

# 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

<b>Scope of the thesis</b>	1
<b>Introduction</b>	3
1. Patho-physiology	4
1.1. Steatorrhea	4
1.2. Urolithiasis	8
2. Analytical methods:	14
2.1. Reference methods:	14
2.1.1. Fecal fat analysis:	14
2.1.1.1. Van de Kamer method	14
2.1.1.2. Gas chromatography	17
2.1.2. Urinary calculus analysis:	21
2.1.2.1. Wet and dry chemical analysis	21
2.1.2.2. X-ray diffraction	24
2.2. Infrared spectroscopy and sample handling techniques	27
3. Chemometrics	39
3.1. General	39
3.2. Multivariate calibration methods	42
3.2.1. PLS regression	42
3.2.2. Neural networks	56
3.3. Spectral library search	69
4. Future trends (state of the art)	73
References	79

<b>Part I. The analysis of fecal fat</b>	89
Chapter 1    New method for fecal fat determination by mid-infrared spectroscopy, using a transmission cell: an improvement in standardization. <i>Ann Clin Biochem 2000;37:343-349.</i>	91
Chapter 2    An enhanced time-saving extraction procedure for the analysis of fecal fat by means of Fourier Transform Infrared spectroscopy. <i>Clin Chem 2000;46(7):1019-1020.</i>	101
Chapter 3    A new NIR spectroscopic method for the determination of fecal fat. <i>Ann Clin Biochem 2001;38:256-263.</i>	105
<b>Part II. Urinary calculus analysis</b>	117
Chapter 4    Partial Least-Squares regression for routine analysis of urinary calculus composition with Fourier Transform Infrared analysis. <i>Clin Chem 1993;39(6):948-954.</i>	119
Chapter 5    Artificial neural network predictions of urinary calculus compositions analyzed with infrared spectroscopy. <i>Clin Chem 1994 ;40(9):1692-1697.</i>	133
Chapter 6    Infrared analysis of urinary calculi by a single reflection accessory and a neural network interpretation algorithm. <i>Clin Chem 2001;47(7):1287-1296.</i>	145
Chapter 7    Neuranet: The computer application for analyzing IR spectra of urinary calculi using an artificial neural network and library search.	163
<b>Summary</b>	193
<b>Samenvatting</b>	199
<b>Dankwoord</b>	205





## 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<sup>2</sup>). 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).

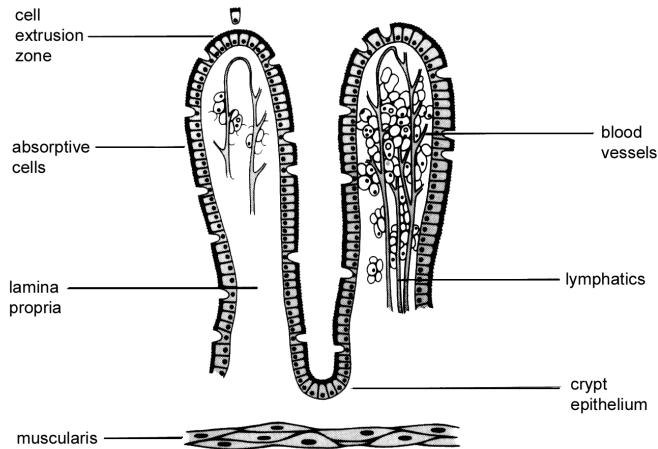


Figure 1. Villus of the intestinal mucosa. The epithelial cells that cover the surface of the villus absorb dietary molecules. At the apex of each villus the cells are sloughed off (2).

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<sub>2</sub>. 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  $\mu\text{mol/ml}$ ), 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 unilateral diffusion is promoted, because fatty acids and monoglycerides of long-chain fatty acids ( $\geq\text{C14}$ ) 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).

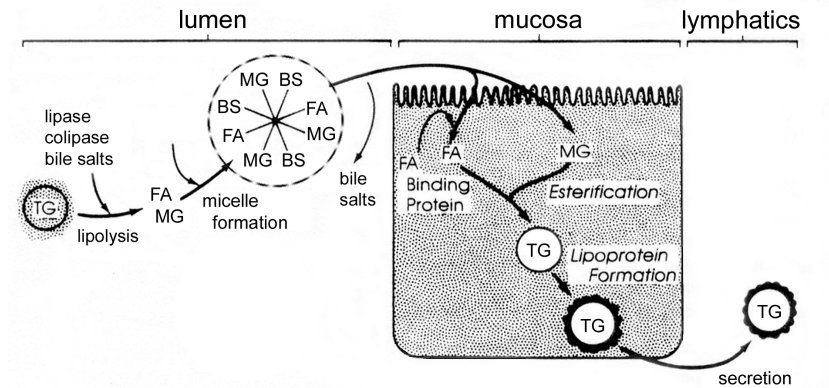


Figure 2. Diagram of intestinal digestion, absorption, esterification, and transport of dietary triglycerides. TG= triglycerides; FA = fatty acids; MG = monoglycerides; BS = bile salts (3).

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  $\geq 20$  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:

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

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 too 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 then 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  $\text{Ca}^{++}$ , urate, oxalate, pH  
Decreased  $\text{Mg}^{++}$ , 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	$\text{CaC}_2\text{O}_4 \cdot \text{H}_2\text{O}$ / Calcium oxalate	75.0	Hyperoxaluria, hypercalciuria, hyperuricosuria, hyperuricaemia and primary hyperparathyroidism
Weddellite	$\text{CaC}_2\text{O}_4 \cdot 2\text{H}_2\text{O}$ / Calcium oxalate	70.7	See whewellite
Carbonate apatite	$\text{Ca}_{10}(\text{PO}_4)(\text{CO}_3\text{OH})_6(\text{OH})_2$ / Calcium phosphate	48.9	Hypercalciuria, renal tubular acidosis (RTA), urinary tract infection (not essential), hyperphosphaturia and immobilization
Brushite	$\text{CaHPO}_4 \cdot 2\text{H}_2\text{O}$ / Calcium hydrogen phosphate	13.0	Hypercalciuria, hyperphosphaturia, RTA and immobilisation
Struvite	$\text{MgNH}_4\text{PO}_4 \cdot 6\text{H}_2\text{O}$ / Magnesium ammonium phosphate	4.3	Urinary tract infection with urease producing bacteria
Uric acid	$\text{C}_5\text{H}_4\text{N}_4\text{O}_3$	3.3	Hyperuricosuria and hyperuricaemia
Ammonium urate	$\text{C}_5\text{H}_7\text{N}_5\text{O}_3$	1.1	Hyperuricosuria and urinary tract infection
Cystine	$\text{C}_6\text{H}_{12}\text{N}_2\text{O}_4\text{S}_2$	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 extracorporeal 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 ( $\text{Ca}^{2+}$  and  $\text{Mg}^{2+}$  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 ( $\leq \text{C6}$ ) 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 ( $\text{mw fatty acids} = \text{mg} / \text{mmol fatty acids}$ ). 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 de Kamer method, is defined as:

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

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 <sup>+</sup> and K <sup>+</sup> soaps	70
Ca <sup>2+</sup> and Mg <sup>2+</sup> 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 derivatives, 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.

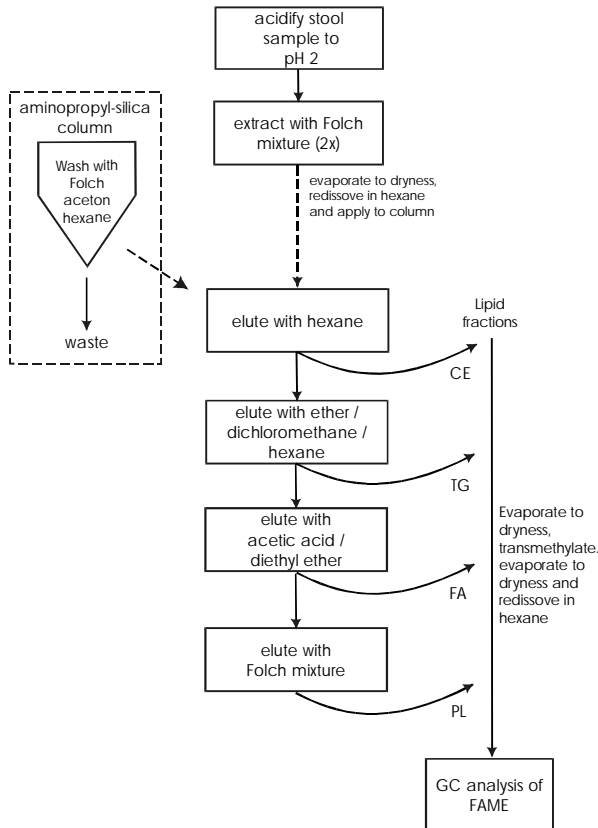
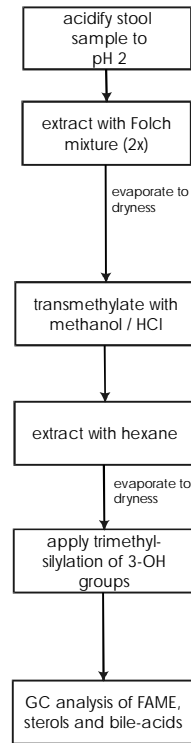


Figure 4. GC analysis of FAME sterols and bile-acids in total fecal lipid.



