# Infrared spectroscopy in clinical chemistry, using chemometric calibration techniques

Volmer, Marcel

# Infrared spectroscopy in clinical chemistry, using chemometric calibration techniques

Volmer, M
Infrared spectroscopy in clinical chemistry,
using chemometric calibration techniques

Proefschrift Groningen. – Met Lit. opg. – Met samenvatting in het Nederlands

ISBN 90-367-1485-0

RIJKSUNIVERSITEIT GRONINGEN

# Infrared spectroscopy in clinical chemistry, using chemometric calibration techniques

Proefschrift

ter verkrijging van het doctoraat in de
Medische Wetenschappen
aan de Rijksuniversiteit Groningen
op gezag van de
Rector Magnificus, dr. D.F.J. Bosscher,
in het openbaar te verdedigen op
dinsdag 11 september 2001
om 13.15 uur

door

Marcel Volmer
geboren op 21 oktober 1952
te Emmen

Promotor : Prof. Dr. S. Poppema

Referenten : Dr. I.P. Kema
Dr. B.G. Wolthers

Beoordelingscommissie : Prof. Dr. F.A.J. Muskiet
Prof. Dr. A.K. Smilde
Prof. Dr. J.L. Willems

Paranimfen: Eddie Ligeon
Jikkemien Volmer

Maar hoe het ook zij, zo al met al
mag een ieder tevreden wezen:
we hebben de schrijver gezien in de hal,
in de zaal, in de aula en overal …
Nu hoeven we ’t niet meer te lezen.

naar A.M.G. Schmidt

# Contents

# Scope of the thesis

The aim of our studies was to further develop the assays of fecal fat and urinary calculi. The development consisted of the investigation of the applicability of new infrared spectroscopic methods for routine use in the clinical laboratory. Because most of these assays made use of authentic sample material, quantification of the analyte concentrations was often hampered, because of the complex sample matrices. Therefore, we also investigated the application of chemometrical methods for quantification of the analyte concentrations from the spectral results. We applied artificial neural networks and partial least-squares regression analysis for both calibration and prediction of the outcome of both kinds of assays. Furthermore, we gave some information about the pathophysiology background of our studies.

# Introduction

# 1. Pathophysiology

## 1.1. Steatorrhea

*Introduction*

Digestion and absorption of nutrients in the gastro-intestinal (GI) tract is a complex process in which a great number of steps are involved. Normally, food is digested, followed by absorption of the nutrients into the bloodstream. The absorption mainly occurs in the small bowel area of the GI tract. In case of impaired absorption, malabsorption occurs, either because a disorder disturbs the digestion of food, or directly disturbs the absorption of nutrients from the small intestine. Malabsorption may occur for many nutrients simultaneously, or for specific nutrients such as carbohydrates, proteins, fats, or micronutrients (e.g. vitamins) separately. Sometimes, secondary nutritional deficiencies develop as a result of primary diseases. For example, malabsorption of fat may lead to impaired absorption of the fat soluble vitamin K, which in turn may lead to hypoprothrombinemia and bleeding disorders (1). Any combination of weight loss, diarrhea and anemia should raise suspicion of malabsorption. Laboratory studies can be useful in the diagnosis of impaired digestion, or absorption. Most of the laboratory tests used in the diagnosis of malabsorption syndromes will indicate the presence of an abnormal absorptive, or digestive process, but only a few tests can lead to a more specific diagnosis. Therefore, it is often necessary to make use of a combination of tests to increase the specificity of the test result. In this introduction, we will limit our discussion to the malabsorption of fat. The impaired absorption of fat leads to increased amounts of fat in stool, which is referred to as steatorrhea. The direct measurement of fecal fat is the most reliable laboratory test for establishing malabsorption, because increased fecal fat concentration is unequivocally associated with impaired absorption. Unfortunately, steatorrhea is not always present in the patients with malabsorption.

*Physico-chemical aspects and pathophysiology of intestinal fat absorption*

The total absorptive area of the small intestine is enormous (200 $m^2$). Not only the length of the gut, but also the surface of this part of the bowel contributes to this. The surface is arranged in small projections, called villi. Each villus is composed of thousands of intestinal absorptive cells, which overlie a core of blood vessels and lymphatics (Figure 1). Each absorptive cell itself is further modified to increase its surface area by the projection of microvilli on its surface, called the brush border. In this way, the adaptations of the intestine increased the surface area over 40-fold, facilitating the absorption of molecules that have been made available by digestive processes.

The motility of the bowel permits the nutrients to remain in close contact with the intestinal cells. Although the proximal intestine is the major absorption area for fat (monoglycerides and free fatty acids), the entire small intestine is involved in this absorption process (3). Most of the ingested dietary fats are in the form of long-chain triglycerides. These triglycerides are composed of both saturated (mainly palmitic and stearic acid) and unsaturated fatty acids (e.g. linoleic acid) and glycerol. About 30% of the dietary triglycerides, mainly medium-chain fatty acids, is digested by lingual and gastric lipase. The particle size of the bulk of the dietary lipid is largely decreased by the peristaltic contractions and temperature of the stomach (4).

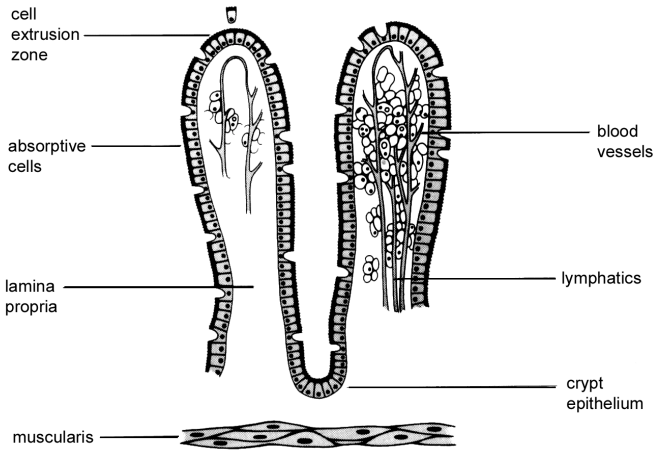



Figure 1. Villus of the intestinal mucosa. The epithelial cells that cover the surface of the villus absorb dietary molecules. At the apex of each villus the cells are sloughed off (2).

After a retention time of about 2–4 hours in the stomach, the partly digested food enters the duodenum. This process, together with the presence of acid, causes release of secretin and cholecystokinin, which in turn leads to a stimulation of the flow of bile and pancreatic juice. The pancreatic lipase acts at the oil-water interface of the emulsified triglyceride substrate. This emulsion is formed by mechanical moulding of fat in the gut in the presence of lipase, bile salts, colipase, phospholipids and phospholipase $A_2$. Pancreas lipase, colipase and bile salts form a ternary complex, which generates lipolytic products from tri- and diglycerides (Figure 2). Under normal circumstances, more than 98% of all ingested triglycerides are hydrolyzed to monoglycerides and free fatty acids by this complex (5).
Bile salts, which are synthesized by the liver and excreted by the gallbladder into the small intestine, not only play an important role in the digestion, but also in the absorption of fat. Bile salts are good detergents, having both polar (hydrophilic) and nonpolar (hydrophobic) groups that have the ability to lower surface tension. This enables the bile salts, to solubilize the free fatty acids, water insoluble soaps and monoglycerides. If the bile salt concentration in the lumen is high enough (Critical Micellar Concentration: 5–15 μmol/ml), the bile salts aggregate to form micelles. The fatty acids and monoglycerides enter these micelles to form mixed micelles. Then, the mixed micelles migrate to the absorbance sites of the intestine, where the fatty acids and monoglycerides are released from the micellar phase and enter the cell by diffusion. This unilaterial diffusion is promoted, because fatty acids and monoglycerides of long-chain fatty acids (≥C14) are promptly reesterified to triglycerides, upon entry into the mucosal cell. The esterification occurs by the interaction of the tryglycerides with apolipoproteins, cholesterol and phospholipid to form chylomicrons and very large density lipoprotein, which in turn are secreted into the intestinal lymph (Figure 2).

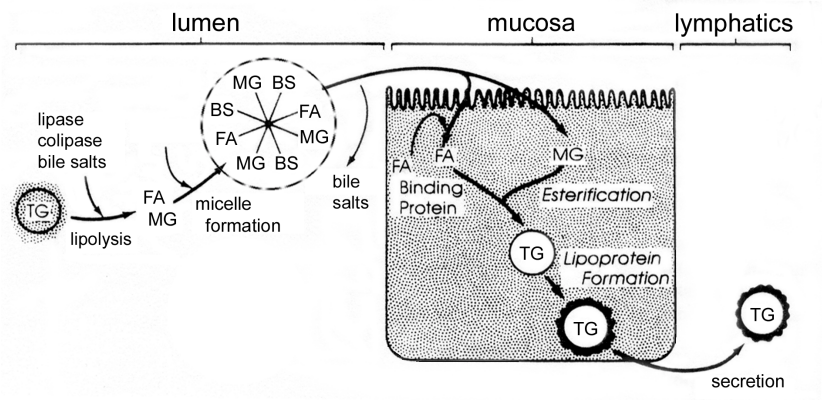



Figure 2. Diagram of intestinal digestion, absorption, esterification, and transport of dietary triglycerides. TG= triglycerides; FA = fatty acids; MG = monoglycerides; BS = bile salts (3).

Most of the bile salts are absorbed in the distal ileum, to reenter the enterohepatic circulation. Although the digestion and absorption of dietary fat is not described in complete detail, it is apparent that the whole absorption process comprises a very complex series of events. Normally, the unabsorbed dietary residue in feces is very small. Disturbances of any of the described events may lead to decreased fat absorption and subsequently give rise to an increased amount of fat in the stool. Therefore, many disorders can result in stools that contain poorly digested nutrients. Since they have diarrhea in common, these disorders are generally lumped together under the title malabsorption syndrome. These disorders may have vastly different etiologies as noted in Table 1. Although this table is far from complete, it shows that there are many etiologies causing steatorrhea. Therefore, it is important to distinguish between digestion and absorption.

*Diagnosis of steatorrhea*

The analysis of fecal fat is the most definite laboratory test for establishing the presence of steatorrhea, but it is not appropriate for the delineation of its cause. Under physiological conditions, about half of the fecal fats are non-absorbed, non-dietary fats, also called metabolic lipids. From these metabolic lipids, significant amounts (~2 g/day) derive from intestinal bacteria and epithelial cells, normally sloughed from the intestinal mucosa (8). Furthermore, a certain fraction of total fecal lipids is composed of unsaponifiable matter such as sterols (mainly cholesterol and coprostanol). On a lipid-free diet, the fecal fat output falls to values between 1 and 4 g/day, representing the sum of non-dietary fat (9). Dietary lipid consists of 92–96% triglycerides (10;11), whereas the remainder consists of cholesterol esters, plant sterols and phospholipids (12). In normal individuals, a daily dietary intake of up to 150 g lipid results in a relatively little change in total lipid excretion. Therefore, fecal fat excretions greater than 7 g/day is considered to be abnormal in adults consuming a usual Western diet with a daily intake of 50 to 150 g lipid. Children up to 6 months normally excrete 0.3–2.0 g fecal lipid per day (8;9). Patients with fecal lipid ≥20 g/day are suspected to suffer from pancreatic insufficiency, whereas fecal lipid contents >9.5 g/day but <20 g/day is thought to be the result of malabsorption of fat in the intestine (13). Often, the percentage of total lipid absorption is calculated. This percentage or coefficient of fat retention is calculated by:

$$\text{lipid absorption (\%)} = \frac{(\text{fat ingested} - \text{fat excreted})}{\text{fat ingested}} \text{ x } 100\%$$

To determine the percentage of total lipid absorption, the method is only claimed to be valid if at least 50 g of lipid is ingested per day (14). If the intake falls below 50 g lipid/day, the proportion excreted dietary lipid to non-dietary lipid becomes to small, to be important.

Table 1. Classification of the malabsorption syndrome (3).

*Inadequate digestion:*
- following gastrectomy (6)
- exocrine pancreatic insufficiency (7):
  - chronic pancreatitis
  - cystic fibrosis
  - pancreatic carcinoma
  - pancreatic resection
- Zollinger-Ellison syndrome (ulcerogenic tumor of the pancreas or gastrinomas, which cause inactivation of the pancreatic enzymes by increased amounts of acid)

*Reduced intestinal bile salt concentration:*
- liver disease
- abnormal bacterial proliferation in the small bowel, causing deconjugation of bile salts
- interrupted enterohepatic circulation, such as ileum resection
- sequestration, or precipitation of bile salts by drugs (e.g. cholestyramine)

*Primary mucosal absorptive defects:*
- celiac disease
- tropical sprue
- giardia lamblia infection
- mastocytosis
- radiation enteritis
- cystinuria

*Impaired lymphatic transport:*
- a-betalipoproteinemia
- obstructions
- cardiovascular disorders

*Accelerated passage through the intestine:*
- short bowel syndrome
- hyperthyroidism

The clinical usefulness of the quantitative analysis of fecal fat may be influenced if no standardized diet is applied during stool collection (see percentage fat retention). It has been noted that fecal fat excretion nearly linearly increases with the intake of dietary fat in patients with steatorrhea (24). Dinning et al. (25) described a standardized diet containing 100 g fat per day, to ensure sufficient accuracy of the test results. Another diet, containing 135 g lipid per day for lipid balance studies, was described by Nothman (9). Patients should be instructed carefully, in order to make them understand the importance of consuming the whole meal. After starting consumption of the diet, a 24–48 hour acclimatization period is needed preceding stool sample collection. Since the intestinal motility is variable, causing erratic frequency of fecal output from day to day, the results on single random sample analysis are generally considered to be useless. Therefore, the variation should be minimized by pooling at least three consecutive days collections (9). Fecal samples can be kept for up to 4 days at refrigerator temperatures (9). If separate lipid classes have to be determined, the fecal samples should be frozen as soon as possible after collection (9). Representative sampling is of general concern for any kind of analysis, but special attention should be given to the homogenization of the stool samples due to their inherent inhomogeneity. Despite of this homogenization, statistical averaging of the outcome of duplicate, or even triplicate samples of the pooled stool collection is often necessary (26).
As stated above, the impaired intestinal absorption of fat is only one of the many intestinal function disorders. More extensive descriptions are out of range of this consideration, but can be found elsewhere (3).

## 1.2. **Urolithiasis**

*Introduction*
Urinary calculi have plagued man over the centuries. Today, approximately 5 % of the population of the western world is thought to have formed at least one renal stone at the age of 70 years, from which they may suffer at some point in their lives (27-29). The mean age of the patients is about 45–50 years and approximately 60–70% of them are male (30-32). In the American population, stones are even three to four times more common in men than in women (30). In Western countries the portion of the population that is affected annually is about 0.5%. The yearly incidence of patients presenting to the hospital with urinary stone colic is about 0.1%–0.2% of the population (33;34). In 1% of the patients with urinary calculi the course of the disorder is without symptoms. Urinary calculi, or renal stones, may occur in different parts of the urinary tract, such as the kidney, renal pelvis, ureter, or urinary bladder (vesicle). In 80% of all cases, the urinary calculi will pass the urinary tract spontaneously, if the stones have a diameter smaller than 8 mm. Vesicle calculosis (bladder stones) are found fairly widespread in Asia. Bladder stones, due to malnutrition in the very early years of life, is currently frequent in areas of Turkey, Iran, India, China, Indochina and Indonesia, although the incidence is decreasing (in proportion) as social conditions gradually improve. At the beginning of the 20th century and beyond bladder stones were relatively frequent in Europe also, but in the course of the last 100 years, there has been a gradual decrease in its incidence, whereas the upper tract kidney stones became more common. This trend, defined as "stone wave", has been explained in terms of changing social conditions and the consequent changes in eating habits (more animal meat and fat). In Europe, Northern America, Australia, Japan, and, more recently Saudi Arabia, affluence

has spread to all social classes, and with it the tendency to eat "rich" food in large quantities (35). In these affluent societies the kidney is the most common site of urinary stones in the urinary tract, estimated to 58% of all cases (36). Stones, originating from the kidney, may traverse the ureter symptomless, but most of the time this passage is accompanied with severe pain and bleeding. Stones in the distal portion of the ureter or the bladder cause frequency, urgency and dysuria that may be confused with urinary tract infections. Classical symptoms of an acute renal colic are excruciating flank pain spreading downwards and anteriorly toward the ipsilateral loin and genitals. Renal colic is also often accompanied by nausea and vomiting, because the pain is so severe. Patients are often restless, tossing and turning in a futile attempt to find a comfortable position, which symptoms are also referred to as a renal colic (27).(30) Another problem of renal stones is that they tend to recur. The recurrence rate is about 50% in 10 years, and 75% in 25 years (34). These serious implications of urinary lithiasis cause high socio-economic cost, which justifies the investigation of its creation and prevention of recurrences. Therefore, much effort has been invested in the research of urinary calculi, comprising a great number of aspects such as the etiology, or pathogenesis of stones, the physico-chemical base of stone formation, risk factors, epidemiology and dietary, or medical treatments of urinary calculi. However, despite intensive research the knowledge of stone pathogenesis, which is the basis of every rational stone metaphylaxis, has remained rather scanty. Stone formation in most patients is probably caused by a coincidence of different environmental and genetic factors.

*Pathogenesis, including risk factors*

Urinary stones usually arise because of disturbance of a delicate balance. On the one hand the kidney must conserve water, on the other hand it is supposed to excrete waste and materials that have a low solubility. These two opposing requirements must be balanced against one another during adaptation to a particular combination of diet, climate and activity. The equilibrium is changed to some extent by the fact that urine contains substances that inhibit crystallization of salts, and others that bind ions into soluble complexes. These protective mechanisms however are less than perfect. When the urine becomes supersaturated with insoluble materials, due to e.g. a combination of excessive excretion rate and excessive water conservation, crystals form and may grow and aggregate with one another to form a stone (27). Except for potent inhibitors, human urine also contains a number of promoters (albumin, globulins, matrix substance A). A list of promoters, inhibitors and other predisposing risk factors is given in Table 2. The predominant risk factor is poor hydration. At least this partially explains the increased incidence of renal stone formation in hot climates (37). In general the etiology of stone formation comprises genetic factors, environmental factors, such as dietary causes (e.g. hyperuricosuria), or urinary tract infection. The most commonly occurring component of stones is cationic calcium, caused by idiopathic hypercalciuria which probably has a genetic origin and occurs in 50–55% of all stones (27;38).

*Physico-chemical factors*

The physico-chemical basis of stone formation is mainly supersaturation. If a solution is in equilibrium with crystals of e.g. calcium oxalate, the product of chemical activities of positive calcium ions and negative oxalate ions in solution is called equilibrium solubility product. If crystals are removed and than either calcium or oxalate is added to the solution,

the activity equilibrium solubility product will increase, but the solution remains clear. Such a solution is considered metastably supersaturated. Alternatively, if new calcium oxalate seed crystals are added, the crystals will grow in size (nucleation). If a critical point, called the upper limit of metastability, is reached, the solid phase begins to develop spontaneously (27;39).
Nucleation is also considered to be a physico-chemical factor in stone formation. If urine is supersaturated, the crystals normally form instable clusters of crystals. However, clusters of at least 100 crystals can remain stable, because attractive forces balance surface losses. These clusters, called nuclei, can create a permanent solid phase if the urine is frequently supersaturated. If supersaturated urine is seeded with nuclei containing crystals of the same structure, this is called homogeneous nucleation, whereas seeding of supersaturated urine with foreign nuclei is called heterogeneous nucleation. Sodium hydrogen urate, uric acid and hydroxylapatite crystals often serve as heterogeneous nuclei that permit calcium oxalate stones to form even though urine calcium oxalate never exceeds the metastable limit. The previously mentioned inhibitors (Table 2) slow down crystal growth and nucleation of calcium phosphate and calcium oxalate (27). Struvite, cystine and uric acid stones often grow too large to pass the ureter. These stones gradually fill the renal pelvis to form staghorn calculi. Calcium stones often grow in the urinary papillae, some of them break loose and cause colic.

*Composition of urinary stones*
The majority of all urinary stones contains calcium (31;32;40;41). Table 3 shows the incidence rates and etiology of the most commonly occurring components in urinary calculi. The distribution of these incidence rates (%) is based on the incidences of the components in mixed stones, as found in the St Elisabeth Hospital in Tilburg, in the southern part of the Netherlands (31). The incidence rate of 70-80% of calcium oxalate (Table 3) was similar to results obtained from an own study (University Hospital in Groningen) and a study in France (32). We presume that these incidence rates are the same in most Western countries. However, the distribution of the incidence rates may differ in certain regions.

*Diagnosis and analysis*
Urinary stones can also be detected by means of abdominal radiographic studies, which however may miss many stones. Therefore X-rays are often followed by an intravenous pyelogram (IVP), which requires an injection of dye. Unfortunately, this dye may cause allergic reactions. Another detection method is renal ultrasound, which sometimes misses stones in the lower half of the ureter. The newest technique is spiral-computed tomography (spiral CT). It is a non-invasive method that produces images of the urinary tract by X-rays (30).

Most diagnostic protocols include the analysis of biochemical parameters in 24h urine collections for the identification of risk factors of urinary stones (32;34). Normally the output of calcium, uric acid, oxalate, citrate, magnesium and urea (as reflection of daily protein intake) are measured in 24h urine. Most centers also measure the urinary cystine output. Furthermore the urinary pH is often determined, together with the 24h urinary volume. In case of high pH (8–9), the urine is sometimes tested for the presence of bacterial infection with Proteus species. In addition, the serum concentrations of calcium, uric acid

and the parathyroid hormone are often measured. Sometimes the assays are repeated after dietary restriction.

Table 2. Promoters, inhibitors and pre-disposing risk factors of stone formation (37).

| |
|---|
| **Promoters** |
| Albumin |
| Globulins |
| Matrix substance A |
| **Inhibitors** |
| Magnesium |
| Citrate |
| Pyrophosphate |
| Tamm Horsfall glycoprotein |
| RNA |
| **Predisposing factors** |
| *Preurinary* |
| Family history |
| Hot climate |
| Stress |
| Decreased fluid intake |
| Protein-rich diet |
| Immobilization |
| *Urinary* |
| Increased $Ca^{++}$, urate, oxalate, pH |
| Decreased $Mg^{++}$, volume, citrate |
| *Metabolic disorders* |
| Primary hyperparathyroidism |
| Renal tubular acidosis type I |
| Hereditary hyperoxaluria |
| Medullary sponge kidney |
| Cushing's disease |
| Cystinuria |
| Milk-alkali syndrome |
| *Bacterial infection* |
| Proteus infection |

The analysis of the composition of the calculi is important for proper treatment of patients with urolithiasis, especially in case of recurrence of stones. The compositions of urinary stones can be determined by means of wet-chemical analysis, infrared spectroscopy, or X-ray diffraction. Unfortunately, wet-chemical analysis is only a semi-quantitative assay. Therefore, infrared spectroscopy and X-ray diffraction are gradually replacing the less specific chemical assay for stone analysis. More detailed information upon the analytical methods is given in paragraph 2.1.2 and 2.2. of the introduction.

Table 3. Incidence rate (%) and etiology of the most commonly occurring components of urinary calculi. The incidence of the components expresses the presence of the component in mixed stones, as found in a hospital in the southern part of the Netherlands (31).

| Component name | Formula / Composition | Incidence rate (%) | Etiology (41) |
|---|---|---|---|
| Whewellite | $CaC_2O_2.H_2O$ / Calcium oxalate | 75.0 | Hyperoxaluria, hypercalciuria, hyperuricosuria, hyperuricaemia and primary hyperparathyroidism |
| Weddellite | $CaC_2O_2.2H_2O$ / Calcium oxalate | 70.7 | See whewellite |
| Carbonate apatite | $Ca_{10}(PO_4)(CO_3OH)_6(OH)_2$ / Calcium phosphate | 48.9 | Hypercalciuria, renal tubular acidosis (RTA), urinary tract infection (not essential), hyperphosphaturia and immobilization |
| Brushite | $CaHPO_4.2H_2O$ / Calcium hydrogen phosphate | 13.0 | Hypercalciuria, hyperphosphaturia, RTA and immobilisation |
| Struvite | $MgNH_4PO_4.6H_2O$ / Magnesium ammonium phosphate | 4.3 | Urinary tract infection with urease producing bacteria |
| Uric acid | $C_5H_4N_4O_3$ | 3.3 | Hyperuricosuria and hyperuricaemia |
| Ammonium urate | $C_5H_7N_5O_3$ | 1.1 | Hyperuricosuria and urinary tract infection |
| Cystine | $C_6H_{12}N_2O_4S_2$ | 1.1 | Cystinuria |

*Medical management of urinary stones*

In most cases stones are lost by time and fluid, allowing passing the stone on its own. However patients with stones larger than 6 mm may often need help. In the past, urinary calculi could only be removed by operating the kidney, renal pelvis, or ureter. Today, alternative methods are available. Stones can be fragmented in situ by exposing them to extracorporal shock wave lithotripsy (ESWL). The patient is submerged in a water bath, after which high-energy sound waves are focused at the center of the stone by means of a parabolic reflector. Subsequently, the stones are fragmented with the use of laser energy, electromagnetic or electro-hydraulic transducers. In this way, most stones are reduced to powder that passes through the ureter to the bladder. A second method is percutaneous ultrasonic lithotripsy. With this method a cystoscope-like instrument is passed into the renal pelvis, where an ultrasonic transducer disrupts the stones. The fragments are washed out directly. A third method is ureteroscopy, by which a cystoscope-like instrument is passed through the bladder into the ureter (27;30).

Rational stone prophylaxis is important, especially in all cases of stone recurrence. Conservative treatment should always be offered to patients with stones, whether or not

additional treatment with drugs or diets is necessary. Traditional treatment always includes high fluid intake of at least 3L/day to ensure a minimum urine volume of 2 L/day, irrespective of the composition of the urinary stone (42). The composition of the stone, as well as the frequency and extent of severity of stone formation determine the kind of additional treatment, which may consist of dietary advise, or medication with drugs. With respect to nutrition many interesting studies are available, such as a study of the effect of drinking French mineral water containing calcium and magnesium (43). One of the remarkable recent findings of new research on urinary stones, is that dietary calcium restriction possibly is detrimental in prevention of stone formation and in fact seems to make things worse (30). Nevertheless, a more extensive review with respect to additional treatment is out of scope of this introduction, but can be found elsewhere (27;34).

# 2. Analytical methods

## 2.1. Reference methods

### 2.1.1. *Fecal fat analysis:*

#### 2.1.1.1. Van de Kamer method:

*Introduction*
The titrimetric Van de Kamer method is the most popular method for the determination of fecal lipids. As a result of a comprehensive study of Van de Kamer as described in his thesis (15), the method was first published in 1949 by Van de Kamer et al (16). Even though the method exists for a very long time, it is still used by a great number of laboratories and is considered by many as the gold standard procedure for the determination of fecal fat. The method is intended for the quantitative measurement of neutral lipids (unsplit), as well as medium- and long-chain fatty acids (split).
With the most common procedure (method A), the determination of the fat content in a homogenized stool sample is performed without drying the sample. The lipids of a weighted amount of stool sample are saponified by boiling under a reflux condenser with concentrated potassium hydroxide in ethanol. After cooling down the alkaline solution, HCl is added to liberate the fatty acids from their salts (soaps). After cooling again, ethanol is added, and the fatty acids are extracted with petroleum ether. Subsequently, the liberated and extracted fatty acids are titrated in a fixed amount of the extract with isobutyl alcoholic KOH and thymol blue as indicator. In this way the split and unsplit fat is measured simultaneously as total fat. The fecal lipid content is normally expressed in mass percent (g%), or g/day wet weight.

Van de Kamer also described an alternative procedure (method B) for the determination of split and unsplit fat separately. To measure the amount of split fat, the stool sample is not treated with alkali for saponification, but boiled with diluted HCl to convert the fecal soaps into free fatty acids. After extraction with petroleum ether, the fatty acids are quantified by titration. After titration an excess, but known amount of isobutyl alcoholic KOH is added and the unsplit fat is saponified by boiling. The excess of alkali is titrated with HCl and thymol blue as indicator, from which the amount of unsplit fat can be quantified. With this method, the free fatty acid index related to the amount of ingested triglycerides can be calculated. Increased amounts of unsplit fat suggest impaired digestion. Unfortunately, this method may lead to false negative results. In the '*diagnosis of steatorrhea'* section of chapter 1.1, we already described the drawback of the free fatty acid index for the detection of impaired digestion, because bacterial lipase can split substantial amounts of triglycerides in the colon (44).
Although Van de Kamer has described his method in great detail, some additional notice will be given to specific issues in the next section that may be helpful for setting up new methods.

