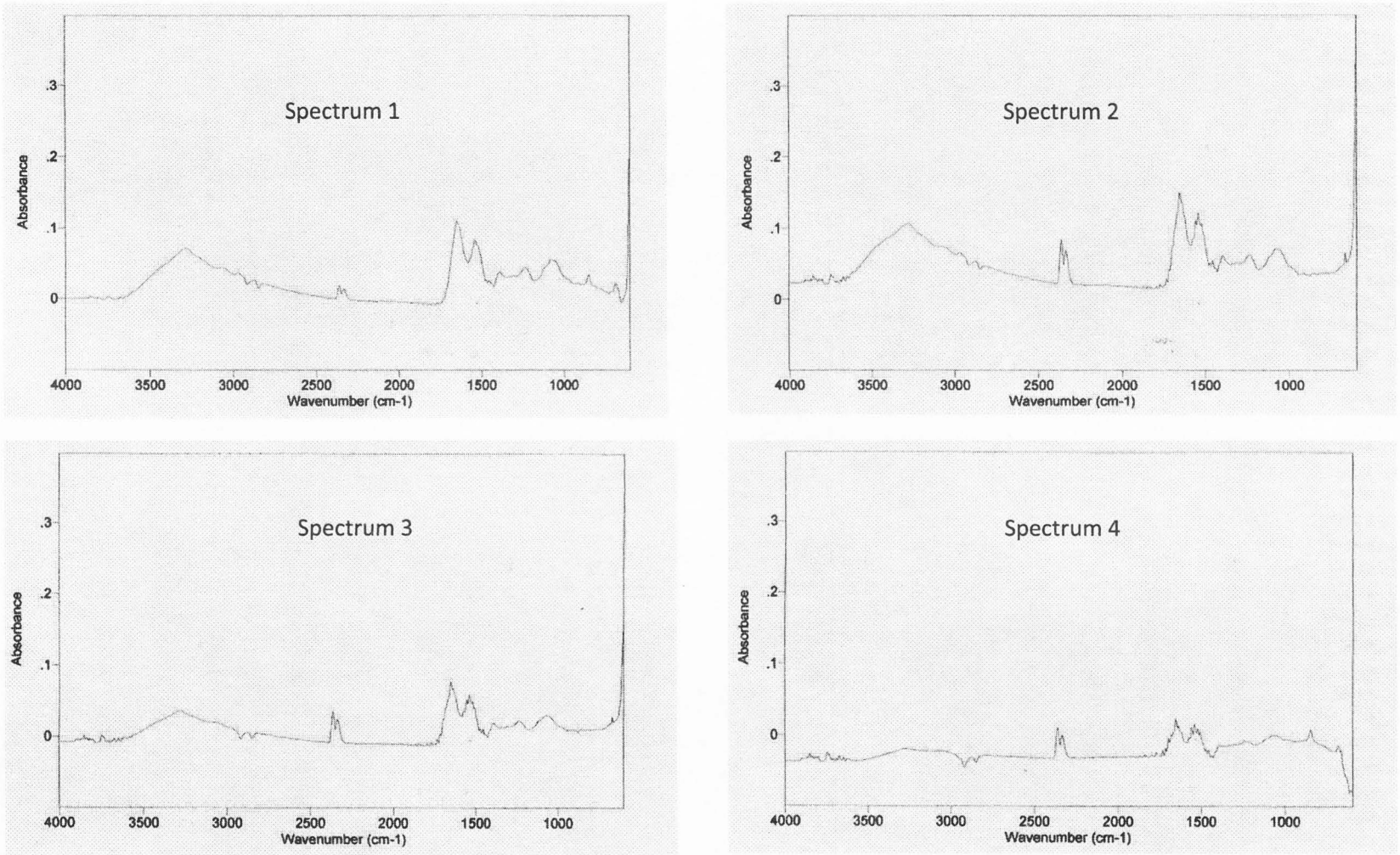


Figure B20 Spectra of sub-samples of *Lactococcus lactis* ssp *lactis*, replicate 1