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

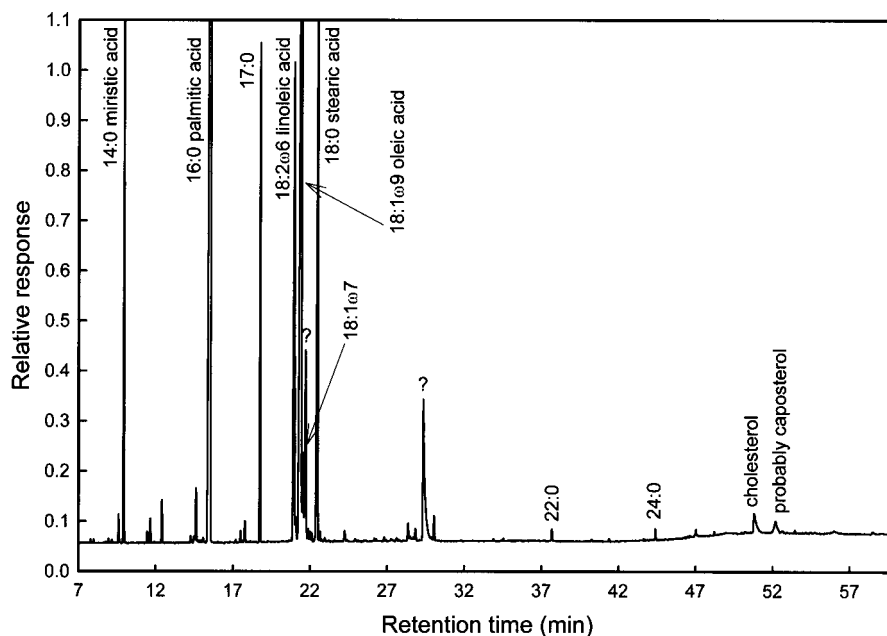


Figure 5. GC transmethyated fatty acid profile of feces. The relative response is defined as the peak height of the respective fatty acids in relation to the height of the internal standard (17:0).

### 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 ( $\text{MgNH}_4\text{PO}_4 \cdot 6\text{H}_2\text{O}$ ) 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<sub>2</sub> 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.

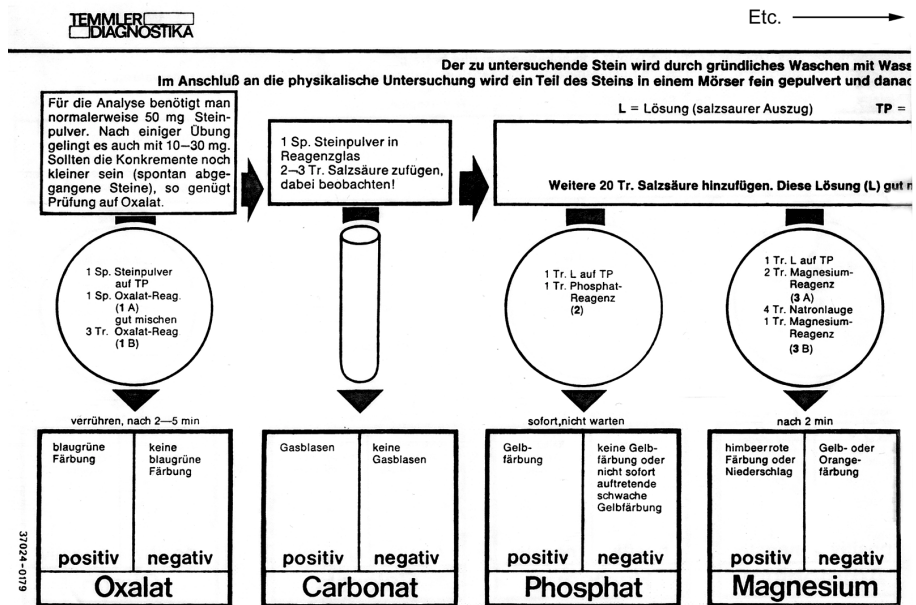


Figure 6. Part of the procedure of the Temmler kit for qualitative analysis of urinary calculus compositions.

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.

2.1.1.2.2. 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 interference 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,  $\theta$ ). 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 Debye-Scherrer-Hull camera on X-ray photographic film (73), or with a solid state electronic detector (74). The films, obtained with the Debye-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 Debye-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  $0^\circ - 50^\circ$  over the sample, the detector records the measured intensity as a function of the diffraction angle.

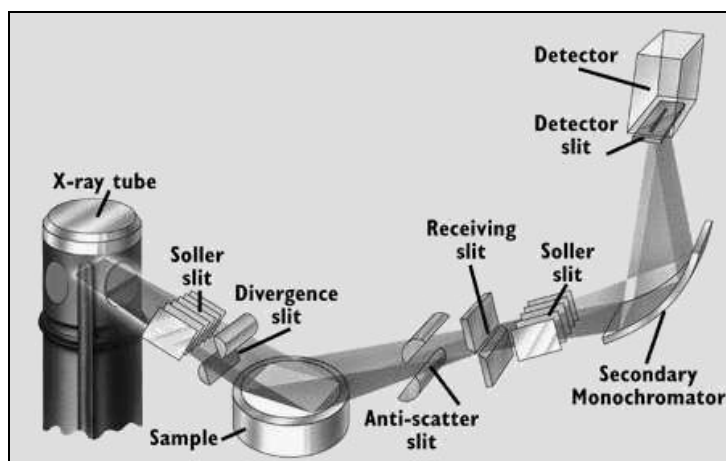


Figure 7. Diagram of a X-ray diffractometer

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  $2\theta$ ), or  $\theta$  (degrees  $\theta$ ), 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  $[\text{Ca}_{10}(\text{PO}_4)_6(\text{OH})_2]$ . The samples are measured from diffractor angles  $7^\circ$  to  $49.5^\circ$  ( $2\theta$ ), in steps of  $0.01^\circ$ . 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).

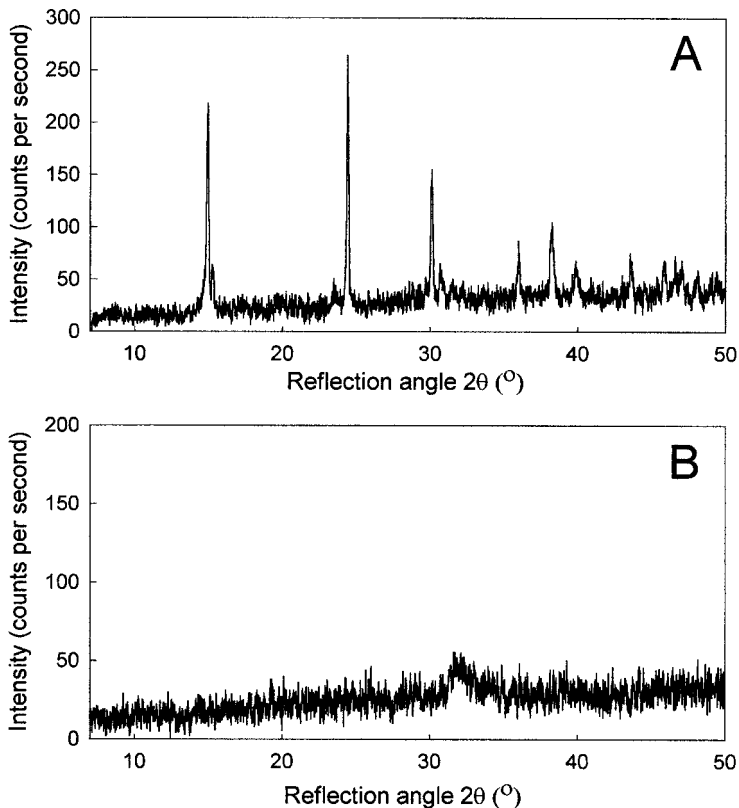


Figure 8. XRD spectrum of whewellite (A) and apatite (B). The data of the diffractograms were obtained from the Clinical Chemical laboratory of the University Hospital of Rotterdam, The Netherlands

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  $\mu\text{m}$  to 1000  $\mu\text{m}$ , 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).

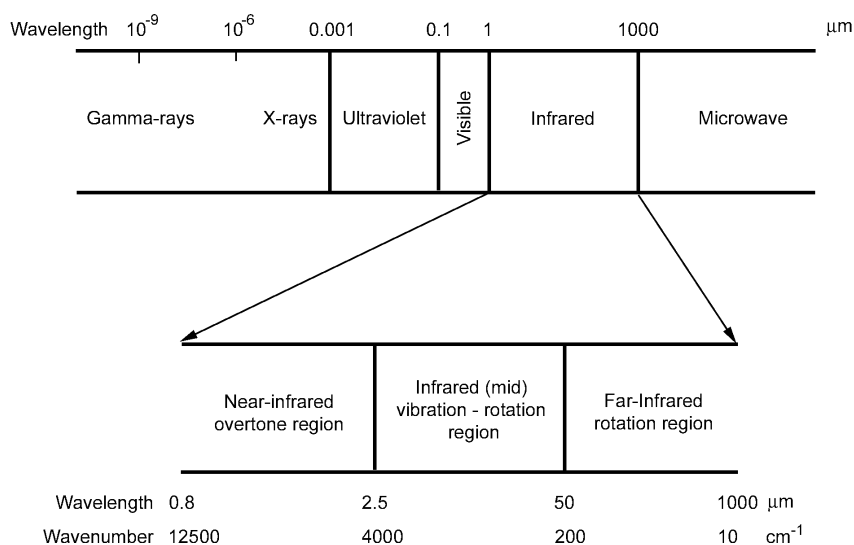


Figure 9. Infrared region of the electromagnetic spectrum.

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 ( $V_0-V_1$ )  $\pm 20 \text{ cm}^{-1}$ . The same is true for higher

overtone. For example, the first overtone band of the strong  $\text{CH}_2$  band at  $890\text{ cm}^{-1}$  is found at  $1780\text{ cm}^{-1}$ .

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 ( $\text{CO}_2$ ) 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  $\text{CO}_2$  molecule (Figure 10A) (80).

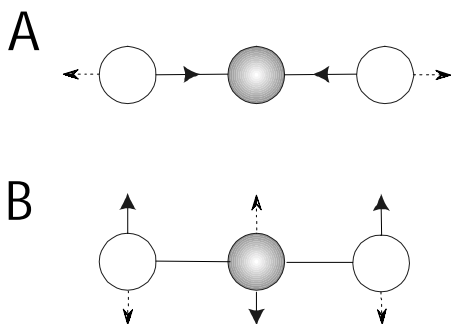


Figure 10. Symmetric stretch vibration (A) of  $\text{CO}_2$ . Infrared inactive. Symmetric bending vibration (B) of  $\text{CO}_2$ . Infrared active.