*Additional remarks in relation to the Van de Kamer method.*
*Extraction procedure*
Many fatty acids in the stool are present as insoluble divalent soaps ($Ca^{2+}$ and $Mg^{2+}$ salts), which cannot be extracted with petroleum ether. Lowering the pH to 2 will liberate all fatty acids from the soaps. Solvents, such as petroleum ether, chloroform and acetone, normally extract triglycerides and fatty acids quantitatively from dry matter. Van de Kamer (15) has found that the extraction recovery of fecal lipid from wet stool samples was low, but that the distribution coefficients remarkably increased by adding ethanol to the acidified solution. A contribution of 60% ethanol (v/v) resulted in nearly 100% recovery of the long chain fatty acids with chain length greater than fourteen (mainly palmitic and stearic acid) using a single extraction. Under these conditions myristic acid (C14) had a recovery of 90 %, whereas the recovery of the short chain fatty acids (≤ C6) was very low (< 25%). Lowering or raising the ethanol contribution lowers the extraction recoveries. About 1% of the 60% ethanol layer dissolves in the petroleum layer. Therefore, Van de Kamer used a small correction factor in his calculation formula.

*Quantitative analysis of fecal lipids*
Using the Van de Kamer method, the amount of lipid is quantified by titration of the free fatty acid COOH group with sodium hydroxide. As a consequence of this titration, the lipid content has to be calculated by using the mean molecular weight of fatty acid. It is important to notice that the total fecal lipid content is normally expressed as triglycerides in g/day, but sometimes the total amount of fecal fat is expressed in fatty acids. Unfortunately, the lipid class (triglyceride, or fatty acid) in which the total lipid is expressed is hardly ever mentioned. Because the titration is applied on the COOH group of the fatty acids, the molecular weight, used in the calculation procedure, must include the molecular weight of 13 from the glycerol residue (CH) of the triglycerides, otherwise an underestimation of the outcome of about 5% will occur if total lipid is supposed to be expressed in triglycerides in g/day. Van de Kamer (15) has found that the mean molecular weight of fatty acids in feces depends on the composition of the dietary lipids. Dietary habit may vary in different countries and may in change in time. Therefore, the exact (mean) molecular weight of the calculation formula is important, in order to obtain accurate results. Van de Kamer used a mean molecular weight of 276 in this standard formula for the determination of total fecal lipid. Van de Kamer estimated his molecular weight by weighing and titration of the fatty acids in purified petroleum extracts of feces (mw fatty acids = mg / mmol fatty acids). His final molecular weight was based on a mean molecular weight of 263 of the fatty acids in normal adults and adding 13 for the glycerol residue. The calculation formula of the Van the Kamer method, is defined as:

$$\text{Fecal fat in g per 100g feces} \approx ml_{\text{titrated NaOH}} \times 276 \times N_{NaOH} \times 1.03$$

The dilution and weighing conversion factors are not specified in this formula, because they strongly depend on the exact procedure that is used. The factor 3% (1.03) is a combined correction factor for the volume increase of ethanol in the petroleum ether layer and an adjustment for the average distribution coefficient (15).

*Origin of fecal lipids*

The aim of the Van de Kamer method is the detection of possible impaired absorption of dietary fat. As a consequence, the method should be restricted to the detection of dietary lipids (15). To judge this fact, some general understanding of the origin of the various lipid classes in feces is necessary. The lipids in stool may be subdivided into dietary lipids, volatile fatty acids (VFA) and endogenous lipids.

Dietary lipid:

The majority of the dietary lipids are triglycerides (up to 150 g/day). The daily diet of an average western adult also contains about 4–8 g phospholipids, predominantly lecithin, and small amounts of sterols (0.5 g), such as cholesterol and sitosterol. The upper gut of an average healthy person normally absorbs over 98% of these ingested lipids. As a consequence, about 1–3 g of the dietary lipids is normally excreted in the feces.

VFA

VFA are fermentation products from carbohydrates produced by the large intestinal bacteria. These VFAs, such as acetic acid, propionic acid and butyric acid, are mostly absorbed by the colonic wall (45) to provide metabolizable energy.

Endogenous lipids:

The endogenous lipids are primarily remnants of biliary lipids (bile acids and sterols) and phospholipids from membranes of sloughed intestinal and bacterial cells. About 15–40 g of endogenous lipids (biliary, sloughed cells and other intestinal secretions) are normally re-absorbed in the small intestine.

- Approximately 1 g of cholesterol (secreted from the gall bladder) is eliminated from the body per day. About half of the cholesterol is excreted in the feces in the form of neutral sterols, whereas the rest is excreted as bile acids. Coprostanol is the major sterol in feces, which is formed from cholesterol by the bacterial flora. The majority of the so-called primary bile acids (e.g. cholic acid) is re-absorbed in the small intestine, whereas the rest (approximately 0.5 g/day) is metabolized by colonic bacteria and are subsequently excreted in the feces (46).
- Most (1–3 g/day) endogenous fecal lipids (phospholipids) derive from membranes of sloughed cells and bacteria. During transit through the colon several bacterial modifications occur, including hydrolysis of the phospholipids by various bacterial lipases (5).

Care et al. (5) have fractionated the fecal lipids based on their different physical properties. Each fraction was successively quantified chromatographically. Table 4 shows the relative amounts of lipids that are usually excreted in stools of healthy man with a daily total output of 4–6 g/day. From this table it can be seen that no glycerides are found. Only in case of severe pancreatic insufficiency glycerides may be present. Furthermore it can be seen that the majority of lipids are in the form of fatty acids or soaps.

The principle of the Van de Kamer method is based on additional saponification of the glyceryl-, sterol- and phospholipid esters, liberating the fatty acids from the soaps by lowering the pH, extraction of the apolar components in petroleum ether and detection by

titration of the carboxyl group (COOH) of the fatty acids with sodium hydroxide. As a consequence, the phospholipids in stool, most of them originating from cellular debris and bacteria, are additionally measured as free fatty acids. Van de Kamer has shown that VFA are not quantitatively extracted into petroleum ether from the alcoholic HCl solution, using a single step (15). As a consequence, the Van de Kamer method measures all non-neutral lipids and other acidic organic components (such as bile acids and other apolar organic acids), but hardly any VFA. Other methods for the determination of fecal fat are even more non-specific. Some of these methods, such as the gravimetric method of Sobel (9) measure the total fecal lipid content (neutral and non neutral lipids).

Table 4. Relative distribution of lipids in feces of individuals with normal lipid excretion rates.

| Lipid fraction | % of total lipid |
|---|---|
| Fatty acids and $Na^+$ and $K^+$ soaps | 70 |
| $Ca^{2+}$ and $Mg^{2+}$ soaps | 10 |
| Glycerides (TG and DG) | 0 |
| Neutral sterols and bile acids | 15 |
| Other (e.g. phospholipids) | 5 |

TG, triglycerides; DG, diglycerides

As shown above, stools contain various lipid classes, many of these lipid classes have different origins and the physical and chemical properties of the lipids may be modified at different sites in the intestine. None of the methods for determination of fecal fat is able to measure the excretion of lipids of dietary origin alone. Under normal physiological conditions, about half of the fecal lipid is endogenous, the remainder of dietary origin (47). Therefore, one should notice that all methods for determination of fecal fat measure a certain amount of background noise (non dietary lipids) with a certain amount of dietary lipids superimposed on top of it. Fortunately, the amount of non-dietary lipids is fairly constant. Therefore, the upper reference limit of 7 g/day fecal fat is based on a combination of excreted endogenous and dietary lipids. Even today, there are no methods available that can measure fecal lipids of only dietary origin in a simple, or inexpensive way. If patients with steatorrhea consume a lipid free diet during the test, increased fecal lipid concentrations will not be found. Therefore, it is recommended that patients consume a standardized diet containing at least 100 g fat per day, to ensure sufficient accuracy of the test results.

2.1.1.2. Gas chromatography:

Introduction

Gas chromatographic analysis of different lipid classes may be used to gain a better insight in the distribution of different lipid classes of fecal lipid. In addition, the GC analysis of FA in fecal lipid may be used for the determination of the mean molecular weight of FA in

stool (48). As stated in the previous chapter, fecal lipids may originate from different sources (diet, bacteria, cell membranes, etc.). Apart from their origin, there is no generally accepted definition of 'lipid'. Christie (49) has defined lipids as 'fatty acids' and their derivates, and substances related biosynthetically or functionally to these components. A major classification is generally made between simple and complex lipids. The simple lipid class contains lipids such as fatty acids (FA), fats (esters of fatty acids with glycerol), whereas the complex lipid class contains lipids like phospholipids and precursors of lipids and derived lipids. The neutral lipids, such as sterols (e.g. cholesterol) and sterol esters belong to this last category (50). In a certain sense, bile acids may be reckoned among the complex lipid class also, since they are derived from cholesterol.
To obtain better insight in the composition of different lipid classes, separation of the different lipid fractions of feces may be performed by a variety of techniques such as preparative thin layer chromatography (51), high-performance liquid chromatography (HPLC) (49), or solid phase extraction (52). Hoving et al (53) have used capillary gas chromatography (GC) for characterizing the fatty acid compositions of cholesterol ester and triglyceride fractions in plasma, using a preceding solid phase extraction with an aminopropyl-silica column. Because of its high separation power, GC analysis, using capillary columns and flame ionization detection (FID), is definitely an important technique available to the lipid chemist for the analysis of FA in various biological fluids (54). This method is suitable for the quantitative analysis of different kinds of fatty acids in total lipid, or separate lipid classes. GC analysis of lipid can be applied on a wide variety of biological materials such as plasma, erythrocytes, amniotic fluid, tissue, or in feces as described hereafter. Contemporary methods for the analysis of FA make use of apolar capillary GC columns. With this column type, the fatty acids are separated in the order of their mass. With the standard GC analysis, the saturated, as well as the unsaturated fatty acids from C14 (myristic acid) up to C26 can be separated in a quantitative manner with sufficient separation power. Cholesterol, bile acids and other sterols elute after the fatty acids from the GC column. By using an adjusted temperature program and an appropriate isolation procedure that is capable of a quantitative extraction of the more volatile FA, the medium chain fatty acids, as low as C6, may also be analyzed.

*FA analysis of different lipid classes, using GC.*
Verkade et al (19) described the determination of the fatty acid composition of the triglyceride, cholesterol ester and free fatty acid fractions of feces, after separation of these lipid classes with solid phase extraction. Their method was adapted from a method of Kaluzny et al (52). Kaluzny and associates used a bonded phase aminopropyl column for the separation of 7 lipid classes (FA, triglycerides, diglycerides, monoglycerides, cholesterol esters, phospholipids and cholesterol) on the basis of lipid polarity, solvent strength and polarity. With this separation method, recoveries of at least 97% were reached for each of the lipid classes.
Before isolation of the lipids from feces in a relative pure state, the pH of the fecal sample has to be brought to pH 2, in order to liberate the fatty acids from their potential soaps. This step was however not used by Verkade. The extraction of the fecal samples is performed with a chloroform-methanol mixture (2:1 by volume) which was described by Folch (55). To prevent auto-oxidation of the polyunsaturated fatty acids, butylated hydroxytoluene (BHT) is usually added during extraction. Once extracted in the Folch solvent mixture, the

fecal lipids can be separated into their respective lipid classes. The Folch extract is evaporated to dryness and redissolved in hexane. The hexane is brought onto the aminopropyl-silica column and eluted with 2 separate aliquots of hexane. The combined hexane eluate contains the cholesterol-ester (CE) fraction. The triglyceride (TG) fraction is subsequently collected by elution (3x) with diethyl ether:dichloromethane:hexane (1:10:89, vol:vol:vol). The FA fraction is collected by eluting the column with 2% acetic acid in diethylether (2x). Finally, the phospholipid (PL) fraction is eluted from the column with methanol. The FA, TG, CE and PL fractions are evaporated to dryness. The fatty acids of the fractions are transmethylated with a methanol:6 mol/l HCl (5:1, vol:vol) mixture to fatty acid methyl esters. The fatty acid methyl esters (FAME) are purified by extraction of the mixture with hexane. Figure 3 shows the separation process of the different fecal lipid classes. The collected FAME in hexane is analyzed by GC with FID, by injecting a small amount of the hexane extract on an apolar cross-linked methyl silicone column. The fatty ester methyl esters are identified on the basis of their retention times, using standard solutions containing even- and odd numbered saturated and unsaturated fatty acids. Quantification of the fatty acids is performed by adding a known amount of C17 fatty acid to the samples at a certain point in the extraction procedure. This C17 is used as an internal standard in the GC procedure (19).
If FAME analysis of the combined lipid classes (total lipid) is required, the fecal samples must be brought to pH 2, extracted with a Folch mixture, evaporated to dryness, transmethylated and extracted with hexane. Except for FAME the hexane may contain neutral sterols and bile acid methyl esters. GC analysis of these components, using a capillary column with an apolar stationary phase, may show tailing peaks. This peak tailing may be caused by the polar 3-OH group of the cholesterol backbone of these components. Therefore, the hexane layer should be dried and the sterol and bile acid fraction should be trimethylsilylated with tri-sil-TBT. After forming of the fatty ester methyl esters and trimethylsilylated sterols and bile acids, the components must be extracted with hexane. Figure 4 shows the extraction process of FAME, sterols and bile acids from stool. Figure 5 shows a GC lipid profile. The fatty ester methyl esters are identified and quantified as described before. A description of the identification and quantification of the sterols (56) and bile acids (57) may be found elsewhere.

Based on the selectivity of GC analysis, the method may be used as a reference method for fecal fat determination. The selectivity could even further be enhanced by a combination of gas chromatography and mass spectrometry (GCMS). Nevertheless, the method should not be used for routine analysis of fecal fat, because of its rather high complexity.

Figure 3. GC analysis of FAME in the fractionated fecal lipid classes.

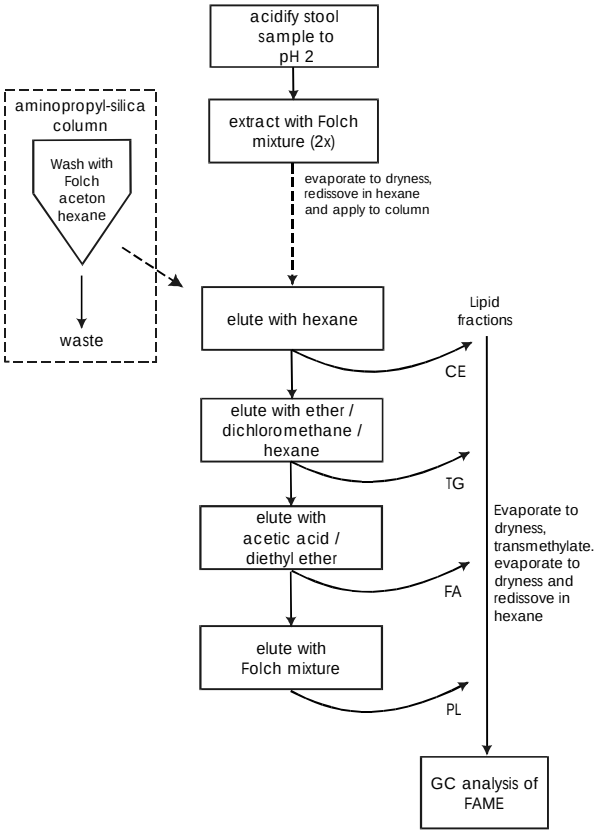
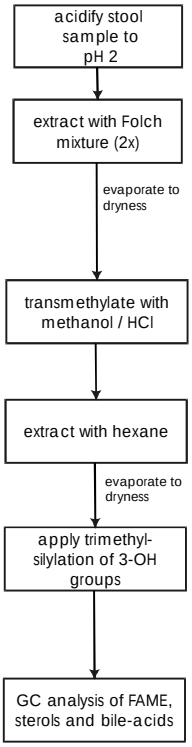



Figure 4. GC analysis of FAME sterols and bile-acids in total fecal lipid.




FAME, fatty acid methyl esters; CE, cholesterol esters; TG, triglycerides; FA, fatty acids; PL, phospholipids

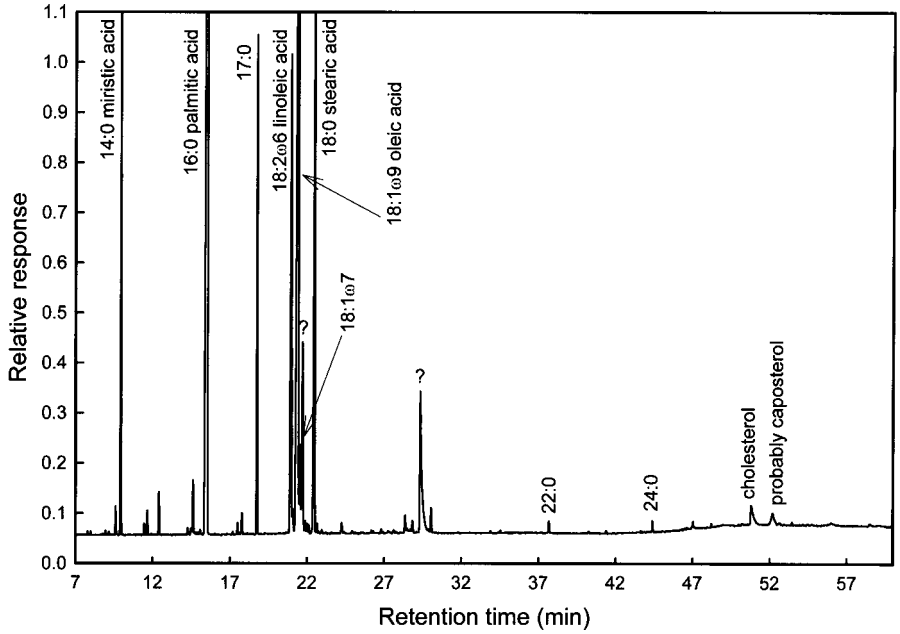



Figure 5. GC transmethylated fatty acid profile of feces. The relative response is defined as the peak height of the respective fatty acids in relation to the height of the internal standard (17:0).

### 2.1.2. *Urinary calculus analysis:*

#### 2.1.2.1. Wet and dry chemical analysis:

*Introduction*

The chemical analysis of urinary calculi has been an often-neglected field in clinical chemistry. However, the quantitative determination of the chemical composition of urinary calculi is important. Accurate analysis of the composition of calculi may provide an indication of the underlying condition and direct efforts towards its identification and treatment. Qualitative methods have dominated the investigation of urinary calculi for more than 100 years. In 1860, J. F. Heller (58) proposed a scheme for chemical investigation of urinary stones. His method was based on the colour of the sample, the odour at ignition of the pulverized material, and a number of chemical reactions performed on the dry sample. Even today, this scheme is utilized in a number of clinical laboratories (59). Although the method has gone through some modifications, it is commercially available in the form of kits, with tests for routine qualification of the composition of the urinary calculus. Of these kits, the Merckognost 11003 kit for urinary calculus analysis (Merck, Darmstadt, Germany) is widely used. Other kits, such as the Oxford and the Temmler kit are no longer commercially available. The tests of these kits all rely on spot colour end-point detection of the ions of the components of dried samples (60). Another method, often combined with the qualitative dry ‘spot test’, is quantitative ‘wet chemistry’. In the wet chemistry method, ions derived from dissolved stone material are quantified using automated chemistry analyzers.

Subsequently, the quantified ions can be combined into salts by calculation (61). In contrast to dry spots tests and wet chemical analysis, which only measure ions, infrared spectroscopy and X-ray diffraction provide information on the actual salts. The latter two physical-chemical analytical methods also provide information about the degree of hydration of the components, and demonstrate better quality in the analysis of spurious calculi. Table 5 shows the absolute and relative distribution of the techniques used by a number of laboratories of the Netherlands in 1986 and 10 year later. These data were reported by the Stichting Kwaliteitsbewaking Ziekenhuis Laboratoria [SKZL, the Dutch quality control society] (62). From this table, it can be seen that the total number of participants using own chemicals or commercially available kits decreased in favor of laboratories using infrared spectroscopy. A similar shift, in the direction of physical-chemical techniques, was also observed in quality control programs, organized by the German quality control society between 1980 and 1989 (63). The physical-chemical methods will be described in the next chapters. No single method of analysis is perfect. Sometimes, the best approach is using a combination of techniques. Below, the dry spot test and wet chemical analysis is described in more detail.

*Dry spot test*

Preceding performance of the spot tests, the colour, shape, size, and consistency (hardness) of the calculus have to be recorded. A description of the specific characteristics of a number of urinary calculus components may be found elsewhere (64;65). If only pulverized sample material is available, the colour and the consistency have to be recorded. After weighing the sample, the calculus must be washed with de-ionized water and completely dried with filter paper, or with silica gel. It is important to dry the calculus at ambient temperatures, because some of the calculus components may lose crystal water when drying at higher temperatures. Struvite ($MgNH_4PO_4.6H_2O$) not only loses crystal water, but also ammonia, when dried at 37 °C (61;66).

Table 5. Absolute and relative distribution of urinary calculus analysis techniques, used in different laboratories in the Netherlands in 1986 and 1996.

| | 1986 | | 1996 | |
|---|---|---|---|---|
| Total number of participants: | 69 | | 37 | |
| | | | | |
| Own reagents (qualitative) | 17 | 25% | 2 | 5% |
| Own reagents (quantitative) | 3 | 13% | 4 | 11% |
| Oxford kit | 16 | 23% | - | |
| Temmler kit | 8 | 11% | - | |
| Merckognost kit | 14 | 20% | 17 | 46% |
| Infrared spectroscopy | 3 | 4% | 10 | 27% |
| X-ray diffraction | 1 | 1% | 1 | 3% |
| Polarization microscopy | 1 | 1% | 2 | 2% |

Urinary calculi are normally formed over a long period. Therefore, the calculus may be layered and the nucleus and each layer may contain different components. However, in the

routine laboratory the calculi from each patient are pulverized and mixed very well and the examination is carried out on samples of this material.
Urinary calculus components can be divided in organic and inorganic components. Urate and xanthine components are classified as organic components, whereas oxalate, phosphate and carbonate containing components belong to the inorganic component class (64). A distinguishing feature of organic components is that they will burn in a flame, resulting in loss of volume. This can be tested by burning a small, but known amount of calculus powder in a flame (oxalate will burn partly). For this reason and in this context, oxalate is classified to the inorganic component class.
About 50 mg of dry, pulverized calculus material is needed to perform the spot tests. With the spot tests of e.g. the Temmler kit the following ions and organic components can be identified: oxalate, carbonate, phosphate, magnesium, ammonium, calcium, uric acid and cystine (Figure 6). For example, the detection of effervescence of $CO_2$ after addition of acid (HCl) to a small amount of stone powder indicates the presence of carbonate in the urinary calculus (present in calcium carbonate, or carbonate apatite).
Unfortunately, often little information can be obtained about the major constituent of the urinary calculus, using this method.

*Wet chemistry analysis*
Wet chemical analysis is based on the quantification of ions and organic components, from which the quantitative composition of the salts and components may be calculated. In mixed stones these calculations may be rather complex. Therefore, it is particularly important that the stone is carefully examined before analysis. As a consequence, it is often undesirable to crush and analyze the whole calculus, as minor components may be diluted out and overlooked. After drying the pulverized calculus sample(s) on silica gel, an accurately weighed amount of sample is dissolved in HCl, sulphuric acid, or nitric acid. Larsson et al (61) found that nitric acid was the only effective agent for complete dissolution, however others have reported (67) that organic components do not dissolve into acid quantitatively. For each analysis, about 10–15 mg of sample material is needed. After dissolution of the calculus the ions (magnesium, calcium, oxalate, ammonium and phosphate) and organic components (urate and cystine) may be measured with automated laboratory analyzers. Especially wet chemistry of urinary stones may be prone to errors, because the quantitative results are always based on the assumption of 100% recovery. Most urinary calculi contain small amounts of protein and mucopolysaccharides (68), causing only a minor decrease of the recovery. However, if other unusual components (such as xanthine and spurious components) are present in the sample, the recovery may definitely not be 100%. If small recoveries are obtained (<70%) the sample should be examined by further chemical or physical-chemical techniques.

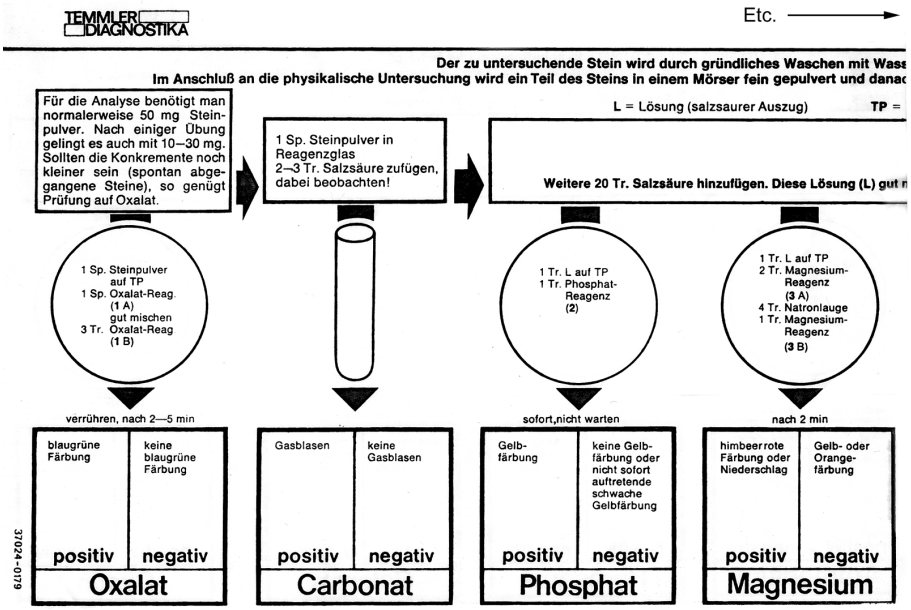



Figure 6. Part of the procedure of the Temmler kit for qualitative analysis of urinary calculus compositions.

Based on the results of external quality control surveys (63), the single use of the qualitative dry spot tests and semi-quantitative wet chemistry analyses is not recommended (69). On the other hand the use of both techniques may provide additional information to infrared analysis or X-ray diffraction for the determination of the composition of urinary calculi.

### 2.1.2.2. X-ray diffraction:

Nearly 95% of all solid urinary calculus materials appears in a crystalline form, whereas the remaining 5% is amorphous. The atoms, ions, or molecules of the crystalline solids are arranged in regular patterns, which are repeated in three dimensions. In amorphous substances, the atoms are ordered in a random way (70). Sometimes, crystals are embedded in amorphous structures of the same material, in which case one speaks of the amount of crystallinity (70).

When an X-ray beam hits an atom, the electrons around the atom will start to oscillate at the same frequency as the incoming wave, resulting in destructive inference in almost all directions. This means that the combining beams are out of phase and no energy will leave the solid sample. Because the atoms in a crystal are arranged in regular patterns, a few directions will have constructive interference. Therefore, according to Bragg's law (71), crystals appear to reflect X-rays when an X-ray beam hits parallel atomic layers at certain angles of incidence (theta, θ). By stepwise changing the angles of incidence, the X-rays

interact with crystalline substances, resulting in highly specific diffraction patterns. As a consequence, every crystalline substance produces its own specific pattern, and in mixtures each substance produces its own pattern together with each of the other substances. An extensive description of X-ray diffraction can be found in a publication of Dosch et al. (72).

An X-ray diffraction (XRD) analyzer (Figure 7) is composed of a source of X-rays, a sample holder and a detector. A narrow beam of X-rays, which strikes the crushed (pulverized) crystalline substance, is scattered in patterns that depend upon the electron densities in the different parts of the crystal. The scattered or diffracted beams can also be detected by means of a Debije-Scherrer-Hull camera on X-ray photographic film (73), or with a solid state electronic detector (74). The films, obtained with the Debije-Scherrer-Hull camera contain patterns of dark concentric rings. The radius of each ring is a measure of the crystal lattice distance, whereas the amount of blackening of the ring on the emulsion indicates the intensity of the reflected radiation, which in turn can be used for the calculation of the relative composition of the components in a mixture (75). Today, the Debije-Scherrer-Hull camera is not used very often anymore for urinary calculus analysis and has been replaced by X-ray diffractometers (Fig. 7). By stepwise moving the detector with an angle from 0° – 50° over the sample, the detector records the measured intensity as a function of the diffraction angle.

