Criteria for Drug Identification by Thin Layer Chromatography and Near Infrared Reflectance Spectroscopy

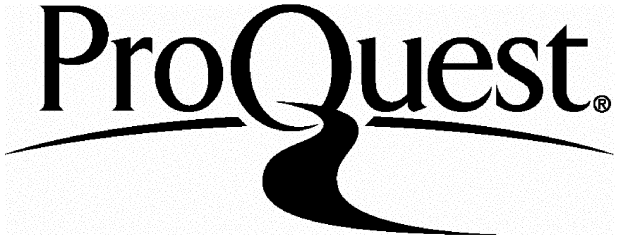
Pervaise R. Khan

A thesis submitted in partial fulfilment of the requirements of
The University of London for the degree of
Doctor of Philosophy
in the Faculty of Medicine

The School of Pharmacy,
University of London

October 2003

Supervisors: Prof A. C. Moffat and Dr R. D.Jee



Abstract

A study was conducted of drug identification by thin layer chromatography using computer search techniques. This investigated the criteria for drug identification as used by pharmacopoeias, and various organisational bodies. The effect of varying standard control conditions, the use of standardising reference standards and running conditions were investigated. Computer based search methods incorporating fixed and moving windows, mean list length and discrepancy index were then used to identify drugs based upon R_f data. A probability of identification and the ability to rank drugs in order of likeliness of identification was produced from the R_f data. Achieving the current pharmacopoeial criteria and the feasibility of setting rigid criteria is discussed. Near Infrared (NIR) spectroscopy is a rapid non-destructive method of analysis which requires little or no sample preparation. The rapid, non-invasive identification of pure drug substances and tablets from a database of over 300 drugs by NIR is described, making use of simple chemometrics. An investigation of sample presentation and physical effects upon NIR spectra is presented. Novel drug profiling studies are described that differentiate between different tablet manufacturers demonstrate identification based upon NIR data. Methods include the analysis of pure drugs and tablets by NIR, the setting up and validation of the database and identification of unknowns against the database. The data were analysed and the methods optimised using a number of chemometric procedures such as second derivatives, correlation spectral matching, wavelength distances, polar co-ordinates, centres of gravity and scanning probability windows.

Identification by NIR spectroscopy is demonstrated for unknown pure drugs and tablets with varied success for the type of substance, tablet and chemometric method used. A low number of mis-identifications were reported. A ranking method of identification based upon spectral match is incorporated with optimisation and criteria to be used in the general method.

Table of Contents

Index of Tables

Index of Figures

Index of Equations

Acknowledgement

I would like to express my gratitude to my supervisors Dr Roger Jee and Professor Tony Moffat for initially giving me the opportunity to undertake this study and for their direction, help and guidance throughout. Their support throughout the writing up process has really made this submission possible.

I would also like to thank Dr Robert Watt for his continued encouragement and counselling throughout, it has been very much appreciated.

I would also like to sincerely thank my laboratory partner and friend Dr Robert Waters for his support and friendship.

I would like to acknowledge the Royal Pharmaceutical Society of Great Britain for the PhD studentship that enabled this work to take place. I would like to thank Foss NIR Systems for the loan of the 6500 instrument and also Sheelagh Halsey for her support.

I would also like to thank the staff of the School of Pharmacy for their help and in particular to Colin James and to Derek Marley, the master of the School's drug register. I would like to say thank you to Andy Jackman who guided many through our initial degrees and onto further study.

I would also just like to add my appreciation to my family in India and England for their support. Finally, I would like say a huge thanks to Mum, Dad and my sister Shama for everything and it is to them that I dedicate this work.

Chapter 1 Introduction

1.1 Thesis layout

It has been attempted to keep the layout of the thesis as simple and logical as possible to aid the reader in accessing information as quickly as possible. Cross referencing between sections has been used to eliminate the use of duplicate information.

This chapter contains the background and overall introduction to the work in this thesis. It is divided into the two analytical techniques i.e. Thin Layer Chromatography (TLC) and Near Infrared (NIR) reflectance spectroscopy. It is then broken down further to relate to the specific topics of each chapter.

Chapter 2 contains the methods employed in each part of the practical work, the sources of substances used and suppliers etc. This chapter is again divided by analytical technique and describes what was actually carried out in each subsequent chapter. All practical details are described here.

Chapters 3 to 7 contain a summary of the aims of the chapters, the experimental results, discussion of results and conclusions drawn. The Appendices contain details of all the drug substances used and are followed by a full list of literature references.

1.2 Background to the thesis topics

The work in this thesis seeks to focus on the relatively simple, novel but quantifiable identification of compounds of pharmaceutical interest. The main focus is on NIR spectroscopy, however, before progressing onto the NIR work the fundamentals of the criteria for drug identification are explored. This is based on TLC, an analytical technique that is tried and trusted and the cornerstone of forensic analysis identification. TLC also has the advantages of being relatively inexpensive and straightforward to set up. It is apparent, upon close examination of TLC, that identification is often based upon subjective criteria and the robustness of TLC systems is open to question. This presents the question as to whether rigid criteria would need to be established for NIR identification or whether the existing TLC criteria need to be strengthened. Whereas TLC is well known and established, NIR spectroscopy as an analytical technique is still relatively untried and in many cases un-trusted. It is for this reason and because of the constant argument of regulators that the technique and its chemometrics are too often presented as a "black box wonder" and not clearly that explained this thesis focuses heavily on NIR spectroscopy. Also for this reason, the detail to the background, theory and introduction sections for NIR spectroscopy is more comprehensive than for TLC.

1.3 Aims and Objectives

Upon examination of the literature (referenced in section 1.4), especially the pharmacopoeias, some questions arise. For example, standard TLC conditions are stated (temperature x °C, run distance y cm, dry the sample spots, saturate chambers with solvent etc) but there is no impact in the case that these standards are not rigorously followed. What is the quantifiable effect on the results of varying temperature by y °C? For example, would a laboratory test set up in an African climate be expected to achieve the same results as a test in a British climate? Are there some factors that must be rigidly controlled and others that need not? This is examined with the aim of quantifying the effect of varying standard control conditions and suggesting guidelines to improve the repeatability of TLC testing. The criteria for the identification of substances by TLC in a number of pharmacopoeias is often wordy, open to interpretation and involves subjective comparisons.

The objective is to determine whether it is feasible to set rigid and quantifiable criteria for the identification of substances based upon TLC analysis.

The investigation involves the use of novel computer based search techniques to identify test substances from a database of published TLC data.

The aims of the work are to examine the robustness of TLC systems, to take them to their limits of performance, to quantify the effects of operating conditions and to recommend optimum conditions.

This work is contained within Chapter 3.

NIR spectroscopy is, in analytical terms, a technique that has much potential for the future of drug identification.

The umbrella of techniques covered by NIR spectroscopy has developed significantly over the past two decades. However, a large percentage of the work undertaken has been focussed on quantitative analysis of materials already known to be present in a sample[1,2]. In fact, algorithms capable of accounting for quantifications down to 0.05% active detection limits[3] have been demonstrated. The chemometrics are also becoming increasingly complex. However, it is important that key assumptions are not taken for granted. Firstly with the rate of advancement of spectral processing and data pre-processing, the effect upon the original information may be overlooked. For example, consideration must be given to spectral shift, loss/ masking of peak information and the introduction of artefacts. Another issue to be addressed is the basis of understanding by the regulatory bodies, who have stated previously that they have no objection to the implementation of NIR spectroscopy as an in house release test method if properly validated[4]. However, suspicion remains within the regulatory bodies that processing advancements of the technique may be being made at such a rate and to such levels of complexity that fundamental analysis of the data may be being overlooked. An example of this occurs when justifying the use of wavelength selective chemometric techniques.

The aim of the initial work with NIR is to determine the factors related to sample presentation that could influence the peaks in a spectrum. This is similar in manner to the TLC work, in order to recommend the factors that must be kept constant when considering identification. This is contained within Chapter 4.

Where the identification work differs from that in the literature is that it seeks firstly to set up a large database of drug actives (approximately 300 actives) which has not been done before. It shows the difficulties of achieving such a task, of testing the actives internally and externally and makes recommendations as to protocols to be followed and problems that may be encountered.

The ultimate aim is to identify powdered drug actives (from a variety of sources) against this database, using test substances with actives in the database and some without.

The vision behind this work is the ability to place an unknown drug active on to a spectrometer, to produce fast analysis (within approximately 10 seconds) and to search a large database of spectral data to produce an identification with a predicted level of assurance. The method should also ensure that a drug that is not contained within the database is not identified.

The initial series of experiments were to examine how accurate the matching of compounds would be if there were a series of substances of closely related structure. The purpose of this was as a small pre-test for the large database of drugs. This was to determine where issues of identification might arise and to decide upon which chemometric method would be the most suitable to carry forward onto the larger database. It would also dictate whether it would be necessary to construct two large databases going forward (i.e. one for spectra and one for second derivatives) or just one. From this work the criteria for a successful identification would be sought.

From the key published work (section 1.8) it can also be seen that many researchers have attempted to identify the active substance within tablets. A

few of these have used the tablet intact and have compared this non-destructive NIR analysis of the tablet active with conventional Pharmacopoeial assay results. None, however, have attempted to identify the active within intact tablets from a large database of drug actives. The aim is to demonstrate this with tablets from multiple batches and manufacturers (Chapter 7). Additionally, none of the key published work indicates that it has been possible to identify the manufacturer from which a batch of tablets has been derived. It is attempted to show this in Chapter 7.

The majority of the papers published also include some fairly complicated chemometric methods, which have caused much debate among the regulatory authorities. There is little evidence of attempts to base identification upon simple peak wavelength position (as is common in mid-infrared analysis). The work in Chapter 6 sets out to address this.

1.4 Background to thin layer chromatography

TLC is a widely used technique for the identification and separation of drugs in their pure form in pharmaceutical preparations and in biological fluids. There were different reasons for this choice. TLC is a technique that is commonly used and well understood, it is widely used in forensic analysis, is a recognised pharmacopoeial method for identification and can generate a sizeable amount of data with relatively little time or cost incurred. In addition, R_f data can be produced from samples analysed across a number of different solvent systems. TLC also produces reproducible systems with a well defined set of influencing variables. A large published database of R_f values across different solvent systems is available (Stead et al[9]) and will be used as the reference database to which the generated R_f values are to be compared.

Criteria for identifying drugs by TLC are used in many different contexts such as pharmacopoeias, for regulatory matters and forensic analysis. When investigating TLC and the criteria upon which identification is based, it is interesting to note that these are based upon methods relying, in many cases upon observation of spot size, colour and intensity and with usually comparing these and the R_f s obtained with known reference standards. The R_f values that are used throughout the thesis are percentage R_f values (as found in Stead et al[9]).

Taking the pharmacopoeial context as an example, Table 1, wording such as "similar to" or "corresponds to" are used in the British[5] and United States[6] Pharmacopoeias to identify compounds where as the European Pharmacopoeia (EP)[7] looks for spots that are "similar" and that have the "same" colour and fluorescence with the EP conferring identification upon spots that move "identical distances." Rarely is it possible to suggest how unique a particular identification appears or to put a statistical value on the reliability of identification. The application of these and other such guidelines also provided the scope for scientific analysis of such data.

Table 1 Examples of the wording for criteria of drug identification in some common Pharmacopoeias.

"the principal spot...is *similar* in size and position to"	British Pharmacopoeia 2002[5]
"the value of the main spot... *corresponds* to that obtained"	US Pharmacopeia 2003[6]
"the spot...is *similar* in position, fluorescence and size" "The spot... is of the *same* colour"	European Pharmacopoeia 2003[7]
"if the two substances move *identical* distances" "*correspond* in position, size and intensity"	International Pharmacopoeia 2001[8]

This part of the thesis concentrates initially on setting up a functional TLC system based upon an established database created by Stead et al[9]. This was followed by optimisation of the system and the demonstration of reproducibility. The effect of varying standard control conditions was examined, as was the effect of introducing standard reference compounds as described by Moffat et al[10]. From this, the database published by Stead et al[9] was recreated electronically. This database was used as the reference point for comparison of various statistical tools and computer methods that were developed to provide probabilities of identification of unknown substances.

1.4.1 Theory

The technique of TLC may be one or two dimensional, depending on the complexity of the analysis. In one dimensional chromatography, a drop of the solution containing the mixture of substances is placed on the chromatographic plate stationary phase. The spot is allowed to dry and the TLC plate placed into a suitable solvent which passes through the spot, so dissolving the substances, and carries them along while moving up the stationary phase. Different substances move with different rates which depend on their size, nature of the substances, solvent and diffusion coefficient. The relative distance moved of solute and solvent is represented by the Retardation factor (R_f) value. This is defined for any component in a mixture as the ratio of the distance travelled by the component to the distance travelled by the solvent. Components are identified by their R_f values, which are specific to the solvent used. Individual components will therefore possess a series of R_f values over a number of different solvent systems.

Solvent selection is dependent upon the nature of the solutes to be separated, such that when the components are mixed together, separated solute spots should result. Most stationary phases (eg silica and alumina) are adsorbents, the separation achieved is by interaction between the drug and the surface of the stationery phase. Other stationery phases, such as cellulose, interact by nature of partition processes between the mobile and stationary phases. Where the stationary phase is more polar than the mobile phase the system is called a normal phase system, wheras where the mobile phase is more polar than the stationery phase, this is called a reverse phased system. The substances are identified by application of various spray reagents, by size, colour, intensity or by illumination under ultra violet light.

1.5 Background to NIR spectroscopy

The NIR region is generally referred to as that portion of the electromagnetic spectrum between 750 nm and 2500 nm. Until the early 1980's, the majority of spectroscopists had avoided this region due to the difficulty in interpreting the absorptions. This was due to the absence of well defined peaks (as commonly peaks overlapped one another) and difficulties in measuring peaks that were not set upon a constant baseline. The presence of numerous combination bands also made assignment very difficult.

However, it was a few pioneers from the early 1900's who laid the foundations upon which the technique is based today. Kubelka and Munk's original paper[11] described the phenomenon of diffuse reflectance and its potential for analysis of samples with the simplest of handling. This is due to the shorter wavelengths

in the NIR region being more beneficial to diffuse reflectance despite the fact that transmission techniques were more favoured at the time. The good reflectance-absorbance ratio possible in the NIR, for solid state systems in particular, leads to more linear and reproducible calibrations.

Other early work by Ellis and Bath[12] focussed on the properties of the high O-H bond absorption to determine the water content of gelatin. Around this time workers in the field concentrated largely upon single wavelength analysis and Kaye[13] produced a review of the main work to that year (1954) which included examples of qualitative work attempted at single wavelengths. It was not until 1968, however, when Ben-Gera and Norris[14,15] applied multi-component regression with multiple wavelengths that the foundation for correlation based practical work had been laid. The work in the analysis of agriculture products involved moisture/ fat determination in soya beans and milk, and laid the foundation for more sophisticated data handling methods to follow. An important factor in the development of the technique in the last twenty years has been the advance in sensitive, precise instrumentation that is now commonly available from a variety of manufacturers. Norris and his co-workers found adequate instrumentation a problem in the 1960s and resorted to designing their own. Allied to the instrumentation (as an integral part), are fast modern day computers capable of handling vast amounts of data in fractions of a second.

It is widely recognised that NIR spectroscopy offers several advantages over traditional analytical methods. It is fast, non-destructive, non-invasive, low in running costs and can be applied through optic fibres. As with any method there are drawbacks, these being the need for a reference method, or construction of

a learning/ calibration set, relative sensitivity for solid-state systems and the need for fast computers to analyse the data - a property leading to regulatory perceptions of a "black box" technique.

In recent years NIR spectroscopy has become an important analytical tool in the pharmaceutical industry. It is being used extensively in the food, agriculture and textile industries[16,17]. Although the actual spectroscopy is not new, the applications and development of chemometrics are new and innovative.

1.6 Fundamentals of NIR spectroscopy

Spectra arise from the emission or absorption of definite quanta of radiation when transitions occur between certain energy levels. The energy of a molecule can be considered to be made up from the collective sum of the electronic, vibrational, rotational energies etc according to the Born-Oppenheimer approximation, Equation 1.

$$E_{total} = E_{elc.} + E_{vib.} + E_{rot.} + \ldots$$

Equation 1 Energy of a molecule following the Born-Oppenheimer approximation.

E_{total} is the total sum of individual energy approximations. E_{elc}, E_{vib} and E_{rot} are the approximated electronic, vibration and rotational energies respectively.

When a molecule undergoes a transition from a lower energy state E_2 to a higher energy state E_1 absorption of energy hv occurs, Figure 1.

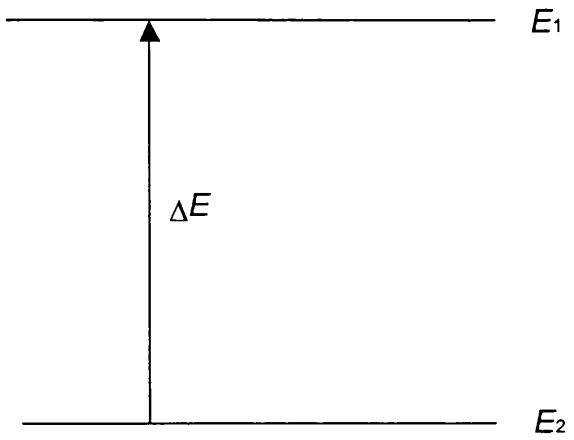


Figure 1 Example of the type of change a molecule can undergo from a lower energy state (E_2) to a higher energy state (E_1).

ΔE is the difference in energy. See also Equations 1 and 2.

$$\Delta E = h\nu = E_1 - E_2$$

$$c = \lambda\nu \qquad \lambda = \frac{c}{\nu} \qquad \tilde{\nu} = \frac{1}{\lambda}$$

Equation 2 Energy transition states

ν is the frequency of the radiation absorbed and h is Planck's constant. The position of a transition is commonly expressed in terms of the wavelength (λ) or wavenumber ($\tilde{\nu}$) of the radiation and where c is the velocity of light.

The intensity of a transition depends upon the population of the energy levels and the strength of the interaction of molecules with the electromagnetic field. For a sample of N molecules at temperature T, the number with energy E is given by the Boltzmann distribution.

$$N_E \propto N g e^{-E/kT}$$

Equation 3 The Boltzmann Distribution

where N is the number of molecules at temperature T, g is the degeneracy of the level (i.e. the number of states corresponding to that level) and k is the Boltzmann constant.

For absorption of radiation to occur, the lower level (E_2) must be populated. The greater the population the more intense the absorption. The absorption of NIR radiation is associated with changes in the vibrational plus rotational energy of molecules. Hence for the molecule to interact with the radiation the electric dipole moment of the molecule must change when the atoms are displaced. Vibrations in which there is no dipole moment change will be NIR inactive.

1.6.1 Molecular vibrations

A typical potential energy curve for a diatomic molecule is shown in Figure 2 (solid line). There is a repulsion between the positively charged nuclei of both atoms, and between the respective negative electron "clouds". In addition to this there is an attraction between the nucleus of each atom and the electrons of the other. The two atoms settle at a mean inter-nuclear distance such that these forces are just balanced and the energy of the whole system is at a minimum. If the bond length is stretched too far the bond will break.

The simplest model for the vibration of a diatomic molecule is to assume that the potential energy curve can be approximated to a parabola (Figure 2, dashed line), i.e. Hooke's law applies and the bond extension/ compression is proportional to the force applied.

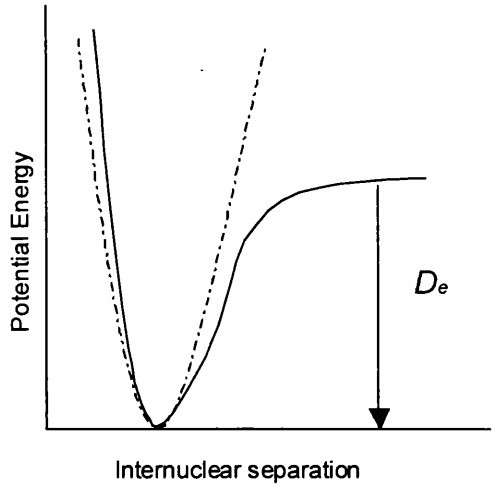


Figure 2 Potential energy curve for a diatomic molecule

D_e is the delta in energy

Figure 2 illustrates the potential energy curve for a simple diatomic molecule. The dotted line represents harmonic potential, obeying Hooke's Law for an ideal harmonic oscillator. This is shown in Equation 4. The solid line represents the extension of bond stretch and where the vibrational energy level reaches the dissociation energy (bond breakage). The allowed evergy levels *E* for an anharmonic oscillator are given in Equation 10

Classically the frequency of vibration, v_0, of a *harmonic oscillator* is given by Hooke's Law:

$$v_o = \frac{1}{2\pi}\sqrt{\frac{k}{\mu}}$$

Equation 4 The frequency of vibration of a harmonic oscillator (v_o)

where k is the classical force constant and μ is the reduced mass of the 2 atoms, ν is the vibrational frequency

For a diatomic molecule the reduced mass is given by:

$$\frac{1}{\mu} = \frac{1}{m_1} + \frac{1}{m_2}$$

Equation 5 The reduced mass of a diatomic molecule (μ)

where μ is the reduced mass and m_1 and m_2 are the masses of the atoms

The potential energy V of the oscillator as a function of the displacement, r, is represented by the parabolic equation:

$$V = \tfrac{1}{2}k(r-r_e)^2$$

Equation 6 The potential energy (V) of the oscillator as a function of the displacement (r)

where r_e is the equilibrium distance, r is the displacement and k is the classical force constant

The quantized vibrational energies of vibration (E_{vib}) are given by the following formula from quantum mechanics:

$$E_{vib} = (v + \tfrac{1}{2})hv_o$$

Equation 7 The quantized vibrational energies of vibration (E_{vib})

where v is the vibrational quantum number and can take the values 0, 1, 2, 3. This represents a series of equally spaced energy levels in the molecule as described in Figure 3. h is Planck's constant. *V_o is the frequency of absorption. When v=0, E≠0.*

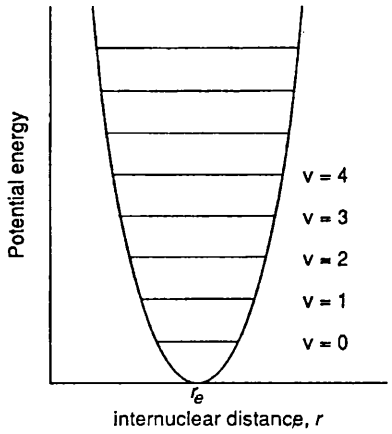


Figure 3 Harmonic Oscillator

where v is the vibrational quantum number. This represents a series of equally spaced energy levels in the molecule.

For a harmonic oscillator the selection rule is $\Delta v = \pm 1$. The allowed absorption transitions are therefore:

$$E_{1\leftarrow 0} = (1 + \tfrac{1}{2})h\nu_0 - \tfrac{1}{2}h\nu_0 = h\nu_0$$

$$E_{2\leftarrow 1} = (2 + \tfrac{1}{2})h\nu_0 - (1 + \tfrac{1}{2})h\nu_0 = h\nu_0$$

$$E_{3\leftarrow 2} = (3 + \tfrac{1}{2})h\nu_0 - (2 + \tfrac{1}{2})h\nu_0 = h\nu_0$$

etc.

Equation 8 Examples of allowed absorption transitions for a harmonic oscillator

h is Planck's constant, *v* is the vibrational quantum number and *E* is the energy level

The predicted vibrational spectrum will consist of a single line of energy $h\nu_0$. At room temperature the majority of the molecules will be in the $v = 0$ state and therefore only the transition $v_{1\leftarrow 0}$ will be important. For a typical molecule the values of k and μ are such that the fundamental frequency of vibration will occur in the mid-IR region of the electromagnetic spectrum. No transitions would be seen in the NIR region.

In reality, molecules do not behave exactly like simple harmonic oscillators. As r increases the chemical bond becomes weaker and eventually dissociates (breaks). A more realistic potential energy curve is that shown by the solid line in Figure 2. A good mathematical approximation to this *anharmonic oscillator* is the Morse potential energy curve:

$$V = D_e\left[1 - e^{-a(r-r_e)}\right]^2$$

Equation 9 The Morse Potential Energy Curve

where D_e is the depth of the potential minimum (Figure 2) and a is a constant related to various molecular parameters.

When the Schrödinger equation is solved for the Morse curve the permitted energy levels are given by:

$$E_{vib} = (v + ½)h\nu_e - (v + ½)^2 x_e h\nu_e$$

Equation 10 Permitted Energy levels when the Schrödinger Equation is solved for the Morse Curve

where x_e is called the *anharmonicity constant.*

The energies of the transitions $\nu_{1\leftarrow 0}$, $\nu_{2\leftarrow 1}$, $\nu_{3\leftarrow 2}$, *etc.* will now be slightly different from one another, however, as noted above only the $\nu_{1\leftarrow 0}$ transition will be important at room temperature. The selection rule for the anharmonic oscillator is $\Delta v = \pm 1\ \pm 2\ \pm 3\ldots$ and consequently transitions such as $\nu_{2\leftarrow 0}$, $\nu_{3\leftarrow 0}$, *etc.* are now possible as in Figure 4.

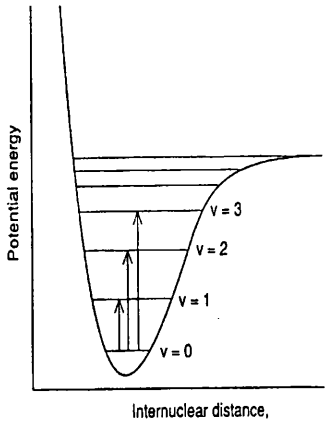


Figure 4 Fundamental, first and second overtones for an anharmonic oscillator

where v is the vibrational quantum number.

For the fundamental 1st and 2nd overtones this is calculated below:-

For $\Delta v = +1$.

$$E_{1\leftarrow 0} = [(1 + \tfrac{1}{2})hv_e - (1 + \tfrac{1}{2})^2 x_e hv_e] - [\tfrac{1}{2}hv_e - (\tfrac{1}{2})^2 x_e hv_e]$$

$$= (1 - 2x_e)hv_e$$

For $\Delta v = +2$

$$E_{2\leftarrow 0} = [(2 + \tfrac{1}{2})hv_e - (2 + \tfrac{1}{2})^2 x_e hv_e] - [\tfrac{1}{2}hv_e - (\tfrac{1}{2})^2 x_e hv_e]$$

$$= (1 - 3x_e)2hv_e$$

For $\Delta v = +3$

$$E_{3\leftarrow 0} = [(3 + \tfrac{1}{2})hv_e - (3 + \tfrac{1}{2})^2 x_e hv_e] - [\tfrac{1}{2}hv_e - (\tfrac{1}{2})^2 x_e hv_e]$$

$$= (1 - 4x_e)3hv_e$$

The anharmonicity constant is typically small ($x_e \approx 0.01$), hence the frequency of these transitions are close to v_e, $2v_e$ and $3v_e$. The line near v_e is called the *fundamental absorption,* while those near $2v_e$ and $3v_e$ are called *the first* and *second overtones,* respectively.

The energies of these overtones are such that for many molecules they lie in the NIR region of the electromagnetic spectrum. The probability of these spectroscopic transitions decreases rapidly with increasing Δv, consequently these absorptions are weak. Typically the first overtone is about 1/10th and the second overtone 1/100th the intensity of the fundamental absorption. The relationship between sample absorbance at specific wavelengths and analyte concentrations can be explained by Beer's law:

$A = €cl$

Equation 11 Beer's law

Where *A* is sample absorbance, *€* is the molar absorption coefficient, *c* is the absorber concentration and l is the path length

There is only one mode of vibration for a diatomic molecule, the bond stretch (stretching involves a change in bond length). In polyatomic molecules there are two types of vibrations namely stretching and bending (bending involves changes in bond angles). For a non-linear molecule that consists of *N* atoms, there are $3N - 6$ fundamental modes of vibration *(*$3N - 5$*,* for a linear molecule). Some vibrations/ bends will be associated with small groups of atoms, e.g. just one bond. Absorptions due to these vibrations in the mid-IR are very useful for identifying functional groups. Most of these vibrations occur at high energies because the reduced mass of the vibrating system will be small. This gives rise to what is known as the *functional group* region in the mid-IR region (4000 to 1300 cm^{-1}). Many vibrations will be associated with the whole structure and will be characteristic of that particular molecule only — this gives rise to what is termed *the fingerprint* region in the mid-IR region (1300 to 600 cm^{-1}). Polyatomic molecules may exhibit simultaneous changes in the energies of two or more vibrational modes: the frequency observed will be the sum of ($v_1 + v_2$, $2v_1 + v_2$ *etc.)* or the difference between ($v_1 - v_2$, $2v_1 - v_2$, *etc.)* the individual frequencies. (Where $2v_1$ represents the first overtone, subscripts refer to different fundamental vibrations). This results in very weak bands that are called *combination* and *subtraction* bands — the latter are possible but rarely observed in room temperature NIR spectra. The vibrational transitions observed will also be accompanied by changes in rotational energy of the molecule

making the absorption lines broad and appear as bands. For even quite small molecules the NIR spectrum will therefore be expected to be complex, though unlike the mid-IR much more difficult to assign bands.

1.7 Diffuse reflectance in the NIR region

Many of the algorithms that have been developed to achieve multi-component determination of a variety of samples have been based upon diffuse reflection. Of these the most accurate have had a dependence upon the linear relationship between band intensity and analyte concentration. The band intensity is described as log $1/R$ where R is the reflectance of the sample compared to a non-absorbing standard (e.g. a ceramic disc). A brief summary of some of the major diffuse reflectance theories is given here. Ciurczak has given a more expanded explanation of the mathematics behind these theories[18].

1.7.1 Kubelka - Munk theory

Kubelka and Munk[11] made the assumption that radiant energy is transferred from the incident NIR beam continuously and is converted to thermal vibrational energy of molecules. The decrease in intensity of the reflected light is related to the absorption coefficient of the reflectance material (i.e. the sample). The reflectance is described as being related to the absorption coefficient (K) and scattering coefficient (S) by

$$K/S = (1 - R)^2 / 2R = f(R)$$

Equation 12 Kubelka - Munk Function

where R, the diffuse reflectance is a fraction of K/S and is proportional to addition of absorbing species to the sample medium.

On this basis is the assumption that the diffuse reflectance of the incident beam is directly proportional to the quantity of the absorbing species that interacts with the incident beam. Therefore, for quantitative determinations the reflectance is dependant upon analyte concentration. Practically, in NIR terms a relative reflectance is measured rather than the absolute diffuse reflectance. The relative reflectance is equal to the ratio of intensity of radiation reflected from the sample/ intensity of radiation from a reference material.
The reference material used in the project is a ceramic disc. The absolute reflectance is described as:

$$I_s/I_o$$

Equation 13 The Absolute Reflectance

where I_s is the radiation intensity reflected from the sample and I_o is the intensity of the incident radiation

Generally it is accepted that the Kubelka - Munk equation is a limiting equation and should only be applied for weakly absorbing bands when absorptivity and concentration are low, due to vibrational overtones and combination bands. The absorptivities of these bands are much weaker than the respective fundamentals. Therefore, most analysis in the NIR region is considered weakly absorbing without dilution, however, many analytes are not isolated but

surrounded by a matrix of other components which are strongly absorbing of the incident radiation at analytical wavelengths. Therefore, unless a proper referencing method is used, the absorption matrix may cause deviations from the Kubelka/ Munk theory.

1.7.2 Lambert Cosine law

Lambert in 1760 sought to demonstrate that the reflected radiation from a perfect reflector (or as near as was physically possible to one) was of the same intensity regardless of the angle of observation or the angle of incidence[19].

$$\frac{dI_r / df}{d\omega} = \frac{CS_0}{\pi} \cos\alpha \cos\vartheta = B\cos\vartheta$$

Equation 14 Lambert Cosine law

where I_r is the remitted radiation flux, f is the surface area of the reflector in cm^2, ω is the solid angle in steradians (sr), α is the angle of incidence, ϑ is the angle of observation, S_0 is the irradiation intensity in W/ cm^2, C is a constant representing the fraction of the incident that is remitted (<1 as some radiation is always absorbed) and B is the surface brightness in W/ cm^2 sr.

There was some debate as to whether this was only possible where the reflector was in fact a black body radiator acting as the ideal diffuse reflector (as used by Lambert), otherwise the angle of distribution of the reflected/ remitted radiation was dependant upon the angle of incidence. It was, however, argued that such a black body could not be deemed an ideal diffuse reflector as it would absorb a significant amount of the incident radiation anyway. As the ideal diffuse reflector has not been found in practice there will be deviations to the law.

1.7.3 Mie Scattering

This theory was developed by Mie in the early 1900s and deals with the scattering of radiation light by isolated particles[19]. Mie described the scattering of plane polarised radiation by a particle that can be both non-conductive and absorbing. The particle was spherical and had no size limit. He showed that the distribution of such scattered radiation was not uniform in all directions.

$$\frac{I_{\vartheta s}}{I_0} = \frac{\lambda^2}{8\pi^2 R^2}(i_1 + i_2) \equiv q(\vartheta_s)$$

Equation 15 Mie Scattering

where $I_{\vartheta s}$ is the scattered intensity at a distance R from the sphere's centre, I_o is the intensity of the incident radiation, λ is the wavelength of the incident radiation and i_1, i_2 are functions of the angle of scattered radiation

Functions of the angle include harmonics or their derivatives with respect to the cosine of the angle of scattered radiation, the refractive index of the sphere and its surrounding medium and the ratio of particle circumference to wavelength. The basic equation (Equation 15) is for a non-conductive non-absorbing particle and for non-polarised incident radiation. If the particle does absorb this must be taken account of in i_1 and i_2. Mie theory encompasses spherical particles of any size but is only applicable to a single incidence of scattering. For example, scattering by a system with well separated molecules. It is applicable where the particle size is smaller than the incident wavelength of radiation.
However, in most NIR applications the scattering will only take place by molecules that are close together (e.g. solids) therefore multiple scattering will take place.

Scattering order was described by Theissing as the number of times a photon is scattered and described how, as the order of scattering increased, Mie scatter decreased and the angular distribution of scattered radiation tended to become isotropic[19]. The conclusion was that for a sufficiently thick sample and for a sufficiently large number of particles, multiple scattering will occur for the majority of samples in NIR reflectance analysis so that an isotropic distribution of radiation should be reached. There was no definition of sufficient number of particles or sufficient thickness, it is difficult therefore to describe the multiple scattering within a densely packed medium in relation to a change in analyte concentration.

1.8 NIR applications

Applications of NIR spectroscopy take many shapes and forms across a variety of industries. Of special interest for this project is the pharmaceutical industry including the regulatory aspects, for drug identification. Providing a rapid, non-invasive, non-destructive method for drug identification that is accurate, reproducible and robust would provide enormous benefits to industry. The general public at large would also benefit be it through quality control or forensic analysis. For this to be achieved, the power of the NIR technique for drug identification must not only be demonstrated, but must incorporate chemometrics that are unambiguous and intelligible to regulators and databases that are fully validated.

Prominent published work within the areas of interest is summarised below.

1.8.1 Regulatory work

Before the mid-1980s the majority of the NIR work that had been published was in the agriculture, textile and petrochemical industries. The pharmaceutical industry was slow to react and the regulatory authorities viewed NIR with more than a hint of suspicion.

The interest from the pharmaceutical industry has arrived from the realisation of the potential production of better products in shorter amounts of time and at lower cost. An increase in regulatory interest has arrived in manufacturing and production control, uniformity and the detection of deviations from approved formulations. For a method to be accepted by the regulatory bodies it must be validated. Official pharmaceutical applications are still few in number. However, Moffat et al[20] have produced a detailed study explaining how to meet the International Conference on Harmonisation's Guidelines on Validation of Analytical Procedures. The quantitative method determined the Paracetamol content of intact tablets and was compared against the British Pharmacopoeia ultraviolet assay. In addition, Laasonen et al[21] have recently developed and validated a method for determining the Caffeine content in intact tablets from a Finnish pharmaceutical product.

An assay entitled "Piperazine in Drugs" was published in the 1990 AOAA Official Methods of Analysis[22] and The United States Food and Drug Administration (FDA) has accepted NIR as the official method for determination of the lincomycin content in an agricultural premix of a soyabean meal.

The FDA have also approved the use of an NIR method in place of compendial methods for moisture content in the identification and assay of ampicillin

trihydrate[23]. The demonstrated advantages of this method compared to the previous method in the submission to the FDA, were faster and more accurate results, the lack of solvents and the production of no waste products. Blanco et al[24] have demonstrated validation of a method for Miokamycin in tablets. The relative standard error of prediction achieved was <1.5 %.

The Health Protection Branch (HPB), Canada's regulatory authority, has approved the use of a NIR method for raw material and packaging identification. The submission for the method used a database of 185 reference spectra of different raw materials and packaging components. In 1995 the European Pharmacopoeia took a compendial lead by approving a monograph[25] on the use of NIR spectroscopy for the identification of organic substances. The monograph deals with sample preparation, instrument control and the establishment of a reference library. European Health Authorities have also approved the identification and assay of Zovirax tablets by NIR[26]. The latest regulatory status within Europe (EMEA, 2003[27]) is that NIR spectroscopy can be used as an alternate method to one or more validated methods specified in the quality part of the submission dossier. However, the validated methods must also be included for comparison and the NIRS method must be challenged by the validated method at least once a year to ensure ongoing validity.

Of the other quantitative methods for approval, some have been published[28,29,30] while others have not been published due mainly to the proprietary nature of the pharmaceutical industry. Presently, if a manufacturer intends to introduce an alternative method to an existing method, validation approval must be gained. For validation to be achieved the error of the

established primary method must be well known and the accuracy, sensitivity, linearity, specificity, robustness and reproducibility of the method must be demonstrated[4]. The accuracy of the NIR data produced is determined with comparison to an approved analytical method. The method specificity can be determined through the use of the instrument software which can qualify the sample. Sample position is an important source of error, therefore the sample should be measured, removed and re-measured repetitively to minimise the error and determine reproducibility. The major pharmacopoeias allow for conformance testing to be carried out using alternative analytical methods but these must be approved. For approval to be granted, validation must be achieved proving that the method will reach the same conclusion as the official or approved method. An obstacle to NIR becoming officially approved is instrument transferability as no one instrument has been designated as the standard and a separate calibration model is required for each sample and instrument type. Satisfying in-house quality control specifications for consistency within one site has been demonstrated[31,32,33] but not adequately across sites. However, Smith et al[34] have compared several correction methods to facilitate the transfer of a validated assay for paracetamol in intact tablets across two instruments.

For replacement of an existing method with a high performance liquid chromatography method, there is little debate from the regulators concerning the actual technique. Whereas with NIR methods, an applicant must first submit substantial evidence on the technique. The advancement of a number of NIR methods to officially accepted status may take some time. This may seem ironic when compared against traditional pharmacopoeial mid-IR methods,

which involve comparison of two spectra often taken from different instruments held to the light and visually compared. The argument against the "black box" technique is still to be overcome from a regulatory perspective.

1.8.2 Pharmaceutical applications

1.8.2.1 Raw material and pure drug identification and qualification

The early 1980s saw a number of breakthrough publications on pharmaceutically related NIR applications and in chemometric approaches. Qualitative analysis by NIR reflectance was virtually unknown until Rose[35] announced findings at the annual Tarrytown symposium in 1982 and showed that a number of parenteral products could be analysed by NIRS. He described results distinguishing 40 raw materials by using the discriminant analysis SAS statistics package. This was based purely on selected wavelengths only at "representative" points in the spectrum. Shenk[36] used an approach called the HAT Matrix for the analysis of forage samples. He demonstrated that the samples forming the basis of his quantitative calibration were from the same set as the test samples to be quantified. Process and further non-invasive methods can be found in Workman's review[37].

Mark and Tunnel[38] used linear discriminant analysis and Mahalanobis distance techniques for the qualitative identification of materials. It was applied to raw materials by Ciurzcak[39]. Both methods are discriminant and require the selection of an optimal set of wavelengths. These were selected either visually by noting wavelengths of apparent discrimination while considering only a few compounds for analysis or by computer analysis when the number became much larger. Linear discriminant analysis classifies spectra of samples by

identifying a linear function that separates categories of spectra based on absorbance at the selected optimal wavelengths. The optimal wavelengths were considered in pairs and were used as X,Y co-ordinates. In the ideal situation, clusters of samples were formed for each particular material and these clusters were easily separable. This method however, could not take outliers into account. Mahalanobis distances were reported as being superior, classifying spectra based on the standardised distance of a spectrum from the central zone of a particular category cluster. Rules were set for classification of a cluster member and of separate cluster categories based on standard deviation distances in multi-dimensional space. The advantages of the technique are cited as speed because of the relatively few wavelengths involved in routine computations and the ability to perform qualitatively and quantitatively. However, the methods still rely heavily upon selected wavelengths and the justification for exclusion of data is questionable when valid chemical information is contained throughout the NIR region. This is another reason why regulators view the technique with apprehension and this must be addressed.

Gemperline et al [40] tested 400 spectra from six raw materials for identification using Soft Independent Modelling of Class Analogy (SIMCA). The acceptable/ unacceptable groupings were determined by *F*-Test probability, the method used the full NIR spectra and could detect outliers. However, physical characteristics did lead to an incorrect failure (due to particle size) and the construction of the training and the test set appeared intensive in time and selection albeit with a large number of spectra, incorporated only 6 different raw materials.

Shah and Gemperline[41] proposed a combination of two techniques of pattern recognition: initially calculating principal components of NIR spectra followed by calculating a Mahalanobis distance model in the subspace defined by the significant principal components. A separate class model was generated for each class of raw material. However, the classification only contained the same six raw materials as previously described. The combined methods were, however, capable of detecting foreign grades of material and low levels of contaminants, but noticeably did not incorporate long-term instrument variation.

Salamin et al[42] also identified difficulties in identification by NIRS due to physical variation. They therefore applied a multiplicative scatter correction calculation transformation before basing identification upon the Mahalanobis Distance method but used only three compounds and only five spectra per compound. The physical variability between samples due to particle size was therefore accounted for.

However, only three compounds were used and once more specific wavelengths were the basis for the identification.

Corti et al[46] investigated the qualitative analysis of 27 primary materials by NIR. Each powdered sample was represented by three spectra which were analysed by Product Identification by Discriminant Analysis (PIDAC). This was based upon an *n*-dimensional map set up from a number of wavelengths. Independent clusters were constructed from a training set within the n-dimensional map. Mahalanobis distances were then used as the basis of identification or non-identification. The study produced positive results including the suggestion of differentiation by NIR between amorphous forms and those containing a crystalline fraction. A general method substance identification method was

proposed but produced a mis-identification between dihydrostreptomycin and streptomycin sulphate. This was discounted as a serious limit to the technique based upon positive results from specifically constructed maps for the two drugs. The method again ignored large regions of the NIR spectra. Kohn and Jeger[43] constructed a library of 37 spectra from drugs and stretching agents of primarily forensic interest and utilised Mahalanobis distances for the identification tool with three to six wavelengths and with multiplicative scatter correction (MSC). The limitations of the distance approach have been highlighted already whereas the MSC approach was heavily calculation intensive, as pointed out by the authors. A third approach was demonstrated in the paper (principal components) and this did overcome the disadvantages of the first two methods but was mainly based on comparison of spectra.

First derivatives were used to reduce physical variation and results were improved. Identification between different carbohydrates was also demonstrated. A draw back of the technique was that several grams of each substance were required as sampling was by fibre optic probe. Gerhausser et al[44] have described strategies for setting up spectral libraries based upon original and second derivative spectra for a small number of substances.

Blanco et al[45] have also recently constructed a spectral library of 125 raw materials. The approach is proposed for the routine identification of a finite set of raw materials.

Corti et al[46,47] in 1991 and 1992 applied discriminant analysis to different qualitative problems and succeeded in achieving good mapping for the identification of compounds of pharmaceutical interest and of different types including organic, inorganic and organometallic compounds. The method also

attempted quantification of two antibiotics: streptomycin sulphate and cloxacillin benzathine, producing a good random distribution of the percentage errors within ±5 %.

Salamin et al[48] and Naes et al[49] enhanced the methodology of pure drug identification by the introduction of the Multiplicative Scatter Correction to reduce misidentification due to the spectral physical variation presented by drugs. Work on the potential for identification of polymorphic and optically active forms of caffeine and amino acids and differentiation of structurally similar compounds: theobromine, theophylline, caffeine, ephedrine and pseudoephedrine has been achieved by Buchanen et al[50].

Despite the amount of work published in this area very little has focused on producing a reliable method of drug identification for unknown substances. A method such as this could be applied to a forensic or an anti-counterfeiting approach (of considerable value to the pharmaceutical industry). However, it would need to be transparent enough to be understandable and acceptable by the regulatory authorities.

1.8.2.2 Identification of dosage forms

The discriminant analysis methodology was expanded to solid dosage forms in 1986 by Ciurczak and Maldacker[51] who attempted to classify ground tablet granulations of aspirin, caffeine and butalbital tablets in a "pass/ fail" manner. This introduced the idea of spectral subtraction of placebo samples (excipient only) from the final granulations to produce spectra of the active components. These were then compared against a library of actives. A separate algorithm based upon modified multiterm linear regression analysis to determine the

rapidly changing component within a complex matrix was described by Honigs et al[52]. Whitfield[53] continued with this approach leading to the identification of veterinary samples of lincomycin and within a certain potency. This led to the first submission to the FDA for NIR analysis of a drug substance in a dosage form for a new drug application (NDA). Lodder and colleagues[54,55,56,57,58] have examined the possibilities of drug identification in tablets and capsules and the potential of identifying adulterants and breakdown products in both. The work includes the use of a multivariate statistical model to define sub-populations in qualitative analysis. This was applied to the control of the presence of unexpected or unwanted inorganic residues in capsules. The capsules were fitted into the reflector and were analysed without opening. Spectral Processing was by a technique known as the Bootstrap Error Adjusted Single sample Technique (BEAST), a non clustering algorithm. Excellent results were reported with a linear response produced down to detection limits of 0.4% of capsule weight for potassium cyanide. The response was found to be dependant upon mass of the adulterant and position within the capsule, due to the presentation method of the sample. Principal Component Analysis (PCA) was also used to separate aspirin tablets with and without adulterant. The results were again good for a minimum aspirin content of 2.0 %. The work did, however, highlight the production of erroneous values (composition values), i.e. the generation of some values regardless of which material had caused the reflection. If a material was not present in the training/ learning set, but was present in the sample set, a composition value was produced but was incorrect (misidentified). The high dependency on the training set with the possibility for

errors of this type being produced is a disadvantage of multiple regression methods.

A piece of work on qualitative and quantitative control in the pharmaceutical sector, although on solutions, was carried out by Dubois et al[59]. They describe the control of the contents of solutions containing varying concentrations of phenazone, lidocaine, sodium thiosulphate, ethyl alcohol and glycerol. Very good qualitative results were obtained for phenazone, glycerol and ethyl alcohol but poor results were obtained for lidocaine and sodium thiosulphate. Multiple linear regression (MLR) was used as the data processing method and the work has been applied to the production instance. Chasseur[60] assayed cimetidine in powders, a calibration was performed again by multiple linear regression and the method was introduced for production control. The MLR helped to alleviate the physical effects of the sample on the spectral characteristics. Spectra were analysed in the second derivative mode giving preferential results compared to the first derivative. The percentage standard error reported with respect to the reference method, ultra violet (UV) spectrophotometry, was about 2.5%.

Jensen et al[61] studied the content qualitatively and quantitatively of amiodarone tablets but used a method involving crushing the tablets first. Precision and reproducibility were both good and the correlation between NIR and the reference method (refractrometry) was reportedly very good.

Molt and Egelkraut[62] studied aspirin, paracetamol and caffeine as a mixture in the solid state producing coefficients of variation ranging from 0.5 to 1.0%. However, they encountered problems related to particle size distribution and geometry producing spurious effects. Corti et al[63] have worked on a variety of pharmaceutical substances and forms and have produced good results when

the active species was greater than 1%. However, the ideal conditions were identified as having the active being greater than 5%. They also stressed the importance of including production calibration samples in the training set. Again methods and analytical software were specific for the three substances studied. A general method for the rapid verification of identity and content of solid dosage forms was devised by Ryan et al[64]. The authors also compared mid-IR and NIR for the analysis of tablets or capsules containing lovastatin, simvastatin, enalapril, finastride or placebo from 0.2 mg to 40 mg of drug. The minimum amount of drug detectable was 1% w/w by both methods. Differentiation by NIR was not possible between the closely related structures of simvastatin and lovastatin. As would be expected, mid-IR did well on structural elucidation of the compounds.

Dempster et al[65] developed a non-invasive test to confirm the identity of blister packed tablets for use in clinical trials. The work involved the examination of three different methods of sample presentation to the NIR instrument: unpackaged single tablets exposed to the NIR window, unopened blister packaged tablets placed on the NIR window and a fibre optic probe used to examine the tablets through the packaging. The chemometric methods used centred around construction of a library of second derivative spectra for each tablet type. Directly analysed tablets of 5%, 10%, and 20% w/w, placebo and comparator product were identified correctly when compared with reference techniques. 2% w/w tablets were indistinguishable from placebo. When analysed through the blister packaging, similar results were obtained with the 5% w/w now indistinguishable from placebo. The probe was more convenient but less sensitive. Broad et al[66] employed Fourier transform-NIR to quantify

ethanol, propylene glycol and water through amber plastic bottles also using a probe. Calibration models were set up using partial least squares regression and multiple linear regression.

MacDonald and Prebble[67] have discussed some applications that have been used internally by a leading pharmaceutical company. They include the use of NIR spectroscopy to determine moisture in freeze dried parenterals by non-invasive measurement through the base of the product vials. Excellent calibration results were produced and there were no significant differences for the moisture in vials compared against Karl Fischer titration at 95% t-test levels.

Blanco et al[68] determined the ascorbic acid content in pharmaceutical preparations by using stepwise multiple linear regression (SMLR) and partial least squares (PLS). The method involved grinding the tablets and comparing the results to the published titrimetric method. The results provided prediction errors of 1-2%. Another successfully developed tablet method is by Aldridge et al[69] who developed a quantitative method for tablet analysis within blister packs. This involved construction of a factor based library model from analysis of 165 tablets as a learning set. Other samples from the learning set were then used to validate the model which was then used to conformance check thousands of clinical trial sample tablets. Time and labour savings against the traditional TLC method were estimated.

Dempster et al[70] have also demonstrated confirmation of the identification of blister packed, film coated and non-film coated tablets for clinical trial supplies. Of the three methods compared, wavelength distances and Residual Variance gave better performance with Mahalanobis distances producing some incorrect results. The conclusion was that the methods involving spectral libraries built

from selected portions of the second derivative spectra gave better performance than using full range spectra. However, all tablets were 'known' to the training set and all had been packaged with the same materials and under the same conditions, no unknowns were introduced. Van Zyl and Louw[71] investigated the viability of using PCA and SIMCA for an actual case of illicit methaqualone tablets taken from the South African market. The limited results presented indicated that the NIRS coupled with PCA and SIMCA could successfully be employed for rapid differentiation between illicit methaqualone containing tablet formulations. However, the different types of tablets examined varied in terms of physical appearance, colour, shape, pattern, imprint and it is unclear whether these differences could have accounted for many of the spectral differences highlighted anyway. Scafi et al[72] have also developed method against counterfeits. The method uses multivariate modelling and principal components/SIMCA to demonstrate that NIR can be used to detect counterfeit products.

A group from SmithKline Beecham Pharmaceuticals followed SB216469-S tablets through the production process taking samples at different stages. Han et al[73] sampled during process moisture analysis right through to the end of process tablet identification in blister packs. Second derivative spectra were used with PLS factors incorporated to build up the calibration equation. Good correlations were obtained between NIR data and those obtained by UV, loss on drying, HPLC and weight gain. The most significant results claimed were the identification of tablets with actives as low as 6.0%, 3.0% and 1.5% through the blister plastic.

Gottfries et al[74] have attempted to assess the active substance within metoprolol tablets while comparing transmission and diffuse reflectance instruments. Interestingly enough, each particular NIR technique had particular strengths. The results showed that despite the transmission mode possessing a narrower wavelength range available, predictions were much better for models based on transmission spectra than for models based upon diffuse reflectance spectra. In the reflectance mode the spectrometry is more susceptible to the sensitivity of inhomogeneity of material - less severe in the transmission mode. However, it is precisely this type of sensitivity that enhances the differentiation between physical properties that is an advantage of diffuse reflectance.

A recent example of rapid intact tablet analysis is a quantitative assay of paracetamol tablets by Trafford et al[75], which gave results comparable to the British Pharmacopoeial UV assay. Eustaquio et al[76] have also recently examined intact paracetamol tablets for content using transmittance combined with multivariate data processing methods.

1.9 Instrumentation

The relatively small changes in absorption produced in the NIR region with respect to concentration and the lower extinction coefficients due to combination and overtone NIR bands have placed the greatest restrictions on the development of NIR instrumentation in terms of stability and noise.

By the 1940's, as the interest in spectrophotometry grew, commercial instrument manufacturers responded with instrumentation capable of taking measurements in the ultraviolet, visible, NIR, mid and far infrared regions. However, the available NIR instruments at the time were optimised for work in the ultraviolet and visible regions with adaptations for NIR. The result being

unsatisfactory in terms of noise. In the late 1960s Karl Norris and workers at the US department of agriculture designed their own spectrophotometer, making use of correlation algorithms to account for the natural spectral variation produced from ground wheat and soya bean. The first commercially available unit by Dickey-John used a tungsten halogen source, six research grade interference filters and lead sulphide detectors placed at 45° to each other. Long term wavelength stability and the ruggedness of the instrument were improved by sealed optics, free from dust and internal temperature control. Another improvement was the implementation of the integrating sphere giving both reference readings and signal integration for each sample measurement. A detailed account of NIR instrumentation history is given by Butler[77].

The two most common instrument types are reflectance and transmittance. Reflectance measurements penetrate between 1-4 mm of the surface of the sample producing greater variation in non-homogenous samples than transmittance. In the transmittance mode, the entire path length is measured, so reducing the error due to non-homogenous formulations. However, any fine particles at the front surface may lead to an increase in scatter and a resulting loss in transmitted signal and lower signal to noise as a result.

1.9.1 The spectrophotometer

Diffuse reflectance

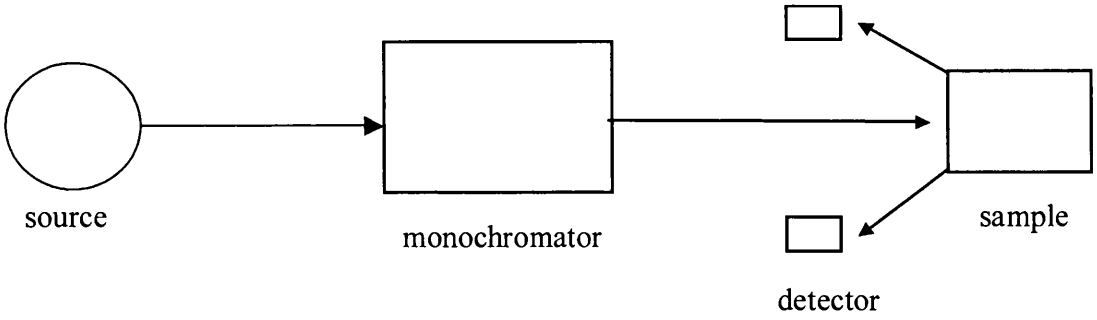


Figure 5 Schematic of basic spectrophotometer instrument configuration.

Light is initiated from the source and transferred to the sample as monochromatic light, via the monochromator.

The instrument used for the work in this thesis was a diffuse reflectance model 6500 spectrophotometer from Perstorp (now Foss) NIRSystems with attached Rapid Content Analyser module. The spectrophotometer is a single beam pre-dispersive diffuse reflectance instrument, a schematic representation of which is shown in Figure 5.The geometry of the system provides for monochromatic light to travel from the source in the monochromator unit to a holographic diffraction grating via a folded mirror through the order sorter filter to strike the sample on the Rapid Content Analyser window at 90° to the surface where upon the reflected light is collected by six un-cooled lead sulphide detectors set at a 45° angle. The analogue signal from the detectors is sent into a low noise, auto gain log amplifier whereupon it is converted into a digital signal which is exported from the instrument to a microcomputer for data processing, calibration and storage. This is repeated for reference and sample materials. The reflected light is measured and stored directly as log values. The ratio

between the sample and reference is plotted as the absorbance values shown on a raw spectrum.

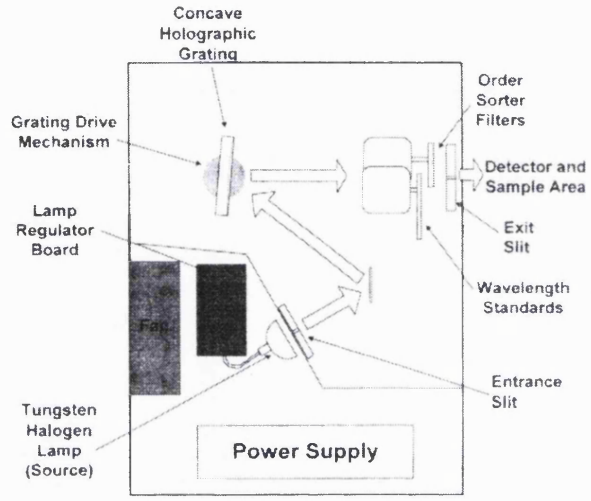


Figure 6 Schematic diagram of the FOSS NIRS 6500 diffuse reflectance spectrophotometer[80].

Light travels from the source via the entrance split to a mirror and onto the grating which is situated on a dive mechanism. The light is dispersed into various wavelengths prior to striking the sample. The order sorter filter alternates between light striking the sample and comparison to a wavelngth standard.

1.9.1.1 The source

The source of energy in NIR spectroscopy is a broad band tungsten halogen lamp the intensity of which is closely regulated by a precise power supply. Rapid scanning allows for the computer averaging of spectra to improve the signal to noise ratio. Additionally the frequent recording of a reference spectrum reduces possible errors due to the components changing with time.

1.9.1.2 The monochromator

The instrument obtains spectral resolution via an oscillating concave holographic grating. The grating consists of a reflecting surface onto which very fine lines have been laser scribed. The incident light enters through a slit/ lens combination and is dispersed as a function of wavelength. Rotation of the grating causes the spatially dispersed radiation to pass the exit slit with only a narrow band of wavelengths passing through the slit at any given time. For each angle of the grating, a different wavelength is produced, rotating the grating produces a set of discreet wavelength bands. The wavelength coverage is regulated by the angle of rotation of the grating.

1.9.1.3 Detectors

The instrument operating at different wavelength ranges utliise different detectors. From 400 to 1100 nm (emanating from the UV/Visible region) the detectors are made from silica and act as photovoltaic cells, from 1100-2500 nm the detection equipment consists of lead sulphide semi conductors set within a high voltage circuit. The voltage is maintained to high precision such that resistance change across the circuit, similar in type to a Wheatstone bridge, is in the micro-ohm range. At zero absorbance the noise is estimated at approximately 10 micro Absorbance units.

A problem with dispersive monochromators is that, in addition to the production of dispersed monochromatic light at a specified wavelength, several spectral orders are also produced. For example, when the grating angle is set to

produce light at 1400 nm, it will also produce it at 350, 700 and 2800 nm. To prevent these spectral orders from reaching the sample, an order sorter filter is used. The filter is a series of band pass filters that move in harmony with the grating allowing a certain region of the spectrum (the first order) for the grating angle to pass unhindered while ensuring all other spectral orders are absorbed.

1.9.1.4 Software

The instrument is controlled via the NIR Systems NSAS software, version 3.50 incorporating IQ^2 version 1.20. IQ^2 is the algorithm component of the software describing identification, qualification and quantification. The mathematical transformations within NSAS used to treat the spectra are described below along with reasons for their use and points of awareness.

Occasionally the spectra need to be treated to make them more amenable to analyses. This is where spectral roughness is produced as a result of sample texture surface effects. Where these were to be discounted, for example when examining a specific chemical property, the digital points making up the spectrum would be treated to produce a smoother looking spectrum.

N-Point Smoothing[78,79] is the simplest of these transformations and can be employed by selecting a segment of 2-10 nm and averaging the points within this range to produce a single point to plot representing this range. This is continued by repetition across the full spectral range to produce a more precise data scan. Since the peaks produced in the raw scan are broad, short term shifts in data points are considered as noise i.e. are not likely to affect the overall character of the spectrum. Point to point differences in absorbance

values are artificial in nature and only diminish the ability of an algorithm to produce a reasonable analysis.

1.10 Chemometrics

1.10.1 Derivatives

Upon viewing NIR reflectance spectra (Figure 7 Spectra of powdered allopurinol between the spectral range of 1100 nm to 2500 nm.) it is apparent that there is a gradual increase in baseline drift across the spectral range. Also, peaks are not sharp or well defined and there is much overlapping of absorption bands. Despite this, the spectra are complex and much physical information can be gleaned from such plots. In general, however, many effects such as these need to be reduced or eliminated in order to proceed with drug identification. These effects can be largely removed by the use of derivatives.

The basis of the technique is that since virtually all peaks in NIR are broad, short term shifts in data point values may be considered noise[80]. Derivatives must used with care, however. O'Haver[81] discusses the advantages and disadvantages of derivative spectroscopy methods and points out the dangers of higher order derivatives with respect to degrading the signal-to-noise ratio. O'Haver and Green's[82] assessment of the utility of derivative spectrometry determined the systemic and random errors of measurement with respect to 0^{th}, 1^{st} and 2^{nd} derivatives for a number of band pairs. While increasing the order of derivative increased the random error, the systematic error (which was larger) decreased and derivatives were found to be useful for reducing band overlap

errors in quantitative analysis. Gill et al[83] in using derivative UV visible spectroscopy also note that while sharp spectral features are emphasized, broader bands are discriminated against. Fell et al[84,85] also provided findings that are favourable with the second derivative and indicate that noise levels in using the fourth derivative are a limiting factor to using such a high order derivative. The second derivative has been found most commonly to provide a

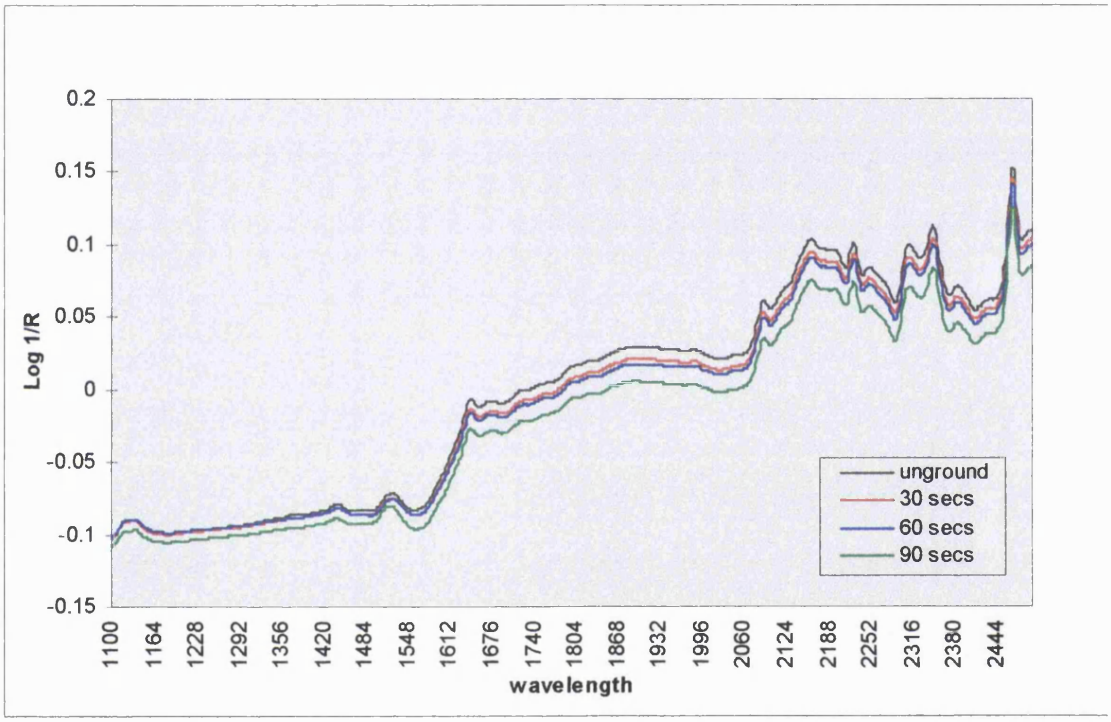


Figure 7 Spectra of powdered allopurinol between the spectral range of 1100 nm to 2500 nm.

compromise between enhancing the advantages while offsetting the disadvantages of derivative spectroscopy.