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.  $\text{CO}_2$ ), 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 inorganic component. For this reason, IR spectra represent one of the unique physical properties of organic and some inorganic 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 ( $200\text{--}10\text{ cm}^{-1}$ ). 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 ( $4000\text{--}200\text{ cm}^{-1}$ ) and the near IR ( $12500\text{--}4000\text{ cm}^{-1}$ ) 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 ( $\epsilon$ ). 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  $\text{CO}_2$ ), 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  $\text{cm}^{-1}$ .

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  $400\text{ cm}^{-1}$ , 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  $\text{cm}^{-1}$ .

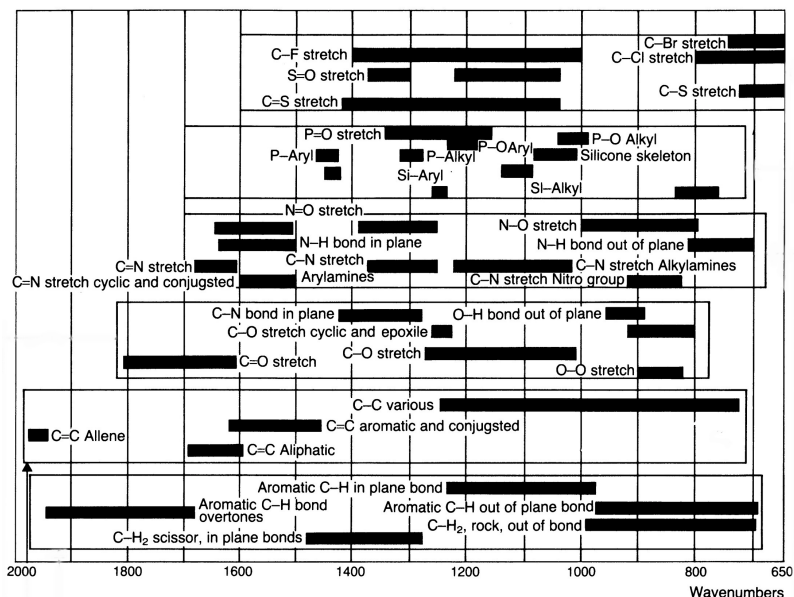


Figure 11. Correlation chart with a number of functional groups in the IR region from 2000–650  $\text{cm}^{-1}$ .

### 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).

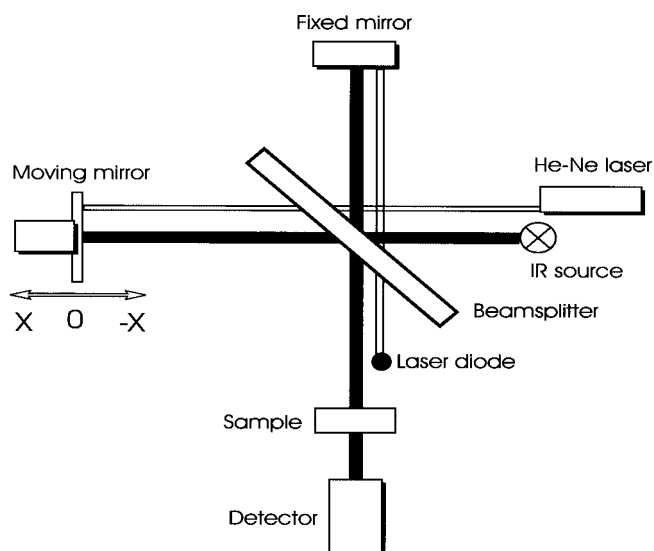


Figure 12. Michelson interferometer for signal-modulation in FT-IR spectroscopy.

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	Short scan times of full scans
Measuring each resolution element separately	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

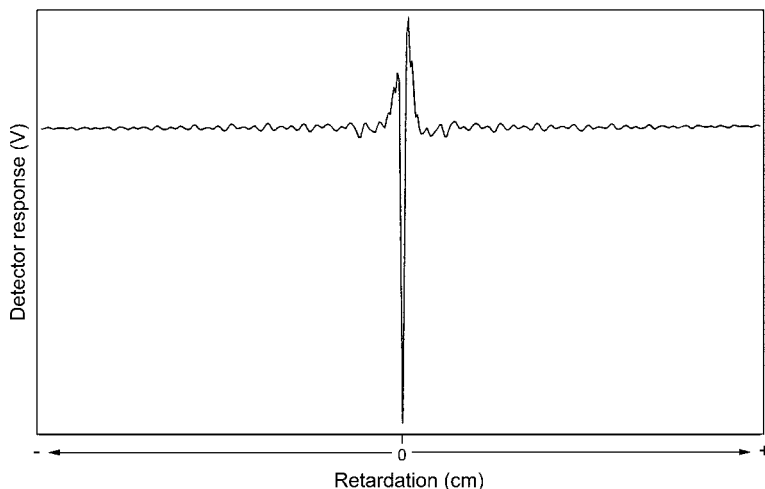


Figure 13. Interferogram of a continuous IR source.



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  $\text{cm}^{-1}$ . 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  $\text{cm}^{-1}$ . 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 $\text{cm}^{-1}$	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.

Sample category	Sampling techniques	
	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 ( $\geq 1$  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 ( $2.5 \mu\text{m} = 4000 \text{ cm}^{-1}$ ). 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  $45^\circ$ . ATR can only be effective if the difference of the RI between the sample and the ATR crystal is large enough ( $\Delta \text{RI} \geq 1$ ). 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 ( $45^\circ$ ,  $30^\circ$  or  $60^\circ$ ) 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).

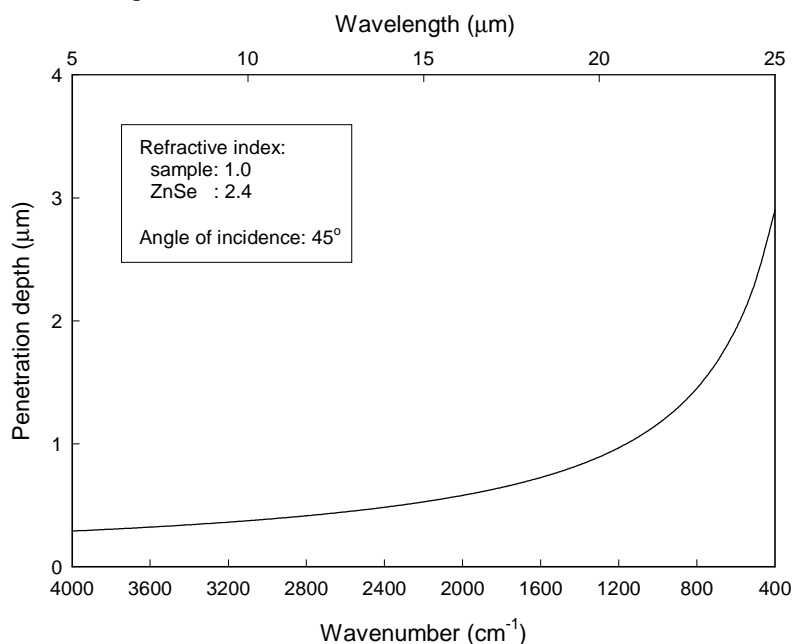


Figure 14. Relation between the depth of penetration and the wavelength using ZnSe as ATR crystal.

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.

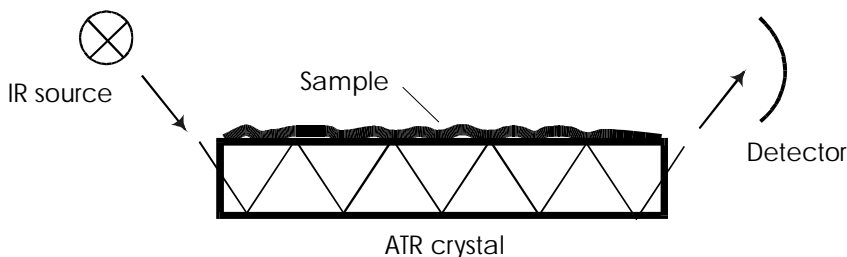


Figure 15. Multiple reflection horizontal ATR system.

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 weakly 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).

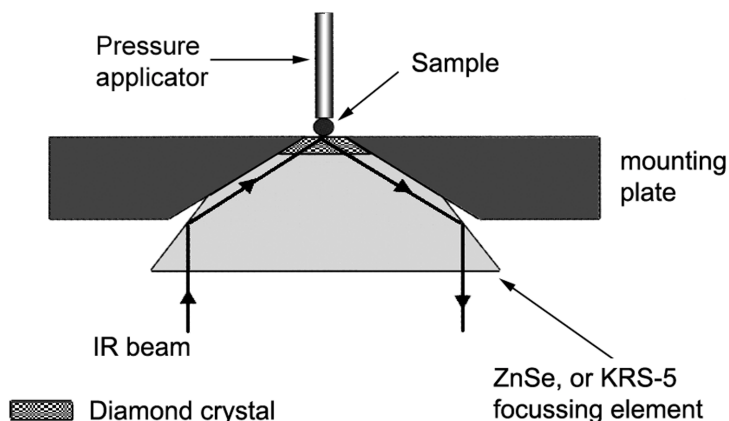


Figure 16. Single reflection micro-ATR sampling device.

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 al (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  $^{13}\text{C}$  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  $^{13}\text{C}$ -urea into ammonia and  $^{13}\text{CO}_2$ , which diffuse into the blood. The  $^{13}\text{CO}_2/^{12}\text{CO}_2$  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<sub>2</sub> 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](#)).