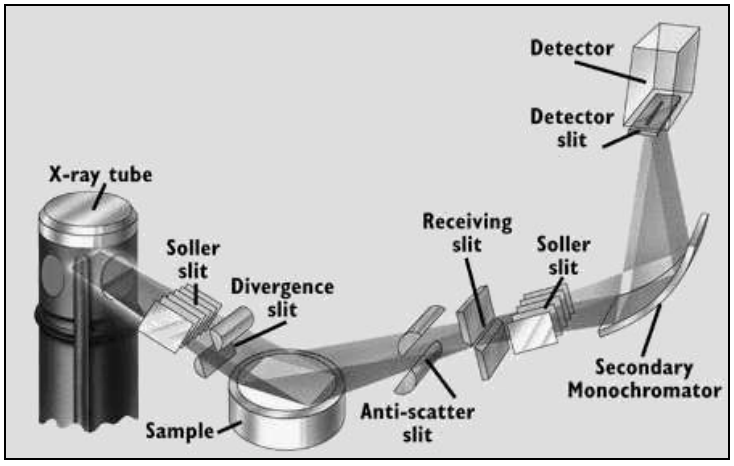



Figure 7. Diagram of a X-ray diffraction analyzer

After mathematical conversion of the detector signals, the typical diffraction spectra (diffractograms) can be shown. These diffractograms consist of a plot of reflected intensities against the detector angle 2-theta (degrees 2θ), or theta (degrees θ), depending on the goniometer configuration. In case of urinary calculus analysis, about 4 mg of the grinded sample is applied on a mono-crystalline silicon powder applicator disk.
Figure 8A shows a typical diffraction spectrum of whewellite, whereas Figure 8B shows the diffractogram of apatite [$Ca_{10}(PO_4)_6(OH)_2$]. The samples are measured from diffractor angles 7° to 49.5° (2θ), in steps of 0.01°. Apatite is known for its microcrystalline structure and therefore lacking sharp and well defined peaks.

After acquisition of the diffractograms, the obtained patterns have to be interpreted for the estimation of the composition of the urinary calculus. The International Center Diffraction Data (ICDD) database, formerly known as Joint Committee on Powder Diffraction Standard (JCPDS) is often used as reference database for comparison of the diffractograms. Despite the availability of the ICDD reference database, the interpretation of the patterns still has to be performed by specialists with many years of experience. Therefore, Wulkan and associates al have developed an expert system (LITHOS) for the evaluation of X-ray diffractograms of urinary calculi (76).

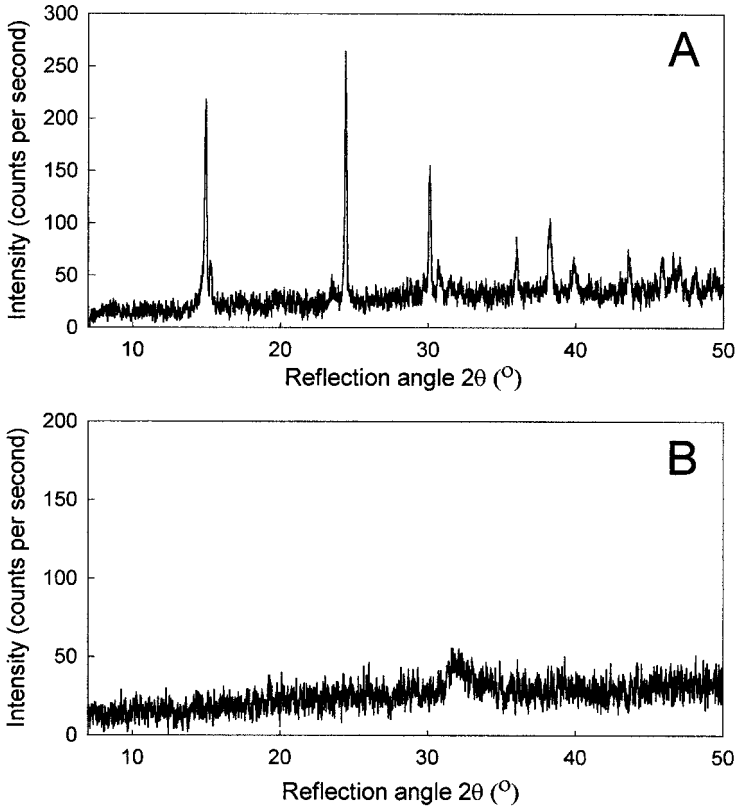



Figure 8. XRD spectrum of whewellite (A) and apatite (B). The data of the diffractograms were obtained from the Clinical Chemical laboratory of the University Hospital of Rotterdam, The Netherlands

X-ray diffraction and infrared spectroscopy both are well suited for the quantitative analysis of the atomic composition of urinary calculi (77). X-ray diffraction can detect crystalline components in low concentrations. Unfortunately, the quantitative determination of amorphous components by XRD may be problematic (see Figure 8B). This is especially true when amorphous substances are present in mixed stones. Moreover, the XRD apparatus, and ICDD reference library are very expensive. The XRD apparatus is potentially dangerous with respect to accidental exposure to X-rays and has no other application in the routine clinical chemical laboratory.

## 2.2. Infrared spectroscopy and sample handling techniques

*Introduction*

In analytical chemistry infrared spectroscopy (IR) is mainly used for the analysis of organic components. The qualitative assessment of organic components is performed for the identification of unknown compounds, or for the determination of the chemical structure of the components. In addition, IR analysis may be used for quantification of the components. IR spectroscopy is also known as vibration spectroscopy, since the spectra arise from transitions between the vibrational energy levels of a covalent bond of a molecule. The infrared spectrum, which ranges from 1 μm to 1000 μm, is part of the electromagnetic spectrum and is surrounded by the visible and microwave regions (Figure 9). The IR region may be further subdivided in the near infrared, the mid infrared and the far infrared regions (78).

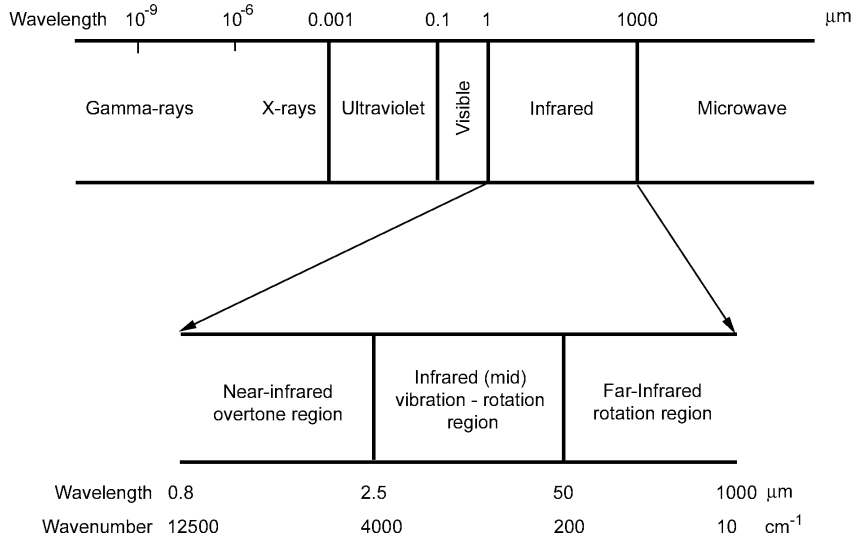



Figure 9. Infrared region of the electromagnetic spectrum.

The energy, associated with the radiation of the IR region, is sufficient to cause rotational and/or vibrational changes of the atomic bonds of the molecule. In order to absorb IR radiation the covalent bond of a molecule must undergo a net change in dipole moment as a consequence of its rotational or vibrational motion (79). According to the quantum theory, when a molecule absorbs IR radiation, a vibration transition occurs from the ground state to the first excited state ($V_0$–$V_1$). This occurs when the frequency of the radiation matches the natural vibrational frequencies of the molecule. Except for this first level transition, other transitions may also occur ($V_0$–$V_2$, $V_0$–$V_3$). In theory, a single absorption band should be observed for each transition level. Second and higher order transitions always give rise to weaker absorbances. The bands causing the higher order transitions are often called the overtone bands. The energy required for the transitions $V_0$–$V_1$, $V_1$–$V_2$, etc, are about equal. Therefore, the first overtone of a band is often found at wavenumbers two times the wavenumber of the first transition level ($V_0$–$V_1$) ± 20 $cm^{-1}$. The same is true for higher

overtones. For example, the first overtone band of the strong $CH_2$ band at 890 $cm^{-1}$ is found at 1780 $cm^{-1}$.

Infrared spectra of substances are characterized by three major properties, namely the number of bands of molecule in the spectrum, the wavenumber positions of the bands and the intensities of the bands.

*The number of bands.*

Except for bands found at different wavenumbers resulting from the various transition levels, different bands may occur as a result of the freedom of movement of the single atoms in the molecule along their X, Y and Z-axis. The relative positions of the atoms in a molecule are not fixed, but fluctuate as a consequence of different types of vibration. Apart from rotation, normally two major modes of vibration occur, namely stretching (Figure 10A) and bending or deformation vibrations (Figure 10B). Both vibrational modes can be further classified in symmetric and asymmetric types of vibration. Carbon dioxide ($CO_2$) is a symmetric molecule. Therefore, no change in the dipole moment will occur by symmetric stretch vibration, and as a consequence no symmetric stretch band will be seen in the IR spectrum of a $CO_2$ molecule (Figure 10A) (80).

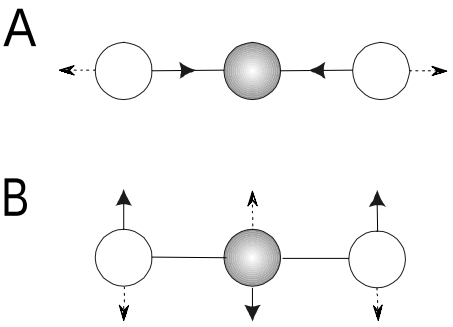



Figure 10. Symmetric stretch vibration (A) of $CO_2$. Infrared inactive.
Symmetric bending vibration (B) of $CO_2$. Infrared active.

Based on the transition levels and vibrational modes of the molecule, a great number of bands should be seen in the IR spectrum. In practice the number of observed bands is frequently much less because the symmetry of the molecules results in no dipole moment at a certain vibration frequency (see e.g. $CO_2$), the absorption intensity is too low to be detected, the energies of two or more vibrations are nearly identical, or the vibrational energy is beyond the wavenumber range of instrument.

*The position of the bands.*

For diatomic molecules it is possible to predict the theoretical position of the bands for the stretch vibrational modes. This can be performed by calculating the vibrational frequencies, by using a formula for harmonic oscillations (79). Usually there is a good agreement between the calculated and experimental values for the wavenumbers. However, in practice the specific groups rarely absorb at definite positions but occur

over a range (band) of wavenumbers, because the simple calculations do not take into account the effects arising from other neighboring atoms of the molecule. While these interaction effects may lead to uncertainties in the identification of the functional groups of a molecule, the combination of the bands of an IR spectrum is very important for a positive identification of a specific organic or anorganic component. For this reason, IR spectra represent one of the unique physical properties of organic and some anorganic components, with the exception of optical isomers.

The amount of energy to cause a change in rotational level is very small, which primarily occurs in the far infrared region (200–10 $cm^{-1}$). In this spectral region, which is mainly used for the determination of gasses, the absorption bands of the gasses are found by discrete, well-defined lines. Rotation is highly restricted in liquids and solids. Their vibrational bands are found in the mid IR (4000–200 $cm^{-1}$) and the near IR (12500–4000 $cm^{-1}$) regions. Because the energy required for deformation is much lower than for stretching, the deformation bands are found at lower wavenumbers than those for the stretching vibrations. In the near infrared spectral region, mainly weak absorption overtone bands of the OH, NH, CH and C=O groups can be found.

*The intensity of the bands.*

Just like in ultraviolet-visible (UV–VIS) spectroscopy, the intensity of an absorption band is usually expressed as the molar extinction coefficient (ε). This intensity is proportional to the square of the change of the dipole moment during vibration. Therefore, if no change in the dipole moment occurs (e.g. symmetrical stretch vibration of $CO_2$), no absorption band will be seen. The intensity of overtone absorption is frequently low and the bands may not be observed. Fortunately the intensity of carbon-carbon single bond stretching vibration is usually very low. Therefore, the majority of the bands, observed in an IR spectrum, arise from the substituent groups and not from the carbon skeleton of the organic molecules.

IR spectra are plots of the absorbance against wavelength, similar to the plots of UV–VIS spectroscopy. However, ordinarily the ordinate of IR plots is expressed in transmittance units (%), whereas the abscissa is expressed in wavenumbers. The wavenumber scale is the reciprocal of the wavelength scale and has the units $cm^{-1}$.

An advantage of the weak absorbance in the NIR region is that sample dilution is often unnecessary and that longer pathlengths may be used. For this reason NIR analysis is well suited for remote analysis (e.g. transcutaneous glucose measurement). Because of the limited number of functional groups that can be detected (OH, NH, CH and C=O) with NIR spectroscopy, more complex data handling routines are often necessary for quantitative analysis. Another specific problem of the NIR region is that the shorter wavelength areas are prone to excessive scattering, causing loss of light. The most important region for identification of organic components is the mid IR region. Today the majority of the analytical IR applications are confined to the IR region between 1700 and 400 $cm^{-1}$, because most of the functional groups have relatively sharp absorption bands in that area. With this region, which is also called the fingerprint area of the spectrum, quantification of the components may often be done with simple data handling methods. Some of the more complex data processing techniques used in IR spectroscopy will be described in more

detail in the next chapters. Structure analysis may be performed by comparing the bands from the spectrum with the approximate band positions of infrared absorption bands. This information may be obtained from spectroscopic software (e.g. Sadtler software, Bio-Rad Laboratories LTd, London, UK) or be found in tables and correlation charts in different publications (41;78;79;81). Figure 11 shows a correlation chart of the IR region between 2000–650 $cm^{-1}$.

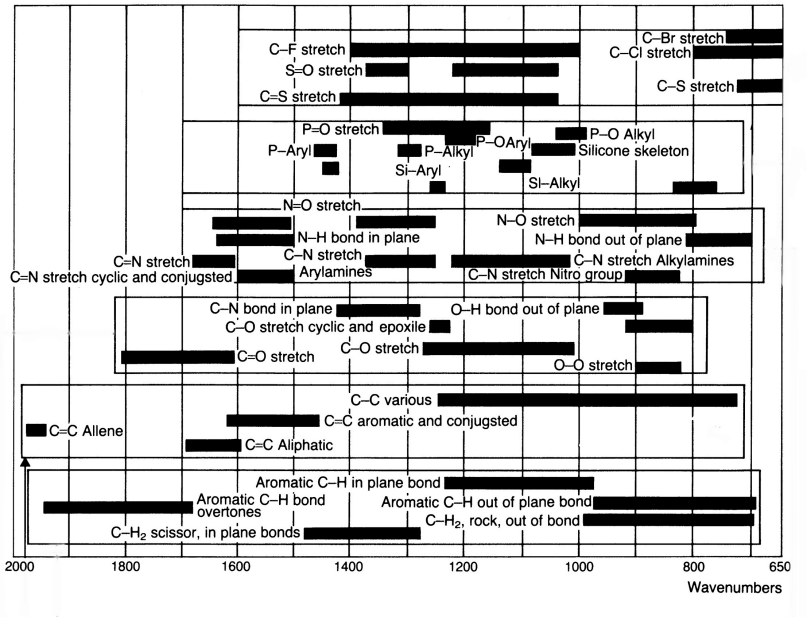



Figure 11. Correlation chart with a number of functional groups in the IR region from 2000–650 $cm^{-1}$.

*IR Instrumentation*

IR spectrometers have the same basic components as the instruments used for UV-VIS spectroscopy. They consist of three basic components: a source to provide IR radiation, a wavelength selector to disperse the source energy and isolate the required wavelength, and a detector to measure the intensity of the dispersed radiation. Interference filters are often used as wavelength selector for the NIR infrared region.

For the (mid) IR region often diffraction gratings are used as infrared monochromators. The energy of infrared sources is generally low. Using a narrower slit width of the monochromator to increase the quality of the spectral resolution will usually be accompanied by a decrease of the signal to noise (SN) ratio. Fortunately, weak spectra may be extracted from noisy environments by means of signal averaging. The SN ratio of a spectrum may be improved by a factor $\sqrt{n}$ by averaging n replicated spectra. This implicates, that averaging of 16 replicated scans gives a four-fold enhancement the SN ratio. Using conventional infrared spectroscopy, the resulting radiant power is recorded as a function of the radiant frequency, which is inversely related to the wavenumber. With this so-called frequency domain spectroscopy, the absorbance intensity is measured at each wavenumber or resolution element separately. Therefore, signal averaging of a number of

replicated scans, consisting of a great number of resolution elements (e.g. full spectrum scans), may be very costly in terms of time.

In contrast with conventional spectroscopy, Fourier transform spectroscopy measures all resolution elements of a spectrum simultaneously. Fourier transform spectroscopy is concerned with changes of radiant power with time, and is also called time domain spectroscopy. It is important to notice that the time domain spectrum contains the same information as the frequency domain spectrum. For that reason, the complex time domain and frequency domain spectra can be interconverted into each other by complex mathematical (Fourier) calculations. In order to obtain a measurable signal for the various wavelength regions of the spectrum in time domain spectroscopy, a signal-modulation (conversion) has to be employed. For this purpose, the Michelson interferometer (Figure 12) has been used extensively for the measurement of the infrared region.

This design of the device for modulation of the infrared radiation was first described by Michelson in 1891. The Michelson interferometer splits the radiation of an infrared source into two beams by means of a semi transparent mirror (beamsplitter) in such a way that the two beams with almost equal power are positioned at right angles of each other. The resulting twin beams are reflected from mirrors, one of which is fixed and the other of which is movable between position X and -X (see Figure 12).

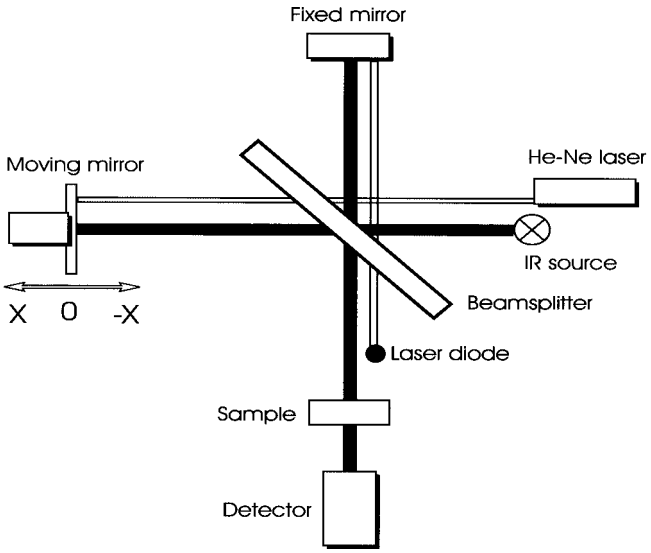



Figure 12. Michelson interferometer for signal-modulation in FT-IR spectroscopy.

These beams meet the beamsplitter again and half of each beam is directed toward the sample and detector. If an absorbing material placed in the sample compartment in these beams, the resulting interferogram (Figure 13) will carry the spectral characteristics of the analyte. An interferogram is a plot of the output power of the detector against the retardation. The retardation is the difference of the path length of the two beams. The actual conversion of the interferogram into a conventional infrared spectrum is very complex and

is done by computers. In modern FT-IR instruments, precise signal sampling is obtained by using a so-called laser-fringe system. It consists of a helium neon laser source, an interferometric system and a laser diode detector (Figure 12). This system gives highly reproducible and regularly spaced sampling intervals. The interested reader is referred to publications of Skoog et al. (79) and Griffith (82) for more information about FT-IR spectroscopy.

FT-IR spectrometers have several benefits over the conventional scanning spectrometers. The differences between both methods are summarized in Table 6. Today, most IR spectrometers employed for measurement in the mid IR region are FT-IR spectrometers. If only absorbance data have to be collected from one, or a limited number of wavenumbers, low cost filter or dispersive spectrometers may still be used.

Table 6. Comparison of conventional and FT-IR spectroscopy.

| Conventional | FT-IR |
|---|---|
| Lengthy scan times of full scans<br>Measuring each resolution element separately | Short scan times of full scans<br>Measuring all resolution elements simultaneously |
| Resolution non continuous over the wavenumber region | Resolution steady over the whole wavenumber region |
| Calibration of wavenumbers by external standards | Internal calibration by means of the laser beam |
| Sensitive to stray light | Insensitive to stray light |
| Low energy throughput | High energy throughput |

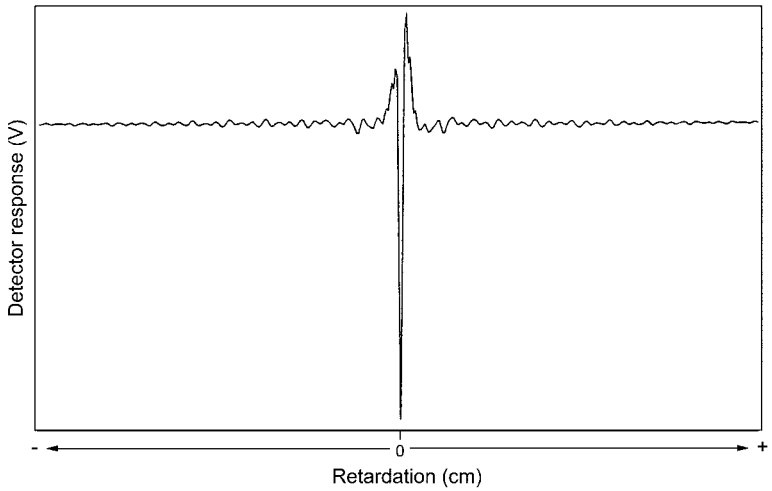



Figure 13. Interferogram of a continuous IR source.

Normal optical materials such as glass or quartz absorb strongly in the infrared region and therefore cannot be used. Quartz can however be used in the near IR region, because it is IR transparent from 5000–2760 $cm^{-1}$. Absorption cells, sampling devices and other optical parts of an (mid) IR spectrometer must be made of infrared transparent material in the region from 4000–200 $cm^{-1}$. The substances most commonly used as optical parts in IR spectroscopy are given in Table 7, together with their useful transmission ranges, refractive indices and relative hardness relative to NaCl.

Table 7. Physical properties of some materials for IR spectroscopy.

| Material | Transmission range | Refractive index at 2000 $cm^{-1}$ | Hardness relative to NaCl |
|---|---|---|---|
| NaCl | 5000 – 590 | 1.52 | 1.0 |
| KBr | 5000 – 340 | 1.53 | 0.4 |
| KRS-5 | 5000 – 250 | 2.37 | 2.2 |
| ZnSe | 5000 – 500 | 2.40 | 8.3 |
| Ge | 5000 – 600 | 4.01 | 160 |
| Diamond | 5000 – 10 | 2.40 | Very hard |

*IR sampling handling techniques*

IR spectroscopy may be used for the analysis of gasses, liquids, pastes, powders and polymer films. The IR sampling techniques may be subdivided in transmission, and reflection techniques. The sampling techniques appropriate for the sample categories that were used in our own studies are summarized in Table 8.

Table 8. Some sampling techniques.

| | Sampling techniques | |
|---|---|---|
| Sample category | Transmission | Reflection |
| Liquid (chloroform extracts) | Liquid cells | |
| Pastes (feces) | Liquid cells | ATR |
| Powder (urinary calculi) | KBr | ATR |

KBr, Potassium bromide; ATR, attenuated total reflection

Analysis of liquids using a liquid cell:

In mid IR spectroscopy water and alcohol are rarely used as solvents, because of their strong absorbance intensities. Furthermore, they are less suitable as solvents because of their interactions with the metal halide cell window materials that are often used. More commonly, organic solvents such as chloroform, carbon tetrachloride, and carbon disulphide are used. In mid IR spectroscopy, sodium chloride windows are often employed, because they are rather cheap, but must be handled with care because of their tendency to absorb moisture. Careful selection of the solvents must also be done to prevent unwanted interaction of the solvent absorbance bands with those of the components of interest.

In mid-IR, the pathlength of infrared liquid cells are normally much smaller (0.1–1 mm) than those employed in UV-VIS spectroscopy because of the relatively high absorbance caused by the organic solvents. The pathlength of the cells is often fixed or may be adapted

by removable spacers. The pathlength in NIR spectroscopy may commonly be much larger (e.g. 1 cm) when compared to mid IR spectroscopy.
We applied NIR spectroscopy, using a liquid cell with polyethylene windows and a pathlength of 1 cm for the determination of fecal fat. We also applied mid-IR spectroscopy for the determination of fecal fat, by measuring chloroform extracts of feces with a liquid cell with NaCl windows and a pathlength of 0.1 mm (both studies are described in Part I).
Sometimes unwanted interference fringes may be observed in the spectra, which are caused by internal reflection of the light by the two cell walls. These fringes may occur when the refractive index of the cell window material and solvent differ too much. The fringes are observed as regular sinusoidal curves, superimposed upon the spectrum. Interference fringes may be beneficially used for the prediction of the exact pathlength of a cell (78).

Analysis of powder with KBr tablets:
The composition of solid samples are often measured by using the potassium bromide (KBr) disk technique. This technique is especially applicable for crystalline material. The solid sample is totally grinded, manually by using an agate pestle and mortar, or by using a mechanical mill. About 0.5–2.0 mg of grinded sample is thoroughly mixed with 100–200 mg dry KBr. This mixture is applied to a special pellet die. The KBr disk is produced by applying about 10 kbar pressure to the pellet die construction with a hydraulic press. The resulting disk is typically 13 mm in diameter and has a thickness of less than 1 mm. The transparent disk is placed in a special tablet holder and measured in transmittance in the mid IR region.
We applied the KBr sampling technique for the determination of the composition of urinary calculi (see Part II).
Also when this technique is used, small inference fringes may be observed in the spectrum. These fringes may be prevented by the preparation of thicker disks (≥ 1 mm), by using more KBr. Spectral distortion may also be caused when the particle size of the sample is too large. In theory, the particle size of the sample should be less than the lowest wavelength to which it is exposed (2.5 μm = 4000 $cm^{-1}$). If the sample is not sufficiently grounded, the spectrum may contain distorted bands and often sloping backgrounds caused by loss of energy by scattering. In such cases, prolonged grinding of the sample may enhance the spectral resolution and sloping background effects. Another spectral distortion may occur when the refractive indexes of the sample and the halide (e.g. KBr) differ too much. In practice this may not be a serious problem because the refractive index of most organic and inorganic components are almost similar to the refractive index of KBr (1.5). In case of serious band distortions, alternative halides (e.g. caesium iodide) may be used for the preparation of the disks. Both distortions due to a combination of the particle size effects and refractive index are known as the Christiansen effect.

Analysis of pastes or powders with the attenuated total reflection technique:
Attenuated total reflection (ATR) is a relatively new technique that can be applied to a wide range of sample materials, such as liquids, pastes and powder (83;84). This reflectance technique can be applied without sample pre-treatment. ATR is based upon the fact that IR radiation propagated through an optically dense medium is reflected when it arrives at an interface with a less optical dense medium. The reflection becomes complete when the angle of incidence is greater than a certain critical angle. It has been shown that, due to the

wave nature of radiation, the reflection does not occur directly at the interface of the two media, but after penetrating a small distance into the less dense medium. The depth of penetration of the radiation is a function of the wavelength of the radiation, the angle of incidence at the interface and the refractive indices (RI) of both media. Figure 14 shows the relationship between the wavelength (wavenumber) and the depth of penetration in a sample with a refractive index of 1.0, using a ZnSe ATR crystal with an angle of incidence of 45°. ATR can only be effective if the difference of the RI between the sample and the ATR crystal is large enough ($\Delta$ RI $\geq$ 1). Table 7 shows the RI values of some of the ATR crystal materials, whereas most sample materials have RI values close to one. Today, the majority of ATR sampling accessories are flat crystal plates with fixed angles (45°, 30° or 60°) employing about 10 reflections (Figure 15). The incident light from the IR source is directed with a fixed angle onto the entrance of the crystal plate by means of a plane mirror system. At the gate of the ATR crystal plate the retarded radiation is directed towards the detector by the same kind of mirror. Because the source radiation is not directed by lenses, the whole sample upon the crystal plate is flooded with the radiation. Absorption and attenuation will take place at each of the internal reflections. The number of reflections may be increased or decreased by obtaining thinner or thicker ATR plates, by changing the length of the ATR plate, or by changing the angle of incidence. We applied the horizontal ATR technique for the determination of fecal fat (see Part I).