The advantages of using the second derivative are illustrated in Figure 8A, which shows a spectral absorption peak with (1) horizontal background, (2) linearly sloping background and (3) a curved background plus zero offset.Taking the first derivative with respect to wavelength removes any offset and reduces a linearly sloping background to a simple offset (Figure 8B).

Taking a further derivative (i.e. second derivative) removes the offset and the spectra are now almost identical (Figure 8C).

The positions of the negative peaks in the second derivative spectrum correspond to the positions of the peaks in the original spectrum. Derivative spectra can be calculated in different ways. The simplest way is to calculate the difference between blocks of data points on either side of the wavelength at which the derivative is required.

Absorbance (*A*)

$\frac{dA}{d\lambda}$

$\frac{d^2A}{d\lambda^2}$

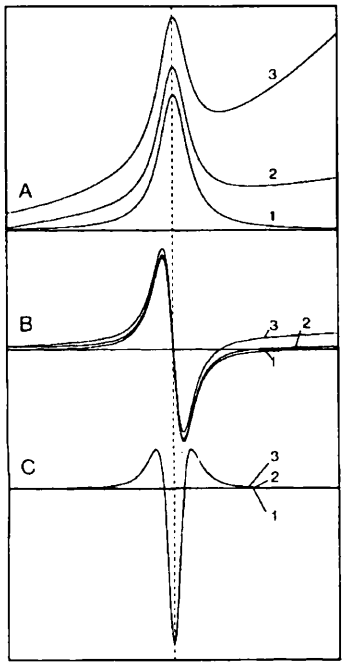


Figure 8 Diagram to illustrate the change from NIR reflectance spectra to second derivative spectra

A- Original spectra, B- First derivative spectra, C- Second derivative spectra.

This point is replaced by the difference between the two blocks.

This would be calculated for all such blocks and data points, moving along one data point at a time. Second derivatives are calculated as the second derivative of the absorbance function at each individual wavelength or $\frac{d^2A}{d\lambda^2}$.

Two adjacent first derivatives are subtracted one from the other to determine the change of slope between these two points. This value is now plotted against the wavelength to produce an inverted maxima with shoulders and inflections formerly not visible but now also present (Figure 8C). When calculating the derivatives of data in this manner the gap and segment sizes should always be quoted, as second derivatives produced with different gap and segment sizes are not directly comparable. Obtaining higher order derivatives are also possible, however, in practice are rarely used because of the tendency to enhance noise and produce peaks in areas where there is no spectral reason for their presence.

Another method of calculating the second derivative is that of Savitsky and Golay[86]. The method is based upon fitting a certain order polynomial to the data by least squares. The order of the polynomial is specified and is applied by using a number of data points before and after the the point where the derivative is required. The estimated derivative is then the derivative of the resulting fitted polynomial. This process is then repeated across the spectrum moving one data point at a time, applying a least squares polynomial fit to each and then calculating the derivative. Fitting all the least squares fits required would be very time consuming, however, providing the datapoints are all equally spaced this can be achieved very efficiently by using a Savitsky-Golay filter. Three to ten datapoints with a polynomial fitting of order two to four are typical parameters used. An illustration of this *moving window average* is shown in Figure 9.

An example definition of a Savitsky-Golay filter in use would be "4 4 4 2". This indicates four points taken to the left of the data point, four points taken to the right of the data point, application of a fourth order polynomial and application of the second derivative.

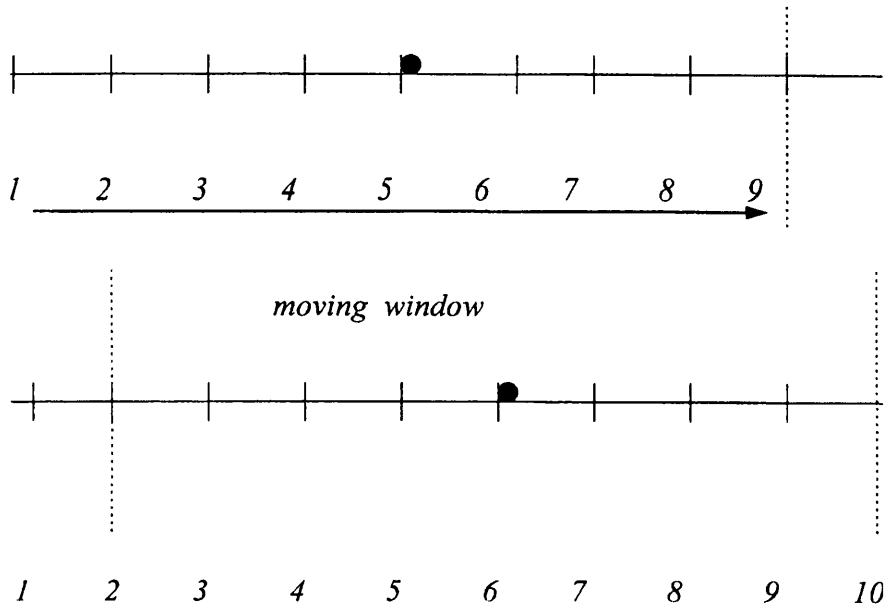


Figure 9 Moving Window Averaging.

The centre point is represented by the shaded dot. Four points to the left and right of the centre point are selected in this example and are fitted with a higher order polynomial. This example would be 4 4 4 2. The function at the centre point is evaluated and the window moves along to the next data point; the procedure is repeated.

These parameters must be applied with care. When higher orders are used, any smaller peaks may be lost as peak width increases and peak intensities change. The Savitsky-Golay method is used in Chapter 6 (Identification of drug substances by peak matching). Care was taken to examine whether an effect was exerted on the six peak identification such that where two peaks are close in intensity (pre application of the method) the order may have changed. For example, the most intense peak may become the second most intense peak after the filter has been applied. As the number of points taken either side of

the data point is increased, resultant peaks are broader and the overall number of peaks is decreased. This is considered in Chapter 6 (section 6.3.1).

1.10.2 Correlation spectral match

Identification of a compound by correlation spectral matching is based upon comparing spectral data from an unknown compound against those that are known and have been used to construct a reference library of spectra. The sample and the reference are compared by calculating a correlation coefficient between the two. If the two spectra possessed different absorbance values ($-\log_{10}R$) over selected wavelengths but were to rise and fall in synchronisation, plotting the amplitude of both spectra at these wavelengths would give a straight line correlation. A numerical measure of this correlation is achieved by calculating either the dot product (r_d) or product moment correlation (r_p) (Equation 16).

$$r_d = \frac{\sum S_i R_i}{\sqrt{\sum S_i^2 \sum R_i^2}}$$

$$r_p = \frac{\sum (S_i - \overline{S})(R_i - \overline{R})}{\sqrt{\sum (S_i - \overline{S})^2 \sum (R_i - \overline{R})^2}}$$

Equation 16 Correlation spectral matching values

r_d is the dot product correlation where S_i and R_i are ordinate values of the two spectra at wavelength *i*. r_p is the product moment correlation and includes $\overline{S}$ and $\overline{R}$ which are the mean values across the wavelength range

An *r* value of 1 indicates a perfect match, but really in practice is unachievable due to noise. During the course of the experiments, rounded values to 1 may be seen, however.

The two correlation coefficients produce very similar results with second derivatives as the mean value is usually close to zero. When the original ($-\log_{10}R$) spectra are being used r_p may give better discrimination. However a problem with the spectra as mentioned earlier is that as the absorbance values tend to increase as the wavelength increases, r values tend to be close to 1, which makes discrimination very difficult. The critical values for r are investigated during the internal and external library searches in Chapter 5. Additional information on correlation coefficients can be found in Mark[87] and the IQ^2 document[80].

1.10.3 Maximum wavelength distance match

The principle behind this method of matching is to determine whether a sample spectrum falls within an 'acceptance band' centred around the mean reference spectrum. The Student's t test is applied at each wavelength to determine the likelihood that the mean and test spectra are within the same acceptance band. This is based on the Distance algorithm (Equation 17) which calculates the greatest distances between two sets of spectra at each wavelength. The distance is actually measured in standard deviations (sd), with the match of lowest standard deviation suggested as the most likely product as shown in (Equation 17). Further details of how to calculate the method can be found in Gemperline and Boyer[88]

$$\bar{x}_i = \frac{1}{n}\sum_{j=1}^{n} x_{ij} \qquad s_i = \left[\frac{1}{n-1}\sum_{j=1}^{n}\left(x_{ij} - \bar{x}_i\right)^2\right]^{\frac{1}{2}}$$

$$t = \max\left[\left(\frac{y_i - \bar{x}_i}{s_i}\right)\left\{\frac{n}{n+1}\right\}^{\frac{1}{2}}\right]_{all\,i}$$

Equation 17 Wavelength distance matching

y is the test sample spectrum, x is the reference, $\bar{x}$ is the mean and s_i the standard deviation of each of the reference spectra (n =the number of spectra), i is each wavelength to be calculated. t is the test, calculated by the student t-statistic

The matching method therefore is really a t test with two samples. Critical values of t are dependant upon the number of reference spectra as well as the number of wavelengths used.

1.10.4 Principal Components Analysis

When considering the geometry of a set of samples measured at two wavelengths only, the values would be plotted against two axes. As principal components are concerned only with variation in data, all reflectance energies will be described as variations from their mean. Considering the original two axes, rotating them 90° would maximise the variation along one of the new axis. The first new axes would be described as the first principal component and the second axis the second principal component. The second axis would be at a right angle to the first axis and both axes would be linear combinations of the original axes. When reflectance values are measured at several wavelengths, each sample can be taken as a single point in a higher number of dimensions. Each dimension corresponds to an axes or wavelength at which the sample is

measured. Principal components analysis in the case of more than two dimensions involves a rigid rotation of the original axis so that there is maximum variation along the first new axis or principal component. The second component is the axis at right angles to the first along which there is maximum residual variation. Subsequent components are defined as the axes at right angles to the preceding principal components that exhibit the maximum amount of remaining or 'unexplained' variation in the data. Consequently, principal components are uncorrelated variables and each principal component has less variation than any of its predecessors. When using a monochromator, the spectra are recorded at 700 wavelengths. Therefore each sample is measured on 700 variables or axes and can be considered as a single point in 700 dimensions. A principal component is a linear combination between the original reflectance values from different wavelengths and so can not be measured directly and so must be derived. Mathematically, principal components are the eigenvectors of the covariance matrix (*S*) of the reflectance energies and the variances along each of the component axes are the eigen vectors of *S*. For a set of *n* sample spectra, E_{ij} can be used to denote the deviation of the *i*th sample reflectance energy from the mean reflectance energy at the *j*th wavelength such that :

$$S = \frac{1}{(n-1)E'E}$$

Equation 18 Principal components as eigen vectors of the covariance *S*

Where *S* is the covariance matrix, *E* is the relectance energy and n is the number of samples. *E'E* as a combination is used here to decrease the size of the possible matrix of 700x700 axes and instead use an axes combination compared for *n* samples

Further information on principal components can be found in Mark[89] and Cowe[106].

1.10.5 Polar Qualification System (PQS)

The polar qualification system is a discriminative method based on mathematical treatment of NIR spectra. It has been employed by Van der Vlies et al[90] to successfully distinguish two bulk pharmaceutical substances from different manufacturing sources. It is used in this thesis in Chapter 7 to investigate the differences in tablets. The method is explained here as there are very few clear descriptions in the literature. In the polar qualification method an entire spectrum is represented by a single 'quality point' or 'centre of gravity'. The quality points calculated by Van der Vlies et al[90] were shown to be a powerful indicator of physical and chemical properties.

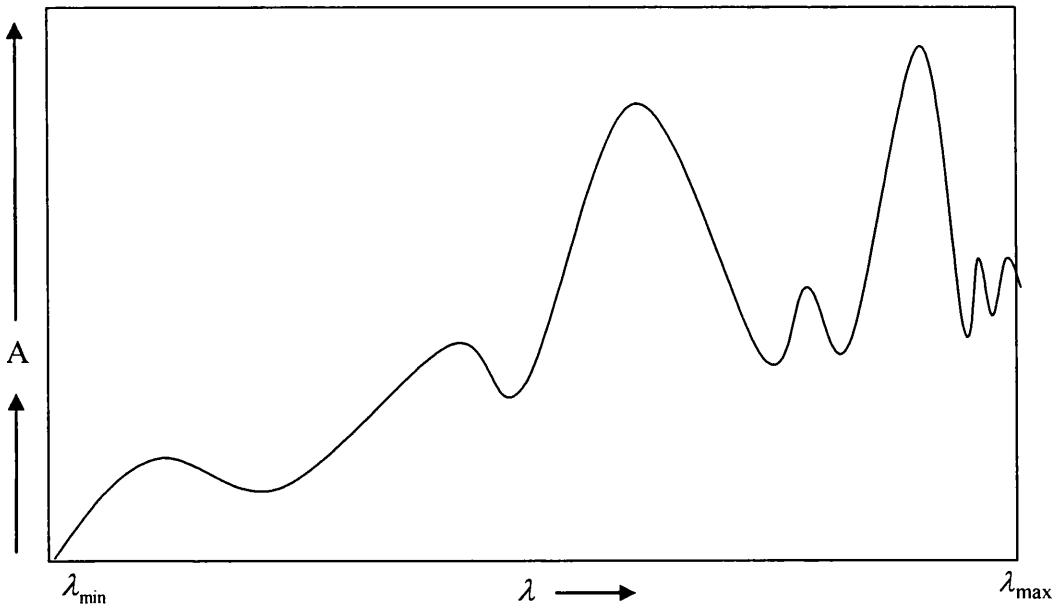


Figure 10 Original spectrum or pre-treated spectrum,

A is the absorbance log (1/*R*). λ is the wavelength range.

The first step towards producing a quality point is achieved by taking a spectrum (e.g.Figure 10) and plotting each data point around a polar axis as illustrated in Figure 11. In the polar co-ordinate plot, the wavelength range (λ_{min} to λ_{max}) is represented between 0 and 2π radians. The absorbance (or derived quantity) becomes the distance from the origin.

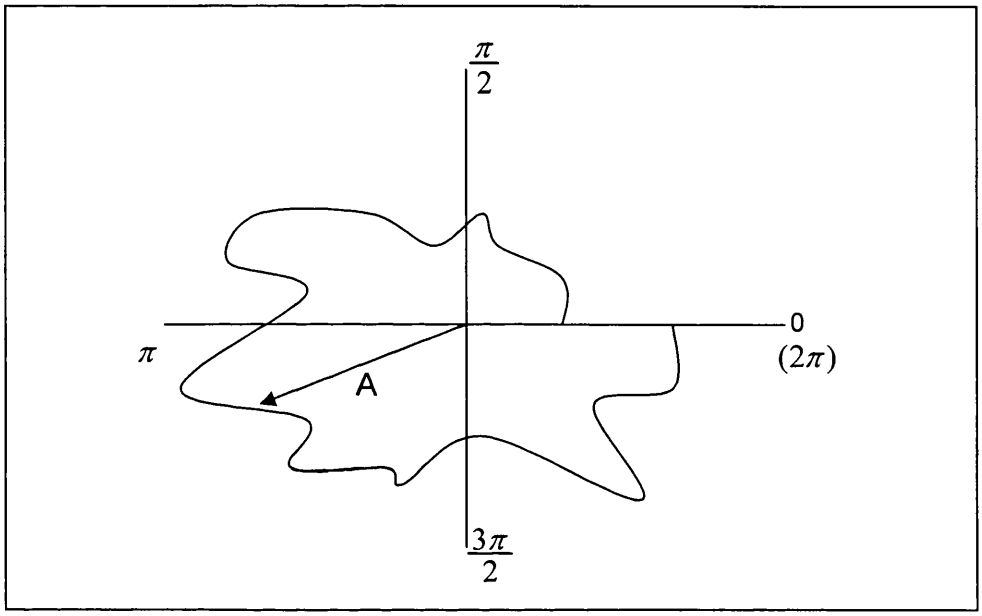


Figure 11 A polar co-ordinate plot

The spectrum is plotted with the wavelength range represented between the angles 0 to 2π radians. The absorbance (A) is the distance from the origin.

Figure 12 shows the relationship between polar and Cartesian co-ordinates. If the spectrum consists of *N* data points (wavelength points 1 to *N*), then θ for the *n*th data point is given by :

$$\theta = \frac{2\pi(n-1)}{N-1}$$

Equation 19 Definition of the angle as a function of wavelength

In terms of Cartesian co-ordinates with Point *P* in Figure 12,

x = A cos θ and y = A sin θ.

As PQS is often applied to second derivative spectra, it is preferable to use absolute values of A to prevent positive and negative peaks from cancelling one another out.

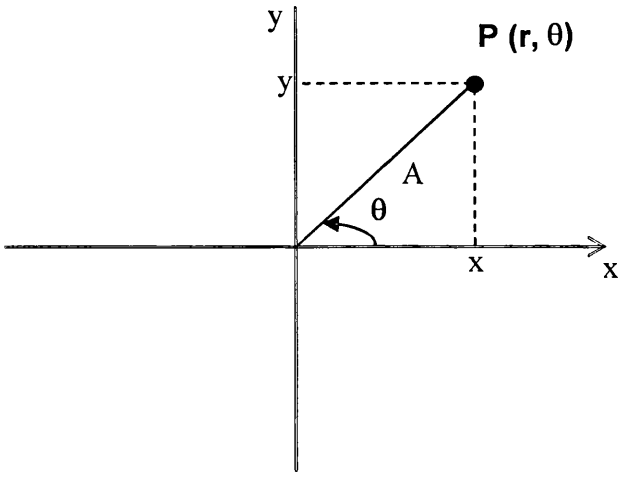


Figure 12 Relationship between polar and Cartesian co-ordinates

The length A is the absorbance, r and θ are functions of the wavelength, and the co-ordinates for the point P are given by x = A cos θ and y = A sin θ.

Hence,

$$x = |A| \cos\theta \quad \text{or} \quad x = |A| \cos\frac{2\pi(n-1)}{N-1}$$

$$y = |A| \sin\theta \quad \text{or} \quad y = |A| \sin\frac{2\pi(n-1)}{N-1}$$

Equation 20 Cartesian co-ordinates for a data point on a polar co-ordinate plot

Where x and y are the co-ordinates, A is the absorbance, *n* is the *nth* datapoint and *N* is the overall number of datapoints

To calculate the Cartesian co-ordinates for the centre of gravity of the polar co-ordinate plot, the mean value of x and y over all datapoints are calculated.

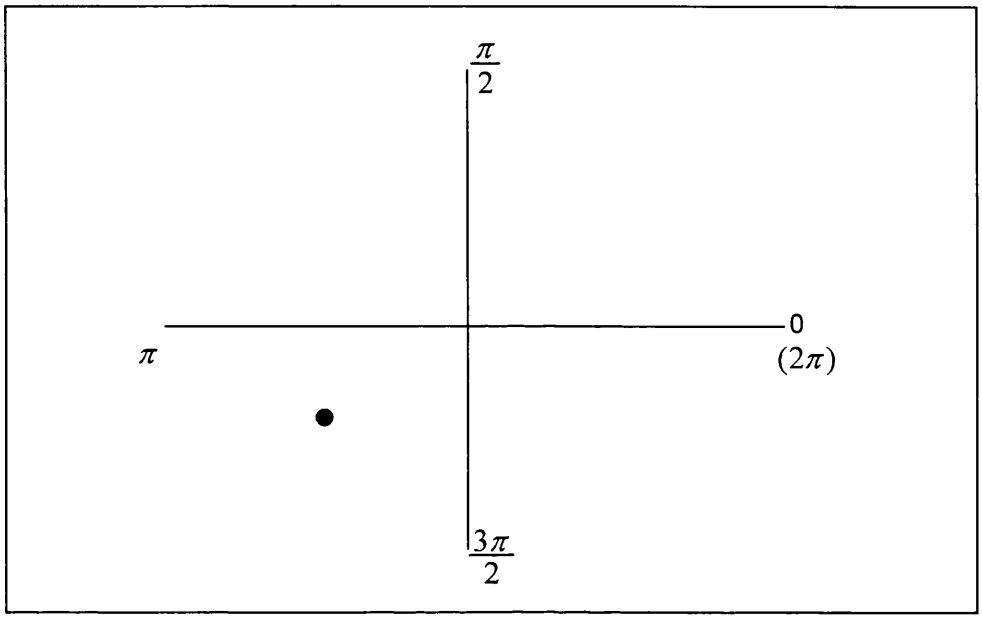


Figure 13 Centre of gravity plot

The centre of gravity of a polar co-ordinate plot is calculated from the mean x and y values from all datapoints

Hence,

$$X_{cg} = \frac{1}{N}\sum_{n=1}^{n=N} x_n$$

$$Y_{cg} = \frac{1}{N}\sum_{n=1}^{n=N} y_n$$

Equation 21 Co-ordinates for the centre of gravity or 'quality point'

Where X_{cg} and Y_{cg} are the co-ordinates for the centre of gravity, *n* is the *nth* datapoint, *N* is the overall number of datapoints, *x* is the x value from

Figure 12 for each data point and *y* is the corresponding y value

When repeat measurements are carried out (either on the same sample or on different samples), the resulting centre of gravity points will generally show a scatter of points, as in Figure 14.

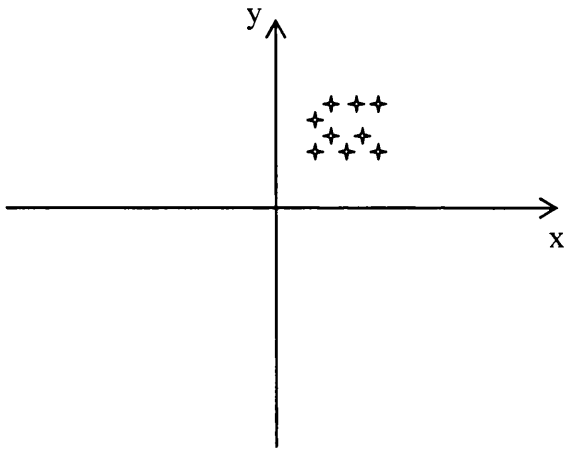


Figure 14 Illustration of centre of gravity points from different samples

The plot represents a common example of repeat measurements taken from samples of a single batch

Samples measured from different batches are likely to give different clusters of points. The mean centre of gravity for each cluster can also be calculated. Hence,

$$\overline{X}_{cg} = \frac{1}{m}\sum_{i=1}^{i=m} x_{cg,i}$$

$$\overline{Y}_{cg} = \frac{1}{m}\sum_{i=1}^{i=m} y_{cg,i}$$

Equation 22 Co-ordinates for the mean centre of gravity of a cluster

Where $\overline{X}$ and $\overline{Y}$ are the mean centre of gravity points for a cluster, *m* is the number of centre of gravity points in the cluster, *i* is the *i*th sample and *x* and *y* are the co-ordinates of each point in the cluster

If it is assumed that the points for a given cluster come from a bivariate normal distribution, then there are two possibilities for determining whether a measurement (or measurements) fit(s) within the same cluster (or sample population). The first method involves calculating the area within which a certain fraction of the population falls and determining whether the measurement would fall within this. The second method considers the area within which there is a given probability of finding the population mean centre of gravity. Both methods result in the production of an ellipse per cluster. The two types of ellipses are slightly different. The first method results in an equal frequency ellipse with the second resulting in a confidence ellipse for the population mean. If a single measurement is to be compared to a cluster, the equal frequency ellipse would be used, wheras if a series of measurements is to be compared, a comparison of population means would be preferable.

To calculate and plot the ellipses, it is necessary to know the equation for an ellipse. Details of how to calculate this for a bivariate normal distribution are given in Sokal and Rohlf[91].

The separation between clusters can often be enhanced by selection of spectal pre-treatment and the use of a particular spectral range. To semi-automate the process a 'scanning probability window' program was written in which a spectral window is scanned across the whole spectrum and a plot of log (probability) of spectral difference produced. This plot describes the probability of a sample spectrum being different from a reference set plotted against the wavelength of the window starting position. The position of the window which gives the best separation between clusters can easily be determined.

Figure 15 displays an example of how a scanning probability window would operate. For each portion of the spectral window the centre of gravity is calculated for each spectrum and the probability of difference of each test spectrum against the reference set is calculated (Figure 16).

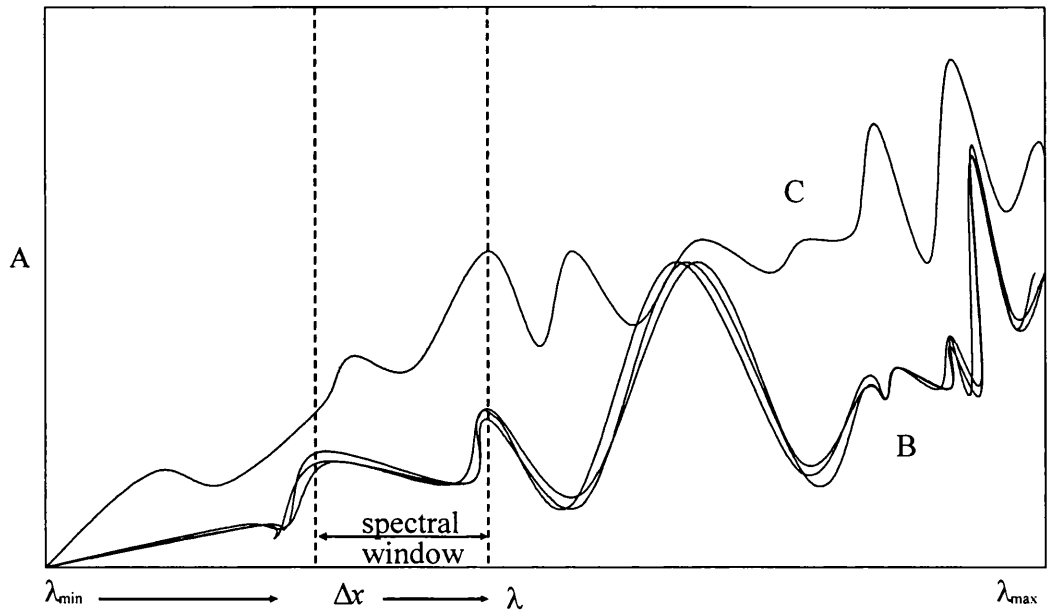


Figure 15 Illustration of reference spectra and a test spectrum with a scanning probabililty window.

A is the absorbance, B is the reference set of spectra, C is the sample spectrum from the test set, Δx is the window size in nm, λ is the wavelength,

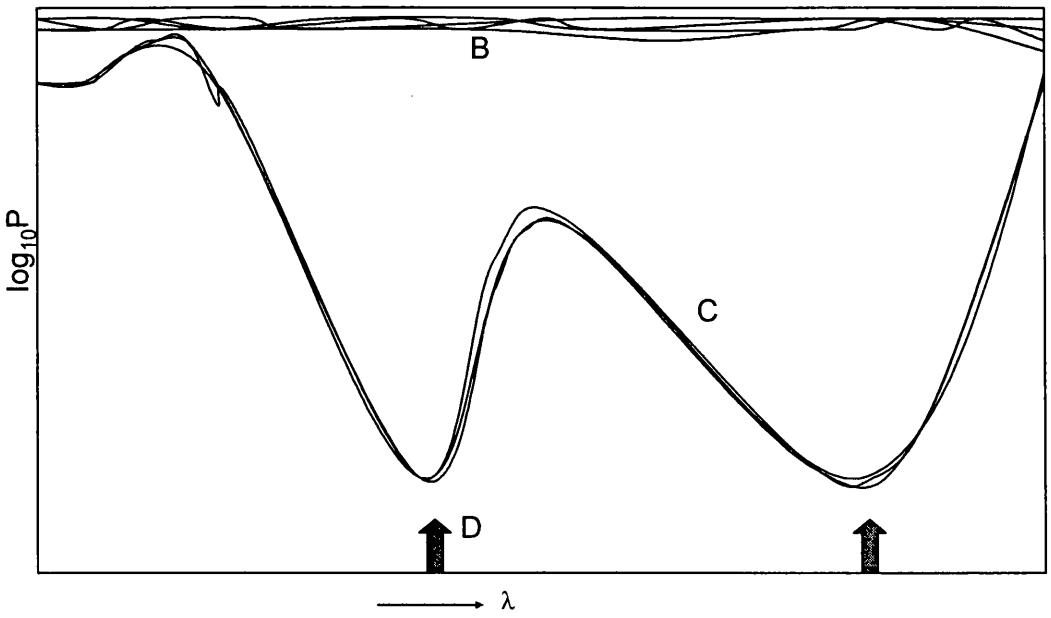


Figure 16 Illustration of scanning probability window

Where B is the reference set with respect to the mean reference set, C is the sample spectrum from the test set and D is one of two suggested starting window positions to give the best cluster separation. λ is the wavelength.

The point D along the wavelength axis in Figure 16 represents one of two points with a high probability of difference against the reference set. This represents an ideal start position for the spectral window to best separate between clusters.

1.11 Sample presentation

An advantage of NIR spectroscopy is the ability to obtain physical as well as chemical information about samples. Due to the precision of the technique and the power of data manipulation software, any influences upon the spectra will effect the interpretation of the data.

It is necessary to understand the factors that could influence the production of NIR spectra. It is also important to determine the impact such factors might have upon the spectra being examined. By gaining an understanding of these external influences it will aid interpretation between genuine chemical differences from effects of sample preparation.

As with any laboratory method, the reliability and integrity of the analysis is dependant upon the sample, the sampling method and its preparation.

The purpose of this part of the chapter is to examine factors that may have an effect on the production of NIR spectra in subsequent parts of this thesis, to demonstrate the effects to be aware of and to suggest either solutions or allowances that will be made during subsequent data interpretation.

The assumption behind any sample testing is that the sample under test is a 'true representative' of the entire batch (overall sample). This really depends on the objective of the analysis. The only true representation, however, is to sample the entire batch. There are many strategies that are employed in the pharmaceutical industry that do not necessitate the analysis of every single sample based upon terms of practicality, but sample selection and sampling strategy are of the utmost importance in routine analysis. When defining a validated method for sample analysis sample selection would therefore be a important factor.

Sample preparation is the second part of understanding the analysis process. This includes sample identification, handling, cleaning, sub-sampling, grinding and presentation. The effect due to sample particles will also be considered in this chapter. Sample presentation involves any factor influencing the position of the sample in relation to the incident ray and detection devices. Anything

interfering with this, for example a sample vial, must be considered. For tablets, there may be lack of uniformity and analysis will be affected depending upon which part of the tablet is in sight of the incident ray. Surface demarcations, scratching and scoring as well as coatings and polish must also be considered due to the same reasons.

1.11.1 Factors

The basis of the factors to be considered during this chapter have been identified by Williams[92] The 30 identified factors that could potentially affect samples and sample preparation could then impact accuracy and precision of the analysis. They are broken down into three areas, sampling, samples and sample presentation. All the factors that were identified were considered with respect to the experiments to be carried out in this thesis. The list is produced below with comments applied with respect to this project. Williams' work was carried out on agricultural products, so some of the factors were not applicable.

Sampling (below) – the samples were supplied from analytical laboratories or registered suppliers and were analysed in the laboratory only.

- Type of sampler – NIRS Systems Rapid Content Analyser used throughout the experiments
- Location of sampler – kept constant
- Material to be sampled – powders and solid samples
- Foreign material – not applicable
- Physical nature of material - powders and solid samples
- Size of sample – considered in terms of number of spectra produced and in terms of physical sample size used. Considered also in section 6.3.4
- Flow characteristics – not applicable
- Sampler transfer method – not applicable
- Blending – not applicable
- Storage – all samples stored in light proof cupboards or refrigerator depending on storage conditions required
- Identification – not applicable, full analysis supplied with each compound

- Variability of population to be sampled - considered during external database validation, section 6.3.4
- Frequency of sampling – not applicable
- Sub sampling – not applicable
- Sample selection – not applicable

Samples

- Type of material – not applicable
- Composition – oil, moisture, fibre – considered during results analysis of chapters 6 and 7
- Physical texture – not applicable
- Foreign material – not applicable
- Grinder type – grinding in general is important to the experiments and is examined further in this chapter
- Blending – not applicable
- Identification – not applicable

Sample presentation

- Instrument – kept constant
- Type of test – intermittent, on-line – intermittent - kept constant
- Cell type – important because vials and direct sampling used, considered further in this chapter
- Cell size – as above, considered further in this chapter
- Particle size – important for the different powdered substances to be analysed – considered further in this chapter
- Bulk density – considered further in this chapter with sample packing
- Composition – oil, moisture, fibre – considered during results analysis of chapters 6 and 7
- Physical nature – not applicable
- Stratification – not applicable
- Static electricity – considered in sampling protocol
- Cell loading – considered in sampling protocol
- Cell cleanup - considered

In addition

- Temperature – always at room temperature for all sample analysis but monitored
- Humidity – always at room humidity but monitored
- Moisture – as above, no drying or sample pre treatment was used
- Light – non excessive, standard laboratory lighting conditions

The factors to be investigated in more detail in this chapter are: varying the default instrumental scans, sample packing and presentation, grinding of powders, vial glass types, vial base diameter sizes and particle size and are described in more detail.

1.11.2 Instrument scans

These are the number of NIR scans of the defined wavelength range undertaken by the instrument to provide a sample spectrum. This is an instrumental parameter which is set from the NSAS software controller screen. The industry standard for quantitative studies is 32 scans[80]. This is the number of scans commonly use in pharmaceutical applications today. Each spectrum produced from a sample is the mean of the defined number of scan passes of the wavelength range.

The scan passes were investigated to determine the consequence of varying from the default setting with respect to spectral reproducibility and signal to noise ratio.

Reference is made to the work of O'Haver et al[81,82] who determined the systemic and random errors of measurement for a number of derivatives and also reported the signal to noise ratio being proportional to $\sqrt{}$ (number of scans).

1.11.3 Sample packing and presentation

When a powdered sample is placed onto a spectrometer and examined under diffuse reflectance the effect upon the incident ray is derived from the surface of the powder and a small penetrable amount below the surface. The sample vial is considered in section 1.11.5. A powder consists of non-regularly shaped

and sized particles the orientation of which when placed in contact with the light will produce varying results. An appreciation is required of the effect on results due to variations within samples as opposed to between samples. The purpose of this work is to investigate whether there is variation caused by examining different portions of the powder or tablet sample. The specific effect of particle size is examined in Chapter 4. This experiment demonstrates the effect of orientation and packing effects within powdered samples of drug actives.
A variety of powdered pure drug substances were used representing a range of particle shapes, textures and sizes. Tablets were examined with an expectation of large amounts of sample variation due to surface curvature, polish and packing effects.

1.11.4 The effect of grinding and particle size

Previously with NIR, crushing/ grinding was common practice for tablets and other samples such as grain. However attention was brought to the effects that grinding in different ways could have[92]. The effect of grinding samples has been investigated in this thesis as it was the intention to crush and grind some samples as part of the identification of tablet samples. Even with powders there must be consideration given to the effect upon mean particle size, particle size distribution, particle shape, energy transfer during the grinding process, leading to temperature rise and moisture loss.
Particle size is important to the pharmaceutical industry as it influences the way powders flow, mix and dissolve[93,94]. Measurements of particle size in pharmaceutical raw materials are commonly taken by forward angle laser light scattering (FALLS) and electrical zone sensing.

The ability of NIR to distinguish between not only chemically but physically different compounds is well reported. The effects of particle size and other physical characteristics of lactose monohydrate performance during pharmaceutical manufacturing are also well known. In recent years the distinction and identification of particle size by a variety of chemometric methods has been reported[95, 96].

What is not certain is the impact particle size will have upon raw and second derivative spectra being identified from a database of reference spectra.

1.11.5 Vial types and base diameter sizes

With NIR being a technique employing diffuse reflectance, the influence of the sampling vials must be noted. The potential for spectral differences caused by the vials must be understood before any examination of the technique with respect to examining samples can be undertaken.

The ideal vial will be either transparent or very low absorbing with a flat base free from any marking or scoring with minimal interaction with the NIR light. The assumption has been made previously that glass does not absorb in the NIR region, however, it is the intention to investigate any influence upon the spectra caused by the material from the vials themselves. It is also the intention to examine the effect of vial base diameter upon the spectra. Vials vary enormously in base diameter and the purpose is to determine if there is any influence caused by a minimum vial diameter.

Neither area has had a great deal of work contributed towards it. If influences upon the spectra are identified, ways of eliminating or reducing the effect will be examined. This is critical to the setting up of the spectral databases.

Chapter 2 Experimental

This chapter describes the methods and experimental details used for the TLC and NIR spectroscopy experiments.

All of the pharmaceutical test substances and reference standards for TLC, the NIR drug actives and solid dosage forms were obtained from the following sources:-

- Medicines Control Agency, London, England
- The Forensic Science Service, Aldermaston, England
- Medicines Testing Laboratory, Edinburgh, Scotland
- The School of Pharmacy, University of London, London, England

A comprehensive list of all the substances used can be found in the Appendices, at the back of this thesis.

2.1 Thin layer chromatography (Chapter 3)

Sixteen substances of pharmaceutical interest were chosen in order to provide a range of R_fs and a variety of chemical functional groups.

The substances selected were phenylpropanolamine, metoprolol, pheniramine, fluphenazine, haloperidol, promazine, metronidazole, oxazepam, cocaine, chlorprothixene, trimipramine, temazepam, papaverine, clobazam, prazepam and methaqualone.

Solutions containing each drug were prepared by dissolving the compound (2 mg) in chloroform (1 ml). The drug solutions were freshly made each day and were applied in 5 µl (10 µg) quantities to Whatman Al Sil G/UV_{254} chromatographic plates of 250 µm thickness. The plates were pre-treated by dipping in 0.1 M methanolic potassium hydroxide and then allowed to air dry. They were run for reproducibility using 4 basic TLC solvent systems namely methanol/ammonia (100:1.5), cyclohexane/toluene/diethylamine (75:15:10), chloroform/methanol (9:1) and acetone, and the drugs were located by their response to short (254 nm) and long (350 nm) wave UV light and by the use of ninhydrin solution, Dragendorff reagent and acidified iodoplatinate solution. The size of each plate was 20 cm by 20 cm with the solutes spotted 1.5 cm from the plate edge.

The experiments to determine factors that could influence production of R_f values were run on the chloroform/methanol system.

The robustness of the technique and the effect that variations in R_f value could have on final drug selection using the computer search techniques was investigated by altering the standard control conditions. The standard control conditions were made up of the following :-

1) Run distance	5 cm
2) Spot mass	10 µg
3) Temperature	≈ 20 °C
4) Chamber saturation	saturated
5) Age of solvent	freshly prepared and used once only
6) Solute spot	dried
7) Angle of plate development	90° angle with baseline

For each experiment, one of the standard control conditions was varied. The conditions that were varied (with alternate settings) are listed below.

1) Run distance	5 cm, 10 cm, 15 cm
2) Spot mass	1 μg, 2 μg, 5 μg, 10 μg, 20 μg
3) Temperature	4 °C, ≈20 °C (room temp) , 37 °C
4) Chamber saturation	saturated, unsaturated
5) Age of solvent	freshly prepared (and used only once), aging solvent (used more than once)
6) Solute spot	un-dried/ dried
7) Angle of plate development	37 ° angle with baseline, 90 °

8) The incorporation of internal reference standards

9) The examination of intra plate differences

The temperature experiments were set up in three areas. The first was set up and maintained at 4 °C in a thermostatically controlled refrigerator, the second at 20 °C was set up in a corner of the laboratory at room temperature, the area being well shaded, the temperature fairly constant and monitored. The final experiment was set up in a laboratory oven with the temperature held at 37 °C. Each chamber plus solvent was allowed an hour to come to equilibrium before the introduction of the TLC plate.

For the experiment varying the application spot mass between 1 and 20 μg. fresh solutions of each drug were prepared for each spot mass (1 μg, 2 μg, 5

μg, 10 μg and 20 μg) in order to allow each drug sample to be applied in one 5 μl aliquot. The sample solutions were prepared using

- 2 mg drug in 10ml Chloroform
- 4 mg drug in 10 ml Chloroform
- 10 mg drug in 10 ml Chloroform
- 20 mg drug in 10 ml Chloroform
- 40 mg drug in 10 ml Chloroform

A solution (5 μl) containing a mixture of specific reference compounds (taken from Moffat et al[10]) was also applied to each plate at three separate points along the baseline. Each set of experiments was replicated six times, and the results analysed by using the *F*-test analysis of variance. The R_fs generated from these standards were compared to literature values (contained within the Moffat reference). Based upon comparison between the two sets of values a factor was calculated and used to standardise the actual R_fs generated from the 16 test substances. Details of the factor calculation can be found in section 3.3.9.

For experiment (8) the incorporation of internal reference standards, ten drugs were run across the four basic TLC systems using the standard method and conditions above. The ten drugs were phenylpropanolamine, metoprolol, pheniramine, fluphenazine, metronidazole, cocaine, trimipramine, papaverine, prazepam and methaqualone. Each substance was run repeated six times on each solvent system.

For the final experiment (intra - plate differences), three samples of a drug were spread 4 cm apart along the baseline and were run on a Whatman plate and

the R_fs compared for intra-plate differences. The experiment was repeated for four drugs, each of which were run six times. The four drugs used were desipramine, dipipanone, caffeine and meclozine. The results and discussion can be found in section 3.3.10.

2.1.1.1 Locating Agents

- Examination under short (254 nm) and long (350 nm) ultraviolet light.

- Spraying a plate sequentially with ninhydrin solution and Dragendorff reagent. Acidified iodoplatinate was also used occasionally when required.

- Ninhydrin solution – Ninhydrin (0.5 g) + concentrated hydrochloric acid (10 ml) + acetone to 100 ml. The solution was prepared daily. After spraying the plate it was heated in an oven at 100 °C for 5 minutes. It was used to detect primary amines (violet/purple/pink colours) and secondary amines (yellow).

- Dragendorff reagent – solution a) contained bismuth subnitrate (2 g) + glacial acetic acid (25 ml) + water (100 ml), solution b) contained potassium iodide (40 g) + water (100 ml). The reagent was prepared by mixing 10 ml of solution a) + 10 ml of solution b) + glacial acetic acid (20 ml) + water (100 ml). The reagent was prepared every two days. It was used to detect alkaloids (yellow, orange, red/orange, brown or brown/orange colours).

- Acidified iodoplatinate solution – Platinum (IV) chloride solution (5% *m/V*) (5 ml) + potassium iodide (5 g) + concentrated hydrochloric acid (5 ml) + water to 100 ml. The solution was prepared daily, when required. It was used to detect tertiary amines and quaternary ammonium compounds (purple, blue/purple, grey/purple, brown/purple) and secondary and primary amines (dirtier colours of the above).

2.1.2 Computer based searches

Ten drugs were analysed across the four TLC systems to produce R_f data. These drugs were then treated as unknowns with computer based search methods incorporating fixed widows (section 3.4.1), moving windows (section 3.4.2), discrepancy index (section 3.4.3) and mean list length (3.4.4) being employed to search a database of 594 drugs (Moffat et al[10]). The purpose was to determine if it was possible to use four sets of experimentally generated data to produce a successful match of identification and if so to determine the most accurate and precise method of searching.

2.2 NIR spectroscopy - general sampling regimen

2.2.1 Powdered samples

The drug sample was placed in glass HPLC sampling vials. The vials were Waters 4 ml clear glass vials (WAT025051) with a flat base of diameter of 14 mm.

The instrument used was the NIRSystems model 6500 diffuse reflectance spectrophotometer with attached Rapid Content Analyser sampling module. The instrument was controlled using a Viglen 486SX Personal Computer (PC) containing 8 megabytes of RAM and a 700 megabytes hard disk drive. The spectral range examined was between 1100 nm and 2500 nm.

The protocol for sampling was as follows:

- Instrument set up diagnostics were completed as stated in the NIRS NSAS/IQ^2 manual[80].
- A ceramic reference spectrum was scanned and stored on the PC local hard disk drive.
- The glass vial containing the powdered sample to be analysed was placed directly onto the quartz sampling window of the Rapid Content Analyser (RCA) Module from the diffuse reflectance spectrophotometer.
- A spectrum was recorded from the drug sample.
- The glass vial was shaken and replaced
- A further spectrum was recorded

Twelve spectra were recorded (each consisting of 32 scans) for each drug sample using the same regimen.

2.2.2 Setting up a spectral database

Each set of twelve spectra were saved as files on the PC of the type .da and .cn. The .da file contained spectral data, the .cn file contained additional information such as constituent details, where applicable. These files were created as a part of the NIRSystems system IQ^2 software[80].

A spectral database was constructed from the spectral files utilising the instrument software package. Each drug substance was represented as a database 'product'. Each database entry consisted of the mean of the twelve spectra for each drug substance. Having entered each database 'product', database internal and external 'validations' were carried out (section 5.2.1.1.).

To construct an equivalent database for second derivative spectra, the spectra were converted into second derivative plots (refer to section 1.10.1, with a gap size of 0 nm and a segment size of 10 nm) and stored as database 'products' in a separate database.

2.2.3 Tablet Samples

The drug sample to be analysed were placed directly onto the quartz sampling window of the Rapid Content Analyser Module (RCA). The spectral range examined was between 1100 nm and 2500 nm.

The sampling order was as follows:

- Instrument set up diagnostics were completed as stated in the NIRS NSAS/IQ^2 manual[80].
- A ceramic reference spectrum was scanned and stored on the PC local hard disk drive.

- The tablet to be sampled was placed directly onto the quartz sampling window of the RCA
- A spectrum was recorded
- The tablet was rotated and a further spectrum was recorded
- After six spectra had been produced, the tablet was removed, turned over and replaced
- A spectrum was recorded from side two of the tablet.
- The tablet was rotated and a further spectrum was recorded
- After six spectra had been produced from side two of the tablet, (making twelve spectra in total) the tablet was removed and the sampling window wiped clean

2.3 Sample presentation (Chapter 4)

2.3.1 Instrument scans

The test samples used were allopurinol 100 mg tablets BP.

The General NIR Sampling Regimen for tablets described in section 2.2.3, formed the basis of the analysis with the following additions:

- Upon completion of recording the reference spectrum, the default instrument scan settings were altered to FOUR.
- 12 Spectra from a tablet sample were then recorded as detailed in section 2.2.3.
- The scan settings were then set to 16 scans with the above step repeated

- This was repeated for 32 and 64 scans.
- The RCA lid was opened, the tablet was removed and the sampling window was cleaned.

2.3.1.1 Sample packing and presentation

The general NIR sampling regimen for powders and tablets described in section 2.2 formed the basis of the analysis with the following alterations :

- Powdered drug samples were placed into glass sampling vials
- The first sample to be analysed was placed onto the quartz sampling window of the Rapid Content Analyser Module
- The sample was scanned by NIR, a spectrum was produced and stored
- The vial was shaken to disturb and rearrange the powdered drug particles
- The vial was replaced onto the sampling window and re-centred using the iris
- The sample was re-scanned to produce a second spectrum which was also stored
- The above steps were repeated until 12 spectra had been recorded

2.3.2 The effect of grinding samples

The general NIR sampling regimen for powders and tablets described in section 2.2 formed the basis of the analysis with the following additions :

Prior to analysis :

- Approximately 5 g of allopurinol BP powdered sample was placed into a pestle and mortar
- The sample was ground in a regular, circular manner
- Aliquots were withdrawn after 0 seconds, 30 seconds, 60 seconds and 90 seconds and transferred to labelled glass sampling vials
- A ceramic reference spectrum was obtained and stored.
- The powdered drug samples were placed into glass vials
- Each aliquot sample was scanned the vial was removed, shaken and replaced
- The sample was scanned again.
- The sample vial was, removed, shaken and re-scanned until 12 spectra had been recorded.

2.3.3 Investigation of vial types and base diameter sizes

The general NIR sampling regimen for powders described in section 2.2 formed the basis of the analysis with the deviations listed below.

The test sample utilised was powdered lactose monohydrate BP.

The sample vials plus sample were placed individually on to the sampling platform with 12 spectra recorded for each sample. The vials examined are listed below.

- Waters 4 ml vial A with base diameter 14 mm containing lactose
- Waters 4 ml vial B with base diameter 14 mm containing lactose
- Waters 4 ml vial C with base diameter 14 mm containing lactose
- Waters 4 ml vial D with base diameter 14 mm containing lactose
- Waters 4 ml vial E with base diameter 14 mm containing lactose

- BDH glass vial with base diameter 22 mm containing lactose
- Samco glass vial (code 6050/28) with base diameter 24 nm containing lactose
- Flat bottom wide rim glass vial with base diameter 26 mm containing lactose
- Waters 1 ml glass vial with base diameter 8 mm containing lactose

2.3.4 Particle size

Sieve fractions of known particle size, varying from <32 microns to >150 microns from batches of alpha-D-lactose monohydrate (97%) were incorporated into the database used in Chapter 5 for the purposes of this experiment only. This was achieved by adding spectra for the samples of each particle size listed in this section into the original spectra and second derivative databases - following the method described in 2.2.2. Having completed this experiment the spectra of the particle size samples were removed from each database. Different batches of sieved and unsieved lactose were tested for identification by different chemometric techniques. Spectra were derivatised and identification repeated using a second derivative database. Water content by loss upon drying was measured at each particle size.

Two 5 kg batches of lactose monohydrate were obtained from Pfizer Ltd., Sandwich, Kent. Sieve fractions were obtained for particle sizes between <32 µm and >150 µm by machine sieving and were examined by the Malvern Forward Angle Laser Light Scattering Diffraction 2600C. Fractions of each particle size were scanned on the Rapid Content Analyser through the Waters

4ml glass vials. 12 spectra were recorded for each fraction. The lactose fractions examined are set out below.

- <32 µm
- 32 µm mean particle size
- 45 µm mean particle size
- 63 µm mean particle size
- 93 µm mean particle size
- >150 µm

2.4 Drug Identification (Chapter 5)

2.4.1 Part 1 – initial study

The first experiment was a small-scale pilot in order to

- optimise the steps for setting up a small spectral database
- determine if structurally related compounds could be identified
- identify the best method for handling closely related compounds when scaling up to a larger database

Thirteen barbiturates, eight tetracyclines and six other compounds of pharmaceutical interest were set up in two different databases one containing spectra and the other with second derivatives. Each database was then examined separately.

The compounds used are contained within Appendix 1.

The identification testing was carried out by taking one of the powdered samples from each barbiturate batch and generating NIR spectra using the standard sampling regimen (above). The mean from the 12 spectra collected

was then compared against the spectral database in order to locate the nearest match.

Second derivative spectra were generated in the same way with the mean of the twelve second derivatives compared against the second derivative spectral database. For both series of compounds, two matching techniques - wavelength distance and correlation spectral matching were used. Detail on these techniques is contained within sections (1.10.2 and 1.10.3) Results and discussion can be found in Chapter 5 in section 5.3.

2.4.2 Part 2 – Large number of compounds

After the work with the closely related substances, the second part of the chapter focuses on scaling up for a large number of different compounds. Over 300 different powdered drugs were analysed by NIR spectroscopy as described in section 2.2 and set up in a second derivative database as described in section 2.2.2. An internal and an external validation of the database were carried out. This is described in Chapter 5 in section 5.3.3.

The following groups of test samples were then compared as unknown test substances, in a blind manner, against the large database.

- Samples of 32 'unknown' substances (the substances were treated as if unknown by not revealing their identity to the tester) were compared against the database

- The wavelength ranges used were 1100 -2498 nm, 2000 -2400 nm and 2200 -2400 nm

- Correlation spectral matching and wavelength distance matching were used to identify the 'unknown' substances.

Results and discussion are located in section 5.3.

2.5 The identification of actives in whole tablets (Chapter 7)

The tablet samples were analysed as described in section 2.2.3

For allopurinol 100 mg tablets compared against the pure drug second derivative database - two spectra were recorded for each tablet, 20 tablets were measured from each batch sampled, five batches were sampled from each of ten manufacturers (except for DDSA, where only four batches were available). Refer to Appendix 4 for full details.

For allopurinol 300 mg tablets compared against the pure drug second derivative database - two spectra were recorded for each tablet, 20 tablets were measured from each batch sampled, five batches were sampled from each of three manufacturers. Refer to Appendix 4 for full details.

For cimetidine 200 mg tablets compared against the pure drug second derivative database - two spectra were recorded for each tablet, 20 tablets were measured from each batch sampled, five batches were sampled from each of three manufacturers. Cimetidine pure drug was not included in the pure drug database as this was used as the control.

For aspirin 300 mg tablets compared against the pure drug second derivative database - two spectra were recorded for each tablet, 20 tablets were

measured from each batch sampled, five batches were sampled from each of three manufacturers.

The 'validated' database used for the identification of pure drugs in Chapter 5 was used to identify the active component within the allopurinol tablets. Details of database construction and validation can be found in section 2.2.2.

Second derivative spectral plots used were produced from the NIRSystems IQ^2 software based on a gap size of zero and a segment size of 10 nm.

Spectra were compared against the second derivative pure drug database using both correlation spectral matching and wavelength distance matching chemometric techniques. Details of both methods can be found in Sections 1.10.2 and 1.10.3 respectively.

Chapter 3 Identification by thin layer chromatography

3.1 Summary of aims

- Generate a sizeable amount of original analytical data with little cost of time involved
- Generate data across multiple TLC systems
- Determine robustness of the technique by varying standard control conditions
- Examine use of internal reference standards upon the data points across multiple TLC systems
- Apply rules based upon computer search results and relate to pharmacopoeial standards of drug identification

3.2 Method

The practical methodology followed can be found in the "Experimental" chapter (section 2.1).

3.3 Results and Discussion

3.3.1 Statistical Analysis of the Data

Significance tests were employed to test if the measured differences in the data could be accounted for by random error or if they were 'true' differences.

The *F*-Test Analysis of Variance (ANOVA) capable of analysing more than two sample means was used to describe variation due to changing the experimental control conditions. The data were tested at the 5% significance level and were calculated as a comparison between the *within* sample variation and the *between* sample variation. The calculations were automated by use of in-house ANOVA software.

Both variations were calculated individually.

$$\sigma^2 = \frac{\sum_i \sum_j (x_{ij} - x_i)^2}{h(n-1)}$$

Equation 23 Calculation of variation within samples.

where there are *h* conditions (e.g. temperature), with *n* members and *i* is the *i*th sample and *j* the *j*th measurement of this sample.

$$\sigma^2 = \frac{n\sum_i (\bar{x}_i - \bar{x})^2}{(h-1)}$$

Equation 24 Calculation of the variation between samples.

Where *n* is the number of samples, *h* is the number of conditions (e.g. temperature), *i* is the *i*th sample and $\bar{x}$ is the sample mean

If there is not a significant difference between the data i.e. any differences are attributable to random error then the two calculated estimates of σ^2 should not

differ significantly. To test if they are significantly different, a one tailed *F*-test was used with the critical value of F at P=0.05. An estimate greater than the critical value denoted a significant difference in the sample means. In the results table the *F*-test results are displayed as the probability that such a difference in the variances could occur by chance if the parent populations have the same variance. Significance is denoted by probabilities less than 5%. The results investigating variation in R_f values under different conditions were analysed for significance by the *F*-test analysis of variance 5% significance level.

3.3.2 Run distance

TLC actually proved to be a difficult and frustrating technique to set up and often produced results that were not of a reproducible and consistent nature. This called into question one of the perceived benefits of being able to quickly set up and generate a large volume of results from TLC. The problems encountered included non-appearance of spots altogether after application of the appropriate reagents, for example when the sample was pethidine, the mobile phase was cyclohexane, toluene and diethylamine and the plates were sprayed with ninhydrin and Dragendorff reagent and viewed under long and short ultra violet light.

To determine if this was due in particular to the batch of sample being used, a standard reference solution containing codeine, desipramine, pethidine and dipipanone was run across the same system with the same conditions and plates. This also encountered similar separation problems.

In summary, in an attempt to stabilise the external factors that might have been affecting reproducibility, all mobile phases were freshly prepared and re-

prepared with plates run many times (approximately 40 times). Different batches of the Whatman Sil G/UV_{254}TLC plates were used, with little effect. The powdered reference standards were obtained from The Forensic Science Service Laboratories in Aldermaston and contained a full analysis history. Alternative samples were obtained and presented the same results. Three different analysts performed the analysis in order to determine if analyst error was a cause. Each analyst prepared solvent and solutes and prepared and spotted the TLC plates, however, similar effects were presented.

Similar difficulties were encountered with solutions of amitriptyline, procaine, mepivacaine and meclozine run on a mobile phase system of acetone. Spot concentrations as high as 25 μg were attempted and slightly improved the separation. However, the results were not reproducible.

An acceptable level of stability was eventually achieved by switching the overall chromatographic plate size from the Whatman 80 mm X 40 mm to 200 mm X 200 mm (same constitution from Whatman). The poor separations would appear to have originated from the plate size itself and possibly the proximity of the samples to each other. Previously the samples had been spotted 10 mm from the plate edge and with a baseline distance of 5 mm between each spot. The new plates were cut to 200 mm X 100 mm, 10 mm was left between the plate edges and the solute spot but spots were 10 mm apart. Using the previous, smaller plates with solute spots 10 mm apart did improve the shape and appearance of the resulting TLC spots but the colour, intensity and shape were not as clear and 'tail free' as with the newer larger plates. The newer plates were therefore used as this also permitted more samples to be loaded onto one plate (up to a maximum of 8 solute spots). The exercise did

demonstrate firstly how difficult it was to control the spot appearance and secondly how much manually intensive the technique is. Raw data from the experiment is presented for 5, 10 and 15 cm run distances. All spot distances were measured with a ruler. The distance being taken from the centre of the spot to the baseline. In further experiments some raw data will be presented but more often the results of the significance tests and trends are displayed.

The purpose behind this experiment was to compare the run distances of 5 cm, 10 cm and 15 cm. However, as can be seen from the results in Table 5, the actual run distance had relatively little impact as compared to the overall size of the plate.

When the TLC systems were therefore set properly, the separation and reproducibility were good.