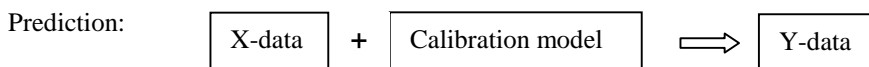
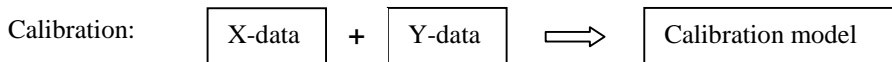
### *Data structures used in chemometrics*

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.





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):

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

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.

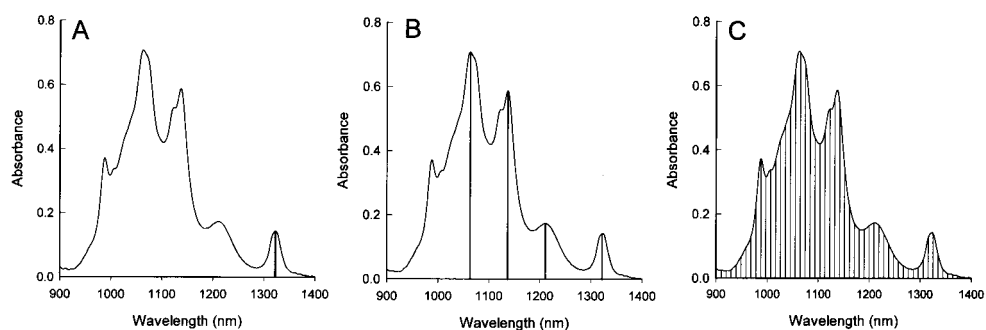


Figure 17. Part of an IR spectrum showing the absorbance band(s) at one or more wavelengths that are used for linear regression (A), multiple linear regression (B) and partial least-squares regression (C).

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:

$$A = \epsilon cl \quad (A, \text{Absorbance}; c, \text{concentration}; \epsilon, \text{molar extinction coefficient}; l, \text{pathlength})$$

- 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 interferences 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:

$$\begin{aligned} T &\leftarrow f_1(X) && (T, \text{represents systematic structure, possibly unidentified}) \\ X &\leftarrow f_3(T) + E && (E, \text{residuals}) \\ Y &\leftarrow f_2(T) + F && (F, \text{residuals}) \end{aligned}$$

Both Principal Component Regression (PCR) and Partial Least 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:

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

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  $\beta$ ) of the previous model may be described as follows:

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

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:

$$\hat{\beta} = (X'X)^{-1}X'y \quad (121)^1. \text{ The } y\text{-values of unknown samples may be predicted by using the}$$

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:

---

<sup>1</sup>  $X'$ , 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,  $\beta_1$  and  $\beta_2$  are calculated as follows:

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

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  $\beta$  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.:

$$x_1 = X_1 - \bar{X}_1, \quad x_2 = X_2 - \bar{X}_2, \dots \dots x_k = X_k - \bar{X}_k \text{ (for } k \text{ } X\text{-variables)}^2$$

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

$$\begin{aligned} t_1 &= v_{11}X_1 + v_{12}X_2 + \dots + v_{1k}X_k & t_1 \text{ is called the first principal component} \\ t_2 &= v_{21}X_1 + v_{22}X_2 + \dots + v_{2k}X_k & t_2 \text{ is called the second principal component} \\ \text{etc.} & & t_k \text{ is called the } k\text{-th principal component} \end{aligned}$$

---

<sup>2</sup>  $\bar{X}$ , mean of variable  $X$

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

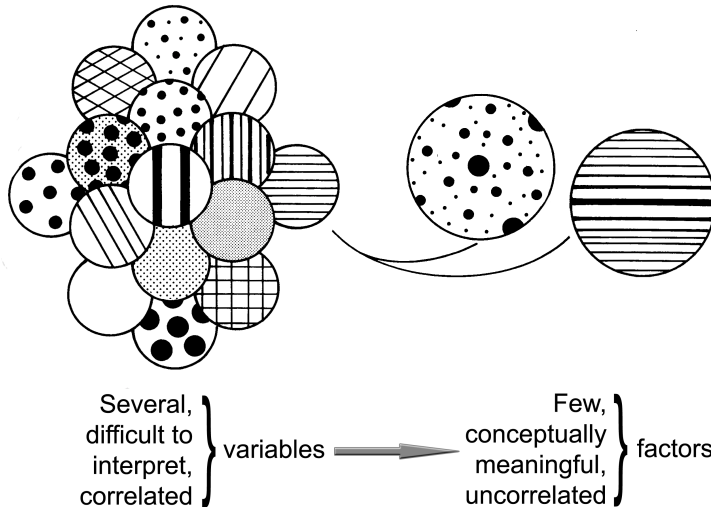


Figure 18. General purpose of principal component analysis.

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 [ $\text{Var}(t_1) \geq \text{Var}(t_2) \geq \dots \geq \text{Var}(t_k)$ ]
2. The  $k$  principal components  $t_1, t_2, \dots, t_k$  are uncorrelated
3.  $v_1'v_1 = v_2'v_2 = \dots = v_k'v_k = 1$  (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  $x_2$  ( $x_i = X_i - \bar{X}_i$ ,  $i=1, 2$ ) (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 ( $ss_{\text{total}}^2 = \sum_{i=1}^k \lambda_i$ ). 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  $v'_j = (v_{j1}, v_{j2}, \dots, v_{jp})$  is the row-vector with the coefficients of the  $p$  original variables corresponding to the  $j$ -th largest eigenvalue  $\lambda_j$ , then 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).

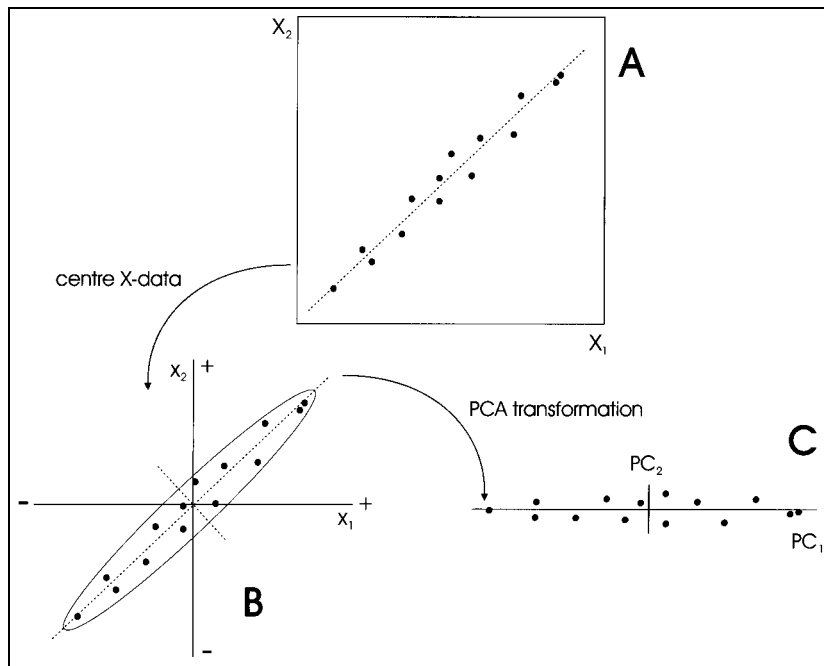


Figure 19. Principle of Principal Component Analysis. Scatterplots of bivariate hypothetical data (A), centered X-data (B) and after PCA transformation (C)



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  $\{x_i, i = 1, 2, \dots, k\}$ . The Kaiser criterion (129) is probably the most widely used criterion to select the number of components  $\hat{T} = \{\hat{t}_1, \dots, \hat{t}_A\}$  ( $A < k$ ). 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:

$$X = TP' + E \quad \text{in which } P' \text{ is the transposed loading matrix and } E \text{ the residuals}$$

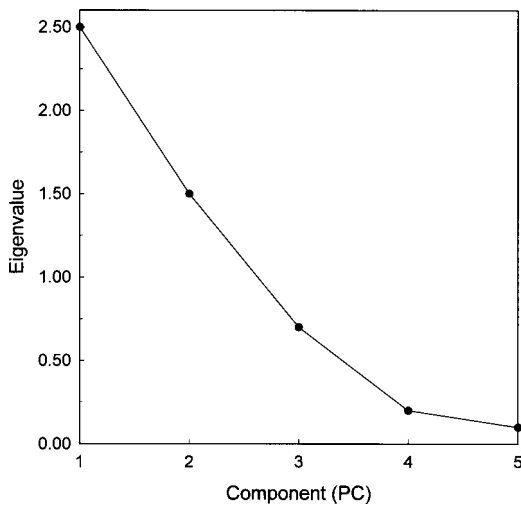


Figure 20. Scree plot expressing the relation between the eigenvalues and the component numbers

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 losing 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).

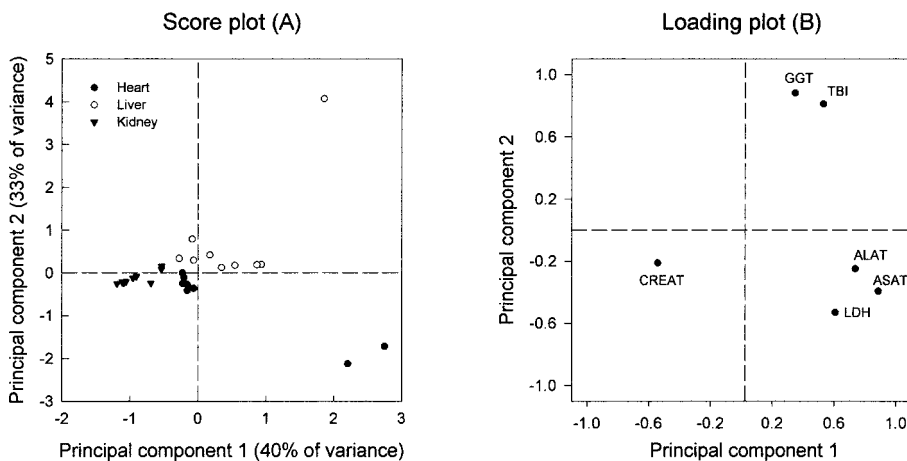


Figure 21. Score and loading plot of the data set of Hemel et al.