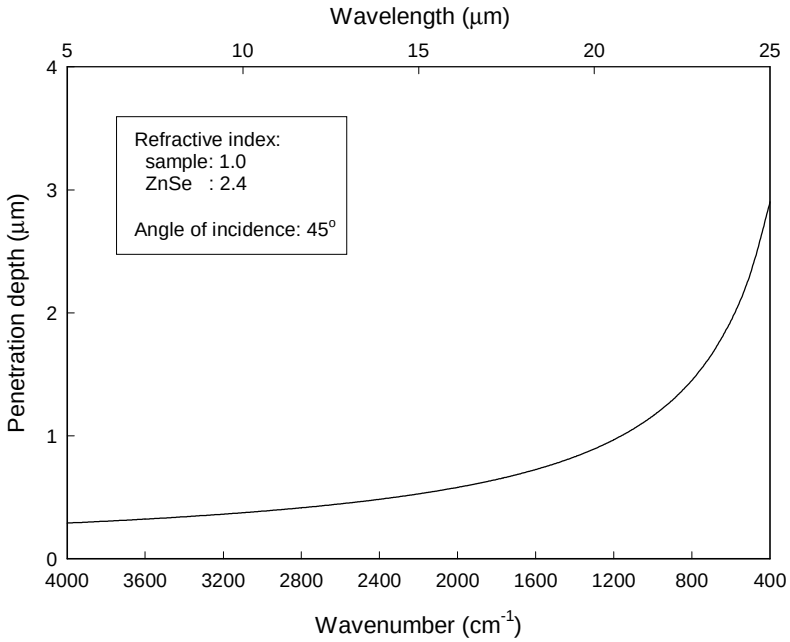



Figure 14. Relation between the depth of penetration and the wavelength using ZnSe as ATR crystal.

Horizontal flat plate ATR crystals require a reasonable amount of sample material. More recently, micro-sampling ATR devices became available (Figure 16). Single reflection micro-ATR devices are usually equipped with diamond crystals and enable the measurement of very small amounts of sample material (liquids, pastes and solids). The

practical benefits of diamond as ATR crystal are its strength and chemical inertness. The intrinsic hardness of diamond, as well as the small sampling area of the micro-ATR (diameter usually < 1 mm) enables the application of high pressure to the sample on the crystal, without the risk of crushing the crystal. This pressure is needed to achieve sufficient sample contact between the solid sample and the sense area of the crystal.

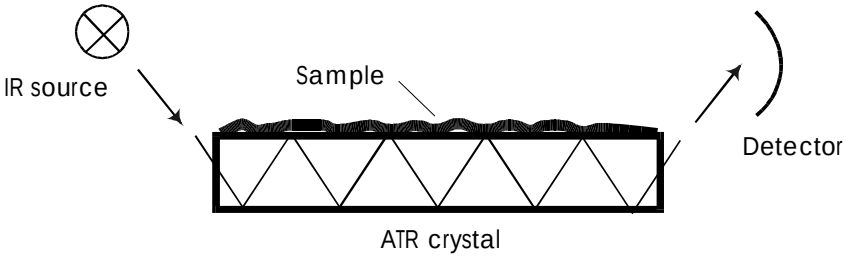



Figure 15. Multiple reflection horizontal ATR system.

For quantitative analysis, it is important that the pressure applicator of the micro-ATR is equipped with a pressure restraint in order to prevent irreproducible outcome as a result of the pressure-induced phase transitions (85). These transitions may result in changes of the polymorph distribution of the sample components. On the other hand, this pressure dependency is sometimes also beneficially used to obtain an extra spectroscopic dimension (see paragraph: Diagnostic applications of IR spectroscopy). We have applied micro-ATR (Golden Gate) for the determination of the composition of urinary calculi (see Part II).
ATR is usually not practical in NIR spectroscopy because the very small penetration depth causes weekly absorbing bands in NIR.

The ATR spectra are similar, but not identical to ordinary transmission spectra. The spectral bands will be the same, but their relative absorbances will differ because the depth of penetration varies as a function of the frequencies (Figure 14).

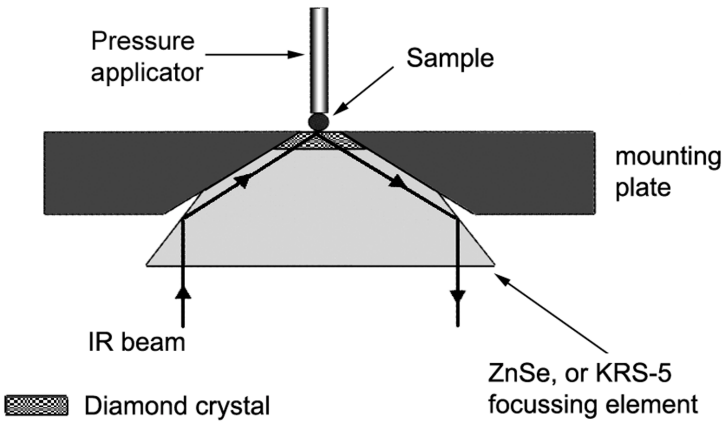



Figure 16. Single reflection micro-ATR sampling device.

ATR sampling may be applied to almost any sample substance. Infrared analysis normally avoids the use of water as solvent because of its strong absorption in the mid IR region and for the reason that many of the materials used in IR analysis are typically non resistant to water. However, by using horizontal ZnSe ATR plate systems, quantitative analysis of samples dissolved in water is possible in mid IR. Liquids and pastes can normally be measured very reproducibly with ATR, but solid samples and powders may cause irreproducible results as a consequence of poor sample contact, as was described above. Therefore, the sample contact of the solid samples must be enhanced by clamping the sample onto the ATR plate with a pressure applicator.

The list of the preceding sampling devices applied in IR spectroscopy is far from complete. Since the 1980s many sampling techniques evolved and became available to handle almost any sample. More information about these accessories, such as specular reflectance, diffuse reflectance (DRIFT), folded path cells for gas measurements, photoacoustic detection, IR microscopes (86) for micro-sampling and hyphenated techniques such as GC-IR. may be found in literature (87;88). More general background information about IR spectroscopy may be found elsewhere (78-80;82;89).

*Clinical and biomedical applications of IR spectroscopy*
IR spectroscopy has been employed for the analysis of several analytes in different biofluids and solid biosamples. IR spectroscopy has been applied for the analysis of pathological samples, for diagnostic applications and for non-invasive in vivo monitoring.

Analysis of patient samples
FT-IR spectroscopy has been used for fast multi-component analysis of different analytes in biological sample materials. Fecal fat, sugar, nitrogen and water contents have been quantified simultaneously using NIR reflectance spectroscopy (90). The fecal fat content has also been analyzed using mid IR spectroscopy (23). FT-IR spectroscopy has been employed for the determination of the composition of urinary stones (31) and human gall stones (91). Shaw et all (92) described an IR spectroscopic method for the simultaneous quantification of serum concentrations of total protein, albumin, triglycerides, cholesterol, glucose urea, creatinine and uric acid. The serum samples were spread as a thin film onto an IR-transparent material. After drying of the serum film, the samples were measured in transmission in the mid IR region. Other authors described a method for the IR determination of the lecithin/sphingomyelin ration in amniotic fluid (93).

Diagnostic applications of IR spectroscopy
IR spectroscopy has been applied for the $^{13}C$ urea breath test for diagnosis of Helicobacter pylori infections inside the stomach (94;95). The Helicobacter pylori bacteria produce large quantities of urease. The test exploits the hydrolysis by urease of orally administrated $^{13}C$-urea into ammonia and $^{13}CO_2$, which diffuse into the blood. The $^{13}CO_2/^{12}CO_2$ enrichment in breath is measured using an isotope-selective nondispersive infrared spectrometer.
Combinations of visible microscopy and IR spectroscopy are used in the development of methods for the diagnosis and identification of cancer cells. FT-IR microscopy (FT-IR-MC) is a hyphenating technique of an optical microscope and an infrared spectrometer. It allows visual and infrared assessment of different spatially localized parts of the sample on a microscopic scale. FT-IR-MC is very sensitive in visual and infrared transmission and

reflection. Healthy and cancerous cells have different infrared spectra. These differences are often based on changes in the DNA/RNA complexes and differences in lipid cellular membranes. The changes involve the phosphate, the C–O stretching bands and the CH stretch region. Another remarkable difference is the dissimilarity of the pressure dependence of the $CH_2$ and C=O stretching modes of the normal and cancerous cells (96). Pressure dependency is a normal functional relationship between an increasing pressure (0 – 20 kbar) applied to the sample and the spectral parameters (e.g. frequency, intensity, band shape) (97).
A recent study, using FT-IR microspectroscopy, described the spectral differences between healthy and cancerous human lung cells (98). Another study described the improved discrimination by IR spectroscopy, between different types of tissue structures of human melanoma and colon carcinoma (96). FT-IR spectroscopy was also applied to samples of normal, and malignant and dysplastic cervical smears (99;100) .

Noninvasive in vivo monitoring in IR spectroscopy
Usually, NIR instrumentation is used for metabolic monitoring, because of its longer pathlength in relation to the mid IR region and availability of fiber-optics for the NIR region. In vitro monitoring is inherently invasive, causing relative lengthy turn-around times because the samples usually have to be analyzed in a central laboratory. In case of in vivo measurement, which is often not invasive, the analysis can be performed near the patient. Unfortunately, the accuracy of the non-invasive in vivo IR measurements does not yet match the accuracy of in vitro measurements (101). Therefore, the in vivo measurements are best suitable for the detection of a trend of change in one patient. Except for non-invasive patient monitoring in intensive care and surgery units, in vivo monitoring might be used for patient self-monitoring.
A well-known example of in vivo monitoring is tissue oxygenation monitoring. Quaresima and associates described two approaches to non-invasive NIR spectroscopic measurements of cerebral hemoglobin oxygen saturation (102). Much effort has been put in near infrared reflectance spectroscopy for non-invasive monitoring of blood glucose. In some of the studies, glucose was measured through the surface of the finger (103), while others measured through the oral mucosa (101). Both kinds of transcutaneous glucose measurements were performed by using fiber optics and diffuse reflectance probes (104;105).

From the previous descriptions we conclude that infrared spectroscopy is very useful for the determination of analytes in complex biochemical samples. The instrumentation is not very expensive, and there is a large number of sampling devices and optics available for invasive, as well as noninvasive measurement of almost any kind of bio-sample. In general, IR spectroscopy can save time and expense in terms of sample preparation.

# 3. Chemometrics

## 3.1. General

Today, clinical laboratories can produce an almost unlimited number of test results on body fluids from each patient sample submitted to the laboratory. Consequently, the clinical and hematological laboratories generate a lot of numerical information, which contributes to the patients' database. Sometimes, the information of the laboratory tests is redundant, what not only may lead to saturation, but even to a decline of information. Furthermore, the interpretation of the results obtained from some of the analytical methods used in the clinical and hematological laboratories has become more and more complex. This is caused by for example the improved separation power of the analytical instrumentation (e.g. capillary GC) and by the increased demand for measurement of analytes in authentic sample material, which is often a complex sample matrix containing many interfering substances. Both phenomena, the increased amount of analyte information obtained from a single sample, and difficulties with the interpretation of the test results obtained from a complex sample matrix has led to the development and application of statistical and mathematical methods in the past decades.

These statistical and mathematical methods have resulted in new analytical applications, often by omitting the otherwise imperative sample pretreatment. Svante Wold was the first investigator who applied the so-called 'chemometrics' to organic chemistry applications. Nowadays, many of the statistical and mathematical methods developed for analytical chemistry applications, are referred to as 'chemometrics'. Chemometrics is concerned with the application of mathematical and statistical methods, as well as those methods based on mathematical logic, to extract useful information from chemical measurements (106). Similar disciplines have emerged in other fields of science, such as biometrics, psychometrics, econometrics or medicometrics. Sometimes, the chemometric techniques are applied to sub-fields of analytical chemistry, such as qualimetrics, which is concerned with the use of chemometric methods to improve the quality control and quality assurance, or pharmacometrics in which the methods are used in the synthesis, analysis and formulation of pharmaceuticals.
Medicometrics has relationship to the medical sciences, because the methods evaluate clinical and laboratory test results from patients (107). Medicometrics is not involved in the administration of patient medical records, but e.g. in the extraction of useful information from chemical and hematological data (108;109), automated pattern recognition of signal processes such as electrocardiograms (110) and electroencephalograms (111), pattern recognition of digitized microscopic images of urothelial cell carcinoma (112) or malignant gastric cells (113), simulation studies of arm movements (114), validation of test results of patient samples by means of a rule based system (115) or a statistical method (116), and other application areas.

During their development, the medicometric methods applied to laboratory test results always need specific background information about the patients (diagnosis, medication, gender, etc.). Chemometrics on the other hand can be performed within the walls of the laboratory, without needing any patient background information. It can be used to optimize the analytical, or post-analytical processes.

Our own studies with respect to the analysis of urinary calculi and fecal fat and concerned with difficult quantitative interpretations of infrared spectra, made use of chemometrics. Apart from these studies, chemometric methods are applied to a great number of analytical techniques, such as gas chromatography, high-pressure liquid chromatography (HPLC), mass spectrometry, infrared spectroscopy, etc. The chemometric methods may be subdivided in the following categories: statistics (e.g. method validation, sampling strategies, detection limits, etc), optimization (minimization or maximization of a function of one or more independent variables, e.g. mixture designs, liquid/liquid extractions, HPLC parameters, etc.), signal processing (digital filtering, smoothing, background correction, domain transformations, trend analysis, image analysis, etc), resolution (e.g. identification of peak patterns in unresolved regions), parameter estimation (curve fitting and mathematical modeling of chemical properties of e.g. spectral band shapes), structure-activity relationships (relation of the molecular structure to its chemical, physical or biological properties), pattern recognition (classification of an unknown into one of a set of predetermined classes), artificial intelligence (e.g. automated chemical workstations, including a scheduler for the initiation and monitoring of parallel experiments), calibration (relating or modeling measured responses to the composition of a set of analytes), exploring chemical data (for understanding/finding underlying phenomena), and library searching (identification of unknowns and qualitative analysis of mixtures). Our own studies only make use of calibration techniques (partial least squares regression and neural networks) and simple library search algorithms. These methods are described in the following chapters. Interested readers in other methods may find reference to a large number of articles concerning the development and application of chemometric methods applied to each of the above-mentioned categories, in a series of review articles (106;117-119). A general introduction to chemometrics may be found in a book of Massart et al (120).

*Data structures used in chemometrics*

Chemometrics is concerned with the extraction of 'useful' information from measurements. The characteristics of these data are generally stored in one or two data sets. Some of the chemometric techniques will, however, work on only one data set. Such a data set is normally referred to as the X data and they may contain the more easily accessible variables, such as spectroscopic data (NIR, UV-VIS), chromatographic data (GC, HPLC, MS), process measurements, image analysis data, etc. Table 9 shows an example of a data set, containing absorbance data of some patient samples. The data set consists of a number of objects (patient samples). Each object is a set of values such as absorbances, measured at different variables (wavenumbers). In this data set each line represents one object and each column represents one variable. Statisticians have found matrix mathematics a very useful concept for the formulation of data sets, because they permit extremely efficient and accurate calculations for carrying out multivariable analyses on large data sets. Therefore, chemometricians always use matrix formulations for there data sets. For this reason the data set of table 9 would normally be defined as matrix X.

Table 9. Data set X containing spectroscopic data (Absorbances).

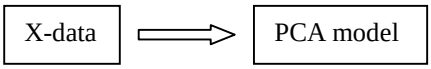
| Patient sample | Wavenumber ($cm^{-1}$) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 2000 | 1996 | 1992 | 1988 | 1984 | 1980 | 1976 | 1972 | …… |
| 1 | 1.8 | 1.9 | 2.0 | 1.8 | 1.6 | 1.2 | 0.8 | 0.4 | |
| 2 | 1.7 | 1.9 | 2.3 | 1.9 | 1.4 | 0.8 | 0.5 | 0.3 | |
| 3 | 1.2 | 1.4 | 1.2 | 1.0 | 0.8 | ….. | | | |
| ……. | | | | | | | | | |

A matrix is defined as a rectangular array of numbers. For example the data of table 9 would look like:

$$X = \begin{bmatrix} 1.8 & 1.9 & 2.0 & 1.8 & 1.6 & 1.2 & 0.8 & 0.4 \\ 1.7 & 1.9 & 2.3 & 1.9 & 1.4 & 0.8 & 0.5 & 0.3 \\ 1.2 & 1.4 & 1.2 & 1.0 & 0.8 & ….. & & \end{bmatrix}$$

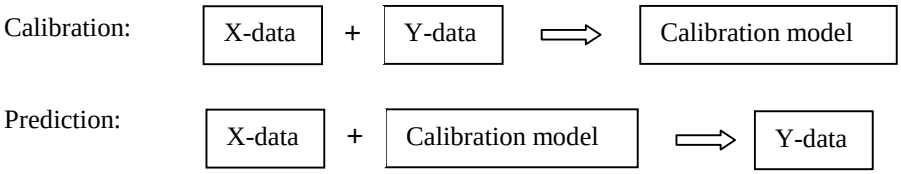
The dimension of a table is often defined as $X_{N\times K}$ in which N is the number of rows (1..N) and K the number of columns (1..K). In case of the example, the dimension of the matrix X would be $X_{3\times 8}$. Matrices may contain only one row, or column. These matrices are normally referred to as a row vector (e.g. $x'_{(1x4)}$ = [-1 15 3 -2]), or a column vector (e.g. $x_{4x1}$). As mentioned before, matrices are used in almost any kind of statistical calculation. More information about mathematical matrix manipulations, such as matrix inversion, multiplication, transposition, etc. can be found elsewhere (121;122).

By definition, the chemometric techniques are applied on matrices for the extraction of information from chemical measurement data. In case of chemical problems where only a single (X) matrix is available, principal component analysis (PCA) is one of the most commonly used techniques for this purpose. This technique extracts systematic variation from the data by turning the interrelated X-variables into unrelated ones. This is performed in such a way that the original dimensionality of the problem may be reduced without losing much of the information (123). The so called data structure modeling, or data decomposition of PCA can be used for e.g. variable reduction by selection of the most important variables, for finding patterns, or for the classification of variables or groups of objects (samples) that are similar. PCA can be schematized as follows:




Except for the analysis of one X-matrix, chemometrics is also often involved in the determination of the relation of the X-matrix with another matrix. The other matrix is often referred to as the Y-matrix. The Y-matrix may be a vector or a matrix containing variables such as the chemical concentrations, product qualities, etc. In order to find a useful relation between the two matrices it is necessary e.g. to determine the regression relationship between the two data sets. In this case the two data sets contain 'known data', also referred to as the training set, to make a model of the relationship. This process is called calibration. The calibration model (comparable with regression coefficients of a regression line) may be

used to predict 'unknown' Y-values from the measurements of new X-variables. Both calibration and prediction is schematized as follows:




Because most of the post-analytical quantifications of our studies described in Part I and II were based on multivariate calibration and prediction, the next two chapters will describe some of the multivariate chemometric quantification methods in greater detail.

## 3.2. **Multivariate calibration methods**

### 3.2.1. **PLS regression**

*Introduction*

Generally speaking, every mathematical calibration model has a structure part representing the systematic variation and a residual part representing the difference between the data and the structure (DATA = STRUCTURE + RESIDUALS). The mathematical description of the well known standard linear regression equation is: $y= b_0 +b_1x + e$, in which $b_0$ is the intercept, $b_1$ the slope of the regression line and e the residuals. This formula may be rewritten as $y = Xb + e$, in which Xb represents the structure. Good calibration modeling requires attention both to the structure and the residual parts. The prior aim to calibrate is to determine a function f( ) that allows quantitative predictions of Y (e.g. one or more concentrations) from X (e.g. measured absorbances):

$\hat{Y} = f(X)$ in which $\hat{Y}$ is a matrix with the predicted y-values and X are the predictor variables

The accuracy of the outcome of this function will only be good enough if there is sufficient correlation between X and Y. Although this is not the primary aim of their studies, chemometricians should always be concerned about the causal relation of their observations, in order to understand what the calibration data mean. The mathematical methods for simple linear regression and sophisticated multivariate calibration methods such as PLS, do not require causal understanding before starting calibration. This lack of insight can be compensated by the choice of empirical, but well considered calibration objects (training data). The training samples should be collected by selection of a set of representative samples with sufficient variability, in order to take care that the training data will span the whole concentration space. Further understanding about the X-Y relation can be obtained during the subsequent calibrations and predictions, by studying the structures and residuals of the objects (124;125). Sometimes it may help if some causal relationship is known in advance. In such cases, applying a proper transformation function could e.g. linearize known nonlinearities before calibration.

Figure 17A shows a part of an IR spectrum where it is presumed that the absorbance band (x) at wavelength 1320 nm accounts for the entire information about the concentration (y) of a certain component. In this case we can calibrate and predict the concentration of that component entirely on the basis of the absorbance of that band. We applied linear regression in our own studies on IR spectra obtained from petroleum ether and chloroform extracts of stool samples, for the quantitative prediction of fecal fat (see Part I).

A well-known problem in clinical chemistry is interference or the lack of selectivity. These interferences, or chemical matrix effects, may occur from other chemical constituents (e.g. interaction by overlapping IR bands), from physical phenomena (e.g. light scattering) or from the measurement process itself (e.g. temperature variation during measurement). Traditionally, interferences had to be removed physically, e.g. by extraction, filtering, or centrifuging, to ensure selectivity and in order to ensure linearity within the narrow range of the instrument scales. Unfortunately, this is time consuming and expensive. Today, there is a general tendency to measure samples with a minimum of sample pre-treatment (purification), but this often results in loss of selectivity. However by applying modern chemometric multivariate calibration techniques, interferences and non-linearity are often less of a problem.

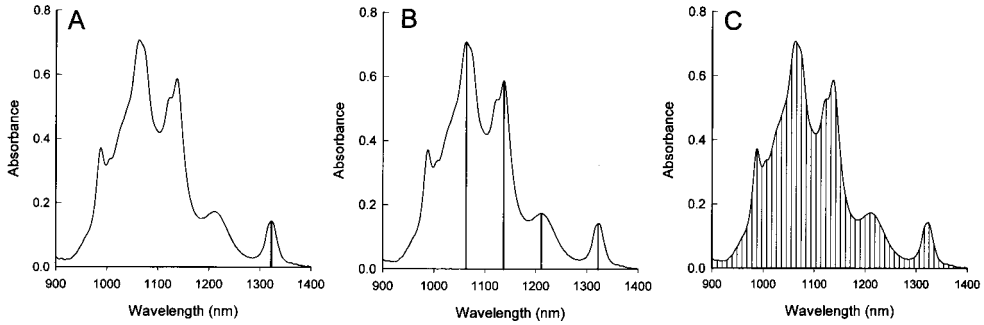



Figure 17. Part of an IR spectrum showing the absorbance band(s) at one or more wavelengths that are used for linear regression (A), multiple linear regression (B) and partial least-squares regression (C).

The calibration models are generally divided in the following distinct ways:

- Inverse calibration. As described before, the major purpose of calibration is prediction, not causal modeling. This is therefore called the forward direction from X to Y, with the predictor formula $\hat{Y} = f(X)$. The predictor can be obtained by regressing the calibration data Y on X, using the following model:

  $$Y = XB + F$$

  where X represents e.g. the spectra (absorbances), Y the concentrations of one or more analytes of a sample, B the regression coefficients and F the residuals. Standard linear regression and multiple linear regression (MLR) models used for prediction of the analyte concentrations are typical examples of inverse calibration. A brief description of MLR will be given later.

- Classical calibration. This is the traditional way to present the functional relationship between X and Y. This relationship resembles the causal structure of most analytical applications. In this case the predicting variables X (e.g. absorbances) are caused by the analytes Y (e.g. concentrations), and can be described with the predictor formula $\hat{X} = g(Y)$. The Beer's law in spectroscopy is a well-known example of this last relationship, namely:

  $A = \varepsilon cl$ (A, Absorbance; c, concentration; ε, molar extinction coefficient; l, pathlength)

- Regression on latent variables. The predictor formula of this kind of models is adapted from the domain of inverse calibrations and is generally described as $\hat{Y} = f(X, U)$ in which U are the unmeasured phenomena. In this relationship both X (e.g. absorbances) and Y (e.g. concentrations) are influenced by unmeasured interferents U. In this case, the concentration is directly predicted from the absorbances, avoiding explicit determinations of the component and interference concentrations. In case of IR spectroscopy, the interference problem may be solved by measuring the absorbances at several different spectral wavenumbers.

  To solve the predictor $\hat{Y} = f(X, U)$, the so-called 'regression on latent variables models' is used as calibration model. In general these models are described as follows:

  $T \Leftarrow f_1(X)$ (T, represents systematic structure, possibly unidentified)
  $X \Leftarrow f_3(T) + E$ (E, residuals)
  $Y \Leftarrow f_2(T) + F$ (F, residuals)

  Both Principal Component Regression (PCR) and Partial Leased Squares (PLS) regression belong to this class of models and will be described in one of the following paragraphs. More information about inverse and classical calibration may be found elsewhere (126).

Before describing PLS regression in more detail, some background information is needed about MLR, PCA and PCR.

*Multiple linear regression*

Classical MLR analysis deals with the estimation of the conditional mean of a random variable y from several X-variables (see Fig. 17B), rather than from a single x as in standard linear regression. The basic equation relating to these variables may be written as:

$$y_i = \beta_0 + \beta_1 X_{i1} + \ldots. + \beta_k X_{ik} + e_i$$

This equation describes that the y-value of the *i-th* object (individual, or sample) is a function of k+1 regression coefficients ($\beta$s) and k independent X-variables of the *i-th* object. Furthermore, the equation describes the residual error $e_i$ of the *i-th* object. The coefficient $\beta_0$ represents the systematic offset, whereas the other $\beta$ coefficients express the rate of change of the respective X-variables. The variable y is usually referred to as dependant, because its value is predicted on the basis of the known values of the independent X-variables.

The matrix (X) and vectors (y, e and β) of the previous model may be described as follows:

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \qquad X = \begin{bmatrix} 1 & X_{11} & X_{12} \cdots X_{1k} \\ \vdots & \vdots & \vdots \quad \vdots \\ 1 & X_{n1} & X_{n2} \cdots X_{nk} \end{bmatrix}, \qquad e = \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix}, \text{ and } \quad \beta = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_k \end{bmatrix} \text{ in}$$

which n is the number of objects (samples) and k is the number of X*s*. The previous equation may be written in a simple matrix notation: $y = X\beta + e$. To obtain reliable predictions from the calibration model, it is important that some assumptions are fulfilled during calibration, that is: the n residuals (e) must be independent and must follow a multivariate normal distribution with constant variance [homoscedastic] (127). If the number of objects (n) is larger than the number of variables (k), the regression coefficients of the previous calibration model can be estimated with the following matrix formula: $\hat{\beta} = (X'X)^{-1}X'y$ (121)[1]. The y-values of unknown samples may be predicted by using the estimated regression coefficients with the following formula: $\hat{y} = X\hat{\beta}$.

When applying MLR, it is important to work with a number of predictors (X) as small as possible. This is needed due to the principle of scientific parsimony, to obtain an optimal n/k (object/variable) ratio and also because the incremental information content of the new variables is often low as the measurements (e.g. absorbances) tend to overlap in content because of their possible intercorrelations (see e.g. Fig 17C). Since MLR is a mathematical maximization procedure, there is a considerable opportunity for capitalization of chance. Therefore, an n/k ratio greater or equal to 15 is needed in order to obtain a reliable regression equation and to provide sufficient reproducible predictive power. As a rule of thumb, it is recommended to select those predictors that highly correlate with the dependent variable (y), but that have low intercorrelations. Most statistical computer programs (e.g. SPSS) for MLR regression contain the three most popular procedures for selection of a good set of predictors, namely: forward, backward and stepwise selection (127;128). Another less commonly used selection procedure is the all possible subset regression procedure, which computes the multiple correlation coefficients ($R^2$) and regression equations for all possible subsets of predictor variables (128).