TLC DATA

Table 2 R_f values from the TLC run against a 5 cm run distance. The solvent system used was chloroform: methanol 1:1. R_f values are defined as percentage R_fs

							MEAN	STD DEV
DRUG								
Phenylpropanolamine	4	4	3	4	5	3	3.8	0.8
Metoprolol	8	8	4	7	9	7	7.2	1.7
Pheniramine	9	10	4	10	17	9	9.8	4.2
Fluphenazine	30	31	28	25	33	24	28.5	3.5
Haloperidol	30	28	31	25	33	27	29.0	2.9
Promazine	26	23	28	29	33	29	28.0	3.3
Metronidazole	23	23	24	19	25	18	22.0	2.8
Oxazepam	13	25	27	21	28	18	22.0	5.8
Cocaine	26	25	25	21	28	24	24.8	2.3
Chlorprothixene	48	49	50	55	54	58	52.3	3.9
Trimipramine	50	52	52	58	57	59	54.7	3.8
Temazepam	60	60	62	60	59	51	58.7	3.9
Papaverine	73	74	70	71	69	72	71.5	1.9
Clobazam	72	70	69	69	68	65	68.8	2.3
Prazepam	76	72	71	73	72	72	72.7	1.8
Methaqualone	78	74	75	71	74	72	74.0	2.4

Table 3 Raw R_f values from the TLC run against a 10 cm run distance. The solvent system used was chloroform: methanol 1:1. R_f values are defined as percentage R_fs

							MEAN	STD DEV
DRUG								
Phenylpropanolamine	3	4	3	4	5	3	3.7	0.8
Metoprolol	6	8	4	7	9	7	6.8	1.7
Pheniramine	7	10	4	10	17	9	9.5	4.3
Fluphenazine	14	22	28	25	33	24	24.3	6.3
Haloperidol	23	26	31	25	33	27	27.5	3.8
Promazine	14	24	28	29	33	29	26.2	6.6
Metronidazole	17	22	24	19	25	18	20.8	3.3
Oxazepam	17	21	27	21	28	18	22.0	4.6
Cocaine	19	26	25	21	28	24	23.8	3.3
Chlorprothixene	49	42	50	55	54	58	51.3	5.6
Trimipramine	56	47	52	58	57	59	54.8	4.5
Temazepam	60	58	62	60	59	51	58.3	3.8
Papaverine	79	70	70	71	69	72	71.8	3.7
Clobazam	72	69	69	69	68	65	68.7	2.3
Prazepam	78	72	71	73	72	72	73.0	2.5
Methaqualone	79	74	75	71	74	72	74.2	2.8

Table 4 Raw R_f values from the TLC run against a 15 cm run distance. The solvent system used was chloroform: methanol 1:1. R_f values are defined as percentage R_f s

							MEAN	STD DEV
DRUG								
Phenylpropanolamine	3	3	3	3	3	3	3.2	0.4
Metoprolol	6	6	7	7	6	6	6.5	0.8
Pheniramine	9	8	10	9	8	7	9.2	2.0
Fluphenazine	25	24	27	25	25	26	25.0	1.3
Haloperidol	26	25	27	26	25	28	26.3	1.1
Promazine	24	23	25	24	21	21	24.0	3.1
Metronidazole	18	13	16	19	12	18	18.9	8.0
Oxazepam	20	15	18	21	12	18	20.5	9.1
Cocaine	25	23	20	28	25	23	27.2	9.0
Chlorprothixene	59	61	58	55	54	51	55.6	3.9
Trimipramine	61	62	61	58	58	55	58.4	3.1
Temazepam	63	71	62	60	51	58	60.6	6.0
Papaverine	73	73	77	72	73	74	72.5	3.6
Clobazam	70	73	69	70	71	69	70.3	1.4
Prazepam	75	77	75	75	73	74	74.7	1.2
Methaqualone	77	77	79	77	76	74	77.2	2.0

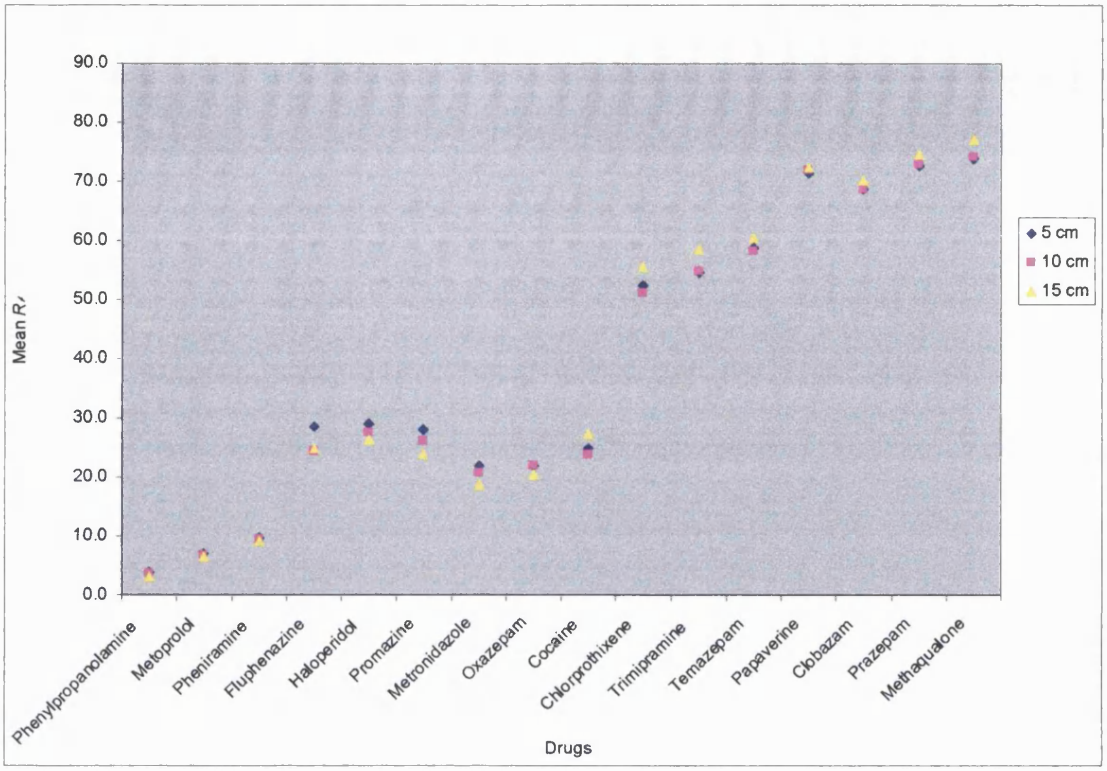


Figure 17 Mean R_f values for three run distances: 5 cm, 10 cm & 15 cm

Each data value is the mean of six R_f values. The solvent system used was chloroform: methanol 9:1. The data can be found in Table 2, Table 3 and Table 4

Table 5 Percentage probability values of the *F*-test analysis of variance for the 5cm,10cm and 15cm run distances.

Significance is denoted by values less than 5%

DRUG	*F*-TEST PROBABILITY
Phenylpropanolamine	43.2
Metoprolol	41.3
Pheniramine	44.2
Fluphenazine	2.4
Haloperidol	24.9
Promazine	41.8
Metronidazole	2.07
Oxazepam	15.7
Cocaine	33.7
Chlorprothixene	17.5
Trimipramine	20.1
Temazepam	19.8
Papaverine	30.0
Clobazam	44.6
Prazepam	48.9
Methaqualone	11.9

Having stabilised the majority of substances, two remained to cause frequently varying results - fluphenazine and metronidazole. Both structures can be seen below, both substances are in the mid-low R_f range and possess ethanolic side chains.

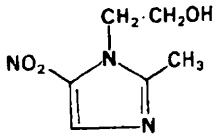


Figure 18 The chemical structure of 2-(2-Methyl-5-nitroimidazol-1-yl) ethanol (metronidazole)

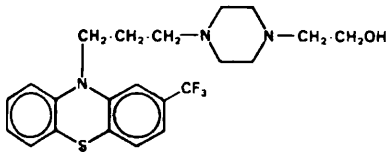


Figure 19 The chemical structure of 2-{4-[3-(2-Trifluoromethylphenothiazin-10-yl) propyl]piperazin-1-yl}ethanol (fluphenazine)

The areas that may have led to the inconsistencies are the solute preparation or the sample spot application on to the TLC plates, despite having tried to standardize the application process, this is still heavily manual.

3.3.3 Temperature

For this experiment the spot mass was 10 µg, a saturated solvent chamber was used with freshly prepared solvents, the solutes were applied to the chromatographic plates with the spots dried with a hair dryer and the plates were placed to stand at 90 ° to the chamber base.

Changes in temperature (Table 6) produced significant changes in drug behaviour. Increasing temperature (from 4 °C to 20 °C and 37 °C) resulted in a rise in R_f values in all of the samples (Figure 20). This increase may be due to evaporation of the mobile phase from the layer and the increased mobile phase flowing over it. The solubility of the solute in the mobile phase and the interaction between this and the sorbent may also be changing.

The best temperature, if defined as that which is most comparable to the database of drug R_f values produced by Moffat et al[10], was 20 °C.

Changing the temperature from 4 °C to 20 °C produced differences in the R_f obtained of up to 76.6 % (Table 6). This represented an actual R_f difference of over 42 values. Of the 16 drugs there was a change greater than 42% in 11 cases and only two drugs (phenylpropanolamine and methaqualone) produced single digit percentage differences. The difference between 20 °C and 37 °C was less dramatic but still statistically significant (Table 7). The highest difference was of 54.4 % but generally the actual differences were lower than the 4 to 20 °C degree range. What is unquestionable, however, is that quite pronounced changes in behaviour do occur with a change in temperature. This has been known to a degree but what is demonstrated here is the impact this would have on pharmaceutical identification by TLC. It is suggested therefore that either analysis must be carried out at a controlled temperature alternatively factors may have to be used to standardise R_f values. For this to occur it would be suggested as further work, that the work of Moffat et al[10] be repeated over a series of different temperatures and that factors for standardisation be built from this. It was anticipated that 1-2 °C either side of a 20 °C (room temperature) would not produce significant differences, but it is clear that the

term 'room temperature' itself needs to be defined with each analysis. However, when considering the overall temperature from 4 °C to 37 °C, a Δt of 33 °C, some of the samples can have seen to have changed in R_f by more than 100%.

Table 6 Analysis of the percentage difference in R_f values caused by altering the temperature from 4 °C to 37 °C.

DRUG	Mean R_f 4 °C	Mean R_f 20 °C	Difference between 20 and 4 °C	% Difference between 20 and 4 °C	Mean R_f 37 °C	Difference between 37 and 20 °C	% Difference between 37 and 20 °C
Phenylpropanolamine	6.3	6.8	0.5	**6.9**	10.5	3.7	**54.4**
Metoprolol	4.7	13.0	8.3	**64.1**	19.3	6.3	**48.7**
Pheniramine	5.8	22.4	16.6	**74.0**	22.3	-0.1	**-0.3**
Fluphenazine	12.8	35.4	22.6	**63.7**	39.2	3.8	**10.6**
Haloperidol	9.2	35.2	26.0	**74.0**	41.7	6.5	**18.4**
Promazine	11.0	32.6	21.6	**66.3**	39.8	7.2	**22.2**
Metronidazole	8.7	31.6	22.9	**72.6**	33.5	1.9	**6.0**
Oxazepam	17.8	31.0	13.2	**42.5**	32.7	1.7	**5.4**
Cocaine	22.8	30.8	8.0	**25.9**	42.2	11.4	**36.9**
Chlorprothixene	10.5	43.4	32.9	**75.8**	59.5	16.1	**37.1**
Trimipramine	30.0	46.0	16.0	**34.8**	64.2	18.2	**39.5**
Temazepam	13.0	55.6	42.6	**76.6**	64.3	8.7	**15.7**
Papaverine	30.7	71.6	40.9	**57.2**	75.7	4.1	**5.7**
Clobazam	22.5	67.8	45.3	**66.8**	72.8	5.0	**7.4**
Prazepam	55.0	70.4	15.4	**21.9**	77.0	6.6	**9.4**
Methaqualone	70.7	69.6	-1.1	**-1.5**	78.3	8.7	**12.5**

Table 7 Percentage probability values of the *F*-test analysis of variance for the 4 °C, 20 °C and 37 °C temperatures.

Significance is denoted by values less than 5%

DRUG	*F*-TEST PROBABILITY (%)
Phenylpropanolamine	0.36
Metoprolol	<0.01
Pheniramine	<0.01
Fluphenazine	<0.01
Haloperidol	<0.01
Promazine	<0.01
Metronidazole	<0.01
Oxazepam	<0.01
Cocaine	<0.01
Chlorprothixene	<0.01
Trimipramine	<0.01
Temazepam	<0.01
Papaverine	<0.01
Clobazam	<0.01
Prazepam	<0.01
Methaqualone	2.54

Room temperature was clearly not adequate for an experiment of this complexity. Although the trend can be seen, quantification can not be accurately determined. The approach would need to be repeated with accurate temperature recording down to 0.1°C, it is suggested.

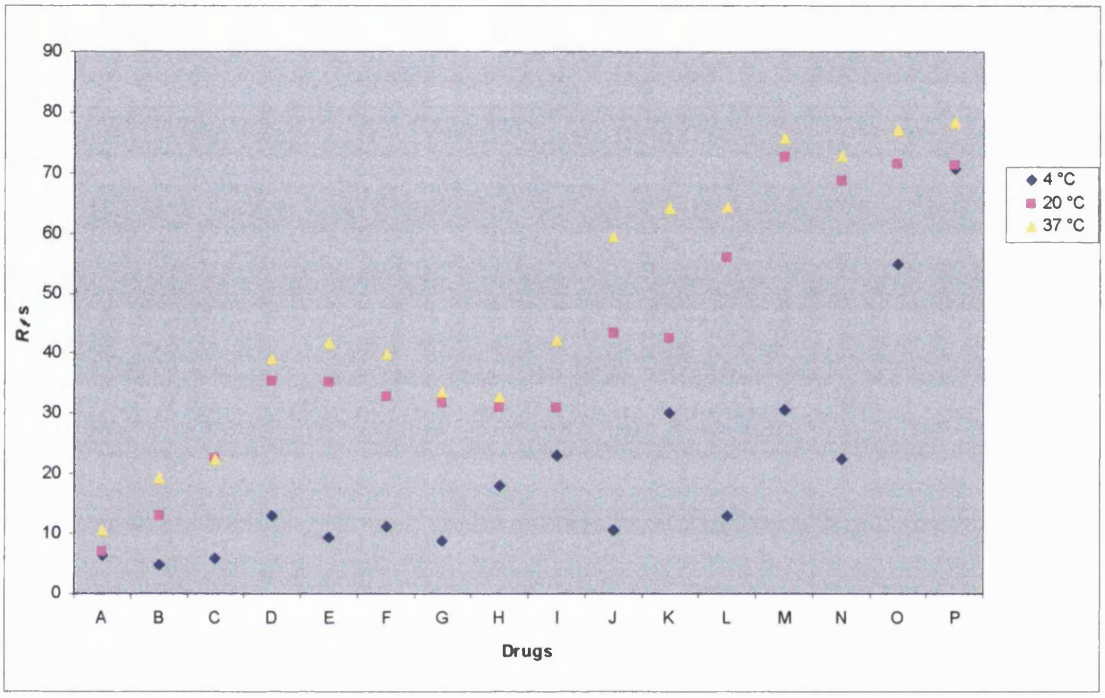


Figure 20 Mean R_f values for 4 °C, 20 °C and 37 °C temperatures.

Key: A-phenylpropanolamine, B-metoprolol, C-pheniramine, D-fluphenazine, E-haloperidol, F-promazine, G-metronidazole, H-oxazepam, I-cocaine, J-chlorprothixene, K-trimipramine, L-temazepam, M-papaverine, N-clobazam, O-prazepam, P-methaqualone.

3.3.4 Spot mass

It has been reported that increasing the mass of the sample drug spots applied, increases the repective R_fs of the substances[97]. In addition, it was also indicated that R_f values are less reproducible in the middle of a chromatogram than at high or low R_fs. It was not clear from the literature, however, whether this was related to the physical position of a sample spot or the R_f value itself (e.g. low/mid/high R_f value).

The test sample spots were applied to two TLC plates, Figure 21 displays data from the first plate and Figure 22 shows data from the second TLC plate. Each data point on the two graphs is the mean of six TLC runs.

The 1 μg spot mass did not produce anything that could be visualised under either short (254 nm) or long (350 nm) UV light or Dragendorff's reagent.

A surprising trend was noted in the results – this was that R_fs for the entire 5 μg run were clearly lower than the other samples. This applies not just to one or two data points but to the mean results displayed in Figure 21 and Figure 22. The standard deviations displayed in Table 9 Table 10 indicate that this was not due to outliers. The results were reproducible and consistently lower that all other results over the 12 plates.

Additionally, the R_fs for spot masses of 2, 10 and 20 μg were very consistent. In other words they were reproducible and there were no obvious outliers (Table 9). Only prazepam (6.6 sd) contained a standard deviation greater than 3.2 and it appeared that the drugs had been equally impacted. It was a strange result because the remaining sample masses between 2 and 20 μg had been so consistent. It was possible that this was something to do with the solvent, or development chambers or even the TLC plates (such as the KOH application). To determine if something had gone wrong or in fact these were genuine results, the 5 μg tests were run again. The same 'stock' sample solutions with new TLC plates used, the solvent was freshly prepared and the development chambers were set up from fresh (and allowed to come to equilibrium). As in the original experiment, twelve TLC plates were run in total, with six sets of spots for each drug sample. The new set of results (Table 10) were in keeping with the original results for the 2, 10 and 20 μg spot masses (Figure 22 and

Figure 23), however there was greater variation between the results (higher standard deviations per drug). It is suggested that there was an issue with the original results for the 5 μg spot mass, although it is not conclusive. It may have added more to the evidence to have run a third batch of samples, however this would not have necessarily highlighted where the inconsistency had been. What can be said however, is that the second batch of results tied in more with the historical literature results).

Table 8 Percentage probability values of the *F*-test analysis of variance for 2 μg, 5 μg, 10 μg, and 20 μg spot masses

Significance is denoted by values less than 5%. Probability values have been calculated from the original data set

DRUG	*F*-TEST PROBABILITY (%)
Phenylpropanolamine	13.84
Metoprolol	16.3
Pheniramine	1.18
Fluphenazine	<0.01
Haloperidol	<0.01
Promazine	<0.01
Metronidazole	8.01
Oxazepam	10.7
Cocaine	10.0
Chlorprothixene	<0.01
Trimipramine	<0.01
Temazepam	<0.01
Papaverine	<0.01
Clobazam	<0.01
Prazepam	<0.01
Methaqualone	19.7

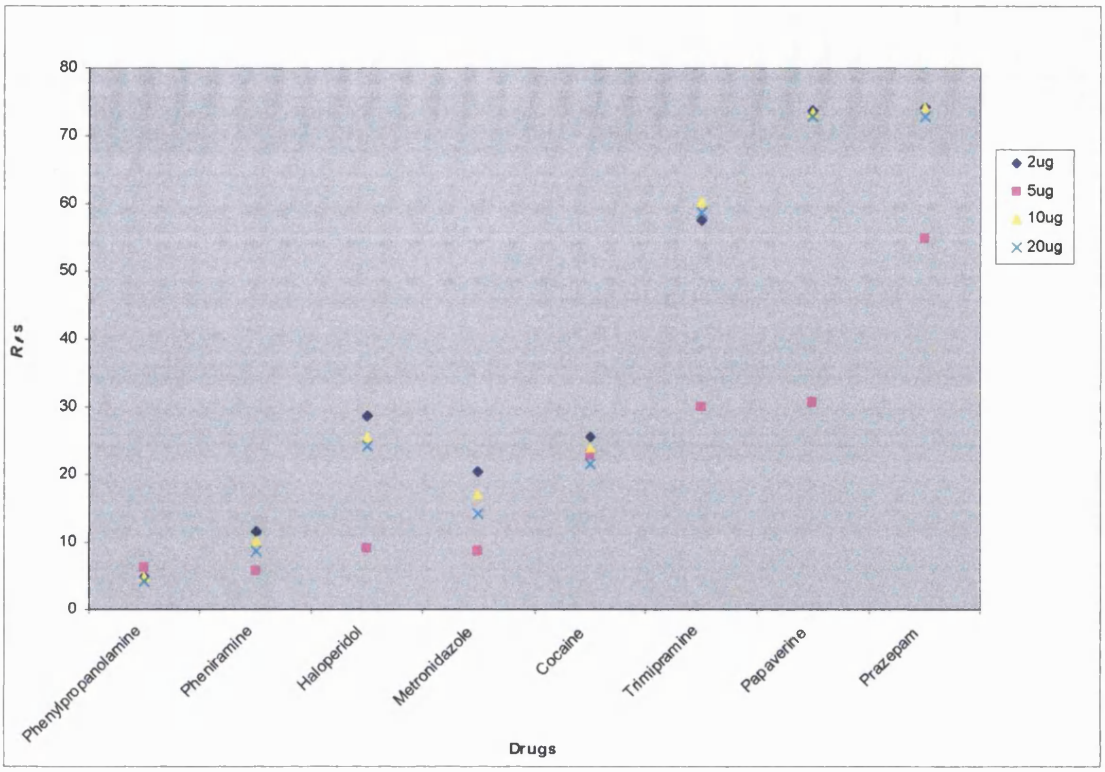


Figure 21 Mean R_f values for spot masses of 2 μg, 5 μg, 10 μg and 20 μg.

Each R_f is the mean of six R_fs. All of the above samples were applied to one TLC plate. This is chart 1 of 2 and contains the first eight substances

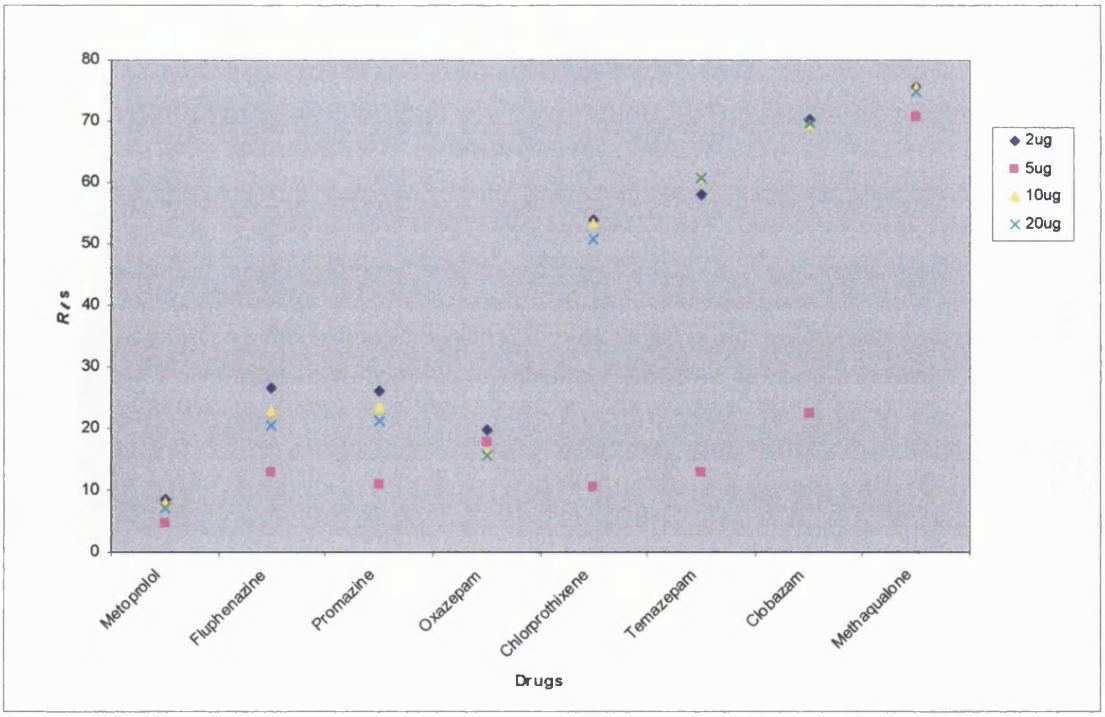


Figure 22 Mean R_f values for spot masses of 2 μg, 5 μg, 10 μg and 20 μg. Chart 2 of 2, containing the second set of eight substances

Table 9 Original R_f data for 5 µg spot mass.

Solvent is chloroform/methanol (90:10), all other conditions were standard as described in section 2.1. R_f (1) – (6) were recorded from six different TLC plates.

							MEAN	STD DEV
Phenylpropanolamine	6	5	6	7	6	8	6	1.0
Pheniramine	8	6	6	6	4	5	6	1.3
Haloperidol	8	9	8	9	12	9	9	1.5
Metronidazole	9	9	8	9	8	9	9	0.5
Cocaine	24	25	23	21	23	21	23	1.6
Trimipramine	29	31	31	30	30	29	30	0.9
Papaverine	32	31	32	30	30	29	31	1.2
Prazepam	46	60	51	63	59	51	55	6.6
Metoprolol	5	5	5	5	3	5	5	0.8
Fluphenazine	12	15	11	13	12	14	13	1.5
Promazine	15	13	9	9	9	11	11	2.5
Oxazepam	20	20	15	18	16	18	18	2.0
Chlorprothixene	8	14	10	15	8	8	11	3.2
Temazepam	11	15	10	17	12	13	13	2.6
Clobazam	24	24	22	23	22	20	23	1.5
Methaqualone	67	75	68	73	69	72	71	3.1

Table 10 Raw R_f data for the re-test batch of 5 µg spot mass.

Each batch consists of six TLC runs for the 16 drugs. The final R_f is the average for each drug from the six runs. The results are plotted on two charts for clarity. The referenced literature values are from Moffat et al[10].

	*Literature**							MEAN	STD DEV
Phenylpropanolamine	***4***	3	4	3	4	5	3	4	0.8
Pheniramine	***13***	7	10	4	10	17	9	10	4.3
Haloperidol	***27***	23	26	31	25	33	27	28	3.8
Metronidazole	***36***	17	22	24	19	25	18	21	3.3
Cocaine	***47***	19	26	25	21	28	24	24	3.3
Trimipramine	***54***	56	47	52	58	57	59	55	4.5
Papaverine	***65***	79	70	70	71	69	72	72	3.7
Prazepam	***74***	78	72	71	73	72	72	73	2.5
Metoprolol	***8***	6	8	4	7	9	7	7	1.7
Fluphenazine	***23***	14	22	28	25	33	24	24	6.3
Promazine	***30***	14	24	28	29	33	29	26	6.6
Oxazepam	***40***	17	21	27	21	28	18	22	4.6
Chlorprothixene	***51***	49	42	50	55	54	58	51	5.6
Temazepam	***59***	60	58	62	60	59	51	58	3.8
Clobazam	***70***	72	69	69	69	68	65	69	2.3
Methaqualone	***80***	79	74	75	71	74	72	74	2.8

Based upon the overall results, there actually appears to be little impact caused by varying the spot mass with even 2 μg of solute producing R_f values that were consistent. It was noticeable, however, that with the 2 μg sample, it had been difficult at times to detect the sample spot under UV light as in some cases, intensities of spot fluorescence were many times less than with the 5 μg spots. It was considered whether to re run the experiments with a 1 μg spot but there did not appear to be great benefits to be gained by the expected scenario of not detecting any spots. In addition 2 μg had demonstrated successfully that this also would not be a recommended spot mass to use routinely because of the possible spot detection problems, but it did show that the final R_f values were not affected if the final spot could be detected.

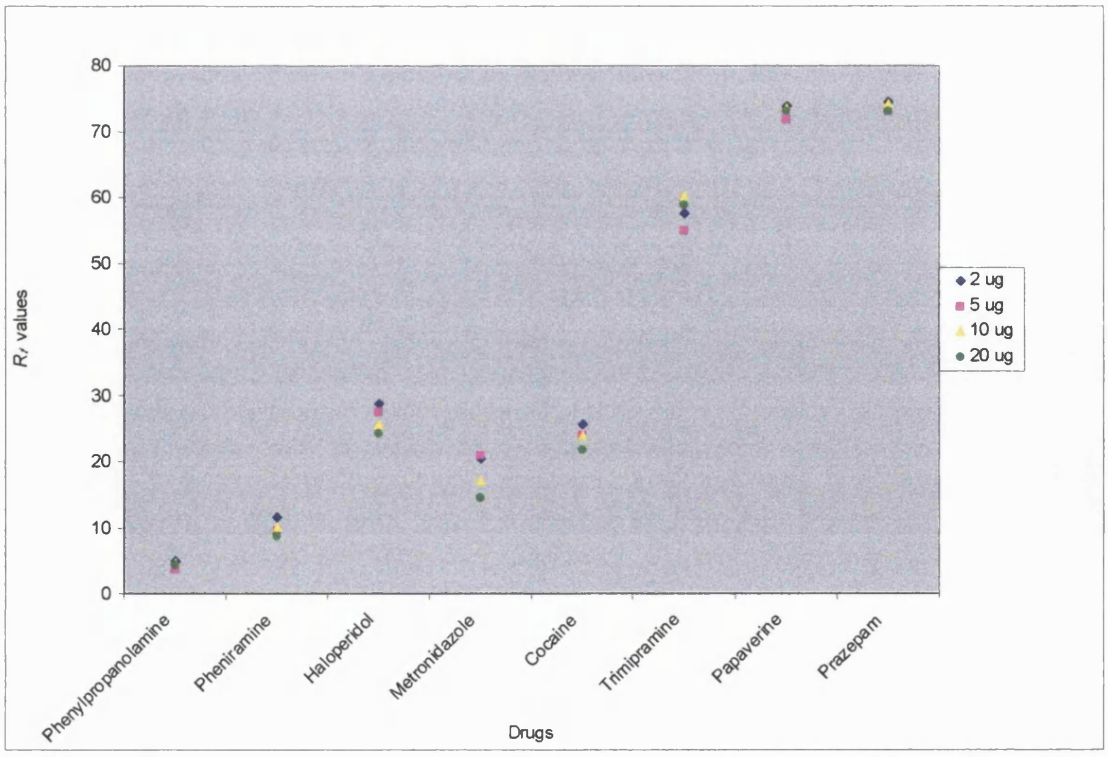


Figure 23 Mean R_f values for spot masses of 2 μg, 5 μg, 10 μg and 20 μg using the re-run data for 5 μg

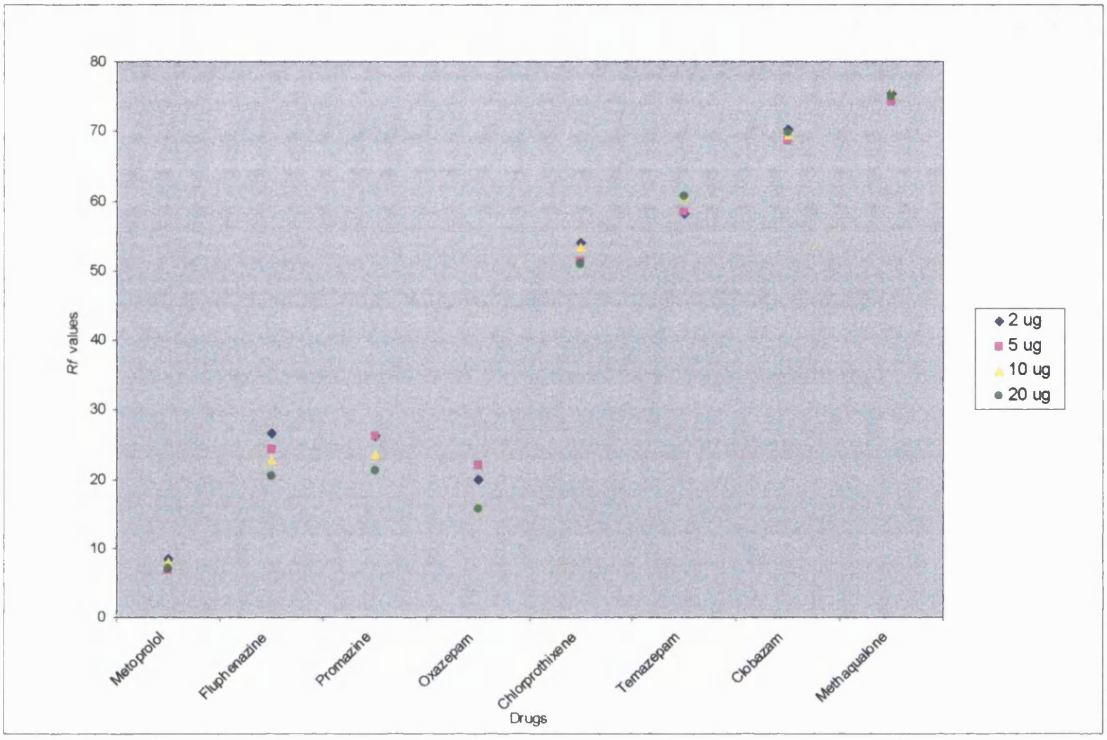


Figure 24 Mean R_f values for spot masses of 2 μg, 5 μg, 10 μg and 20 μg using the re-run data for 5 μg

It should also be pointed out that that at the other end of the scale with 20 μg spots it was often difficult to determine exactly the centre of the final spot position as the spots were larger in diameter than any other and in many cases had quite intense tails present as well. Even though it was not demonstrated in these results, inaccuracies may result from not being able to determine the centre of the TLC spot, this will be an issue if attempting to separate a mixture of compounds. The spots used in these experiments were single drug spots. It would not be recommended to use a spot mass of below 5 μg for routine analysis and also a spot mass of 20μg or above - 5 and 10 μg spots provided the best results.

In summary, varying the mass of the sample applied to the TLC plate did not affect the R_f values produced between spot masses of 2 and 20 μg. Spot masses below 2 μg did not produce visible spots.

3.3.5 Angle of chromatographic plate development

When considering the part of the TLC process generally described as 'running the plates' very little is described in the literature about the actual position of a plate in a development chamber, or indeed if this is important or not. Common and generally accepted TLC texts and standard references[9,10,97] describe the setting up of TLC experiments with words similar to "the TLC plate is placed in a vertical position in the tank so that the application line is above the level of the mobile phase". What if the plate is not vertical ? What if the plate is in fact as 'un-vertical' as possible ? Quite possible if a TLC was place against the wall of a development chamber but had fallen over.

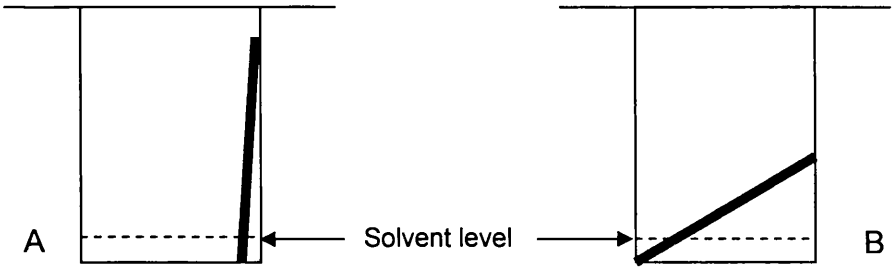


Figure 25 Diagram of the two development chambers.

The first chamber (A) contains a vertical TLC plate making an angle of approximately 90° with chamber base with the (B) showing the TLC plate making an angle of 37° with the chamber base. Note in (B) the baseline containing the solute samples was above the solvent level

The two scenarios set up were the traditional setting of the vertical plate resting against the wall of the development chamber and the plate making as acute an angle possible with the baseline of the chamber (see Figure 25). The remaining

TLC conditions were the experimental standard (see section 2.1). The graph (Figure 26) contains the results from both vertical and angled TLC plates. It is found that overall, the development angle does not have a clear impact effect upon the final R_f value produced. Statistically comparing the full sets of data by *F*-test statistical analysis (Table 11) the two data sets are found not to be statistically different from each other. However, this is not clear cut and there were three exceptions to this - as can be seen in both the graph and from the statistical probability table. The exceptions were metronidazole, oxazepam and prazepam which have varied in value. Additionally, trimipramine and temazepam are borderline to this. So there appears to be a trend but does not involve all of the benzodiazepines (clobazam does not display this). Of the substance involved, the solubility in water, ethanol and chloroform are not entirely dissimilar, the dissociation constants are also close (between 1 and 3). However, thinking this through, it is somewhat unexpected when considering that the main movement of the drugs across the TLC plate is by capillary action and that only the gravitational affect has been altered.The interaction between the individual drug, the mobile phase and the finely divided adsorbent bonded to the TLC plate surface should not have been changed by merely altering the tilt angle of the plate. There should not have been any chemical disturbances. The message that is taken from this experiment is that the standard conditions must be kept as consistent as possible in a bid to decrease the amount the variation that could be introduced into further experimentation.

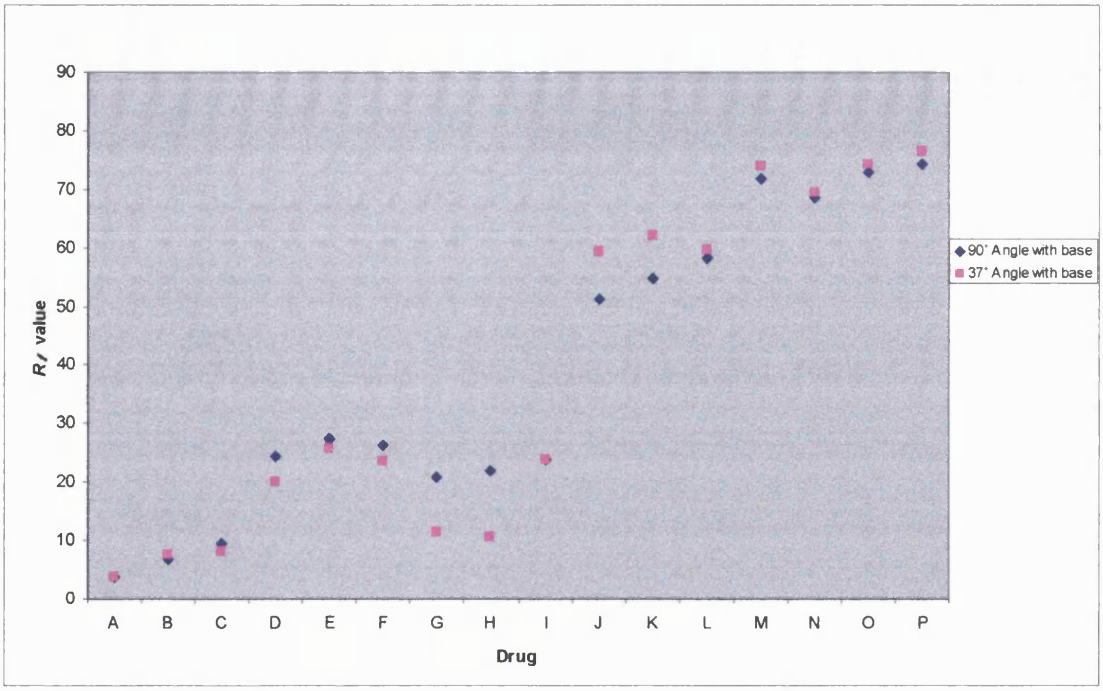


Figure 26 The effect of development angle upon R_f values for 37° and 90° angles of plate development.

Key: A-phenylpropanolamine, B-metoprolol, C-pheniramine, D-fluphenazine, E-haloperidol, F-promazine, G-metronidazole, H-oxazepam, I-cocaine, J-chlorprothixene, K-trimipramine, L-temazepam, M-papaverine, N-clobazam, O-prazepam, P-methaqualone. Each data point is the mean of six values.

Table 11 Percentage statistical values of the *F*-test analysis of variance for 37° angle with baseline and 90° angle with baseline.

Significance is denoted by values less than 5%

DRUG	*F*-TEST PROBABILITY (%)
Phenylpropanolamine	30.9
Metoprolol	47.3
Pheniramine	34.1
Fluphenazine	10.0
Haloperidol	14.3
Promazine	34.0
Metronidazole	0.175
Oxazepam	3.72
Cocaine	12.0
Chlorprothixene	6.1
Trimipramine	11.3
Temazepam	40.8
Papaverine	34.3
Clobazam	21.9
Prazepam	0.21
Methaqualone	21.5

3.3.6 Effect of chromatographic spots that have not been dried

Application of the sample spots requires a high degree of manual skill, the quality of the end result being heavily dependant upon successful execution of this stage. As part of the standard TLC methodology[97], the sample spots are applied to the TLC plate and dried - as is commonly found in such texts, the spots must be dry prior to the placement of the TLC plate into the development chamber. This experiment sought to find out whether there were any consequences of not doing this properly and what they were. The only difference employed the standard TLC method (section 2.1) was that six sets of TLC plates were prepared but did not utilise the hair dryer to dry the spots, instead the visibily non-dry plates containing 'damp' spots were placed into the development chambers. Six sets of these plates were compared in this manner against six sets of plates prepared using the standard method, as a control.

Table 12 R_fs for drug substance spots that were not dried prior to development of the TLC plate

The table displays the substances, reference letter use on subsequent graphs, standard literature R_f values taken from Moffat et al[10], the six sets of raw data, mean and standard deviation. As described earlier (section 3.1), the solvent system was chloroform/methanol, 9:1

DRUG	Non-dried samples						MEAN	STD DEV
Phenylpropanolamine	4	3	3	3	3	4	3.3	0.5
Metoprolol	6	6	7	8	6	9	7.0	1.3
Pheniramine	10	10	9	10	10	11	10.0	0.6
Fluphenazine	30	23	20	27	26	26	25.3	3.4
Haloperidol	25	25	24	29	26	29	26.3	2.2
Promazine	22	23	20	29	26	26	24.3	3.3
Metronidazole	22	20	11	21	17	26	19.5	5.1
Oxazepam	22	20	15	25	20	32	22.3	5.8
Cocaine	27	25	37	16	18	17	23.3	8.1
Chlorprothixene	49	52	58	52	58	50	53.2	3.9
Trimipramine	55	57	58	52	53	50	54.2	3.1
Temazepam	59	61	58	52	58	56	57.3	3.1
Papeverine	74	74	70	68	70	70	71.0	2.4
Clobazam	69	71	65	62	65	67	66.5	3.2
Prazepam	74	75	70	68	67	72	71.0	3.2
Methaqualone	74	76	70	65	65	72	70.3	4.6

Table 13 R_fs from the drug substances using the standard TLC method with dried sample spots

The table displays the substances, reference letter as previous, standard literature R_f values taken from Moffat et al[10], the six sets of raw data, mean and standard deviation. As described earlier (section 3.1), the solvent system was chloroform/methanol, 9:1

DRUG	Standard method						MEAN	STD DEV
Phenylpropanolamine	3	4	3	4	5	3	3.7	0.8
Metoprolol	6	8	4	7	9	7	6.8	1.7
Pheniramine	7	10	4	10	17	9	9.5	4.3
Fluphenazine	14	22	28	25	33	24	24.3	6.3
Haloperidol	23	26	31	25	33	27	27.5	3.8
Promazine	14	24	28	29	33	29	26.2	6.6
Metronidazole	17	22	24	19	25	18	20.8	3.3
Oxazepam	17	21	27	21	28	18	22.0	4.6
Cocaine	19	26	25	21	28	24	23.8	3.3
Chlorprothixene	49	42	50	55	54	58	51.3	5.6
Trimipramine	56	47	52	58	57	59	54.8	4.5
Temazepam	60	58	62	60	59	51	58.3	3.8
Papeverine	79	70	70	71	69	72	71.8	3.7
Clobazam	72	69	69	69	68	65	68.7	2.3
Prazepam	78	72	71	73	72	72	73.0	2.5
Methaqualone	79	74	75	71	74	72	74.2	2.8

The mean data from the two tables above was plotted in Figure 27

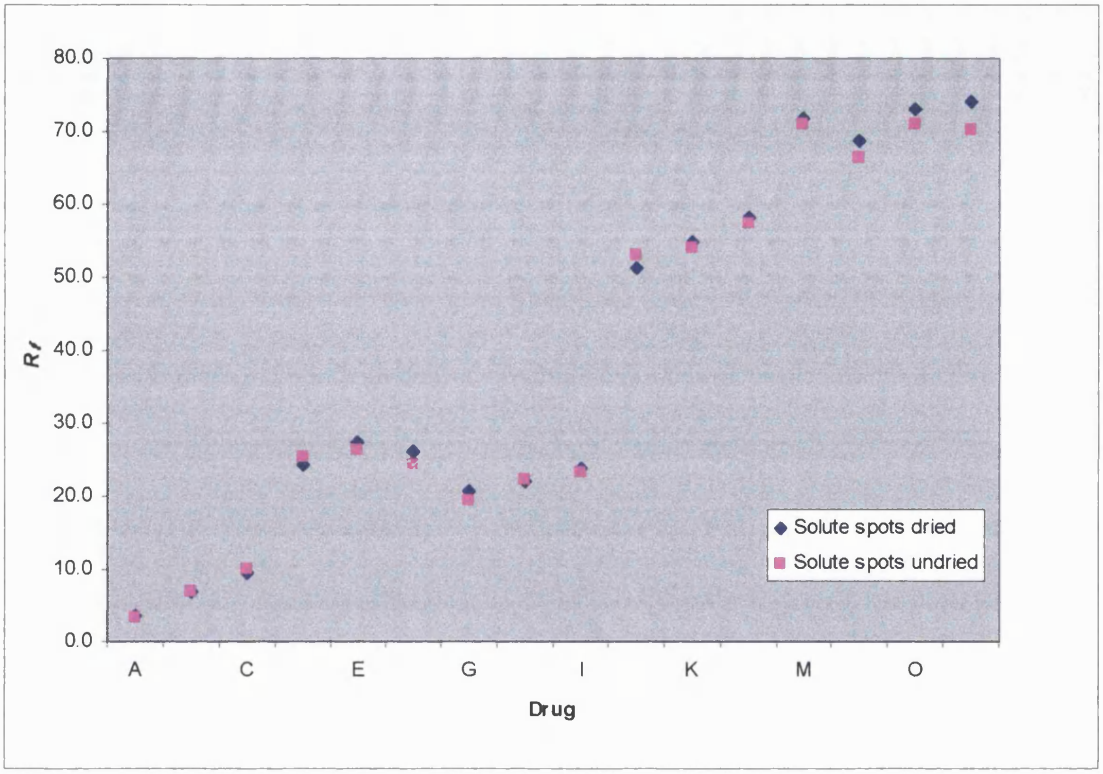


Figure 27 The effect of "Wet" or non-dried solute samples upon the R_f values of pharmaceutical substances.

Key: A-phenylpropanolamine, B-metoprolol, C-pheniramine, D-fluphenazine, E-haloperidol, F-promazine, G-metronidazole, H-oxazepam, I-cocaine, J-chlorprothixene, K-trimipramine, L-temazepam, M-papaverine, N-clobazam, O-prazepam, P-methaqualone

The mean R_f data between the two sets of conditions were fairly similar, in that there were no significent differences. However, both sample sets displayed a higher degree of variability with standard deviations within the non-dried TLC runs as high as 8.1 and within the control set as as high as 6.6.

It is probably a little surprising that more substances were not affected as even a small amount of chloroform remaining in the sample solution and having not been evaporated off would automatically alter the balance of the solvent/ sample/stationary phase interaction.

As with other experiments in this set, the results did not demonstrate a major

difference between the test and the control condition perse, but did highlight the need to control the conditions and decrease the variables as much as possible in a bid to decrease the variability of the results produced.

3.3.7 The effect of solvent chamber saturation

A saturated chamber is defined as one that has had mobile phase placed into it up to a level of 1 cm, a lid placed over the chamber with mobile phase able to saturate the chamber for 45 minutes, prior to introducing the TLC plates. This allows the solvent to come to equilibrium in the chamber.

In this experiment, the term 'unsaturated' is used to describe a chamber into which mobile phase is added and the TLC plate placed immediately after.

Six sets of TLC plates were run in unsaturated development chambers while six further sets were run in the standard (control) saturated conditions.

The results obtained (Figure 28) produced a difference in the behaviour of two mid range R_f drugs only, metronidazole and oxazepam. These two drugs are also commonly the ones that have displayed variation in other experiments and have been quite difficult to contol, from a reproducibility perspective.

Providing a saturated chamber would allow an even distribution of solvent vapour across the surface of the TLC plate and ensure that the chemical reactions taking place between each sample spot and the solvent were at an equal concentration. This must be the ideal scenario.

Despite not being evident in this experiment, unsaturated chambers have the potential to display high variation between samples from run to run. An unsaturated chamber is likely to contain an unstable incomplete interaction between air, solvent and stationery phase in the chamber at the same time as

the interaction with the sample. This could not be regarded as a controlled environment and comparison for identification purposes would have to be questioned.

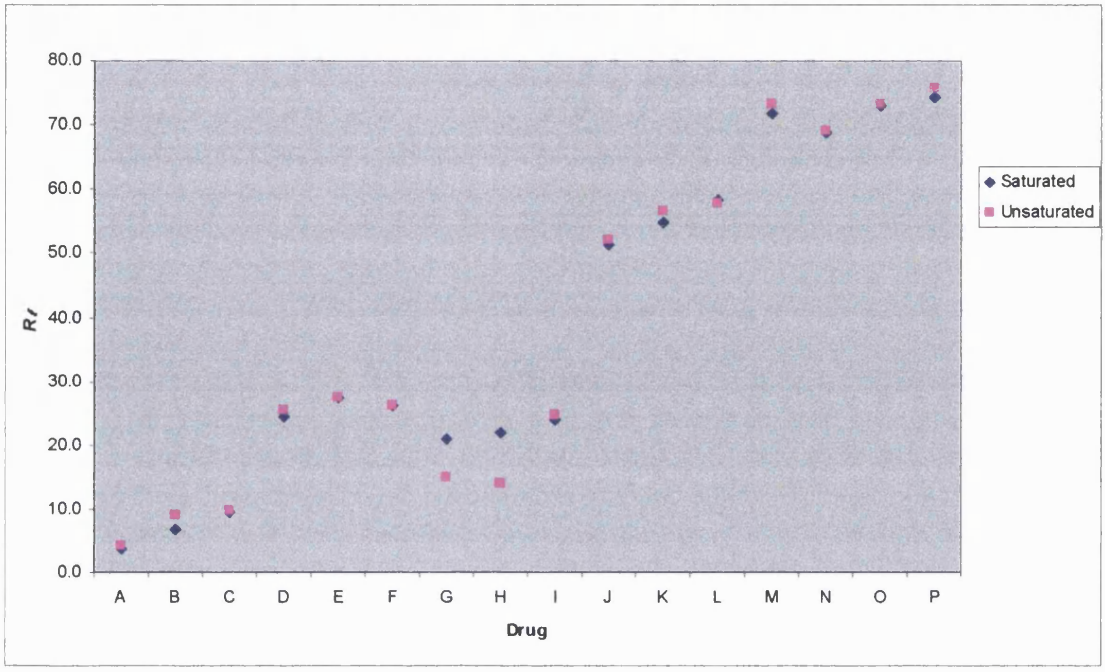


Figure 28 The effect of chamber saturation upon the R_fs for pharmaceutical substances.

Key: A-phenylpropanolamine, B-metoprolol, C-pheniramine, D-fluphenazine, E-haloperidol, F-promazine, G-metronidazole, H-oxazepam, I-cocaine, J-chlorprothixene, K-trimipramine, L-temazepam, M-papaverine, N-clobazam, O-prazepam, P-methaqualone

3.3.8 The use of non-freshly prepared (old) solvents

The definition of a freshly prepared solvent is where the solvent is mixed (e.g. chloroform and methanol) on the day of the experiment and is placed in a development chamber that has been thoroughly cleaned and rinsed with the solvent mix. The solvent will be used for one TLC run and then washed away for fresh solvent to be used as described above.

Non-freshly prepared solvent are used on more than one occasion. For this experiment the same development chamber was used for the 'non-freshly prepared solvent' condition. There were six sets of TLC plates and the solvent

was not changed between the six runs. The control condition was freshly prepared solvent and chambers.

The use of solvents not freshly prepared provided variation in R_f values for a series of drugs towards the middle of the R_f range with fluphenazine to cocaine showing the greatest differences (Figure 29).

As the unsaturated chamber is re-used and is opened between runs the composition of the multi-component mobile phase will change. Solvent will leave the chamber between each run and the concentration of solvent will decrease for each subsequent run. Another impact on the mobile phase is that the polarity may well have change with certain components of a multi-component mix dispersing at greater rates than the other components.

One of the components will be preferentially absorbed by the sorbent and the moisture content within the system is likely to change as well. The complexity of the interactions taking place therefore becomes complex and difficult to control, certainly mobile phases with more than one component should only be used once.

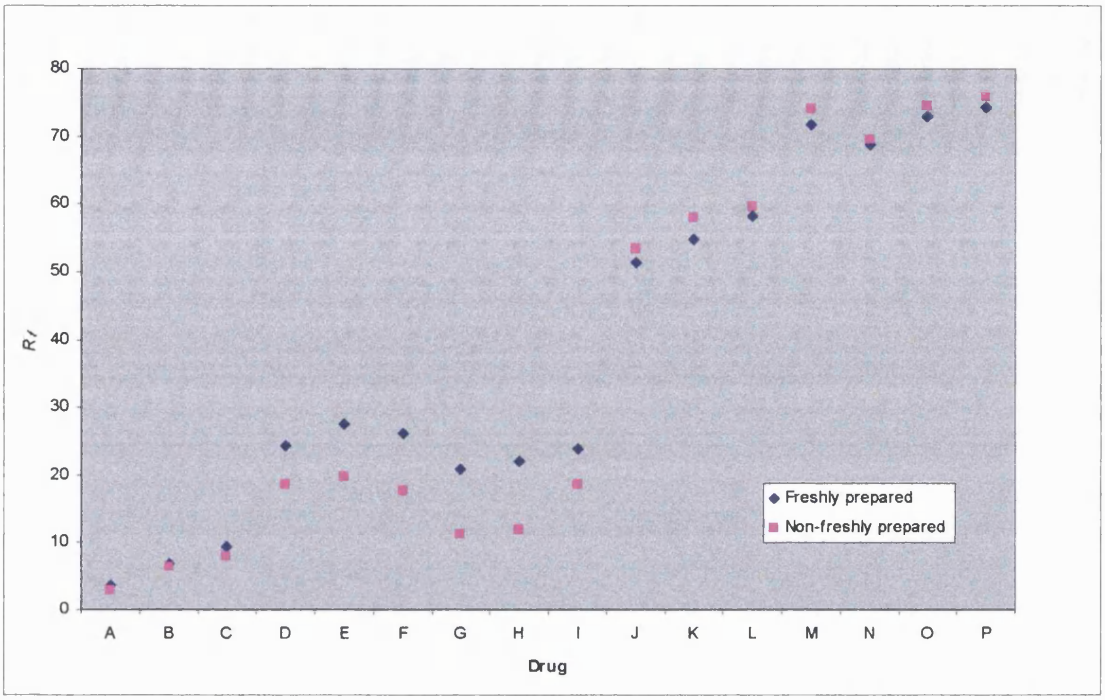


Figure 29 The effect of using old solvents compared to freshly prepared solvents.

Key: A-phenylpropanolamine, B-metoprolol, C-pheniramine, D-fluphenazine, E-haloperidol, F-promazine, G-metronidazole, H-oxazepam, I-cocaine, J-chlorprothixene, K-trimipramine, L-temazepam, M-papaverine, N-clobazam, O-prazepam, P-methaqualone

3.3.9 Chemical reference standards

Stead et al[9] and Moffat et al[10] have described not only the analysis of a large number of substances by TLC but also methods to standardise TLC systems. One of these methods employs the use of internal reference standards to account for the variation in day to day running conditions for a system within any given laboratory. An internal reference standard is a mixture of four substances that have been predetermined because of their ability to produce round and distinct spots and be equally spaced across the chromatographic range. Their purpose is to be run at the same time as the test samples and be used to convert the practically obtained R_f values to corrected ones by a graphical method. Four reference compounds are used because correction graphs are often found to be non-linear. The reference standards selected for

each TLC system (Moffat et al[10]) were run alongside the test samples and were used to produce six-point calibration curves to standardise the values. The curve was obtained by plotting the known literature (or expected) values for the four reference compounds against the experimentally determined values. These four points were taken together with the origin (0,0) and the solvent front (100,100) as shown in Figure 31.

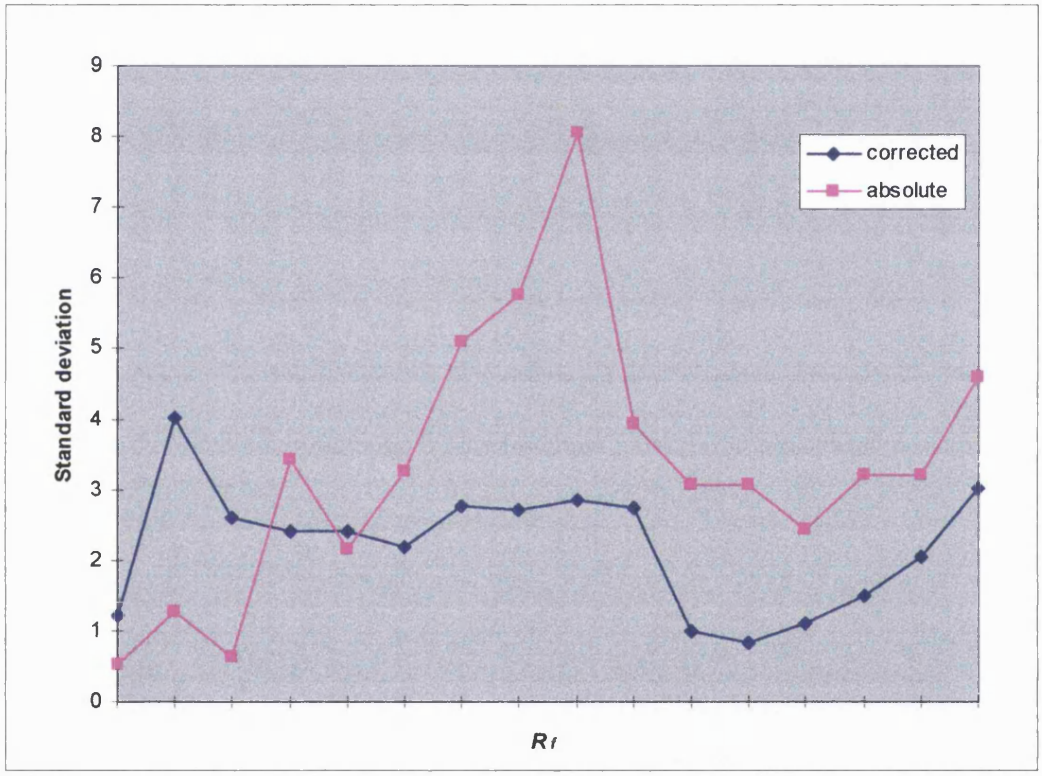


Figure 30 Comparison of standard deviations against R_f value from the 16 test samples each run on six different chromatographic plates.

This is repeated for values that have been corrected after the inclusion of internal reference standards. The standard substances from (Moffat et al[10]) are desipramine, physostigmine, trimipramine and lidocaine. The 16 test samples were *phenylpropanolamine, metoprolol, pheniramine, fluphenazine, haloperidol, promazine, metronidazole, oxazepam, cocaine, chlorprothixene, trimipramine, temazepam, papaverine, clobazam, prazepam and methaqualone*

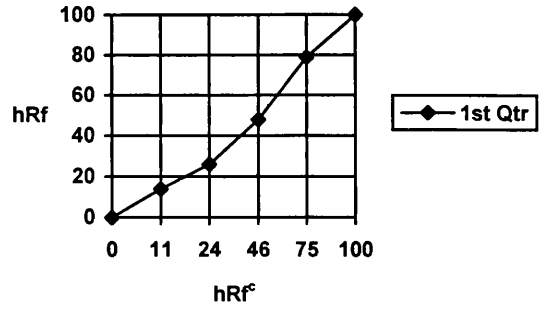


Figure 31 The six point calibration graph

$$hR_f^c(X) = hR_f^c(A) + \frac{\Delta_c}{\Delta}[hR_f(X) - hR_f(A)]$$

where

$\Delta_c = hR_f^c(B) - hR_f^c(A)$ and

$$\Delta = hR_f(B) - hR_f(A)$$

where hR_f = $\frac{\text{distance drug travels from the origin}}{\text{distance the solvent front travels from origin}}$ X 100

hR_f^c = R_f corrected by reference compounds

(A) and (B) are the reference compound R_f values either side of the observed value (X) on the six point calibration curve.

For example, the effect can be witnessed by comparing the variation (standard deviation) in the results for the non-dried solute samples, using absolute (practically obtained) and corrected values (Figure 30). The corrected values were more concistent that the absolute values which varied most towards the centre of the TLC plate. When using corrected values the inter plate variation was clearly less.

Table 14 provides a comparison between the R_f data generated and those given in the database. The data compare well with the literature values. There

were some exceptions namely metoprolol on System 2, cocaine on System 3 and metronidazole – clearly different on both System 3 on System 4.

The reference standards do provide a method of governing some of the factors of variation. While controlling factors such as temperature and spot mass will be still important, the use of internal standards on each TLC plate may reduce the variation.

Table 14 R_f data, the mean of six runs, for ten drugs run on four TLC systems with the literature values in brackets

Drug	System 1	System 2	System 3	System 4
Phenylpropanolamine	47 (44)	4 (4)	4 (4)	33 (31)
Metoprolol	51 (49)	18 (8)	8 (5)	9 (7)
Pheniramine	47 (45)	34 (35)	13 (10)	3 (4)
Fluphenazine	67 (63)	6 (6)	23 (23)	10 (10)
Metronidazole	60 (58)	4 (2)	36 (20)	40 (18)
Cocaine	67 (65)	47 (47)	47 (28)	54 (58)
Trimipramine	63 (59)	63 (52)	54 (57)	37 (39)
Papaverine	60 (61)	12 (8)	65 (69)	47 (50)
Prazepam	70 (65)	36 (36)	74 (73)	63 (64)
Methaqualone	73 (70)	39 (37)	80 (78)	56 (58)

3.3.10 Intra-plate variation

Pharmacopoeial methods of drug identification involve comparison of the test solution with the appropriate chemical reference material. Intra-plate R_f differences within TLC plates were observed by running samples of a known drug side by side on a plate. The mean differences produced by drugs on the same plate ranged from 1.3 up to 3.0 (Table 15) with up to a difference of 8 between the maximum and minimum for desipramine. Despite the fact that samples of the same drug (from different source) were being used, differences in R_f were being produced.

Table 15 Intra-plate variations demonstrating reproducibility in R_f over 18 runs for four drugs

Drug	Maximum	Mean difference	Minimum
Desipramine	32	3.0	24
Dipipanone	57	1.8	50
Caffeine	66	1.3	62
Meclozine	82	1.8	78

The question arises as to when an R_f is "identical" or "similar" to another as stated in the British Pharmacopoeia and the International Pharmacopoeia. As there is such intra-plate variation it will be difficult to set rigid criteria whether these are quantitative or not, unless the variation can be taken into account.

3.4 Retrieving substance identities using stored R_f values

3.4.1 Fixed window method

The window method is used commonly to identify compounds by use of a database. The fixed window approach sets out pre-determined categories within which the generated data for a drug is fitted. The size of the window is determined by the inter-laboratory standard deviation for a given solvent system (Stead et al[9]).

For System 1, the window is ± 7. Therefore, the window bands are set as 0-14, 15-28, 29-42, 43-56 etc. Taking phenylpropanolamine as an example, the R_f on system 1 was found to be 47; only substances within the 43-56 band of the database would be considered as possible identities. The results of this would then be carried in turn across the remaining three systems to produce a list of

possible outcomes. The search is conducted taking into account all four systems the result being that from the 594 drugs in the database, 3 possibilities are suggested (Table 16).

The advantage of using such an approach is that a large database can be searched with speed and effectiveness and it can be done manually. A major disadvantage is the "outlier", a substance falling just outside the border and being rejected. The pre-determined window makes no distinction between any drug R_f falling inside a band. For example, pheniramine having a database R_f of 45 on System 1 and hence falling in the 43-56 band; if an R_f of 42 had been produced this would now fall outside the band and be rejected despite this value being merely 3 units away from the database value. However, R_f values up to and including 56 (11 units away from the "true" value) would still be included as possibilities. There is no distinction between values at either end of the band and those in the middle of it. Another problem is where there is no R_f value for a drug in one of the solvent systems in the database as this is automatically interpreted as a zero result by the search which discards the drug from the search and regardless of the accuracy of the remainder of the data, the drug will never be found.

Table 16 Comparison of searches for ten drugs using Fixed and Moving windows

The summary is for ten drugs using data from the four solvent systems. The obtained results were compared against published literature database values and were based upon window methods of identification

Drug	Fixed window		Moving window	
	Possibilities	Drug included	Possibilities	Drug included
Phenylpropanolamine	3	YES	3	YES
Metoprolol	4	NO	6	NO
Pheniramine	7	YES	8	YES
Fluphenazine	8	YES	9	YES
Metronidazole	11	NO	17	NO
Cocaine	0	NO	0	NO
Trimipramine	8	YES	8	YES
Papaverine	5	YES	4	YES
Prazepam	17	YES	17	YES
Methaqualone	16	YES	13	YES

3.4.2 The Moving window

This is similar to the fixed window, the difference being that the window's bands are set around the observed R_f value as opposed to the R_f being made to fit into pre-determined bands. Using the example of phenylpropanolamine in system 1 (Table 9), the window would be 47 ± 7 (between 40 and 54). From Table 16, phenylpropanolamine again produced 3 possibilities from the 594 drugs with the drug included as one of these, whereas metoprolol was again excluded based upon the result for system 2 still falling outside the window. The advantage of this method over the fixed window approach is the elimination of the problem outlier, but the drawback of missed results, a disadvantage of both window methods, still remains. Additionally, the search result i.e. a list of possibilities produced are not ranked, there being no differentiation between the outcomes, merely a list of possibilities. To overcome this, statistical methods are required to rank possible identities in an ordered fashion.