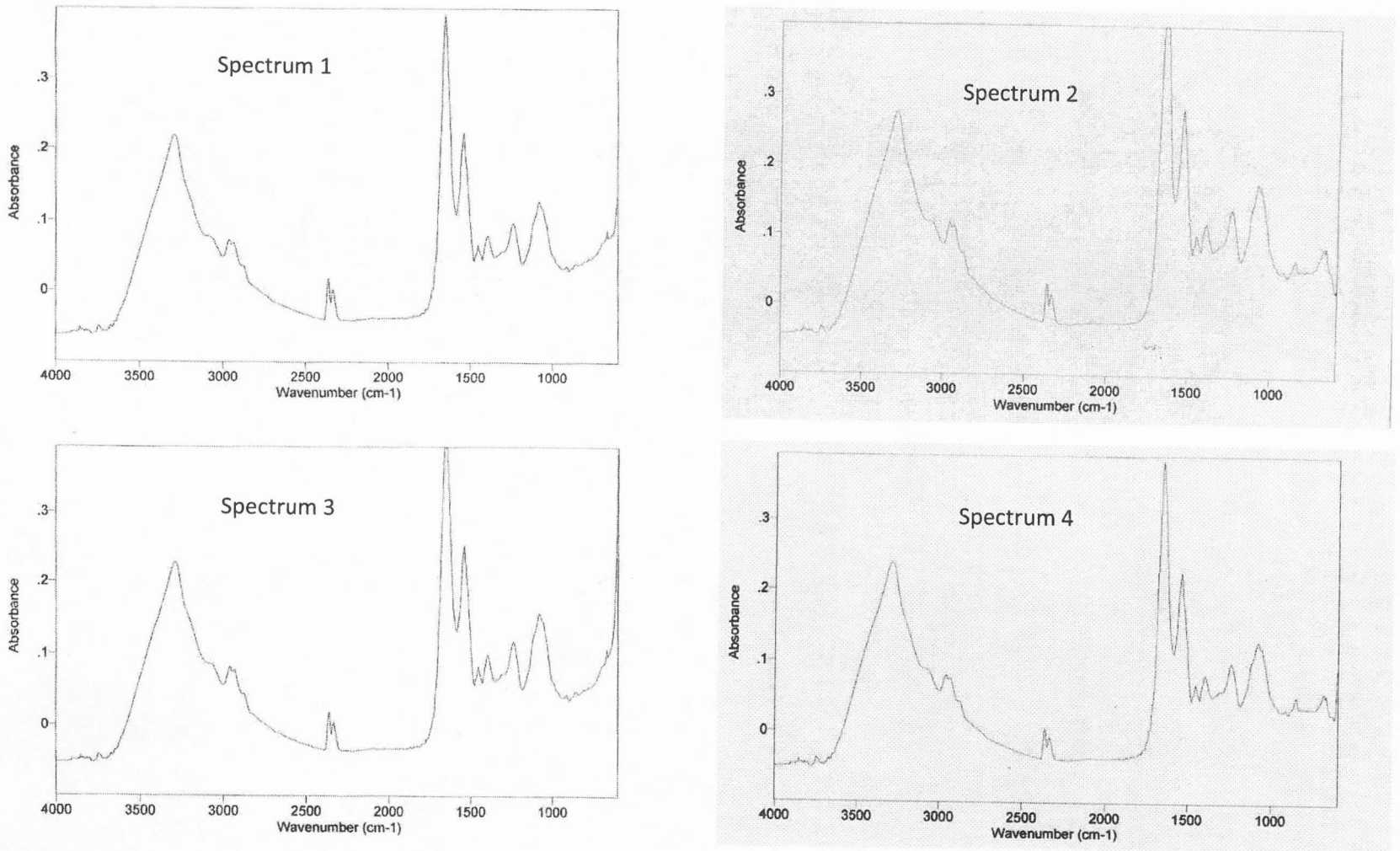


Figure B21 Spectra of sub-samples of *Lactococcus lactis* ssp. *lactis*, replicate 2

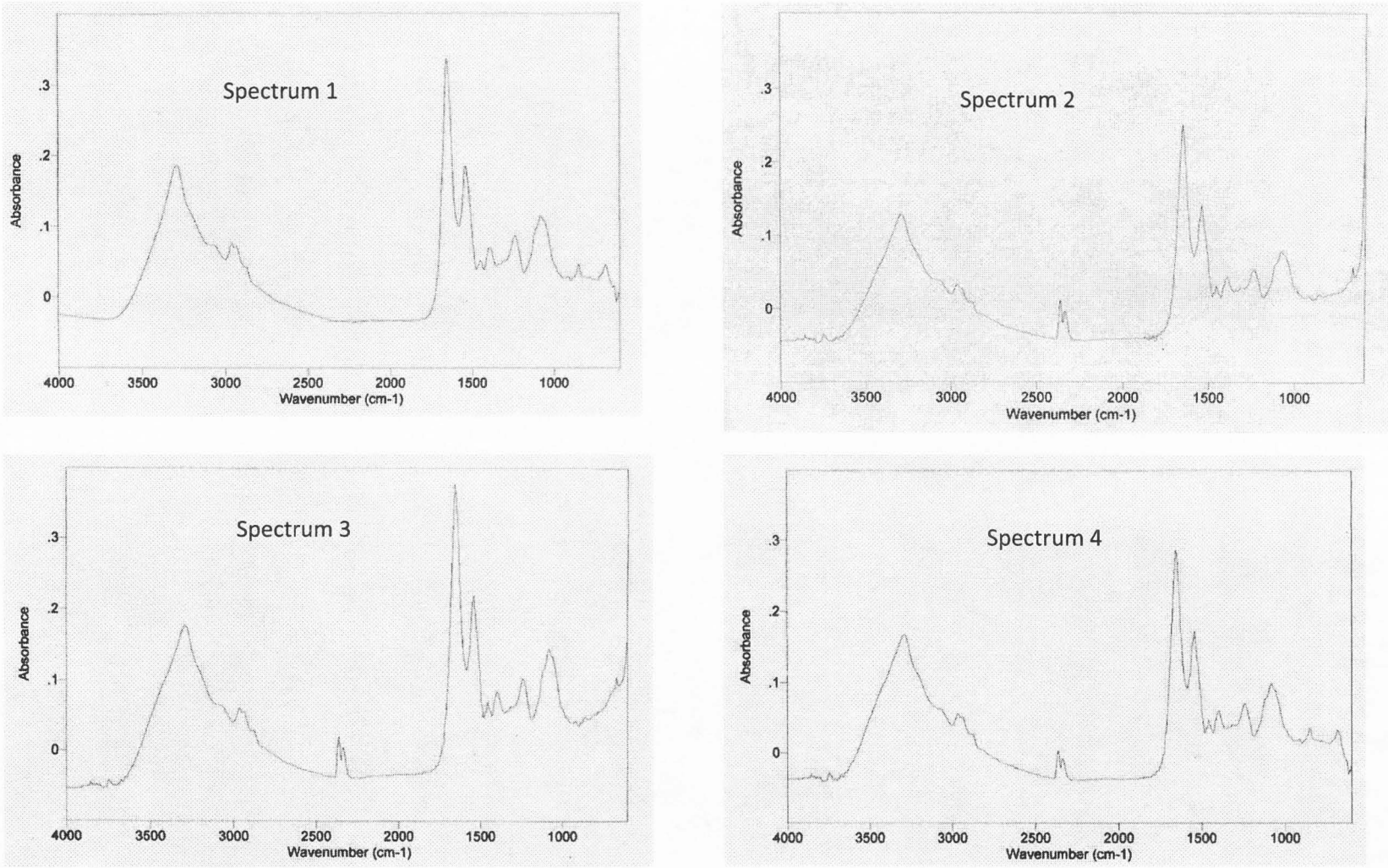


Figure B22 Spectra of sub-samples of *Lactococcus lactis* ssp *lactis*, replicate 3