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  $\rightarrow$  two components), and can be used for classification.

It 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  $\leq 5$  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).

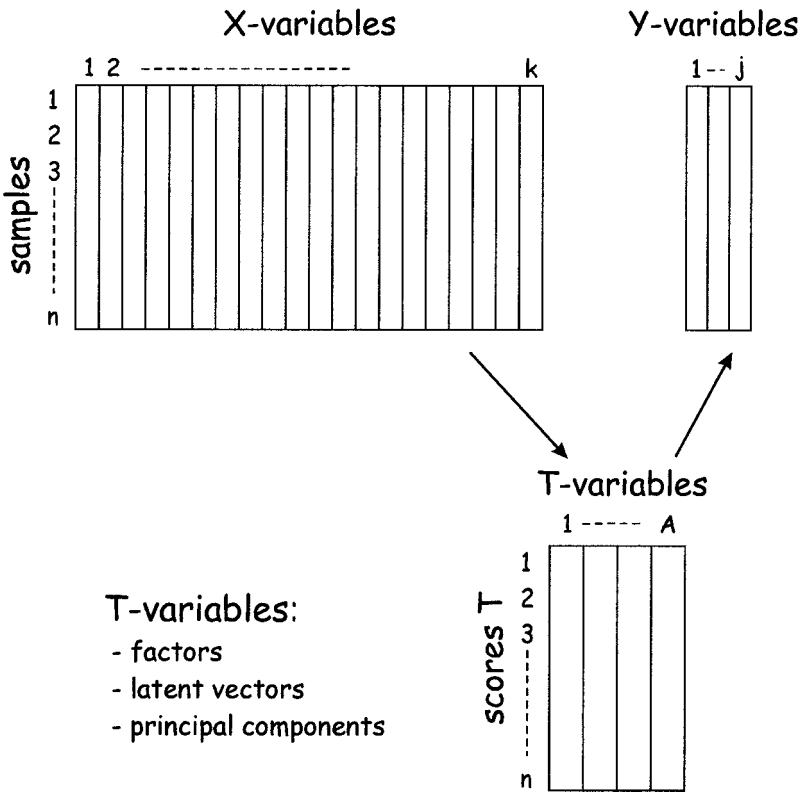


Figure 22. Principal component regression by regressing the Y-variables onto the latent vectors ( $\hat{T}$ ) representing the hidden main variations of the X-data.

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;  $\hat{V}$ , loadings of the X-variables
- $X = \hat{T}\hat{P}' + E$                       decomposition of the X-variables; in PCR the loadings P are equal to V; E, residuals
- $Y = \hat{T}\hat{Q}' + F$                       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  $X$ , then  $X$  can be written as  $X = \hat{T}P'$ . In this case  $Q$  would be equal to the regression coefficients  $\beta$  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 ( $RMSE = \sqrt{MSE}$ ). 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.

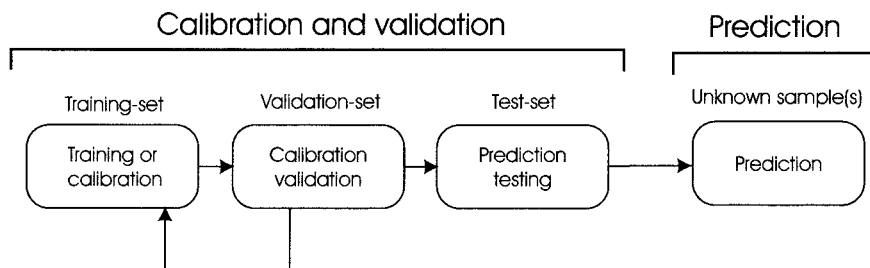


Figure 23. Calibration and validation steps necessary before prediction.

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  $\geq 25\%$  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 carefully 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.

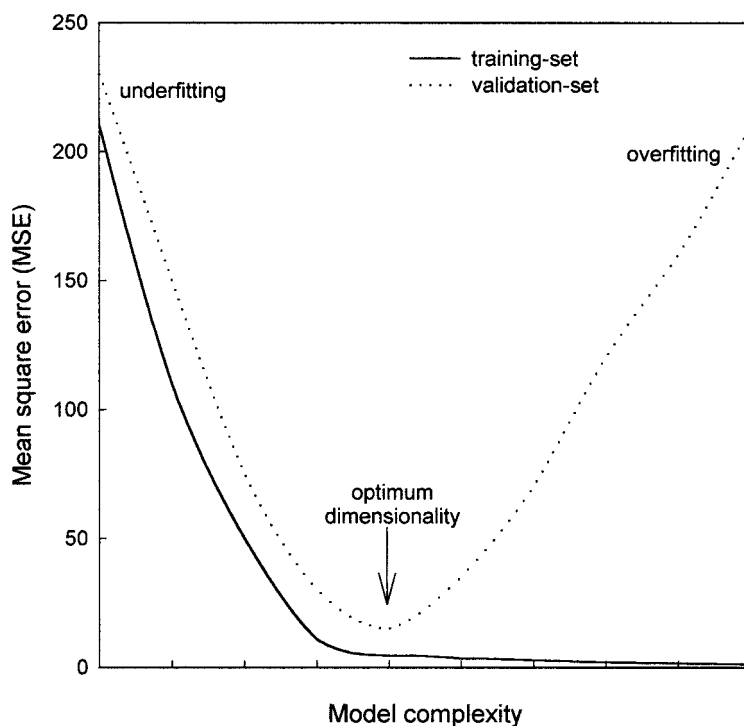


Figure 24. Prediction error as function of the complexity (dimensionality) of the training- and the validation-data.

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 from 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.



From other neurons

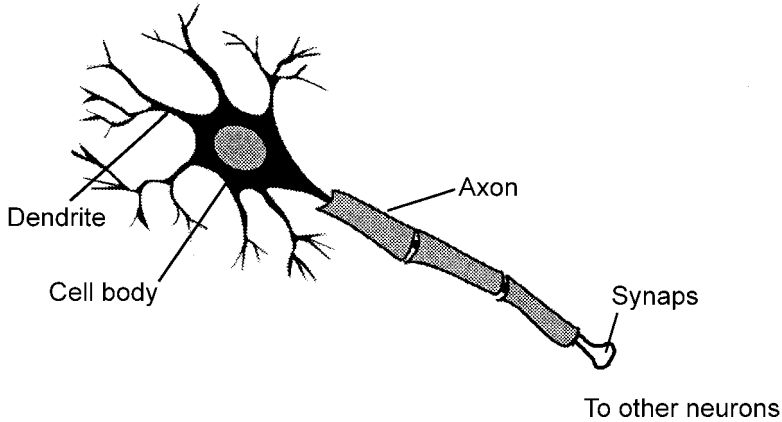


Figure 25. Illustration of a biological neuron.

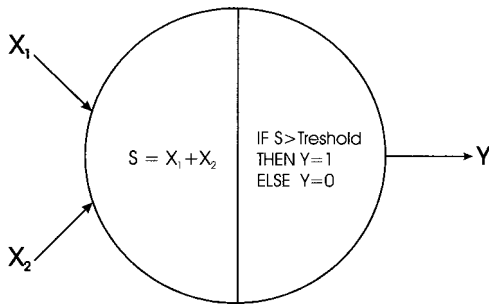


Figure 26. Artificial neuron with a simple threshold function.

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:

$$\text{net}_j = \sum_{i=1}^n w_{ij} x_i \quad w_{ij} \text{ denotes the weight connecting the neuron } i \text{ in the previous layer}$$

(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:

$$\text{net}_j^l = \sum_{i=1}^n w_{ij}^l x_i^l = \sum_{i=1}^n w_{ij}^l \text{out}_i^{l-1}$$

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:  $\text{out} = f(\text{net})$ .

These functions are categorized in three classes: linear, threshold and nonlinear (136). If a neuron does not transform its net input ( $\text{net}_j = \sum_{i=1}^n w_{ij} \text{out}_i$ ), 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.

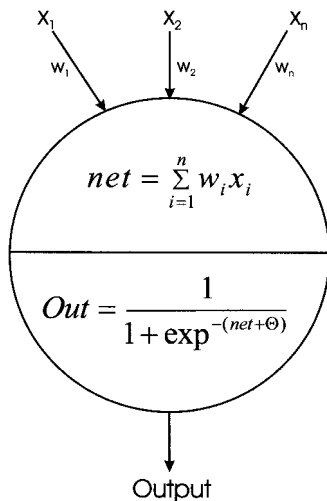


Figure 27. Artificial neuron. The sum of all weighed input signals is computed before the transfer function (sigmoid) is evaluated, using the latest value of net.

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:

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

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.