3.4.3 The Discrepancy Index (DI)

The DI search gives a quantitative estimate of the match comparing the generated data with those in the database. The technique has been based upon Discriminating Power described by Smalldon, Moffat and co-workers[98].

$$DI = \sum \left(\frac{X - \mu}{\sigma} \right)^2$$

Equation 25 Equation for the Discrepancy Index Search.

where X - observed value, μ - expected value, σ - mean standard deviation of measurement of R_f values in that system

The DI is based upon the chi-squared distribution and can be limited to those values below a critical chi-squared value i.e. for $p>0.01$ given as a normalised probability. From Table 17, it can be seen that phenylpropanolamine has $p=0.836$ and is ranked as the highest of the 3 possibilities from 594 drugs. Metoprolol produces a probability of identification of 0.013 whereas with the window searches it was discarded on the basis of the one poor result. Despite possessing a relatively low probability it was still ranked 6th in the entire database. Cocaine, another example of a drug not appearing the window search results, was not suggested by the DI search because $p<0.01$, but a probability value was produced. The advantage of the DI system is the quantitative estimate produced and the ability to rank the data in descending order of probability. Missing values are handled by requiring the analyst to compare manually the results obtained for all four systems with the search while omitting the missed system.

Table 17 Summary of the Discrepancy index and Mean List Length searches for ten drugs, based upon R_f data from four different solvent systems compared against published values from the database of 594 basic drugs

	Discrepancy index			Mean iist length		
	Probability	Rank	Possibilities	Probability	Rank	Possibilities
Phenylpropanolamine	0.836	1st	3	0.486	1st	3
Metoprolol	0.013	6th	10	<0.01	6th	6
Pheniramine	0.797	1st	10	0.435	1st	5
Fluphenazine	0.777	1st	11	0.411	1st	5
Metronidazole	<0.01	-	20	<0.01	-	14
Cocaine	<0.01	-	0	<0.01	-	0
Trimipramine	0.504	1st	9	0.189	1st	8
Papaverine	0.168	2nd	7	0.034	2nd	4
Methaqualone	0.558	3rd	22	0.223	4th	15
Prazepam	0.675	1st	28	0.311	1st	17

3.4.4 The Mean List Length (MLL)

The MLL is approximated by the probability density function of the normal distribution (Schepers et al[99]).

$$P = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(X-\mu)^2}{2\sigma^2}}$$

Equation 26 Equation for the Mean List Length Search

where X - observed R_f
μ - expected R_f
σ - standard deviation

In TLC it is assumed that the R_f value of a substance will vary from day to day following a normal gaussian distribution with the listed value (or literature value) as the mean and with a system dependant standard deviation[10]. Producing an R_f from an unknown gives a R_f value.

Each substance in the database is therefore a potential match to this value. The closer the listed value of a substance in the database to the produced R_f, the higher the probability that the substance matches (up to a 'perfect' match of

1.0). The further the listed value from the R_f, the lower the probability. The mean and standard deviation are used to ensure all substances utilise the same factor of variation for each solvent system. Taking one solvent system, it will be possible to produce a listing of the probabilities of match of each substance in the database to the produced R_f value. After normalising the list so that the sum of the probabilities becomes 1, the substances can be ranked in decreasing order of probability, with the highest probabilities at the top and a limit set to those at the bottom. For example if the limit was set to a cumulative probability of 95 %, this would give the list length - which shows which substances qualify for identification at the 95 % confidence level. The procedure can then be applied to other R_fs in the database to produce a mean list length.

In the experiments reported here, the mean list length was calculated for a combination of four solvent systems.

Metoprolol was run on four different solvent systems and produced four mean R_fs. For each solvent system the probabilities were calculated to the 95 % confidence interval and the list lengths from each one were combined to produce a mean list length. From Table 17, metoprolol was found to have been ranked 6th out of the 594 drugs, so it was not actually identified correctly, whereas phenylpropanolamine gave 3 possibilities and a probability of 0.486 which was ranked the highest of the candidates. This was identified successfully. Further results in Table 17, for phenylpropanolamine and fluphenazine for example show there were successes, however there were failures as well. The method was not totally robust but this is based upon a

single system. Combining results from further solvent systems would increase the accuracy of the method.

The statistical methods are shown to provide a quantitative response to the search and demonstrate that results can be positive, however the variation in the R_fs produced compared with the listed values makes the application of a quantitative rule difficult to apply.

3.5 Conclusions

Firstly, the need to standardise the running conditions is extremely important when carrying out the experiments. Secondly, rigid criteria can not be set in order to base identification. Finally, search routines based upon probability provide the optimal method of identification using TLC data.

Because the high degree of changes made in the experiments were beyond the capabilities of the TLC systems used, i.e. the systems were not robust enough to sustain the changes made, standardised conditions need to be used.

The standardised conditions I would recommend are:

- A good sized TLC plate, at least 20 cm by 20 cm. The Whatman Al Sil G/UV_{254} plates of 250 µm thickness provide an excellent ready-made solution.
- A run distance of 10 cm, which allows spots to clearly separate and be well defined.
- A spot mass of 10 µg provided by a single spot from a 5 µl spotting pipette. The spot should be dried prior to addition of the plate to the TLC chamber.

- Temperature of 20 °C to be controlled and monitored throughout. Plates must be kept away from direct sunlight or other factors liable to alter the temperature.
- The use of internal reference standards on each TLC plate to correct for intra plate variation.
- A saturated TLC chamber brought to equilibrium for one hour prior to adding the TLC plate. A lid must be used to keep the chamber sealed throughout the course of the experiment.
- Solvents should be freshly prepared and when added to the TLC chamber should be at a level lower than the spotting baseline of the TLC plate.
- The TLC plate should be as near to vertically standing as possible in the TLC chamber.

For the search routine I would recommend the probability based searches. The reasons are outlined below:

- Fixed windows – an efficient search method but rigid and inflexible. It does not take account of position of drug within a band or standard deviation of measurement. When run across multiple solvent systems, the lack of an individual R_f value on any system is regarded as a zero R_f value.
- Moving windows – has the advantage over the fixed window by using a pre-determined ± value around the expected result. This pre-determined value is still independent of standard deviation of measurement, missing

results are still treated as zero and a list of drug possibilities is presented without any form of ranking.

- Mean list length – provides a quantitative response to the search and is approximated by the probability density function of the normal distribution. The ranked probabilities of match provide a very good indication to the actual substance but did not work in all cases.
- Discrepancy index – produces a probability of match based upon the chi-squared distribution and provides a ranking of candidates. The method was found to be equally as good as the mean list length but again did not work in all cases.

For the criteria to be used for identification, mean list length and discrepancy index give a ranking in order of probability. It is difficult to set absolute criteria for all matches but the use of the probability methods in conjunction with standard conditions would improve the current standards.

Chapter 4 Sample presentation for NIR spectroscopy

4.1 Summary of aims

- To determine if a basic parameter setting such as the number of instrumental scan passes exerts any influence upon the reproducibility of the results.
- To investigate if there are any spectral effects caused by examining different portions of the powdered sample or tablet sample.
- To determine if sample grinding can affect the resultant spectra.
- To investigate any influence upon the spectra caused by the material from the glass sampling vials.
- To investigate the relationship between vial base diameter size and reproducibility of NIR spectra.

To investigate the influence of particle size upon the reproducibility of raw and second derivative spectra and upon identification from a database.

4.2 Method

The practical methodology followed can be found in the "Experimental" chapter (section 2.3).

4.3 Results and Discussion

4.3.1 Instrument Scans

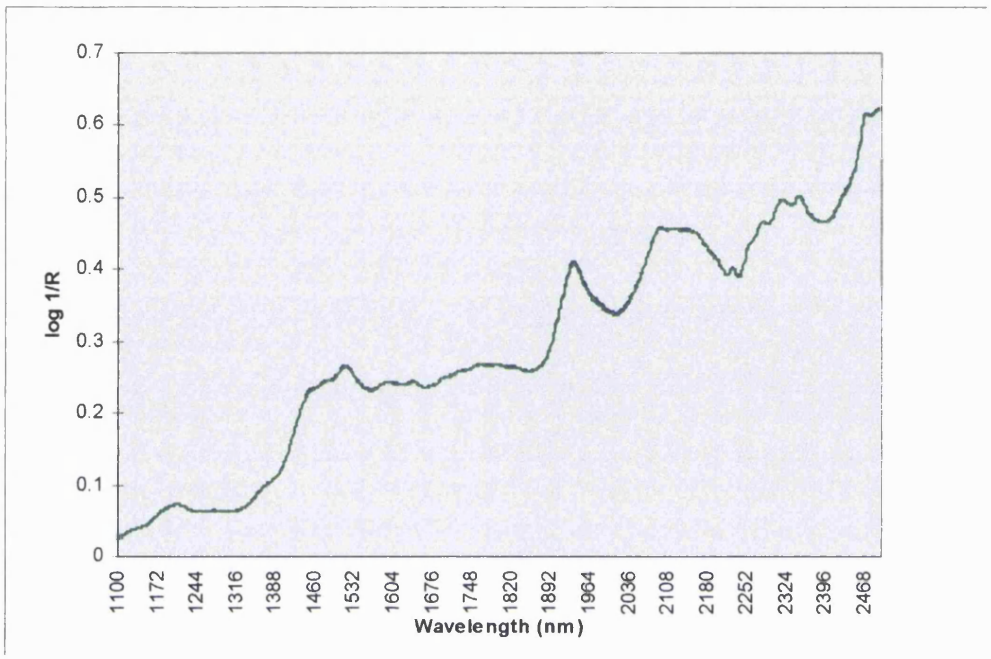


Figure 32 Scan passes for 4, 16, 32 and 64 scans

Each series of scan passes is represented by 12 spectra. Samples are allopurinol 100 mg tablets BP

Each spectrum produced was the mean of a certain number of scan passes. The greater the number of scans, the greater the time taken to produce each spectrum. The standard instrument (and industry) setting is 32 scans to produce one reported spectrum. The sample was not moved between scans.
Having varied the number of scans passes between 4 and 64 scacs,there appeared to be no increase in noise or decrease in reproducibility with decreasing the number of scans. There appeared to be no impact upon the spectra produced by varying the number of scan passes between 4 and 64 scans. This was most likely to be because the spectra needed to be analysed

to a greater level of detail than was used in the experiments, probably down to a range of 50-100 nm. O'Haver[81] suggests that increasing the number of scans from 1 through 4, 16, 64 would improve the signal to noise 1, 2, 4, 8 fold respectively.

For the purposes of this practical work the default number of scans will be 32, the commonly used value in the pharmaceutical industry[80].

4.3.2 The effect of sample packing of a powder

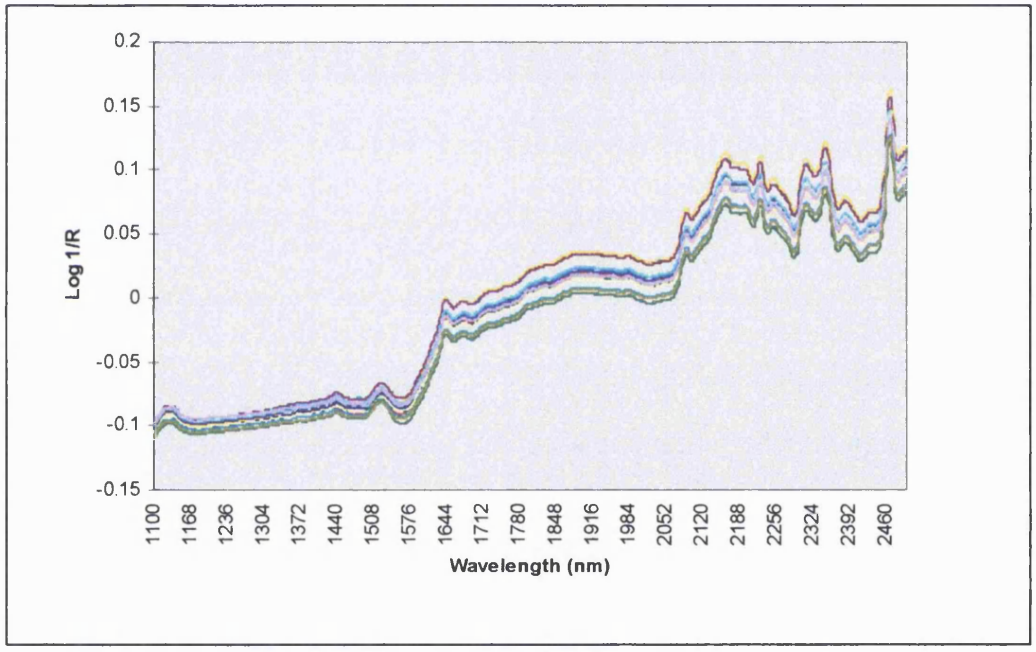


Figure 33 Sample packing effects upon allopurinol powder

The example show packing effects caused by shaking same the sampling vial between measurements of the allopurinol powdered pure drug. Spectra between 1100 and 2500 nm

The effect of shaking the powdered drug sample between each measurement can be seen in Figure 33. There are differences in the spectra which are not apparent when repeatedly measuring the powder and sampling vial held in the same position. By removing the sampling vial from the instrument and shaking

the contents, different areas of the powder are being examined and different portions of the powder sample surface are brough intpo contact with the light. This has led to greater light scattering.

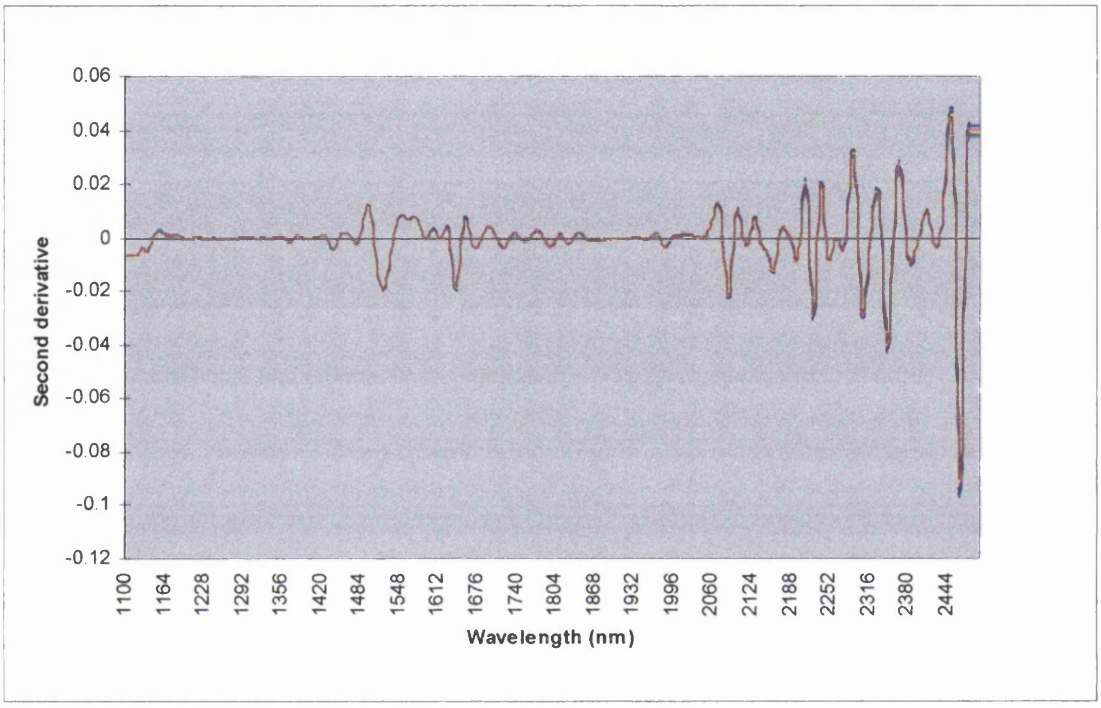


Figure 34 Sample packing effects – Second derivatives

The example shows packing effects caused by shaking the sampling vial between measurements of the allopurinol powdered pure drug. Second derivative spectra are shown between 1100 and 2500 nm

A full spectral shift of this nature is caused by the re-orientation of the way the powder particles are packed and is due to the non-regular shape of such particles and changes in bulk density. Irregularly dispersed particles reflect light at different angles and produce a different response on the detector.
This is a consequence of sampling that will have to be considered when comparing results within all other experiments. The effect produced does not

affect the actual peak position in terms of wavelength but does have a bearing on the intensities and relative intensities of all peaks.

To take this effect into account when constructing spectra to be stored in the database, all samples should be shaken in between measurements and sufficient measurements should be made to take account of these effects. Physical effects such as this can be reduced by applying the mathematical second derivative to the spectra. A description of derivatives is contained within section 1.10.1. A second derivative plot of the spectra is contained within Figure 34. The effects evident in the spectra are no-longer noticeable as the peaks at a particular wavelength now appear to the same order of magnitude and there is no variation in the baseline. The same principle can be applied to tablets. Measurements of tablets rotated between measurements and sampled on both sides produced variation between the recorded spectra (Figure 36).

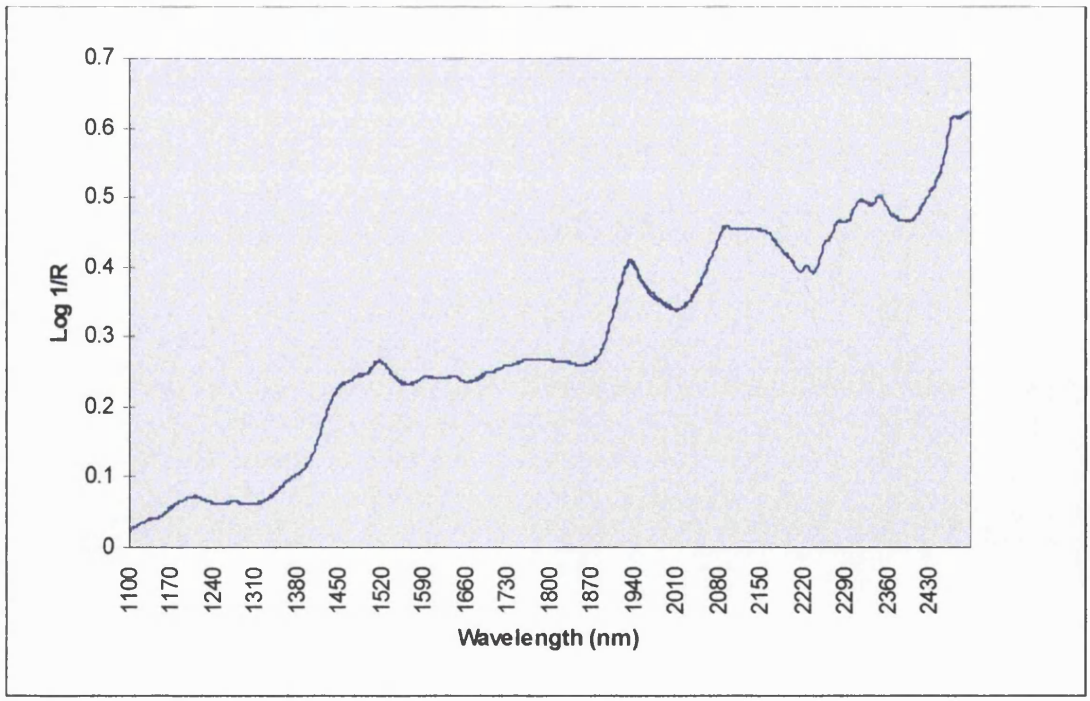


Figure 35 Twelve spectra from one allopurinol 100mg tablet BP, held in one position

Spectra produced between 1100 and 2500 nm

Areas of variation include the curvature of the tablet surface, the packing within the tablet and surface effects such as scratching, scoring and embossing. These effects have occurred because the portion of the tablet directly in contact with the incident ray has varied. This spectral variation is not present when sampling the same portion of one tablet repeatedly (Figure 35).

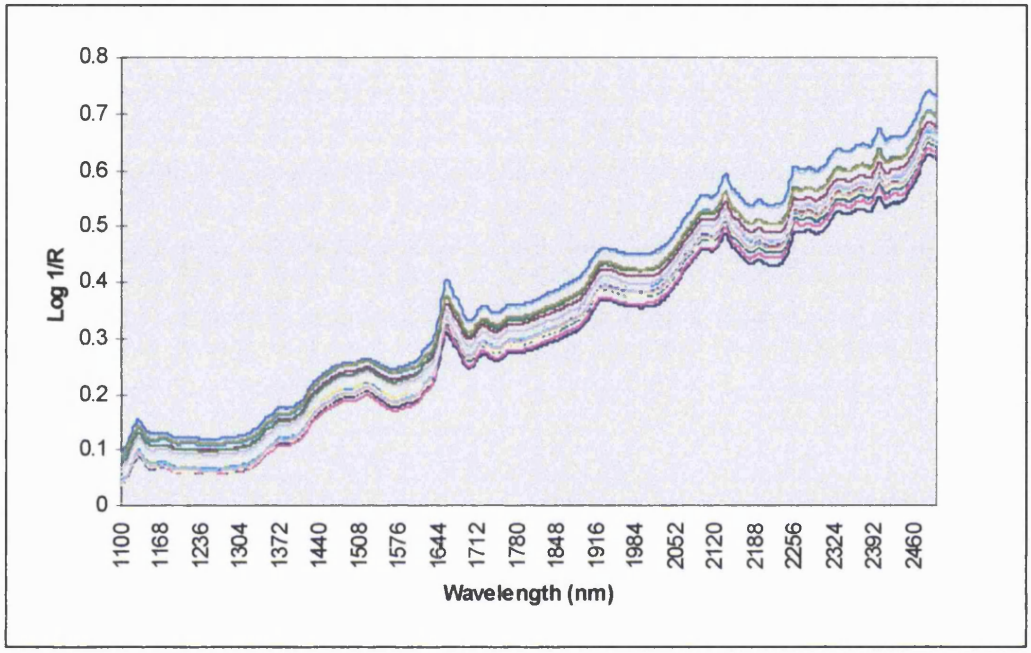


Figure 36 Aspirin 300 mg caplet, 10 rotations on each side of the caplet.

Spectra between 1100 and 2500 nm

The effect of varying tablet position can be seen by measurements taken of one aspirin tablet (caplet). The tablet is capsule like in shape and contains a curved surface with scoring on one side of the caplet. The variation in spectra produced (Figure 36) is of the same type to that noted when shaking powdered samples between measurements. A spectral shift is noted but the magnitude of each individual spectrum is the same - the baseline has shifted, not the peak heights or peak wavelength positions. The detector iris was positioned in the middle of the sampling platform with the iris diameter being less than the

diameter of the tablets being sampled. Stray light as a result of the tablet not covering the iris was not thought to have contributed to the light scatter.

As with the powdered drugs the spectral shift effect was reduced by using second derivative plots of the spectra (Figure 37). Experiments in this thesis will attempt to build the maximum amount of variation possible for each sample into the spectral libraries.

Measurements to produce pure drug spectra will therefore continue to incorporate the shaking of sample vials between each measurement and tablets will be rotated between measurements and sampled on both sides of the tablet.

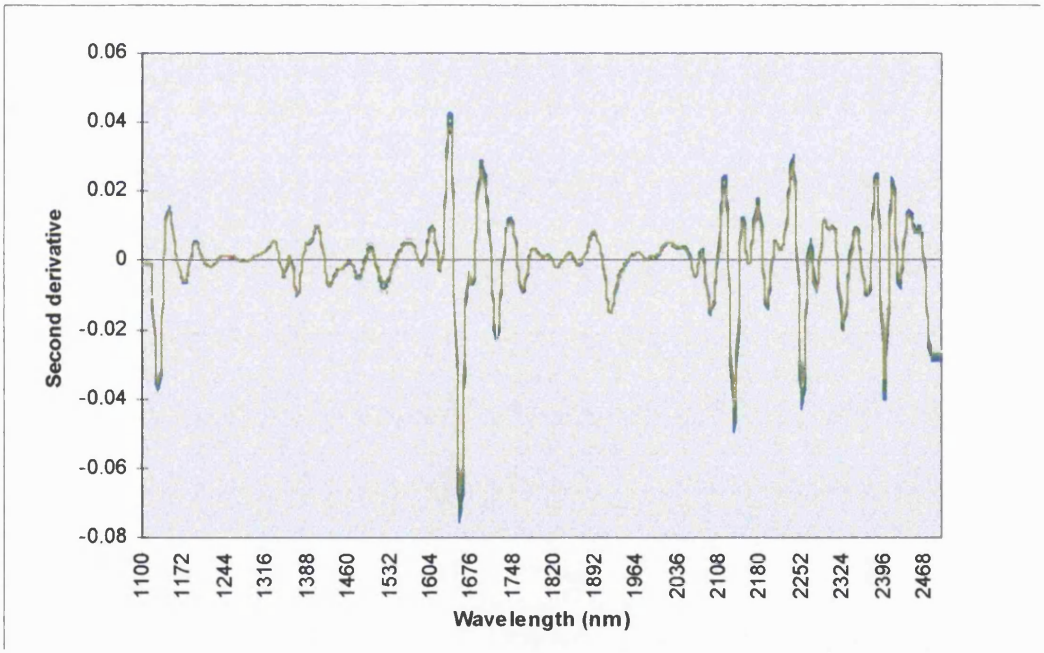


Figure 37 Aspirin 300 mg caplet, 10 rotations on each side of the caplet.

Second derivative spectra between 1100 and 2500 nm

It would appear that where differences in chemical characteristics are important, for example when seeking to identify samples, second derivative spectral matching is likely to be more successful.

4.3.3 The effects of grinding powders

The purpose of this work was to be able to gain an understanding of the types of effects that do exist based upon grinding and in terms of particle size and how these would manifest themselves and what impact this might have on further work in this thesis.

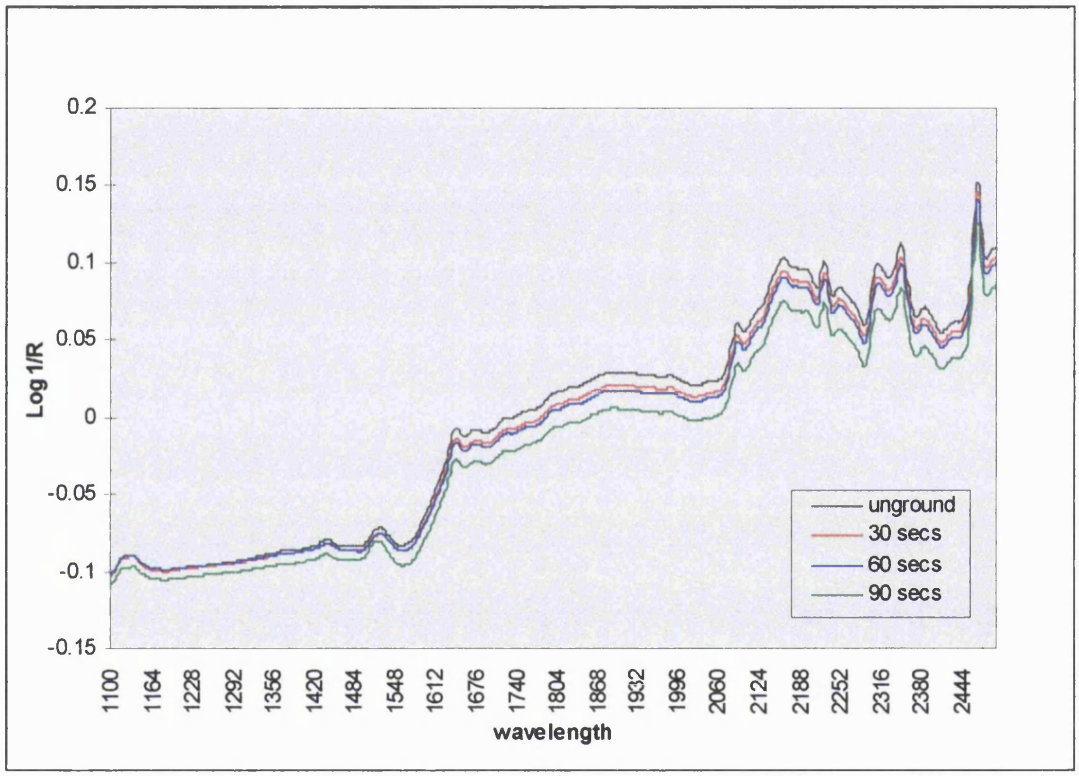


Figure 38 Powdered allopurinol pure drug after various stages of manual grinding

Amount of grinding is represented by time of taking aliquot from grinding vessel. Spectra between 1100 and 2500 nm

Grinding was performed by taking approximately 5 g of allopurinol pure drug, placing in a mortar and by grinding with a pestle in a regular manner. Aliquots were removed prior to grinding and at 30 second intervals up to 90 seconds. Aliquots were transferred to glass sampling vials and labelled. Having collected

the necessary aliquots, NIR measurements were taken. The effects upon the pure drug in the form of original spectra are displayed in Figure 38.

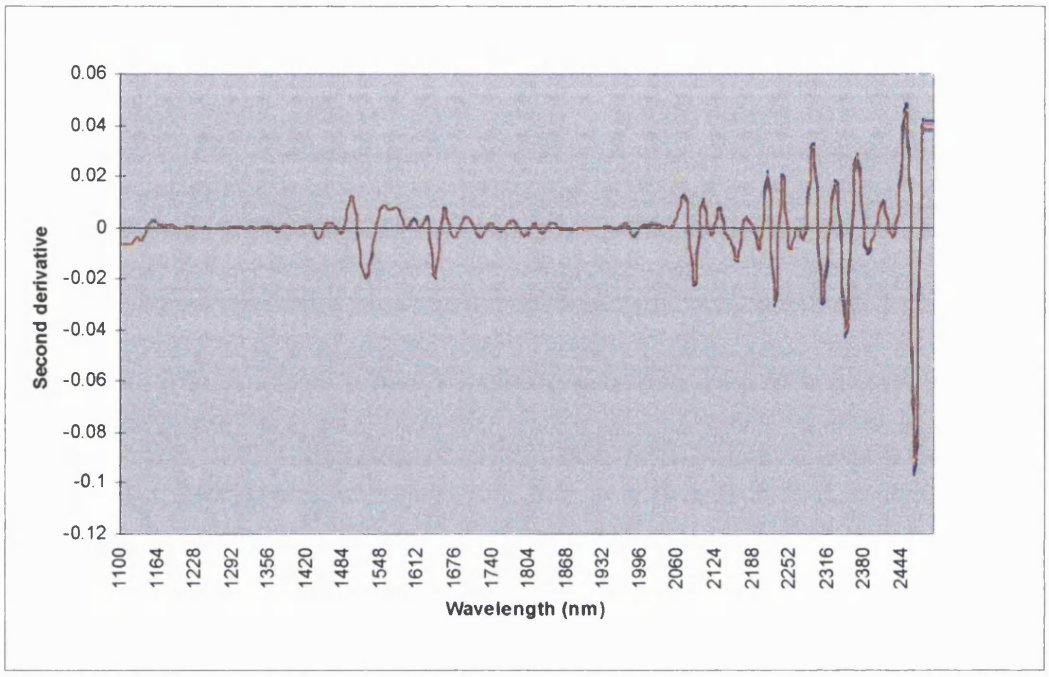


Figure 39 Powdered allopurinol pure drug after various stages of manual grinding

Amount of grinding is represented by time of taking aliquot from the grinding vessel. Second derivative spectra between 1100 and 2500 nm

Firstly it is noticeable that there is variation between each spectrum which at first inspection is similar to that produced by shaking the samples between each measurement as reported in section 4.3.2. However, upon more detailed inspection it is noticeable that there is a relationship between the spectral shift and the time of sampling the aliquot. Grinding can be seen to exert an influence upon the original spectra in terms of a negative baseline shift. Peak wavelength positions are unaffected.

Particles are becoming crushed and distorted and are changing in particle size and shape. This effect is reduced when referring to second derivative plots (Figure 39) as the spectra are now indistinguishable from each other.

The ability of NIR spectroscopy to distinguish between chemical and physical characteristics of compounds is well known as are the benefits of referring to higher order derivatives to overcome the physical effects, which is confirmed here. Also, the distinction and identification of particle size has also been investigated with different chemometric methods[96, 100, 101, 102].

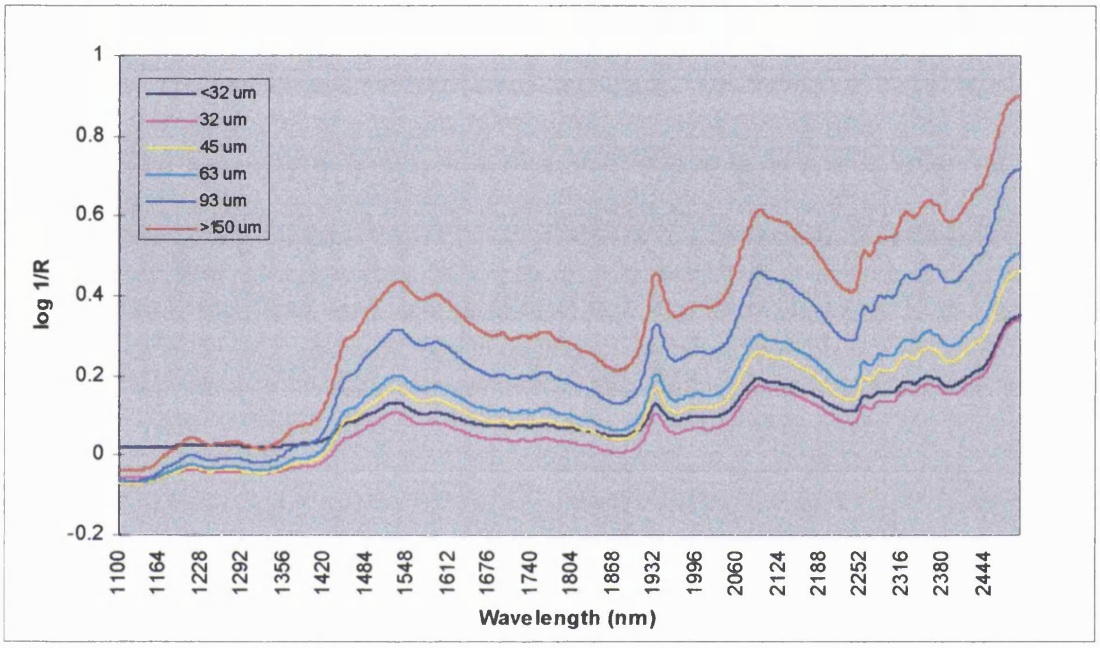


Figure 40 Spectra of different particle size samples of lactose monohydrate

Spectral range is between 1100 nm and 2500 nm

To determine the impacts of different particle sizes, sieve fractions of lactose monohydrate were analysed by Malvern Forward Angle Laser Light Scattering Diffraction. This provided sample aliquots of different mean particle sizes. The particle size fractions obtained were <32 µm, 32 µm, 45 µm, 63 µm, 93 µm, >150 µm.

Table 18 Correlation spectral match results for sieved and unsieved lactose monohydrate compared against the database of pure drug spectra

Spectral range examined was 1100 nm - 2500 nm

	<32 µm	32 µm	45 µm	63 µm	93 µm	>150 µm	Next best
<32 µm	1.000	0.948	0.975	0.983	0.990	0.990	0.940 Tetracycline
32 µm	0.948	1.000	0.992	0.983	0.962	0.941	0.805 Tetracycline
45 µm	0.975	0.992	1.000	0.998	0.989	0.976	0.868 Tetracycline
63 µm	0.983	0.983	0.998	1.000	0.996	0.987	0.894 Tetracycline
93 µm	0.990	0.983	0.989	0.996	1.000	0.997	0.928 Tetracycline
>150 µm	0.990	0.941	0.976	0.987	0.997	1.000	0.950 Tetracycline
Unsieved	0.965	0.994	0.999	0.996	0.984	0.968	0.852 Tetracycline

Spectra are displayed in Figure 40. What is unclear is the impact this would have on identification of substances. To determine this, the samples of the different lactose batches were compared against the pure drug spectra and second derivative databases. (Details of constructing the databases can be found in 2.2.2 and 2.3.4 while the use of databases for pure drug matching is described in Chapter 5.) From the matches using the databases recognition of particle size batches was high (Table 18).

Table 19 Correlation spectral match results for sieved and un-sieved lactose monohydrate

This was compared against the database of pure drug second derivative spectral plots. Spectral range examined was 1116 nm - 2484 nm

	<32 µm	32 µm	45 µm	63 µm	93 µm	>150 µm	Next best
<32 µm	1.000	0.998	0.993	0.990	0.984	0.977	0.531 Orciprenaline
32 µm	0.998	1.000	0.998	0.997	0.993	0.987	0.532 Orciprenaline
45 µm	0.993	0.998	1.000	1.000	0.998	0.994	0.533 Orciprenaline
63 µm	0.990	0.997	1.000	1.000	0.999	0.996	0.535 Orciprenaline
93 µm	0.984	0.993	0.998	0.999	1.000	0.999	0.535 Orciprenaline
>150 µm	0.977	0.987	0.994	0.996	0.997	1.000	0.538 Orciprenaline
Unsieved	0.991	0.997	1.000	1.000	0.999	0.996	0.533 Orciprenaline

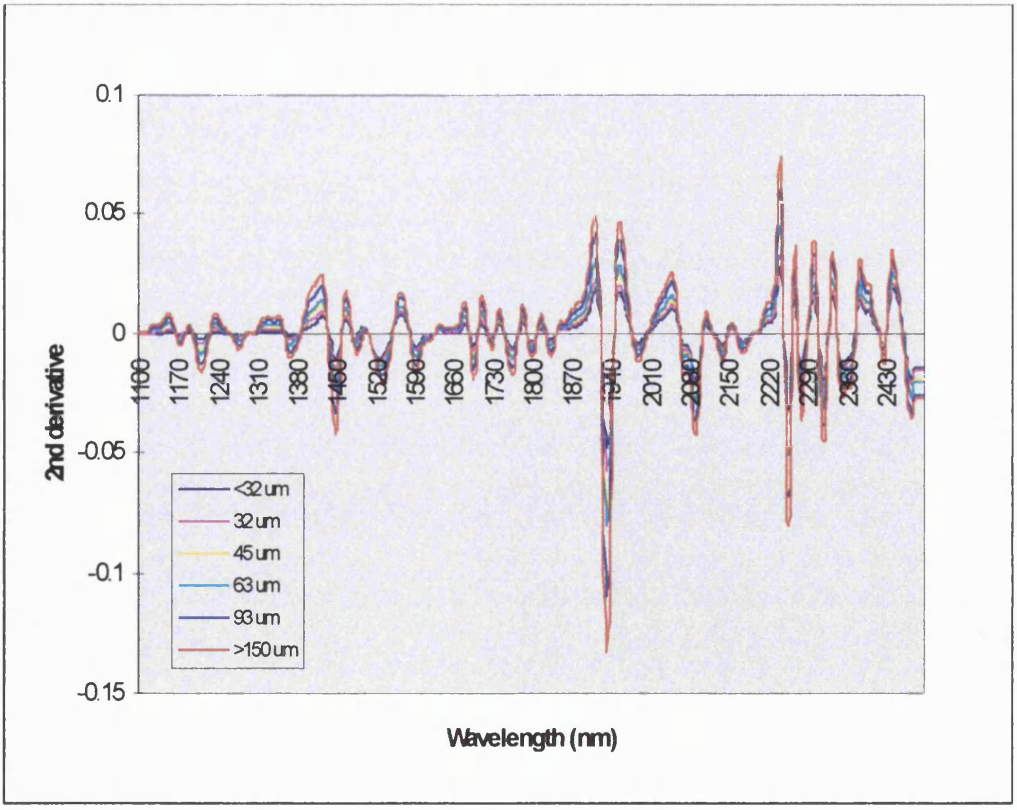


Figure 41 Second derivative spectral plots from different particle size samples of lactose monohydrate

Spectral range examined is between 1100 nm and 2500 nm

Matches were very good, but differentiation between the next best match (tetracycline) was not good, i.e. selectivity was not high; the next best match had an unacceptably high match value. When reverting to the second derivative database comparisons (Figure 41 and Table 19) the correct lactose matches remained high but the next best selected, orciprenaline, the next highest match in all cases was 0.538. Second derivatives are shown to be more selective.

Table 20 Correlation spectral matching of different lactose monohydrate sieve fractions from the database of second derivative spectra

Spectral range was 1116 - 2484 nm

Sieve fraction	Database selections	Correlation
< 32 µm	Lactose	0.991
	Orciprenaline	0.528
	Amylobarbitone	0.278
32 µm	Lactose	0.997
	Orciprenaline	0.532
	Meprobamate	0.274
45 µm	Lactose	1.000
	Orciprenaline	0.533
	Meprobamate	0.276
63 µm	Lactose	1.000
	Orciprenaline	0.535
	Meprobamate	0.278
93 µm	Lactose	0.999
	Orciprenaline	0.535
	Meprobamate	0.278
> 150 µm	Lactose	0.996
	Orciprenaline	0.538
	Meprobamate	0.280

Comparisons against the spectra and second derivative database were run with both un-sieved and sieved fractions. The results can be found in Table 19 and Table 20. Un-sieved lactose fractions were identified correctly but with varying correlations down to 0.991 for the different test samples.

Particle size distribution within the batches varied with the lowest correlations reported for the samples of <32 µm and >150 µm. This may be directly related

to the overall distribution of each particle size within the un-sieved database reference sample for lactose monohydrate.

Searching the database produced correlation spectral matches from 1.000 to 0.991 for lactose from un-sieved samples. Spectral Correlation correctly identified individual sieve fractions ranging from 1.000 to 0.990 (Table 19). The next best drug matches were 0.535 (orciprenaline) and 0.278 (meprobamate). Wavelength distance produced similar results.

Scanning probability windows and centre of gravity plots distinguished between different particle sizes based upon 1100-2500 nm with 95% confidence limit ellipses. Second derivatives improved the lactose identification range from 1.000 to 0.996 sieved and un-sieved. Low correlation values of other drugs were not altered significantly. Water content between particle sizes was also examined with each sieved fraction examined for water loss at 120° C. Sieve fraction water loss varied between 5.13 to 5.20% m/m and appeared unrelated to particle size. Thus, particle size affects drug identification based upon raw NIR spectra and could lead to mis-identification as other test samples reported high correlations. Second derivative spectral plots remove many of the physical spectral shifts, producing more positive identification.

4.3.4 The effect of different vial types and sizes

Table 21 Description of vial types and base diameters under investigation

Vial Reference	Description	Base diameter (mm)
Vial A	Waters 4ml	14
Vial B	Waters 4ml	14
Vial C	Waters 4ml	14
Vial D	Waters 4ml	14
Vial E	Waters 4ml	14
Vial F	BDH clear glass	22
Vial G	Samco clear glass	24
Vial H	Flat bottomed wide clear glass	26
Vial I	Waters 1ml clear glass	8

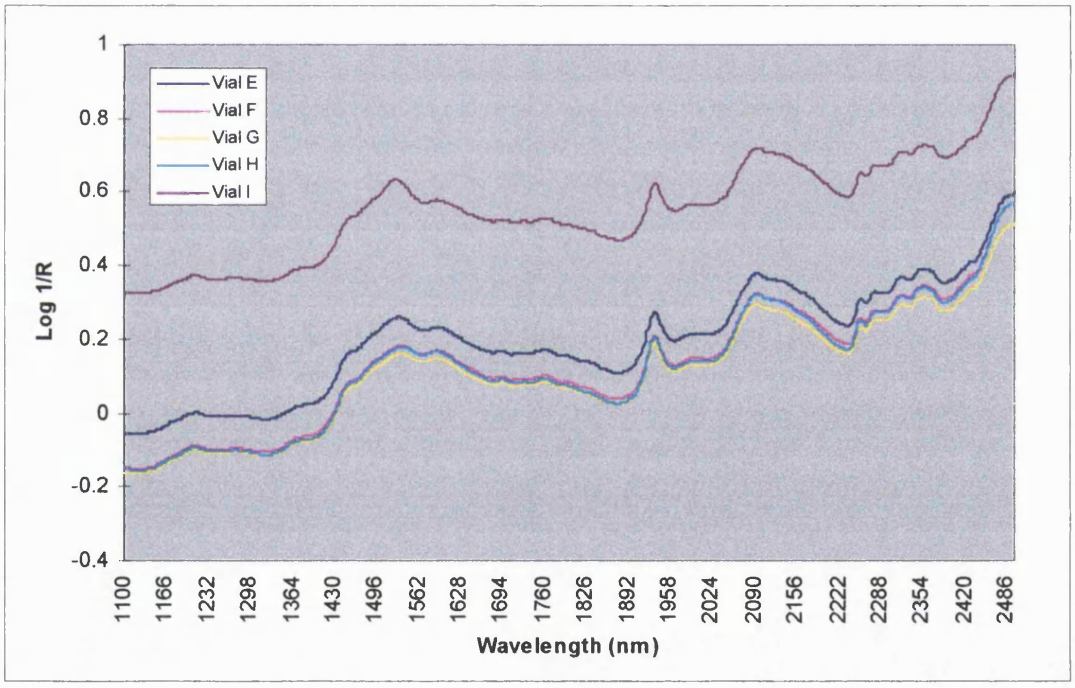


Figure 42 Plot of lactose monohydrate spectra contained within glass vials from different manufacturers

Spectral range is 1100 nm to 2500 nm

A description of vial types, manufacturers and base diameter sizes can be found in Table 21. Spectra were produced by measuring lactose monohydrate powder in each of the vial types described. The spectra are displayed in Figure 42 and

are consistent in terms of shape and peak wavelengths. However, there was considerable variation produced in spectral absorption with spectral offsets/ shifts evident.

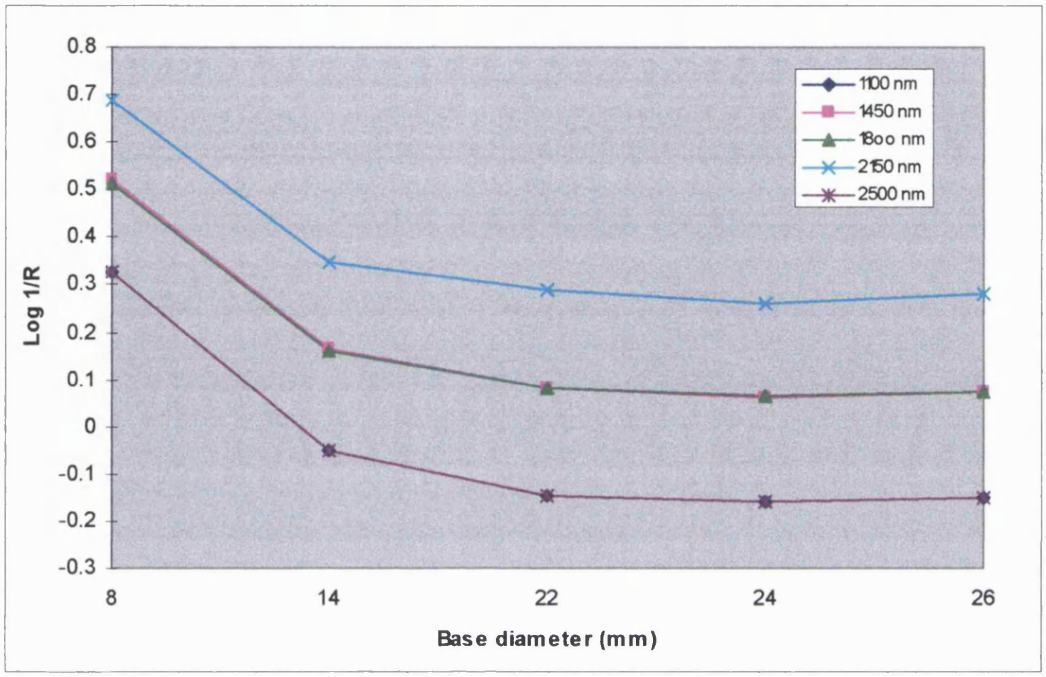


Figure 43 Plot of vial base diameters against absorbance at different wavelength points of the spectrum

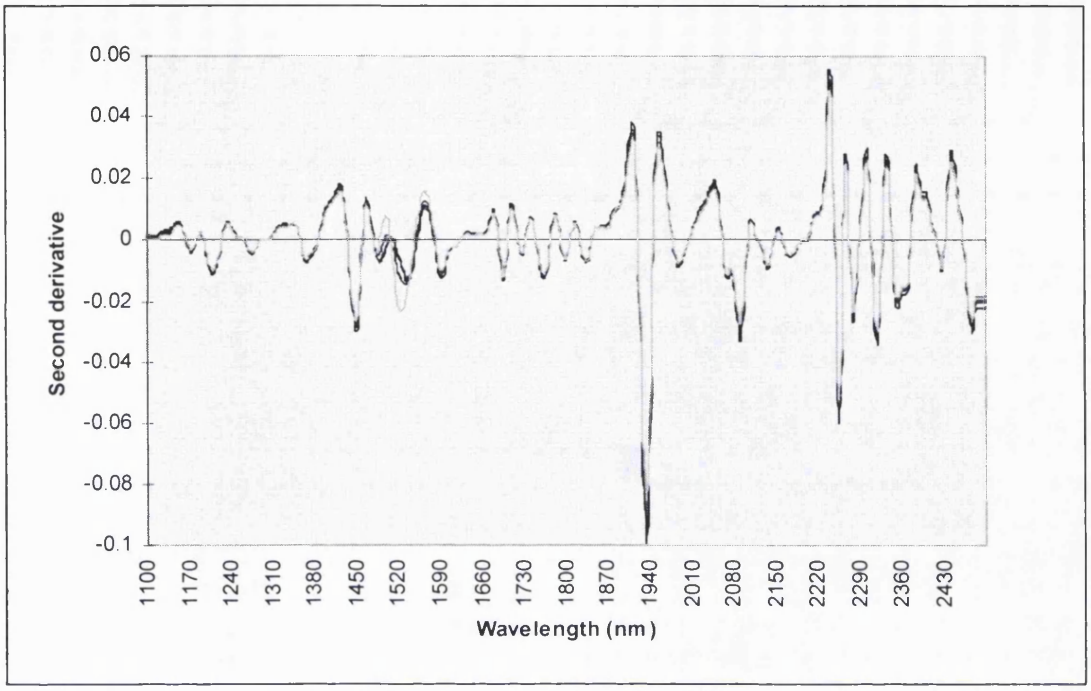
The X-axis represents vial base diameter in mm and the Y-axis represent absorbance.

The largest offest from the origin was produced by the Waters vial with the smallest base diameter (8 mm) and overall volume size (1 ml). There will be an effect upon the spectra where the base diameter size approaches (or is smaller than) the diameter of the instrument iris and therefore incident NIR light ray. Being able to narrow the slit or iris diameter should have decreased the amount of stray light, however this was not possible with the instrument. Where the light ray has a greater diameter than the base size of the sampling vial there will be incident light that will not fall upon the sampling vial and will therefore not fall upon the sample itself. This light will therefore be 'wasted' as it will not be active for the purposes of the sampling. Even if the incident diameter is equal to the

vial base diameter size or slightly less, there may be diffraction of light at the vial edges and there is a increased chance of spectral distortion occurring if the vial is not placed exactly in the centre of the sampling platform of the Rapid Content Analyser module. The effects, however, result only in spectral absorption shifts and in overall amount of light detected, there appears to be no effect upon spectra position which could mean sample identification would be unaffected. A plot of the effect of vial diameter upon the spectra can be seen in Figure 43. The figure suggests that there is an absorbance difference when considering only base diameters, however we need to look further before jumping to conclusions that all vials with base diameters less than 14 mm will produce higher absorbance readings. Between 14 and 26 mm the vials were from different manufacturers and can be assumed to contain different glass mixtures, thickness etc. However these results are remarkably consistent. This plot also shows that this is consistent at a number of different wavelength points across the wavelength range between 1100 nm to 2500 nm. The Waters vial with 14 mm base diameter varied from these higher diameters only slightly, but the greatest deviation from this was the lowest diameter vial of 8 mm. Bearing in mind that the 8 mm vial and the 14 mm vial are made by the same manufacturer and that the only difference in the vials was base diameter and volume capacity, it does appear that an absorbance effect is related to base diameter. Again, this is consistent with each of the wavelengths examined. The effect caused by diameter is reduced when using second derivative plots of the spectra (Figure 44).

The most reliable identification would therefore be achieved by using second derivative spectral plots and sample vials of greater than 14 mm.

Figure 44 Second derivative spectral plots for lactose monohydrate contained within



sampling vials from different manufacturers.

Spectral range examined is between 1100 nm and 2500 nm

4.4 Conclusions

There appeared to be no increase or decrease in reproducibility by varying the number of scans between 4 and 64. As the only potential benefit of reducing the number of scan passes was a time saving and only by a very few seconds, the instrument default of 32 scan passes was used unaltered. This is used as standard in the pharmaceutical industry and is taken from a Foss recommendation[80].

Spectra are effected when different portions of powders or tablets are examined. This is manifested by spectral absorption shifts and does not alter the peak position in terms of wavelength but does impact intensities and relative intensities of all peaks.

Powders must be sampled with the sampling vial shaken between each measurement. Tablets must be measured with the tablet rotated in between each measurement.

The effect is minimised by the use of second derivative spectral plots. These are predicted to be more successful for identification purposes.

Grinding exerts an influence upon the spectra produced in terms of a baseline shift. Peak wavelength positions are unaffected.

This effect is anyway reduced when referring to second derivative plots.

There is considerable variation produced in spectra by use of different sampling vial types. Glass material type was found not to be a factor for the glass vials used in this study, however has been shown to exert an influence as a part of sample presentation[103]. The most important factor in the differences produced is vial diameter. The greatest spectral shift is produced by the smallest base diameter. Critical base diameter value was below 14 mm

Particle size does affect drug identification if based upon NIR spectra and could lead to mis-identification as other test samples reported high correlations.

Second derivative spectral plots remove many of the physical spectral shifts, producing more positive identification.

Chapter 5 Drug identification using simple chemometrics

5.1 Summary of aims

Part 1 – Closely related compounds

- Set up a small database of NIR spectra of closely related drug substances
- Use simple chemometric techniques to identify the drug substances using external sample batches
- Determine the most successful techniques and highlight any potential problem areas
- State learning points for scaling up database for larger product numbers

Part 2 – Large number of compounds

- Collect a large number of pharmaceutical drug substances (300 +), record NIR spectra and construct an NIR spectral database in pure drug form
- Test the database for robustness (internally)
- Examine the identification of sets of external test drug substances run as 'unknowns'
- Report mis-identifications from drugs which are (or are not) contained within the database
- Create recommendations for NIR systems used for identifying drugs

5.2 Method

Part 1 – Closely related compounds

The purpose of the first part of this chapter was to set up a small database of 27 substances and to determine if external batches could be used to correctly identify the sample from the data already stored in a database. To do this two separate databases have been constructed – one houses spectra collected from the 30 batches, the second contains second derivatives.

A small database containing 13 barbiturates, 8 tetracyclines and 6 other substances was constructed. The details can be found in section 2.4.1.

The experimental details for how the powdered drug samples were analysed can be found in the Experimental chapter (section 2.2). Details of how the actual databases were constructed can be found in section 2.2.2.

Samples of each tetracycline and barbiturate were used as external test samples and were compared against the spectral database to determine the nearest match in each case. Two substances (4-epitetracycline, 4-epianhydrotetracycline) were also used to challenge the spectral databases but were not included within them. The purpose of this was to determine what would occur if a closely related substances was presented to a spectral database but was not contained in the database.

Part 1 results can be found in section 5.3.1 and 5.3.2.

The two chemometric methods used in this chapter are wavelength distance matching and correlation spectral matching. These are described in 1.10.2 and 1.10.3.

Part 2 of this section deals with the second part of the experiments, namely scaling up to the large number of substances in the database. The detailed method is contained within section 2.4.2. The first step when scaling up to the larger database was to include database 'validations'.

5.2.1.1 Internal Validation

The database was validated internally by ensuring that each substance in the database product could be distinguished within it utilising the chemometric methods available. Internal validation is the comparison of every compound in the database against every other compound, the purpose being to determine the correlation value ('threshold value') at which it is possible to identify all compounds in the database uniquely from each other. If two compounds could not be uniquely identified from another, then the validation will fail. The options in this case are either to alter identification parameters (such as wavelength range) and re-execute the validation or to remove one, both or all offending substances. Naturally, simply removing products limits the scope of the method to the substances that are successfully validated, but this is no different to being limited to the finite number of substances contained within a database. Internal validation was attempted by correlation spectral matching and by wavelength distance matching. Database validation was attempted for the following wavelength ranges: 1100 - 2498 nm, 2000 - 2400 nm, 2200 - 2400 nm. These ranges represent the ranges that are to be used for identification. Validation was attempted on both spectra and second derivative databases. The initial threshold correlation coefficient was set to 0.850 for each validation. This was then increased in increments of 0.005 where a validation attempt was

unsuccessful. This was repeated until either a critical correlation value was determined (i.e. validation achieved) or until reaching a correlation coefficient of 1.000. If validation was still unsuccessful at this point it was concluded that it was not possible to validate the database and this was investigated further.
The same method was used to determine the critical wavelength distance match value, commencing with ten standard deviations and incrementing by 0.500 standard deviations. This was repeated until validation was achieved or until reaching zero standard deviations (database validation unsuccessful). If unsuccessful, an investigation was commenced.

5.2.1.2 External Validation

To test the robustness of the database it must be used to identify spectra from different samples or batches of drug substances.
The samples of drugs from external sources were treated as unknown samples and were placed into the Waters sampling vials. The samples were then placed onto the RCA sampling window and measured. Twelve measurements were recorded per sample. Second derivative spectra for each 'unknown' were compared in the same manner against the second derivative database. This was repeated for each of the wavelength ranges already identified.

5.3 Results and Discussion

The reason for choosing the specific mathematical transformations of raw and second derivative spectra can be referred to section 1.8.2, where the majority of the work already conducted utilise either raw or second derivative plots – providing this to be a directly comparable piece of work.

As soon as the databases were set up, the external samples were used – firstly samples of all the barbiturates contained within the database and then samples of the tetracyclines.

5.3.1 Identification of the barbiturates

The barbiturates are of the general structure:

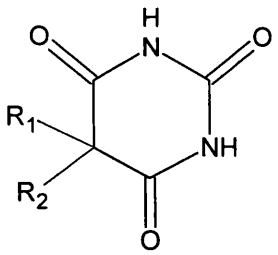


Figure 45 General barbiturate structure

The first series of results from the external barbiturate samples uses the correlation spectral matching to search the spectra database. The results are contained in Table 22.

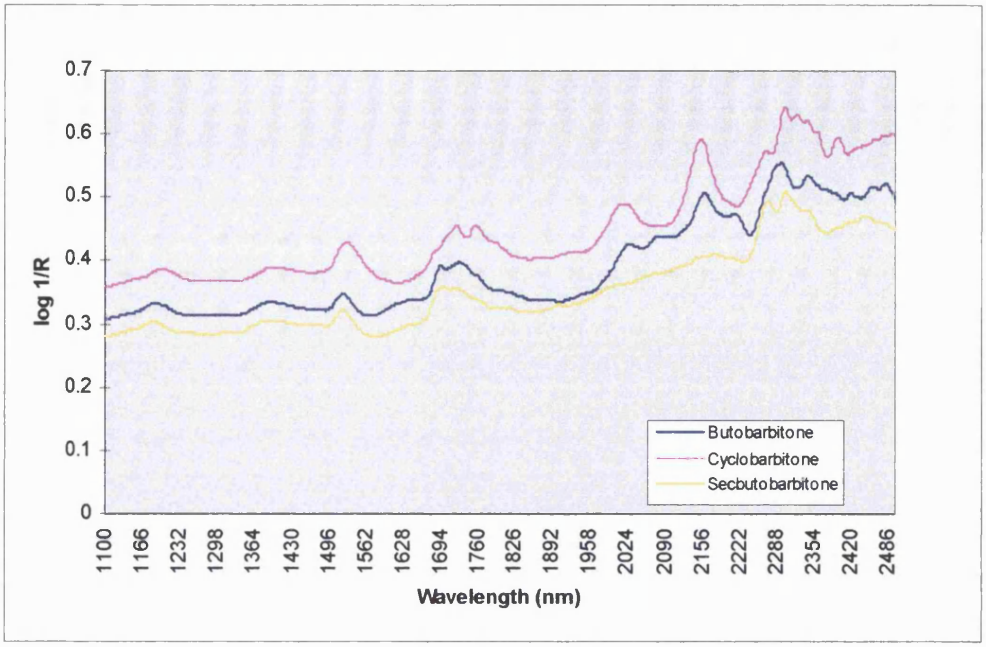


Figure 46 Comparison of spectra of butobarbitone, secbutobarbitone and cyclobarbitone between 1100 nm and 2500 nm

The spectra of the barbiturates (Figure 46) showed many spectral similarities between 1100 nm and 1700 nm, 1760 nm and 2050 nm with prominent peak positions noted at 1180 nm, 1500 nm, 2000 nm and 2300 nm. The main areas of spectral differences were observed between 1700 nm and 1750 nm and 2000 nm to 2400 nm.

Table 22 Correlation spectral match results for six of the thirteen barbiturates compared against the database of pure drug spectra in the range 1100 - 2500 nm

True identity	Correlation spectral match value	Best match
Cyclopentobarbitone	1.000 0.999 0.999	Cyclopentobarbitone Pentobarbitone Brallobarbitone
Brallobarbitone	1.000 0.999 0.999	Brallobarbitone Butobarbitone Anhydrochlortetracycline base
Barbituric Acid	1.000 0.996 0.994	Barbituric acid Allobarbitone 4-Epianhydrochlortetracyline
Butobarbitone	1.000 0.999 0.999	Brallobarbitone Secbutobarbitone Cyclobarbitone
Cyclobarbitone	1.000 0.999 0.998	Cyclobarbitone Secbutobarbitone Butobarbitone
Vinbarbitone	1.000 1.000 0.999	Vinbarbitone Butobarbitone carboxylic acid Talbutal

There were notable points to report. Firstly, the correctly identified compounds were not significantly different from the next best match in nearly all cases. Selectivity of the correct match was not established. Other suggested substances possessing a correlation spectral match value of 0.999 or greater would be classified as matches and following this criteria a number of mis matches are therefore being reported. In one case (butobarbitone), the substance was not selected from the database within the three best matches. These were identified as brallobarbitone (1.000), secbutobarbitone (0.999) and cyclobarbitone (0.999). Clearly the spectra in this case are not providing the selectivity to differentiate between the closely related compounds.

It may be suspected that the nature of matching by correlation between spectra where the baseline is inconsistent, in turn making a magnitude of change difficult to establish; would be the source of this problem.

Comparison can then be made against the second chemometric method of wavelength distances. The same sample spectra were used, only the mathematical method of comparison against the database spectra was altered. The detail for a sample of the same barbiturates are shown in Table 23.

Table 23 Wavelength distance match for six of the thirteen barbiturates against the database of pure drug spectra in the range 1100 - 2500 nm

True identity	Wavelength distance match value	Best match
Cyclopentobarbitone	0.668 0.902 1.071	Cyclopentobarbitone Talbutal Allobarbitone
Brallobarbitone	0.192 1.005 1.791	Brallobarbitone Allobarbitone Talbutal
Barbituric acid	0.736 0.762 2.763	Barbituric acid Allobarbitone Talbutal
Butobarbitone	0.103 0.932 1.888	Butobarbitone Allobarbitone Anhydrochlortetracycline
Cyclobarbitone	0.498 0.922 5.501	Cyclobarbitone Allobarbitone Brallobarbitone
Vinbarbitone	0.173 0.965 2.353	Vinbarbitone Allobarbitone Pentobarbitone

Each barbiturate was correctly selected as the best match but because the matches are so close together, it is still difficult to claim to have obtained successful matches in each case.