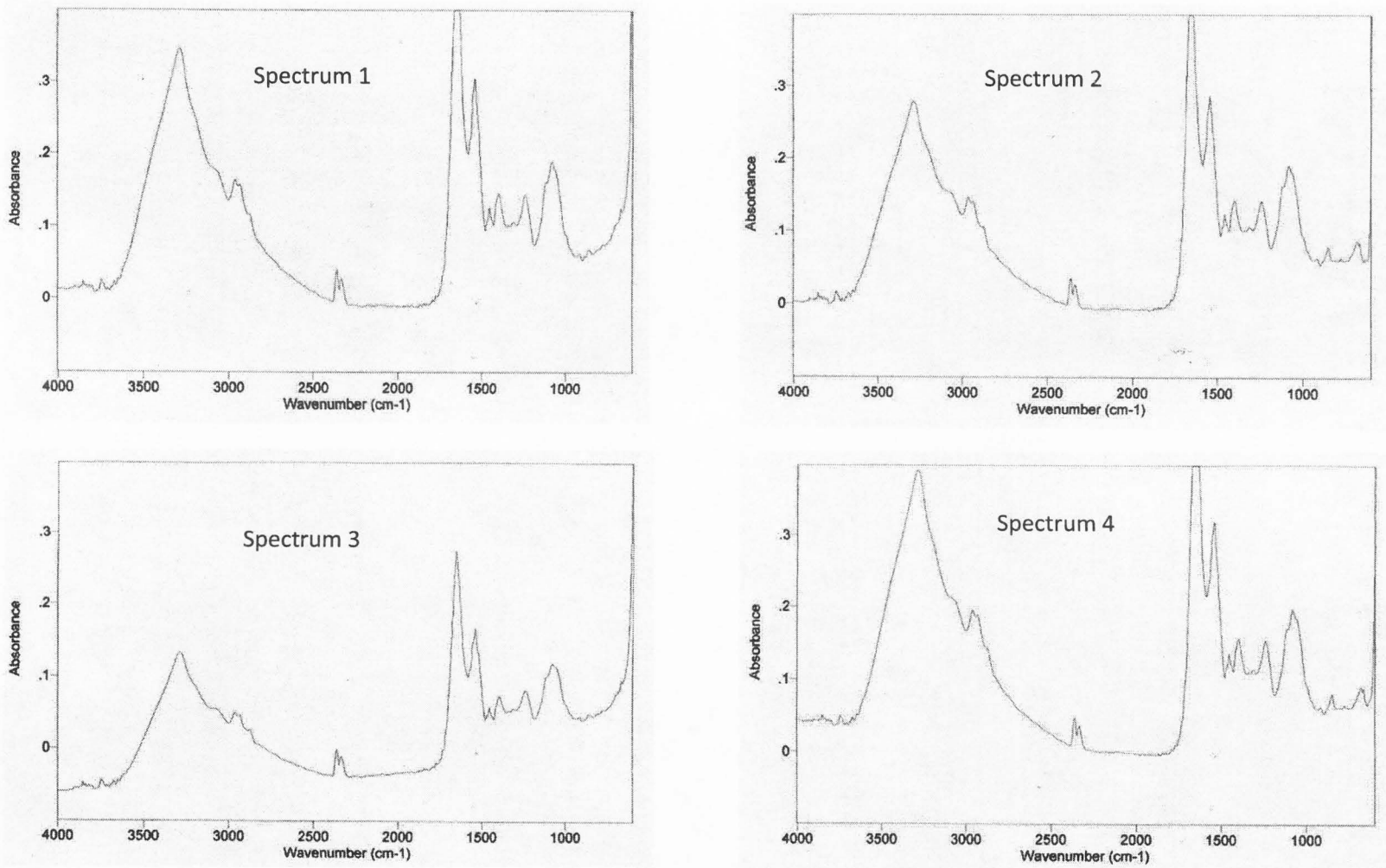


Figure B23 Spectra of sub-samples of *Lactococcus lactis* ssp. *lactis*, replicate 4