In practice, a problem called collinearity or multicollinearity may occur when some of the X-variables are redundant, because they are highly intercorrelated. The term collinearity is used to indicate that one or more of the predictors are approximately or exactly linear dependent of the others (121). As mentioned before, ideally a high $R^2$ would be obtained when each of the predictors is significantly correlated with the dependent y-variable and being uncorrelated with each other, so that they are able to predict different parts of the variance of y. Unfortunately, in practice this does not often occur, since almost all so-called 'independent' X-variables are intercorrelated to some degree. The consequence of collinearity can be illustrated with the following two-independent variable regression example:

[1] $X'$, transpose of matrix X; $(X'X)^{-1}$, the inverse of the matrix product $X'X$

$y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + e_i$ In general, $\hat{\beta}_1$ and $\hat{\beta}_2$ are calculated as follows:

$\hat{\beta}_j = c_j \left[ \frac{1}{1 - r^2(X_1, X_2)} \right]$ for j=1 and j=2, $c_j$ is a value depending on the

data and $r^2(X_1,X_2)$ is the squared correlation coefficient between both predictors. The term between the square brackets is also referred to as the variance inflation factor (VIF). From this equation it follows that the estimated regression coefficients become indeterminate if the correlation coefficient of $X_1$ and $X_2$ is one (VIF = 1/0 = ?). Thus, for the MLR solution, multicollinearity in X may have a detrimental effect on the computed coefficients of $\hat{\beta}$ and render them useless for prediction of $\hat{y}$. Fortunately, certain kinds of collinearity such as those involved in polynomial regression can be expressed as scaling problems and therefore can easily be resolved by subtraction of a constant (centering). In e.g. IR spectroscopy it is sometimes necessary to retain a great number of X-variables in order to get calibration models that include all spectral information and in order to stabilize the predictions against noise. In such cases, a great number of consecutive absorbances is used for the calibration and prediction of the component(s) of interest (Fig. 17C). PCA is an alternative computational method for rank reduction to avoid the impasse created by collinearity or near collinearity and forms the basis of the multivariate calibration methods that will be described in the following paragraphs.

*Principal Component Analysis*

There are three common problems when we want to predict Y from X (126):

- Collinearity: There is interrelation and hence redundancy between the X-variables.
- Lack of sensitivity: No single X-variable is sufficient to predict y.
- Lack of knowledge: A priori information of the mechanisms behind the data may be incomplete or wrong.

The most important way of dealing with these problems is PCA, which essentially transforms the correlated X-variables into new uncorrelated ones. Besides creating uncorrelated variables, PCA is a general framework for 'rank-reduction' or 'data-compression'. The general characteristic of PCA are summarized and depicted in Figure 18. In PCA the original X-variables are treated equally, i.e. they are not divided into dependent and independent variables, as in regression variables.

The new variables of PCA are called the 'principal components' (or factors). To simplify the interpretation, the data of the original variables are normally scaled by subtraction of the sample mean from each observation (centering), thus obtaining e.g.:

$x_1 = X_1 - \overline{X}_1$, $x_2 = X_2 - \overline{X}_2$, …….. $x_k = X_k - \overline{X}_k$ (for k X-variables)[2]

Each principal component is a linear combination of the original X-variables:

| | |
|---|---|
| $t_1 = v_{11}x_1 + v_{12}x_2 + \ldots + v_{1k}x_k$ | $t_1$ is called the first principal component |
| $t_2 = v_{21}x_1 + v_{22}x_2 + \ldots + v_{2k}x_k$ | $t_2$ is called the second principal component |
| etc. | $t_k$ is called the *k-th* principal component |

[2] $\overline{X}$, mean of variable X

In matrix terminology the linear combinations (PCs) are described as: T = XV
(V, PCA coefficients).

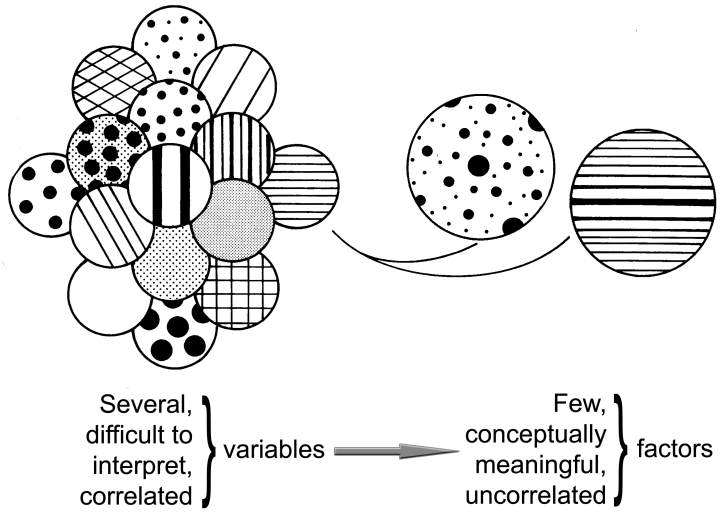



Figure 18. General purpose of principal component analysis.

One measure of the amount of information represented by each principal component is its variance. PCA is performed in such a way that the principal components are arranged in order of decreasing variance. Thus the most informative principal component is the first, and the least informative is the last (a variable with zero variance does not distinguish between the members of the population). The coefficients V of the PCA model are chosen in such a way to satisfy the following requirements:

1. The variance of $t_1$ is as large as possible [$Var(t_1) \geq Var(t_2) \geq ..... \geq Var(t_k)$]
2. The k principal components $t_1$, $t_2$, … $t_k$ are uncorrelated
3. $v_1^{'}v_1 = v_2^{'}v_2 = \cdots = v_k^{'}v_k = 1$ (the sum of the squared coefficients or scalars are one)

The normalization of the coefficients (see point 3) is needed as a constraint to prevent that the variances of the principal components ($t_{1..k}$) become arbitrarily large. A plot to illustrate the transformation of bivariate hypothetical data to principal components is given in Figure 19. This figure shows the scatter plot of the original variables $X_1$ and $X_2$ (Fig 19A) and the centered variables $x_1$ and $x_2$ ($x_i = X_i - \overline{X}_i$, i=1, 2) (Fig. 19B). The probability ellipse describes the relation between $x_1$ and $x_2$. The straight (dotted) line coinciding with the longest axis of the ellipse is called the first principal axis of the ellipse, and it is not surprising that the projection onto this axis is identical to the first principal component ($t_1$ in Fig. 19C). The second coordinate axis of the new coordinate system is uniquely defined by the following two conditions: it has to pass through the origin of the ellipse ($x_1 = x_2 = 0$), and it has to be perpendicular or orthogonal (uncorrelated) to the first axis. This second principal axis of the ellipse is the second principal component ($t_2$). Figure 19C shows the new coordinate system with the first and second principal components. The data points in

this plot are usually called the factor scores. These scores express the relation between the objects and are the projected locations of the objects on the components. In case of higher dimensions, the ellipse is replaced by a sphere (k=3) or a hypersphere (k>3).

Mathematical treatment of PCA consists mainly of the computation of the so-called *eigenvalues* of the covariance matrix (or correlation matrix). The *eigenvalue ($\lambda_i$)* is the variance of the principal component ($t_i$). The eigenvalues of all principal components always add up to the total variance of the k original independent X-variables ($ss^2_{total} = \sum_{i=1}^{k} \lambda_i$). Another characteristic of PCA is the so-called loading. PCA actually performs a redistribution of the variance of the original X-variables. The coefficients (v) of the principal components (linear combination of the X-variables) are usually transformed to factor loadings by dividing the coefficients by the square root of the corresponding eigenvalue of the component. Thus, if $v_j' = (v_{j1}, v_{j2}, \ldots, v_{jp})$ is the row-vector with the coefficients of the p original variables corresponding to the *j-th* largest eigenvalue $\lambda_j$, than the loading of the *k-th* original variable on the *j-th* component is calculated by $v_{jk}\sqrt{\lambda_j}$. These factor loadings express the variable/component relation and therefore reveal which of the X-variables are dominant in determining the model, and tell how they are related to each other. The set of loadings is also referred to as the *i-th* eigenvector (latent vector or factor).

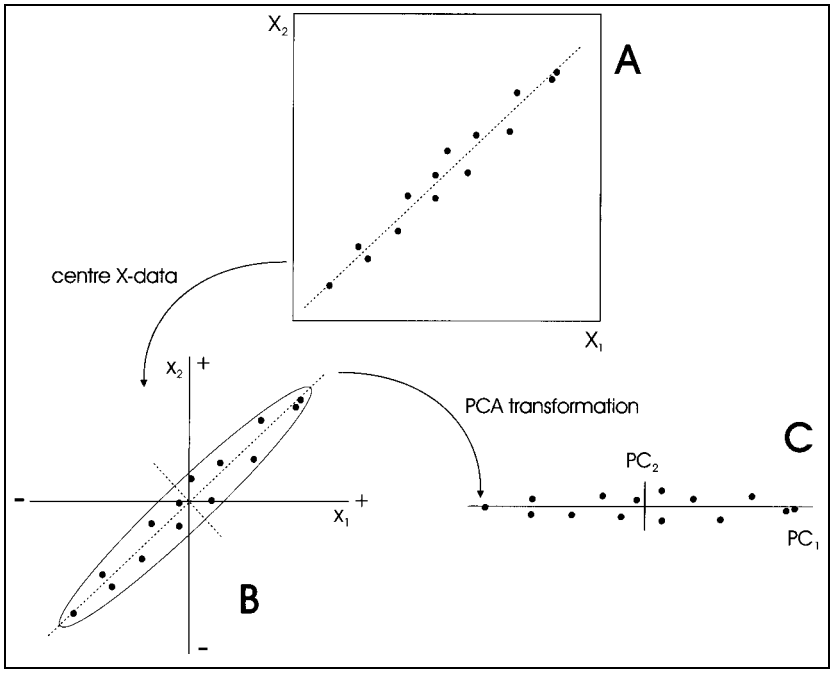



Figure 19. Principle of Principal Component Analysis. Scatterplots of bivariate hypothetical data (A), centered X-data (B) and after PCA transformation (C)

As mentioned before, one of the objectives of PCA is reduction of the dimensionality. Because the principal components are arranged in decreasing order of their variances (eigenvalues), it is common to select only the first few as representatives of the original set of X-variables $\{x_i, i = 1, 2, \ldots k\}$. The Kaiser criterion (129) is probably the most widely used criterion to select the number of components $\hat{T} = \{\hat{t}_1, \cdots, \hat{t}_A\}$ (A<k). According to the Kaiser criterion, the principal components whose eigenvalues are less than the average, i.e. less than one if a correlation matrix of the X-variables has been used, have to be excluded. Another often-used method is the so-called scree test (130). With this method the magnitude of the eigenvalues (vertical axis) are plotted against the ordinal component numbers (Figure 20). Using this method, a recommendation is to retain only eigenvalues (and hence components) in the sharp descend and to discard these where the rate of change between the successive eigenvalues starts to become small. Unfortunately, the method is sometimes slightly conservative by retaining too much components. Several other rules for deciding how many components to retain exist. The selection of the dimensionality needs much attention, but provision of a detailed review of these rules is out of scope of this introduction. A good summary of these rules may be found elsewhere (127). The PC model, which describes the decomposition of the X-variables, is generally denoted as:

$X = TP' + E$ in which P' is the transposed loading matrix and E the residuals

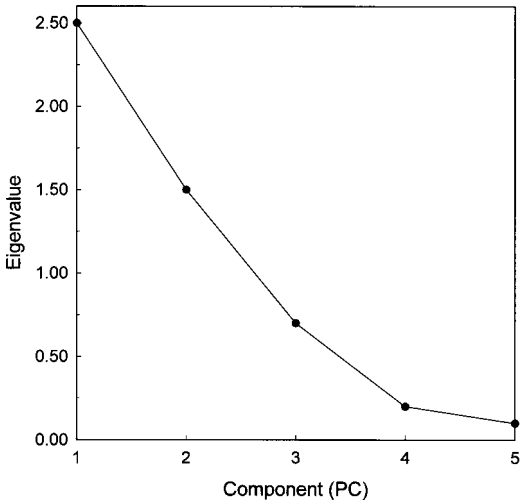



Figure 20. Scree plot expressing the relation between the eigenvalues and the component numbers

The PCA method is illustrated with a simple data set used by Hemel et al. (131). Although these data are intended for classification of patients, they illustrate the technique very well. Clinical chemical parameters creatinine (CREAT), glutamic pyruvic transaminase (GGT), total bilirubin (TBI), lactate dehydrogenase (LDH), aspartate aminotransferase (ASAT) and alanine aminotransferase (ALAT) were measured in serum samples from 27 patients suffering from heart (n=9), liver (n=9) and kidney (n=9) disease. PCA was performed using the correlation matrix of the six analyte results of the 27 patients samples. Table 10 shows

the eigenvalues, the proportional and cumulative proportional variances of the 6 principal components obtained from the 6 analytical parameters. From this table it can be seen that the last two latent vectors (component 5 and 6) only contain 1% and 0.2% of total variance. They can be removed without loosing any information. Based on the Kaiser criterion only the first two components should be retained, because their eigenvalues are both greater than one.

Table 10. Eigenvalues of the patient data set.

| Component | Eigenvalue | % of variance | Cumulative % |
|---|---|---|---|
| 1 | 2.394 | 39.9 | 39.9 |
| 2 | 1.983 | 33.1 | 73.0 |
| 3 | 0.830 | 13.8 | 86.8 |
| 4 | 0.710 | 11.8 | 98.6 |
| 5 | 0.062 | 1.0 | 99.7 |
| 6 | 0.019 | 0.2 | 100.0 |

Figure 21 shows the score plot (A) and the loading plot (B) of the first two components of these data. The first and second eigenvalues are about equally important with 40% and 33% of total variance, respectively. From the score plot (Fig 21A) it can be seen that the data of the 3 diagnosis groups are perfectly separated from each other. Furthermore, it can be seen that the group of liver patients contains an extreme outlier (probably a patient with a viral infection) and the group of heart patients contains 2 outliers (probably caused by acute myocardial damage).

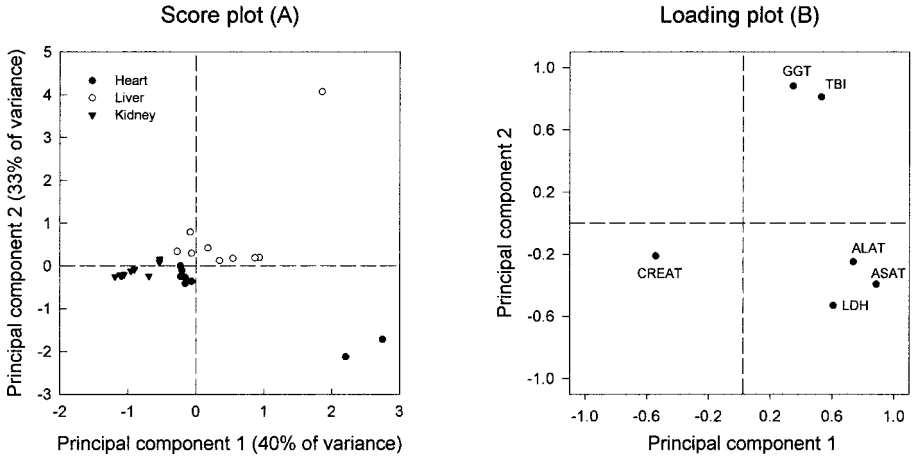



Figure 21. Score and loading plot of the data set of Hemel et al.

From the loading plot (Fig. 21B) it can be seen that all variables have moderate or high loadings on at least one of the two components. From these loadings it can be concluded that all analytes are important in this PCA model. It should be noticed that the two

components are totally uncorrelated to each other (inherent to PCA), so that no information can be obtained from that point of view. The first component of the loading plot shows that the loading value of CREAT is opposite to the loadings of LDH, ALAT and ASAT. In this case we have a component that is also called a bipolar factor. These findings can be related to the results of the score plot. Both the loading of the CREAT, as well as the scores of the kidney patients can be found at the left side of the loading and score plot, respectively. These findings are in line with the expectation that a high outcome of CREAT is found in kidney patients. The same is true with the LDH, ALAT and ASAT parameters and their relation to hearth diseases. The second component shows the contrast between the GGT, TBI and the ALAT, ASAT and LDH loadings, respectively (Fig. 21A). Therefore, this component is mainly responsible for the separation between the scores of the liver and hearth disease patients (Fig. 21A).

From this example it can be concluded that PCA can: graphically depict outliers, show the relevance and the relation of the original variables to the scores, result in less uncorrelated, but more meaningful new variables (six X-variables → two components), and can be used for classification.

If has been suggested that the sample size to obtain a reliable number of factors with PCA should be at least 5 samples per X-variable, and not less than 100 samples per analysis (132). Because of its data-reduction qualities, PCA forms the basis of the multivariate calibration methods as described in the next paragraphs.

More general background information about PCA can be found in books of e.g. Stevens (127), Flury et al. (133), or Afifi et al. (128), whereas more mathematical oriented background can be found in Mardia et al. (122) and Morrison (123).

*Principal component regression (PCR)*
PCR, a so-called bilinear calibration method, performs the regression of Y on selected principal components of X (see Fig. 22). PCR is most suitable in case of a single Y-variable (just like MLR), since it handles one Y-variable at a time. In case of several Y-variables it is possible to perform several subsequent PCR runs for one Y-variable at a time, but PLS is a better choice if the Y-variables are correlated.

In the discussion about MLR, it was noted that the estimated regression coefficients will be very imprecise if the independent variables are highly interrelated. In the previous paragraph dimension reduction was obtained by performing PCA. Using PCA, often ≤ 5 components will account for most of the variance of the X-variables and become the new predictors in the regression analysis. As a consequence, much better N/k (object/variable) ratios will be obtained for the regression analysis. PCR is in essence a MLR analysis on the selected principal components, instead of the original X-variables. The choice of the components in the regression context is however somewhat different from that in PCA. In contrast to MLR, the principal components with the largest variances are selected in PCA in order explain as much of the total variation of the X-variables as possible. In MLR the correlations (explained variances) with each of the dependent X-variables must be defined.

Fortunately, the components with the largest variation (highest eigenvalues) often have good correlation with the dependent Y-variable(s).

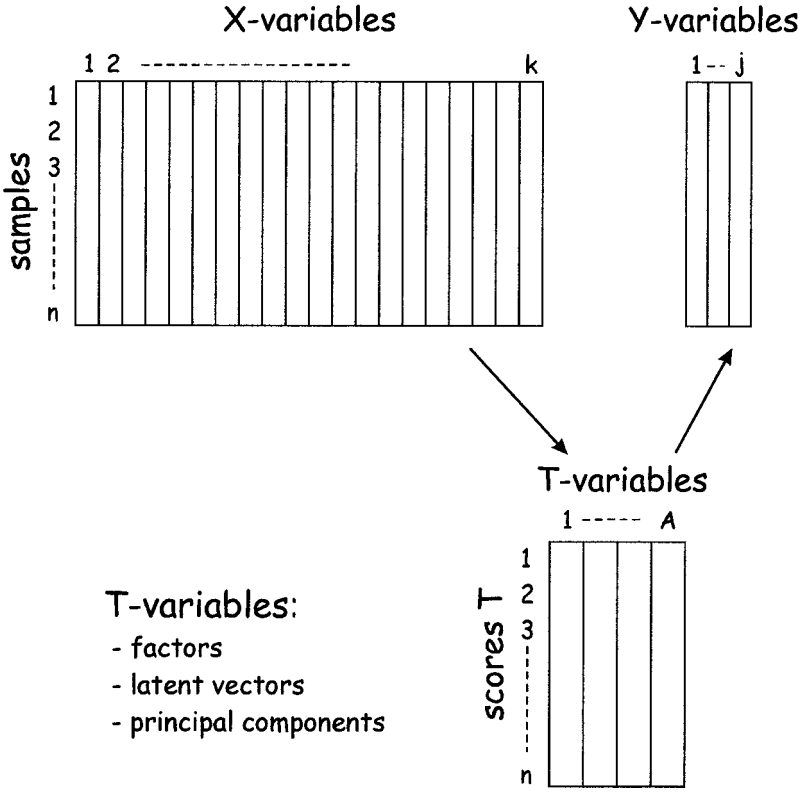



Figure 22. Principal component regression by regressing the Y-variables onto the latent vectors ( $\hat{T}$ ) representing the hidden main variations of the X-data.

PCR on the centered X and Y data can be formulated as follows:

Calibration:

| | |
|---|---|
| $\hat{T} = X\hat{V}$ | The scores $\hat{T}$ obtained from the linear combinations of X; V, loadings of the X-variables |
| $X = \hat{T}P' + E$ | decomposition of the X-variables; in PCR the loadings P are equal to V; E, residuals |
| $Y = \hat{T}Q' + F$ | the loadings Q may be compared to $\beta$ in MLR; F, residuals |

Prediction:

| | |
|---|---|
| $\hat{t}_i = x_i'\hat{V}$ | $x_i'$, the measured values (e.g. Abs.) of prediction sample i |
| $\hat{y}_i = \hat{t}_i'\hat{Q}$ | $\hat{y}_i$, the predicted outcome of sample i |
| $\hat{e}_i = x_i' - \hat{t}_i'\hat{P}'$ | $\hat{e}_i$, the residual of sample i |

If only the first few factors are collected in $\hat{T}$, $\hat{E}$ represent the residuals in the model because X is approximated by $\hat{T}P'$. But, if all the A=k factors would have been extracted from X, than X can be written as $X = \hat{T}P'$. In this case Q would be equal to the regression coefficients β in MLR.
Selection of the optimal number of factors (eigenvectors) is PCR also an important issue. This selection is described in a next paragraph treating some of additional calibration features. More information about PCR can be found elsewhere (122;126).

*Partial least-squares regression*
PLS is also a bilinear calibration method. It usually handles several Y-variables better than both PCR and MLR, and should be chosen in case several intercorrelated Y-variables have to modeled. Another important difference of PLS from PCR is that PLS uses the Y-variance actively during the decomposition of the X-variables. By balancing the X-, and Y-information, the PLS method reduces the effect of large but unrelated X-variations in the calibration model. PLS therefore, produces a calibration model with as few dimensions as possible and in such a way that these dimensions are as relevant as possible. As a consequence, the PLS method has somewhat greater flexibility compared to PCR, but at the expense of the need of an extra loading vector, referred to as the loading weights W. As a drawback, PLS has a stronger tendency to overfit noisy Y-data than PCR (126). Usually there are two PLS techniques employed: PLS1 for one Y-variable and PLS2 for the simultaneous calibration of several Y-variables. As described, in PLS calibration the Y-data affect the data compression modeling of X. The different Y-variables will therefore give somewhat different modeling of X, and hence different regression factors ($\hat{T}$). With the PLS1 regression algorithm, each y-variable is modeled separately. With the PLS2 regression algorithm, a jointly optimized calibration is accommodated for several Y-variables by using a linear combination of the Y-variables. The PLS2 analysis may be especially useful during calibration, if the Y-variables are strongly intercorrelated with each other. The PLS2 algorithm uses this intercorrelation structure to stabilize the random noise of the individual Y-variables. However, if the different Y-variables have different types of curvature in their relationship to the X-variables, the PLS2 solution will find a suboptimal approximation solution. In such cases it may be advantageous to use separate PLS1 modeling for each separate y-variable. Good mathematical descriptions of both PLS algorithms, including the use of the methods with several samples can be found in Martens et al. (126).

*General features of calibration*
Data pretreatment
Except for the previously described centering (see PCA paragraph) various pretreatments exist to obtain more easily interpretable models. This includes linearization of strong non-linearities, weighting, normalization and other transformations. The book of Martens et al. (126) treats these issues in a comprehensive way, including applications.

Calibration and prediction
Every calibration model has to be validated before it can be used for prediction purposes. The average prediction error is often used to get an impression about the predictive performance of the calibration model. This average prediction error is frequently denoted as

the mean squared error (MSE) or the root of the mean square error (RMSE = √ MSE). The MSE is generally formulated as:

$$MSE = E(y - \hat{y})^2$$

in which E is the expectation of the squared differences $(y - \hat{y})$; y the real outcome; $\hat{y}$ the estimated outcome of a sample.

To obtain reliable predictions of the 'unknown' samples, the calibration modeling process is usually subdivided in three phases, the training-, the calibration validation and the prediction-testing phase (see Figure 23). Each step is associated with a separate dataset. In the context of (multivariate) calibration two kinds of validations should be distinguished. The first validation step concerns the validation of the calibration data themselves and is also referred to as 'internal validation'. In case of bilinear calibration models (PCR and PLS) internal validation is used to assist in identifying the optimal number of factors, which should be retained. With bilinear calibration models, the MSE (SEC: standard error of calibration) of the calibration data (training-set) will continue to be reduced, when more factors are included (see Figure 24). The SEC, however, does not reflect the real predictive ability of a calibration model. Therefore a separate set of samples is needed, referred to as validation samples, to determine the actual predictive ability. At low model complexity (not enough factors), the MSE (SEP: standard error of prediction) of the validation samples is high (Fig. 24). In this case, the calibration model is underfitted due to e.g. unmodeled interferences. After increasing the number of factors, the SEP will begin to rise again and this indicates how many factors should be retained – usually one factor less than the minimum SEP. Calibration beyond the optimum model complexity will result in an opposite trend with increasing SEP values, because the model starts to adapt to noise instead of to the relevant features of the data.

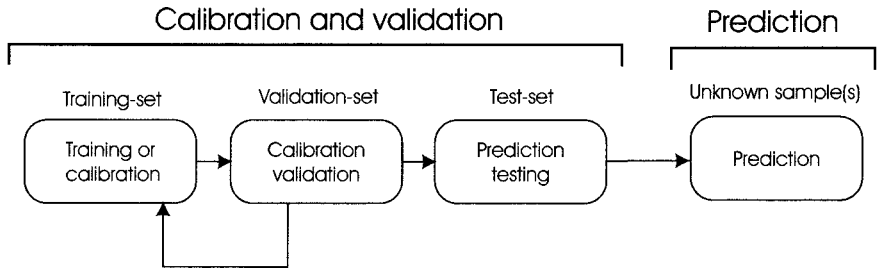



Figure 23. Calibration and validation steps necessary before prediction.

There are several strategies for selecting the validation samples for the internal validation. Ideally, the best strategy is to have a set of samples that are independent of the training samples. This method can be used if sufficient calibration samples are available to split the dataset into two halves, one for the calibration and one for the calibration validation. Unfortunately, it is often difficult to obtain a sufficient number of authentic samples to form a representative sample set for the calibration process itself. The size of the independent validation set is ideally as large as the training set, but must be at least ≥ 25% than the

number of training samples. Another validation strategy is the so-called cross-validation (CV) method (134), which allows to validate the calibration without using an extra set of validation samples. CV is performed by partitioning the calibration set into various careful selected data subsets. With the current powerful computers the calibration set is often partitioned into subsets of size one, which is also referred to as full CV or the leave-one-out-method. A drawback of the crossvalidation methods is that the results are often suboptimal (slightly overoptimistic), when compared to the results obtained with a set of independent validation samples. In this case, the samples are repeatedly taken out and replaced out of the calibration set (n) one by one and the calibration is successively performed on the remaining n-1 samples.

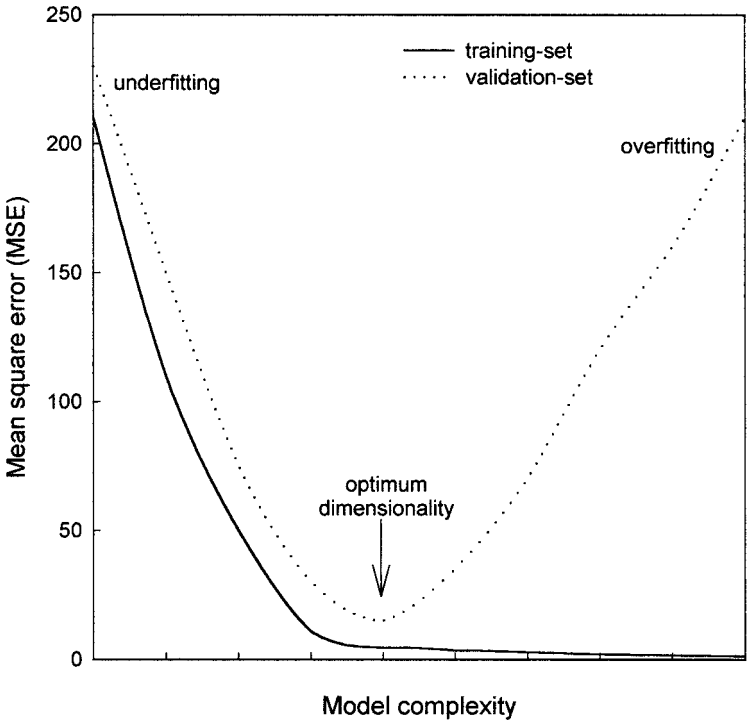



Figure 24. Prediction error as function of the complexity (dimensionality) of the training- and the validation-data.