It raises the question as to what is the definition of a successful match, because being the 'best' match from a database of compounds is firstly limited by the scope of the database and secondly has been shown to be influenced by other factors - as discussed in Chapter 4 (Sample Presentation).

At this point there is no criterion under which matches are being accepted other than being the best match. This is discussed further in Part 2 of this chapter where more substances are involved and thought is put to when not to rely on the best match that is presented. However, in any case if guidelines were being constructed in a forensic scenario to describe an acceptable/ unacceptable match, in addition to the actual match value itself would be the pre-requisite that no other substance was matched within some acceptable range of standard deviations.

The results are similarly as inconclusive as the correlation results other than being able to say that the matches are working, appear to be equally good, but are not discriminating enough based upon untreated spectra.

Comparing the substances as second derivative plots (Figure 47) provided an altogether more encouraging outcome. Distinct areas of specificity can be seen on this plot despite the structural similarities of the substances. This can be witnessed with reference to where each barbiturate can be seen to have been successfully identified from the database (Table 24). The distinction between correct and next best match was also marked with the next highest match for any substance being 0.783 (with the highest correlation match in this case of 1.000).

Table 24 Correlation spectral match results for the full thirteen barbiturates compared against the database of pure drug second derivative spectra in the range 1100 - 2500 nm

True identity	Correlation spectral match value	Best match
Cyclopentobarbitone	1.000 0.782 0.730	Cyclopentobarbitone Allobarbitone Talbutal
Brallobarbitone	1.000 0.723 0.514	Brallobarbitone Allobarbitone Talbutal
Butobarbitone	1.000 0.693 0.656	Butobarbitone Cimetidine Pentobarbitone
Cyclobarbitone	1.000 0.540 0.454	Cyclobarbitone Pentobarbitone Vinbarbitone
Butobarbitone carboxylic acid	1.000 0.582 0.523	Butobarbitone carboxylic acid Butobarbitone Cimetidine
3'Hydroxy butobarbitone	1.000 0.422 0.413	3'Hydroxy butobarbitone Butobarbitone Secbutobarbitone
Pentobarbitone	1.000 0.657 0.591	Pentobarbitone Butobarbitone Cyclobarbitone
Talbutal	1.000 0.739 0.724	Talbutal Cyclopentobarbitone Allobarbitone
Vinbarbitone	1.000 0.694 0.585	Vinbarbitone Secbutobarbitone Thiopentone
Allobarbitone	1.000 0.783 0.721	Allobarbitone Cyclobarbitone Talbutal
Secbutobarbitone	1.000 0.798 0.695	Secbutobarbitone Thiopentone Vinbarbitone
Methylphenobarbitone	1.000 0.509 0.399	Methylphenobarbitone Ampicillin Allopurinol
Quinalbarbitone	1.000 0.507 0.488	Quinalbarbitone Butobarbitone Talbutal

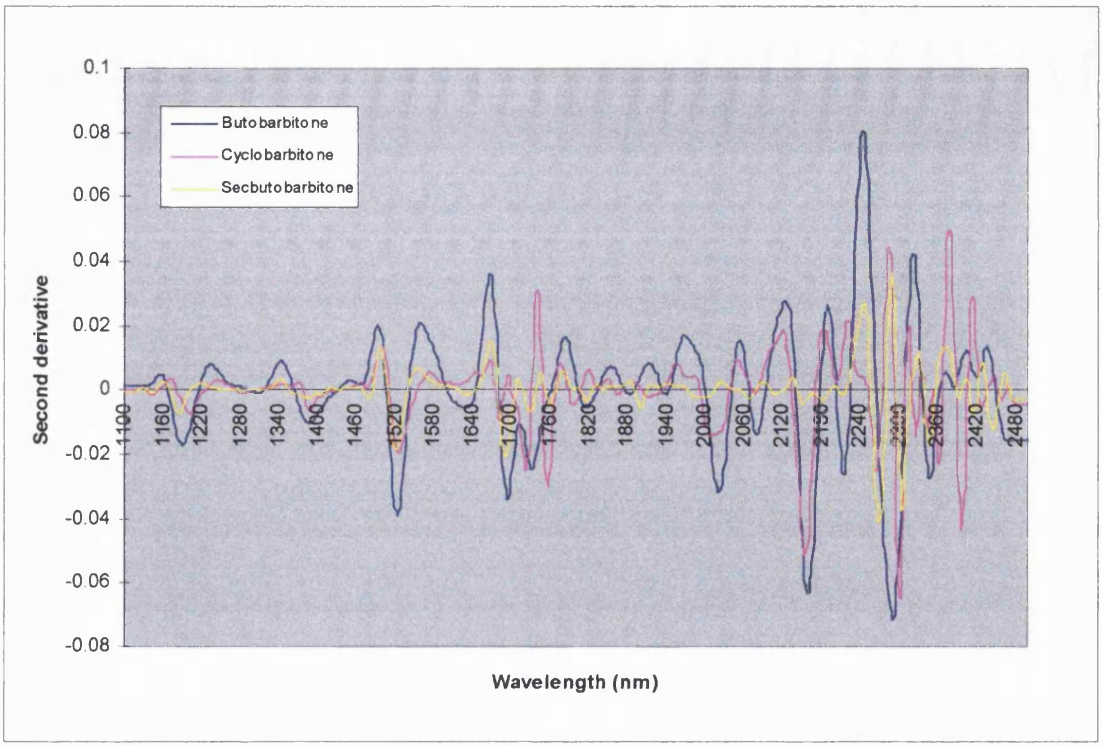


Figure 47 Comparison of second derivative spectra butobarbitone, secbutobarbitone and cyclobarbitone between 1100 nm and 2500 nm

Butobarbitone was correctly identified with second derivatives with a match correlation of 1.000. The next nearest matches were cimetidine (0.693) and pentobarbitone (0.656).

In the case with the matches by the wavelength distance method, a similar trend is noted, whereby the correct substance was identified, with the second derivatives, the discrimination is seen to be quite marked (see Table 25).

All barbiturates have been correctly identified and the next highest match in any case was 56.1 standard deviations for allobarbitone when the best match was 1.867 standard deviations for talbutal.

In addition to the evidence for the greater selectivity of identification from the second derivatives, the wavelength distance method demonstrated better

identification characteristics with the closely related barbiturates. The same is considered for the tetracyclines.

Table 25 Wavelength distance match for thirteen barbiturates

The matches were run against the database of pure drug second derivative spectra between 1100 nm and 2500 nm

True identity	Wavelength distance match value	Best match
Cyclopentobarbitone	1.6 100.1 190.9	Cyclopentobarbitone Allobarbitone Brallobarbitone
Brallobarbitone	1.7 149.0 313.9	Brallobarbitone Allobarbitone Cyclobarbitone
Butobarbitone	1.5 132.1 165.2	Butobarbitone Butobarbitone carboxylic acid Cyclopentobarbitone
Cyclobarbitone	1.3 182.7 274.0	Cyclobarbitone Allobarbitone Talbutal
Butobarbitone carboxylic acid	1.6 133.3 134.9	Butobarbitone carboxylic acid Allobarbitone Brallobarbitone
3'Hydroxy butobarbitone	1.9 613.3 853.9	3'Hydroxy butobarbitone Allobarbitone Pentobarbitone
Pentobarbitone	2.0 141.6 153.6	Pentobarbitone Cyclobarbitone Allobarbitone
Talbutal	1.9 56.1 84.0	Talbutal Allobarbitone Brallobarbitone
Vinbarbitone	2.3 97.5 124.5	Vinbarbitone Allobarbitone Brallobarbitone
Allobarbitone	1.6 236.0 532.7	Allobarbitone Brallobarbitone Cyclopentobarbitone
Secbutobarbitone	2.2 111.0 144.6	Secbutobarbitone Thiopentone Allobarbitone
Methylphenobarbitone	1.9 269.2 432.2	Methyl phenobarbitone Cyclopentobarbitone Benzyl penicillin
Quinalbarbitone	2.1 418.9 445.1	Quinalbarbitone Allobarbitone Talbutal

5.3.2 Identification of the tetracyclines

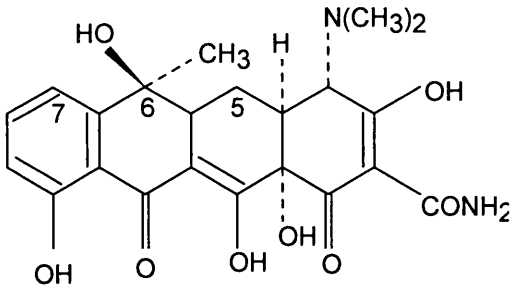


Figure 48 Structure of Tetracycline

The initial comparisons were again based upon zero order spectra but produced poor matches (Table 26). The correlation spectral matches based upon the zero order spectra produced the correct tetracycline as the highest match in each case, so the identification was correct, but these matches were not demonstrably different to the second and third best in each case. The same trend was apparent with wavelength distances (Table 27) where there were also mis-identifications. For example, chlortetracycline hydrochloride was incorrectly identified as anhydrochlortetracycline with a match value of 0.670. Table 26 demonstrates a very similar pattern to that from the barbiturates. The similarity in the spectra for oxytetracycine, chlortetracylince and anhydrochlortetracylcine can be compared directly in Figure 49 and Figure 50 (second derivatives).

With the zero order spectra, even visually differentiating between the spectra is difficult, when taking away the overall absorbance shift, whereas on the second derivative plot, 'negative' peaks can be clearly seen to separate the three

drugs. The second derivative plot accentuates specific differences between the spectra, for example between 1850 nm and 1916 nm and 2000 nm to 2400 nm. This region between 2000 and 2400 nm has already been shown to be highly discriminating for searches conducted earlier in this chapter with the other pharmaceutical 'unknowns'.

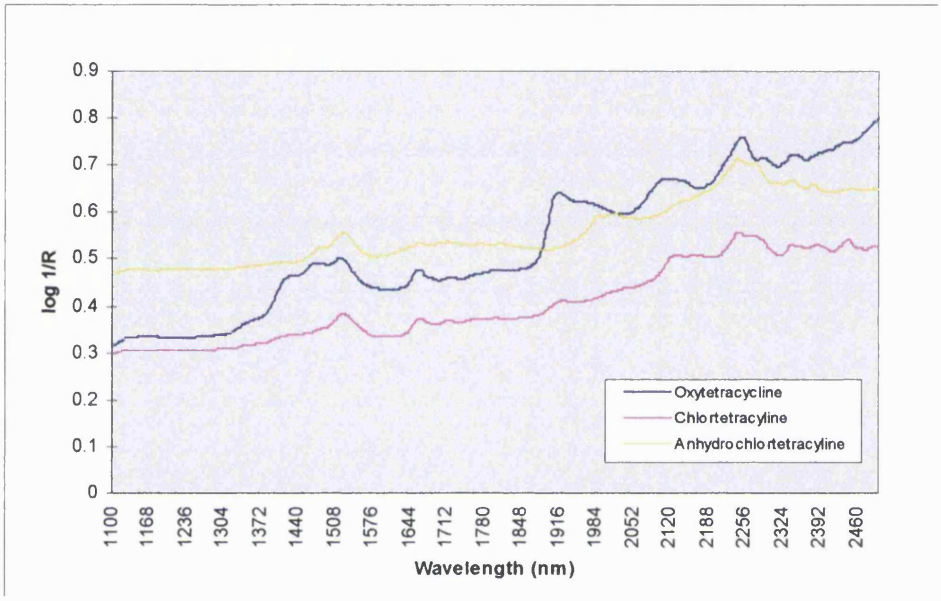


Figure 49 Comparison of zero order spectra of oxytetracycline, chlortetracycline and anhydrochlortetracycline between 1100 nm and 2500 nm

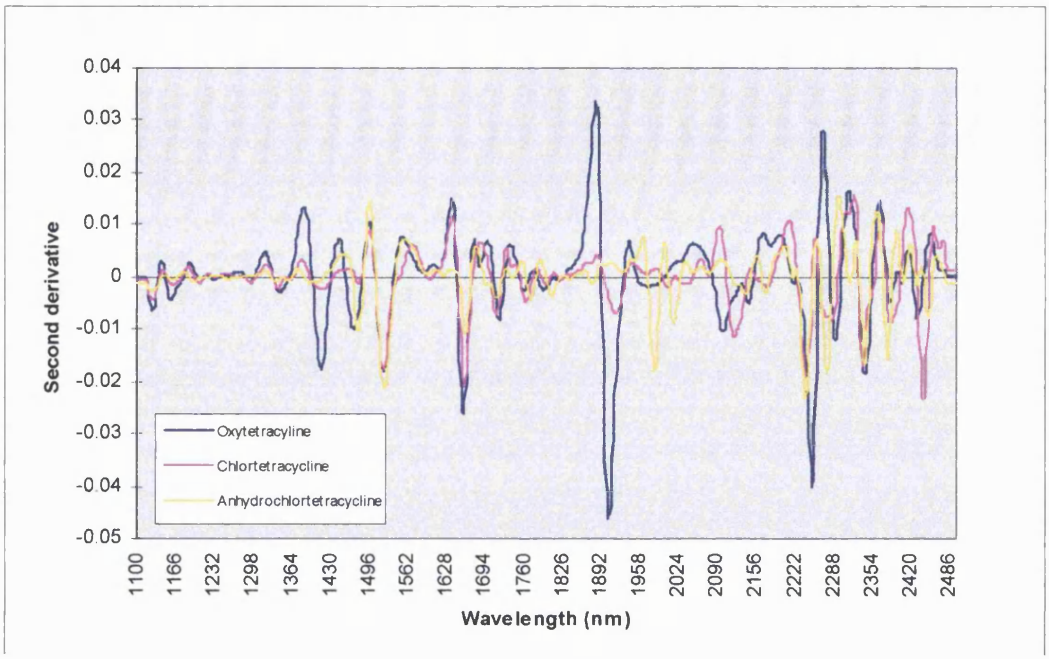


Figure 50 Comparison of second derivative spectra of oxytetracycline, chlortetracycline and anhydrochlortetracycline between 1100 nm and 2500 nm

Table 26 Correlation spectral match results for five tetracyclines.

The samples were compared against the database of pure drug spectra in the range 1100- 2500 nm

True identity	Correlation spectral match value	Best match
Oxytetracycline hydrochloride	1.000	Oxytetracycline hydrochloride
	0.999	Oxytetracyline dihydrate
	0.999	Limecycline
Demeclocycline hydrochloride	1.000	Demeclocycline hydrochloride
	0.999	Chlortetracycline hydrochloride
	0.999	Methacycline hydrochloride
Tetracycline hydrochloride	1.000	Tetracycline hydrochloride
	0.998	Methacycline hydrochloride
	0.996	Papaverine
Methacycline hydrochloride	1.000	Methacycline hydrochloride
	0.998	Tetracycline hydrochloride
	0.997	Demeclocycline
Chlortetracycline hydrochloride	1.000	Chlortetracycline
	0.999	Demeclocycline
	0.999	Tetracycline hydrochloride

The spectra, as has previously been noted are difficult to base identification upon. The reason for the non-selectivity of the spectra is the wealth of information contained within them. The spectra provide not only chemical and structural information but a variety of physical data too. It is to this end that the spectra are not set about a constant baseline but display baseline shifts. Peaks are often not well defined but are a cumulate of overtones and combinations where peak areas often offer direct relationships to physical attributes of the sample such as particle size. This makes it very difficult to use zero order spectra for identification. The range of matches reported for the chlortetracycline test substance were anhydrotetracycline hydrochloride (0.670), allobarbitone (0.720) and chlortetracycline third with 1.150. Allobarbitone was also selected as the match to the oxytetracycline test substance with a wavelength distance match of 1.375 standard deviations.

These anomalies were alleviated when using second derivative spectra from both identification methods (Table 28 and Table 29).

Table 27 Wavelength distance matches for five tetracyclines

The matches were run against the database of pure drug spectra between 1100 nm and 2500 nm.

True identity	Wavelength distance match value	Best match
Oxytetracycline hydrochloride	0.503	Oxytetracycline hydrochloride
	1.741	Allobarbitone
	2.267	Talbutal
Demeclocyline hydrochloride	0.752	Demeclocycline
	1.032	Allobarbitone
	1.478	Anhydrochlortetracycline hydrochloride
Tetracycline hydrochloride	1.497	Allobarbitone
	1.744	Tetracycline hydrochloride
	3.361	Talbutal
Methacycline hydrochloride	0.455	Methacycline hydrochloride
	0.713	Allobarbitone
	3.804	Brallobarbitone
Chlortetracycline hydrochloride	0.670	Anhydrochlortetracycline hydrochloride
	0.720	Allobarbitone
	1.150	Chlortetracycline
Oxytetracycline dihydrate	1.375	Allobarbitone
	1.378	Oxytetracycline hydrochloride
	2.681	Talbutal

Table 28 Correlation spectral match results for the ten tetracyclines

The samples were compared against the database of pure drug second derivative spectra between 1100 nm - 2500 nm

True Identity	Correlation spectral match value	Best Match
Limecycline	1.000 0.765 0.611	Limecycline Oxytetracyline hydrochloride Demeclocycline
Oxytetracyline hydrochloride	1.000 0.769 0.473	Oxytetracyline hydrochloride Limecycline Chlortetracycline hydrochloride
Demeclocycline hydrochloride	1.000 0.729 0.649	Demeclocycline hydrochloride Tetracycline hydrochloride Chlortetracycline hydrochloride
Tetracycline hydrochloride	1.000 0.747 0.730	Tetracycline hydrochloride Chlortetracycline hydrochloride Demeclocycline hydrochloride
Methacycline hydrochloride	1.000 0.433 0.422	Methacycline hydrochloride Demeclocycline hydrochloride Imipramine
Chlortetracycline hydrochloride	1.000 0.749 0.646	Chlortetracycline hydrochloride Tetracycline hydrochloride Demeclocycline hydrochloride
Oxytetracyline dihydrate	1.000 0.603 0.596	Oxytetracycline dihydrate Limecycline Demeclocycline hydrochloride
Anhydrochlortetracycline hydrochloride	1.000 0.580 0.471	Anhydrochlortetracycline hydrochloride Chlortetracycline hydrochloride Demeclocycline hydrochloride
4-Epitetracycline*	0.461 N/A N/A	Limecycline none available none available
4-Epianhydrotetracycline*	0.632 N/A N/A	Brallobarbitone none available none available

*Note: * indicates products not included in the database but used to challenge the products contained within it. N/A indicates no match obtained greater than or equal to 0.1*

Each test substance was correctly identified with the highest match and each of the next best matches was outside the threshold for identification. With correlation spectral matching the matches were 1.000, with the wavelength distance method the reported highest standard deviation distance for a 'best' match was 2.247 standard deviations.

Two tetracyclines were also submitted as test substances but were not included in the database. These were 4-epichlortetracycline and 4-

epianhydrotetracycline (Table 28 and Table 29). One match for each substance was in fact suggested from each identification method but the identification values were very low. The products were only run against the second derivative database and not against the original spectra because the standard test results were inconclusive for original spectra. The values for acceptance criteria are investigated further later in the chapter with the internal validation aimed at setting a threshold value and then external test samples used to test against this.

It is clear from this work, however, that Identification of the closely related compounds is possible only when using the second derivative database. For both classes of drug, barbiturate and tetracycline and for both methods of identification: correlation and wavelength distance ambiguity and mis-identification exist if identifying compounds based solely on original spectra. These were removed when using their second derivatives. Because of these, the key learning point for scaling up to a larger number of compounds is that original spectra are unsuitable for identification purposes. However both wavelength distance and correlation are both simple to operate, understand and interpret, have both performed well and should be examined further.

Table 29 Wavelength distance match of the tetracyclines against the database of pure drug second derivative spectra between 1100 nm and 2500 nm

True identity	Wavelength distance match value	Best match
Limecycline	1.7 72.2 96.2	Limecycline Brallobarbitone 3'Hydroxy butobarbitone
Oxytetracyline hydrochloride	1.8 252.1 359.2	Oxytetracycline hydrochloride Brallobarbitone 3'Hydroxybutobarnitone
Demeclocycline hydrochloride	1.6 136.4 150.8	Demeclocycline hydrochloride Brallobarbitone Tetracycline hydrochloride
Tetracycline hydrochloride	2.2 252.7 330.7	Tetracycline hydrochloride Brallobarbitone Allobarbitone
Methacycline hydrochloride	2.0 247.5 247.8	Methacycline hydrochloride 3'Hydroxy butobarbitone Cyclobarbitone
Chlortetracycline hydrochloride	2.2 105.4 112.1	Chlortetracycline hydrochloride Cyclopentobarbitone Brallobarbitone
Oxytetracyline hydrochloride	1.8 272.5 372.5	Oxytetracycline hydrochloride Brallobarbitone Cyclobarbitone
Anhydrochlortetracycline hydrochloride	1.5 83.2 103.9	Anhydrochlortetracycline hydrochloride Cyclobarbitone Allobarbitone
4-Epitetracycline*	84.9 N/A N/A	Brallobarbitone none available none available
4-Epianhydrotetracycline*	178.9 N/A N/A	Brallobarbitone none available none available

Note: * indicates products not included in the database but used to challenge the products contained within it. N/A indicates no match obtained less than or equal to 650 s.d.

5.3.3 Database validation

This part of the work carries on from section 5.2.1.1. A database of second derivative spectra from 300 plus compounds has been constructed (as detailed in section 2.4.2). With so many different substances the first problem was always going to be whether such a large number of substances could be differentiated from each other. The first part of this chapter has already showed that setting up a database of a few compounds was quite achievable and that by using simple chemometrics such as correlation matching even closely related substances could be identified correctly when using second derivative spectra.

An important part of creating a database upon which identification is to be based is to validate that database. This involves the inclusion of closely related substances to those of interest in addition to spectra from different batches so that some robustness can be built into the database. This part is built into the sampling method (section 2.2). This will also increase the chances of a correct identification, when using different external sample batches.

Database validation was determined both internally and externally. The internal validation by correlation spectral matching and wavelength distance matching is a part of the IQ^2 software package supplied from the instrument manufacturer, NIRSystems.

The internal library validation carried out two functions.

1) To check that all possible pairs of spectra used to create the database can be uniquely identified by correlation spectral matching and wavelength distance matching. Critical identification values were therefore determined.

2) To check that pairs of spectra from different substances do not contain areas of identification that overlap and that therefore are indistinguishable.

The purpose of the external validation was to examine the robustness of the identification method through the use of external samples.

5.3.4 Determination of criticality

The critical value (r) is the threshold value between identification and mis-identification. For correlation spectral matching this is the match value above which it is deemed that identification has occurred and below which it is accepted that identification has not taken place. This value is derived from the internal and external database validation. The critical value defines the standard equal to or above which a match result is considered to be acceptable (or a pass). It is intended to be used as a guide rather than the 'hard and fast rule' because of the lack of experience of data, i.e. this will be determined as the identification of the external test samples is carried out.

A mis-identification is where a correlation spectral match value greater than the critical r value has been obtained but where the true identity of the test substance and the database match differ.

For wavelength distance matching the critical value (r) is the value below which it is deemed that identification has occurred. Therefore wavelength distance match values above this critical r value are deemed as non-identifications. A mis-identification is where a wavelength distance match below the critical value has been obtained where the true identity of the test substance and the database match differ.

The first step was to determine if it was possible to distinguish between all the substances in the database. This was performed within the IQ^2 software. For

correlation spectral matching the 300+ compounds were included within the database.

When matching by distance mode the algorithm uses the maximum calculated distance between the data points of interest i.e. the database substance and the test samples. Samples must be both identified and qualified for the database to be validated as suitable for matching.

Problems were encountered while attempting to validate the database internally. Both identification by correlation and identification by distance produced numerous repeated errors across all wavelength ranges attempted. The closest to full validation was the correlation matching between 2000 and 2400 nm. This still produced two failures (4-epianhydrotetracycline and 4-epichlortetracycline hydrochloride). These two products are remarkably similar in structure and behaviour. They were not identified as difficult substances in Part 1, the earlier test because they were not included in the database of substances. This was because they were used as test substances to challenge the few products in the database and as such were not actually compared against each other. Admittedly, this was overlooked but the point here is that both products fail the internal validation due to the similarity of their spectra, which is in turn because of the structural similarities of the two products, which is correct.

However in the validation stage here, their structural similarity made detection between the two difficult and caused a failure when attempting to validate the database by either correlation or distance matching. The detection settings for each correlation internal validation were altered from 0.850 to 0.900, 0.950. 0.975, 1.000. The wavelengths attempted were 1100-2500 nm and 2000 –2400

nm – however, distinct identifications were still not possible for the two failing products. The internal validation was also unsuccessfully attempted for wavelength distances (15.0, 10.0, 7.5, 5.0, 2.5 standard deviations). Each validation attempt took between 4-6 hours to run which was fairly time consuming and frustrating. It is envisaged that this task would have taken less time with enhanced computer processing capabilities but would remain the most time consuming part of the practical NIR process. It is also envisaged that the time taken would be longer as the database is increased in size. It must be repeated each time a new sample is added to the database.

In the case of the tetracyclines it was found that the only way of completing the full database validation with zero failures was by removing one of the products – 4-epianhydrotetracycline. The successful correlation validation threshold occurred at 0.850. The database was able to be validated over a range of wavelengths - separately between 1100 - 2500nm, 2000 - 2400 nm (and later 2200 - 2400 nm.) The distance threshold value was found to be 10.00 standard deviations.

However, what this demonstrated is that neither identification method can be classified as a 'one method fits all' identification technique. Failing the validation at this stage is highlighting the fact that there are, or will be 'problem cases' and exceptions can be built around these.

34 product test samples each comprising different batches of drug substance were used to externally validate the spectral database by correlation spectral matching and wavelength distance matching. The samples were run in a blind manner (i.e. the analyst did not know the identity) and included samples for which there were no spectra in the database (see Table 64 in the Appendix for

the full list of samples used). The test samples compared against the database were obtained from different sources to those used to construct the database. The variations in particle size, particle shape, potency and age which were introduced by the use of external test samples provided a more stringent and 'real life' challenge to the database than alternative batches of the same samples used to construct the database.

5.3.5 Comparison using second derivative spectra

Only second derivative spectra were used (as discussed in Part 1). The ratio of an inflection "peak" to a "true" peak is approximately 1:2. The resultant second derivative plot (Figure 51) has, however, lost the majority of physical information that would have been present in the origin spectral plot.

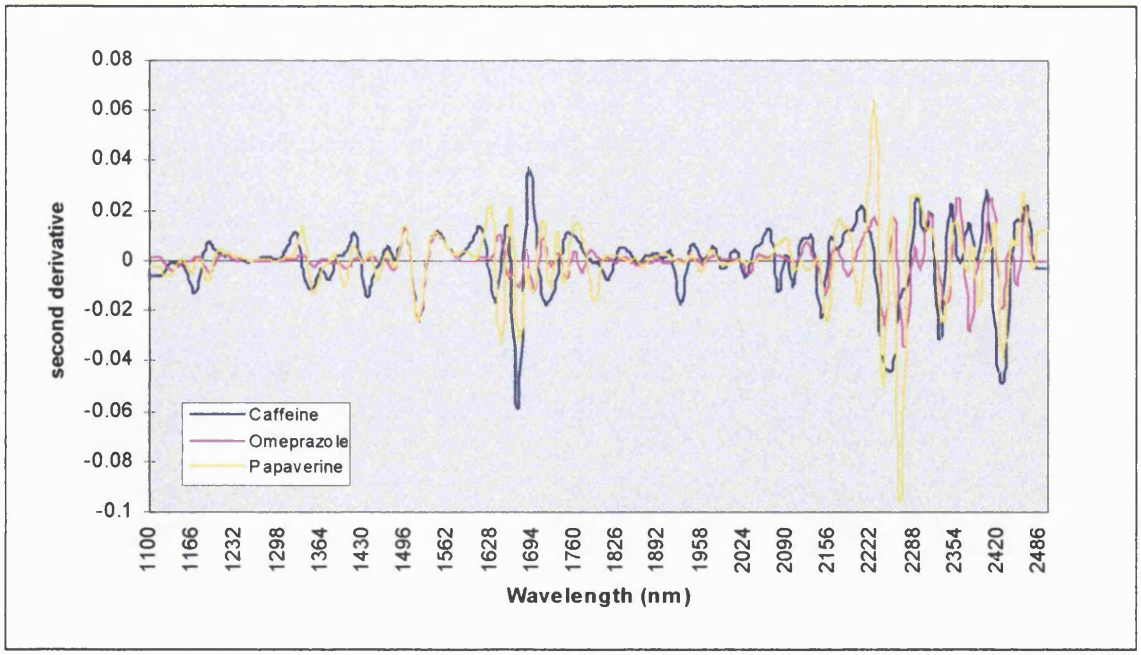


Figure 51 Comparison of second derivative spectra of papaverine, omeprazole and caffeine drug substances in the range 1116 nm to 2484 nm

The searches against the database of second derivative spectra were made initially using the full spectral range (1100 nm – 2500 nm). This range was then altered to determine whether including less information would be of advantage or not. A selection of the search results are shown in Table 30

The highest match values ranged from 1.000 to 0.988, throughout all the searches carried out. A further noteworthy point is the difference between the highest and next highest correlations in each search. The best second placed match reported was 0.614 for progesterone (also displayed above). For each search there was now a clear distinction between the highest and next highest matches.

Pheniramine and paracetamol were not included in the spectral databases. Both substances are the best matches in Table 30 but the spectral match values are low (0.565 and 0.508 respectively). This is below the critical threshold value of 0.850 and would not be accepted as a high enough correlation upon which to base a match. No other searches reported any mis-identifications at this wavelength range for spectral matching by correlation or distance.

Due to a number of overtones being present between 1100 and 2500 nm carrying duplicate information (section 1.6.1), it was decided to run the identification matches across a single overtone. The range selected was 2000 - 2400 nm which is towards the mid infrared region. Theoretically towards the higher energy area of the NIR region there is a greater possibility of differentiating between different substances by peak comparison. Hence mid-IR visual peak identification.

Table 30 identification of pure drugs by correlation spectral matching from the database of second derivative plots of spectra.
Spectroscopic range examined 1100 nm - 2500 nm (Top three matches displayed)

True identity	Correlation spectral match value	Best match
Desipramine	1.000 0.591 0.582	Desipramine Imipramine Propranolol
Allopurinol	1.000 0.557 0.410	Allopurinol Phenytoin Glutethimide
Testosterone	1.000 0.614 0.323	Testosterone Progesterone Naproxen
Papaverine	1.000 0.541 0.397	Papaverine Omeprazole Caffeine
Cefuroxime	0.999 0.238 0.193	Cefuroxime Amylobarbitone Phenobarbitone
Benzyl penicillin	1.000 0.387 0.283	Benzyl penicillin Ampicillin Phenobarbitone
Ampicillin	0.988 0.522 0.441	Ampicillin Glutethimide Phenytoin
Thiopentone	0.993 0.546 0.535	Thiopentone Progesterone Amylobarbitone
Pheniramine - drug not included in database	0.565 0.557 0.509	Ampicillin Benzyl penicillin Orciprenaline
Paracetamol - drug not included in database	0.508 0.496 0.484	Phenytoin Allopurinol Glutethimide

Also by eliminating variation liable to decrease the correlation spectral match from other areas of the spectrum, the match value should be optimised for the correct compound. The counter argument is that a fuller (wider) spectral range would contain more spectral information that would :

- increase the match value to the correct compound as the match would be tighter

- decrease the matches of incorrect compounds as there would larger areas of variation
- eliminate/ ignore data (i.e. it is not justified)

It is likely that a combination of both cases exist. However, it has already been seen that some data analysis modes are indeed wavelength specific e.g. multivariate discriminant analysis by Mahalanobis distances[38]. Arguments for and against have been raised between this and techniques employing the full wavelength range such as (SIMCA)[40].

The representative results displayed in Table 31 however, appear to indicate that the maximum or matched compounds are unchanged and that the match value for the next highest match decreases slightly, increasing the gap between best match and next best. The proposed value for paracetamol (not included in the database) in this case has also decreased. However it is noticeable that there were now no further selections suggested greater than 0.314 (Table 31). The deduction from this is clear, either the drug is not contained within the database or it has not been identified by this technique. The results from Table 30 and Table 31 and the database validations would suggest that the substances are not present in the database.

Visual inspection of the spectra from Figure 51 suggests that there may be specific active peaks that are substance specific between 2200 and 2400 nm. Test substances were therefore compared between 2200 and 2400 nm with the expectation of enhanced separation between the substances and therefore higher and more selective match values.

Table 31 Identification of pure drugs by correlation spectral matching from the database of second derivative plots of spectra. Spectroscopic range examined 2000 nm - 2400 nm (Top three matches displayed)

True identity	Correlation spectral match value	Best match
Papaverine	1.000 0.538 0.401	Papaverine Omeprazole Caffeine
Desipramine	1.000 0.590 0.581	Desipramine Imipramine Propranolol
Cimetidine	1.000 0.512 0.474	Cimetidine Procaine Ephedrine
Aspirin	1.000 0.699 0.593	Aspirin Orciprenaline Isoprenaline
Allopurinol	1.000 0.532 0.459	Allopurinol Chlorpromazine Isoniazid
Quinidine	1.000 0.660 0.316	Quinidine Quinine Phenylephrine
Testosterone	1.000 0.606 0.369	Testosterone Progesterone Naproxen
Pheniramine - Drug not included in database	0.361 N/A N/A	Benzyl penicillin No other selections No other selections
Paracetamol - Drug not included in database	0.314 N/A N/A	Chlorpropamide No other selections No other selections

Note 'N/A' denotes that there were no match values obtained greater than or equal to 0.1

The results in Table 32 actually show very little in terms of the highest match value as this appears as the maximum of 1.000 anyway. In actual terms, there may have been slight change in this value but it is not visible as the match value is rounded up to 1.000. However, the other second best, third best compounds have now taken on higher correlations. This may indicate that 200 nm is insufficient a region and that 100 is an insufficient number of data points for selective identification compared to either 1100 - 2500 nm or 2000 - 2400 nm.

Table 32 Identification of pure drugs by correlation spectral matching from the database of second derivative plots of spectra. Spectroscopic range examined 2200 nm - 2400 nm (Top three matches displayed)

True identity	Correlation spectral match value	Best match
Testosterone	1.000 0.743 0.732	Testosterone Progesterone Glutethimide
Desipramine	1.000 0.706 0.613	Desipramine Aspirin Isoprenaline
Cimetidine	1.000 0.706 0.575	Cimetidine Procaine Ephedrine
Acetylsalicylic acid	1.000 0.706 0.663	Aspirin Desipramine Orciprenaline
Quinidine	1.000 0.618 0.512	Quinidine Quinine Nicotinamide
Allopurinol	1.000 0.581 0.538	Allopurinol Chlorpromazine Isoniazid
Pheniramine - Drug not included in database	N/A	No suggestion
Paracetamol - Drug not included in database	N/A	No suggestion

Note 'N/A' denotes that there were no match values obtained greater than or equal to 0.1

Generally, the results for the wavelength distance match mirrored those by Correlation. Over the greater wavelength range namely 1100 - 2500 nm all drug actives were now identified correctly with the highest standard deviation match no greater than 2.0 (Table 33), but clearly discernible from the next best matched compound with a wavelength distance Match of 197.3. The method itself is highly discriminating as it is based upon the calculation of maximal standard deviations between the test sample spectra and the referenced database spectra at every other wavelength.

Table 33 Identification of pure drugs by wavelength distance from the database of second derivative spectra (Top three matches displayed). Spectral range 1100 – 2500 nm

True identity	Wavelength distance match value	Best match
Desipramine	1.2 231.2 274.9	Desipramine Procaine Carbamazepine
Cimetidine	1.4 495.2 587.5	Cimetidine Procaine Phenylephrine
Allopurinol	1.9 271.0 328.5	Allopurinol Procaine Oxprenolol
Papaverine	1.2 323.5 354.5	Papaverine Ephedrine Procaine
Aspirin	1.3 197.3 409.1	Aspirin Procaine Ephedrine
Ampicillin	1.5 282.8 381.0	Ampicillin Procaine Carbamazepine
Imipramine	1.2 346.9 637.1	Imipramine Procaine Ephedrine
Pheniramine - Drug not included in database	N/A	No suggestion
Paracetamol - Drug not included in database	N/A	No suggestion

Note 'N/A' denotes that there were no match values obtained greater than or equal to 650 sd

The difference between the two matching techniques is that the distance method, being based upon point distance differences at each wavelength considered is more discriminating than the correlation spectral match, which is based upon a simple correlation (see section 1.10.2). The correlation spectral match is less sensitive to changes in peak intensity if the magnitude of absorption changes uniformly across the spectral range, so this would be thought to lead to mis-identifications of extremely closely related compounds – however, this was not found to be the case in sections 5.3.1 and 5.3.2. However, both spectral matching and comparison by distances were found to be equally effective when examined earlier in these contexts.

5.4 Conclusions

It is possible to identify drug substances use of NIR. It is not possible to provide a 'one method fits all' to achieve this.

Criteria for identification by NIR should have the following objectives:

- Reduce the number of candidates to a small number
- Identify the candidates by chemometrics
- Identify by comparison to reference spectra

In order to this I recommend the following:

Firstly, use second derivative spectra. Secondly, standardise the parameters to be used. Finally, base the approach upon a hierarchy of methods.

The identification methods would be based upon second derivative spectra. These were found to be the best order of spectra to provide reliable identification in this experiment. This is consistent with other identification techniques such as ultra violet spectrophotometry and high performance liquid chromatography and the pioneers of modern derivative spectroscopy such as O'Haver[81,82] and Fell[84,85].

The method order to be used would be :

- Correlation matching – this proved to be the most effective method of identification, is probability based and accounts for variation in peak intensity from different samples of the same substance. It is the method suggested by the CPMP.
- Wavelength distance matching – more effective with substances that are closely related in structure because it is very sensitive to small changes in peak wavelength. The method is most effective when dealing with a small number of candidates.
- Wavelength selection – examine the spectra for wavelength ranges displaying differences peak position and re-execute both the correlation and wavelength distance methods over this reduced wavelength range
- Visually examine the spectra – finally visual comparison of the second derivative spectra of the remaining candidates with the test sample spectrum would now be possible due to the expected (low) number of candidates.

In an industrial scenario, it would be expected that the first and second methods (correlation and wavelength distance) could be run by relatively unskilled workers, such as in a warehouse, following standard operating procedures. The final two methods (wavelength selection and visual comparison) would need skilled workers and would be carried out in a laboratory, such as a quality control laboratory.

I would recommend criteria for matches or critical match values based upon experience of:

- Correlation – 0.95 indicates a pass, with values between 0.85 and 0.95 as a potential pass requiring corroboration by one of the remaining three methods (above)
- Wavelength distance – 5.0 standard deviations indicates a pass, with values between 5.0 and 10.0 standard deviations requiring corroboration by one of the remaining two methods (wavelength range selection or visual examination)

The proviso I would make for these values is that they are based upon the experience of a limited number of batches per sample and from a database of 301 substances. Many more batches per sample and substances overall would need to be investigated to provide a robust figure for the critical match value.

Chapter 6 Identification of drug substances by peak matching

6.1 Summary of aims

The aims of the study are to:

- Determine the effect of window size upon the spectra
- Construct library lists of the 300 pharmaceutical substances based upon peak position and peak intensity relative to the most intense peak in each case, from NIR data
- Validate the library internally and through the use of external samples
- Demonstrate identification of external test drug substances run as 'unknowns' and in a blind manner
- Determine the optimum number of peaks required to demonstrate identification from the library of 300 substances
- Report mis-identifications from drugs presented which do exist within the library
- Report any mis-identifications (based upon the optimum number of peaks selected) from presenting test substances for which there is no reference substance contained within the library

6.2 Introduction

Chapter 5 on Drug Identification demonstrated that it is possible to collect NIR data from a large number of compounds and to store them in a database. The database was validated both internally and externally and a series of unknown substances were identified by two relatively simple chemometric methods, namely correlation spectral matching and wavelength distance matching. Conclusions from the chapter based much of the success of the matches upon use of second derivative calculated spectra and not original spectra. Original spectra were found to be unreliable and led to mis-identifications.

This chapter will propose an identification technique based upon the matching of peak positions. The purpose is to relate peak matching in NIR to the "six peak method" employed in conventional mid infrared spectra[107,108].
The purpose of this study is to apply the sound principles of peak identification that has been routinely applied to mid infrared spectroscopy, on to NIR spectroscopy and to determine how many peaks would be required to facilitate such identification from a database of 300 drug substances. The basis of this work has been founded on the principles of the TLC work (Chapter 3) where library lists of substances were constructed based upon TLC data from between four and eight different solvent systems. Unknowns were then run against the library of reference data with identification based upon the search criteria of mean list length and discrepancy index. The principles of identification from the chapter on TLC have been applied in this chapter.

6.2.1 The six peak method in mid infrared spectroscopy

The six peak method that is used in mid IR involves the use of the six major peaks in the region of 2000 cm^{-1} to 650 cm^{-1} for the identification of substances[104]. The region of 1490 cm^{-1} to 1320 cm^{-1} are omitted to avoid complications arising due to absorbance by Nujol. Peaks are chosen only if they are completely resolved and the maximum of the peak is easily detectable, i.e. shoulders will not be used. If the major peak is a hump and the maximum point is not easily detectable, the wavelength of the peak mid-point is used.

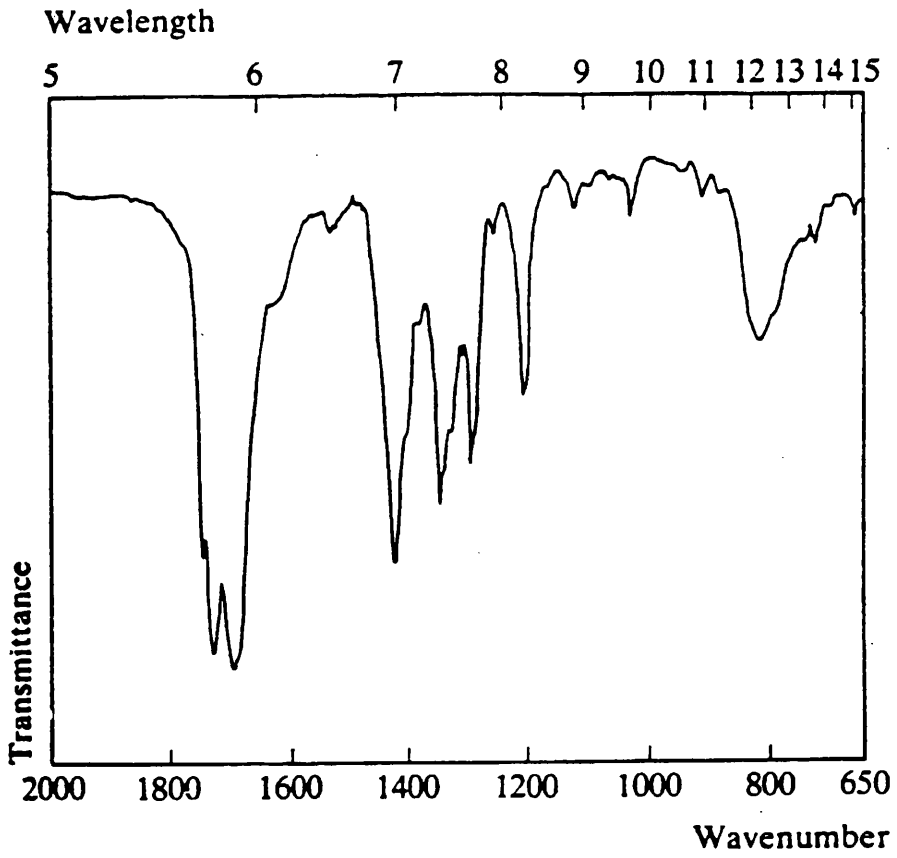


Figure 52 Mid infrared spectrum of cyclobarbitone $C_{12}H_{16}N_2O_3$ in Nujol with principal peak positions at 1693, 1725, 1745, 1300, 1210, 830 cm^{-1}.

The region between 1490 cm^{-1} to 1320 cm^{-1} has been excluded from the peak positions due to Nujol interference.

Differences in instruments and conditions used, lead to variations in peak amplitude. This can reverse the order of the peaks and has to be considered when using the index.

Identification is obtained by matching the wavenumbers of the (first, second, third etc) most intense peaks in order to produce a unique match.

The wavenumber positions of the six most intense peaks for a series of reference compounds are recorded in a table, in order of intensity. To identify a compound, the wavenumber of the most intense peak is compared against the table to within $\pm$ 4cm^{-1}. From the resulting matches, the remaining peaks would then be used to determine the compound. If no match is obtained after using the most intense peak, the remaining peaks are compared in turn against the table in order to find any peak matches and suggest any possible compounds. Visual comparison of the spectra would be recommended here.

Figure 52 shows an infrared spectrum from a chemical substance, giving wavenumbers of the six most intense peaks. To identify the compound, the wavenumber of the most intense peak is compared against Table 34 and the compounds within $\pm$ 4cm^{-1} selected. All compounds displayed in Table 34 can be seen to match the criteria. The remaining wavenumbers from the compound in Figure 52 were then used to determine which was the best match against the table. The compound was therefore identified as cyclobarbitone. The index table was taken from the list of infra red peaks from the index produced from Clarke's Isolation and Identification of Drugs[104]. These methods will be modified slightly in this study by automating the searches to be able to identify peak

positions rapidly and to be able to identify peak positions from a large number of compounds within a reasonable time period.

Table 34 Index of the six principlal mid infrared peaks from compounds with the most intense peak between 1689 and 1697 cm-1.

Wavenumbers cm^{-1}						Compound
1689	1718	1550	4640	1285	672	Enallylpropymal
1689	1232	1270	857	1606	1149	Methallenoestril
1690	1720	1740	1310	1290	1200	Butalbital
1690	1610	1041	1170	1150	1010	Emylcamate
1690	1235	1497	1598	1155	750	Fluanisone
1690	1640	1602	835	665	880	Muzolimine
1690	1225	1527	1120	1496	1310	Nifedipine
1690	1610	698	1536	745	784	Nitrazepam
1690	1088	1064	700	1605	1623	Phenprobamate
1690	1274	1605	1174	1116	772	Procaine Hydrochloride
1690	1205	1170	1310	1620	1270	Salsalate
1690	1665	1221	1550	1595	680	Theobromine
1692	1587	916	1224	1235	956	Allopurinol
1692	1730	1750	1318	1220	1630	Vinylbitone
1693	1720	1745	1316	1255	860	Aprobarbitone
1693	1725	1745	1300	1210	830	Cyclobarbitone
1694	757	1587	1162	1204	1219	Benperidol
1694	660	1600	1094	1212	834	Carbromal
1694	729	1162	1282	1298	934	Ethebenecid
1694	1149	1587	1086	1612	684	Sulphaurea
1694	1260	1520	1155	1140	1222	Tetrabenazine
1694	1176	884	819	675	1086	Tolazamide
1695	1148	1220	1267	1000	1252	Alphaxalone
1695	1185	1200	1515	1623	1267	Aminoglutethimide
1695	1215	1199	1153	1587	1280	Bumetanide
1695	1527	1075	1246	1101	1319	Carisoprodol
1695	1520	1229	1610	1175	1270	Chlorambucil
1695	1613	1497	1263	1205	1558	Decoquinate
1695	1220	707	930	773	960	Flurbiprofen
1695	1752	1265	833	1205	775	Nealbarbitone
1695	1505	753	1220	1230	825	Pimozide
1695	1543	766	754	1299	1258	Prilocaine Hydrochloride
1695	1250	1065	1538	1600	1140	Tybamate
1696	1727	1760	1242	850	1215	Butobarbitone
1696	1728	1755	835	1290	1207	Idobutal
1696	747	1142	1120	712	1027	Ketamine
1696	1619	1574	1260	1100	1038	Nabilone
1697	1620	1490	1528	1107	783	Flunitrazepam

A peak table such as this is commonly available in sources such as Clarke's Isolation and Identification of Drugs[104].

A significant difference between infrared and NIR is that the NIR absorbances are 10-1000 times weaker, and consist of many more peaks in the 2nd derivative (commonly greater than 50 in a powdered sample).

As was noted in Chapter 3, the criteria for pharmacopoeial drug identification by TLC involved wording such as "similar to" or "corresponds to". Rarely was it possible to suggest how unique a particular identification appears or to put a value on the reliability of identification. The same is true of pharmacopoeial identification by mid infrared, with commonly used wording such as "correspond in position and relative intensity..." (British Pharmacopoeia[5]) and "correspond in position and relative size..."

Published identifications using NIR employ search methods based upon a variety of complex mathematics algorithms and criteria as discussed earlier. Examples of this are Mahalanobis distances[105], principal component analysis[106], wavelength distance[80] matching and correlation in wavelength space[80]. When considering other analytical techniques, it is common practice to use peak positions for identification. For example there is a listing of the six most intense mid IR peaks for over 1000 compounds in Clarke's Isolation and Identification of Drugs[104]. This is not uncommon: listings such as these are available for many techniques such as UV – visible spectroscopy, mass spectrometry and nuclear magnetic resonance spectroscopy.

Curry et al[107] described the use of a simple infrared retrieval system based upon peak identification. Peaks were defined as those nearest to 0% transmission and shoulders were not accepted as they were not completely resolved. Two methods of peak selection were utilised. In the first, the six most intense peaks were selected, in the second the most intense peaks from each of five band ranges distributed evenly across the spectrum were selected. This ensured that the fingerprint region would be represented. 50 spectra were used, comprising 25 duplicated spectra from within the library and 25 spectra for which reference curves were present in the library. The retrieval was based upon manual matching of peak wavelength positions to $\pm$ 0.1 μm.

The six peak identification provided the best performance, correctly retrieving all 25 spectra based upon either five or six peak matches. The five-band range method failed to retrieve the correct spectra in four cases. The cause of failure was attributed to the inclusion of unreliable peaks of low absorption that would not have qualified for inclusion into the six peak method. The group achieved

over a 95 % success rate using the six most intense peaks when searching through up to 10,000 spectra of forensic interest.

Ingle et al[108] ran a study based upon sampling and retrieving from a spectral list of 230 commonly used drugs. Test samples were produced by a variety of personnel using different instruments in 11 different laboratories. A 90 % success rate was achieved based upon use of the six strongest absorption bands in conjunction with visual comparison of unknown against authentic spectra.

6.3 Method

The spectra of over 300 powdered drug substances (the list of drugs can be found in the Appendix section) were analysed by the standard procedure described in Chapter 2 in section 2.2.

6.3.1 Effect of window size upon the spectra

The effect of using different smoothing parameters was determined on 14 barbiturate samples. The smoothing parameter window size was varied using an in-house software program called 'SG' written by Dr Roger Jee, School of Pharmacy, London. This provided a convenient way of selecting the smoothing parameter before applying it to a list of spectra. The parameter was varied from 4 4 4 2, 10 10 4 2 through to 20 20 4 2. These meaning of these parameters is defined in section 1.10.1.

Peak positions were determined for each of the barbiturates using each of the above smoothing parameters.

6.3.2 In-house programs

In-house programs written by Dr Jee, were used to create a list of the spectral files and retrieve spectral information. The programs created are listed below in Table 35.

Based upon the library of spectra from over 300 substances that were initially scanned (section 2.4.2) the individual spectra files were converted into data interchange format ('.dif') files using the NSAS standard application software (version 3.26) and IQ^2 chemometric software (version 1.13).

The 'DNA' program was used to calculate the mean spectrum from the spectra contained within the '.dif' files.

The 'SG' program was used to convert the mean zero order spectrum contained within each of the 'dif' files from each of the 300 substances into second derivative format and apply the Savitsky-Golay smoothing parameters 5 5 4 2. These parameters were kept constant throughout the course of the study.

The 'PEAK' program (Table 35) was then used to identify the position of each peak within each of the spectral files. This involved normalising the peaks by searching each of the files, locating the most intense minimum 'peak' and allocating a relative intensity of –1.000. Each of the remaining minima is then compared to this and allocated a comparative value relative to –1.000. The range of values was between –1.000 and –0.001 and relative in terms of intensity. The wavelength position of each 'peak' was also recorded.

From this output a library of substances could then be constructed containing a list of substance names in association with the positions (in nm) and relative

intensities of each peak. Three libraries were built. The first contained the list of drugs identified by the six most intense peaks, the second library contained drugs listed by the twelve most intense peaks and the third used the twenty five most intense peaks. The peaks were ordered in ascending order of most intense peak position.

Note: The most intense peaks in the second derivative are defined as the most intense minima.

Table 35 Details of the programs created in-house and used as part of the study into peak identification.

Program Name	Version Number	Version Date	Brief Description
SG	0.13PC	24/04/98	Program for converting spectra into derivatives and applying smoothing parameters
PEAK	0.10PC	05/03/97	Program for normalising the peaks and selecting the *n* most intense peaks from a sample file. *n* is defined as an integer between 1 and 99
PMATCH	V0.11A	13/02/98	Program to match the *n* number of most intense peaks derived from the PEAK program against the library list of 300 substances
RMATCH	0.07PC	06/06/98	Program to score the peak matches based upon peak position, relative peak intensity and peak order.
DNA	VO.58	26/04/98	Program to calculate polar co-ordinates, centres of gravity and scanning probability windows

6.3.3 Internal validation of the library

The purpose of validating a database is firstly to determine if it is possible to distinguish each substance within the library from each other and secondly to

determine if the library is robust enough to be able to identify spectra from different external samples and batches. Internal validation covers the first part – distinguishing between the library reference samples. To achieve this, the spectra from each substance in the library must be compared against the spectra from each other substance.

The comparisons were carried out by taking the first product in the library of six most intense peaks and running a comparison against all the products in that library to gain a list of possible peak matches. This was repeated for the twelve and twenty five peak libraries.

6.3.4 External validation of the library

The second part of the validation cycle is to ensure that the library is robust enough to be able to identify external samples and batches that will contain variations in physical characteristics and possibly in purity of compound. This was carried out by running a series of 59 external samples against the library and comparing peak matches based upon six, twelve and twenty five most intense peaks.

The samples were analysed by NIR and then compared against each of the three drug substance libraries. The regimen used to analyse and compare the samples is the same as described below in section 6.3.5.

6.3.5 Test substances

34 test samples of powdered drug substances (Appendix 3, Table 64) were taken from a variety of external sources and used as unknowns in a blind identification study. Each sample was placed on the NIR instrument and

scanned with the resultant spectra compared against the library of drug substances for six, twelve and twenty five peaks.

The actual spectra were saved as '.da' and '.cn' files and converted in '.dif' format within the NSAS program. As with the library list of substances, each program defined within Table 35 was used in turn to :

- calculate the mean spectrum for each substance
- convert to second derivative
- apply Savitsky-Golay parameters of 5 5 4 2 (see 6.4.1)
- identify the peak position of the most intense peak, allocating a relative intensity of –1.000
- to identify all remaining peaks and allocate relative intensities accordingly while recording all peak positions
- match each test substance against the library list of substances created earlier

When comparing the test substances against the library the number of peaks required for a match were varied. This was defined by a mandatory user input field in the 'PMATCH' program. A peak match was defined as being ± 2 nm, as the Savitsky-Golay calculation was based upon consecutive data points. The distance between two consecutive data points was 2 nm.

Table 36 List of drug substance database and sample files constructed and used in the study

File name	File type	Description
pdrugs6	Database	6 most intense peaks
pdrugs12	Database	12 most intense peaks
pdrugs25	Database	25 most intense peaks
pun6	Unknown test samples	6 most intense peaks
pun12	Unknown test samples	12 most intense peaks
pun25	Unknown test samples	25 most intense peaks
Pkn6	Unknown test samples	6 most intense peaks
Pkn12	Unknown test samples	12 most intense peaks
Pkn25	Unknown test samples	25 most intense peaks

6.4 Results and Discussion

6.4.1 Effect of 2nd derivative parameters

The peak intensities, wavelengths at which peaks appear and the number of peaks are affected by varying the smoothing parameter window size. Calculation of the second derivative by the Savitsky-Golay method introduces the use of smoothing parameters as a consequence. It is not possible to use the method for calculation of the second derivative without the consequence imposed by the smoothing parameters. The effects of these parameters are examined below.

If the there is an effect upon peak intensity, the six principal peaks selected may vary, the order of the peaks may also change and identification using the six peak method will be altered.

The effects are demonstrated for three of the barbiturates in Figure 53, Figure 54 and Figure 55.

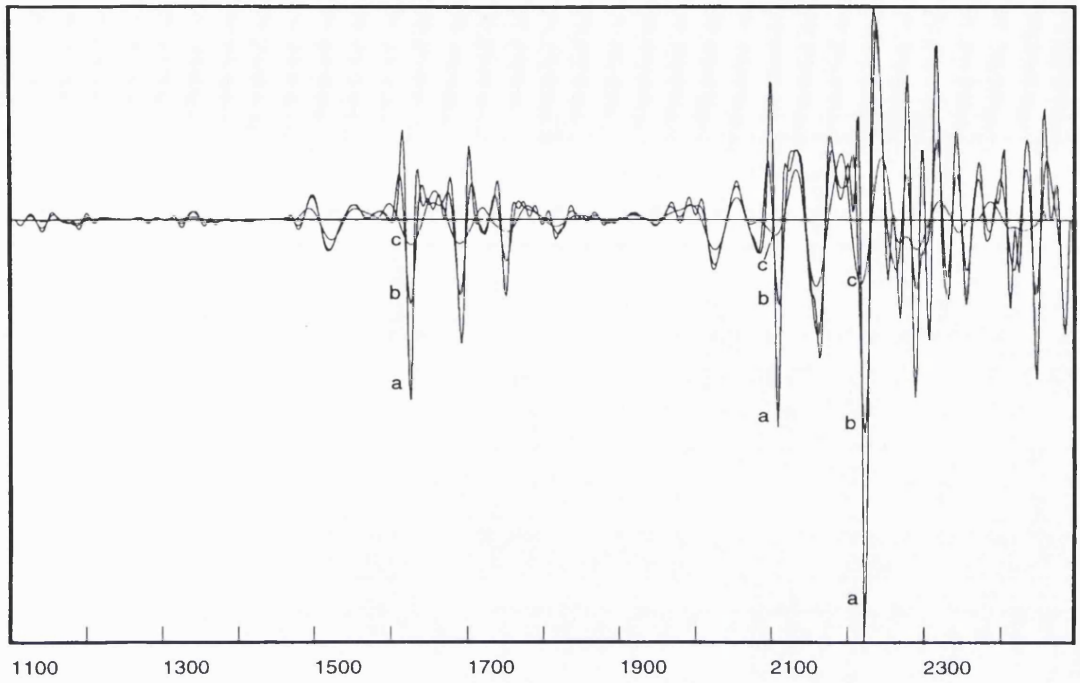


Figure 53 Different smoothing factors applied to cyclopentobarbitone. The Savitsky-Golay parameters are a) 4 4 4 2, b) 10 10 4 2 c) 20 20 4 2

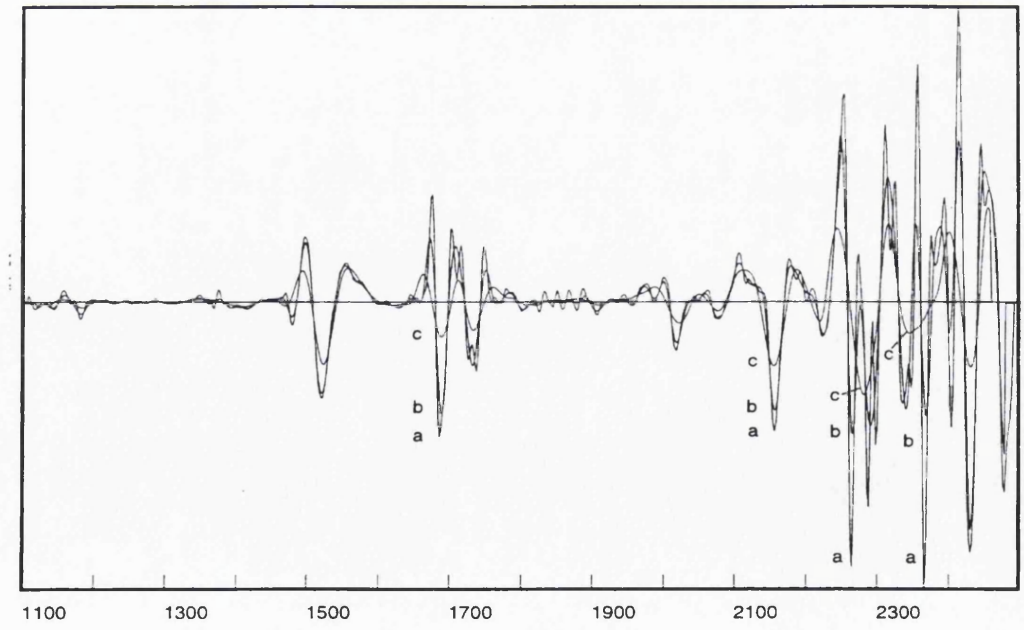


Figure 54 Different smoothing factors applied to butobarbitone carboxylic acid. The Savitsky-Golay parameters are a) 4 4 4 2, b) 10 10 4 2 c) 20 20 4 2

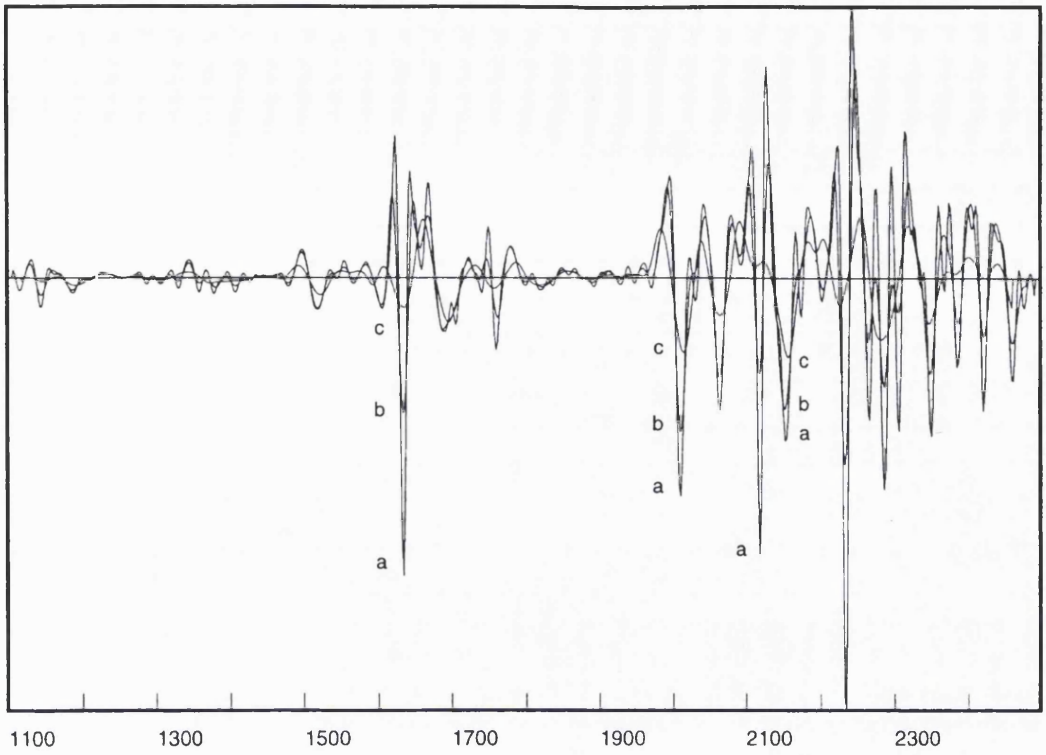


Figure 55 Different smoothing factors applied to quinalbarbitone. The Savitsky-Golay parameters are a) 4 4 4 2, b) 10 10 4 2 c) 20 20 4 2

Table 37 Peak intensities for Barbiturate 1, cyclopentobarbitone

Comparison of Savitsky-Golay parameters 4 4 4 2, 10 10 4 2 and 20 20 4 with the spectra from cyclopentobarbitone

Barbiturate 1 (cyclopentobarbitone)						
	4 4 4 2		**10 10 4 2**		**20 20 4 2**	
	Wavelength	Intensity	Wavelength	Intensity	Wavelength	Intensity
PEAK No.						
1	2222	-1.000	2222	-1.000	2158	-1.000
2	2113	-0.492	2159	-0.547	2218	-0.969
3	1626	-0.433	2111	-0.405	2026	-0.453
4	2288	-0.426	1625	-0.403	2290	-0.444
5	2448	-0.382	1698	-0.358	1626	-0.374
6	2161	-0.333	2445	-0.355	1689	-0.356

In Table 37, the most intense peak from the barbiturates appears to be affected the least. However in butobarbitone carboxylic acid (Table 38) the most intense peak becomes the fifth most intense when varying parameters from 4 4 4 2 to 10 10 4 2.