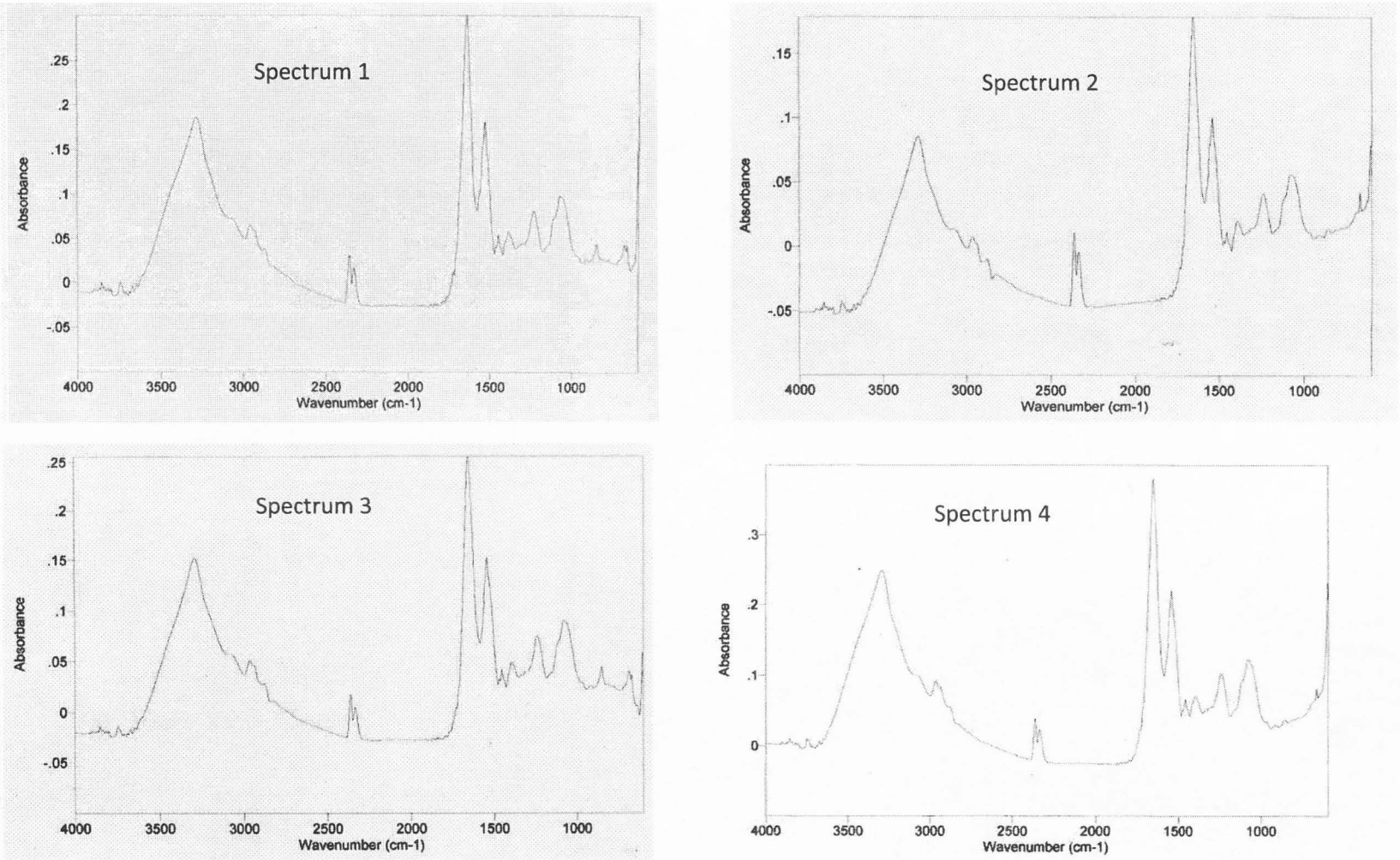


Figure B24 Spectra of sub-samples of *Lactobacillus casei* 201, replicate 1

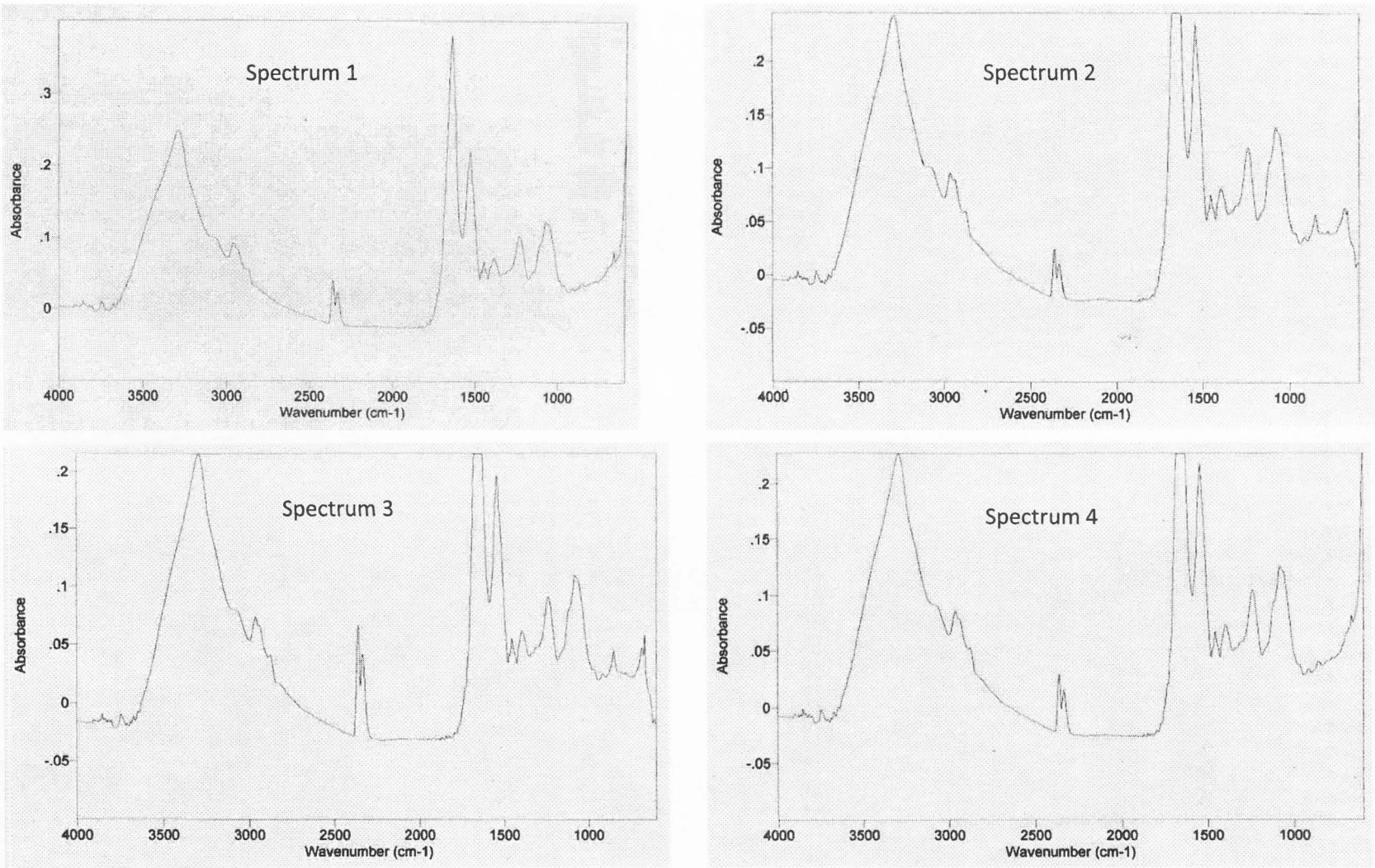


Figure B25 Spectra of sub-samples of *Lactobacillus casei* 201, replicate 2

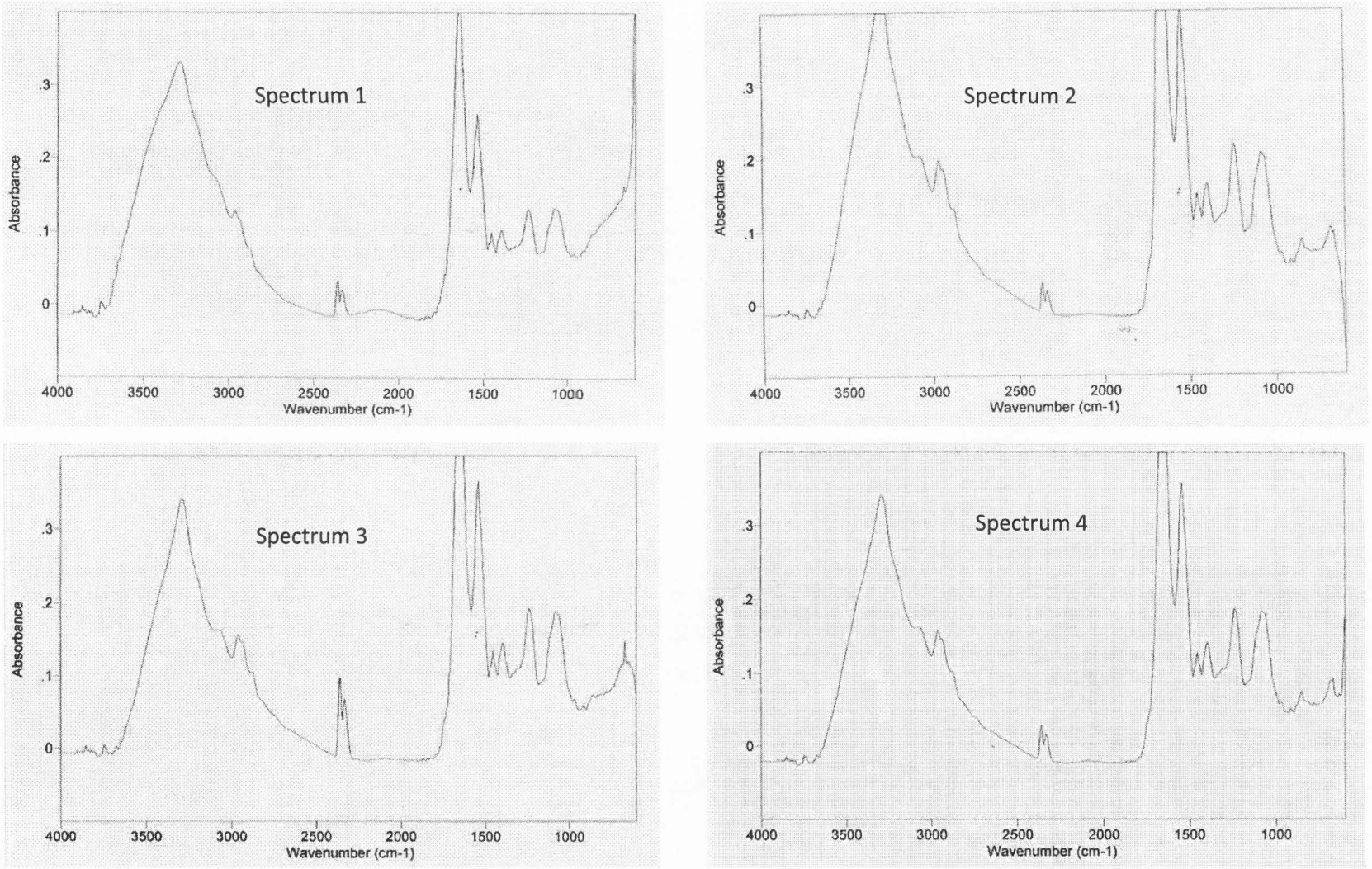


Figure B26 Spectra of sub-samples of *Lactobacillus helveticus*, replicate 1

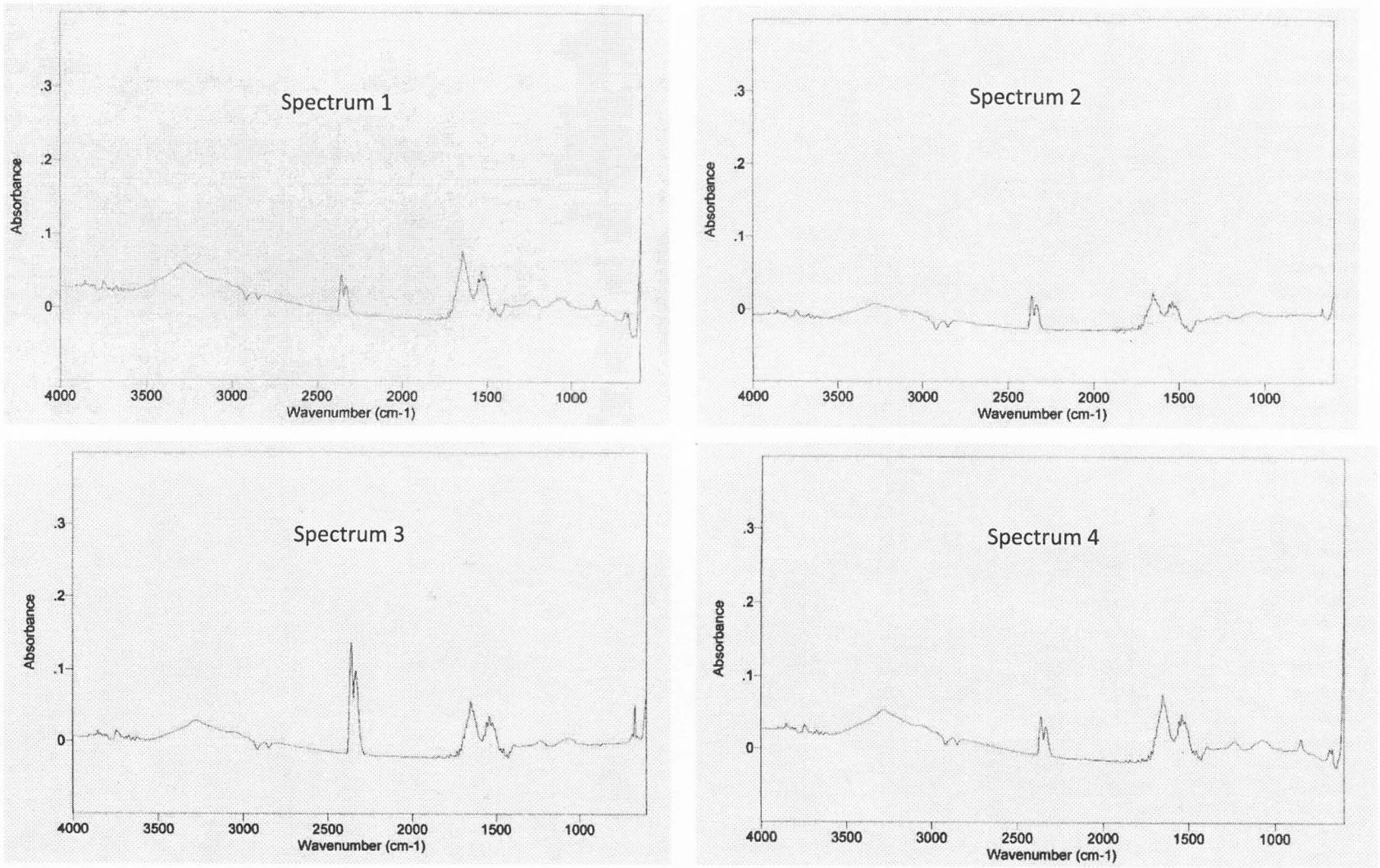


Figure B27 Spectra of sub-samples of *Lactobacillus helveticus*, replicate 2

APPENDIX C: Spectral Distances

Table C1a Spectral Distances Codes used can be found at end of tables

NAME	**A100**	**A111**	**A121**	**A122**	**A130**	**A131**
A100	0	914.8267	92.65536	85.2563	363.3253	262.0135
A111	914.8267	0	578.7174	592.5657	207.9689	315.0659