The sample left out of the calibration set is treated as an independent prediction or validation sample. This process is repeated until all the samples have been left out and predicted. The SEP of the cross-validated samples is also referred to as the standard error of cross-validation (SECV) and can be plotted in a graph as Figure 24.

The second important validation step after the calibration modeling is needed to obtain an objective assessment of the magnitude of the prediction errors (Figure 23). After the optimal dimensionality has been determined, the predictive ability of the model must always be determined by using an independent set of data, also referred to as the test- or

external validation-set. If the results form the prediction testing is satisfactory, the calibration model may be used for the prediction of the results of 'unknown samples'.
More information about some of the sampling strategies for validation and recommendations about the sample set sizes may be found in a paper of Davies (135).

Outlier detection
In calibration modeling outlier detection is very important. Outliers are measurement values that do not fit with the rest of the dataset. These outliers may arise from different kinds of measurement errors or misreadings. The outliers may be detected in the X-, or Y-matrix or occur in the X-Y relationship. These outliers may seriously affect the future predictions using the calibration model. Detection of the abnormal observations is therefore important. Once detected, the observations should be corrected or be removed from the dataset, if they appear to damage the calibration model. However, whenever possible one should try to understand the reason for every outlier. Outliers can be detected in the score and loading plots (see the SMAC dataset in the PCA paragraph) or by using other outlier detection criteria (e.g. studentized residuals, leverage warnings, etc.). For comprehensive background information with respect to outliers statistics we refer to e.g. Kleinbaum et al. (121) and Martens et al. (126).

### 3.2.2. **Neural networks**

*Introduction*
In the previous chapter about PLS regression some methods were described that were based on precisely defined mathematical and statistical algorithms. Artificial neural networks (ANNs) use a different approach. They are based on algorithms that are capable of storing the various characteristics of different input patterns (e.g. absorbance and concentration patterns), in a system of multiple connections between so-called neurons. As can be guessed by the name, artificial neural networks are models adapted from the structures in the brain that makes thoughts possible. The brain interprets imprecise information from the senses and learns − without any explicit instructions − to create the internal representations enabling many skills. In the brain, a typical neuron collects many signals through a host of fine structures, called dendrites (Figure 25). The neuron sends spikes of electrical activity through a thin strand known as the axon, which ends in thousands of branches. At the end of each branch, a structure called synapse converts the activity from the axon into electrical effects that inhibit or excite activity to the connected (downstream) neurons (136).

*Artificial neurons*
ANNs do not reflect the detailed geometry of the dendrites, axon and synapses, but are made of much simpler, structured patterns of interconnected artificial neurons. These neurons (nodes) express a single number, similar to the electrical output of the biological neurons that represents the rate of firing or activity. Figure 26 illustrates the concept of a type of an artificial neuron. Both $X_1$ and $X_2$ represent a measured value (e.g. absorbance), whereas Y is the outcome (e.g. concentration) of the neuron. After applying a certain threshold function to the sum of the X-values, the outcome of the neuron will be forced to either zero or one (Fig. 26). Normally, an artificial neuron is slightly more complex.

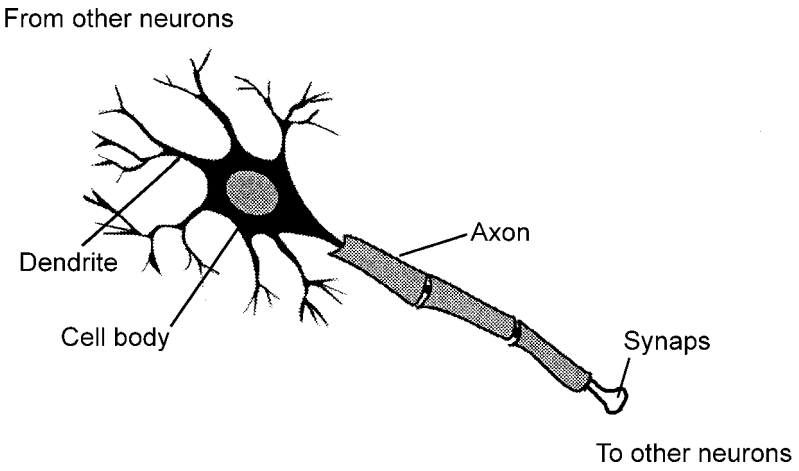



Figure 25. Illustration of a biological neuron.

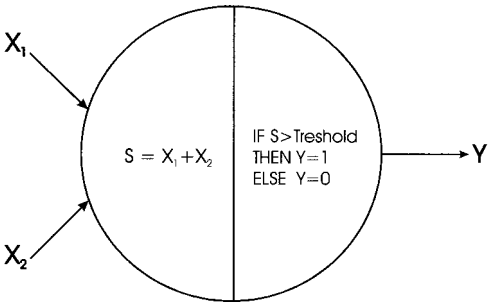



Figure 26. Artificial neuron with a simple threshold function.

The nodes of an ANN are made up of mathematical formulas, in two interconnected units. In the first unit of the artificial neuron the function of the synapse is modeled by a modifiable weight, which is associated with each connection. This part of the node is a computational device, which receives a number of input signals (values). Each input is associated with a weight (number), which represents the stimulating or inhibiting influence of the input signal. These weighted inputs are added together to create a quantity, which is called the net input (Figure 27). This weighted sum is formulated as:

$$net_j = \sum_{i=1}^{n} w_{ij} x_i$$

$w_{ij}$ denotes the weight connecting the neuron i in the previous layer (see later) to neuron j in the current layer; $x_i$ denotes the *i-th* input value of the n input signals.

Because in multilayer networks (see later) the input value $x_i$ of the *l-th* (current) layer is usually the output (out) of the *(l-1)-st* previous layer, the last equation can be written as:

$$net_j^l = \sum_{i=1}^{n} w_{ij}^l x_i^l = \sum_{i=1}^{n} w_{ij}^l out_i^{l-1}$$

The second unit (bottom half of Fig. 27) of the artificial neuron uses an input-output function that transforms the total input value to an outgoing activity.

*Transfer functions*

The behaviour of a node depends on both the weights and the input-output functions. These functions, also called activation functions, may be defined as: out = f(net).

These functions are categorized in three classes: linear, threshold and nonlinear (136). If a neuron does not transform its net input ( $net_j = \sum_{i=1}^{n} w_{ij} out_i$ ), it is said to have an identity or linear activation function. (Figure 28A). The so-called hard-limiter function (Figure 28B) is a threshold function that sets the output at one of two levels (0 or 1), depending on whether the total input value (net) is greater or less than some threshold value. The neuron of Figure 25 is an example of a typical hard limiter function. The threshold function can be used where binary output values are used. In these cases an output of 1 signifies a Boolean 'true' and 0 a Boolean ‘false’. The threshold logic function (Figure 28C) is another activation function, which is in some respect similar to the hard-limiter function but has in addition a swap interval, within which out is linearly proportional to net. These threshold functions should not be used for direct quantitative analysis, where continuous input and output values directly represent the desired values such as concentrations. The third group of functions is the nonlinear functions. The nonlinear functions bear greater resemblance to the real neurons than do linear or threshold functions, but all functions must be considered as rough approximations. For the nonlinear functions the output varies continuously but not linearly, as the input changes.

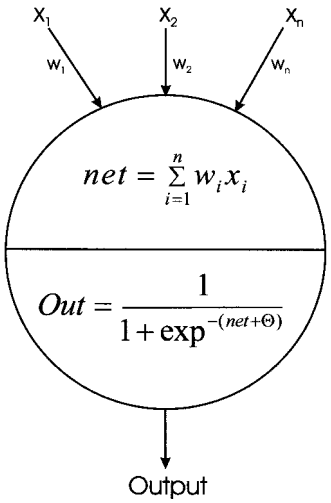



Figure 27. Artificial neuron. The sum of all weighed input signals is computed before the transfer function (sigmoid) is evaluated, using the latest value of net.

In principle the form of the function of this category is quite arbitrary, with only three conditions attached to it:
- Its outcome must be confined to the interval [0, 1]
- It must increase monotonically
- Its must be possible to define a derivative of the function (see later)

Because these activation functions have a bounded range they are also referred to as squashing functions. From all squashing functions representing this group, the sigmoid or logistic function is probably most used (137). A standard sigmoid neuron j having an input $net_j$ is described as:

$$out_j = \frac{1}{1+\exp^{[-net_j+\Theta]}}$$ in which $out_j$ is the outcome of the squashing function and $\Theta$ is the bias

The shape of this function is an S-curve, scaled between 0 and 1 and it has a threshold value $\Theta$. Negative and positive values of the $\Theta$ just move the S-curve to the left and the right and can therefore be regarded as a threshold value at which the output of the neuron is released (Figure 29). The middle most S-curve of Figure 29 has a $\Theta$ value of zero. The goal of the network training is to change most of the weights (and $\Theta$) so that most of the neurons will have net-values spread around the non-linear parts of the S-curve (output-values between about 0-0.2 and 0.8-1). The use of the parameter $\Theta$ is not limited to the sigmoid activation function, but is applied to most of the activation functions (137), and is generally denoted as: f(net, $\Theta$).

Another nonlinear function is the symmetric sigmoid squashing function, which has the same input ($net_j$) as the sigmoid neuron, but has outputs, which are scaled between –1 and 1. This function is formulated as follows:

$$out_j = \frac{2}{1+\exp^{[-net_j+\Theta]}} - 1$$

Bos et al (138) has found that this activation function needs much lower learning rates (see later) than standard sigmoid neurons.

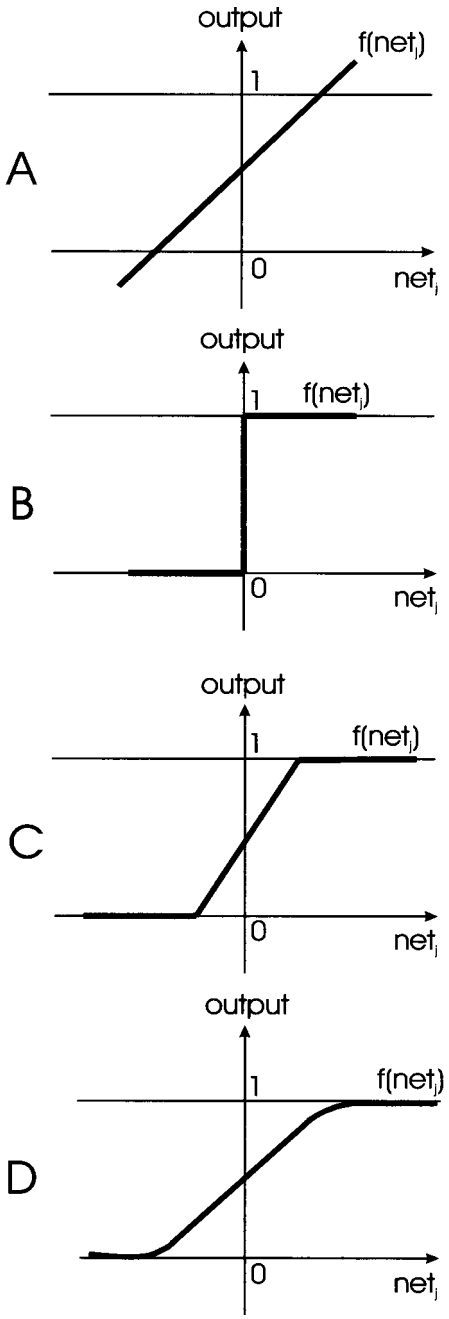



Figure 28. Activation functions. Linear (A), hard-limiter (B), threshold logic (C) and sigmoid (D).

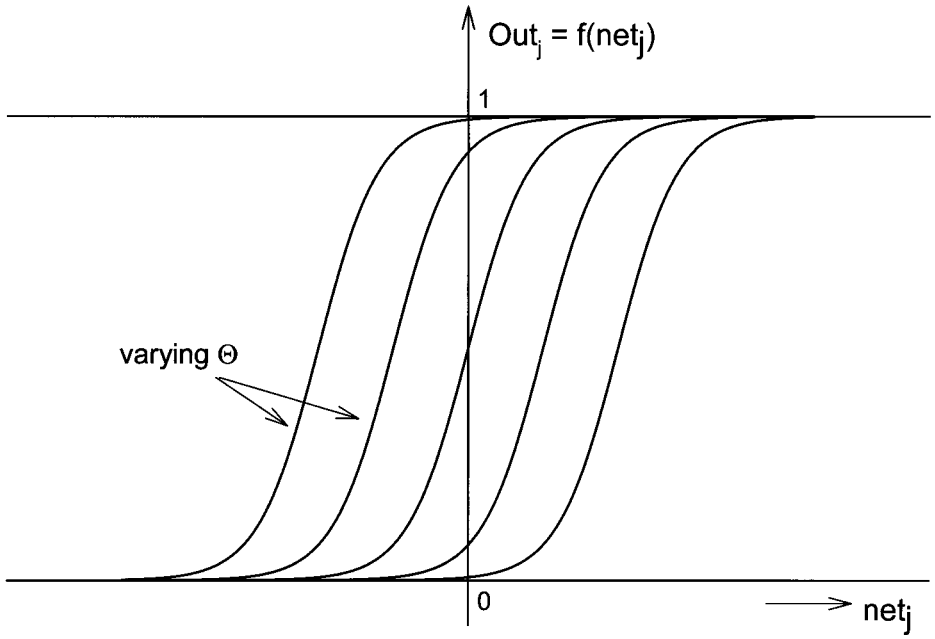



Figure 29. Effect of the parameter Θ on the sigmoid activation function.

*Network topology*

Like biological neurons, a single neuron is not sufficient to perform a specific task. In ANNs, the neurons have to be connected to one another and the set of weights have to be set properly. The way in which the neurons are connected is called the network topology. The topology determines whether it possible for one neuron to influence another and the weights indicate the strength of the influence. In a certain sense, the whole set of weights represents the knowledge of a specific learning task. Artificial neural networks can roughly be subdivided in four groups, namely networks that are capable of association (e.g. character recognition), classification, transformation (mapping a multivariate space into a smaller dimensionality), and modeling. Modeling, one of the most frequently used mathematical applications in science, is the search for a function or model that can predict a specified output from any input pattern. The advantage of a neural network model is that it does not require any knowledge of the mathematical function. Using a sufficiently large number of parameters (weights) ensures enough freedom to adapt the neural network to any relation between the input and output data. Modeling always requires the so-called supervised learning. Supervised learning is a training process in which a mechanism is used to make a neural network associate the target values with the input values (e.g. associate concentrations with absorbances). As a consequence, the training must contain the target values for supervised learning to take place. After training, these networks are capable of predicting the values of output patterns of new samples, which is actually a form of interpolation, and sometimes extrapolation. The so-called backpropagation networks are the most widely applied supervised networks.

Hopfield networks (139) are often used for solving problems related to association, whereas Kohonen networks (140) are often applied for classification and mapping problems. In our own studies we were only interested in the modeling capabilities of neural networks (see Part II of this thesis). Therefore, the properties of the backpropagation networks will be explained in some greater detail in the next paragraphs. More information about the other kind of network applications can be found in e.g. Zupan et al (137).

*Backpropagation networks*
The name of these networks is derived from the learning algorithm, the backward error propagation algorithm. Backpropagation networks may be subdivided in two kinds of network systems: the classifier systems and the function approximation systems. The first system is the oldest one and is used for classification purposes with dichotomized output values (only 0 and 1 values). With the function approximation systems the capabilities of the backpropagation networks are used for the approximation of continuously valued functions. Because our studies are concerned with calibration and prediction of analytes analyzed with spectroscopic methods, we have only given attention to the use of the function approximation backpropagation network systems. The majority of this group of artificial neural network systems consists of three groups, also called layers, of neurons (Figure 30). With these neural networks, which are also referred to as multi-layer perceptrons (MLPs), the neurons of the input layer are connected to the neurons of the so-called hidden layer, which in turn are connected to the neurons of the output layer. Note that all neurons of the hidden layer have every possible connection with the input and the output neurons (Fig. 30).

As a consequence, a large number of neurons (nodes) will result in a large number of connections in the ANN. The signals at the input neurons represent the raw-input data that are fed into the network. Each connection carries the signal from the input neuron to a node deeper into the network, and each connection applies its own weight (w) to the signal (s) so that the received signal is the product w•s. In this way the hidden neurons are free to construct their own representations of the input data. After applying a non-linear transfer function, the same process is repeated between the hidden and the output neurons.

Because the weights of a network are not known in advance, the starting values of the weights are normally randomized between –1 and +1 before assignment to the nodes of the neural network. The process of carrying the signals through the network from the input- to the output-neurons is also called the *forward* step of the network processing. In supervised neural networks, however, the weights are not totally free to construct their own representations of the input data, because the patterns of input activities (e.g. absorbances of a spectrum) have to be mapped to the patterns of the output activities (e.g. concentrations of one sample). The adaptation of the input to the output data, is a process that is generally referred to as learning by *back-propagation*.

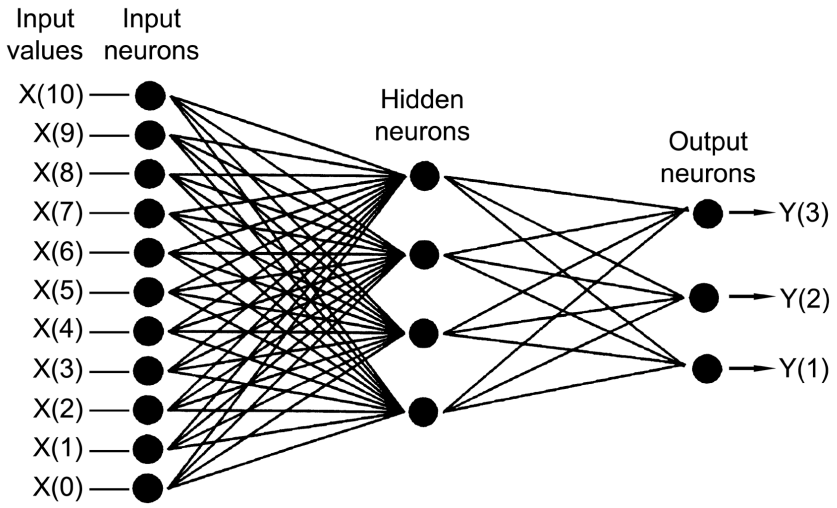



Figure 30. Three layer 'multi-layer perceptron' neural network. The input signals X(0)..X(10) represent the values of an input pattern (e.g. the absorbances of a spectrum) and Y(1)..(Y3) the values of an output pattern (e.g. the concentrations of different analytes of a sample).

With back-propagation, the ANN is provided with a set of training samples (e.g. spectra) with known outcome (e.g. concentrations). The ANN is set to iterate around a loop where for each sample of the training set it predicts the outcome in the forward step. Subsequently, the ANN compares the predicted outcome (Out) to the real sample outcome (t, target) and changes the weights by either strengthening or weakening their values proportional to the size of the prediction error. The backpropagation network derives its name from the fact that the errors propagate back into the network from the output layer to the preceding layers. The global prediction error function E of the network for a pattern p in the training set is defined as (141):

$$E = \frac{1}{2}\sum_j (t_j - Out_j)^2$$

in which $t_j$ is the required (target) value of output neuron j of pattern p, $Out_j$ the calculated value of the neuron of pattern p, and ½ is a factor added for mathematical reasons.

The forward and backward steps occur for each sample in turn, and then are repeated many times over the complete training data set in order to reduce the prediction error, or ultimately eliminate it completely. What actually happens is that the weights are juggled around, so that the output becomes closer and closer to the actual solution. This is called convergence. In order that the weights do not fluctuate (oscillate) wildly, the change of the weights is controlled by application of the so-called *delta-rule*. This delta-rule, presented by Widrow and Hoff (142) as the 'least mean square' learning procedure, extended the original perceptron learning rule (143) to continuously valued inputs and outputs. The principle of the generalized delta-rule is based on gradient descent (137) and is only applicable to MLPs with differentiable activation functions. This training algorithm that was popularized by

Rumelhart et al. (141) and remains the most widely used supervised training method for neural nets.
In a 3 layer MLP, the first step in the application of the delta-rule is the calculation of the error term $\delta$ for the neurons of the output layer:

$\delta_j = (t_j - Out_j) f'(Net_j)$ $Out_j\ (= f(Net_j))$ is the outcome of the *j-th* neuron of the output layer, and $f'$ is the derivative of the transfer function of $Net_j$, which is $Out_j(1 - Out_j)$ in case of a sigmoid function.

For each neuron *i* from the preceding (hidden) layer the term $\delta_i$ is calculated using the $\delta_j s$ from the succeeding output layer and the weights connecting the neuron *i* in the hidden layer to the neurons *j* in the output layer:

$\delta_i = f'(Net_i) \sum_{j=1}^{r} w_{ij}\delta_j$ in which $\sum_{j=1}^{r}$ is the summation over the r connected output neurons

Using these error terms ($\delta$), the re-adjustments of the weights can be calculated. The correction of the weights $\Delta w_{ij}^n = (w_{ij}^{(new)} - w_{ij}^{(old)})$ of the neurons, at any layer *n* during the learning process is defined by the delta-rule as:

$\Delta w_{ij}^n = \eta \delta_j^n Out_i^{n-1}$ in which $Out_i^{n-1}$ is the outcome of the *i-th* neuron of the previous layer (n-1) and hence one of the inputs of the *j-th* neuron in the current layer *n* ($x_j^n$), $\delta_j^n$ is the error term of the *j-th* neuron of the current layer (n) and $\eta$ the so-called learning rate (see later)

From this formula it can be seen that the change of the weight $\Delta w_{ij}^n$ on the layer *n* is proportional to the error $\delta_j^n$ and to the signal $Out_i^{n-1}$ coming from the neuron *i* of the preceding layer.

The described backpropagation algorithm can be used for both batch training (in which the weights are updated after processing the entire training set) and incremental training (in which the weights are updated after processing each training sample) (144;145). For batch training, the weight adjustments ($\Delta w$) are temporarily saved by summing them in an adaptation-array. After each of the training patterns has been processed in this way, the summed adaptations of the weights are added to the weights. After this, the adaptation-array is zeroed and the process is repeated. Each processing of the entire training set in this fashion is referred to as an epoch or iteration. As a rule hundreds or even thousands of epochs are necessary to achieve convergence to a global minimum. The summation of the weight adjustments in an adaptation-array and re-adjustment of the weights at the end of an epoch avoids that the network becomes skewed to the last presented training pattern(s) to

the network. For incremental training, the standard backpropagation algorithm does not converge to a stationary point of the error surface (see later). To obtain convergence, the learning rate must be slowly reduced. This methodology is called 'stochastic approximation' or 'annealing' (146).

This delta-rule is normally associated with two extra variables $\eta$, the learning rate and $\mu$ the momentum. The learning rate $\eta$ (see the equation about the weight adaptations $\Delta w_{ij}^{n}$) is normally an empirical fixed step size that determines the size of the steps taken in the weight adaptations. The learning rate must have a value between 0 and 1 and needs to be chosen in such a way that it ensures the most rapid learning without the $\Delta w_{ij}^{n}$ values oscillating wildly. In order to achieve faster convergence, and to avoid getting trapped in local minima (see Figure 31), the general equation for correction of the weights is generally augmented by an additional term:

$$\Delta w_{ij}^{n} = \eta \delta_{j}^{n} Out_{i}^{n-1} + \mu \Delta w_{ij}^{n(previous)}$$

where $\mu$ is called the momentum and $w_{ij}^{n(previous)}$ is a change of the weight $w_{ij}^{n}$ from the previous learning cycle. The first term of the equation refers to the current cycle.

Unfortunately, the addition of the momentum in backpropagation networks requires doubling of the computer space, because all weights have to be stored for both the current and the previous cycle (epoch), as can be seen from the last equation. This generalized delta-rule including the momentum is called the 'heavy ball method' in the numerical analysis literature (147). Figure 31 shows the effect of the standard backpropagation training on the global network error E. The error surface on the X-axis is defined by the set of training samples and the applied network topology. The starting point of the training is illustrated in this figure by the black ball (circle), which is rolling down from a hill. This hill, however, does not only have a downhill slope (gradient descent), but also has a bumpy surface with several peaks and valleys before the lowest point (valley) is reached. In unfavorable circumstances, the ball may settle into a local minimum (valley) instead of finding the global minimum. In network terminology, such an awkward situation may be prevented by choosing the right learning rate and momentum. In such a case, the 'heavy ball' is pushed out of the local minimum and continues its way along the error surface towards the optimal solution.

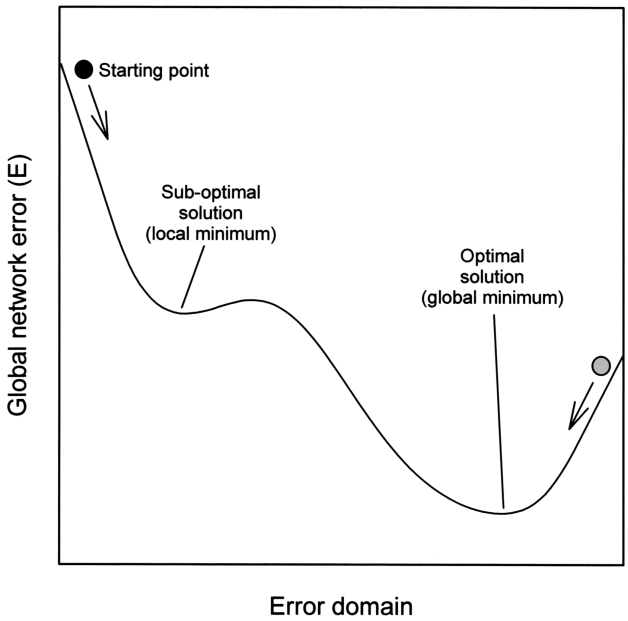



Figure 31. Illustration of the change of the global error rate of the network during training.

*Practical aspects of backpropagation networks*

Data collection

All training-, validation- and test-data must be representative for the problem that has to be modeled. It is important to have more training samples than the number of neurons of the input layer. The number of training samples required depends on the amount of noise in the targets and the complexity of the function that has to be learned. As a starting point, it is a good principle to have at least 10 times as many training samples as input neurons. It should be noted that the data collection and the next step – data pre-treatment – are critical aspects in the development of a network system and can account for most part of the time of the whole development cycle.

Data pretreatment

In order to permit the generation of an accurate network model it is sometimes necessary to perform pretreatment of the input data. For more information is referred to the same passage in the 'general features of calibration' section of the previous chapter.

Rescaling the input and output values

Rescaling a vector means that a constant is added or subtracted from the vector values and subsequently multiplied or divided by (another) constant. This scaling is often applied to obtain values between 0 and 1, or between –1 and +1.

When backpropagation networks with fixed learning rates and sigmoid neurons in the hidden layer are used for the approximation of continuous valued functions, scaling of the input values is often necessary. Without scaling, large input values will result in extreme

(positive or negative) values of the net input function ( $net_j^{hidden} = \sum_{i=1}^{n} w_{ij} out_i$ ), which in turn results in output values of the sigmoid squashing function close to either 0 or 1. These output values ( $Out_j$ ) result in derivative values close to 0 [ $Out_j(1-Out_j)$ ] and hence small adaptations and learning will occur. A larger learning rate may be chosen to compensate, but this also affects the weight adaptations of the succeeding layers as well and may lead to adverse behaviour. Another aspect is that the order of magnitude of an input value with respect to other inputs values should be the same, otherwise the learning may be dominated by the input with the larger magnitude. For that reason, each of the inputs of the network is often scaled separately. On the other hand, scaling the input values may have adverse consequences as it can amplify noise (e.g. in case of NIR spectra).
Linear output functions are often suggested when the mapping of the input and output patterns are highly linear (e.g. absorbances mapped to concentrations). When the output of a sigmoid neuron approaches the limits of 0 and 1, the derivative and thus the delta term approach to 0. In contrast to sigmoid neurons there is no damping of the magnitude of the delta term $\delta_j = (t_j - Out_j) f'(Net_j) = (t_j - Out_j)$ of the linear neurons, and large errors, or even floating point overflow may occur. More information about this scaling subject may be found elsewhere (144).