The second most intense peak shows distortion and begins to appear as the third or even fifth most intense peak (and vice versa), for example in cyclopentobarbitone the sixth most intense becomes the second most intense when changing from 4 4 4 2 to 10 10 4 2.

Table 38 Peak intensities for Barbiturate 2, butobarbitone carboxylic acid

Comparison of Savitsky-Golay parameters 4 4 4 2, 10 10 4 2 and 20 20 4 with the spectra from substance butobarbitone carboxylic acid

Barbiturate 2 (butobarbitone carboxylic acid)						
	4 4 4 2		**10 10 4 2**		**20 20 4 2**	
	Wavelength	Intensity	Wavelength	Intensity	Wavelength	Intensity
PEAK No.						
1	2365	-1.000	2429	-1.000	2282	-1.000
2	2264	-0.900	2477	-0.676	2432	-0.690
3	2429	-0.856	2265	-0.586	2155	-0.685
4	2288	-0.702	2291	-0.548	1524	-0.665
5	2477	-0.654	2358	-0.503	1689	-0.377
6	2298	-0.494	1689	-0.492	2347	-0.341

As can be seen from Figure 53, Figure 54 and Figure 55 there are numerous intensity changes, however of potentially greater significance are the changes in peak position. For example in butobarbitone carboxylic acid, the peak at wavelength 2264 in 4 4 4 2 is noted at 2282 in the 20 20 4 2 set. Reproducibility of peak positions was key to the identification described in the chapter describing drug identification (Chapter 5) and was found to be reproducible despite varying several physical factors (Chapter 4, Sample presentation for NIR spectroscopy).

Table 39 Peak intensities for Barbiturate 3, quinalbarbitone

Comparison of Savitsky-Golay parameters 4 4 4 2, 10 10 4 2 and 20 20 4 with the spectra from quinalbarbitone.

Barbiturate 3 (quinalbarbitone)						
	4 4 4 2		**10 10 4 2**		**20 20 4 2**	
	Wavelength	Intensity	Wavelength	Intensity	Wavelength	Intensity
PEAK No.						
1	2235	-1.000	2231	-1.000	2155	-1.000
2	1634	-0.696	2010	-0.756	2014	-0.942
3	2117	-0.645	1635	-0.668	2282	-0.783
4	2009	-0.508	2151	-0.648	2347	-0.580
5	2285	-0.495	2285	-0.535	1697	-0.567
6	2153	-0.381	2063	-0.488	2065	-0.477

Peak position was also identified as being directly linked to the chemical structure of the compound therefore it is extremely important to state the smoothing parameter being used and to keep this constant throughout a series of identifications. When constructing tables or libraries of reference NIR data it is important that the smoothing parameter being used must be stated. Generally, as the smoothing parameter is increased the peaks become broader and the narrowest peaks less intense. Information is also lost as the total number of peaks is seen to decrease. The lower number of points, i.e. 4 4 4 2 provided the most well defined peaks and with the highest relative intensity and would therefore be best for identification purposes. The important point, however, is that whenever comparing spectra or basing identification, the database samples and the test samples must have been derived using the same parameters.

The parameter used for the database of the 300 substances was 5 5 4 2. This was selected in order to produce well defined and relatively 'sharp' peaks as in

Figure 53, however the important point is that the 5 5 4 2 parameter was kept constant throughout this experiment.

6.4.2 Internal validation

The purpose of the internal validation is to determine if the library of reference drug substances is suitable for use as such a library. A key criterion for this is that all the reference substances within the library must be distinct from each other, in other words must be identifiable from each other. To determine this it is necessary to compare each substance within the library against every other substance.

For peak identification, three reference libraries were compiled. These were based upon the original database of drug substances used in Chapter 5, Drug identification. Some compounds, however, for which there were only small amounts available (e.g. a few milligrams) were removed, leaving 301 compounds. The list of compounds used in this chapter can be found in Appendix 2. Identification tests were run by comparing the six, twelve and twenty five most intense peaks of these substances and test samples (Appendix 3).

When performing the internal validation, results differed only slightly across the three libraries. Each match set contained 45150 pairs. The tables (below) show the number of different pairs of compounds giving 0, 1 etc peak matches.

No two compounds for the six peak library had all six peaks in common, which therefore allowed all compounds to be distinguished by use of all six peaks. There were however, five pairs of compounds with five peaks in common and upon further investigation it was clear that some closely related substances

(e.g. hydrocortisone acetate and hydrocortisone alcohol) and palmitic acid, triacontane, triheptadecanoin and triacontane would be difficult to distinguish between. This was a consistent theme through the higher number of peak matches.

Table 40 Internal validation of a library of 301 different pharmaceutical substances based upon six peak match.

The total number of possible pairs was 45150. Peaks were defined as matching if within ± 2nm of each other. Peak order was not considered. Database file: pdrugs6, Unknowns file: pdrugs6

Peaks	Number of Pairs	Percentage of Total	Correct Pairs
0	34253	75.9	-
1	9596	21.3	-
2	1179	2.6	-
3	103	0.23	-
4	14	0.031	
5	5	0.0011	
6	0	0	

There were also further pairs of substances with four peaks in common. These included ethyl para-aminobenzoate and benzocaine, lignocaine hydrochloride and iso-chlortetracycline hydrochloride, orphenadrine citrate and atropine, chlorpropamide and 4-chlorobenzene sulfonamide, chlorpromazine and prochlorperazine, benzyl cinnamate and pamidronate disodium, amylase lipase and protease, palmitic acid and triheptadecanoin, triacontane, cetyltrimethyl ammonium chloride, cetylpyridinium bromide monohydrate, sodium valproate and penicillamine. For four or five peaks to have thrown up the number of pairs of compounds as is seen, this suggests that translating an entire spectrum of information into merely six peaks and basing identification upon this introduces a high element of risk of a mis-identification. This risk would increase as the size of the reference library increases and would be extremely high if a libraray were to be constructed of thousands of substances including many that are

closely related. This validation has shown however, that as few as six peaks are enough to produce uniqueness of identity from the 300 plus substances.

Table 41 Internal validation of a library of 301 different pharmaceutical substances based upon twelve peak match.

The total number of possible pairs was 45150. Peaks were defined as matching if within ± 2nm of each other. Peak order was not considered. Database file: pdrugs12, Unknowns file: pdrugs12

Peaks	Number of Pairs	Percentage of Total
0	17375	38.48
1	17128	37.94
2	7760	17.19
3	2279	5.05
4	483	1.07
5	87	0.19
6	20	0.04
7	9	0.02
8	2	0.004
9	2	0.004
10	3	0.007
11	2	0.004
12	0	0.33

The twelve peak validation (Table 41) provides a greater opportunity for specificity with more peaks required for a match and more spectral information being considered. The results were similar as for the six peak validation, with no conflicting pairs at twelve peaks. In addition to this, far less pairs of compounds existed at nine, ten and eleven peaks.

Table 42 Internal validation of a library of 301 different pharmaceutical substances based upon twenty five peak match.

The total number of possible pairs was 45150. Peaks were defined as matching if within ± 2nm of each other. Peak order was not considered. Database file: pdrugs25, Unknowns file: pdrugs25

Peaks	Number of Pairs	Percentage of Total
0	2028	4.5
1	6925	15.3
2	10810	23.9
3	10571	23.3
4	7580	16.7
5	4215	9.3
6	1928	4.3
7	723	1.6
8	235	0.52
9	74	0.16
10	23	0.05
11	14	0.03
12	9	0.02
13	2	0.004
14	0	0
15	4	0.009
16	1	0.002
17	1	0.002
18	4	0.009
19	0	0
20	2	0.004
21	0	0
22	0	0
23	0	0
24	1	0.002
25	0	0

Twenty five peaks provided the greatest amount of information within this study upon which to uniquely identify the substances from each other.

All compounds (Table 42) were differentiated correctly with twenty five peaks.

In addition only one pair of substances provided and confliction with twenty four peaks similar – these being hydrocortisone acetate and hydrocortisone alcohol.

There were no further conflicts until twenty peaks, providing a greater comfort level between like compounds than the lower peak validations.

In general across all three libraries all compounds were successfully distinguished from each other with the maximum number of peaks in each

case. However, some examples involving closely related compounds did produce conflicting matches and in some cases the number of peaks was also high. It is anticipated that issues over closely related compounds will always be expected. To minimise these, visually comparing spectra is required as a backup which is no worse than other spectroscopic techniques. NIR, however, does possess other significant advantages over the other techniques in terms of sample preparation, analysis and the ability to derive physical as well as chemical information.

6.4.3 External validation

Whereas the internal validation compares each reference substance in the library against each other reference substance, the external validation seeks to compare a set of external test samples against the library. This then introduces the type of factors that would vary between batches and manufacturers such as substance purity, quality, water content and physical variances such as particle size and shape. A degree of robustness can then be either assured or built into the library.

59 substances were chosen as a test set being approximately 20 % of the overall size of the library.

From the external validation upon the six peak library (Table 44) 98.2 % of test substances were successfully identified with the number of peaks ranging from six to four. All compounds with six peak matches were uniquely and correctly matched. With five peaks, there were three anomalies (Table 43).

Table 43 External validation using 59 external powdered drug samples with the six peak match.

The external samples were compared against the library of 301 different pharmaceutical substances. Searches were based upon matching the six most intense negative peaks. Peaks were defined as matching if within ± 2nm of each other. Peak order was not considered. Database file: pdrugs6, Unknowns file: pkn6. Data displayed is 4 or more peaks matched using second derivative spectra.

Test Sample Id	Number of Peaks matching library	Library id suggested	Suggested compound	Result
Rw003	6	Sa016	Chlororquine phosphate	Correct
Rw006	5	Rw001	Cefuroxime	Correct
Rw007	5	Rw002	Ampicillin	Correct
Rw009	6	Rw001	Cefuroxime	Correct
Rw010	5	Rw004	Thiopentone	Correct
Rw012	6	Rw005a	Benzyl penicillin	Correct
Sa018	5	T0504	Cetyltrimethyl ammonium chloride	Correct
T0052	6	T0337	Caffeine	Correct
T0069	6	Rw005	Procaine	Correct
T0085	5	Sa007	Ascorbic acid	Correct
T0251	6	T0255	Allopurinol	Correct
T0252	6	T0255	Allopurinol	Correct
T0253	6	T0255	Allopurinol	Correct
T0254	6	T0255	Allopurinol	Correct
T0256	6	T0262	Lactose monohydrate	Correct
T0257	6	T0262	Lactose monohydrate	Correct
T0258	6	T0262	Lactose monohydrate	Correct
T0259	6	T0262	Lactose monohydrate	Correct
T0260	6	T0262	Lactose monohydrate	Correct
T0261	6	T0262	Lactose monohydrate	Correct
T0277	6	T0255	Allopurinol	Correct
T0278	6	T0255	Allopurinol	Correct
T0329	6	T0337	Caffeine	Correct
T0340	6	Sa019	Chlorthalidone	Correct
T0349	6	T0341	Diltiazem	Correct
T0352	6	T0317	Trans-2-[4-(1,2-Diphenyl-1-butenyl)phenoxy]-N,N-dimethylethylamine	Correct
T0355	6	T0262	Lactose monohydrate	Incorrect* *Correct is Glyceryl trinitrate
T0376	5	T0262	Lactose monohydrate	Incorrect *Correct is Isosorbide-5-mononitrate
T0377	6	T0382	Iminodibenzyl	Correct
T0384	6	T0373	Irgasan	Correct
T0392		*Sa041 is the correct match and was present in the database but not matched*	No match	Incorrect* *Correct is methyl hydroxybenzoate
T0397	6	Sa044	Nalidixic acid	Correct
T0398	5 6	T0370 T0383	Hydrocortisone acetate Hydrocortisone alcohol	Incorrect Correct
T0406	6	T0262	lactose Monohydrate	Incorrect* *Correct is Nitroglycerin
T0435	4	T0420	Palmitic acid	Incorrect

Test Sample Id	Number of Peaks matching library	Library id suggested	Suggested compound	Result
	6 5 4 4	T0451 T0453 T0504 T0506	Triheptadecanoin Triacontane Cetyltrimethyl ammonium chloride Cetylpyridinium chloride monohydrate	Correct Incorrect Incorrect Incorrect
T0444	6	Sa054	Thioridazine hydrochloride	Correct
T0448	6	T0408	Reserpine	Correct
T0452	5	T0413	Lorazepam	Correct
T0454	6	T0408	Reserpine	Correct
T0477	4 6 5 4 4	T0420 T0451 T0453 T0504 T0506	Palmitic acid Triheptadecanoin Triacontane Cetyltrimethyl ammonium chloride Cetylpyridinium chloride monohydrate	Incorrect Correct Incorrect Incorrect Incorrect
T0483	6	T0224	Cyclobarbitone	Correct
T0488	6	Sa020	Carbromal	Correct
T0499	6	T0500	Avicel	Correct
T0502	6	T0060	Quinine	Correct
Ch001	4	T0238	Chlortetracycline	Correct
Ch002	5	T0238	Chlortetracycline	Correct
Ch003	4	T0238	Chlortetracycline	Correct
Ch004	5	T0238	Chlortetracycline	Correct
Ch005	4	T0238	Chlortetracycline	Correct
Sa055	5	Sa004	Potassium hydrogen phthalate	Correct
Sa056	4	Sa004	Potassium hydrogen phthalate	Correct
Sa057	5	Sa004	Potassium hydrogen phthalate	Correct
Sa058	6	Sa004	Potassium hydrogen phthalate	Correct
Sa059	6	Sa004	Potassium hydrogen phthalate	Correct
Sa060	5	Sa004	Potassium hydrogen phthalate	Correct
Sa061	4	Sa004	Potassium hydrogen phthalate	Correct
Sa062	5	Sa004	Potassium hydrogen phthalate	Correct
Sa063	5	Sa004	Potassium hydrogen phthalate	Correct
Sa064	6	Sa004	Potassium hydrogen phthalate	Correct

Table 44 External validation of a library of 301 different pharmaceutical substances based upon a six peak match of 59 external samples and batches – Summary of pairs

Peaks were defined as matching if within ± 2nm of each other. Peak order was not considered.
Database file: pdrugs6, Unknowns file: pkn6

Peaks	Number of Pairs	Percentage of Total	Number Correct	Percentage Correct
0	13751	77.4	-	-
1	3540	19.9	-	-
2	365	2.1	-	-
3	36	0.20	-	-
4	11	0.06	5	8.4
5	17	0.10	14	23.7
6	39	0.22	39	66.1
Total	17759	100	58	98.2

The nitrate based glyceryl trinitrate and isorbide-5-mononitrate were all identified as lactose monohydrate, not totally unexpected because of the low active concentrations (<1 % m/m) were mixed with lactose. Of the 59 tests there were only three occasions of more than one compound being suggested as a match (at four or more peaks). Two of these examples were for palmitic acid, triheptadecanoin, cetyltrimethyl ammonium chloride and cetylpyridinium chloride monohydrate when the external sample was triacontane.

The fact that three substances had four peaks in common with the actual test substance suggests that four out of six peaks is not enough to be sure of a match. The correct substance was, however, identified in the above cases as being the only match with all six peaks.

The third example was also a very close match between hydrocortisone acetate and hydrocortisone alcohol, with five peaks out of six unable to discriminate the two forms. They were discriminated by the sixth peak, so this is not regarded as a mis-match rather than requiring the full six peaks. One substance could not be matched and this is considered below. However, for 55 out of the 59 samples to have been identified correctly based solely upon a simple six peak match is fairly impressive.

There was one substance that could not be indentified (T0392) which required further investigation. Firstly the reference substance spectra and test spectra were compared visually. This was to ascertain if the two substances were in fact similar or if there had been any pollution to either sample. Both samples should have been methyl para-hydroxybenzoate.The results can be seen. Visually, peak and shoulder positions match. The differences are in intensity with the spectra in the external sample being about three times weaker in relative intensity than the reference sample (Figure 56).

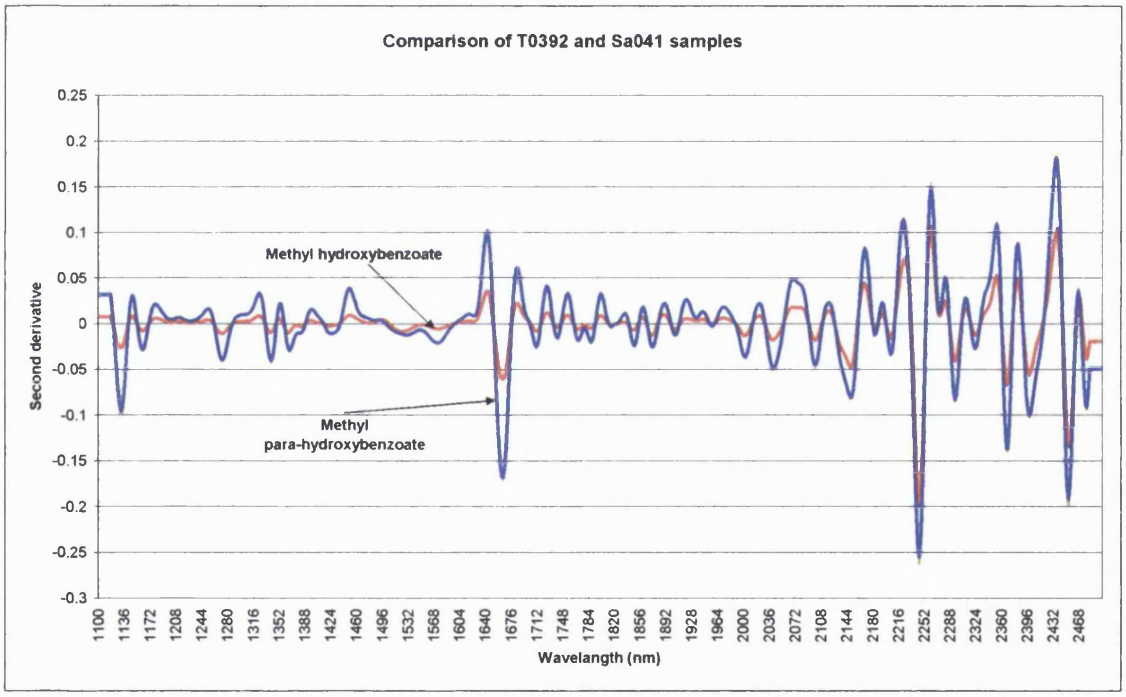


Figure 56 Comparison of the second derivatives of the external sample of methyl hydroxy benzoate and the internal reference of methyl para-hydroxybenazoate.

The external sample was derived from the Medicines Testing Laboratory, Edinburgh and the reference sample was obtained from the internal reference standards from The School of Pharmacy.

This may have accounted for no match being detected, no other sample was selected during the database search, so this was not a mis-match, there was no

match. The evidence does suggest that the samples are the same substance but that the peak absorptions from the external sample were not strong enough to be detected. To ascertain the exact identity of the samples a primary analytical technique such as mid Infrared or UV spectrophotometry would need to be used, it was not the important factor as both samples appeared to be the same substance. The question is why did they not match in the search?

Both sets of spectra were mathematically transformed to their second derivatives and compared. The sets of derivatives are overlayed in Figure 56, where the troughs, peaks and shoulders again show a near identical match. The differences in relative intensity, however, can be seen clearly. To confirm whether the low absorption of the external sample was the cause, the sample should be re-run. The suspicion is that a low amount of powdered sample in the sampling vial may have caused the problem (the reference standards were in short supply.) This was not actually undertaken but with access to the materials with hindsight would have been.

The utlisation of more peaks improves some of the scenarios encountered with the six peaks of compounds with peaks in common, for example cetyltrimethyl ammonium chloride and cetylpyridinium chloride monohydrate are removed from the list of similarities with triacontane (Table 45).

All compounds that produced a full twelve peak match were identified correctly and for all except one sample a match was found with 10 or more peaks. The remaining sample was matched with nine peaks.

The table also shows the number of peaks that were common in some cases to the incorrect sample in addition to the correct match having been supplied with a higher number of peaks. For example, where the external sample was

triheptadecanoin (t04350) this was identified correctly with twelve peaks but also produced matches to triacontane and palmitic acid with ten and nine peaks respectively. This is due once again to the structural similarity of these compounds as discussed with the six peak match.

This does highlight that when considering the twelve most intense peaks, ten peaks are not enough to guarantee a match in all cases. However, there was one sample, chloroquine phosphate that was correctly identified with nine peaks. So a lower number of peaks may produce a correct match but visual comparison would be highly recommended before drawing conclusions in these cases.

Perfect matches (i.e. six, twelve, twenty five) peaks can not be guaranteed for a correct match because physical characteristics will impact the peak intensities and make it likely that not all six, twelve etc peaks will be the same most intense in all cases. The case of methyl para-hydroxybenzoate is interesting because (as discussed earlier) there was no match with six peakes. However, when using the twelve most intense peaks, ten peaks do match to the reference sample of methyl para-hydroxybenzoate. In addition, looking forward to the twenty five peak match, there is also a match here. Potentially, the six peaks lacked sufficient absorbance to make suitable matches but a higher number of peaks, in other words more spectral information, has led to a more accurate search and also suggests that lower concentrations of sample could be identified. This has not worked, however, for the glyceryl trinitrate/ isosorbide nitrate samples where the instensity of the strongly absorbing lactose has been the overriding factor. This raises questions for substance matrices, with lactose being a common excipient, and suggests that low concentrations of active (<

1%) may not be detectable if mixed with lactose. This will become apparent when examing whole tablets in the next chapter (Chapter 7).

Table 45 External validation using 59 external powdered drug samples with the twelve peak match.

The external samples were compared against the library of 301 different pharmaceutical substances. Searches were based upon matching the twelve most intense negative peaks. Peaks were defined as matching if within ± 2nm of each other. Peak order was not considered. 8 or more peaks displayed using second derivative spectra. Database file: pdrugs12, Unknowns file: pkn12.

Test Sample Id	Number of Peaks matching library	Library id suggested	Suggested compound	Result
Rw003	9	Sa016	Chlororquine phosphate	Correct
Rw006	11	Rw001	Cefuroxime	Correct
Rw007	11	Rw002	Ampicillin	Correct
Rw009	12	Rw001	Cefuroxime	Correct
Rw010	11	Rw004	Thiopentone	Correct
Rw012	11	Rw005a	Benzyl penicillin	Correct
Sa018	10	T0504	Cetyltrimethyl ammonium chloride	Correct
T0052	11	T0337	Caffeine	Correct
T0069	12	Rw005	Procaine	Correct
T0085	11	Sa007	Ascorbic acid	Correct
T0251	12	T0255	Allopurinol	Correct
T0252	12	T0255	Allopurinol	Correct
T0253	12	T0255	Allopurinol	Correct
T0254	12	T0255	Allopurinol	Correct
T0256	11	T0262	Lactose monohydrate	Correct
T0257	11	T0262	Lactose monohydrate	Correct
T0258	12	T0262	Lactose monohydrate	Correct
T0259	12	T0262	Lactose monohydrate	Correct
T0260	12	T0262	Lactose monohydrate	Correct
T0261	12	T0262	Lactose monohydrate	Correct
T0277	11	T0255	Allopurinol	Correct
T0278	11	T0255	Allopurinol	Correct
T0329	11	T0337	Caffeine	Correct
T0340	11	Sa019	Chlorthalidone	Correct
T0349	12	T0341	Diltiazem	Correct
T0352	11	T0317	Trans-2-[4-(1,2-Diphenyl-1-butenyl)phenoxy]-N,N-dimethylethylamine	Correct
T0355	12	T0262	Lactose monohydrate	Incorrect* correct is Glyceryl trinitrate
T0376	12	T0262	Lactose monohydrate	Incorrect* correct is Isosorbide-5-monohydrate
T0377	12	T0382	Iminodibenzyl	Correct
T0384	12	T0373	Irgasan	Correct
T0392	10	Sa042	Methyl para-hydroxybenzoate	Correct
T0397	11	Sa044	Nalidixic acid	Correct
T0398	11 11	T0370 T0383	Hydrocortisone acetate Hydrocortisone alcohol	Incorrect Correct

Test Sample Id	Number of Peaks matching library	Library id suggested	Suggested compound	Result
T0406	12	T0262	Lactose Monohydrate	Incorrect* *correct is Nitroglycerin
T0435	9 12 10	T0420 T0451 T0453	Palmitic acid Triheptadecanoin Triacontane	Incorrect Correct Incorrect
T0444	12	Sa054	Thioridazine hydrochloride	Correct
T0448	12	T0408	Reserpine	Correct
T0452	12	T0413	Lorazepam	Correct
T0454	12	T0408	Reserpine	Correct
T0477	9 12 10	T0420 T0451 T0453	Palmitic acid Triheptadecanoin Triacontane	Incorrect Correct Incorrect
T0483	11	T0224	Cyclobarbitone	Correct
T0488	12	Sa020	Carbromal	Correct
T0499	12	T0500	Avicel	Correct
T0502	10	T0060	Quinine	Correct
Ch001	11	T0238	Chlortetracycline	Correct
Ch002	11	T0238	Chlortetracycline	Correct
Ch003	11	T0238	Chlortetracycline	Correct
Ch004	11	T0238	Chlortetracycline	Correct
Ch005	11	T0238	Chlortetracycline	Correct
Sa055	12	Sa004	Potassium hydrogen phthalate	Correct
Sa056	12	Sa004	Potassium hydrogen phthalate	Correct
Sa057	12	Sa004	Potassium hydrogen phthalate	Correct
Sa058	11	Sa004	Potassium hydrogen phthalate	Correct
Sa059	11	Sa004	Potassium hydrogen phthalate	Correct
Sa060	12	Sa004	Potassium hydrogen phthalate	Correct
Sa061	12	Sa004	Potassium hydrogen phthalate	Correct
Sa062	12	Sa004	Potassium hydrogen phthalate	Correct
Sa063	12	Sa004	Potassium hydrogen phthalate	Correct
Sa064	12	Sa004	Potassium hydrogen phthalate	Correct

In summary 56 out of 59 searches produced correct matches with 9 or more peaks.

Table 46 External validation of a library of 301 different pharmaceutical substances based upon a twelve peak match of 59 external samples and batches – Summary of pairs

Peaks were defined as matching if within ± 2nm of each other. Peak order was not considered. Based on second derivative spectra. Database file: pdrugs12, Unknowns file: pkn12

Peaks	Number of Pairs	Percentage of Total	Number Correct	Percent Correct
0	6877	38.7	-	-
1	6820	38.4	-	-
2	2932	16.5	-	-
3	853	4.8	-	-
4	168	0.9	-	-
5	35	0.2	-	-
6	7	0.04	-	-
7	3	0.02	-	-
8	0	0	-	-
9	3	0.02	1	1.7
10	5	0.03	3	5.1
11	24	0.1	23	39.0
12	32	0.2	29	49.2
TOTAL	17759	100	56	95.0

The trends of the twelve peak match were followed by the twenty five peak match. (See summary, Table 48). Of the forty two matches with twenty five, twenty four or twenty three peaks, only two were not straight matches and these can be explained by the presence of lactose (Table 47). With less than twenty three peaks more anomalies start to appear as the actual intensities of some of the peaks are now fairly low. This is similar to the findings of Curry et al[107] who also realised problems when looking to expand out identification from their original six peaks and ended up including a number of peaks of low absorbance.

Similar to the twelve peak matches, from the 59 external test samples, 56 were identified correctly and three samples were misidentified as lactose.

Table 47 External validation of a iibrary of 301 different pharmaceutical substances based upon a twenty five peak match of 59 external samples and batches.

Peaks were defined as matching if within ± 2nm of each other. Peak order was not considered. 12 or more peaks displayed using second derivative spectra. Database file: pdrugs25, Unknowns file: pkn25.

Test Sample Id	Number of Peaks matching library	Library id suggested	Suggested compound	Result
Rw003	22	Sa016	Chlororquine phosphate	Correct
Rw006	25	Rw001	Cefuroxime	Correct
Rw007	24	Rw002	Ampicillin	Correct
Rw009	25	Rw001	Cefuroxime	Correct
Rw010	23	Rw004	Thiopentone	Correct
Rw012	25	Rw005a	Benzyl penicillin	Correct
Sa018	12 21	T0453 T0504	Triacontane Cetyltrimethyl ammonium chloride	Incorrect Correct
T0052	25	T0337	Caffeine	Correct
T0069	24	Rw005	Procaine	Correct
T0085	24	Sa007	Ascorbic acid	Correct
T0251	25	T0255	Allopurinol	Correct
T0252	25	T0255	Allopurinol	Correct
T0253	24	T0255	Allopurinol	Correct
T0254	25	T0255	Allopurinol	Correct
T0256	24	T0262	Lactose monohydrate	Correct
T0257	23	T0262	Lactose monohydrate	Correct
T0258	25	T0262	Lactose monohydrate	Correct
T0259	25	T0262	Lactose monohydrate	Correct
T0260	23	T0262	Lactose monohydrate	Correct
T0261	24	T0262	Lactose monohydrate	Correct
T0277	24	T0255	Allopurinol	Correct
T0278	24	T0255	Allopurinol	Correct
T0329	24	T0337	Caffeine	Correct
T0340	23	Sa019	Chlorthalidone	Correct
T0349	22	T0341	Diltiazem	Correct
T0352	24	T0317	Trans-2-[4-(1,2-Diphenyl-1-butenyl)phenoxy]-N,N-dimethylethylamine	Correct
T0355	23	T0262	Lactose monohydrate	Incorrect* Correct is Glyceryl trinitrate
T0376	21	T0262	Lactose monohydrate	Incorrect* Correct is Isosorbide-5-monohydrate
T0377	25	T0382	Iminodibenzyl	Correct
T0384	25	T0373	Irgasan	Correct
T0392	22	Sa041	Methyl para-hydroxybenzoate	Correct
T0397	21	Sa044	Nalidixic acid	Correct
T0398	22 23	T0370 T0383	Hydrocortisone acetate Hydrocortisone alcohol	Incorrect Correct
T0406	23	T0262	Lactose Monohydrate	Incorrect* Correct is Nitroglycerin
T0435	16 25 20	T0420 T0451 T0453	Palmitic acid Triheptadecanoin Triacontane	Incorrect Correct Incorrect

Test Sample Id	Number of Peaks matching library	Library id suggested	Suggested compound	Result
	15	T0504	Cetyltrimethyl ammonium Chloride	Incorrect
T0444	24	Sa054	Thioridazine hydrochloride	Correct
T0448	25	T0408	Reserpine	Correct
T0452	22	T0413	Lorazepam	Correct
T0454	25	T0408	Reserpine	Correct
T0477	16 25 20 15	T0420 T0451 T0453 T0504	Palmitic acid Triheptadecanoin Triacontane Cetyltrimethyl ammonium Chloride	Incorrect Correct Incorrect Incorrect
T0483	23	T0224	Cyclobarbitone	Correct
T0488	25	Sa020	Carbromal	Correct
T0499	24	T0500	Avicel	Correct
T0502	16	T0060	Quinine	Correct
Ch001	20	T0238	Chlortetracycline	Correct
Ch002	23	T0238	Chlortetracycline	Correct
Ch003	21	T0238	Chlortetracycline	Correct
Ch004	22	T0238	Chlortetracycline	Correct
Ch005	20	T0238	Chlortetracycline	Correct
Sa055	23	Sa004	Potassium hydrogen phthalate	Correct
Sa056	22	Sa004	Potassium hydrogen phthalate	Correct
Sa057	23	Sa004	Potassium hydrogen phthalate	Correct
Sa058	23	Sa004	Potassium hydrogen phthalate	Correct
Sa059	24	Sa004	Potassium hydrogen phthalate	Correct
Sa060	23	Sa004	Potassium hydrogen phthalate	Correct
Sa061	22	Sa004	Potassium hydrogen phthalate	Correct
Sa062	23	Sa004	Potassium hydrogen phthalate	Correct
Sa063	22	Sa004	Potassium hydrogen phthalate	Correct
Sa064	22	Sa004	Potassium hydrogen phthalate	Correct

Table 48 External validation of a library of 301 different pharmaceutical substances based upon a twenty five peak match of 59 external samples and batches – Summary of pairs

Peaks were defined as matching if within ± 2nm of each other. Peak order was not considered. Based upon second derivative spectra. Database file: pdrugs25, Unknowns file: pkn25

Peaks	Number of Pairs	Percentage of Total	Number Correct	Percent Correct
0	792	4.5	-	-
1	2724	15.3	-	-
2	4238	23.9	-	-
3	4251	23.9	-	-
4	2876	16.2	-	-
5	1708	9.6	-	-
6	742	4.2	-	-
7	263	1.5	-	-
8	67	0.4	-	-
9	11	0.06	-	-
10	17	0.10	-	-
11	3	0.02	-	-
12	1	0.006	-	-
13	0	0	-	-
14	0	0	-	-
15	2	0.01	-	-
16	3	0.02	1	1.7
17	0	0	-	-
18	0	0	-	-
19	0	0	-	-
20	4	0.02	2	3.4
21	4	0.02	3	5.1
22	9	0.05	8	13.6
23	15	0.08	13	22.0
24	13	0.07	13	22.0
25	16	0.09	16	27.1
Total	17759	100	56	94.9

6.4.4 Identification of external 'Unknown' drug samples

The 'unknown' drugs were run as test samples by an analyst unfamiliar with the study. The actual identities of the unknown compounds were also not disclosed to the analyst during the course of the experiments thus creating a blind experiment scenario. The comparisons were run for 34 test samples across the library. Prior to running the comparisons, spectra from the test samples were recorded, the spectra were converted into second derivatives, the Savitsky-

Golay 5 5 4 2 smoothing parameter was applied and the samples were run through the PEAK program to identify the six, twelve and twenty five most intense peaks respectively.

The results varied slightly based on the number of peaks used to identify the compounds. For the six peak matches (Table 50) all the unknowns for which there was a corresponding reference in the library were identified successfully with peak matches being at five or six peaks in all cases except two. The two cases, for which the best match was four peaks, were the chlortetracycline hydrochloride, tetracycline hydrochloride 50/50 mix (both constituents identified) and Avicel for which two different grades were used. For all of the above, only three compounds contained cases where other possible matches were suggested. In every case, the highest peaks suggested produced the correct match.

Ideally there would be a discernable gap between the true and false outcomes but in the three examples in Table 49, that was not the case. Careful visual examination of the spectra would be required prior to confirming any matches of four peaks. The results, however, were impressive, the limitations of the size of the database (301 substances) and of the test set (34 samples) are noted but nonetheless, to identify all samples correctly and additionally not match any of the substances not included in the database raises simple peak matching as a powerful method.

Table 49. Table of results for the 34 powdered drug samples run as unknowns against the library of 301 different pharmaceutical substances.

Searches were based upon matching the six most intense negative peaks. Peaks were defined as matching if within ± 2nm of each other. Peak order was not considered. Four or more peaks displayed using second derivative spectra. Database file: pdrugs6, Unknowns file: pun6.

Test Sample Id	Number of Peaks matching library	Library id suggested	Suggested compound	Result
u001	No match	None	N/a	Correct
u002	6	RW005	Procaine hydrochloride	Correct
u003	5	T0236	Tetracycline hydrochloride	Correct
u004	6	SA045	Nitrafurantoin	Correct
u005	6	SA054	Thioridazine hydrochloride	Correct
u006	6	T0262	Lactose monohydrate	Correct
u007	No match	None	N/a	Correct
u008	6 4	SA049 T0461	Penicillamine Sodium valproate	Correct Incorrect*
u009	No match	None	N/a	Correct
u010	No match	None	N/a	Correct
u011	5	SA004	Potassium hydrogen pthalate	Correct
u012	No match	None	N/a	Correct
u013	6	T0066	Papaverine hydrochloride	Correct
u014	6	T0500	Avicel	Correct
u015	No match	None	N/a	Correct
u016	5 4	SA046 T0318	Orphenadrine citrate Atropine	Correct Incorrect*
u017	6	T0479	Phenylmercuric nitrate	Correct
u018	6	T0088	Metronidazole	Correct
u019	4 4	T0236 T0238	Tetracycline hydrochloride Chlortetracycline hydrochloride	Correct** Correct**
u020	6	T0055	Barbitone sodium	Correct
u021	6	T0046	Naproxen	Correct
u022	4 6	T0281 T0476	Benzyl cinnamate 2-phenylbutyric acid	Incorrect* Correct
u023	No match	None	N/a	Correct
u024	5	T0314	Aneurine hydrochloride	Correct
u025	5	T0049	Salicylic acid	Correct
u026	6	SA016	Choroquine phosphate	Correct
u027	No match	None	N/a	Correct
u028	6	T0362	Edetate disodium	Correct
u029	No match	None	N/a	Correct
u030	6	T0313	4-acetamidophenol	Correct
u031	6	T0235	Demeclocycline hydrochloride	Correct
U032	No match	None	N/a	Correct
U033	4	T0500	Avicel	Correct
U034	No match	None	N/a	Correct

NOTE *The highest peak match was correct, this is the important point, **Both are correct as mixture was tetracycline hydrochloride/chlortetracycline hydrochloride 50/50

Table 50 Comparison of 34 powdered drug samples run as unknowns in a blind manner against a library of 301 different pharmaceutical substances.

Searches were based upon matching the six most intense negative peaks. The total number of possible pairs was 10234. Peaks were defined as matching if within ± 2nm of each other. Peak order was not considered. Database file: pdrugs6, Unknowns file: pun6

Peaks	Number of Pairs	Percentage of Total	Number Correct	Percent Correct
0	7776	76.0	-	-
1	2165	21.2	-	-
2	245	2.4	-	-
3	21	0.2	-	-
4	6	0.06	2	5.9
5	5	0.05	5	14.7
6	16	0.2	16	47.1
No match			11	32.4
Total	10234	100	34	100

The matching outcomes with twelve peak matches (Table 52) could be interpreted as being more discernible than the six peak match because test substances u008, u016 and u022 no longer produced alternative peak match suggestions other than the correct result. However, now there was one genuine mis-identification, where there had previously been none. This was in the case of potassium chloride being suggested as sodium chloride and titanium dioxide (eight and nine matching peaks respectively). Both the matching peak numbers were low, however.

Additionally, the mixture of chlortetracycline hydrochloride and tetracycline hydrochloride was only matched as chlortetracycline hydrochloride. It is particularly difficult to identify inorganics because they absorb outside of the near or mid infrared regions, what is detected are trace elements of water and peaks relating to this. In fact the only peaks in inorganics are going to be due to the presence of water so identifications would not be expected to be possible.

When trying to relate the results to setting criteria for identification, it would seem reasonable to set a limit of acceptance based upon these and the previous results of ten matching peaks, but it is proving increasingly difficult to set fixed criteria for peak matching by NIR. This was also witnessed in the TLC chapter (Chapter 3) where it was also concluded that it would not be always possible to set rigid criteria for identification (this based upon thin layer chromatography). It was also suggested that visual inspection of the TLC plates would also be required. Much the same is evident here where common spectroscopic problems of closely related compounds are difficult to differentiate between.

Table 51 Table of results for the 34 powdered drug samples run as unknowns against the library of 301 different pharmaceutical substances. Searches were based upon matching the twelve most intense negative peaks.

Peaks were defined as matching if within ± 2nm of each other. Peak order was not considered. Eight or more peaks displayed using second derivative spectra. Database file: pdrugs12, Unknowns file: pun12.

Test Sample Id	Number of Peaks matching library	Library id suggested	Suggested compound	Result
u001	No match	none	N/a	Correct
u002	11	RW005	Procaine hydrochloride	Correct
u003	10	T0236	Tetracycline hydrochloride	Correct
u004	12	SA045	Nitrafurantoin	Correct
u005	12	SA054	Thioridazine hydrochloride	Correct
u006	12	T0262	Lactose monohydrate	Correct
u007	No match	none	N/a	Correct
u008	12	SA049	Penicillamine	Correct
u009	No match	none	N/a	Correct
u010	No match	none	N/a	Correct
u011	12	SA004	Potassium hydrogen pthalate	Correct
u012	No match	none	N/a	Correct
u013	11	T0066	Papaverine hydrochloride	Correct
u014	12	T0500	Avicel	Correct
u015	No match	none	N/a	Correct
u016	12	SA046	Orphenadrine citrate	Correct
u017	12	T0479	Phenylmercuric nitrate	Correct
u018	12	T0088	Metronidazole	Correct
u019	10	T0238	Chlortetracycline hydrochloride	Correct
u020	12	T0055	Barbitone sodium	Correct

Test Sample Id	Number of Peaks matching library	Library id suggested	Suggested compound	Result
u021	11	T0046	Naproxen	Correct
u022	11	T0476	2-phenylbutyric acid	Correct
u023	No match	none	N/a	Correct
u024	12	T0314	Aneurine hydrochloride	Correct
u025	12	T0049	Salicylic acid	Correct
u026	10	SA016	Choroquine phosphate	Correct
u027	No match	none	N/a	Correct
u028	10	T0362	Edetate disodium	Correct
u029	No match	none	N/a	Correct
u030	12	T0313	4-acetamidophenol	Correct
u031	11	T0235	Demeclocycline hydrochloride	Correct
u032	No match	none	N/a	Correct
u033	10	T0500	Avicel	Correct
u034	8 9	T0459 T0501	Sodium chloride Titanium dioxide	**Incorrect** **Incorrect**

Table 52 Comparison of 34 powdered drug samples run as unknowns In a blind manner against a library of 301 different pharmaceutical substances. Searches were based upon matching the twelve most intense negative peaks.

Peaks were defined as matching if within ± 2nm of each other. Peak order was not considered. Database file: pdrugs12, Unknowns file: pun12

Peaks	Number of Pairs	Percentage of Total	Number Correct	Percent Correct
0	4116	40.2	-	-
1	3830	37.4	-	-
2	1663	16.2	-	-
3	479	4.7	-	-
4	98	1.0	-	-
5	18	0.18	-	-
6	4	0.04	-	-
7	1	0.01	-	-
8	1	0.01	0	0
9	1	0.01	0	0
10	5	0.05	5	14.7
11	5	0.05	5	14.7
12	13	0.13	13	38.2
No Match	-	-	10	29.4
Incorrect	-	-	1	2.9
Total	10234	100	34	100

The move to the even higher number of peaks led to greater selectivity with the matches. With the identification by twenty five peaks, all compounds contained in the library were successfully identified (including the chlortetracycline hydrochloride, tetracycline hydrochloride mix.) Again, predictably, the inorganics caused a problem, but the number of peaks matched in the mismatch was comparatively low (17 peaks). Based upon these and the previous results an automatic acceptance limit of 22 peaks would be fairly comfortable in preventing mismatches while not including the vast majority of correct matches. In this manner the 'mismatch' of sample u001a would not be deemed an acceptable match anyway. By this criteria, u019 would have been rejected, but would have been investigated; it being preferable to have rejected this sample first time, rather than let it pass and open up the possibility of passing non-correct samples.

Table 53 Table of results for the 34 powdered drug samples run as unknowns against the library of 301 different pharmaceutical substances. Searches were based upon matching the twenty five most intense negative peaks.

The total number of possible pairs was 10234. Peaks were defined as matching if within ± 2nm of each other. Peak order was not considered. 12 or more peaks displayed using second derivative spectra. Database file: pdrugs25, Unknowns file: pun25

Test Sample Id	Number of Peaks matching library	Library id suggested	Suggested compound	Result
u001	13	T0252A	4-Epianhydrotetracycline	**Incorrect**
u002	24	RW005	Procaine hydrochloride	Correct
u003	22	T0236	Tetracycline hydrochloride	Correct
u004	25	SA045	Nitrafurantoin	Correct
u005	24	SA054	Thioridazine hydrochloride	Correct
u006	25	T0262	Lactose monohydrate	Correct
u007	No match	none	N/a	Correct
u008	23	SA049	Penicillamine	Correct
u009	No match	none	N/a	Correct
u010	No match	none	N/a	Correct
u011	23	SA004	Potassium hydrogen pthalate	Correct
u012	No match	none	N/a	Correct
u013	22	T0066	Papaverine hydrochloride	Correct

Test Sample Id	Number of Peaks matching library	Library id suggested	Suggested compound	Result
u014	24	T0500	Avicel	Correct
u015	No match	none	N/a	Correct
u016	24	SA046	Orphenadrine citrate	Correct
u017	25	T0479	Phenylmercuric nitrate	Correct
u018	22	T0088	Metronidazole	Correct
u019	18 13	T0238 T0236	Chlortetracycline hydrochloride Tetracycline hydrochloride	Correct Correct
u020	24	T0055	Barbitone sodium	Correct
u021	22	T0046	Naproxen	Correct
u022	25	T0476	2-Phenylbutyric acid	Correct
u023	No match	none	N/a	Correct
u024	25	T0314	Aneurine hydrochloride	Correct
u025	23	T0049	Salicylic acid	Correct
u026	24	SA016	Choroquine phosphate	Correct
u027	No match	none	N/a	Correct
u028	25	T0362	Edetate disodium	Correct
u029	No match	none	N/a	Correct
u030	24	T0313	4-Acetamidophenol	Correct
u031	23	T0235	Demeclocycline hydrochloride	Correct
u032	No match	none	N/a	Correct
u033	20	T0500	Avicel	Correct
u034	17 16 14	T0459 T0501 T0419	Sodium chloride Titanium dioxide Potassium molybdate	**Incorrect** **Incorrect** **Incorrect**

Table 54 Comparison of 34 powdered drug samples run as unknowns in a blind manner against a library of 301 different pharmaceutical substances. Searches were based upon matching the twenty five most intense negative peaks

The total number of possible pairs was 10234. Peaks were defined as matching if within ± 2nm of each other. Peak order was not considered. Database file: pdrugs25, Unknowns file: pun25

Peaks	Number of Pairs	Percentage of Total	Number Correct	Percent Correct
0	478	4.7	-	-
1	1598	15.6	-	-
2	2470	24.1	-	-
3	2396	23.4	-	-
4	1663	16.2	-	-
5	962	9.4	-	-
6	408	4.0	-	-
7	149	1.5	-	-
8	58	0.57	-	-
9	16	0.16	-	-
10	6	0.06	-	-
11	2	0.02	-	-
12	0	0	-	-

Peaks	Number of Pairs	Percentage of Total	Number Correct	Percent Correct
13	2	0.02	1	2.9
14	1	0.01	0	0
15	0	0	-	-
16	1	0.01	0	0
17	1	0.01	0	0
18	1	0.01	1	2.9
19	0	0	-	-
20	1	0.01	1	2.9
21	0	0	-	-
22	4	0.04	4	11.8
23	4	0.04	4	11.8
24	8	0.08	8	23.5
25	5	0.05	5	14.7
No match	-	-	9	26.5
Incorrect	-	-	2	5.9
Total	10234	100	35*	102.9*

Note *35 samples identified as there were two possible correct matches for chlortetracycline hydrochloride/tetracycline hydrochloride

Discrimination between compounds was improved by increasing the number of peaks used for identification, although there was not much difference between the twelve and twenty five peak matches, all three matching levels provided very good detection. Important parameters to consider are also the spectral range (as discussed in section 1.10), degree of spectral smoothing and the peak matching window. In addition, validation must be carried out on internal and external data.

Introducing a scoring system whereby the relative peak intensities and order are considered may improve the matches and is a consideration to investigate for the future. As far as Curry et al[107] used this system with mid infrared, they found the system restrictive and led to less successful results than the six peaks method. Relying on peak intensities does, however, lessen the robustness of the method to changes in physical characteristics (these have a

direct bearing on peak intensity as discussed in Chapter 4 "Sample Presentation").

Increasing the number of peaks increases the ability to distinguish between compounds but it is important to verify that the peaks are genuine and not due to noise. Setting a signal to noise threshold or a minimum absorbance setting on the spectrometer would help to eliminate this. With modern day instrumentation and when using the powdered samples in particular, this is rarely a problem as there are frequently well in excess of 25 well defined usable peaks, however spectra produced from liquids often contain relative few well defined negative peaks so this can be a problem. Visual inspection is also necessary to ensure that there are also no missing or unaccounted for peaks.

6.5 Conclusions

Identification based upon matching positions of the most intense negative second derivative peaks provides an extremely simple and effective means of identification. Criteria for the identification of substances are based upon the following approach :

Firstly, use second derivative spectra. Secondly, standardise the smoothing parameters and number of peaks to be used. Finally, use a hierarchy of methods.

The importance of second derivatives has already been stated in Chapter 5. Varying the Savitsky Golay window size can affect the peak position of second derivative spectra. As the peak position is the basis of identification, a standard

window size must be used. I used 5 5 4 2 (five data points either side of the data point of interest, a fourth order polynomial fit and the second derivative). I would strongly suggest the fourth order polynomial and second derivative. This can be used in conjunction with four to ten data points - whichever is chosen must be kept constant throughout. As also noted in the conclusions of Chapter 5, is that the second derivative method itself and all smoothing parameters must be kept constant.

The method hierarchy that I would recommend is

- Use of the peak identification method – the 25 peak identifications provided the best overall matching when compared with the 12 and 6 peak methods. However, there was very little to choose between the methods. A 12 or 6 peak match could be used in the case where peak intensities are low or where a substance does not possess 25 peaks
- Critical examination of spectra – in cases where there is more than one substance suggested as a match, when greater than or equal to the acceptance peak number. To be used when the highest matching number of peaks is below the acceptance peak number.

The criteria for identification for the 25 peaks should be based upon matching 22 or more peaks out of the 25 peaks. This is independent of peak order and is based on the experience of the 301 substances. Many more batches of samples and substances would be needed to determine the robustness of this value. Setting of rigid criteria based upon matching peaks could lead to inflexibility in this technique and may increase the chance of mis-matches. The importance of visually inspecting the spectra can not be overstated.

Chapter 7 The identification of actives in whole tablets

7.1 Summary of aims

- To demonstrate identification of the active component in an intact tablet by NIR
- To optimise the method for identification by an algorithmic peak matching technique
- To demonstrate differentiation between batches of tablets from different manufacturers
- To identify the manufacturer of an "unknown" batch of tablets to stated levels of confidence

7.2 Method

The methods are described in detail in the Experimental chapter in section 2.5 Details of the tablets, manufacturers and batch references can be found in Appendix 4, Table 65.

7.3 Results and Discussion

Chapter 5 and Chapter 6 described the setting up of a database of over 300 compounds of pharmaceutical interest and demonstrated that the compounds could be reliably distinguished from one another using certain mathematical techniques and under certain conditions.

The same principals have been employed in this chapter with the focus on the identification of the active drug in solid dosage forms.

Tablets, comprising a complex blend of chemical and physical properties unsurprisingly produce NIR spectra that are complex in nature and difficult to interpret. Additionally, the physical shape of tablets decrease the efficient use of light. This is because tablets are often curved in shape which increases both reflection and the difference between incident and detected light. The effect of sample presentation has been covered in Chapter 4 "Sample presentation for NIR spectroscopy."

7.3.1 Tablet Spectra

Spectra collected from 20 intact allopurinol 100mg tablets BP, placed directly onto the quartz window of the Rapid Content Analyser module are displayed in Figure 57. The tablets were taken from the same batch and the variation between them is relatively small. Second derivative plots were calculated from the spectra using the NSAS IQ^2 software utilising a gap size of 0 and a segment size of 10 nm. These plots are displayed in Figure 58. Also displayed in Figure 57 and Figure 58 are spectra collected from the allopurinol pure drug.

Different approaches can be taken to the examination of tablet spectra. One approach would be to attempt to match each peak and trough to a distinct chemical group or relate to a particular absorption band or bond resonance. A

similar approach is taken to the interpretation of mid infrared data. However, two key elements in the NIR argue against such an approach in this case. Firstly, absorption (related to log 1/*R*) is many fold less in the NIR region and as discussed earlier, peaks are therefore often not well defined. The second point is that tablets contain many excipients in addition to the main drug active(s) and these may well share bands of absorption leading to co-absorption. A more rational approach is to attempt to identify areas of unique absorption. The task is to identify the active in a tablet from a database of drug actives, the unique element is therefore the drug active itself.

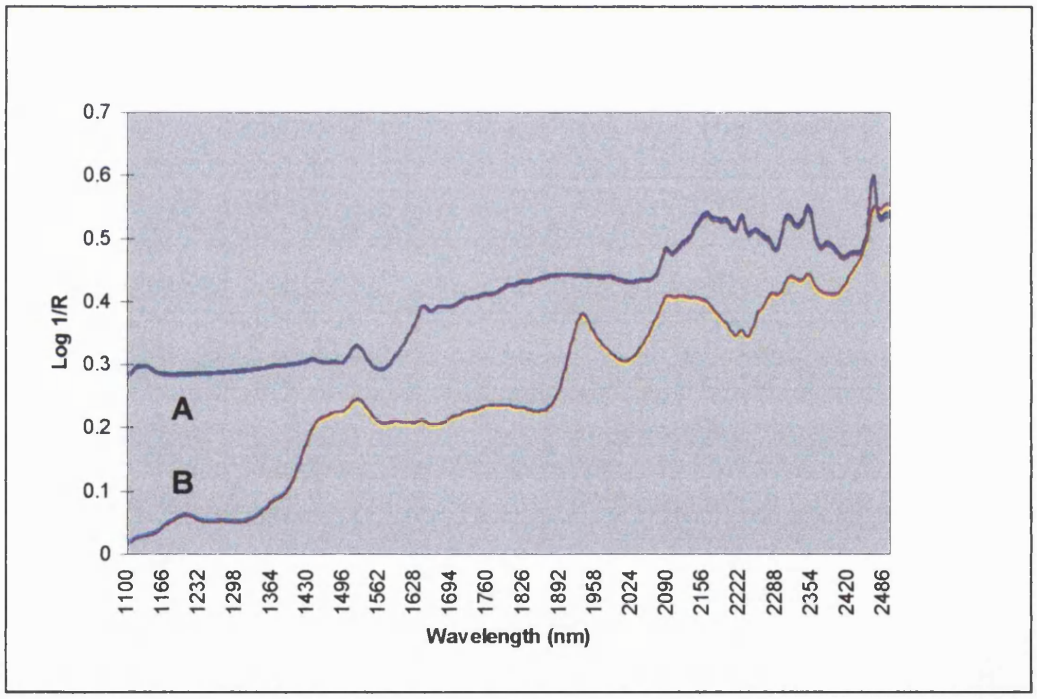


Figure 57 Spectra from 20 intact allopurinol 100 mg tablets BP and allopurinol BP pure drug.

Spectra recorded between 1100 nm and 2498 nm. A = allopurinol BP pure drug, B = allopurinol 100 mg tablets BP

7.3.2 Identification of the tablet active

Visual comparison between the two sets of spectra show common areas between 1500 nm to 1570 nm, 2060 nm to 2100 nm and 2150 nm to 2400 nm. Based upon these wavelengths and experience of the best matching wavelengths from Chapter 5, correlation spectral matching was employed to match the active ingredients in tablet samples to those contained in the drugs substances database, constructed and validated as part of Chapter 5. Table 55 shows the results of analysing 10 tablets from ten batches of allopurinol 100mg tablets (manufactured by Wellcome), aspirin 300mg caplets (manufactured by Tesco) and cimetidine 200 mg tablets (manufactured by Clonmel). Correlation spectral match was used across 1100 nm to 2500 nm and identified matches for the actives allopurinol and aspirin to the respective tablets. The drug active for cimetidine was not included in the database and was used to ensure no matches were made when an active was not included in the database.
Both the positive results for allopurinol and aspirin were ranked as the highest matches in each case with a significant gap to the next highest matches which were orciprenaline 0.551 and diflunisal 0.623 respectively.

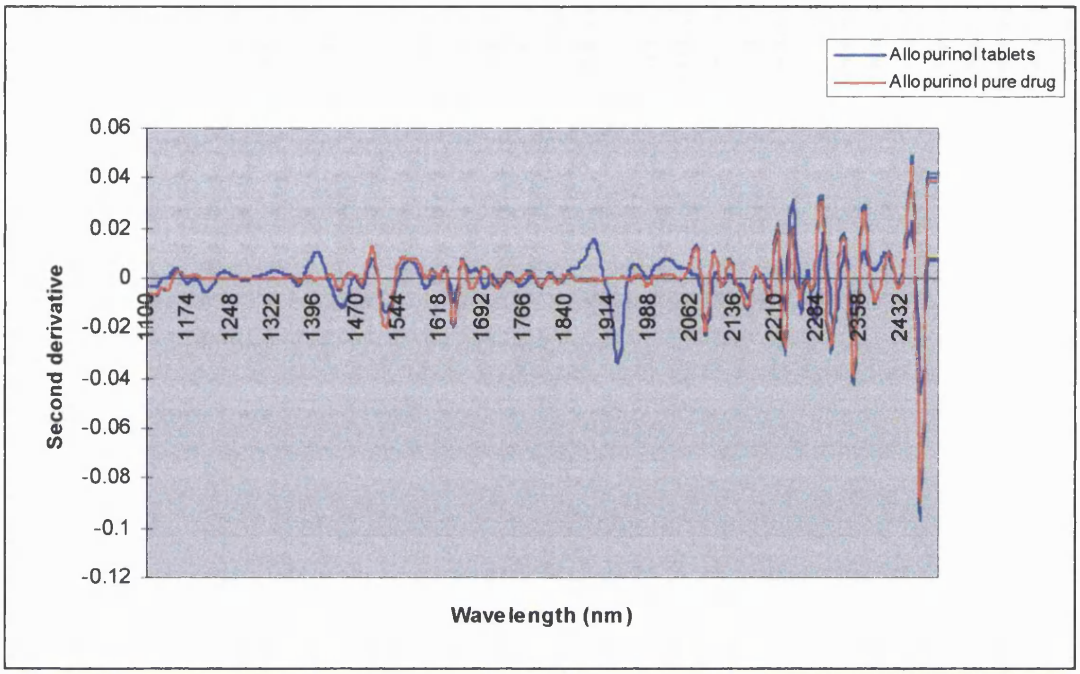


Figure 58 Second derivative spectra from 20 intact allopurinol 100mg tablets BP and allopurinol BP pure drug.

Spectra recorded between 1100 nm and 2500 nm

Table 55 The Identification of the Raw Active in Tablets by correlation spectral matching.

Range is 1100 - 2500 nm and Match values displayed are the mean of ten batches.

Tablets	Correlation spectral match (mean)	Standard deviation
allopurinol 100 mg	0.754 allopurinol	0.076
	0.551 orciprenaline	0.041
	0.425 Amylobarbitone	0.019
aspirin 300 mg	0.923 Acetylsalicylic acid	0.018
	0.623 Diflunisal	0.027
	0.483 Nicotinic Acid	0.021
cimetidine 200 mg	no matches found*	n/a
	no matches found*	n/a
	no matches found*	n/a

*The cut-off point to indicate "no matches found" was for matches < 0.1

Similar results are obtained when applying the same comparison criteria between 2000 nm and 2400 nm. The range was selected as an effective identification range from the results of raw active identification from Chapter 5

and from the visual comparisons in Figure 57 and Figure 58. The results of the database search are contained within Table 56.

The overall trend is the same as with the 1100 nm to 2500 nm comparison, however the allopurinol (0.784) and aspirin (0.948) matches were enhanced. Cimetidine is again not matched.

This method was repeated once again narrowing the range to 2200 nm - 2400 nm. Also included in the test were tablets from batches of allopurinol 300 mg tablets (Table 56). The reason was two fold. Firstly to use a strength of tablet similar to that of the aspirin in terms of the ratio of active to overall tablet weight (approximately 30 % in both cases) and secondly to determine if increasing the strength of the active within the tablets would increase the match value.

The results in Table 57 indicate that the match for allopurinol in the 300 mg tablets is greater than that of the 100 mg tablets and now comparable to that obtained by the aspirin tablets.

Table 56 The identification of the active in tablets by correlation spectral matching.

Range is 2000 - 2400 nm and match values displayed are the mean of ten batches.

Tablets	Correlation spectral match (mean)	Standard deviation
Allopurinol 100 mg	0.784 Allopurinol	0.069
	0.462 Amylobarbitone	0.026
	0.390 Chlorpromazine	0.040
Aspirin 300 mg	0.948 Aspirin	0.026
	0.601 Orciprenaline	0.041
	0.513 Nicotinic acid	0.038
Cimetidine 200 mg	no matches found*	n/a
	no matches found*	n/a
	no matches found*	n/a

*The cut-off point to indicate "no matches found" was for matches < 0.1

As the range of database comparison has been narrowed from 1100 nm - 2500 nm down to 2200 nm to 2400 nm, the identification of both actives contained within the database has increased. Aspirin 300 mg tablets produced higher match values than the allopurinol 100 mg tablets and both were selected as the highest matches in each of the respective database searches. There were no misidentifications to report. Comparison of the second derivative spectra from both strengths of tablet compared to the active (Figure 60) suggests that the increased concentration of the active is the reason for the increased match value.

An explanation for the increasing match value as the wavelength range is narrowed is the elimination of variation (for example overlapping absorption bands from different components such as excipients and physical characteristics such as particle size and light scattering) within and between the tablets. Comparison with Figure 58 and Figure 59 shows that it is possible to obtain areas of absorption unique to the drug active within the tablet despite the complex tablet matrix and adds to the probable answer as to why this range performed the best.