A121	92.65536	578.7174	0	0.345221	144.2393	82.91543
A122	85.2563	592.5657	0.345221	0	148.3573	84.54668
A130	363.3253	207.9689	144.2393	148.3573	0	13.15037
A131	262.0135	315.0659	82.91543	84.54668	13.15037	0
A132	399.2027	174.6404	183.9906	189.9291	12.90345	39.24296
A133	383.8191	188.9464	166.0412	170.953	4.077943	24.32068
A200	580.9337	148.9062	258.0234	272.236	122.7387	163.6316
A201	667.5596	1884.109	969.9916	954.7706	1442.196	1292.242
A202	676.2254	1822.708	927.8761	918.358	1428.488	1285.343
A203	544.4716	1758.739	784.0745	772.2416	1286.665	1133.699
A300	105.2154	561.1038	45.50698	47.62963	164.1585	115.5848
A400	293.1633	261.1181	110.7415	117.1517	36.5876	43.12145
A402	72.52281	604.4677	8.499223	8.618046	162.8767	99.71661
A403	133.8045	452.6158	18.20088	20.19656	83.49217	45.72483
A500	758.3509	1876.646	1014.556	1002.97	1487.168	1343.798
A600	416.3293	150.2154	168.1583	176.7352	23.45253	53.63725
A601	194.2278	375.3926	34.05509	37.31802	51.4458	26.73014
A602	192.2691	378.4069	32.76458	36.27619	54.601	29.61791
A700	237.6268	378.1741	45.22441	51.67212	90.83351	68.15673
A701	119.4562	607.4459	30.83539	27.30011	140.0318	71.39188
A702	151.2076	513.1469	27.35588	26.09458	93.06618	39.8961
A703	124.3579	1179.708	213.0104	203.0124	570.5969	432.5922
A801	98.01992	608.4081	11.07236	8.994102	149.7433	81.58859
A802	69.13953	800.9086	43.62086	38.19843	268.3275	172.0432
A901	122.7916	1060.247	167.9927	158.1269	478.6005	353.7141
A902	598.1168	1636.616	750.3255	746.005	1247.055	1111.38
ASaline	1015.089	43.98504	723.1808	733.2013	302.1969	421.9306