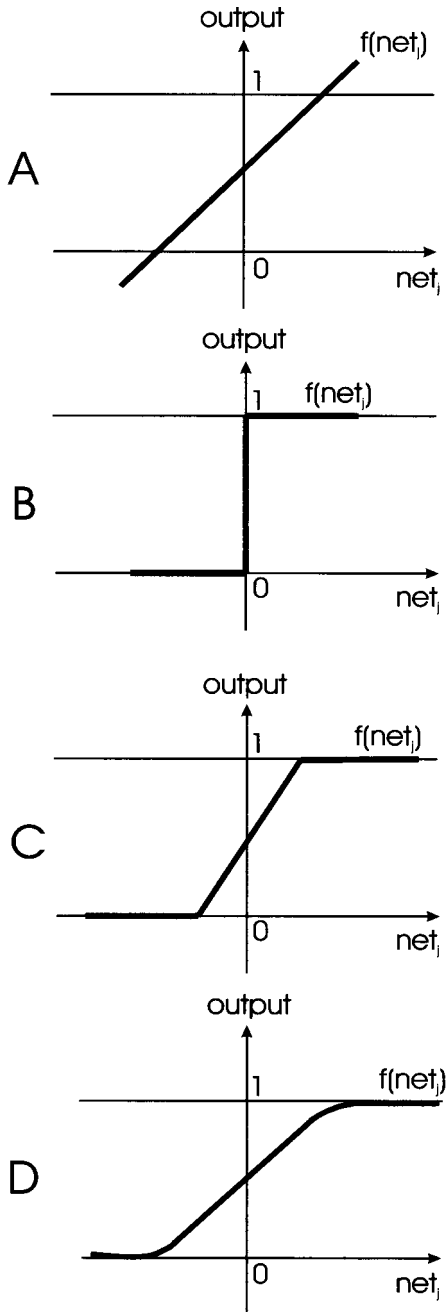


Figure 28. Activation functions. Linear (A), hard-limiter (B), threshold logic (C) and sigmoid (D).

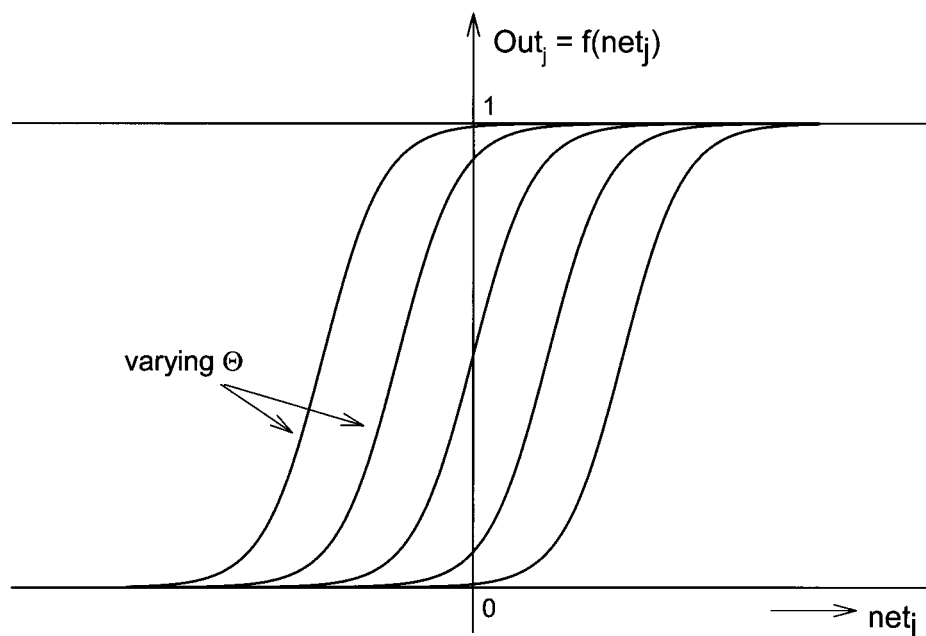


Figure 29. Effect of the parameter  $\Theta$  on the sigmoid activation function.

### *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 is 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 into 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  $w \cdot s$ . 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*.

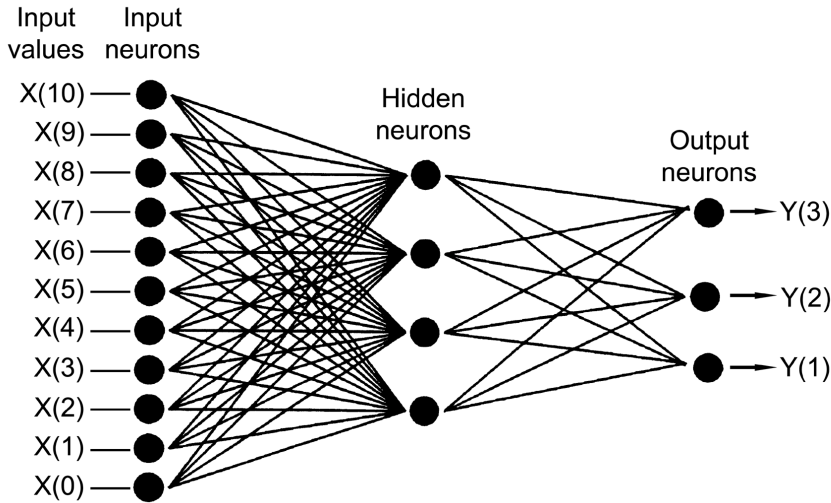


Figure 30. Three layer 'multi-layer perceptron' neural network. The input signals X(0)..X(10) represent the values of an input pattern (e.g. the absorbances of a spectrum) and Y(1)..Y(3) the values of an output pattern (e.g. the concentrations of different analytes of a sample).

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):

$$E = \frac{1}{2} \sum_j (t_j - \text{Out}_j)^2$$

in which  $t_j$  is the required (target) value of output neuron  $j$  of pattern  $p$ ,  $\text{Out}_j$  the calculated value of the neuron of pattern  $p$ , and  $\frac{1}{2}$  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:

$$\delta_j = (t_j - \text{Out}_j) f'(\text{Net}_j)$$

$\text{Out}_j (= f(\text{Net}_j))$  is the outcome of the  $j$ -th neuron of the output layer, and  $f'$  is the derivative of the transfer function of  $\text{Net}_j$ , which is  $\text{Out}_j(1 - \text{Out}_j)$  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:

$$\delta_i = f'(\text{Net}_i) \sum_{j=1}^r w_{ij} \delta_j$$

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  $\Delta w_{ij}^n = (w_{ij}^{(\text{new})} - w_{ij}^{(\text{old})})$  of the neurons, at any layer  $n$  during the learning process is defined by the delta-rule as:

$$\Delta w_{ij}^n = \eta \delta_j^n \text{Out}_i^{n-1}$$

in which  $\text{Out}_i^{n-1}$  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  $\text{Out}_i^{n-1}$  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:

$$\Delta w_{ij}^n = \eta \delta_j^n \text{Out}_i^{n-1} + \mu \Delta w_{ij}^{n(\text{previous})}$$

where  $\mu$  is called the momentum and  $w_{ij}^{n(\text{previous})}$  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.

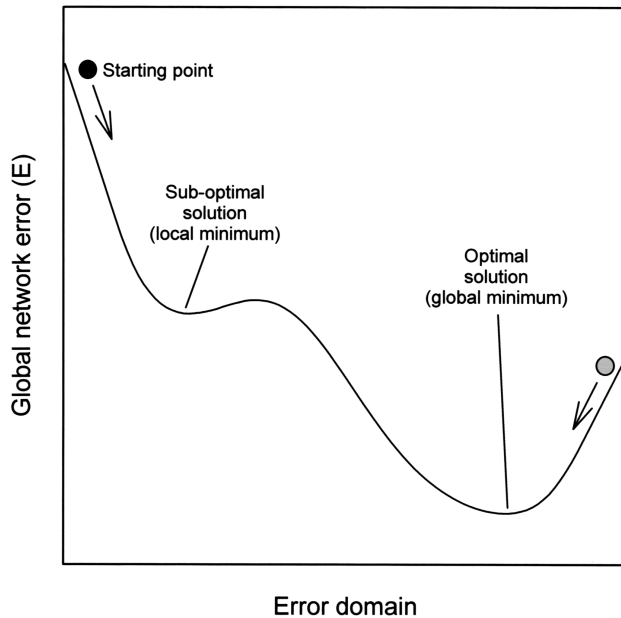


Figure 31. Illustration of the change of the global error rate of the network during training.

### *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 ( $\text{net}_j^{\text{hidden}} = \sum_{i=1}^n w_{ij} \text{out}_i$ ), which in turn results in output values of the sigmoid squashing function close to either 0 or 1. These output values ( $\text{Out}_j$ ) result in derivative values close to 0 [ $\text{Out}_j(1 - \text{Out}_j)$ ] 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  $\delta_j = (t_j - \text{Out}_j) f'(\text{Net}_j) = (t_j - \text{Out}_j)$  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:

$$\text{NSE} = \frac{1}{PJ} \sum_p \sum_j (t_{p,j} - \text{Out}_{p,j})^2$$

in which  $t_{p,j}$  is the required (target) value of neuron  $j$  of pattern  $p$ ,  $\text{Out}_{p,j}$  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. Spectral 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  $4\text{ cm}^{-1}$  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  $400\text{ cm}^{-1}$ ), 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 similarity 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.

### <sup>13</sup>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 <sup>13</sup>C medium-chain triglyceride breath test (153;154) as alternative for the analysis of fecal fat. The principle of the <sup>13</sup>C labeled triglyceride tests is based on lipolysis-dependent <sup>13</sup>CO<sub>2</sub> excretion via the breath after the ingestion of a certain amount of <sup>13</sup>C-enriched triglycerides. The <sup>13</sup>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<sub>2</sub> production and pulmonary excretion may obscure the relation between the tracer compound expired and the metabolic process studied. Up to now, none of the <sup>13</sup>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 <sup>13</sup>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).

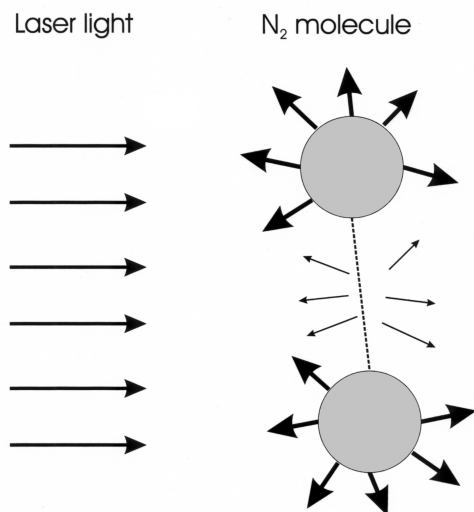


Figure 32. Raman scattering (thin arrows around the molecular bond) and Rayleigh scattering (dark arrows around the N-atoms) of a N<sub>2</sub> molecule.