Training and prediction

The development cycle of the training and prediction of a backpropagation network is essentially the same as the process described in the previous chapter about the PLS regression (see Figure 23). One of the known problems with backpropagation networks is their tendency to over-fit training sets with noisy data. Therefore, it is important to validate the performance of the network training with a separate validation set or full cross-validation and subsequently monitor the prediction error of the training- and validation-data as a function of the number of epochs (training cycles), analogous to the plot of Figure 24 in the previous chapter. Generally, the normalized standard error (NSE) is used to express the performance of the prediction. The NSE is an extension of the global prediction error function E and is expressed as:

$$NSE = \frac{1}{PJ} \sum_{p} \sum_{j} (t_{p,j} - Out_{p,j})^2$$

in which $t_{p,j}$ is the required (target) value of neuron *j* of pattern p, $Out_{p,j}$ the calculated value of the neuron of pattern p, and P and J are the total number of patterns and neurons of the output layer, respectively.

In practice, the square root is often taken from the NSE. This characteristic is also referred to as the root mean squared error (RMSE), which is called RMSEC in case of the error of the training samples (in which C stands for calibration) and RMSEV in case of the validation samples. Except for validation, the predictive performance of the network should be tested with an independent test set.

Initial weights

As mentioned earlier, the starting weight values are normally randomized between –1 and +1. The function of the backpropagation learning algorithm is to move the network system towards a lower error state. By changing the initial conditions of a neural network (other randomized weights), the starting point on the error surface can be modified. The gray circle in Figure 31 illustrates a different starting point on the error surface. In contrast to the former starting point (black circle) the optimal solution is likely to be found in a shorter time with fewer epochs. It is recommended to use at least five different starting conditions (sets of random initial weights) for each neural network model. Typical validation results (RMSEV) should be within a few percentage points of each other.

Early stopping rule

Overfitting is one of the most serious problems in neural network training. Generally, many thousands of cycles are necessary to achieve convergence, if convergence is achieved at all. Because the number of connections and thus weights is large in even medium-sized networks, the iteration can be very lengthy or even be unattainable. A method, which is called early stopping or stopped training, is the most common solution to solve this iteration problem. The method is popular because it is fast. Briefly, the method proceeds as follows:

- Judge the error values of the validation set for this method
- Use relatively small initial weights (e.g. between –0.3 and + 0.3)
- Use a relatively small learning rate (e.g. < 0.2)
- Compute the validation error (RMSEV) periodically during training
- Stop training when the RMSEV value starts to go up

If the validation error goes up and down several times during training the safest approach is to train towards convergence of the RMSEV and then go back to see which iteration (epoch) had the lowest validation error. More information about the early stopping rule can be found in a manuscript of e.g. Sarle (148).

Number of hidden units

The optimal number of hidden neurons depends in a complex way on several issues such as:

- The amount of noise in the targets
- The complexity of the function to learn
- The hidden unit transfer function
- The number of training samples
- The number of input and output neurons

If too few hidden units are selected, the training error (RMSEC) and validation error (RMSEV) will be too high due to underfitting (see Fig. 24). If too much neurons are selected, the training error will be low, but the validation error will be high due to overfitting (Fig. 24). However there are several rules of thumb described for the selection of the optimal number of hidden neurons, none of these is based on heuristic rules. In most applications of backpropagation networks, the topology is determined empirically by training several times with different numbers of neurons in the hidden layer. As the computational effort is large with this approach, an efficient and automated procedure is very desirable.

Learning rate and momentum

In normal backpropagation networks, too low learning rates make the network train very slowly. Too high learning rates make the weights diverge because the change of the weights $\Delta w$ (step sizes) becomes too large, so there is no learning at all. The learning rate may be constant during the training of each network model, or be changed in the course of the training process. In our studies we only applied a fixed learning rate during training of each network model. The determination of the optimum (fixed) learning rate is often found by trial and error. As a starting point, the learning rate may be set proportional to the number of connections with the neurons of the previous layer. In attempt to speed up the backpropagation, methods are developed to change the learning rate in the course of the training. Many of these methods produce erratic behaviour because they change the weight as a function of the magnitude of the gradient (see Fig. 31). The reason for their erratic behaviour is that in some cases large step sizes are needed in areas with small gradients (e.g. to get out of a local minimum) and in other cases small step sizes are needed in areas with small gradients (e.g. to stay into a global minimum). Other methods have been developed, such as the Quickprop (149) and RPROP (150) methods, that do not have this excessive dependence on the magnitude of the gradient.

The momentum is usually set to zero. In calibration problems, the application of the learning rate has proved to be sufficiently fast, without ever leading to oscillation (138).

*Further reading*

An excellent starting point for further reading is a book of Zupan et al. (151) which provides the reader with an overview about artificial neural networks including some chemical applications. A manuscript of Smits et al. (152) is good general description of MLP feed-forward networks, whereas Bos (144) studied some theoretical aspects of backpropagation networks and applied them to a number of quantitative analysis problems (e.g. spectroscopy).

## 3.3. S**pectral library search**

*Introduction*

Visual interpretation of IR spectra for the determination of the composition of a sample may be difficult and time-consuming. The use of FT-IR spectroscopy coupled with a method for searching spectra within a database provides an efficient methodology to the identification problem. There are several commercial spectral databases (e.g. Sadler), each containing large numbers of FT-IR spectra with various components. The components not only occur in their pure state, but also as part of a mixture. For accurate identification of the components, the reference spectra of the library have to be sampled under the same conditions (sample handling techniques, equipment used, etc.) as the sample spectrum. The quality of the library not only depends on the conditions of the reference spectra in the library, but also upon their number. For accurate identification of the component(s), the search method needs a library with a large number of reference spectra. In practice, however, it is often difficult to obtain a commercially available general-purpose library that contains reference spectra of all possible components and mixtures of interest. Therefore, it is often necessary to build a large in-house reference database specially adapted to more

specific analytical problems. The benefit of these in-house databases is that they may be used for the prediction of the composition of a sample, provided that the database contains a sufficiently large number of reference spectra with known compositions of a limited number of components that cover the whole concentration range of the sample spectra.

*Computerized library search*
A search program (e.g. SearchMaster from Sadler) calculates which library spectrum is the most similar to the sample spectrum. Except for library searching, most programs are also provided with features for library creation, development and maintenance. Three issues are especially important for obtaining accurate results from library search: the library resolution, the search area and the kind of search algorithm.

Library resolution
For computerized library search, the number of datapoints to be compared between the sample spectrum and those in the library must match. The resolution of the library determines the size of the library and the processing time. In the mid-IR region, resolutions better than 4 $cm^{-1}$ are rarely if ever necessary. At higher resolutions, the processing time may become too long and searching on minor irrelevant bands caused by noise may occur. If the resolution of the sample spectrum does not match the resolution of the library, the quality of the algorithm for reducing the resolution of the sample spectrum is very important. Just dropping datapoints will definitely distort the spectral bands.

Search area
All library search methods depend on peak matching or full spectrum matching. In any case, it is important to choose the correct wavenumber range(s) for searching. All bands of interest that originate from the components of the sample spectra must be present in the search area. In mid-IR spectroscopy, library search is often performed on the so-called fingerprint area of the spectrum (2000 to 400 $cm^{-1}$), because this area often contains all characteristic and unique spectral properties of the components.

Search algorithm
A number of different search algorithms exist. Each of these algorithms compares the absorbances of the selected datapoints of the sample spectrum with the absorbances of the matching datapoints of each of the reference spectra. The result of the algorithm is always a single characteristic, expressing the degree of resemblance between the sample and reference spectrum. Depending on the kind of search algorithm, the characteristics may be classified into two groups, namely those expressing a measure of dissimilarity and those expressing a measure of similarity. Small values ideally zero or close to zero, express good matching in case of the dissimilarity measures. The correlation coefficient is the only member of the second group of similarity measures and amounts to 1.000 in case of perfectly matching spectra and to 0.000 if no similarly is obtained at all. After the search algorithm is applied to each of the reference spectra, the calculated dissimilarity or similarity values are ordered in magnitude. The sort direction (ascending or descending) depends on the type of algorithm, but the most corresponding reference spectrum is always placed on top the ordered list. This list, which is also referred to as the hit list, contains the resulting value of the algorithm together with a description of the reference spectrum.

Depending on the search program, the entries of the hit list also may be completed with a drawing of the chemical structure, the CAS registry information and the physical-chemical properties of the component. The quality of the search method may depend on the chosen search algorithm. The formula of the six most commonly used criteria will be presented here, followed with a short description of each of these methods.

- The absolute difference: $\sum_{i=1}^{n} |s_i - r_i|$

- The squared difference: $\sum_{i=1}^{n} (s_i - r_i)^2$

- The absolute derivative: $\sum_{i=1}^{n} |\Delta s_i - \Delta r_i|$ with $\Delta s_i = s_i - s_{i-1}$ and $\Delta r_i = r_i - r_{i-1}$

- The squared derivative: $\sum_{i=1}^{n} (\Delta s_i - \Delta r_i)^2$ with $\Delta s_i = s_i - s_{i-1}$ and $\Delta r_i = r_i - r_{i-1}$

- The Euclidean distance: $\sum_{i=1}^{n} \sqrt{s_i^2 - r_i^2}$

- The correlation coefficient: $\frac{\sum_{i=1}^{n} (s_i - \bar{s})(r_i - \bar{r})}{\sqrt{\sum_{i=1}^{n} (s_i - \bar{s})^2 \sum_{i=1}^{n} (r_i - \bar{r})^2}}$

In these formulas, $s_i$ is the absorbance of the *i-th* datapoint of a sample spectrum, $r_i$ the absorbance of the *i-th* datapoint of a reference spectrum and n the number of selected datapoints.

*Absolute difference*
The absolute difference algorithm emphasizes band heights and has the shortest processing time. Results obtained using the absolute difference algorithm may be especially uncertain when the unknown spectrum has a sloping or offset baseline.

*Squared difference*
The squared difference algorithm is a least-squares metric, which tends to weigh bands in the sample spectra more heavily than in the case of the absolute difference algorithm. This means that the squared difference tends to minimize the effects of a noisy baseline. The squared difference algorithm is also a fast search algorithm. Similar to the absolute difference algorithm, the results may be inconclusive when the sample spectrum has a sloping or offset baseline.

*Absolute derivative*
This algorithm emphasizes band positions more than band heights and hence tends to minimize the effect on the match value when the sample spectrum has a sloping baseline or broad non-specific features. If the component of the sample spectrum is not in the library, the first matches may not be alike. The processing time for the absolute derivative algorithm is longer than the processing time for the difference algorithms.

*Squared derivative*
The squared derivative algorithm emphasizes band positions and weights them more heavily than the absolute derivative algorithm. The squared derivative algorithm also reduces the effect of a sloping or offset baseline in the sample spectrum. If the component of the sample spectrum is not in the library the first matches may not be alike.

*Euclidean distance*
This algorithm calculates the match values as the sum of vector differences (120). The Euclidean distance algorithm is especially suitable for prediction of the composition of mixtures. If the sample spectrum has a sloping or offset baseline, the results may be inconclusive. The processing time for the Euclidean distance algorithm is larger than that of the difference and derivative algorithms. Because the Euclidean distance is not as sensitive to differences between spectra as are the other algorithms, it may produce inconclusive results for sample spectra that are not mixtures.

*Correlation coefficient*
The standard correlation coefficient algorithm (121), can automatically account for factors such as baseline drift, differences in scaling and so on. No data preprocessing, such as baseline correction, is needed. Furthermore, because the calculated match value is a correlation coefficient, the search results are not a measure of relative best fit, but rather absolute values with a statistical significance (similarity). It is unlikely that dissimilar compounds in the library will display correlation coefficients higher than 0.95. The processing time for the correlation coefficient is the longest of all the available algorithms.

*Conclusion*
The difficulty with library search is that the method needs a library with a large number of reference spectra to obtain reliable results. In practice, it is impossible to build a library with all possible compositions of components. So, interpolation and combination of library search results is often necessary for the prediction of the composition of sample mixtures.

# 4. Future trends

This chapter briefly describes a few promising techniques related to some of the subjects, which are described in this thesis.

$^{13}$C triglyceride breath test as alternative for fecal fat determination

Both the traditional (e.g. Van de Kamer) methods as well as some of the new IR methods are generally accepted analytical methods for the routine determination of fecal fat for the diagnosis of steatorrhea. Nevertheless, the outcome of neither of these methods is a direct function of the metabolic state of dietary lipid intake. This is because the amount of fecal fat is based on a combination of dietary and metabolic lipids as described in chapter 1.1. Additionally, the treatment of the feces samples in each of these methods remains rather cumbersome. To avoid the above-mentioned shortcomings, some studies described a $^{13}$C medium-chain triglyceride breath test (153;154) as alternative for the analysis of fecal fat. The principle of the $^{13}$C labeled triglyceride tests is based on lipolysis-dependent $^{13}CO_2$ excretion via the breath after the ingestion of a certain amount of $^{13}$C-enriched triglycerides. The $^{13}$C enrichment may be measured by means of isotope ratio mass spectrometry, also referred to as IRMS (154), or gas isotope ratio measurement with FT-IR spectroscopy (155). A general problem of breath tests using labeled lipids for the diagnosis of steatorrhea is the poor sensitivity and specificity, probably caused by the various steps involved in the metabolism of the labeled compound. Differences in e.g. the gastric emptying rate, intraluminal lipolysis, mucosal absorption, lipid metabolism, endogenous $CO_2$ production and pulmonary excretion may obscure the relation between the tracer compound expired and the metabolic process studied. Up to now, none of the $^{13}$C triglyceride breath test studies has been clinically validated, so further investigations have to be done. Because of the large number of compartments (such as stomach, intestinal lumen, blood and lung) involved in the metabolism of the labeled compound, suggestions have been made to measure the $^{13}$C tracer compound in plasma (156) instead of breath.

Raman spectroscopy as alternative for FT-IR

Almost all biological samples contain water. Without sample pretreatment, water may be a serious problem in FT-IR spectroscopy, especially in the mid-IR region (see chapter 2.2). Water attenuation is not a problem for Raman spectroscopy. With Raman spectroscopy, no elaborate specimen preparation is needed and the samples can be remotely detected by back-reflection even through glass windows. Raman spectra are also insensitive to temperature. This is in marked contrast to FT-IR, in which often very short pathlength is required and contamination of the sample probes can cause serious practical limitations. In spite of these advantages and even though it is a rather old technique (C.V. Raman received the Nobel prize for his work in 1930), Raman spectroscopy has not received much attention from analytical and clinical chemists until the invention of the laser.

Raman spectroscopy is a form of optical spectroscopy in which the energy is exchanged between the light and the matter. When light impinges upon a substance it can be scattered or absorbed (157). Most of the scattered light will have the same frequency as that of the incidence light, and is also referred to as elastically scattered light (Rayleigh). Raleigh scattering occurs by the interaction of the incident light and an atom. However, a small

fraction of the incident light can set the molecules in the material into vibration when it impinges upon a molecular bond (Figure 32).

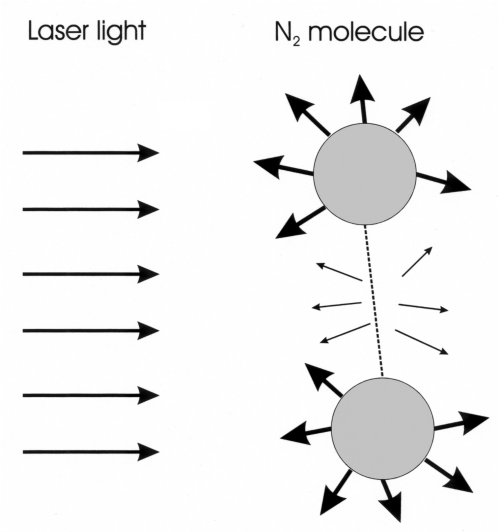



Figure 32. Raman scattering (thin arrows around the molecular bond) and Raleigh scattering (dark arrows around the N-atoms) of a $N_2$ molecule.

The interaction between the incident light and the molecular bond causes a wavelength shift, which is known as the Raman effect (158). The resulting Raman spectrum is a plot of the scattered light intensity (y-axis) versus its change in frequency (x-axis), relative to that of the incident light. The Raman frequency shifts are conventionally measured in wavenumbers ($cm^{-1}$). The Raman spectra differ from IR spectra, in that the Raman bands are sharp and locatable to 1 $cm^{-1}$ for both organic and inorganic components. The Raman and IR bands do not necessarily coincide because different rules govern possible spectral transitions. Fewer bands appear in Raman spectra since fewer combinations of fundamental frequencies occur, resulting in sharper and less crowded spectra (159). Raman spectra may be used for the identification of a sample component as well as for its quantification. Because Raman spectroscopy is based on weak inelastic scattering (typically $10^8$ weaker than Rayleigh scattering), modern Raman spectrometers are equipped with irradiation lasers and sensitive detectors to obtain spectra in a reasonable time. In Raman spectroscopy, visible, ultraviolet and infrared light can be used for Raman excitation. Unfortunately, many samples (particularly organics) fluorescence quite strongly when excited with visible laser light, and as a consequence hide the weak Raman spectra. Therefore, the newest generation of compact FT-Raman/IR instruments eliminate fluorescence as much as possible by using an infrared laser coupled with the Fourier transform technique by which the whole spectral range is sampled and the data are processed in real time. The current list of Raman probes include sample (fiber) optics for noninvasive point-and-shoot monitoring with working distances from 0.01 to 17 inch, immersion probes for use in process streams,

standard (glass) cuvettes and microscopes (160). Because conventional glass does not interfere in Raman spectroscopy at all, standard microscope optics may be used.

In clinical chemistry, Raman spectroscopy may provide quantitative chemical information of analytes in either the cellular or extracellular compartments of the body. Biochemical analysis, using Raman, may be performed in vitro and in vivo. Urinary stones and gallstones have extensively been studied in vitro using Raman spectroscopy. Most of these studies used visible excitation (161;162), whereas others used near-IR laser light (159). Furthermore, Raman spectroscopy has been applied to numerous other bio-medical studies, such as in the diagnosis of arteriosclerosis (163) by in vitro invascular tissue investigation of human coronary artery samples obtained during autopsy, for the detection of breast cancer (164), the classification of drug-resistant bacteria (165), and the analysis of several blood analytes, such as glucose, cholesterol, total protein, albumin, triglyceride and urea (166).
One of the most challenging areas of spectroscopy is in non invasive glucose monitoring in diabetics, which is one of the fastest growing segments of diagnostic testing. Raman spectroscopy has also been applied to the measurement of the blood glucose concentrations (167). Currently, glucose specimen sampling is often performed by finger pricking and collecting a drop of blood. Therefore, Raman spectroscopy is under investigation as an optical technique for non-invasive glucose measurement (168). However, it now appears that individual patient calibration models are needed to overcome the physical effects in noninvasive patient IR and Raman monitoring (101). Further development of the probes aimed at continuous monitoring is also needed. Therefore, the present spectroscopic technology and chemometrics still require further improvements. Miniaturizing monitoring techniques in diabetes therapy including insulin pumps provide further scientific impetus for research into noninvasive glucose assays by Raman and NIR spectroscopy.

Raman spectroscopy is a very promising technique for biomedical applications. Nevertheless, some analytical problems need special attention, especially in case of in vivo measurements, such as laser wavelength and intensity stability, spectral acquisition times, fluorescence blocking and laser heating of the sample.

Genetic algorithms as a method for wavelength selection
In analytical and clinical chemistry, the purpose of developing a calibration model is mainly prediction of the concentration of the components of new samples. It is well known that high spectral overlap may cause large prediction errors (see also chapter 3.2.1). When multiple linear regression (MLR) is used, selection of wavelengths is the most popular method to attempt to reduce the error in the prediction. When a large number of input variables is used (e.g. 700 points of an absorbance spectrum) there will be a serious problem using MLR without selection (very large prediction error). The simplest method of selection would be to examine all possible combinations of the variables exhaustively. Using this selection method with 700 variables one has to choose from 244650 possible two-term (X-variable) equations. In case of three-term equations there would even be 56921900 equations to examine.

In general the number of possible equations can be calculated by:

Number of possible subsets = p!/(m!(p!-m!)) in which p is the number of initial variables; m the number of best variables; and ! is the faculty of p or m.

Of course it is difficult to predict the number of best variables in the subsets. For this reason, and given the large number of possible subsets using even only two- and three-term equations in case of large number of initial X-variables (e.g. 700), this selection method is computationally expensive and in most situations virtually impossible. The disqualification of this selection method has led to the development of other methods, such as PCR and PLS (see chapter 3.2.1.). These techniques became popular because they by-pass the selection problem by using the whole spectrum or large parts of the spectrum. However, because relevant information is often restricted to a few areas of the spectrum, whole spectrum selection will potentially cause many variables to be completely irrelevant to the objectives of the calibration model. The selection of a large number of variables also necessitates the use of a large number of calibration samples in relation to the number of variables to obtain reliable estimates of the regression parameters. Genetic algorithms (GAs) are techniques that circumvent the use large parts of a spectrum for calibration.

The basic principles of GAs were first described by Holland (169). Since the mid 1980s, GAs have been applied to numerous scientific fields, such as to solve search (e.g. wavenumber selection) and the optimization of chemical systems (e.g. the optimization of temperature and solvent gradients in chromatographic processes). In the problem of wavelength selection in multicomponent analysis, GA have shown to be useful in the selection of the most important variables for the calibration model (170-172). From this literature there is an indication that MLR yields models with the same number of regression variables after GA variable selection as PCR or PLS regression, and usually with the same or somewhat better predictive ability.

Genetic algorithms are general-purpose search algorithms based upon the principles of evolution observed in nature. Analogous to genetic evolution, GAs combine cross-over, mutation and selection operators with the goal of finding the best solution to a problem (survival of the fittest). Genetic algorithms search for this optimal solution until a specific termination criterion is met. In case of GAs, the chromosomes are represented by a number of strings. The GAs operate on the populations of the strings, with the strings coded to represent some underlying parameter set (see later). Ideally, these strings are coded with binary values (0 and 1), though other possibilities exist. Selection, cross-over, and mutation operators are applied to the successive string populations to create new string populations. These operators are very simple, involving nothing more than random number generation, string copying, and partial string exchanging.

*The standard genetic algorithm for variable selection*

The first stage of GA is initialization. During this stage, a coding plan and fitness or target function have to be defined, followed by a preliminary selection of a subset of strings (chromosomes) from the population. After the initialization stage, the evolution phase of GA has to be performed. Every evolutionary step in GAs is known as a generation. The

generations in GAs consist of a combination of a cross-over, a mutation and a selection step. These generation steps are repeated over and over until a certain criterion has been reached. The successive steps of GA are briefly described in the next sections.

Initialization of the population

For a spectral wavenumber selection problem, the general way to code the solution is to generate a set of binary coded strings (0100111010…). Such a string is also referred to as a subset. The length of each string is the number of wavelengths of e.g. a spectrum. In this way each possible subset of wavelength (points) can be represented as a string of 0s and 1s, with a 1 in position *i* if the *i-th* wavelength is present in the subset and 0 if it is not. The number of 1s in any particular string is the number of wavelengths in the subset. An initial population of strings is always generated at random. The number of subsets (strings) is generally equal to the number of calibration samples available. The whole set of coded strings is also referred to as the population, and the strings in this population are also called the parents.

Evaluation of the population

To select good subsets of wavelengths, a fitness function has to be defined for each GA selection problem. In spectroscopy, the outcome of such a function is a single numeric fitness value that expresses the predictive ability of the calibration model. In principle, any sensible fitness measurement can be used. Often, the RMSE values of the validation or test sets are used. The fitness value is calculated for each subset of the population. After this process, the fitness values and accompanying subsets are sorted in order of their fitness rating.

Selection

During this phase, the subsets with good fitness ratings are selected from the population. These subsets have a higher chance of producing offspring (new subsets) with even better fitness ratings. The number of selected subsets (strings) is generally only a fraction of the number of calibration objects available (e.g. one third). These subsets are allowed to produce offspring in the next step of the GA process. The rest of the substrings of the population with worse fitness values will 'die' off.

Cross-over

Cross-over is a genetic operator that combines (mates) two parent strings to produce a new subset (offspring). The idea behind the cross-over operation is that the new offspring may be have better characteristics (selected variables) than both of the parents if it takes the best characteristics from each of the parents. The cross-over procedure has two steps, namely the strings are mated randomly and the mated strings couples cross-over, using a randomly selected crossing site (Figure 33).

Parent 1: 11001|010 offspring 1: 11001111

**X**

Parent 2: 00100|111 offstring 2: 00100010

Figure 33. Example of two parent strings creating two offsprings with a single random cross-over point. The "|"symbol denotes the randomly chosen cross-over point.

Cross-over is not necessarily applied to all mates. A choice for crossing is made depending on a probability value selected by the user (typically between 0.6 and 1.0). If cross-over is not applied, the offsprings are simply duplicates of the parents.

Mutation

Mutation is a genetic operator that alters one or more string values in a subset from its initial state (0 to 1 or 1 to 0). Mutation is an important part of the genetic search procedure as it helps to prevent the population from stagnating in any local optimum. The mutation occurs during the generations according to a user-definable mutation probability. This probability is generally fairly low (0.01 or 1%). If this value is set to high, the search will turn into a primitive random search.

The described genetic algorithm is a stochastic iterative process that is not guaranteed to converge. The termination condition may be specified at some fixed, maximal number of generations or as the attainment of an acceptable fitness value. Whatever approach is chosen, when the process is stopped the final population should at least contain some subsets that perform well.

In practice the GA process is usually more complex. For example, other coding strategies may be followed, as well as the use of different mutation and cross-over operators. For this and other information about GAs, the interested reader is referred to other sources with more in depth information about the subject (151;170-176).

## References


1. Malabsorption syndromes. The Merck Manual Section 3, Chapter 30. 22-12-2000. Merck. 22-12-2000. Ref Type: Electronic Citation
2. Lower digestive tract. In: Boyd W, Sheldon H, eds. An introduction to the study of disease. Philidelphia: Lea & Febiger, 1977:282-94.
3. Greenberger NJ, Isselbacher KJ . Disorders of absorption. In: Braunwald E, Isselbacher KJ, Petersdorf RG, Wilson JD, Martin JB, Fauchi AS, eds. Principles of Internal Medicine. New York: McGraw-Hill Book Company, 1987:1260-75.
4. Mayes PA. Nutrition, digestion, & Absorption. In: Murray RK, Granner DK, Mayes PA, Rodwell VW, eds. Harper's Biochemistry. Norwalk: Appleton&Lange, 1988:571-88.
5. Carey MC, Small DM, Bliss CM. Lipid Digestion and absorption. Ann Rev Physiol 1983;45:651-77.
6. Sategna-Guidetti C, Bianco L. Malnutrition and malabsorption after total gastrectomy. J Clin Gastroenterol 1989;11:518-24.
7. Bruno MJ, Haverkort EB, Tytgat GN, van Leeuwen DJ. Maldigestion associeted with exocrine pancreatic insufficiency: implications of gastrointestinal physiology of enzyme preparations of cause-related and patient-tailored treatment. Am J Gasteroenterol 1995;90:1338-93.
8. Henderson R, Rinker AD. Gastric, pancreatic, and intestinal function. In: Burtis CA, Ashwood ER, eds. Tietz textbook of Clinical Chemistry. Philidelphia: W.B. Saunders Company, 1999:1283-97.
9. Wybenga DR, Inkpen JA. Lipids. In: Henry RJ, Cannon DC, Winkelman JW, eds. Clinical Chemistry: Principles and Techniques. New York: Harper & Row, 1974:1421-93.
10. Brindley DN. Metabolism of triacylglycerols. In: Vance DE, Vance J, eds. Biochemistry of lipids, lipoproteins and membranes. Amsterdam: Elsevier Science Publishers, 1991:171-203.
11. Carey MC, Hernall O. Digestion and absorption of fat. Sem Gastroint Dis 1992;3:189-208.
12. Friedman HI, Nylund B. Intestinal fat digestion, absorption, and transport. A review. Am J Clin Nutr 1980;33:1108-39.
13. Greenberger NJ, Toskes PP. Disorders of the pancreas; Approach to the patient with pancreatic disease. In: Braunwald E, Isselbacher KJ, Petersdorf RG, Wilson JD, Martin JB, Fauchi AS, eds. Principles of Internal Medicine. New York: McGraw-Hill Book Company, 1987:1368-84.
14. Van de Kamer JH. Standard methods of clinical chemistry. In: Seligson D, ed. New York: Academic Press, 1958:34.
15. Van de Kamer JH. Vet in faeces. In Dutch. 1-100. 3-5-1948. University of Utrecht, The Netherlands. Ref Type: Thesis/Dissertation
16. Van de Kamer JH, Ten Bokkel Huinink H, Weyers HA. Rapid determination of fat in feaces. J Biol Chem 1949;177:349-55.
17. Phuapradit P, Narang A, Mendonca P, Harris DA, Baum JD. The steatocrit: a simple method for estimation stool fat content in newborn infants. Arch Dis Child 1981;56:725-7.