Table C1b Spectral Distances Codes used can be found at end of tables

NAME	A132	A133	A200	A201	A202	A203
A100	399.20267	383.81909	580.93366	667.55961	676.22543	544.47155
A111	174.64039	188.94645	148.90618	1884.1094	1822.7076	1758.7388
A121	183.99061	166.04124	258.0234	969.99161	927.8761	784.07454
A122	189.92906	170.95296	272.23604	954.77061	918.35805	772.2416
A130	12.903448	4.0779426	122.73868	1442.1956	1428.4882	1286.6652
A131	39.242956	24.320678	163.63159	1292.2418	1285.3432	1133.6993
A132	0	3.1473302	129.66641	1511.9609	1492.5915	1368.4016
A133	3.1473302	0	128.59183	1481.1949	1468.5011	1334.357
A200	129.66641	128.59183	0	1539.6376	1406.9978	1322.3783
A201	1511.9609	1481.1949	1539.6376	0	38.445426	34.101373
A202	1492.5915	1468.5011	1406.9978	38.445426	0	22.3531
A203	1368.4016	1334.357	1322.3783	34.101373	22.3531	0
A300	162.96238	165.55418	288.13078	1031.4798	983.39797	869.56784
A400	23.805279	29.243584	130.90453	1389.1125	1344.9143	1226.5425
A402	189.84332	178.29832	285.13647	949.88116	907.50548	772.78371
A403	102.54409	94.090928	200.4145	1121.2673	1080.4098	941.66192
A500	1557.8632	1526.5974	1501.9592	21.57168	22.635719	39.971653
A600	20.407208	21.533466	59.318072	1521.0154	1460.1824	1341.6883
A601	75.989005	64.35259	143.1085	1202.7826	1154.7993	1014.5507
A602	77.716459	66.920551	142.63019	1200.1732	1149.2869	1011.0544
A700	119.28467	109.16119	102.06521	1199.5373	1108.0045	984.27926
A701	203.08577	172.64548	310.0973	933.69678	932.01195	767.86395
A702	147.01458	121.0808	241.6909	1037.4037	1023.055	863.60703
A703	658.6029	618.88032	735.37216	389.53096	392.58954	268.48231
A801	207.27403	180.01006	285.78919	928.43735	905.14962	748.13365
A802	341.63225	307.05946	423.67555	729.15043	719.32088	563.50988
A901	575.95382	529.98115	634.8106	487.33907	493.98869	347.59387
A902	1318.9087	1291.6189	1140.7864	140.11766	46.454868	52.858531
ASaline	266.50596	281.27888	311.77751	1921.9802	1916.5468	1847.3549

Table C1c Spectral Distances Codes used can be found at end of tables

NAME	A300	A400	A402	A403	A500	A600
A100	105.21537	293.16328	72.522814	133.80449	758.35092	416.32925
A111	561.10382	261.11805	604.46767	452.61583	1876.6459	150.21542
A121	45.506975	110.7415	8.4992227	18.200884	1014.5559	168.15825
A122	47.629634	117.15173	8.6180459	20.196562	1002.9698	176.73523
A130	164.15848	36.587605	162.8767	83.49217	1487.1682	23.452529
A131	115.58482	43.121454	99.716605	45.724834	1343.798	53.637248
A132	162.96238	23.805279	189.84332	102.54409	1557.8632	20.407208
A133	165.55418	29.243584	178.29832	94.090928	1526.5974	21.533466
A200	288.13078	130.90453	285.13647	200.4145	1501.9592	59.318072
A201	1031.4798	1389.1125	949.88116	1121.2673	21.57168	1521.0154
A202	983.39797	1344.9143	907.50548	1080.4098	22.635719	1460.1824
A203	869.56784	1226.5425	772.78371	941.66192	39.971653	1341.6883
A300	0	78.476507	20.948136	26.9075	1087.5859	167.01602
A400	78.476507	0	106.5963	46.928155	1429.028	27.821244
A402	20.948136	106.5963	0	15.651004	1001.9584	180.83529
A403	26.9075	46.928155	15.651004	0	1173.7816	94.944751
A500	1087.5859	1429.028	1001.9584	1173.7816	0	1545.3392
A600	167.01602	27.821244	180.83529	94.944751	1545.3392	0
A601	58.306698	39.827797	41.374933	9.6883868	1244.2521	59.773823
A602	54.502791	38.620597	39.334615	8.825237	1240.3362	60.6958
A700	84.840446	66.303024	61.556508	34.166175	1213.7445	74.669376
A701	108.505	161.16065	48.827614	54.862922	992.2878	206.29288
A702	94.867778	113.41338	46.479259	36.044861	1089.1329	146.44062
A703	306.94075	552.52297	216.78529	318.65033	439.19859	659.79697
A801	81.034101	147.36535	24.95308	36.480273	978.40939	193.87553
A802	131.40561	263.00828	56.130315	100.65178	781.79064	331.2162
A901	286.00421	486.74246	179.87746	262.2589	536.84362	562.44989
A902	828.60498	1159.3043	740.47891	900.63239	91.21611	1248.1377
ASaline	709.45291	379.53566	753.37193	588.20594	1948.5701	262.95033