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 ( $\text{cm}^{-1}$ ). The Raman spectra differ from IR spectra, in that the Raman bands are sharp and locatable to  $1 \text{ cm}^{-1}$  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) fluoresce 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* intravascular 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
<b>X</b>	
Parent 2: 00100 111	offspring 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. Philadelphia: Lea & Febiger, 1977:282-94.
3. Greenberger NJ, Isselbacher KJ . Disorders of absorption. In: Braunwald E, Isselbacher KJ, Petersdorf RG, Wilson JD, Martin JB, Fauci 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 associated with exocrine pancreatic insufficiency: implications of gastrointestinal physiology of enzyme preparations of cause-related and patient-tailored treatment. *Am J Gastroenterol* 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. Philadelphia: 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, Fauci 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 faeces. *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. *Gastroenterol* 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, Caliani 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. *Gastroenterol* 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, Fauci 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. Joul 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*. Philadelphia: 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 diëtiëk*. 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 fecal fat. *Gut* 1969;10:400-3.
48. Jakobs BS, Volmer M, Swinkels DW, Hofstetter 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. Christie WW. *High performance liquid chromatography and lipids. A practical guide*. Oxford: Pergamon 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, Merritt 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 human biological fluids. *J Lipid Res* 1983;24:491-511.
58. Die Harnsteinconcretionen, ihre Entstehung, Erkennung, und Analyse mit besonderer Rücksicht auf Diagnose und Therapie der 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. Larsson 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 - enquête 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 Geneesk* 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 Geneesk* 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. Mirabellella 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. Paluszkiwicz 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  $^{13}\text{CO}_2/^{12}\text{CO}_2$ -measurements: a clinically feasible 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-invasive 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 diabetic 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 Engng 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 decision 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 electrocardiograms. *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 screening and neural network-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 Wiley & Sons, 1991.
127. Stevens J. *Applied multivariate statistics for the social sciences*. London: Lawrence Erlbaum Associates 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 Jersey: Lawrence Erlbaum, 1983.

133. Flury B, Riedwyl H. Multivariate statistics. A practical approach. New York: Chapman and Hall, 1988.
134. Stone M. Cross-validated 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. Widrow G, Hoff ME. Adaptive switching circuits. *IRE WESCON Convention Record, Part 4* 1960;96-104.
143. Minsky ML, Papert S. *Perceptrons, an essay in computational 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*. Belmont, 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  $^{13}\text{C}$ -mixed triglyceride breath test in healthy adults: determination of the  $^{13}\text{CO}_2$  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 spectroscopy. The  $^{13}\text{C}/^{12}\text{C}$  isotope ratio of  $\text{CO}_2$ . *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 microstructure of gallstones. *Appl Spectrosc* 1987;41:407-12.
162. Kodati VR, Tomasi GE, Turmin JL, Tu AT. Raman spectroscopy identification of phosphate-type kidney stones. *Appl Spectrosc* 1991;45:581-3.
163. Buschman HPJ, Puppels GJ, Römer TJ, Laarse van der A, Bruschke AVG. In vitro intravascular histochemistry of human atherosclerosis by Raman spectroscopy. *Proceedings of the Vibrational spectroscopy: New tool in Medicine Potsdam Germany* 1998.
164. Redd DCB, Feng ZC, Yue KT, Gansler TS. Raman spectroscopy 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. *Proceedings 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 spectroscopy. *Diabetes Technol Therapeu* 1999;1:153-7.
168. Klonoff. Editorial: Noninvasive laser measurement of blood glucose in the eye: A bright idea or an optical illusion. *Diabetes Technol Therapeu* 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. Singapore: 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.



## Part I

### The analysis of fecal fat



## Chapter 1

### **A new method for faecal fat determination by mid-infrared spectroscopy, using a transmission cell. An improvement of standardisation**

B.S. Jakobs<sup>1</sup>, M. Volmer<sup>2</sup>, D.W. Swinkels<sup>1</sup>, M.T.W. Hofs<sup>1</sup>, S. Donkervoort<sup>3</sup>,  
M.M.J. Joosting<sup>3</sup>, B.G. Wolthers<sup>2</sup>, P. de Peinder<sup>4</sup>, H.A.M. Voorbij<sup>3</sup>

*Departments of Clinical Chemistry, <sup>1</sup> University Hospital Nijmegen St Radboud, PO box 9101, 6500 HB Nijmegen, <sup>2</sup> University Hospital Groningen, PO box 30.001, 9700 RB Groningen, <sup>3</sup> University Hospital Utrecht, PO box 85.500, 3508 GA Utrecht, <sup>4</sup> Department of Vibrational Spectroscopy, Faculty of Chemistry, Sorbonnelaan 16, 3584 CA Utrecht, The Netherlands.*

## ABSTRACT

Current techniques used in clinical laboratories for faecal fat determination, such as the Van de Kamer method, are not very accurate and precise. This became apparent when results obtained by different laboratories were compared, and could explain the disappointing performance of near-infrared and mid-infrared spectroscopy since the accuracy of these techniques depends upon the accuracy of the calibration used (i.e. inaccurate wet chemical analysis). In order to improve standardization we developed and tested a new quantitative method in three laboratories based on Fourier Transform infrared (FT-IR) spectroscopy. Fatty acids were extracted from faecal samples with acidified petroleum ether–ethanol and the extracts were dried and dissolved in chloroform. An infrared spectrum of the extracts was recorded in the range 4000–650  $\text{cm}^{-1}$ , using an IR transmission cell. Standard mixtures of stearic and palmitic acid (65:35) were used for calibration. Quantification was based on the absorbance band of the  $\text{CH}_2$  group (2855  $\text{cm}^{-1}$ ) of free fatty acids and fatty acid glycerol esters. The calibration curve showed excellent linearity. The correlation coefficient between the titrimetric Van de Kamer and FT-IR methods was 0.96 ( $y=1.12x-0.02$ , standard error of prediction = 0.89 g% fat). No significant difference was found when the FT-IR results of 28 faecal samples from patients were compared between two different university hospital laboratories. The new FT-IR method, using primary standards, is simple and rapid, and provides satisfactory intra- and interlaboratory precision for diagnosis and monitoring of steatorrhoea.

---

## INTRODUCTION

Steatorrhoea is defined as an increase of fat in stool. Intestinal malabsorption, which is the most frequently occurring cause of steatorrhoea, is a problem with absorptive functions of the bowel (1,2). Furthermore gastric, pancreatic and biliary diseases can cause maldigestion, which in turn can also lead to steatorrhoea. The gold standard laboratory method to diagnose steatorrhoea is faecal fat analysis: stools are collected for 72 h and the faecal fat is measured by traditional techniques such as titrimetric (3), gravimetric (4) or acid steatocrit (5) methods. These techniques are less suitable for serial routine analyses as they are inaccurate, imprecise, time-consuming and require unpleasant and prolonged handling of stools (6). In recent years new techniques have been introduced to determine faecal fat content by means of infrared (IR) spectroscopy. Several authors have reported on the use of near-infrared (12.500–4000  $\text{cm}^{-1}$ ) reflectance (NIRR) analysis as a method suitable for the investigation of steathorrhoea (7-9). Calibration is performed by NIRR using a considerable number of reference samples in which the faecal fat content has been previously derived from wet chemical analysis (secondary reference samples), using, e.g. the Van de Kamer method (3). Multivariate analysis methods such as partial least squares (PLS) or multiple linear regression (MLR), are then used to relate the NIRR absorbance data to the known fat content of the reference samples, thus enabling the fat content of unknown samples to be calculated from their NIRR data. Recently, Franck et al. (10) introduced a similar method for faecal fat analysis, using mid-infrared reflectance (MIRR) spectroscopy (4000–400  $\text{cm}^{-1}$ ). Their method is based on the use of a horizontal attenuated total reflectance accessory on which the stool sample is spread. Samples that are too liquid, too solid, or that contain visible food or other fragments have to be excluded.

The disappointing performance of these NIRR or MIRR results can be attributed to

inaccurate wet chemical analysis of the calibration set. Nevertheless, IR spectroscopy is potentially useful for faecal fat determination and can be improved by the introduction of standardization and calibration using primary standards and lipid extracts of faecal samples, instead of unprocessed stool samples. Such lipid extracts, using chloroform as solvent, were put in a liquid transmission cell, from which the IR spectrum is recorded. We used primary standards consisting of a mixture of the most prominent fatty acids in human stools (stearic and palmitic acids, 65:35) and compared the FT-IR results obtained by the Van de Kamer method. We studied the consequence of primary standardization for faecal fat analysis and the resultant interchangeability of laboratory results generated by the new method.

## MATERIALS AND METHOD

### *Stool collection*

Seventy-two-hour stool samples, collected from 97 consecutive patients who were suspected of having malabsorption or maldigestion, were homogenised in a blender. Samples that could not be analysed immediately after collection were stored at  $-20^{\circ}\text{C}$ .

### *Sample preparation and determination of faecal fat with FT-IR*

Stool samples were treated according to the extraction procedure derived from the gravimetric method, as described by Wybenga et al. (4), with minor modifications. A faecal sample (0.5 g) was suspended in 1 mL water, 100  $\mu\text{L}$  HCl (37%) and 3 mL of ethanol (96%). Then 5.0 mL petroleum ether was added and the mixture was vigorously shaken for 10 min followed by centrifugation for 5 min at 3000 g. A 4-mL portion of the organic layer was transferred to a new tube, and evaporated for at least 30 min at  $45^{\circ}\text{C}$  under a stream of nitrogen. The dried lipid extracts were dissolved in 1.0 mL chloroform (gradient grade) and transferred to a transmission flow cell with sodium chloride crystals (path length 0.1–0.025 mm). The spectra were measured in the mid-infrared region from  $4000\text{--}650\text{ cm}^{-1}$ . Sixteen scans were co-added at an optical resolution of  $4\text{ cm}^{-1}$  (strong apodization), using a Perkin-Elmer Spectrum 2000 spectrometer (Perkin Elmer, Norfolk, CT, USA), or a Bio-Rad FTS-7 spectrometer (Bio-Rad, Cambridge, MA, USA).