18. Khouri MR, Huang, G, Shiau YF. Sudan stain of fecal fat: new insight into an old test. Gastrolenterol 1989;96:421-7.
19. Verkade HJ, Hoving EB, Muskiet FA, Martini IA, Jansen G, Okken A et al. Fat absorption in neonates: comparison of long-chain-fatty-acid and triglyceride composition of formula, feces and blood. Am J Clin Nutr 1991;53:643-51.
20. Benini L, Caliari S, Valona B, Talamini G, Vantini I, Guidi GC. Near infrared spectrometry for fecal fat measurement: comparison with conventional gravimetric and titrimetric methods. Gut 1989;30:1344-7.
21. Bekers O, Postma C, Lombarts AJPF. Determination of faecal fat by near-infrared spectroscopy. J Clin Chem Clin Biochem 1995;33:83-6.
22. Neumeister V, Jaross W, Henker J, Kaltenborn G. Simultaneous determination of fecal fat, nitrogen and water by Fourier transform near infrared reflectance spectroscopy through a polyethylene/polyamide film. J Near Infrared Spectrosc 1998;6:265-72.
23. Franck P, Sallerin JL, Schroeder H, Gelot MA, Nabet P. Rapid determination of fecal fat by Fourier transform infrared analysis (FTIR) with partial least-squares regression and an attenuated total reflectance accessory. Clin Chem 1996;42:2015-20.
24. The total solids, fat and nitrogen in the feces, 1: a study of normal persons and of patients with duodenal ulcer on a test diet containing large amounts of fat. Gastrolentelol 1946;6:83-104.
25. Dinning JD, Bowers JM, Trujillo MA. A standardized diet for performing fecal fat excretion. Arch Intern Med 1997;157:245-6.
26. Muller P. Klinische methoden. Scheikunde en microscopie, second ed. Utrecht: Erven J. Bijleveld, 1934:90-91pp.
27. Harrison's. Nephrolithiasis. In: Braunwald E, Isselbacher KJ, Petersdorf RG, Wilson JD, Martin JB, Fauchi AS, eds. Principles of Internal Medicine. New York: McGraw-Hill Book Company, 1987:1211-8.
28. Boer PW, van Geuns H, van der Hem GK, Blickman JR. Population survey in a community on the occurrence of stone disease. Proc VIIIth Congr Soc Int Urol 1979;2:64-5.
29. Knight J. Diagnosis and treatment of renal and uteral calculi. Alaska Med 1998;40:27-30.
30. Phyllips J. Kidney stones. Harv Mens Health Watch 4[9], 1-4. 2000. Palm Coast, Florida, Harvard Health Publications. Ref Type: Magazine Article
31. Volmer M, Vries JCM de, and oldschmidt HMJ. Fourier transform infrared analysis for the analysis of urinary calculi, using a single reflection diamond accessory and a combination of library search and a neural network interpretation algorithm. Clin Chem . 2001. Ref Type: In Press
32. Leonetti F, Dussol B, Berthezene P, Thirion X, Berland Y. Dietary and urinary risk factors for stones in idiopathic calcium stone formers compared to healthy subjects. Nephrol Dial Transplant 1998;13:617-22.
33. Laerum E. Urolithiasis in clinical practice. Occurrence, etiology, investigation and preventive treatment. Tidsskr Nor Laegeforen 1996;116:2897-902.
34. Pak CYC. Kidney stones. Lancet 1998;351:1797-801.
35. Trinchieri A. Epidemiology of urolithiasis. Arch Ital Urol Androl 1996;68:203-49.

36. Joual A, Rais H, Rabii R, el Mrini M, Benjelloun S. Epidemiology of urinary lithiasis. Ann Urol (Paris) 1997;31:80-3.
37. Newman DJ, Price CP. Renal function and nitrogen metabolites. In: Burtis CA, Ashwood ER, eds. Tietz textbook of Clinical Chemistry. Philidelphia: W.B. Saunders Company, 1999:1204-70.
38. Scheinman SJ. Nephrolithiasis. Semin Nephrol 1999;19:381-8.
39. Achilles W. In vitro crystallisation systems for the study of urinary stone formation. World J Urol 1997;15:244-51.
40. Funez AF, Cuerpo GE, Castellanos LF, Barrilero EA, Padilla AS, Palasi VJ. Epidemiology of urinary lithiasis in our unit. Clinical course in time and predictive factors. Arch Esp Urol 2000;53:343-7.
41. Hesse A, Sanders G. Atlas of infrared spectra for the analysis of urinary concrements, 1 ed. Stuttgart: George Thieme Verlag, 1998.
42. van der Hem GK, Beusekamp BJ, Rientjes MA. Nierstenen. In: Carbasius Weber EC, Post GB, Swager TW, eds. Informatorium voor voeding en diëtiek. Houten: Bohn Stafleu Van Loghum, 1989:27-30.
43. Rodgers AL. Effect of mineral water containing calcium and magnesium on calcium oxalate urolithiasis risk factors. Urol Int 1997;58:93-9.
44. Alford JA, Pierce DA, Suggs FG. Activity of microbial lipases on natural fats and synthetic triglycerides. J Lipid Res 1964;5:394.
45. McNeil NI, Cummings JH, James WPT. Short chain fatty acid absorption by the human large intestine. Gut 1978;19:819-22.
46. Miettinen TA. Intestinal and faecal bile-acids in malabsorption. Lancet 1968;10:358.
47. Wiggins HS, Howell KE, Kellock TD, Stalder J. The origin of feacal fat. Gut 1969;10:400-3.
48. Jakobs BS, Volmer M, Swinkels DW, Hofs MTW, Donkervoort S, Joosting MMJ et al. New method for faecal fat determination by mid-infrared spectroscopy, using a transmission cell: an improvement in standardization. Ann Clin Biochem 2000;37:343-9.
49. Cristie WW. High performance liquid chromatography and lipids. A practical guide. Oxford: Pergamom Press, 1987.
50. Mayes PA. Lipids of physiologic significance. In: Murray RK, Granner DK, Mayes PA, Rodwell VW, eds. Harper's Biochemistry. Norwalk: Appleton&Lange, 1988:130-41.
51. Lipid analysis: Isolation, separation, identification and structural analysis of lipids, Second ed. Oxford: Pergamon, 1982.
52. Kaluzny MA, Duncan LA, Merrit MV, Epps DE. Rapid separation of lipid classes in high yield and purity using bonded phase columns. J Lipid Res 1985;26:135-40.
53. Hoving EB, Jansen G, Volmer M, Van Doormaal JJ, Muskiet FAJ. Profiling of plasma cholesterol esters and triglyceride fatty acids as their methyl esters by capillary gas chromatography, preceded by rapid aminopropyl-silica column chromatographic separation of lipid classes. J Chromatogr 1988;434:395-409.
54. Muskiet FAJ, Van Doormaal JJ. Fatty acids, fatty acids containing lipids, and their gas chromatographic analysis. In: Clement RE, ed. Gas chromatography: Biochemical, biomedical and Clinical applications. John Wiley & Sons Inc., 1990:217-41.

55. Folch J, Lees M, Loane-Stanley GH. A simple method for the isolation and purification of total lipids from animal tissues. J Biol Chem 1957;226:497-509.
56. Miettinen TA. Gas-liquid chromatographic determination of fecal neutral sterols using a capillary column. Clin Chim Act 1982;124:245-8.
57. Street JM, Trafford DJH, Makin HLJ. The quantitative estimation of bile acids and their conjugates in huma biological fluids. J Lipid Res 1983;24:491-511.
58. Die Harnsteinconcretionen, ihre Entstehung, Erkennung, und Analyse mit besonderer Rücksicht auf Diagnose und Therapie de Nieren-und Blasenkrankung. Berlin: Trendler u Comp, 1860.
59. Hesse A, Voigt U, Hesse A, Breuer H. Ergebnisse aus Ringversuchen für Harnsteinanalysen. J Clin Chem Clin Biochem 1982;20:851-9.
60. Hahim IA, Zawawi TH. Wet vs. dry chemical analysis of renal stones. Ir J Med Sci 1999;168:114-8.
61. Larrson L, Sörbo B, Tiselius HG, hman S. A method for quantitative wet chemical analysis of urinary calculi. Clin Chim Act 1984;140:9-20.
62. Baadenhuijsen H. Niersteen - enquete 962. In Dutch. 962. 4-11-1996. Stichting Kwaliteitsbewaking Klinisch Chemische Ziekenhuislaboratoria. Ref Type: Report
63. Blijenberg BG, Baadenhuijsen H. De resultaten van verschillende analysetechnieken met betrekking tot urinesteenonderzoek. In Dutch. Ned Tijdschr Geneeskd 1990;134:1402-4.
64. Gorter E, de Graaff WC. Klinische diagnostiek. Deel 1. In Dutch. Leiden: Stenfert Kroese's, 1941:298-301pp.
65. La Granga TS. Calculi. In: Henry RJ, Cannon DC, Winkelman JW, eds. Clinical Chemistry. Principles and technics. New York: Harper & Row, 1974:1569-83.
66. Paulik F, Paulik J. TG and EGA investigations of the decomposition of magnesium ammonium phosphate hexahydrate by means of the derivatograph under conventional quasi-isothermal/quasi-isobaric conditions. J Thermal Anal 1975;8:557-66.
67. Corns CM. Infrared analysis of renal calculi: a comparison with conventional methods. Ann Clin Biochem 1983;20:20-5.
68. Uldall A. Strategies and methods for the analytical investigation of urinary calculi. Clin Chim Act 1986;160:93-101.
69. Röhle G, Voigt U, Hesse A, Breuer H. Ergebnisse aus Ringversuchen für Harnsteinanalysen. J Clin Chem Clin Biochem 1982;20:851-9.
70. Macromolecules, polymers and polymerization. Structure and properties of macromolecules. In: Morrison RT, Boyd RN, eds. Organic Chemistry. Boston: Allyn and Bacon, Inc, 1974:1045-9.
71. Bragg WH, Bragg WL. X-rays and crystal structure. London: G. Bell and Sons, ltd, 1915.
72. Dosch W. Röntgendiffraktometrie. In: Hesse A, Classen A, Röhle G, eds. Labordiagnostik bei Urolithiasis. Stuttgart: Seminar Urolithiasis, 1988:53-64.
73. den Boer NC, Bakker NJ, Leijnse B. Structuuranalyse van nier- en blaasstenen door röntgendiffractie. In Dutch. Ned Tijdschr Geneeskd 1972;116:373-7.
74. Exploring proteins. X-ray crystallography reveals three-dimensional structure in atomic detail. In: Stryer L, ed. Biochemistry. New York: W.H. Freeman and Company, 1988:59-62.

75. Wulkan RW. An expert system for interpretation of x-ray diffractograms. In: Expert systems and multivariate analysis in clinical chemistry. 16-24. 11-11-1992. Erasmus University, Rotterdam. Ref Type: Thesis/Dissertation
76. Wulkan RW, Zwang L, Liem TL, Blijenberg GB, Leijnse B. Renal stone analysis: LITHOS, an expert system for evaluation of X-ray diffractograms of urinary calculi. J Clin Chem Clin Biochem 1987;25:719-22.
77. Vergouwe DA, Verbeek RM, Oosterlinck W. Analysis of urinary calculi. Acta Urol Belg 1994;62:5-13.
78. Osland RC. Principles and practices of Infrared spectroscopy, third ed. Cambridge, England: Philips Scientific, 1989:5-100pp.
79. Skoog DA, West DM. Infrared absorption spectroscopy. Principles of instrumental analysis. Japan: Holt-Saunders, 1981:210-59.
80. Infraroodspectrometrie. In Dutch. In: de Galan L, ed. Analytische spectrometrie. Amsterdam, The Netherlands: Agon Elsevier, 1969:71-88.
81. Osborne BG, Fearn T, Hindle PH. Practical NIR spectroscopy with applications in food and beverage analysis, second ed. Essex, England: Longman Scientific & Technical, 1993.
82. Griffiths PG. Chemical Infrared Fourier Transform Spectroscopy. New York: John Wiley & Sons, 1975.
83. Harrick NJ. Internal reflectance spectroscopy. New York: Interscience Publishers, 1967.
84. Mirablella FM. Internal reflection spectroscopy: Theory and Applications. Practical Spectroscopy Series Volume 15. New York: Marcel Dekker, 1993.
85. Coates J, Sanders A. A universal sample handling system for FT-IR spectroscopy. Spectroscopy Europe 2000;12:12-22.
86. Hartcock MA, Messerschmidt RG. Infrared microspectroscopy theory and applications. New York: Marcel Dekker, 1998.
87. Coates J. A review of sampling methods for infrared spectroscopy. In: Workman J, Springsteen A, eds. Applied spectroscopy: A compact reference for practioners. San Diego: Academic Press, 1998.
88. Coleman PB. Practical sampling techniques for infrared analysis. Boca Raton: CRC Press, 1993.
89. van de Maas JH. Basic infrared spectroscopy. London: Heyden & Son, 1972.
90. River-Marcotegui A, Olivera-Olmedi JE, Valverde-Visus FS, Palacios-Sarrasqueta M, Grijalba-Uche A, Garcia-Mierlo S. Water, fat, nitrogen, and sugar content in feces: reference intervals in children. Clin Chem 1998;44:1540-4.
91. Paluszkiewicz C, Kwiatek WM, Galka M SD, Wentrup-Byrne E. FT-Raman, FT-IR spectroscopy and PIXE analysis applied to gallstones specimens. Cell Mol Biol 1998;44:65-73.
92. Shaw RA, Kotowich S, Leroux M, Mantch HH. Multianalyte serum analysis using mid-infrared spectroscopy. Ann Clin Biochem 1988;35:624-32.
93. Liu KZ, Ahmed MK, Dembinski TC, Mantsch HH. Prediction of fetal lung maturity from near-infrared spectra of amniotic fluid. Int J Gynecol Obstetrics 1997;57:161-8.
94. Savarino V, Vigneri S, Celle G. The C13 urea brath test in the diagnosis of Helicobacter pylori infection. Gut 1999;45:I18-22.

95. Braden B, Caspary WF, Lembcke B. Nondispersive infrared spectrometry for 13CO2/12CO2-measurements: a clinically feasable analyzer for stable breath tests in gastroenterology. Z Gastroenterol 1999;37:477-81.
96. Rigas B, Morgello S, Goldman IS, Wong PTT. Human colorectal cancers display abnormal Fourier-transform infrared spectroscopy. Proc Natl Acad Sci USA 1990;87:8140-4.
97. Wong PTT. High pressure chemistry and biochemistry. In: van Eldik, Jonas R, eds. NATO ASI Series. C Vol. 197. Dordrecht, The Netherlands: Reidel, 1987:381-400.
98. Yano K, Ohoshima S, Gotou Y, Kumaido K, Moriguchi T, Katayama H. Direct measurement of human lung cancerous and noncancerous tissues by fourier transform microscopy: can an infrared microscope be used as a clinical tool. Anal Biochem 2000;287:218-25.
99. Lasch P, Naumann D. FT-IR microspectroscopic imaging of human carcinoma thin sections based on pattern recognition techniques. Cell Mol Biol 1998;44:189-202.
100. Cohenford MA, Godwin TA, Cahn F, Bhandare P, Caputo TA, Rigas B. Infrared spectroscopy of normal and abnormal cervical smears: evaluation by principal component analysis. Gynaecol Oncol 1997;66:59-65.
101. Heise HM, Bittner A, Marbach R. Near-infrared reflectance spectroscopy for non-inasive monitoring of metabolites. Clin Chem Lab Med 2000;38:137-45.
102. Quaresima V, Sacco S, Totaro R, Ferrari M. Noninvasive measurement of cerebral hemoglobin oxygen saturation using two near infrared spectroscopy approaches. J Biomed Opt 2000;5:201-5.
103. Robinson MR, Eaton RP, Haaland DM, Koepp GW, Thomas EV, Stallard BR. Noninvasive glucose monitoring in diebetic patients: a preliminary evaluation. Clin Chem 1992;38:1618-22.
104. Kajiwara K Uemura T, Kishikawa H, Nishida K, Hashiguchi Y, Uehara M. Non-invasive measurement of blood glucose concentrations using FT-IR spectra through the oral mucosa. Med Biol Enggn Comput 1993;31(suppl):17-22.
105. Heise HM, Marbach R. Human oral mucosa studies with varying blood glucose concentration by non-invasive ATR-FT-IR-spectroscopy. Cell Mol Biol 1998;44:899-912.
106. Brown SD, Blank TB, Sum ST, Weyer LG. Chemometrics. Anal Chem 1994;66:315R-59R.
107. Goldschmidt HMJ, Leijten JF. Medicometrics: a new promising discipline. In: Trendelenburg C, Brackenheim, eds. Computing in clinical laboratories, Proceedings of the fifth international conference. Stuttgart, FRG: Georg Kohl GmbH and Co, 1985:117-8.
108. Hemel JB, van der Voet H. The CLAS program for classification and evaluation. Anal Chim Acta 1986;191:33-45.
109. Beck JR. Laboratory deciision science applied to chemometrics: strategic testing of thyroid function. Clin Chem 1986;32:1707-13.
110. Holst H, Ohlsson M, Peterson C, Edenbrandt L. A confident decision support system for interpreting electrocarrdiograms. Clin Physiol 1999;19:410-8.

111. Sun M, Sclabassi RJ. The forward EEG solutions can be computed using artificial neural networks. Trans Biomed Eng 2000;47:1044-50.
112. Vriezema JL, van der Poel HG, Debruyne FM, Schalken JA, Kok IP Boon ME. Neural network-based digitized cell image diagnosis of bladder wash cytology. Diagn Cytopathol 2000;23:171-9.
113. Kok MR, Boon ME, Schreiner-Kok PG, Koss LG. Cytological recognition of invasive squamous cancer of the uterine cervix: comparison of conventional light-microscopical screeening and neural networtwork-based screening. Hum Pathol 2000;31:23-8.
114. Vos JE, Scheepstra KA. Computer-simulated neural networks: an appropriate model for motor development. Early Hum Dev 1993;34:101-12.
115. Valdiguie PM, Rogari E, Corberand JX, Boneu B. The performance of the knowledge-based system VALAB revisited: an evaluation after 5 years. Eur J Clin Chem Clin Biochem 1996;34:371-6.
116. Oosterhuis WP, Ulenkate HJ, Goldschmidt HMJ. Evaluation of LabRespond, a new automated validation system for clinical laboratory test result. Clin Chem 2000;46:1811-7.
117. Brown SD, Sum ST, Despagne F. Chemometrics. Anal Chem 1996;68:21R-61R.
118. Levine BK. Chemometrics. Anal Chem 1998;70:209R-28R.
119. Levine BK. Chemometrics. Anal Chem 2000;72:91R-7R.
120. Massart DL, Vandeginste BGM, Deming SN, Michotte Y, Kaufman L. Chemometrics: a text book. Amsterdam: Elsevier, 1988.
121. Kleinbaum DG, Kupper LL, Muller KE. Applied regression analysis and other multivariate methods, second ed. Boston: PWS-KENT Publishing Company, 1987.
122. Mardia KV, Kent JT, Bibby JM. Multivariate analysis. London: Academic Press, 1988.
123. Morrison DF. Multivariate statistical methods, Second ed. London: McGraw-Hill, 1988.
124. Davies AMC. The value of pictures. Spectroscopy Europe 1998;10:28-31.
125. Davies AMC. More pictures from PLS regression analyses. Spectroscopy Europe 1998;10:20-2.
126. Martens H, Næs T. Multivariate calibration. Chichester: John Whiley & Sons, 1991.
127. Stevens J. Applied multivariate statistics for the social sciences. London: Lawrence Erlbaum Assosiates Inc., 1986.
128. Afifi AA, Clark V. Computer aided multivariate analysis, third ed. London: Chapman & Hall, 1996.
129. Kaiser HF. The application of electronic computers to factor analysis. Educational and Psychological Measurement 1960;20:141-51.
130. Cattell RB. The meaning and strategic use of factor analysis. In: Cattell RB, ed. Handbook of multivariate experimental psychology. Chicago: Rand McNally, 1966:174-243.
131. Voet H van der and Hemel J. Multivariate classification methods and their evaluation in applications. 1988. University of Groningen, The Netherlands. 30-5-1988. Ref Type: Thesis/Dissertation
132. Gorsuch RL. Factor analysis. New Yersey: Lawrence Erlbaum, 1983.

133. Flury B, Riedwyl H. Multivariate statistics. A practical approach. New York: Chapman and Hall, 1988.
134. Stone M. Cross-validatory choice and assessment of statistical prediction. J Roy Stat Soc 1974;B:111-33.
135. Davies AMC. The tender trap: sample selection programs and their misuse. Spectroscopy Europe 1996;8:27-9.
136. Hinton GE. How neural networks learn from experience. Scientific American 1992;September:105-9.
137. Zupan J, Gasteiger J. Neural networks: A new method for solving chemical problems or just a passing phase. Anal Chim Acta 1991;248:1-30.
138. Bos A, Bos M, Linden WE van der. Artificial neural networks as a tool for soft-modelling in quantitative analytical chemistry: the prediction of water content in cheese. Anal Chim Acta 1992;256:133-44.
139. Hopfield JJ. Neural networks and physical systems with emergent collective abilities. Proc Natl Acad Sci USA 1982;79:2554-8.
140. Kohonen T. Self-organized formation of topologically correct feature maps. Biol Cybernetics 1982;43:59.
141. Rumelhart DE, Hinton GE, Williams RJ. Learning internal representation by error propagation. In: Rumelhart DE, McClelland JL, eds. Parallel distributed processing: Exploration in the microstructure of cognition. Vol1: Foundations. Cambridge, MA: MIT Press, 1986:318-62.
142. Widraw G, Hoff ME. Adaptive switching circuits. IRE WESCON Convention Record, Part 4 1960;96-104.
143. Minsky ML, Papert S. Perceptrons, an essay in computional geometry. MIT Press, 1969.
144. Bos A. Artificial neural networks as a tool in chemometrics. 27-5-1993. University of Twente, The Netherlands. Ref Type: Thesis/Dissertation
145. Bertsekas DP, Tjistiklis JN. Neuro-dynamic programming. Belmond, MA: Athena Scientific, 1996.
146. White H. Some asymptotic results for learning in single hidden layer feedforward network models. J of the American Statistical Assoc 1989;84:1008-13.
147. Bertsekas DP. Nonlinear programming. Belmont, MA: Athena Scientific, 1995.
148. Sarle WS. Stopped training and other remedies for overfitting. Proceedings of the 27th symposium on the interface of computing science and statistics 1995;352-60.
149. Fahlman SE. Faster-learning variations on back-propagation: An empirical study. In: Touretzky D, Hinton G, Sejnowski T, eds. Proceedings of the 1988 Connectionist Models Summer School, Morgan Kaufmann. 1989:38-51.
150. Riedmiller M, Braun H. A direct adaptive method for faster backpropagation learning: The RPROP algorithm. Proceedings of the IEEE International Conference on Neural Networks. San Fransisco: IEEE, 1993.
151. Zupan J, Gasteiger J. Neural networks in chemistry and drug design, second ed. Weinheim: Wiley-VCH, 1999.
152. Smits JRM, Melssen WJ, Buydens LMC, Kateman G. Using artificial neural networks for solving chemical problems. Part I. Multi-layer feed-forward networks. Chemom Intel Lab Syst 1994;22:165-89.

153. VanTrappen GR, Rutgeerts PJ, Ghoos YF, Hiele MI. Mixed triglyceride breath test: a noninvasive test of pancreatic lipase activity in the duodenum. Gastroenterology 1989;96:1126-34.
154. Kalivianakis M, Verkade HJ, Stellaard F, Werf van der M, Elzinga H, VOnk RJ. The $^{13}$C-mixed triglyceride breath test in healthy adults: determination of the $^{13}CO_2$ response. Eur J Clin Invest 1997;27:434-42.
155. Esler MB, Griffith DW, Wilson SR, Steele LP. Precision trace gas analysis by FT-IR spectrocopy. The $^{13}C/^{12}C$ isotope ratio of $CO_2$. Anal Chem 2000;72:216-2.
156. Kalivianakis M, Minich DM, Bijleveld CMA, Aalderen van WMC, Stellaard F, Lasuer M et al. Lipid malabsorption in cystic fibrosis patients on enzyme replacement therapy is due to impaired intestinal uptake of long-chain fatty acids. Am J Clin Nutr 1999;68:127-34.
157. Diem M. Introduction to modern vibrational spectroscopy. New York: Wiley, 1993.
158. Hanlon EB, Manoharan R, Koo TW, Shafer KE, Motz JT, Fitzmaurice M et al. Prospects for in vivo Raman spectroscopy. Phys Med Biol 2000;45:R1-R59.
159. Sudlow K, Woolf A. Identification of renal calculi by their Raman spectra. Clin Chim Act 1991;203:387-94.
160. Pelletier M, Davis K. Raman spectroscopy: The next generation. International Laboratory 1996;may:11D-G.
161. Ishida H, Kamoto R, Ishitani A, Ozaki Y, et al. Raman microprobe and Fourier transform-infrared microsampling studies of the microstructur eof gallstones. Appl Spectrosc 1987;41:407-12.
162. Kodati VR, Tomasi GE, Turmin JL, Tu AT. Raman spectroscopi identification of phosphate-type kidney stones. Appl Spectrosc 1991;45:581-3.
163. Buschman HPJ, Puppels GJ, Römer TJ, Laarse van der A, Bruschkle AVG. In vitro intravascular histochemistry of human atherosclerosis by Raman spectroscopy. Preceedings of the Vibrational spectroscopy: New tool in Medicine Potsdam Germany 1998.
164. Redd DCB, Feng ZC, Yue KT, Gansler TS. Raman spectroscopi characterization of human breast tissues: Implications for breast cancer diagnosis. Appl Spectrosc 1993;47:787-91.
165. Choo-Smith LP, Maquelin K, Endtz HP, Bruining HA, Puppels GJ. The challenge of rapidly characterizing (drug resistant-)microorganisms: A confocal Raman microspectroscopic approach. Preceedings of the Vibrational spectroscopy: New tool in Medicine Potsdam Germany 1998.
166. Berger AJ, Koo TW, Itzkan I, Horowitz G, Feld MS. Multicomponent blood analysis by near-infrared Raman spectroscopy. Appl Opt 1999;38:2916-26.
167. Koo TW, Berger AJ, Itzkan I, Horowitz G, Feld MS. Reagentless blood analysis by near-infrared Raman spectrocopy. Diabetes Technol Terapeu 1999;1:153-7.
168. Klonoff. Editorial: Noninvasive laser measurment of blood glucose in the eye: A bright idea or an optical illusion. Diabetes Technol Terapeu 1999;1.
169. Holland JH. Adaptations in natural and artificial systems. New York: MIR Press, 1975.
170. Jouan-Rimbaud D, Massart D-L, Leardi R, Noord de O. Genetic algorithms as a tool for wavelength selection in multivariate calibration. Anal Chem 1995;67:4295-301.

171. Broadhurst D, Goodacre R, Jones A, Rowland JJ, Kell DB. Genetic algorithms as a method for variable selection in multiple linear regression and partial least squares regression with applications to pyrolysis mass spectroscopy. Anal Chim Acta 1997;348:86.
172. Arcos MJ, Ortiz MC, Villahoz B, Sarabia LA. Genetic-algorithm-based wavelength selection in multicomponent spectrometric determination by PLS: application on indomethacin and acemethacin in mixture. Anal Chim Acta 1997;339:63-77.
173. Coley DA. An introduction to genetic algorithms for scientist and engineers. Signapore: World Scientific, 1999.
174. Fearn T, Davies T. Genetic algorithms: the evolutionary solution to an old problem. Spectroscopy Europe 1997;9:25-7.
175. Beasley D, Bull RD, Martin RR. An overview of genetic algorithms: Part 1, Fundamentals. University of Cardiff, Cardiff, UK: Department of computing mathematics, 1993.
176. Goldberg DE. Genetic algorithms in search, optimization and machine learning. New York: Addison-Wesley, 1989.