Table C1d Spectral Distances Codes used can be found at end of tables

NAME	**A601**	**A602**	**A700**	**A701**	**A702**	**A703**
A100	194.22777	192.2691	237.62677	119.45625	151.20757	124.35787
A111	375.39259	378.40694	378.17405	607.44589	513.14686	1179.7078
A121	34.055086	32.764581	45.224414	30.83539	27.355879	213.01042
A122	37.318023	36.276187	51.67212	27.300108	26.094583	203.01241
A130	51.445801	54.600999	90.833512	140.03178	93.066178	570.5969
A131	26.730141	29.617906	68.15673	71.391877	39.896104	432.59222
A132	75.989005	77.716459	119.28467	203.08577	147.01458	658.6029
A133	64.35259	66.920551	109.16119	172.64548	121.0808	618.88032
A200	143.1085	142.63019	102.06521	310.0973	241.6909	735.37216
A201	1202.7826	1200.1732	1199.5373	933.69678	1037.4037	389.53096
A202	1154.7993	1149.2869	1108.0045	932.01195	1023.055	392.58954
A203	1014.5507	1011.0544	984.27926	767.86395	863.60703	268.48231
A300	58.306698	54.502791	84.840446	108.505	94.867778	306.94075
A400	39.827797	38.620597	66.303024	161.16065	113.41338	552.52297
A402	41.374933	39.334615	61.556508	48.827614	46.479259	216.78529
A403	9.6883868	8.825237	34.166175	54.862922	36.044861	318.65033
A500	1244.2521	1240.3362	1213.7445	992.2878	1089.1329	439.19859
A600	59.773823	60.6958	74.669376	206.29288	146.44062	659.79697
A601	0	0.1864575	22.155852	60.835127	34.023601	368.42488
A602	0.1864575	0	20.422778	63.25093	36.172963	368.19016
A700	22.155852	20.422778	0	90.403808	59.223384	378.76023
A701	60.835127	63.25093	90.403808	0	6.0686484	173.3264
A702	34.023601	36.172963	59.223384	6.0686484	0	233.65293
A703	368.42488	368.19016	378.76023	173.3264	233.65293	0
A801	45.511349	46.473979	67.571875	10.168577	14.356365	177.39116
A802	124.32322	125.21396	145.77764	28.872732	54.713936	81.628349
A901	293.67196	295.65247	313.42889	123.25471	176.68381	23.657871
A902	957.22749	950.91045	883.82956	778.74061	854.07714	319.89553
ASaline	513.72033	518.84518	537.93441	730.3806	640.27457	1312.0651

Table C1e Spectral Distances

NAME	A801	A802	A901	A902	ASaline
A100	98.019924	69.139531	122.79164	598.11683	1015.0891
A111	608.40807	800.90857	1060.2469	1636.6158	43.985036
A121	11.072363	43.62086	167.99267	750.32548	723.18083
A122	8.9941018	38.198427	158.12695	746.00503	733.20126
A130	149.74333	268.32751	478.60053	1247.055	302.19687
A131	81.588586	172.04318	353.71414	1111.3795	421.93062
A132	207.27403	341.63225	575.95382	1318.9087	266.50596
A133	180.01006	307.05946	529.98115	1291.6189	281.27888
A200	285.78919	423.67555	634.8106	1140.7864	311.77751
A201	928.43735	729.15043	487.33907	140.11766	1921.9802
A202	905.14962	719.32088	493.98869	46.454868	1916.5468
A203	748.13365	563.50988	347.59387	52.858531	1847.3549
A300	81.034101	131.40561	286.00421	828.60498	709.45291
A400	147.36535	263.00828	486.74246	1159.3043	379.53566
A402	24.95308	56.130315	179.87746	740.47891	753.37193
A403	36.480273	100.65178	262.2589	900.63239	588.20594
A500	978.40939	781.79064	536.84362	91.21611	1948.5701
A600	193.87553	331.2162	562.44989	1248.1377	262.95033
A601	45.511349	124.32322	293.67196	957.22749	513.72033
A602	46.473979	125.21396	295.65247	950.91045	518.84518
A700	67.571875	145.77764	313.42889	883.82956	537.93441
A701	10.168577	28.872732	123.25471	778.74061	730.3806
A702	14.356365	54.713936	176.68381	854.07714	640.27457
A703	177.39116	81.628349	23.657871	319.89553	1312.0651
A801	0	21.17552	120.8495	735.08162	745.31634
A802	21.17552	0	44.984153	576.70404	936.87694
A901	120.8495	44.984153	0	390.32486	1187.6601
A902	735.08162	576.70404	390.32486	0	1795.3268
ASaline	745.31634	936.87694	1187.6601	1795.3268	0

Table C2 Codes Used in Spectral Distances Tables

Code	Name of Bacteria
A100	*Bacillus subtilis*
A111	*Escherichia coli*
A121	*Lactobacillus casei 202* (replicate 1)
A122	*Lactobacillus casei 202* (replicate 2)
A130	*Pseudomonas fluorescens* (replicate 1)
A131	*Pseudomonas fluorescens* (replicate 2)
A132	*Pseudomonas fluorescens* (replicate 3)
A133	*Pseudomonas fluorescens* (replicate 4)
A200	*Lactococcus lactis* ssp. *cremoris* (replicate 1)
A201	*Lactococcus lactis* ssp. *cremoris* (replicate 2)
A202	*Lactococcus lactis* ssp. *cremoris* (replicate 3)
A203	*Lactococcus lactis* ssp. *cremoris* (replicate 4)
A300	*Leuconostoc*
A400	*Pediococcus pentosaceus* (replicate 1)
A402	*Pediococcus pentosaceus* (replicate 2)
A403	*Pediococcus pentosaceus* (replicate 3)
A500	*Escherichia coli*
A600	*Pediococcus acidiladici* (replicate 1)
A601	*Pediococcus acidiladici* (replicate 2)
A602	*Pediococcus acidiladici* (replicate 3)
A700	*Lactococcus lactis ssp lactis* (replicate 1)
A701	*Lactococcus lactis ssp lactis* (replicate 2)
A702	*Lactococcus lactis ssp lactis* (replicate 3)
A703	*Lactococcus lactis ssp lactis* (replicate 4)
A801	*Lactobacillus casei 201* (replicate 1)
A802	*Lactobacillus casei 201* (replicate 2)
A901	*Lactobacillus helveticus* (replicate 1)
A902	*Lactobacillus helveticus* (replicate 2)
ASaline	Background spectrum