This is not entirely consistent with the results for drug substance identification where the optimum range was determined to be 2000 nm to 2400 nm. Drug substance identification based upon databases searches with the narrower range were also good but match values were not as high as with the 2000 - 2400 nm range (Table 57). The difference, however, with tablets is the variation factor mentioned earlier. Also the coatings and polish applied to tablets have a bearing. The allopurinol and aspirin tablets under examination are white, non coated and non shiny tablets. Aspirin spectra are displayed in Figure 61. The

Cimetidine tablets are brown in colour but also non shiny. Cimetidine second derivative spectra are displayed in Figure 62. The difference in colour would be apparent towards the ultra-violet end of the NIR wavelength range but a limitation on the detectors the NIRS 6500 instrument that was used for this work meant that it was not possible to analyse the tablets down to the 400 nm that is possible by some NIR spectrophotometers. This method is expected to be limited when considering enterically coated tablets due to the nature of the coatings themselves. Work specifically on tablet coatings has been conducted by Buchanen et al[50] and Kirsch and Drennen[109].

Database searches on wavelength segments less than 200 nm have proved to be less effective. The results for 2300 nm - 2400 nm are shown in Table 58. Ranges less than 100 nm were also attempted but were less effective still and are not reported here. The limit has probably been reached between elimination of variation and having enough datapoints upon which to base a sensible identification. However the question is once again raised whether from a regulatory perspective it is acceptable to analyse tablets between 1100 nm and 2500 nm (a generally peer accepted NIR range) but to only conduct database searches on a subset of this data, e.g. 2200 nm to 2400 nm. The counter argument is that between 1100 nm to 2500 nm there are three overtones of absorption and identification is based on one specific overtone. The rationale for this selective subset match area is identification of the unique wavelength range that characterises the absorption profile of a particular drug active. Cimetidine was not matched in any of the searches and was used as a control to determine if there were any mis-identifications, which there were not.

Table 57 The identification of the active in tablets by correlation spectral matching.

Range is 2200 - 2400 nm and match values displayed are the mean of ten batches.

Tablets	Correlation spectral match (mean)	Standard deviation
Allopurinol 100 mg	0.786 Allopurinol	0.063
	0.469 Amylobarbitone	0.031
	0.387 Chlorpromazine	0.043
Aspirin 300 mg	0.953 Aspirin	0.057
	0.525 Orciprenaline	0.051
	0.483 Nicotinic acid	0.036
Cimetidine 200 mg	No matches found*	n/a
	No matches found*	n/a
	No matches found*	n/a
Allopurinol 300 mg	0.955 Allopurinol	0.032
	0.547 Chlorpromazine	0.046
	0.517 Isoniazid	0.036

*The cut-off point to indicate "no matches found" was for matches < 0.1

Table 58 The identification of the active in tablets by correlation spectral matching.

Range is 2300 - 2400 nm and match values displayed are the mean of ten batches.

Tablets	Correlation spectral match (mean)	Standard deviation
Allopurinol 100 mg	0.758 Allopurinol	0.060
	0.568 Amylobarbitone	0.054
	0.477 Chlorpromazine	0.049
Aspirin 300 mg	0.931 Acetylsalicylic acid	0.026
	0.613 Orciprenaline	0.037
	0.441 Nicotinic acid	0.035
Cimetidine 200 mg	no matches found*	n/a
	no matches found*	n/a
	no matches found*	n/a
Allopurinol 300 mg	0.932 Allopurinol	0.056
	0.522 Chlorpromazine	0.023
	0.510 Isoniazid	0.064

*The cut-off point to indicate "no matches found" was for matches < 0.1

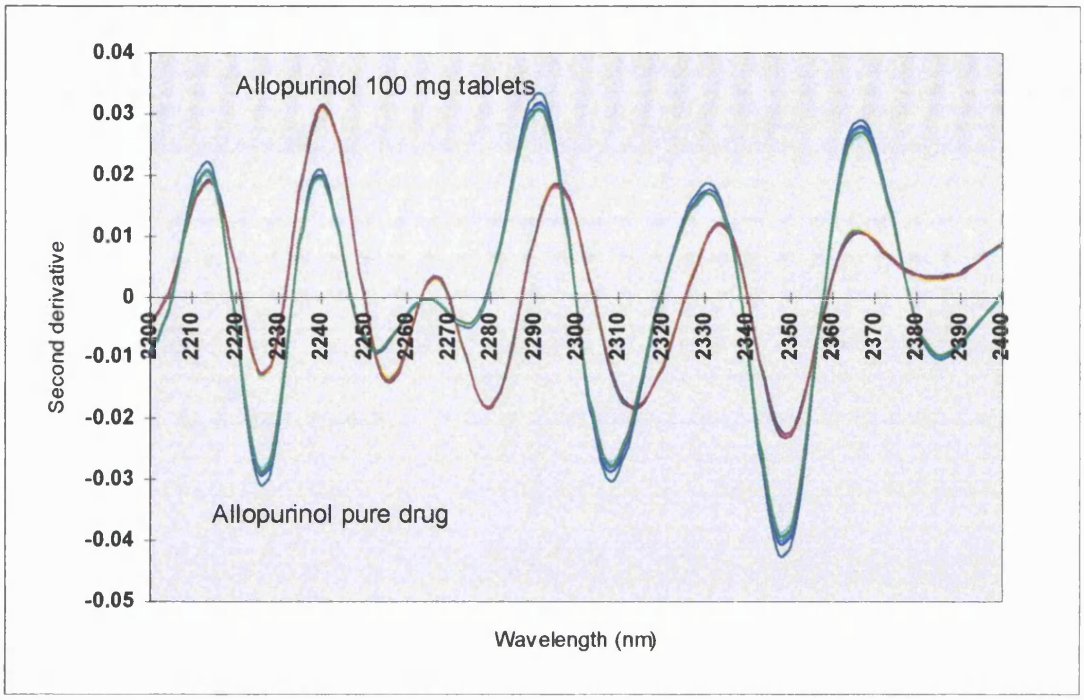


Figure 59 Second derivative spectra of allopurinol 100mg tablets BP and allopurinol BP pure drug between 2200 nm and 2400 nm

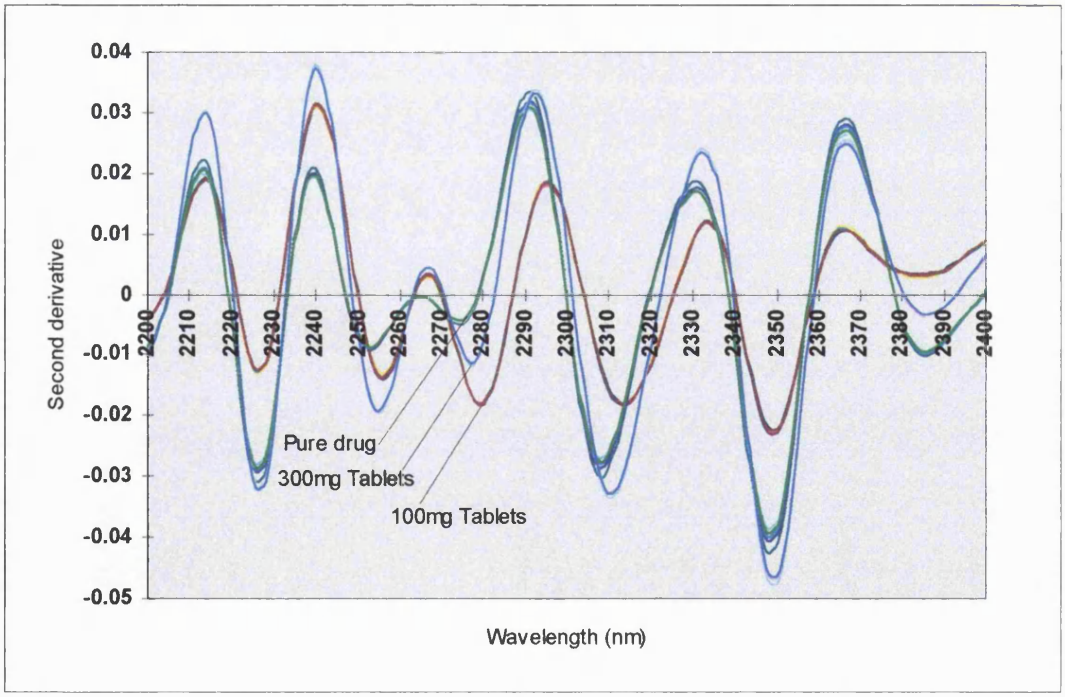


Figure 60 Second derivative spectra of allopurinol 100mg tablets BP, allopurinol 300 mg tablets BP and allopurinol BP pure drug between 2200 nm and 2400 nm

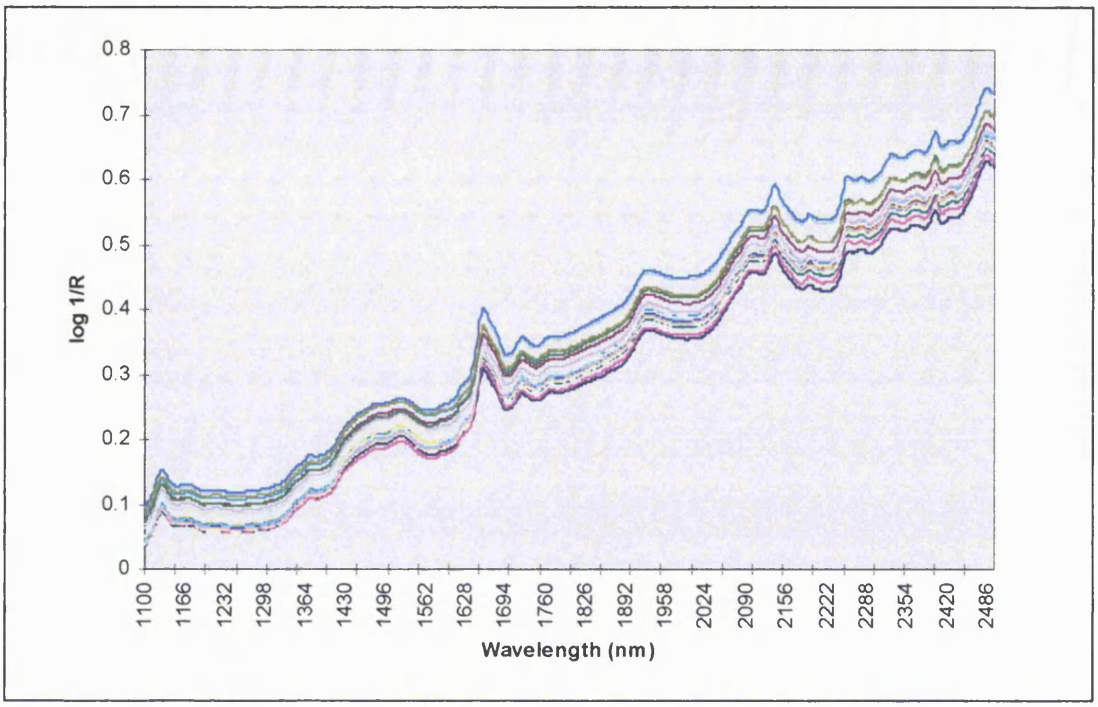


Figure 61 Original spectra from 20 aspirin 300 mg tablets BP in the range 1100 nm - 2500 nm

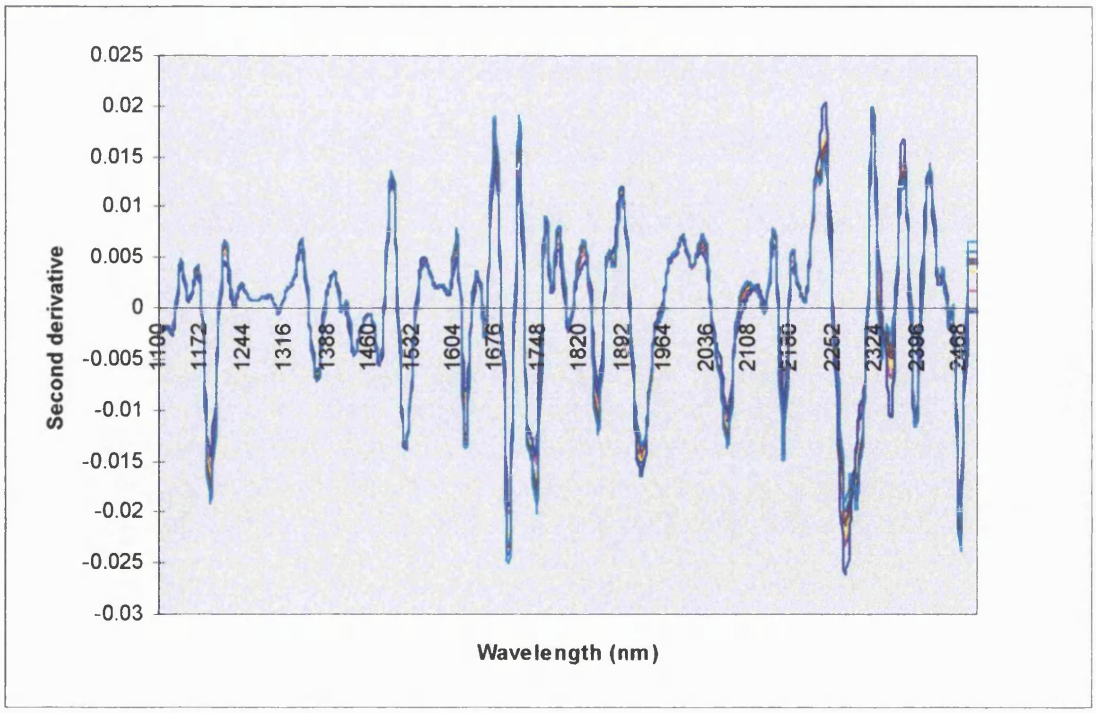


Figure 62 Second derivative spectra from 20 cimetidine 200 mg tablets in the range 1100 - 2500 nm

Wavelength distances were also examined with respect to identification of the actives described above but were found to be extremely poor in gaining any matches. Results from attempting to use wavelength distances to identify the active in the allopurinol tablets between 2000 nm and 2400 nm are shown in Table 59. It appears that the physical variation introduced by the tablets renders the distance match ineffective: no matches were correct and misidentifications occurred in each case. The wavelength distance matching algorithm when applied in Chapter 5, demonstrated sensitivity to variation in intensity, commonly related to physical variation so it is not a surprise for this method to be deemed to be inappropriate here. When considering tablets, there are numerous complexities to be dealt with before considering detecting the drug active.

Table 59 The identification of the raw active in tablets by wavelength distance matching

Range is 2000 - 2400 nm and match values displayed are the mean of five batches. Based upon second derivatives.

Tablet description	Wavelength distance match
APS allopurinol 100mg	47.4 Procaine
	no others matched*
Ashbourne allopurinol 100mg	61.2 Procaine
	no others matched*
Norton allopurinol 100mg	59.7 Procaine
	no others matched*
Cox allopurinol 100mg	47.3 Procaine
	no others matched*
Aspirin Caplet	37.7 Procaine
	159.3 Ascorbic Acid
	no others matched*

*The cut-off point to indicate "no matches found" was for matches >1000

These include excipients with overlapping bands of interest to the active, complex tablet matrices, coatings and variations in manufacturing process from one tablet brand to another.

It has been demonstrated that it is possible to classify tablets based upon their active ingredient but is it possible to take this further? The tablets that have been studied were derived from a number of different batches and were produced by a variety of manufacturers. This has resulted in standard deviations of the correlation spectral match values as high as 0.076, but ensured that the studies carried out took in representative sample of products in the market place today. An example of the results, by manufacturer for allopurinol, that went into the study of the tablet active identification can be found in Table 60. The variation in the identification of the highest match in each case illustrates the variation across the tablets themselves, however the standard deviations displayed demonstrate that variation within the tablets produced by a single manufacturer is very small. Spectra produced by tablets from different batches but from the same manufacturer are displayed in Figure 63. The shape of the spectra remain constant, as do the peak wavelength values, the variable is the peak height. As described in Chapter 4, this effect is related to variation in physical attributes rather than chemical ones. As discussed in section 5.3, displaying the spectra as second derivative plots would eliminate much of this physical effect and this can be seen in Figure 64. Correlation spectral match values of tablets from different batches from the same manufacturer are displayed in Table 61. The breakdown at the bottom of the table displays the variation in match values for the allopurinol tablets. The standard deviation of values is of the order of approximately 0.1, which can be compared against the similar breakdown of results in Table 60. There is a noticeable increase in the range when comparing batches from different manufacturers (Table 60). This is allied to an increase in the standard

deviation. This suggests that differences in the tablets produced by different manufacturers may being detected here, and that the differences are greater than those produced by the inter-batch variation from a single manufacturer. Spectra from tablets from nine different manufacturers are presented in Figure 65.

Table 60 The identification of the raw active in tablets by correlation spectral matching.

Database search range is 1100 - 2500 nm and match values displayed are the mean of 20 tablets from five different batches. Matches based on second derivatives.

Tablet Description	Correlation spectral match
APS allopurinol 100mg	0.796 Allopurinol
	0.479 Orciprenaline
	0.399 Amylobarbitone
Ashbourne allopurinol 100mg	0.801 Allopurinol
	0.554 Orciprenaline
	0.422 Amylobarbitone
Cox allopurinol 100mg	0.643 Allopurinol
	0.605 Orciprenaline
	0.402 Amylobarbitone
CP Allopuroninol 100mg	0.697 Allopurinol
	0.559 Orciprenaline
	0.435 Amylobarbitone
DDSA allopurinol 100mg[1]	0.759 Allopurinol
	0.467 Amylobarbitone
	0.456 Phenytoin
Norton allopurinol 100mg	0.805 Allopurinol
	0.558 Orciprenaline
	0.424 Amylobarbitone
Pharmvit allopurinol 100mg	0.702 Allopurinol
	0.568 Orciprenaline
	0.419 Amylobarbitone
Rima allopurinol 100mg	0.695 Allopurinol
	0.574 Orciprenaline
	0.429 Amylobarbitone
Zyloric (Wellcome) allopurinol 100mg	0.788 Allopurinol
	0.559 Orciprenaline
	0.429 Amylobarbitone
Generics allopurinol 100mg	0.681 Allopurinol
	0.572 Orciprenaline
	0.421 Amylobarbitone
Regent allopurinol 300mg	0.924 Allopurinol
	0.471 Phenytoin
	0.466 orciprenaline

[1] Results based on samples from four batches

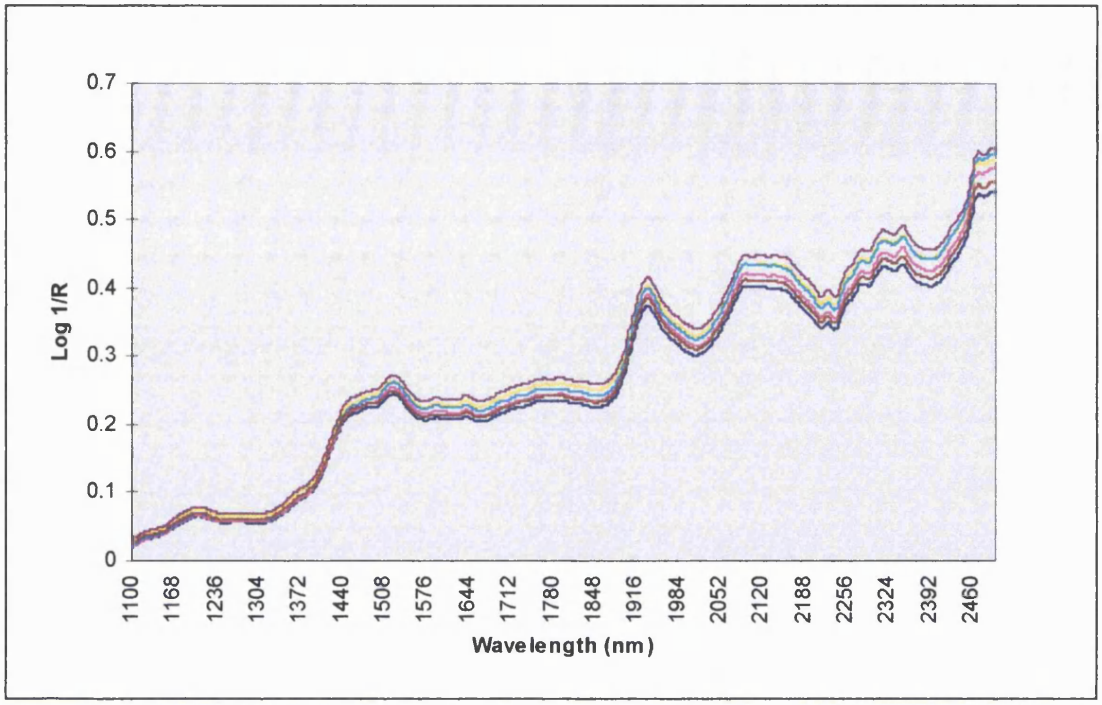


Figure 63 Original spectra from different batches of allopurinol 100 mg tablets BP from one manufacturer (Generics)

The spectral range is 1100 – 2500 nm

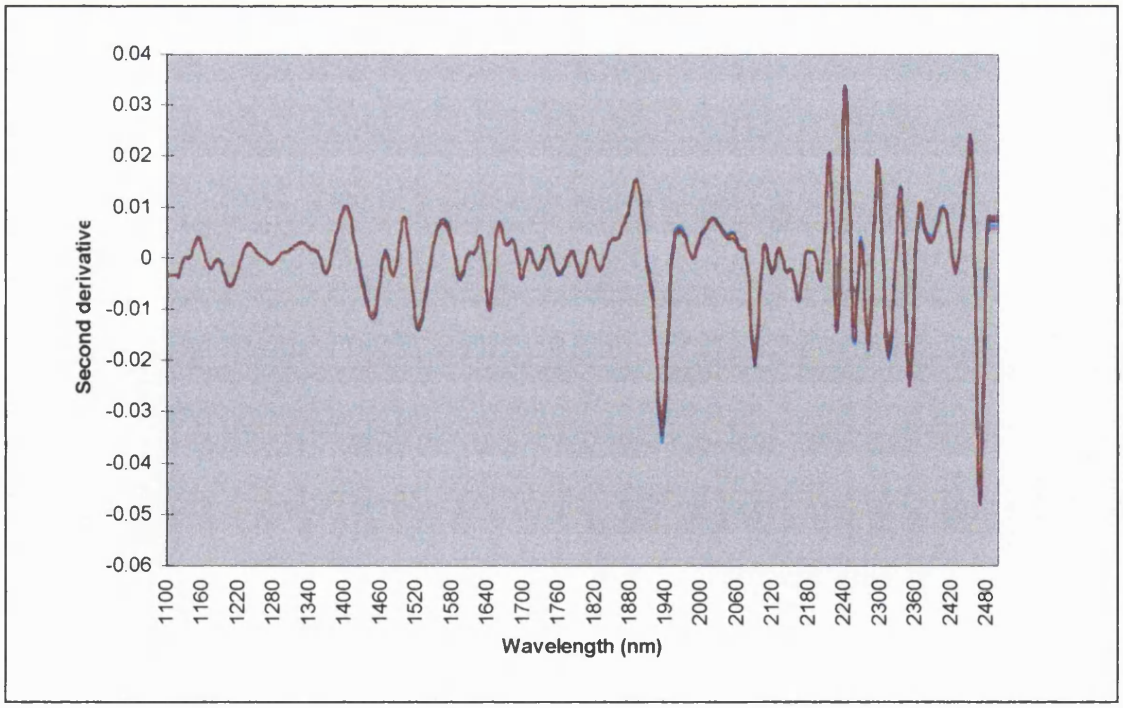


Figure 64 Second derivative spectra from different batches of allopurinol 100 mg tablets BP from one manufacturer (Generics)

Table 61 The Identification of the raw active in tablets by correlation spectral matching.

Comparison of different batches from one manufacturer (Generics) of allopurinol 100 mg tablets Range is 2000 - 2400 nm and match values displayed are the mean of 20 tablets. Matches based upon second derivates.

Batch ID	Correlation spectral match
ESRW0H4	0.700 Allopurinol
	0.482 Amylobarbitone
	0.331 Quinidine
5293M2	0.690 Allopurinol
	0.480 Amylobarbitone
	0.324 Doxepin
EWEAL3	0.709 Allopurinol
	0.485 Amylobarbitone
	0.334 Quinidine
EDOBIE1	0.720 Allopurinol
	0.486 Amylobarbitone
	0.336 Chlorpromazine
EYWLEV1	0.706 Allopurinol
	0.486 Amylobarbitone
	0.328 Chlorpromazine
2412K2	0.744 Allopurinol
	0.480 Amylobarbitone
	0.345 Chlorpropamide
EWBRL7	0.690 Allopurinol
	0.437 Amylobarbitone
	0.305 Quinidine
3M60	0.634 Allopurinol
	0.470 Cimetdiine
	0.411 3' OH Butobarbitone
P31004	0.711 Allopurinol
	0.486 Cimetidine
	0.451 3' OH Butobarbitone
P31469	0.710 Allopurinol
	0.483 Cimetidine
	0.437 3' OH Butobarbitone
Max value for allopurinol	0.744
Min value for allopurinol	0.634
Mean Match value	0.701
Range	0.110
Standard deviation	0.027

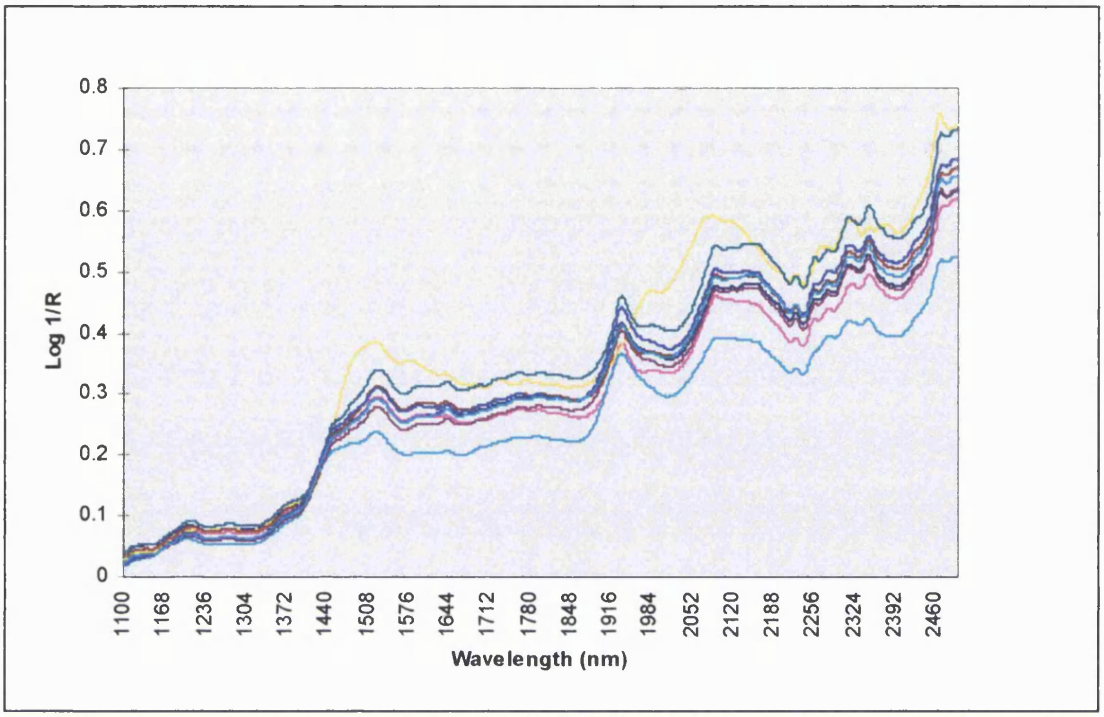


Figure 65 Comparison of spectra from nine different manufacturers of allopurinol 100 mg tablets.

Range is 1100 nm - 2500 nm. Each spectrum represents the mean of 20 spectra which were taken from 20 tablets. The tablets were derived from 10 different batches

The variations in the spectra from different manufacturers (Figure 65) are not restricted to only physical differences but also include peak wavelength shifts in some cases. These differences would be expected from excipients such as diluents, binders etc that make up the constituent of a tablet. In the preceding chapters, limitations with using un-derivatised spectra for identification of drug actives were noted. However, for the purposes of examining differences in tablets, the un-derivatised spectra were found to be ideal.

Based upon this, an additional method (polar qualification) was used to describe each spectrum as a single quality point or centre of gravity.

7.3.3 Identification of the tablet manufacturer

By using correlation spectral matching it was possible to detect the active within a tablet. It was also noted visually that there were differences in spectra relating to tablets from different manufacturers (Figure 65). The 'un-derivatised' spectra contain physical as well as chemical properties relating to tablets. The polar qualification method was used to determine whether it was possible to separate centre of gravity points relating to tablets from different manufacturers.

The polar qualification method is described in section 1.10.5.

The first step of wavelength selection involved constructing a scanning probability window for two of the manufacturers' tablets. Ten spectra were taken from the sets of Cox and Wellcome tablets with the scanning probability window calculated using a window size of 200 nm (Figure 66).

The plot shows the log (probability) of difference between the centre of gravity points for each wavelength, calculated using 200 nm segments of spectra, starting at 1100 nm. This was calculated for each datapoint.

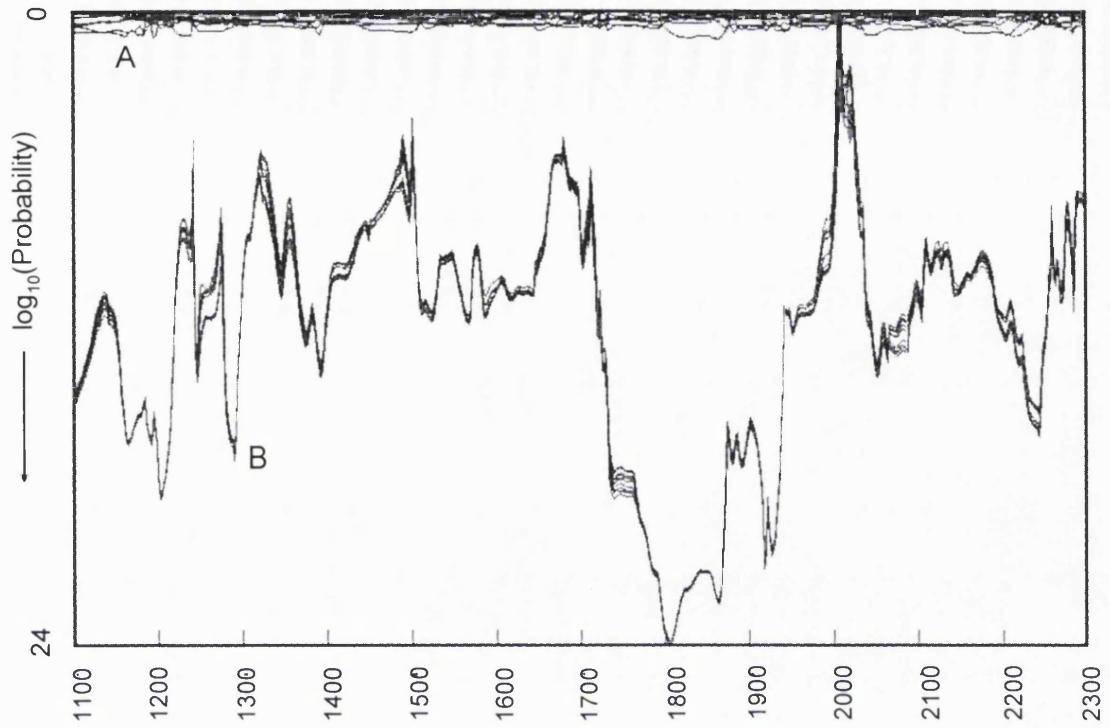


Figure 66 Scanning probability window based upon data for allopurinol 100 mg tablets BP from two different manufacturers (Cox and Wellcome)

X min. = 1100 nm X max. = 2298 nm delta X = 100 nm Y min. –23.998541 Y max. = -0.000018. Window size was 200 nm. A is the reference set (Wellcome) and B is the test set (Cox).

The region between 1750 nm and 1800 nm exhibits the largest variation from the scanning probability window. The range on which to focus the centre of gravity identification was therefore taken to be between 1750 and 1850 nm for comparison of the centre of gravity points.

The equal frequency ellipses constructed from the centre of gravity points for the Cox and Wellcome allopurinol 100 mg tablets can be seen in Figure 67 (C and D). The ellipses represent the area within which 95, 99 and 99.9 % of the population falls.

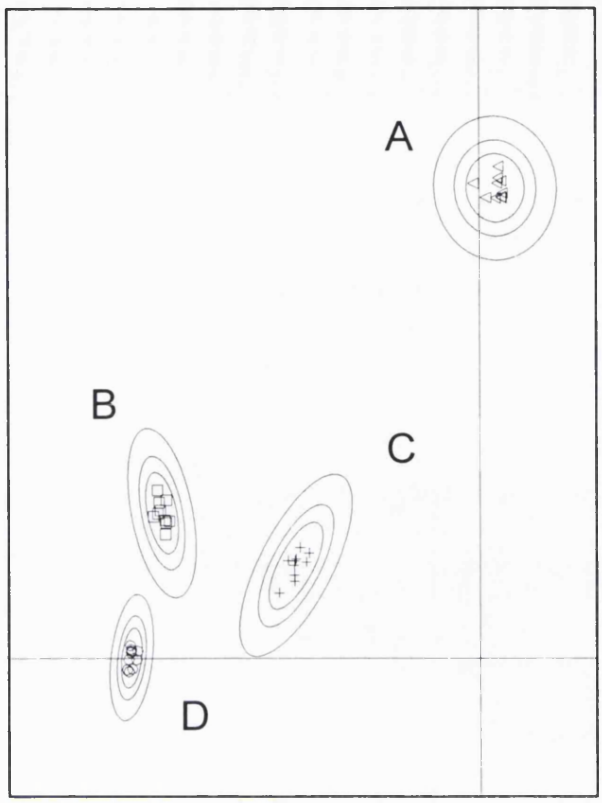


Figure 67 Centre of gravity plots (1750 - 1850 nm) of allopurinol 100 mg tablets BP from four different manufacturers (APS, Cox, CP, Wellcome)

A=APS, B=CP, C=Cox and D=Wellcome. The equal frequency ellipses for a given manufacturer are 0.95, 0.99 and 0.999 (inner, middle and outer respectively). The scales for the axes are X min.= -0.015760 X max. = 0.073238 Y min.=-0.027013 Y max.=0.108089. Range = 1750 - 1850nm, p=0.95, 0.99, 0.995

The plots can be seen to seperate well the centre of gravity sets of the Cox and Wellcome tablets. Two further sets of centre of gravity points were calculated for allopurinol tablet spectra from APS and CP manufacturers and these are seen labelled as A and B in Figure 67.

The four ellipse groupings are distinct from each other.

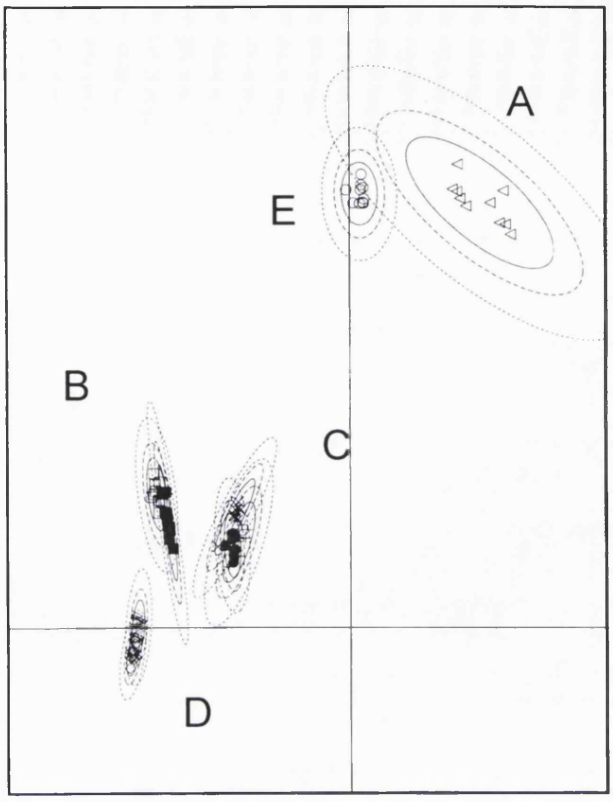


Figure 68 Centre of gravity plots (1750 - 1850 nm) of allopurinol 100 mg tablets BP from five different manufacturers (APS, Cox, CP, Rima, Wellcome)

The manufacturers are represented by the letters A – D; A=APS, B=CP, C=Cox, D=Wellcome, E=Rima. Three additional batches of tablets were used from CP, Cox and Wellcome.. The equal frequency ellipses for a given manufacturer are 0.95, 0.99 and 0.999 (inner, middle and outer respectively). The scales for the axes are X_{min}=-0.020446; X_{max}=0.075959; Y_{min}=-0.095931; Y_{max}=0.128032

Having separated the centre of gravity points from the four manufacturers, variation among batches was included by the introduction of tablets from three further batches from Cox, CP and Wellcome (B, C, and D). This is shown in Figure 68. A batch from a fifth manufacturer of allopurinol 100 mg tablets is also added (E). The ellipses for the additional batches from Cox, CP and Wellcome are still distinct from each other, even at the 99.9% level. The final step of adding in tablet data from all nine manufacturers of allopurinol 100 mg tablets was not as successful because ellipses from the additional four manufacturers

products produced much overlap with the ellipses in Figure 69. The scanning probability window was rerun to include eight manufacturers in the test set against the ninth used as the reference.

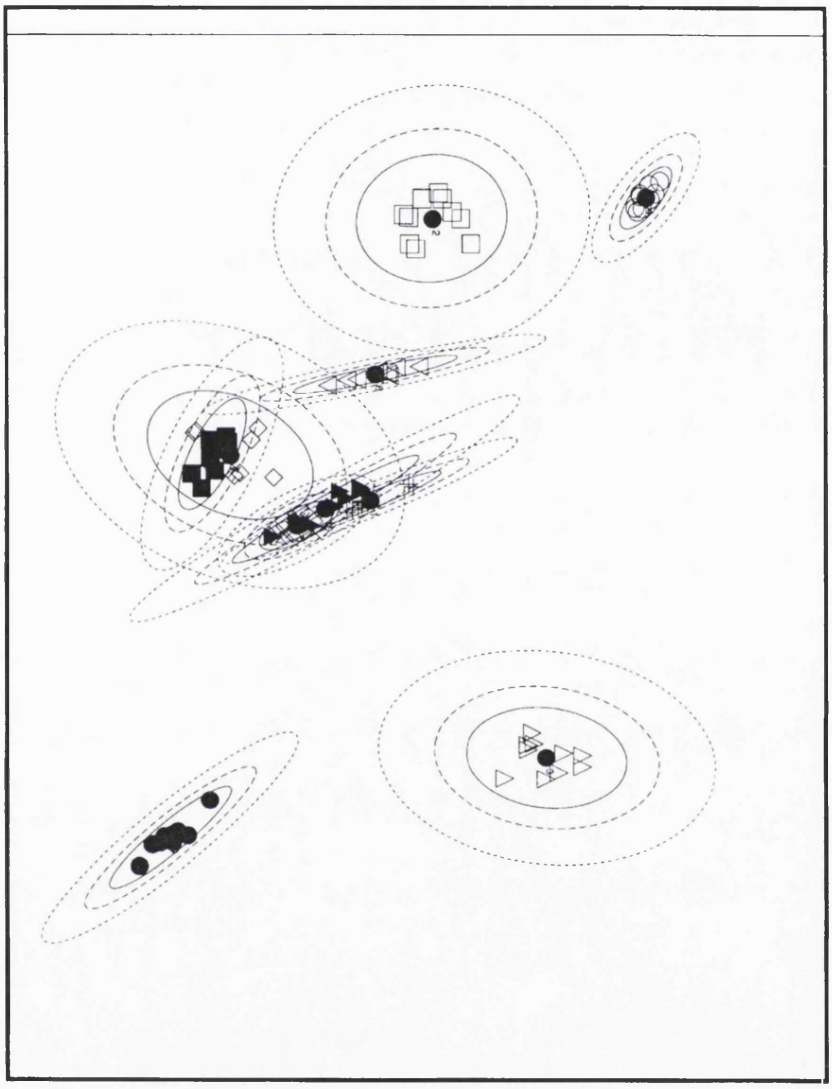


Figure 69 Centre of Gravity plot comparing tablets from nine different manufacturers between 1100 nm and 2500 nm for allopurinol 100 mg tablets

The equal frequency ellipses for a given manufacturer are 0.95, 0.99 and 0.999 (inner, middle and outer respectively). The scales for the axes are Xmin=-0.000072; Xmax=0.002808; Ymin= -0.002520; Ymax= -0.000608

The difficulty that was found was that there was no clear single wavelength region of difference common to all nine manufacturers. Five manufacturers were found to be the most that could be separated in a single comparison (as shown above). A common result of investigating various wavelength regions of three to five separate sets of data and the remainder overlapping can be be seen in Figure 69. Van der Vlies et al[90] found that PQS can distinguish between identical chemical entities that differ only slightly in terms of purity of physical properties. The ability of the polar qualification technique in being able to represent a spectrum as a single centre of gravity point and to define ellipses for populations of these points proved successful here for a small number of different manufacturers of like tablets. This was in combination with a scanning probability window which provided a semi automated method of determining specific wavelength ranges to examine for the best separation. Greater than five manufacturers proved to be difficult to compare on a single plot. However all of the nine manufacturers could be separated when examining pairs of manufacturers at different wavelength ranges.

7.4 Conclusions

This chapter demonstrates the identification of allopurinol and aspirin in intact tablets. The objectives and approach to identification used in Chapter 5 are applicable here. This bases criteria for identification on second derivatives, standard conditions (such as second derivative type and smoothing parameters) and a hierarchy approach to identification.

To identify the active component in the tablets from a database of drug actives, the correlation matching provides the best method. Selection of an appropriate wavelength range can optimise the match. A higher % strength of drug active in the tablets increases the matching statistics. The best range for the identification of allopurinol is 2000 nm – 2400 nm.

Identification of tablet manufacturer is undertaken by using :

- Original spectra – rich in physical information and appropriate for this type of study rather than chemical identification.
- Scanning probability windows – provide the optimal wavelength range of difference between two or more sets of data.
- Polar co-ordinates – a polar plot representing the original spectral absorbance values and wavelength points.
- Centres of gravity – mathematical transformation of the polar co-ordinate plot into a single point.
- Ellipses – spectra from tablets from a single manufacturer will fall within a calculated equal frequency or confidence ellipse. Further samples from

the same manufacturer are seen to fall within the ellipse. Samples from different manufacturers fall outside of the ellipse. Ellipses can be calculated for tablets from different manufacturers.

Specific criteria to identify a certain type of tablet from specific manufacturers can be determined with a large number of tablets, batches and manufacturers' products. Individual manufacturers could provide this data for given products. Collation of this data for all manufacturers of a product into a central database would provide an effective means of detection of any products not from these sets of data i. e. counterfeit products.

Chapter 8 Conclusions

Based on the existing standards for the identification of drug substances, criteria are often open to interpretation, wordy and in some cases unclear.

For TLC, varying the conditions found the TLC systems themselves to lack robustness. This emphasised the need to adhere to standard control conditions. My recommendation for these standard conditions is stated in Chapter 3. TLC systems can also be difficult to stabilise but the use of internal reference standards can improve this.

It is very difficult to set absolute criteria for identification but search systems should be based upon probability. The methods using probability, such as mean list length and discrepancy index, give better confidence of results than techniques like fixed and moving windows.

NIR is a method that can incorporate variation in sample presentation and treatment of spectra. With NIR, again, adherence to standard operating procedures is essential to get reproducible and comparable spectra.

Mathematical pre-treatment of spectra, such as second derivatives are more reproducible and provide increased discrimination than original spectra.

A hierarchy approach of identification procedures gives enhanced capabilities.

My recommendation for the criteria for identification by NIR are described in Chapter 5. This is applicable to the identification of drug substances using methods such as correlation and wavelength distance matching and to the identification of the drug in a tablet.

A simple six peak index can also provide good confirmation for a large number of pure drugs. This is enhanced further by using more peaks such as twelve

and twenty five peak matching and using these in conjunction with visual inspection of troublesome substance spectra.

Potential can also be seen towards defining criteria for the identification of tablets; not just the identity of the tablet, but the identity of the tablet manufacturer. A continuation of the work involving scanning probability windows, polar co-ordinates, centres of gravity and ellipses will provide more information towards the robustness of these techniques. It will also ultimately define the criteria upon which identification of these can be based.

Appendix 1

Table 62 The list of closely related compounds used for the identification of barbituratetes and tetracyclines in Chapter 5

Substance
Allobarbitone
Allopurinol
Ampicillin
Anhydrochlortetracycline hydrochloride
Benzyl penicillin
Brallobarbitone
Butobarbitone
Butobarbitone carboxylic acid
Chlortetracycline hydrochloride
Cimetidine
Cyclobarbitone
Cyclopentobarbitone
Demeclocycline hydrochloride
3'Hydroxy butobarbitone
Imipramine
Limeciclina
Methacycline hydrochloride
Methylphenobarbitone
Oxytetracyline hydrochloride
Oxytetracyline hydrochloride
Papaverine
Pentobarbitone
Quinalbarbitone
Secbutobarbitone
Talbutal
Tetracycline hydrochloride
Vinbarbitone

Appendix 2

Table 63 The 301 drug substances contained within the spectra and second derivative pure drug databases. Included are the peak positions of the twelve most intense peaks.

File Id	Substance	Peak 1	Peak 2	Peak 3	Peak 4	Peak 5	Peak 6	Peak 7	Peak 8	Peak 9	Peak 10	Peak 11	Peak 12
T0287	2-Amino 1-nitrophenyl-1,3 propane diol	2030.8	1533.9	2456.9	2208.5	2110	2138.5	2268.2	1658.8	2242.1	1492	2190.1	2160.1
T0507	2-Amino phenol	2122.2	2213.7	2244.7	2064.8	2339.8	2390.5	1677.1	1529.6	2154.2	1140	1492.8	2179.2
T0279	2-Amino pyridine	1473.1	1662.5	1523.5	1134.2	1969.9	2094.4	1637.2	1981.8	1192.3	2044.4	2272.4	2025.3
T0313	4-Acetamidophenol	2462.5	2042.8	2378.4	2021.8	2151.9	2328.7	2117.3	2310.5	1638.2	2432.2	1669.9	1135.7
T0493	4-Amino-n-butyric acid	2307.6	2376.7	2268.3	2252.5	2342.7	2285	2434.1	1696.3	2481.2	2461.8	1737.7	1714
T0495	Acetomenapthone	2250.8	2131.8	1664.3	2416.8	1923.6	2304.3	2402.7	2072.6	2364.4	2180.6	2487.2	2089.6
Sa005	Alginic acid	1917.3	1437.5	2481.9	2322.1	2097.9	2476.7	2341	2303.6	1737.8	1723.4	1201	2133.4
T0280	Allantoin	1987.3	2049.8	2125.4	2319	1963.1	1463.6	1521	2079.8	2267	1732.6	2341.4	2194.9
T0230	Allobarbitone	2223.4	1626.9	2110	2291.2	1692.6	2167	2334.4	2374.7	2413.3	2443.9	2272.3	1744.2
T0255	Allopurinol	2466.9	2226.7	2348.7	2305.6	2087.6	1640.9	2163.9	2200.8	2382.6	2249.1	2277.5	1117.3
Rw008	Ampicillin	2068	2242.9	2371.2	2333.7	2205.6	2313.6	2472.6	2010.4	1713.6	2299.7	1521.9	2165.2
Rw002	Ampicillin sodium slat	2140.7	1527.9	2466.3	2028.4	2307.5	2271.3	1677.5	1140.8	1726.7	2164.6	1695.3	1906.6
T0319	Amylase lipase	2309.4	2349	1728.9	2057	2467.3	1517.7	1693.4	2166.2	1763.3	1932.9	2260	1188.3
T0045	Amylobarbitone	1692.4	2473.8	1733.9	2157.5	2262.1	2314.7	2280	2347.8	2374.8	2218.6	2420.6	2458.6
T0314	Aneurine hydrochloride	1944	2276.8	1970.7	2314.8	2375	2408.2	2477.8	1691.9	1438.6	2296.5	2204.3	2254
T0249	Anhydrochlortetracycline	2359.7	1524.8	1768.5	2322.9	2280.4	2092	1660.9	2243.9	2434.7	2378.5	2122.1	2487.3
T0248	Anhydrotetracycline hydrochloride	2239.4	2343.1	2382.2	2278.4	1985.1	2449.9	1662.7	1521.7	1479.7	2316.9	2015.1	2430.6
Sa015	Antazoline	1897.6	2265.1	2469.8	2138.4	2450.6	2246.2	1398.8	2163.8	2184.8	1463	1141.6	1730.4
Sa007	L-Ascorbic acid	2249.7	1456.9	2480.3	1749.5	2096.7	2144.8	1353.2	2012.9	1486.7	2385.6	1687.4	2275.4

File Id	Substance	Peak 1	Peak 2	Peak 3	Peak 4	Peak 5	Peak 6	Peak 7	Peak 8	Peak 9	Peak 10	Peak 11	Peak 12
Sa011	Aspartic acid	1710.3	2363.2	2112.3	2382	2166.2	2464.3	2309.1	2427.5	2291.6	2336.3	1685	2452.6
T0081	Aspirin	2400.2	2253.6	2137.3	1656	2326.9	2483.5	2192.7	2424.6	1128	1678.2	1719.5	2281.7
T0284	Astemizole	2457.6	2136	1664.1	2158	2342.7	1750	2408.8	2075.1	2289.6	2189.8	1704.7	2252.6
T0316	Atenolol	2466.9	2163.1	2303.2	1550.8	1432	2272.5	2095.4	2133.5	2447.9	2229.5	2188.4	1986.8
T0318	Atropine	2141.1	2474.2	2246.2	1943	2167.2	2288	2458.5	1894.6	1670	1692.2	2186.1	2389.2
T0500	Avicel PH-101	2273.1	2483.9	1918.6	2336.7	2105.7	1424.8	2078.3	1479.9	1704.8	1366.6	2358.2	1590
T0055	Barbitone sodium	2287.5	2174.4	2406.2	2364.8	2341.3	2304.3	1694	2052.6	2269.5	2469.1	2421.9	1741.2
T0222	Barbituric acid	2404.6	2480.9	2357.3	2336.8	2300.2	2215.4	2161.7	1771	2017.4	1519.5	1712.9	2036
T0315	Benzil	2145	2467.4	2165.8	2453.9	1666.2	2281.7	2341.1	2018.4	2409	1136.7	2215.9	2364.5
T0056	Benzocaine	2262.3	1982.6	2295.4	2161.3	2220.1	2465.8	2098.3	2053.8	2135.7	1533.2	2188.9	2347.6
Sa012	Benzoic acid	2462.5	2445.5	2138.3	1658.6	2159.5	2382.3	2418	2186.4	1134.2	1680.2	2280.4	2211.9
T0497	Benzoin	2143.5	2462	2164.5	2214.3	1667.9	2187.3	1514.6	2307.1	1137.1	2420.2	2341.2	2445.5
T0286	Benzophenone	2146.4	2217.7	1664.6	2166.6	2374.3	2303.2	2191.8	2468.6	1137	2036.8	1765.3	1683.8
T0282	Benztropine mesylate	2264.3	2140.5	1937	2241.9	2471.9	1679.5	2313.6	2186.7	2169.2	2457.9	1658.6	2398.9
T0281	Benzyl cinnamate	2141.2	2473.5	2250.2	1670.3	2164.9	2187.9	2457.4	2291.6	2217.2	2353.2	1138.6	1690.7
Rw005a	Benzyl penicillin sodium salt	2063	2187.9	2303	1526.2	2218.3	2458.7	2476.2	2136.6	2249.4	1713.5	2166.3	2400.3
T0289	Betamethasone valerate	2108.2	2294	2273.5	2045.8	1477.5	2153.8	1705.8	2463.2	2125.4	2351.2	1678.7	2335.1
T0221	Brallobarbitone	2227.8	1630.3	2115.6	2286.8	1692.5	2373.8	2414.3	2171.3	2344	2265.6	1742.4	2153
T0223	Butobarbitone	1691.3	2299.7	2283	2342.7	2159.1	2419.6	2481.5	2262.7	2218.3	2393.8	2376.3	2455.4
T0225	Butobarbitone carboxylic acid	2366.6	2430	2263.7	2477.7	2287.1	1686.9	2156.2	2298.9	2403.7	2336.5	1520.6	2347.2
Sa037	Iso-Butyl para-aminobenzoate	1980.2	2161.5	2463.6	2137.3	2220.1	2270.6	1532.6	2188.3	1662	2097.8	2053.2	1469
T0337	Caffeine	1670.7	2328.3	2431.9	2254	2417.9	2238.7	2144.7	2078	1635.2	1924.7	2099.3	2279.1
T0331	Calcium pantothanate	2308.1	2268.7	1684.3	2028.4	1498.6	2180.4	1696.8	2403	1188.6	2463.7	1934	2359.6
T0047	Carbamazepine	1980.8	1951.1	1468.4	1518.8	2194.3	2054.4	2403.8	2152.1	2446.3	1674.6	2176.7	2367.7
T0332	Carbenoxolone sodium	1891.6	2292.4	2256.8	1701.1	1694.8	1390.1	2305.8	2409.1	2326.8	2135.3	1746.7	1182.6
T0330	Carbidopa	2044.6	1952.6	2404.8	2101.6	1573.1	2143.3	2280.7	1678.1	2252.3	1552.6	2289.2	2230.5
Sa020	Carbromal	2225.6	2487.4	2186.5	2295.9	2070.6	2023.4	2265.3	1691.7	2426.3	2087.5	2387.7	2349.1
T0336	Carprofen standard	1497.9	2459.5	2257	2295.1	2391.6	1670.9	2143.6	1983.8	2327.2	2161.3	2037.4	2177.6
Rw001	Cefuroxime sodium salt	2219.6	2320.5	1764.3	1926.6	1747.7	2264.2	2158.2	2068.2	1630	2454.9	2010.7	2390.9

File Id	Substance	Peak 1	Peak 2	Peak 3	Peak 4	Peak 5	Peak 6	Peak 7	Peak 8	Peak 9	Peak 10	Peak 11	Peak 12
Sa018	Cetrimide	2312.3	2272.1	2246.9	2351.2	1730	1898.4	1672.1	2387.2	1764.3	2447.1	2219.6	1391.5
T0505	Cetylpyridinium bromide monohydrate	2310.7	2350.4	1984.7	1729.2	2184.3	2421.7	2264.9	2155.8	1762.8	2133.7	2100.3	2475.4
T0506	Cetylpyridinium chloride monohydrate	2312.5	2351.4	1730.5	1994.6	2181.9	1764.2	2135.8	2157.6	1649.4	2439.5	2484.1	2386.7
T0504	Cetyltrimethyl ammonium chloride	2312.5	2271.8	2246.8	2351.6	1729.9	1764.3	1672.1	2447.6	2386.8	2487.1	1655.6	2219.4
Sa013	Chloramphenicol	2374.5	2254.5	2338	2060.8	1697	1723.8	2446.2	2128.3	2458.4	1651.7	2038.6	2316.4
T0334	5-Chloranilin-4,6-disulfur-saedredamid	1992.5	1960.4	2183.2	1447.8	2200.7	2051.8	2079.6	1517.9	2035.9	1493.6	2476	2245
T0339	Chlorcresol	2311.5	2432.3	2451.8	2266.8	1668.3	2136.4	2169.2	1697.6	2404.5	2294	2487.4	1771
T0338	Chlorhexidine acetate	2339.4	1668.6	2458.4	2300	2289.4	2382.3	1134.5	2132	2271	1995.8	1723.3	2163.4
T0320	4-Chlorobenzene sulfonamide / calcium pantothanate	2452.7	2147.2	2268.4	2044.5	2403.3	1662	2073.8	2318.9	2350.2	2183.2	1583.4	2202
Sa021	Chlorocyclizine hydrochloride	2145.7	2274	1669.9	2459.8	1718.1	2186.1	2175.1	2362.4	2410.6	2382.4	2475.8	2248.8
Sa017	Chlorohydroxy quinoline	2374.2	2482.9	1664.1	2425.5	2249	2137.9	2190.8	2319.8	1132.2	2233	2264.5	2163.3
Sa016	Chloroquine phosphate	1897.5	2253.9	2284	2301.6	2337.7	1409.7	1667.5	2423.4	2455	2389.7	1177.5	1683.7
Sa014	Chloroxylenol	2314	1698.4	2454.6	1768.3	1735.9	2280	2154.3	1187.3	1872	2385.1	1679	2013.9
T0333	Chlorphenoxamine hydrochloride	2265.1	2146.3	2300.1	2380.7	2170.2	2413.7	2470.8	2245.8	2459.2	2338	2352.2	1672.5
T0083	Chlorpromazine	2433.4	2350.3	1667.8	2249.7	2266.4	2159.6	2386	1914.2	2228	2479	1514.7	1135.2
T0079	Chlorpropamide	2452	2044.8	2147.4	2402.6	2220.3	1524.3	2186.9	2282.2	1662.3	2307	2348.8	1562.5
T0238	Chlortetracycline hydrochloride	2245.2	2337.8	2441.8	1660	1520.2	2354.7	2487.4	2381.1	2117	2131.5	2270.9	2290.9
T0487	Iso-Chlortetracycline hydrochloride	2258.7	2335.7	2236.9	2292.8	1957.6	2450.1	2397.1	2114.6	2365.5	1683.7	2321.1	1903.6
Sa019	Chlorthalidone	2436.9	2039.3	2248.6	1663.2	2389.3	2311.6	2129.7	2405.6	2078.7	1131.3	2188.8	2487.2
T0042	Cimetidine	2469.1	2293.1	2329.3	2308.1	2248	1746.9	1697.6	2390.3	2168.5	1729.9	2408	1851
T0335	Cisapride	2164.7	2203	2032.7	1964.4	2459.7	1507.2	2262.6	1465.8	1660.7	2080.6	2045.1	2282.3
T0326	Clindamycin phosphate	2266.8	2243.3	1744.6	2300.2	1910.8	2222.8	1711.9	2355.6	2454.7	2320.2	2065.4	1422.1
T0327	Clobetasol propionate	2257	2299.4	2285.3	2241.3	2098.1	2122.3	2342.7	1688.6	2155.1	2386.7	2320.5	2462.4
T0385	Codeine phosphate	2253.5	1926.5	2050	2033.7	2396.6	2153.5	1719.1	2093.5	2279.4	2295.7	1679.5	2125.9
T0224	Cyclobaritone	2301.6	2362.4	2396.8	2344.2	2155.6	1754.3	2325.1	2265.1	1723	1523.4	1682.5	2424.2
T0220	Cyclopentobarbitone	2222.3	2108.1	1625.6	2288	2447.3	2162.4	1691.7	2305.4	2485.4	2267.6	2354.1	2329.2

File Id	Substance	Peak 1	Peak 2	Peak 3	Peak 4	Peak 5	Peak 6	Peak 7	Peak 8	Peak 9	Peak 10	Peak 11	Peak 12
T0328	Cyclopentolate hydrochloride	2140.1	2355.2	2263	2164.5	2385.6	2467.5	2252	2448.3	1672.1	2289.9	2417.8	1658.2
T0084	Dapsone	2202.9	1906.7	2167.3	2466.3	1977.3	2143.6	1447.7	2075.4	1667.2	1413.6	1521.8	2033.2
T0344	Dehydro-isoandosterone	2043.9	2272.2	2338	2300.7	1746.9	2094.7	1926.1	2371.8	2149.7	1697.9	1949.5	1713.3
T0235	Demeclocycline hydrochloride	1926.9	1662.8	2243	2437.5	2256.6	2336.8	1520.4	2479.4	2352.8	2276.8	2380.2	1717.8
T0486	Demethylchlortetracycline base	2240.6	2202.2	1505.5	1955.6	1470.4	1991.6	2016.7	1984.8	2344.1	1653	2330.7	2462.7
T0078	Desipramine	2258.1	2148.5	2320.5	2399.9	1678.8	1142.7	2295.4	2177.3	2485.8	2437.6	1763.4	1514.5
T0345	Dextromethorphan	2251	2302.3	2338.8	2438	2285	1756.6	2262.4	2403.6	1676.3	2127.7	1977	1714
T0073	Dichlorophen	2400.9	2139.7	1668.3	2276.4	2365.4	1134.3	2172	2152.5	1518	1742	2329.7	1182.2
T0241	Diclofenac	2042.7	2406.6	2014.4	1668.4	2084.3	1508.1	2291.6	2062	2159.3	2142.2	2209.1	2230.9
T0346	Dicoumerol	2385.6	1674.6	2458.9	1646.5	2339.8	2203.7	2318.7	2127.3	1138	2287.1	2139.7	2229.5
Sa022	Dienoestrol	2461.5	2157.5	2338.2	1704.8	2369.1	1670.3	2137.8	2229.7	2324.3	2274.7	2404.8	1138
Sa023	Diethylcarbamazine citrate	2249.9	2394	1697.5	2462.3	2294.9	2307.2	2331.3	1664.4	1680.5	2358.1	2261.2	1719.2
T0077	Diflunisal	2398	1655.9	2429.8	1514.8	2472.2	2322.8	2190	2144.3	2124.9	2299.6	1126.8	2281.8
T0342	Digoxin	1449.6	2346	2104.3	2295.9	2407.8	1714.8	2030	1733.2	2325.8	2088.9	2282.3	2475.4
T0341	Diltiazem hydrochloride	2432.4	2452.2	2161.6	2259	2351.1	2132	1674.2	2390.9	1742.1	1659.4	2335.2	1725.3
Sa024	Dimenhydrinate	2144.6	2464.4	2250.7	1678.2	2262.5	2480.7	2347.1	2172.7	2418.4	2383.2	2189.1	1140.3
T0343	Dipenefrin	2309	2270.1	2383.2	1681.9	2419.1	1695.2	2460.3	2156.8	1185	1733.2	2356.5	2398.8
T0288	Diphenhydramine hydrochloride	2263	2145.7	2248.1	2230.1	2487.1	2349.2	2389.3	2191.4	2463	1684.1	2163.5	2428.3
T0348	Disulfiram	2268.7	2285.1	2304.2	2396.9	2421.8	1735	2353.3	1693.3	1682.9	2317.7	2460.4	1183.9
Sa025	Dithranol	2345.8	1670.5	1724.2	2458.7	2423.7	1136.2	2148.4	2304.3	2136.3	1632.6	1787.1	1768.5
Sa026	Domiphen bromide	2309.5	2468.9	2142.4	2349.3	2240.7	1674	1728.7	1967.4	2000.9	1763	2256.9	2418
T0347	Doxcycline	1972.4	2232.5	2077.2	2385.8	1662.2	2278.9	2241.4	2252	2335.7	2432.1	2347.9	2022.7
T0053	Doxepin	2250.3	2267.6	2353.5	2176.3	1669.7	2320.3	2387.9	2144.3	2405.8	2452.1	2485.7	2229.8
T0350	Econazole nitrate	1606.8	2458.8	1665.2	2278.1	2307.7	2383.2	2349.2	2434.8	2141	2180.7	2408.2	1711.9
T0362	Edetate disodium	2234.4	1930.6	2420.1	2262	2281	2328.9	2460.7	2301.6	1673.1	1716.1	2246.9	1445.4
T0363	Embramine	2266.8	2466.5	2251.9	2378.7	2142.6	1673.3	2453.1	2345.2	2324.9	2413	2229.6	2170.4
T0054	Ephedrine hydrochloride	2469.8	2290.3	2387.1	2249.8	1667.6	2142	2311.7	2188.5	2436.2	2456.7	2360.2	1138.7
T0252a	4-Epianhydrotetracycline	2241.7	1909.5	1520.7	2335.9	1663.8	2487	2210	2259.9	2383.8	1478.6	1416.6	2086.1
T0250	4-Epichlortetracycline hydrochloride	1906.3	2337.1	2240.9	2255.2	1659.8	1521	1397.9	2487.1	2279.2	2024.5	2383.6	1925.5

File Id	Substance	Peak 1	Peak 2	Peak 3	Peak 4	Peak 5	Peak 6	Peak 7	Peak 8	Peak 9	Peak 10	Peak 11	Peak 12
T0251a	4-Epitetracycline	2248.5	2338.7	2430.2	2223.7	2142.8	1519.9	1952.3	2470.4	1679.7	1737.8	2375.3	2276.3
T0351	Erythomycin ethylsuccinate	2301.8	2257.8	1693.1	1678.6	2350.3	1973.7	1734.5	1721.4	2451.2	1182	2054.2	2399.4
T0356	1,3,5(10) Estradien-3,17b diol	2291.1	2171.2	2342.2	2311.1	2413.1	1184.7	1744.2	2381.9	1759.2	2273.5	1960.9	1717.3
T0364	Ethoprapazine hydrochloride	2259.4	2305.1	2438.6	1677.8	2168.2	2158.8	2339.4	2403.8	2240.2	2388.5	1728.4	2187.2
Sa027	Ethosuximide	2263.7	2292.4	2398	1689.2	2456.7	1730	2343.5	2304.7	2370.1	1183.4	2106.4	2439.9
Sa029	Ethyl para- hydroxybenzoate	2256.6	2306.5	2448.8	2290.4	2344.2	2407.4	1690.8	1135	2487.8	1663.6	1181.9	1677.6
Sa028	Ethyl para-aminobenzoate	2262.5	1983.9	2299.7	2160.8	2221.8	2466.1	2100.3	2055.5	2444.3	2136	2348.8	1532.4
T0361	Fluconazole	2348.9	2272.9	2232.5	2257.9	2419	2377.6	2154	1691.7	2307.3	2444.7	2451.4	1663.9
T0365	Fluocinolone	2117.5	2274.6	2056.9	2355.5	1735.9	2289.9	1681.3	2239.9	2325.7	1902.2	2466.8	2381.9
T0360	Fluoxymesterone	2283.6	1425.4	1744.3	2036.5	2124	1676.8	2298.6	2333.8	2168.9	2240.7	1697.8	2425.5
T0353	Fluphenazine	2429.7	2270.3	1922.1	2384.7	1673.8	2160.6	2323	2143.6	1135.1	2460.1	2477.8	2303.3
T0354	Flurazepam hydrochloride	2261.3	2149.7	2304.3	1661.7	2136.2	2398.3	2475.8	1690.5	2197.6	2456.8	2330.9	1722.2
T0359	3-Formyl Rifamycin	2258.4	1681.9	2292.3	1899.4	1695.8	2454.2	1770.4	1727.5	1179.9	2396.7	2203.7	2360.2
T0051	Frusemide	2033.7	2105.9	2176.2	1527.5	2270	2015.5	2383.4	2304.9	2478	2061.8	2422.3	2090
T0358	Glibenclamide	2193.1	2236.4	2042.4	2458.3	2215.3	2067.5	2307.3	2167.5	2276.3	2143.5	2121.3	2339.8
T0058	Glutethimide	2472	2145.3	2297	2168.5	2267.8	2190.2	2341	2458.3	1696.2	2215.6	1665.8	1741.8
T0491	Glycine	2251.2	2321.9	2454.5	2431.8	1669.1	2262.6	2151.6	2340.1	1163.3	1716.6	1367.9	2100.3
T0498	Glycine hydrochloride	2269.9	2460	2436	2320.5	2150.9	2112.3	2390	2370.3	1606.6	2082.6	2248.9	2343.8
T0357	Griseofulvin	2264	2418.6	2301.5	2384.1	2113.8	2331.1	2470.7	2229.2	2248.3	2151.8	2357.1	1681.5
T0375	Heptabarbitone	2303.6	2416.2	1754.4	2342.1	1725	2267.9	2384.8	2352	2140.3	1672.4	1785.3	2430.2
Sa033	Hexobarbitone	2289.3	2248.1	2317	2141.2	2034.5	1763.8	2001.8	2359.9	2340.3	1748.4	1677.4	1721.2
Sa031	Hexoestrol	2468.4	2305.5	2399.1	1686.8	2166.8	2344.2	2266.3	2449.6	2153.9	2132.3	1722.3	1767
Sa030	4-n-Hexyl resorcinol	2310	2348	1983.2	1762	2046.1	1729.3	2157.1	2458.6	2395.3	2186.3	2476.6	2433
Sa032	Hydralazine hydrochloride	2095.6	2387.6	1678.6	2045	2154	2481.3	2324.4	2423.6	2274.1	2254.3	2340.5	2182.2
Sa034	Hydrochlorthiazide	2034.7	2456.3	2172.8	1517.6	1663.2	2108.1	2247	2013.6	2387.4	2084.2	2356.2	2485.7
T0370	Hydrocortisone acetate	2472.1	2276.4	2130.9	2289.6	2052	2347.2	1744.7	1967.1	2243.9	1708.1	1675.3	2105.7
T0383	Hydrocortisone alcohol	2472.1	2276.4	2130.9	2289.6	2052	1744.7	2347.3	1966.7	1708.1	2243.8	1675.3	2384.9
T0290	Para-Hydroxybenzoate	2455.6	2009.3	2106.9	2298.7	2167.4	2068.6	1973.7	2267.8	1657.2	2148.4	2185.9	2391.7
T0366	L-Hydroxyproline	2228.5	1726.5	2123.5	2318.5	2270.6	1680.9	2369	2393.7	2442.5	2296.5	2474.5	2487.1

File Id	Substance	Peak 1	Peak 2	Peak 3	Peak 4	Peak 5	Peak 6	Peak 7	Peak 8	Peak 9	Peak 10	Peak 11	Peak 12
T0369	Hydroxyzine	2142	2272.7	2454.6	2473.6	1673.9	2170.7	2254.6	1968.1	2388.6	1137	2408.6	2074.3
T0059	Hyoscine hydrobromide	2144.7	1656.8	1720.4	1691.7	1676	2476.9	2273.5	2186.4	2338.8	1938.3	2207.3	2402.6
Sa035	Ibuprofen	2462.6	2172.1	2313.5	2351.4	2259.6	2294.7	2277.1	2139.9	1691.6	2421.6	2196	1674.8
T0382	Iminodibenzyl	1999.6	2035.4	1503.8	2175.9	2356.8	2474.6	2073.9	2314.8	2139	2407	1678.1	1770.8
T0075	Imipramine	2268.4	2353.2	2253.3	2144.5	2179.8	2392.9	2320.7	1676.4	2293.9	2487.1	1764.9	2420.3
T0371	Indapamide	1906.2	2085.8	1408.7	2312.3	2144	2270.9	2429.4	1960	1681.8	2166.8	2354.9	1737
T0367	Indomethacin	2453.9	2388	2308.5	1695.3	2142.5	2281.8	2266.4	1727.8	1665.8	2343.7	2236.1	1652.5
T0368	Iodamide	2279.4	2030.3	1511.1	2148.8	2262.4	1959.7	1689	2104.7	1737.3	2233.4	2321.7	2183.4
Sa036	Iodohydroxyquinoline	2148.7	2388.6	1671	2261.4	2172.4	2323.5	2436.4	2204.6	1136.4	2250.1	2354.6	2410
T0374	5-Iodouracil	1653	2270.3	2452.8	2138.1	2198.9	2328.2	2420.3	2488	2377	2214.6	2104	1785.4
T0379	Ioxacilic acid	1905.4	2263.3	2295.3	2054.2	2317	1685.8	2433.6	1406.8	2385.2	1724.4	1941.4	2113.7
T0372	Ioxhexol	2285	1908.3	2057.2	1940.3	1430.9	1687	2325	2469.1	1725.5	1518	2184.2	1772.4
T0378	Iproniazid	1642.8	2291.3	2164.3	2423.3	2258	1666.5	1683.5	1722.8	2141.9	1176.6	2118.3	1135.1
T0373	Irgasan	2428.3	2397.9	1661.9	2304.6	2140.3	2152.9	2226.9	1131	2176.7	2201.6	2359.5	2487.2
T0381	Isoetharine hydrochloride	2402	2292.6	2156.1	2272.1	1688.2	2336.4	2302.7	1731.3	2258.7	2124.8	2315	2179.6
T0050	Isoniazid	2433.3	2227.2	2156.9	2413.4	1565.4	1646.2	1674.6	2032.2	2255.6	2322.3	2464.3	1543.6
T0380	Isoprenaline hydrochloride	2149.4	2250.1	2292.4	2367.3	2411.8	2212.9	1674.7	2329.5	2436.4	2312.2	1690.1	2191
T0065	Isoprenaline sulphate	2261.4	2295.5	1514.8	2141.5	2361.6	2404.6	2317.6	1915.9	1676.7	1721.7	1177.4	1477.3
T0407	Ketamine hydrochloride	2426.8	2293.4	2272.9	2146.4	1662.1	2126.9	2380.5	2236.3	2333.7	2247.7	1692	2363.4
T0262	Lactose monohydrate	2255.5	1933.8	2277.9	2475.6	2371.8	2346.2	2358.7	2428	2092.6	2316.7	1451.3	1697.5
T0409	Lapetalol	2256.7	1987.1	2140.8	2299.4	2463.6	2478.6	1676.2	2165.6	2032.1	2381.6	2343.4	2229.6
T0394	Levamisole hydrochloride	2257.1	2273.3	1669.6	2479.4	2403.3	1700.1	2168.4	1724.3	2191.9	2146.7	1367.1	1641.5
T0412	Lidocaine	2461.6	1700.1	1682.9	2364.9	2096.7	1740.6	1188.6	2317.2	1772.7	2175.1	1144.9	2014.5
Sa038	Lidocaine hydrochloride	2450.3	1956	2258.7	1987.8	2293.5	1674.7	2426.1	2350.9	1458.2	2154.9	2402.2	2314.1
T0240	Limecycline	1520.5	2247.7	1907.9	2290.8	2339.8	1662.8	1479	1715.8	2143.1	1680.3	1414.4	2390.3
T0405	Lithium clavulanate	2283.9	2342.6	2055	2425.4	2408.5	1688.7	2112.4	2371.5	2266.3	2166.5	2251.6	2454.9
T0411	Loratadine	2257.9	2332.6	2276.3	2178.3	2161.1	2364.6	2401.2	1693.2	2309.8	2437.5	2298.3	1662.6
T0413	Lorazepam	2423.3	1668.1	2147.2	2445.3	2487.7	2371.7	1134.1	2200	2332.5	2387.2	2296.5	2175.2
Sa001	Magnesium acetate	2315.3	2256.9	2292.7	2472.5	2337.4	1449.1	1683.9	1950.8	1525.1	1350.3	1181.6	1667.5

File Id	Substance	Peak 1	Peak 2	Peak 3	Peak 4	Peak 5	Peak 6	Peak 7	Peak 8	Peak 9	Peak 10	Peak 11	Peak 12
Sa002	Magnesium acetate tetrahydrate	1448.8	1350.3	1525.3	1684	2255.6	2315.9	1181.8	2337.5	1159.3	1667.2	2292.6	1376.5
T0403	Magnesium sulphate monohydrate	2413.7	1518.8	1521.3	2066.4	2056.1	1934.3	2322.9	2102.7	2339.3	2111.4	2336.7	1243.4
T0090	Mannitol	2285.7	2446.7	2333.4	2368.7	1515.4	2483.4	2466.9	1603.3	1695.7	1194.4	2131.9	2353.3
T0393	Mebendazole	2190.7	2252.2	1500.4	2142.2	2348	1661.4	2462.8	2226.5	2166.1	2370.6	2280	2025.8
T0401	Meclizine	2455.1	2141.2	2168.6	2269.4	1935.1	2476.4	1673.1	2256.8	2414.6	2296.7	2315.3	2347.4
T0399	Medroxy progesterone acetate	2147.2	2306.2	2116.6	2352.2	1715.5	2261.4	1683.2	2273.1	2385.1	2337.5	1750.3	2461.9
T0215	Mefenamic acid	2044.9	2316.2	2072.9	2273.6	2456.1	2407.3	2364.9	2122.3	2159	2433.1	1665.2	2184.6
Sa043	Menaphthone	2143	2443.6	2406.7	2120.2	1697.8	2366.5	2271.9	2321.1	2200.7	2264.9	1665.2	2310.1
T0076	Meprobamate	1987	2251.7	1459.2	2462.8	1938	1689.4	2023.2	1498.9	2307.7	2054.8	2400.2	1517.8
Sa042	Mepyramine maleate	2243.1	2452.2	2266	2341	1674.3	2385.6	2351.2	2411.8	2135	2282.4	2466.9	2186
T0237	Methacycline hydrochloride	2259.6	1616.5	1663.3	2348.3	2238.9	2209.4	2441.6	2146.4	1720.6	2104.2	1519.9	2071.5
T0396	Methadilazine hydrochloride	2256.1	2435.5	1973.3	2162.7	1670.3	2388	2150.3	1729.4	1947.4	1137.2	1632	2317.5
Sa039	Methoin	2470.6	2455.2	2164.4	2140.4	2267.9	1998.8	2243.5	1684.6	1665.1	1971.1	2344.2	2302.7
T0391	L-a-Methyl DOPA	2245.8	2283.2	1953	2270	2340.3	2447.8	2479.7	2136.8	2077.6	1981	1682	2368.1
Sa041	Methyl para-hydroxybenzoate	2245.2	2366.7	2450.1	2476.7	1664.1	2292.7	2395.8	1130.9	2151.1	2271.6	2205.6	2099.2
Sa040	Methyl sulphonal	2357.6	2290.5	2467.8	2255.4	1669.2	1706.1	1683.2	1725.8	2405.1	2432.1	1177.7	1753.9
T0232	Methylphenobarbitone	2465.5	2414.9	2142.6	2165.9	2259.6	2444.6	2276.4	2295.7	1678.7	2324.8	2371.1	2189.2
T0048	Metoclopramide	2156.5	2024.2	2258	1506.2	2205	1674.3	1959.4	2300.6	2081.7	1997.9	2476.6	1653.8
T0088	Metronidazole	2290.2	2252.5	2440.5	2457.3	2361.7	2394.1	2333.3	1684.6	2270.3	1663.9	2211	2010
T0389	Mexiletene hydrochloride	2259.6	2451.1	1679.2	2296.9	2330.5	1899.7	2156.1	1726.4	2352.8	2389.2	1142.4	2116.5
T0087	Midazolam	2363.6	2261.2	2401.5	1661	2145.1	1636.9	2308.5	2343.4	1697.1	2293.9	1514.9	1679.4
Sa044	Nalidixic acid	2247.5	2331.8	2295.5	1690.4	1665.8	2461.5	2394.8	2263.1	2137.8	2164.4	2099.6	2194.1
T0390	Naloxone	1943	2232.3	2113.1	1629.8	2278.4	2343.3	2298.3	2392.8	1968.4	1745.5	1677.8	2419.5
T0046	Naproxen	2384.9	2343.1	2437.6	2300.4	2362.8	2252.4	1674.4	2267.6	2153	2244.1	2284.7	2125.4
T0386	Neomycin sulphate	1905.8	1427.9	1424.5	1166	1708.8	1469.2	1207.5	1935.7	2242.2	1361.1	1765.4	2306.5
T0043	Nicotinamide	2444.6	2227	2424.4	2137.8	1671	2098.1	1982.5	2194.4	2007.2	2273.8	2484.7	1644
T0089	Nicotinic acid	2138.5	2438.3	1656.9	2284.4	2222.9	2346.6	2198.9	2463.5	1515.2	1124.9	2387.7	2479.5
T0402	Nifedipine	2255.9	2072.3	2372.8	2411.3	2141.9	1658	2318.2	2278.9	2193.5	1531.2	1994	2009.5
T0410	Nifedipine nitrophenyl pyridine	2251.4	2136.7	1950.5	1660	2434.3	2463.1	1689.4	2412	2364.9	2274	1726.4	2401.7

File Id	Substance	Peak 1	Peak 2	Peak 3	Peak 4	Peak 5	Peak 6	Peak 7	Peak 8	Peak 9	Peak 10	Peak 11	Peak 12
Sa045	Nitrofurantoin	1633.7	2403.4	2164.6	2340.1	2300.5	2487.7	1748.1	2317.1	2271.8	1701.2	2429.4	1615.5
T0388	Nitrofurantoin monohydrate	1920.9	1421.6	2339.1	1975.4	2431.3	2163.7	2298.1	2404.2	1633.8	1748.8	1699.7	2105.7
T0492	5-Nitro isophthalic acid	2391.5	2441.1	1643.1	2487	2429.8	1947.1	2305.2	2375.2	2120.9	2339.3	2262.3	1446.5
T0494	4-Nitrophenyl acetate	2445.8	2130	2248.6	2187	2276	1654.5	2477.4	2085.1	1910.3	2321.7	2422.3	2385.4
T0496	p-Nitrophenyl trimethyl acetate	2303.6	2129.9	2446.2	2265.3	2185.9	1651.6	2388.8	1922.1	2221.8	2354.1	1692.6	1639
T0484	5-Nitrosalicyclic acid	1642.4	2477	2405.2	2268.1	2134.7	2374.6	2391.5	2108.1	2189.8	1114.8	2216	2359.2
T0057	Omeprazole	2377.5	2243.5	2425.2	2273.9	2450.7	2298.3	2339.8	1515.1	1673.8	1696.9	2150.3	1656.3
T0068	Orciprenaline	2258.4	2465.3	1933.2	2137.1	2396.7	2313.7	2284.2	1517.2	1684.5	1671.6	1721.6	1174.8
Sa046	Orphenadrine citrate	2246.8	2092	2290	2472.3	2141.2	2230	2458	1665.6	2350.6	1684.5	2271.1	2336.1
T0400	Oxazepam	2468.2	2085	1522.6	2141.8	1664.9	2192.5	1984.6	2168.5	2385.3	1136	2369.4	2118.4
T0074	Oxprenolol	2143.2	2285.1	2112.3	2223.4	2248	1631.8	2444.3	2411.4	2389.4	2321.6	2471.8	2205.3
T0404	Oxyclozanide	2049.8	2200.9	2025.2	2142.8	2168	2237	1528.1	2076.9	2342.2	1627.8	2376.4	1655.7
T0239	Oxytetracycline dihydrate	1919.8	2247.5	2338.5	2219	1662	1428	2281.6	1520.9	2233.4	1132.6	2439	2300.7
T0234	Oxytetracycline hydrochloride	2254.6	1905	1653.2	2339.8	1663.3	2292.2	1414.9	1520.2	1719.2	2431.6	2100.1	1465.7
T0420	Palmitic acid	2313.2	2352.1	1731.2	1764.2	2385.5	2487	2441.7	1215.1	1392.5	2300.6	2282.8	2369.4
T0475	Pamidronate disodium	2241.6	2299.8	2265.2	2283.5	2318.8	2443.7	2101.3	1952.3	1718.2	1668.3	2139.4	2424.9
T0066	Papaverine hydrochloride	2268.7	2240.6	2390.9	2428.8	1645.2	2406.6	1674.1	1723.1	2153.5	2373.6	1694.7	1742.5
Sa049	Penicillamine	2287	2397.1	1693.6	1739.2	2308	2446.5	2243.1	1184.8	1675.4	2487.3	2268.4	2024
T0227	Pentobarbitone	2296.6	2260	2157	2453.8	1688.2	1724.9	2477.4	2351.4	1520.7	1753.5	2320.8	2217.8
T0478	Perindopril salt	2302.3	1543.1	2266.8	2331.9	2437.2	1750.3	2347.2	2385.6	1680.5	2077.1	2465.9	1181.9
T0217	Phenacetin	2446	2304.4	2265.1	2281.2	2350.2	2149.9	1689.2	2124.8	2178	2480.3	1642.5	1731
T0216	Phenazone	2458.5	2105.3	2140.3	2313	2278.2	2444.4	2370.7	2351.3	1655.4	2427.3	2195.9	2410.9
Sa047	Phenindione	2138.9	2205.3	2452.9	2156.1	2483.2	2411.7	2361	2183.6	1672.3	2469.5	1937.9	2290
T0417	Phenobarbitone	2464.2	2140	1901.6	2303.3	2418.6	2268.3	2165.7	2347.5	2443	2186.6	1414.2	1683.2
T0082	Phenobarbitone	2457.5	2302.3	2265.4	2137.4	2412.5	2360.1	2209.8	2160.4	2438.3	2324.1	2181.5	1662
T0476	2-Phenylbutyric acid	2164.8	2143	2304.1	2471.6	1670.7	1141.7	2186.4	1739.3	2382.9	1697.3	2456.6	2402.7
T0072	Phenylephrine	2476.1	2257.2	2234.6	1670.2	2057.4	2384.7	2313	2143.2	2092	2422.1	1139	1513.1
T0479	Phenylmercuric nitrate	2158.2	2458.6	2476.2	2427.9	2204.3	2182.9	1674	2232.1	2404.1	1141.9	2380.9	1637.7
T0218	Phenylpropanolamine	2466.2	2259.2	2140.8	2298.4	2390.2	2162.9	1685.2	2111.9	2450	2317.6	1140.2	1726.6

File Id	Substance	Peak 1	Peak 2	Peak 3	Peak 4	Peak 5	Peak 6	Peak 7	Peak 8	Peak 9	Peak 10	Peak 11	Peak 12
T0086	Phenytoin	2464.2	2139	2190	1667.2	2346.9	2417.5	2372.7	2164.1	1137.5	2091.3	1515.9	1681.5
Sa003	Phthalic acid	2405.3	1652.6	2138.1	1125.4	2201.7	2447.2	2422.2	2348.9	2180.9	1620	2296	2366.5
Sa050	Phthalyl sulphatiazole	1917.4	2455.3	2415.2	2393.7	1634.5	1659.6	2142.4	2190.8	2323.8	2437.4	1673.8	2486.8
Sa051	Pindolol	1545	2109.9	2317.4	2093.9	2278.5	2206.1	2248.6	2336.2	1623.2	2232.5	2156.7	2419.9
T0418	Pipamazine	2432.7	2350	1666.2	2152.6	1961.5	2002.9	2207.3	2284.4	2385.6	2217.2	1747.9	2074.4
T0063	Piperazine	2248.2	1666.7	2295.8	2352.9	2380	1514.8	1172.7	2325.9	1940.7	2467.6	2308	2410.4
T0416	Pirbuterol hydrochloride	2257.6	2296.9	2391.5	2308	2151.5	1707.8	1181.2	1677.8	2439.6	1587	2368.8	1688.6
T0423	Piretanid	2467	2025.5	2266.4	2236.6	1641.5	1515.8	2141.9	1672.4	2076.1	2175.6	2334.8	2320
T0414	Piroxicam	2485.4	2056.5	2199.4	2258.8	2439	1665.2	2421.1	1529.1	2031.3	2374.1	2234	2303.5
T0424	Polymixin B sulphate	2303.7	2344.3	1932.1	2165.6	1939.7	2275.8	2463	1520.1	2056.6	2473	2183.6	1693.5
T0432	Potassium bicarbonate	1879.9	1525.3	1663.4	2390.4	2386.4	2266.6	1617.8	2380.5	2269.1	1694.6	2276.3	2324.4
Sa004	Potassium hydrogen phthalate	2412.7	2145.3	1659.6	2371.5	1135.6	2437.1	1673.8	2164.8	1630.8	2217.3	2486.7	1720.8
T0427	Potassium hydroxy quinoline	1666.9	2144.2	1134.9	1642.1	2288.4	2437.5	2327.9	2374	2204.7	2233.6	2354.6	2450.1
T0419	Potassium molybdate	1961.9	1520.2	1523	2322.5	2359.6	2020.7	1477.7	1714.2	2338.9	1677	1927.5	2387.7
Sa010	Potassium sorbate	2331	2461.6	2138.2	1710	1749.8	2369.9	1778.1	2398.9	2185.8	2443.1	2409.4	2165
Sa052	Primaquin phosphate	2290.8	2240.7	1656.4	2363.3	2388.8	2342.4	1716.1	2131.1	2457.2	1520.4	1123.6	2328.8
Sa048	Primidone	2467.5	2292.2	2146.2	2168.8	2259.4	2315.3	2230.7	2191.6	2271.3	2410.1	1668.2	1676.4
T0219	Procainamide	1988.7	2259.8	1677.6	2470.1	1480.7	2170.4	1685.7	1547.9	2307.4	2071.9	2063	2196.3
Rw005	Procaine hydrochloride	2254.7	2007	2461.7	2302.1	2161.1	2135.4	2333	2078.6	1679	2391.8	2371.7	1727.9
Rw011	Procaine penicillin	2303.8	2465.8	1942.2	2187.2	2062.9	2139.4	2164.2	1726.2	2255.6	2393.8	1978.8	2268.4
T0422	Prochlorperazine	2251.4	2432.2	1681.2	1666.7	2159.5	2110.7	2321.5	2346.5	1140.2	2375.6	2278.2	2385.7
T0474	Prochlorperazine methyl sulphonate	2263.6	2430.6	2315.8	2469.8	1677.1	1662.9	1723.4	2162.1	2239.9	2148.2	2379.9	2072.8
T0080	Progesterone	2131.4	2265.3	2351.6	2442.2	1692	2164.4	2389.1	2193.7	2295.1	1710.5	1984.2	1740.2
T0071	Propanolol	2294.6	2265.4	2403.3	2249.2	2350.8	2318.9	1669.8	2133.3	2151.1	1682.7	1514.6	2177
T0395	Propyl paraben	2269.1	2451.9	2298	2415	2485.1	2340.6	2153.3	2214.6	2388.8	1693.3	1655.6	1135.2
T0415	Proquamazine fumarate	2261.3	2433.8	2154.6	1675.5	2382.2	2475.3	2336.2	2374.6	2302.1	2237.4	1907	1725.9
T0421	Protease	2309.7	2349.4	2465.8	1728.9	1693.7	1763.1	1519	2165.9	2056.7	1188.3	2259.9	1212.2
T0481	2,5-Pyridine dicarboxylic acid monohydrate 98%	2437.8	2139	2358.9	1949.8	2381.8	2476.5	2414.4	2167.2	2118	1646.6	2189.6	2269.6

File Id	Substance	Peak 1	Peak 2	Peak 3	Peak 4	Peak 5	Peak 6	Peak 7	Peak 8	Peak 9	Peak 10	Peak 11	Peak 12
T0482	2,6-Pyridine dicarboxylic acid 99%	1647.4	2154.8	2444.3	2335.4	1114.8	2134.8	2115.6	2216.8	2390.7	2290.9	2271.9	2191.8
T0425	Pyridoxine hydrochloride	2310.6	2170.6	1516.3	1686.5	2098.9	1642.9	2406.7	1725.5	2195.4	2116.2	2277.1	2009.8
T0233	Quinalbarbitone	2233.1	1633.8	2118.2	2009.7	2286	2153.1	2349.4	2305.3	2265.4	2063.9	2420.2	2459
T0064	Quinidine	1893.5	2230.6	2119.8	2258.5	1664.1	2277.3	1636.3	2454.1	1914	1514.9	2339.6	1700.5
T0060	Quinine	2231.1	2286.7	2344.2	2121.2	1636.3	2162.8	2266.8	2138.9	1516.2	1750.6	1698.3	2377.3
T0503	Quinine sulphate	2272.2	2221.8	2115.7	2252.8	1633.5	2330.9	2397.3	1661.3	2351.8	2156.8	2312.4	1693.5
T0426	Ranitidine diamine	2380.1	1634.9	2294.6	2250.5	2440.2	1749.6	2329	2403.5	1729.9	1698.5	2342.4	2223.3
T0408	Reserpine	1483	2342	2233.7	2274.4	2435.9	2134.1	2158.6	2308.4	1646	2250.6	2384.4	2362.4
T0428	Rifampicin	1904.9	2259.6	1684.7	2391.2	1666	2274.9	1693.2	2299.3	2361.5	2452.7	1730.4	2121.9
T0440	Rifampicin quinone	1906.2	2259.1	1682.6	2299.9	1695.3	2394	2453.2	1730.4	1771.9	1419.6	1517.3	2360.1
T0449	Salbutamol sulphate	1434.1	2292.3	2257.5	2020.1	2042.8	1673.6	2111.8	2152.1	2398.6	2136	2384.5	1178.2
T0049	Salicylic acid	2400.9	1655.6	2188.1	2247.1	2153.1	2478.9	2069.8	2439.1	1129.7	2318.1	2107.7	2281.5
T0231	Secbutobarbitone	2303.8	2274	2263.4	1688.3	2344.5	1520.4	2417.4	2442.9	1900.5	1185.3	1756.8	1728.2
Sa006	Sodium alginate	1920.9	2487.3	1440.3	2340.8	2322.9	2359.8	1476.8	2474.1	1519.1	2460.5	2302	1201.6
Sa008	Sodium ascorbate	2266.8	2438.7	2316	2363.9	1743.5	2469.6	2410.1	1697	2184.3	2157.5	2337	2393.4
T0442	Sodium benzoate	2461.7	2138.7	2379.6	2449.4	1659.3	1911.6	2288.7	2422.4	2193.7	2216.9	2344.8	1624
T0445	Sodium bicarbonate	2322.5	1522	1861.1	2218	1719	1878.1	1677.7	2339.4	2487.2	2032	1479.6	2413.6
T0459	Sodium chloride	1909.8	1522.7	1925.5	2322.7	2360	1947.1	1416.1	1478.8	1715.2	2338.4	1676.5	1825.7
T0460	Sodium chromoglycate	2150.5	2321.3	2277.7	1656.7	2263.3	2409.4	2447	1927.5	1675.7	2369.3	1910.9	2474.4
T0461	Sodium valproate	2308.8	2396.7	2446.5	1718.9	2272	1692.1	2459.2	2352.9	2285.3	2337.3	1702.4	1190.7
Sa009	Sorbic acid	2316.1	1701.5	2371.3	1189.4	1738	2433.2	1156.5	2279.4	2394.6	2479.1	1687.8	1770.9
T0429	Starch	1926.7	2284.3	1433.8	2320.9	2485.3	1698.6	2103.4	1200.2	2084.9	1778.3	2263.6	1474.7
T0455	Sulfanilic acid	2456	2336.8	1664.7	2137.1	2161.4	2392.7	2369.1	2287.8	2228.4	1132.6	2407.1	2187.8
T0062	Sulphacetamide	2203.5	1986.4	1960.9	2036	1452	1517.6	2172.3	2079.3	2464.6	2250.3	2122.6	1491.9
T0441	Sulphathiazole	2459.7	2394.3	2041.8	1632.7	2139.4	2105.4	2311.2	1667.9	2340.9	2361.9	2163.7	2196.6
T0456	Sulpiride	2027.7	2162.9	2266.5	1511.7	2143.8	2241.5	2090.5	2402.1	2197.3	2472.3	1720.1	1667
T0438	Suxamethonium chloride	2259	1974.5	2230.9	1950.4	2421.7	2316.6	1451.7	2341.1	2402.7	2354	1668.6	2295.7
T0228	Talbutal	2225.6	2113.1	2450	2296	1629.6	2263.9	2155.9	2277.4	2350.3	1520	2168.6	2421.4
T0443	Tamoxifen citrate	2172	2256.1	2306.1	2243.8	2469.7	2454.5	2274.1	2147.7	1671.2	2224.7	2378.9	2194

File Id	Substance	Peak 1	Peak 2	Peak 3	Peak 4	Peak 5	Peak 6	Peak 7	Peak 8	Peak 9	Peak 10	Peak 11	Peak 12
T0485	Tetracycline base	2234.4	1927.2	2250.7	2338.2	1970.3	2177	1482.6	2391.2	2002.3	1447.1	1681	2358.8
T0236	Tetracycline hydrochloride	1925.5	2294.8	2439.1	2243.3	1659.2	2347.5	2380.6	2108.4	2261.5	1426	2403.4	2127.9
T0044	Theophylline	2475.8	2323.7	2293.7	1662.6	2271.2	2429.1	2254.2	2134.1	1681.2	1624.9	2189.5	2091
T0433	Thiazesin	2350.4	2265.5	2474.9	2249.4	1672.9	2163.8	2434.8	2143.7	2185.2	2452.9	1140	1748.9
Rw004	Thiopental sodium salt/sodium carbonate mixture	1504.3	2303.6	2269.8	1689.4	1186.9	1762.4	2099.8	2346.3	2340.6	2454.2	1717.3	1900.6
Sa054	Thioridazine hydrochloride	2328.2	2436.2	2283.1	2272.8	2165.7	1914.1	2253.1	1694.2	2401.7	1670	2367.4	2150.8
T0430	Tinin P	2413.8	2385.5	2458.5	2153.5	1658.7	2435.6	2133.8	1135.2	2266.2	1701.5	1675.1	2315.3
T0501	Titanium dioxide	1879.4	1520.9	1381.6	2323.2	2359.5	2339.8	1714.6	1719	1772.5	1676.3	1479.1	2390.7
T0439	Tolbutamide	2048.2	2456.3	1529.4	1670.3	2198.4	2143.7	2165.2	2271.6	1565.2	1699.1	2339.2	2408.9
Sa053	Tolnaftate	2471.9	1675.6	2167.1	2148.8	2449.3	1139.6	2391.9	2314.1	2423.1	2283.8	2328.5	1634.5
T0317	Trans-2(4-(1,2-Diphenyl-1-butenyl)phenoxy-N,N dimethyl ethylamine	2172	1672.8	2149.3	2476.6	2466.2	2309.2	2196	2287.2	1141.4	1735.5	2223.8	1770.7
T0453	Triacontane	2313.3	2352.5	1730.8	1764.5	2442.1	2486.5	2385.4	1392.6	1215.1	1691.5	2273.6	2012.8
T0434	Trifluoperazine	1921.9	2272.4	2426.9	2012	2335.1	1659.9	1430.9	2153.4	2397.5	2305.9	1481.6	2487.5
T0431	Trifluoperazine sulphoxide	2254.1	1679.7	2272.9	2108.9	1139.8	2426.8	2142.9	2319.8	1894.3	1524	2381.6	2171.2
T0447	Triflupromazine hydrochloride	2355.6	2435.3	2266.3	2250.1	2385.4	2137	1662.7	2452.4	2161.8	2324.8	2226.7	1727.9
T0451	Triheptadecanoin	2312.4	2351.5	1730.4	2252.4	1764.2	2442	2485.9	2385.1	1392	1214.8	2273.2	2144
T0458	Trimethoprim	1957.8	1466.8	2215.7	2245	2071.5	2111.3	2086.9	1987.9	2383.5	2166.4	2018.9	2272.7
T0436	Trimipramine base	2146.3	2175.3	2299.9	2328.4	1685.2	2419	2271.9	2288.3	1729	2384.3	2257.2	2360.9
T0437	Triprolidine hydrochloride	1969.2	2467.4	1674.1	1451.9	2444.6	2173.1	2250.9	2283.6	1950.9	1988.4	2310.7	1723.1
T0229	Vinbarbitone	2269.6	2304.7	2466.6	2364.1	1686.3	2435.7	2391	2341.1	1520.5	2138	2422.2	1726.5
T0446	Voltarol	1917.4	1898	1666.5	2407.4	2287.1	1412.5	2324.8	1393.7	2370.6	2464.4	1134.6	2205.7
T0070	Warfarin sodium	1904.2	2142	2469.9	2168	1665.1	1679.5	2401.7	2189.9	1517.7	1403	2319.9	1139.4
T0457	Zinc sulphate monohydrate	2422.1	2112.5	2108.4	1522.6	1979.8	2322.5	2175.7	2077.2	2081.1	2133.5	1861.5	1718.8

Appendix 3

Table 64 Test samples that were run as unknowns to externally validate the second derivative spectral database in Chapter 6 (Identification of drug substances by peak matching)

CODE	TEST SAMPLE
U001	Anhydrotetracyline base
U002	Procaine hydrochloride
U003	Tetracycline hydrochloride
U004	Nitrofurantoin
U005	Thioridazine hydrochloride
U006	Lactose monohydrate
U007	L-Alanine
U008	Penicillamine
U009	Sugars (mixed)
U010	2-Thiobarbituric acid
U011	Potassium hydrogen phthalate
U012	8-Hydroxyquinoline
U013	Papaverine hydrochloride
U014	Avicel
U015	L-Glutamine
U016	Orphenadrine citrate
U017	Phenylmercuric nitrate
U018	Metronidazole
U019	Chlortetracycline hydrochloride and Tetracycline hydrochloride
U020	Barbitone sodium
U021	Naproxen
U022	2-Phenylbutyric acid
U023	Salbutamol sulphate
U024	Aneurine hydrochloride
U025	Salicylic acid
U026	Chloroquine phosphate
U027	Talcum powder
U028	Edetate disodium
U029	Sodium tartrate dihydrate
U030	4-Acetamidophenol
U031	Demeclocycline hydrochloride
U032	Deptropine citrate
U033	Avicel PH102
U034	Potassium chloride

Appendix 4

Table 65 Description of tablets, manufacturers and batch references for all products used within Chapter 7 (The identification of actives in whole tablets)

*Batch Reference Ids correspond to Medicines Testing Laboratory reference codes

Tablet desciption	Manufacturer	Batch Reference Id*
Allopurinol 100mg tablets BP	APS/ Berk	31092
Allopurinol 100mg tablets BP	APS/ Berk	31303
Allopurinol 100mg tablets BP	APS/ Berk	31213
Allopurinol 100mg tablets BP	APS/ Berk	31501
Allopurinol 100mg tablets BP	APS/ Berk	31857
Allopurinol 100mg tablets BP	Ashbourne	32170
Allopurinol 100mg tablets BP	Ashbourne	31396
Allopurinol 100mg tablets BP	Ashbourne	31753
Allopurinol 100mg tablets BP	Ashbourne	31756
Allopurinol 100mg tablets BP	Ashbourne	31882
Allopurinol 100mg tablets BP	Cox	31336
Allopurinol 100mg tablets BP	Cox	31532
Allopurinol 100mg tablets BP	Cox	31161
Allopurinol 100mg tablets BP	Cox	31297
Allopurinol 100mg tablets BP	Cox	31298
Allopurinol 100mg tablets BP	CP	31755
Allopurinol 100mg tablets BP	CP	31757
Allopurinol 100mg tablets BP	CP	32750
Allopurinol 100mg tablets BP	CP	32759
Allopurinol 100mg tablets BP	CP	31746
Allopurinol 100mg tablets BP	DDSA	31335
Allopurinol 100mg tablets BP	DDSA	31337
Allopurinol 100mg tablets BP	DDSA	31395
Allopurinol 100mg tablets BP	DDSA	31394
Allopurinol 100mg tablets BP	Norton	31012
Allopurinol 100mg tablets BP	Norton	31184
Allopurinol 100mg tablets BP	Norton	31066
Allopurinol 100mg tablets BP	Norton	31304
Allopurinol 100mg tablets BP	Norton	31060
Allopurinol 100mg	Pharmvit	31534
Allopurinol 100mg tablets BP	Pharmvit	31758
Allopurinol 100mg tablets BP	Pharmvit	31861
Allopurinol 100mg tablets BP	Pharmvit	31879
Allopurinol 100mg tablets BP	Pharmvit	32766
Allopurinol 100mg tablets BP	Rima	31073
Allopurinol 100mg tablets BP	Rima	31095
Allopurinol 100mg tablets BP	Rima	31334
Allopurinol 100mg tablets BP	Rima	31067
Allopurinol 100mg tablets BP	Rima	31187
Zyloric 100mg tablets	Wellcome	31089
Zyloric 100mg tablets	Wellcome	31186
Zyloric 100mg tablets	Wellcome	31207
Zyloric 100mg tablets	Wellcome	30999
Zyloric 100mg tablets	Wellcome	31090
Allopurinol 100mg	Generics	31162

Tablet desciption	Manufacturer	Batch Reference Id*
Allopurinol 100mg	Generics	31469
Allopurinol 100mg	Generics	31529
Allopurinol 100mg	Generics	31069
Allopurinol 100mg	Generics	31183
Allopurinol 100mg	Generics	
Allopurinol 300mg tablets BP	Regent	31185
Allopurinol 300mg tablets BP	Regent	31295
Allopurinol 300mg tablets BP	Regent	31302
Allopurinol 300mg tablets BP	Regent	31338
Allopurinol 300mg tablets BP	Regent	31530
Allopurinol 300mg tablets BP	APS/ Berk	31397
Allopurinol 300mg tablets BP	APS/ Berk	32171
Allopurinol 300mg tablets BP	APS/ Berk	31398
Allopurinol 300mg tablets BP	APS/ Berk	31773
Allopurinol 300mg tablets BP	APS/ Berk	31767
Allopurinol 300mg tablets BP	Ashbourne	32176
Allopurinol 300mg tablets BP	Ashbourne	31407
Allopurinol 300mg tablets BP	Ashbourne	31767
Allopurinol 300mg tablets BP	Ashbourne	31883
Cimetidine 200mg Tablets BP	Clonmel	31525
Cimetidine 200mg Tablets BP	Clonmel	31287
Cimetidine 200mg Tablets BP	Clonmel	31412
Cimetidine 200mg Tablets BP	Clonmel	32739
Cimetidine 200mg Tablets BP	Clonmel	31174

References

[1] Jouan-Rimbaud D., Walczak B., Massart D. L., Last I. R., Prebble K. A. "Comparison of multivariate calibration methods based on latent vector and methods based on wavelength selection for the analysis of near-infrared spectroscopic data" Analytica Chimica Acta (1995) 304: 285-295

[2] Workman J. "Near infrared spectroscopic measurement: special challenges related to quantitative analysis" NIR News (1996) 7 (6) 5-6

[3] Van Zyl E. F., Louw M. "The differentiation of illicit methaqualone tablet formulations using principal component and soft independant modelling of class analogy analysis of their near-infrared reflectance spectra" Journal of Forensic Sciences. (1995) 40 (6) 1072-1076

[4] Morrisseau K.M., Rhodes T.C., "Pharmaceutics of near Infrared spectroscopy" Drug Development and Industrial Pharmacy (1995) 21 (9) 1071-1091

[5] British Pharmacopoiea 2002, Appendix IIIA, A128, HMSO, Norwich, UK

[6] United States Pharmacopeia and National Formulary 2003. 26/NF21 Supplement 1, Rockville, United States Pharmacopoieal Convention.

[7] European Pharmacopoeia 2002, 4th edition, Supplement 4.5, EDQM, Strasbourg Cedex, France

[8] The International Pharmacopoeia 2001, Volume 5, World Health Organization, Geneva, XVIII

[9] Stead A. H., Gill R., Wright T., Gibbs J. P., Moffat A. C. "Standardized thin-layer chromatographic systems for the identification of drugs and poisons. Analyst (1982) 107: 1106-1168

[10] Moffat A. C., Franke J. P., Stead A. H., Gill R., Finkle B. S., Moller M. R., Wunsch F., de Zeeuw R. A. "Thin-layer chromatographic rf values of toxicologically relevant substances on standardized systems." VCH, Weinheim (1987), pp 14-17

[11] Kubelka P., Munk F., "Ein beitrag zur optik der farbanstriche" Zeitshrift für technische Physik (1931) 12, 593-604

[12] Ellis J., Bath J., "Modifications in NIR absorption spectra of protein and of light and heavy water molecules when water is bound to gelatin" Journal of Chemical Physics (1938) 6, 723-729

[13] Kaye W., "NIRS, a review. 1. Spectral identification and analytical applications" Spectrochimica Acta (1954) 6: 257-287

[14] Ben-Gera, I., Norris, K.H. "Determination of moisture in soybeans by direct spectrophotometry" Israel Journal of Agricultural Research (1968) 18(3) 125-132

[15] Ben-Gera, I., Norris, K.H. "Influence of fat concentration on the absorption spectrum of milk in the NIR region" Israel Journal of Agricultural Research (1968) 18(3), 117-124

[16]Osbourne B.G. in "Near infrared spectroscopy in food analysis," (1986) Longman Scientific and Technical, UK, p.2

[17] Ghosh S. in "Handbook of near-infrared analysis"(1992) ed by Burns and Ciurczak, Marcel Dekker Inc., New York, USA, p.495

[18] Ciurzcak E.W. in "Handbook of near-infrared analysis" (1992) ed by Burns and Ciurczak, Marcel Dekker Inc., New York, USA, p.13

[19] Burns D. A., Margoshes M. in "Handbook of near-infrared analysis"(1992) ed by Burns and Ciurczak, Marcel Dekker Inc., New York, USA, p.14-15

[20] Moffat A. C., Trafford A. D., Jee R. D., Graham P. "Meeting the International Conference on harmonisation's guidelines on validation of analytical procedures: Quantification as exemplified by a near-infrared reflectance assay of paracetamol intact tablets" Analyst (2000) 125 (7) 1341-1351

[21] Laasonen M., Harmia-Pulkkinen T., Simard C., Rasanen M., Vuorela H. "Development and validation of a near infrared method for the quantitation of caffeine in intact single tablets" Analytical Chemistry (2003) 75 (4) 754-760

[22] "Piperazine in drugs", Official Methods of Analysis (AOAC) (1990) Vol 1, 533

[23] Plugge W., Van der Vlies C. "NIRS as an alternative to assess compliance of ampicillin trihydrate with compendial specifications" Journal of Pharmaceutical & Biomedical Analysis (1993) 11 (6) 435-442

[24] Blanco M., Coello J., Eustaquio A., Iturriaga H., Maspoch S. "Development and validation of methods for the determination of miokamycin in various pharmaceutical preparations by use of near infrared reflectance spectroscopy" Analyst (1999) 124 (7) 1089-1092

[25] European Pharmacopoeia, (1995) Chapter 2.2.40, 3rd ed p.43

[26] Dempster, M.A., MacDonald, B.F., Gemperline, P.J., Boyer, N.R. "A NIR reflectance analysis method for the non-invasive identification of film-coated and non-film-coated, blister-packed tablets" Analytica Chimica Acta (1995) 310(10) 43-51

[27] EMEA, The European Agency for the Evaluation of Medicinal Products, "Note for guidance on the use of near infrared spectroscopy by the

pharmaceutical industry and the data requirements for new submissions and variations" February, 2003

[28] Broad N. W., Jee R. D., Moffat A. C., Smith M. R. "Application of transmission near-infrared spectroscopy to uniformity of content testing of intact steroid tablets" Analyst (2001) 126 (12) 2207-2211

[29] Blanco M., Coello J., Iturriaga H., Maspoch S., Pou N. "Influence of the procedure used to prepare the calibration sample set on the performance of near infrared spectroscopy in quantitative pharmaceutical analysis" Analyst (2001) 126 (6) 1129-1134

[30] Eustaquio A., Graham P., Jee R. D., Moffat A.C., Trafford A. D. "Quantfication of paracetamol in intact tablets using near-infrared transmittance spectroscopy" Analyst (1998) 123 (11) 2303-2306

[31] Corti, P., Dreassi, E., Corbini, G., Lonardi, S., Gravina, S. "The application of NIRS to pharmaceutical control 1. Preliminary investigation of the uniformity of tablets content" Analusis (1990) 18: 112-116

[32] Plugge W., Van der Vlies C., "The use of NIRS in the quality control laboratory of the pharmaceutical industry" Journal of Pharmaceutical & Biomedical Analysis (1992) 10, 10-12: 797-783

[33] Corti, P., Dreassi, E., Murratzu, C., Corbini, G., Ballerini, L., Gravina, S. "Application of NIRS to the quality control of pharmaceuticals. Ketoprofen assay in different pharmaceutical formulae" Pharmaceutical Acta Helvetica (1989) 64, 5-6, 140-145

[34] Smith M. R., Jee R. D., Moffat A. C. "The transfer between instruments of a reflectance near-infrared assay for paracetamol in intact tablets" Analyst (2002) 127 (12) 1682-1692

[35] Rose J.R., Prusik T., Mardekian J. "NIR multicomponent analysis of parenteral products using the InfraAlyzer 400" Journal of Parenteral Science & Technology (1982) 36(2): 71-78

[36] Shenk J.S., Landa I., Hoover M.R., Westerhaus M.O. "Grain analysis" Crop Science (1981) 21 (3), 355-358

[37] Workman J. J., "Review of process and non-invasive near-infrared and infrared spectroscopy: 1993-1999" Applied Spectroscopy Reviews (1999) 34 (1-2) 1-89

[38] Mark H. L., Tunnell D. "Qualitative near-infrared reflectance analysis using Mahalanobis Distances" Analytical Chemistry (1985) 57: 1449-1456

[39] Ciurczak E.W. "Qualitative analysis of pharmaceutical raw materials via NIRA spectroscopy" in proceedings of the 7th Annual Symposium on NIR, Technicon, New York

[40] Gemperline P. J., Webber L. D., Cox F. O., "Raw materials testing using Soft Independant Modelling of Class Analogy analysis of near-infrared reflectance spectra" Analytical Chemistry (1989) 61: 138-144

[41] Shah N. K., Gemperline P. J. "Combination of Mahalanobis Distance and Residual Variance pattern recognition techniques for classification of near infrared reflectance spectra" Analytical Chemistry (1990) 62: 465-470

[42] Salamin P. A., Cornelis Y., Bartels H. "Identification of chemical substances by their NIR spectra" Chemometrics and Intelligent Laboratory Systems (1988) 3: 329-333

[43] Kohn W. H., Jeger A. N. "Identification of drugs by their near infrared spectra"

Journal Forensic Scicence (1992) 37 (1): 35-41

[44] Gerhausser C. I., Kovar K. A. "Strategies for constructing near infrared spectral libraries for the identification of drug substances" Applied Spectroscopy (1997) 51 (10) 1504-1510

[45] Blanco M., Romero M. A. "Near-infrared libraries in the pharmaceutical industry: a solution for identity confirmation" Analyst (2001) 126 (12) 2212-2217

[46] Corti, P., Savini, L., Dreassi, E., Ceramelli, G., Montecchi, L., Lonardi, S. "Application of NIRS to the control of pharmaceutical identification and assay of several primary materials" Pharmaceutical Acta Helvetica (1992) 67(2) 57-61

[47] Corti, P., Dreassi, E., Ceramelli, G., Lonardi, S., Viviani, R., Gravina, S. "NIRS applied to pharmaceutical QC Identification and assay of cephalosporins" Analusis (1991) 19 198-204

[48] Salamin, P.A., Cornelis, Y., Bartels, H. "Identification of chemical substances by their NIR spectra" Chemometrics and Intelligent Laboratory Systems 3(1988) 329-333

[49] Næs, T., Irgens, C., Martens, H. "Comparison of linear statistical methods for calibration of NIR instruments" Applied Statistics (1986) 35, 195-206

[50] Buchanan, B.R., Ciurczak, E.W., Grunke, A.Q., Honigs, D.E. "Determination of enantiomeric purity of valine via NIRS" Spectroscopy (1986) 3 (5), 54-56

[51] Ciurczak, E.W., Maldacker, T.A. "Identification of actives in multicomponent pharmaceutical dosage forms using NIRS" Spectroscopy (1986) 1(1), 36-39

[52] Honigs, D.E., Hieftje, G.M., Hirschfeld, T.B. "A new method for obtaining individual component spectra from those of complex mixtures" Applied Spectroscopy (1984) 38(3), 317-322

[53] Whitfield, R.G. "Near infrared reflectance analysis of pharmaceutical products" Pharmaceutical Manufacturing (1986) April, 31-40

[54] Lodder, R.A., Selby, M., Hieftje, G.M. "Detection of capsule tampering by NIR reflectance analysis" Analytical Chemistry (1987) 59(15),1921-1930

[55] Lodder, R.A., Hieftje, G.M. "Analysis of intact tablets by NIRS" Applied Spectroscopy (1988) 42(4), 556-558

[56] Lodder, R.A., Hieftje, G.M. "Quantile BEAST attacks the false sample problem in NIR analysis" Applied Spectroscopy (1988) 42(8), 1351-1365

[57] Lodder, R.A., Hieftje, G.M. "Detection of sub-populations in NIR reflectance analysis" Applied Spectroscopy (1988) 42(8), 1500-1512

[58] Drennen, J.K., Lodder, R.A. "Non-destructive NIR analysis of intact tablets for determination of degradation products" Journal of Pharmaceutical Science (1990) 79(7), 622-627

[59] Dubois, P., Martinez, J-R., Levillain, P. "Determination of five components in a pharmaceutical formulation using NIRS" Analyst (1987) 112(12), 1675-1679

[60] Chasseur, J.C. "On-line assay of cimetidine granules by NIRS" Chimicaoggi (1987) June, 21-24

[61] Jensen, R., Peuchant, E., Castagne, I., Boirac, A.M., Roux, G. "One step quantification of active ingredient in pharmaceutical tablets using NIRS" Spectroscopy International Journal (1988) 6, 63-72

[62] Molt, K., Egelkraut, M. "Quantitative multi-component analysis of powder mixtures by NIRS" GIT Fachz. Lab. (1988) 32(12), 1311-1313

[63] Corti, P., Dreassi, E., Lonardi, S. "NIR reflectance analysis : features and applications in pharmaceutical and biomedical analysis" Il Farmaco, (1993) 48, 3-20

[64] Ryan, J.A., Compton, S.V., Brooks, M.A., Compton, D.A.C. "Rapid verification of identity and content of drug formulations using mid infra red spectroscopy" Journal of Pharmaceutical & Biomedical Analysis (1991) 9(4) 303-310

[65] Dempster, M.A., Jones, J.A., Last, I.R., MacDonald, B., Prebble, K.A. "NIR methods for the identification of tablets in clinical trial supplies" Journal of Pharmaceutical & Biomedical Analysis (1993) 11(11/12) 1087-1092

[66] Broad N.W., Jee R. D., Moffat A. C., Eaves M. J., Mann W. C., Dziki W. "Non-invasive determination of of ethanol, propylene glycol and water in a multi-component pharmaceutical oral liquid by direct measurement through amber plastic bottles using Fourier transform near-infrared spectroscopy" Analyst (2000) 125 (11) 2054-2058

[67] MacDonald, B.F., Prebble, K.A. "Some applications of NIR reflectance analysis in the pharmaceutical industry" Journal of Pharmaceutical & Biomedical Analysis (1993) 11(11/12) 1077-1085

[68] Blanco, M., Coello, J., Iturriaga, H., Maspoch, S., de la Pezuela, C. "Determination of ascorbic acid in pharmaceutical preparations by NIRS" Talanta (1993) 40(11), 1671-1676

[69] Aldridge, P.K., Mushinsky, R.F., Andino, M.M., Evans, C.L. "Identification of tablet formulations inside blister packages by NIRS" Applied Spectroscopy (1994) 48(10) 1272-1276

[70] Dempster, M.A., MacDonald, B.F., Gemperline, P.J., Boyer, N.R. "A NIR reflectance analysis method for the non-invasive identification of film-coated and non-film-coated, blister-packed tablets" Analytica Chimica Acta (1995) 310(10) 43-51

[71] Van Zyl, E.F., Louw, M. "The differentiation of illicit methaqualone tablet using principal component and soft independent modeling of analogy analysis of their NIR reflectance spectra" Journal of Forensic Sciences (1995) 40(6) 1072-1076

[72] Scafi S. H. F., Pasquini C. "Identification of counterfeit drugs using near-infrared spectroscopy" Analyst (2001) 126 (12) 2218-2224

[73] Han, S.M., Faulkner, P.G. "Determination of SB216469-S during tablet production using NIR reflectance spectroscopy" Journal of Pharmaceutical & Biomedical Analysis (1996) 14(12) 1681-1689

[74] Gottfries, J., Depui, H., Fransson, M., Jongeneelen, M., Josefson, M., Langkilde, F.W., Witte, D.T. "Vibrational spectrometry for the assessment of active substance in metoprolol tablets: a comparison between transmission and diffuse reflectance NIR spectrometry" Journal of Pharmaceutical & Biomedical Analysis (1996) 14, 1495-1505

[75] Trafford A D., Jee R. D., Moffat A. C., Graham P. "A rapid quantitative assay of intact paracetamol tablets by reflectance near infrared spectroscopy" Analyst (1999) 124:163-167

[76] Eustaquio A., Blanco M., Jee R.D., Moffat A.C. "Determination of paracetamol in intact tablets by use of near infrared transmittance spectroscopy" Analytica Chimica Acta (1999) 383: 283-290

[77] Butler, L.A. "The history and background of NIR" Cereal Foods World (1983) 28(4) 238

[78] Hruschka, W.R., "Spectral reconstruction" in "Handbook of Near-infrared Analysis" ed by Burns and Ciurczak, Marcel Dekker Inc., New York, USA, p 365

[79] Barnes R. J., Dhanoa M. S., Lister S. J. "Standard normal variate and de-trending of near-infrared diffuse refelectance spectra" Applied Spectroscopy (1989) 43 (5) 772-777

[80] NIRSystems, Inc. (1995) IQ^2 Version 3.52 Software Manual, Silver Spring, MD

[81] O'Haver T. C. "Derivative and wavelength modulation spectrometry" Analytical Chemistry (1979) 51 (1) 91A

[82] O'Haver T. C., Green G. L., "Numerical error analysis of derivative spectrometry for the quantitative analysis of mixtures" Analytical Chemistry (1976) 48 (2) 312-318

[83] Gill R., Bal T. S., Moffat A.C. "The application of derivative UV-visible spectroscopy in forensic toxicology" Journal of the Forensic Science Society (1982) 22: 165-171

[84] Fell A. F., Scott H. P., Gill R., Moffat A. C. "Computer-aided multichannel detection in high-performance liquid chromatography" Chromatographia (1982) 16: 69-78

[85] Fell A. F., Scott H. P., Gill R., Moffat A. C. "Novel techniques for peak recognition and deconvolution by computer-aided photodiode array detection in high-performance liquid chromatography" Journal of Chromatography (1983) 282: 123-140

[86] Savitsky A., Golay J. E., "Smoothing and differentiation of data by simplified least squares procedures" Analytical Chemistry (1964) 36 (8), 1627-1639

[87] Mark H. L., in "Handbook of near-infrared analysis"(1992) ed by Burns and Ciurczak, Marcel Dekker Inc., New York, USA, p.119-120

[88] Gemperline P. J., Boyer N. R., "Classification of near-infrared spectra using wavelength distances: Comparison to the Mahalanobis Distance and residual variance methods" (1995) 67, 160-166

[89] Mark H. L. in "Handbook of near-infrared analysis"(1992) ed by Burns and Ciurczak, Marcel Dekker Inc., New York, USA, p.107

[90] Van der Vlies C., Kaffka K. J., Plugge W., "Qualifying pharmaceutical substances by fingerprinting with NIR spectroscopy and PQS" Pharmaceutical Technology Europe (1995) Apr, 43-56

[91] Sokal R. R., Rohlf F. J., in 'Biometry', 2nd edition, 1981. Freeman and Company. Chapter 15, 15.7, 594

[92] Williams P. C., Thompson B. N. "Influence of wholemeal granularity on the analysis of HRS wheat for protein and moisture by near-infrared reflectance spectroscopy (NIR)" Cereal Chemistry (1978) 55, 1014-1037

[93] Aulton M. E. "Pharmaceutics: the science of dosage form design" Churchill Livingstone, Edinburgh, 1988

[94] Barth H. G., Sun S., Nickol R. M. "Particle size analysis" Analytical Chemistry (1987) 59, 142

[95] Plugge W., Van der Vlies C. "Near-infrared spectroscopy as a tool to improve quality" Journal of Pharmaceutical and Biomedical Analysis (1996) 14: 891-898

[96] Ciurczak E. W., Torlini R. P., Demkowicz M. P., "Determination of particle size of pharmaceutical raw materials using near-infrared reflectance spectroscopy" Spectroscopy (1986) 1: (7) 36-39

[97] Moffat A.C., "Thin-layer chromatography" in Clarke's Isolation and Identification of Drugs 2nd ed, (1986), The Pharmaceutical Press, England

[98] Smalldon K.W, Moffat A.C., "The calculation of discriminating power for a series of correlated attributes." Journal of Forensic Science (1973) 13, 291-295

[99] Schepers P. G., Franke J. P., de Zeeuw R. A., "Systemic evaluation and substance identification in systemic toxicological analysis by the mean list length approach" Journal of Analytical Toxicology (1983) 7: 272-278

[100] Norris K. H., Williams P. C., "Optimisation of mathematical treatments of raw NIR signal in the measurement of protein in hard red spring wheat. Influence of particle size." Cereal Chemistry (1984) 61, 158-165

[101] O'Neil A. J., Jee R. D., Watt R. A., Moffat A. C. "Particle size analysis of pharmaceutical powders using near infrared spectroscopy" Journal of Pharmacy and Pharmacology (1997) 49 Suppl 4, 19

[102] Ilari J. L., Martens H., Isaksson T. "Determination of particle size in powders by scatter correction in diffuse NIR reflectance" Applied Spectroscopy (1988) 42(5), 722-728

[103] Yoon W., Jee R. D., and Moffat A. C. "Optimisation of sample presentation for the near infrared spectra of pharmaceutical excipients" (1998), Analyst, 123: 1029

[104] Clarke's Isolation and Identification of Drugs 2nd ed, (1986), The Pharmaceutical Press, England

[105] Mark H. L., Tunnell D., "Qualitative near-infrared reflectance analysis using Mahalanobis Distances" Analytical Chemistry (1985) 57: 1449-1456

[106] Cowe I. A., McNicol J. W., "The use of principal components in the analysis of near infrared spectra" Applied Spectroscopy (1985) 39 (2) 257-266

[107] Curry A.S., Read J. F., Brown C., "A simple infrared spectrum retrieval system" Journal of Pharmacy and Pharmacology, (1969) 21: 224-231

[108] Ingle P.H.B., Mathieson D. W., "Use of authentic spectra for identification of drug substances" Pharmaceutical Journal (1976) Jan 24; 73

[109] Kirsch J. D., Drennen J. K., "Near infrared spectroscopic monitoring of the film coated process", Pharmaceutical Research (1996)13 (2) 234-237