Both spectrometers were equipped with a deuterium triglycine sulphate detector. Chloroform was used for background subtraction.

### *Faecal fat determination using the conventional method*

The faecal fat analyses were done in three university hospital laboratories, according to the conventional Van de Kamer (3) method. This method is the current 'state of the art' in Dutch hospitals and is based upon a single acid extraction of saponified fatty acids from the stool followed by titration of the fatty acid COOH group with sodium hydroxide to quantitate the amount of fat present. Quantitation is based on the mean molecular weight of fatty acids in faeces (276 based on  $\text{C}_{18}$ ).

### *Quantitative analysis and calibration curves*

Before storing the spectra on disk a baseline correction was performed with an integration area in the range  $3400\text{--}2400\text{ cm}^{-1}$  for the 2855 absorbance band and  $1900\text{--}1600\text{ cm}^{-1}$  for the 1709 absorbance band. MLR (11) (Systat 7.0, SPSS Inc. Chicago, IL, USA) using stepwise variable selection was done to select the best-performing absorbance band(s) (1458, 1467, 1709, 2855, 2928 and  $2957\text{ cm}^{-1}$ ) to predict the quantity of fatty acids in the stool. For MLR, 30 of the 97 patient samples were selected in a range 0.5–17.9 g% fat as estimated

using the Van de Kamer method. The composition of the primary standards, for generation of the calibration curve, was based on a previous study in which 69 faecal samples were analysed by gaschromatography (12). In this study, petroleum ether extracts of the faecal samples (prepared as described in the sample preparation section) were transmethylated and injected in a gas chromatograph to obtain the fatty acid composition. Most of the faecal fat samples were composed of mainly C<sub>16</sub> and C<sub>18</sub> fatty acids (median 83%) and small amounts of C<sub>14</sub> and C<sub>20</sub> fatty acids (median 17%). After omitting the small amount of C<sub>14</sub> and C<sub>20</sub> fatty acids, the mean C<sub>18</sub>:C<sub>16</sub> composition of the 69 samples showed to be 65:35. Therefore each hospital used primary standard mixtures of stearic/palmitic acid (Sigma-Aldrich Chemie BV, Zwijndrecht, The Netherlands) with this composition for the calibration curve, in the range 0–15 g%, using simple linear regression analysis. The primary standards were handled in the same way as the faecal samples (prepared as described in the sample preparation section). Results of faecal fat analysis are normally expressed as g% total fat (triglycerides). As our new method is based on calibration with free fatty acids it gives a slight underestimation if the results are expressed as g% fat. This underestimation is 4.5% (13/287), because the mean molecular weight of total fat is 287 (13 for triglyceride for the part glycerol residue (1/3) and 274 for the mean molecular weight of the stearic-palmitic acid mixture).

To test the linearity of the regression equation of the calibration the runs test (13) was used, each calibration point being measured four times. The runs test compares the number of series of consecutive points below and above the regression line to a number of series found at random.

#### *Imprecision*

The imprecision of the FT-IR method was tested for two stool samples with a low (2.6 g%) and high (7.6 g%) fat contents, respectively. Intra-assay reproducibility was tested by a 10-fold measurement of each stool. Analysing one sample from each stool during 11 consecutive days enabled inter-assay reproducibility to be determined.

#### *Accuracy*

Passing & Bablok regression analysis (14) was used to compare the results of the Van de Kamer and FT-IR methods, using the 97 patient samples from the patients.

#### *Inter-laboratory differences*

To obtain information on inter-laboratory differences in faecal fat measurements using both the Van de Kamer method and the new FT-IR sample preparation method in combination with calibration standards, 28 faecal samples were selected from patient mentioned above. Selection of these samples was based on their fat content, over the range 0–15 g% fat, as determined by the Van de Kamer method.

The 28 samples were analysed in the two university hospital laboratories possessing a FT-IR spectrometer with flow cell, and the differences between the laboratories were statistically tested with a paired student-t test.

The same 28 samples were analysed in all three hospitals by the Van de Kamer method, and the results were analysed statistically using the Friedman test for pairwise differences (13). In all cases, P-values  $\leq 0.05$  were considered to be statistically significant.

## RESULTS

Figure 1 shows the spectral region of interest in the IR spectrum of an extracted stool sample. Bands reflecting the CH-stretch vibration of the fatty acid methylene residues, are observed at  $2928\text{ cm}^{-1}$  (antisymmetric), and  $2855\text{ cm}^{-1}$  (symmetric), whereas the C=O stretch vibration of the carboxylic acid functionality is observed at  $1709\text{ cm}^{-1}$ . Chloroform showed no interference in the spectral region of interest ( $2950\text{--}1650\text{ cm}^{-1}$ ), in contrast to petroleum ether. The petroleum ether was therefore evaporated from the lipid extract and the dried extract was re-dissolved in chloroform.

Of all the measured IR-absorbance bands ( $1458$ ,  $1467$ ,  $1709$ ,  $2855$ ,  $2928$ , and  $2957\text{ cm}^{-1}$ ) of an extracted stool, the one at  $2855\text{ cm}^{-1}$  appeared to be the best at predicting the quantity of fatty acids, directly followed by the band at  $1709\text{ cm}^{-1}$ . The single band at  $2855\text{ cm}^{-1}$  proved to be sufficient for calibration (highest F-value).

The results using the Van de Kamer method for the 28 stools differed significantly between the three university hospital laboratories (Friedman test,  $P \leq 0.0001$ ), demonstrating the interlaboratory imprecision of the Van de Kamer method in practice. Because the accuracy of the FT-IR method depends upon the accuracy of the calibration used, we decided to use primary standards composed of stearic and palmitic acids (65:35) instead of the Van de Kamer method and created a calibration set over the concentration range  $0\text{--}15\text{ g}\%$ .

The calibration curve of the primary fatty acid standards was linear from  $0$  to  $12\text{ g}\%$  and departed slightly ( $P=0.012$ ) from linearity when the standard mixture  $15\text{ g}\%$  was included (runs test, Fig. 2). Thus, the primary fatty acid standards used to generate the calibration curve for FT-IR can serve as an accurate alternative to the conventional Van de Kamer method.

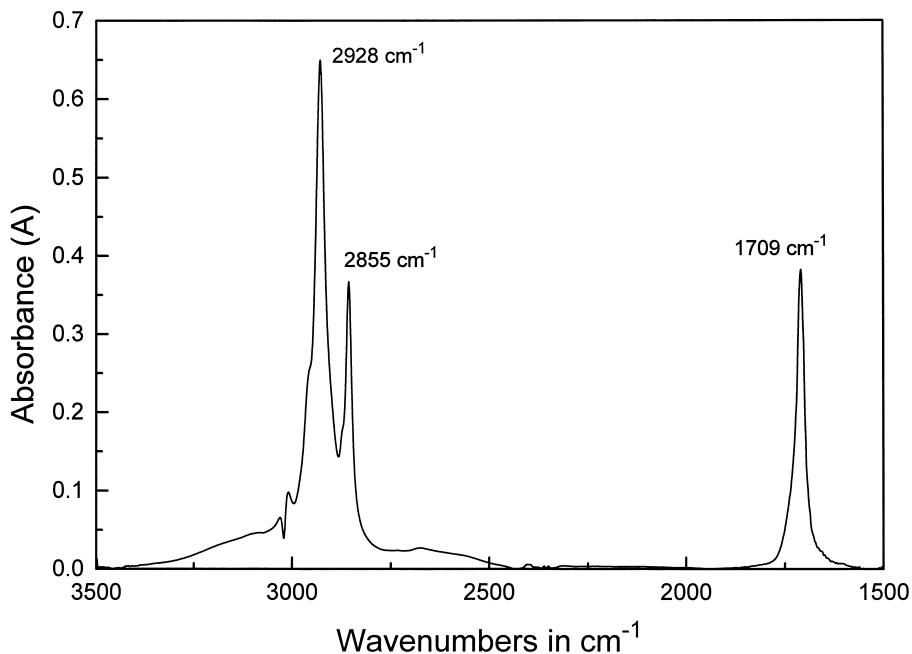


Figure 1. Fourier transform infrared spectrum of a faecal fat extract in chloroform obtained from a patient's stool containing  $12.8\text{ g}\%$  fat.

Comparison of the results obtained in one university hospital for faecal fat in 97 stool samples analysed with the Van de Kamer and FT-IR methods, using the stearic–palmitic acid calibration curve, showed a slope of 1.16 (95 % CI 1.100–1.222), an intercept of  $-0.023$  (95% CI 0.122–0.073), a correlation coefficient of 0.96 and a standard error of prediction (SEP) of 0.89 g% fat (Fig. 3: Passing & Bablok regression analysis).

The intra- and inter-assay coefficients of variation (CVs) of faecal fat estimation by means of FT-IR were 4.7% and 10.0%, respectively, for the stool sample with a faecal fat content of 2.6 g%. Intra- and inter-assay CVs were 2.6% and 8.5%, respectively, for a faecal fat content of 7.6 g%. Comparison of the FT-IR measurements in the two university hospital laboratories of the fat (g%) in 28 stool samples, using the stearic–palmitic acid calibration curve for quantification, in the two university hospital laboratories showed no statistically significant difference (paired Student's t-test,  $P=0.915$ ).

## DISCUSSION

Modern mid-infrared FT-IR spectrometers have a high potential for applications in clinical chemistry, in that they provide high signal-to-noise ratios, high resolution and extensive data processing possibilities by means of computer software packages (10;15).

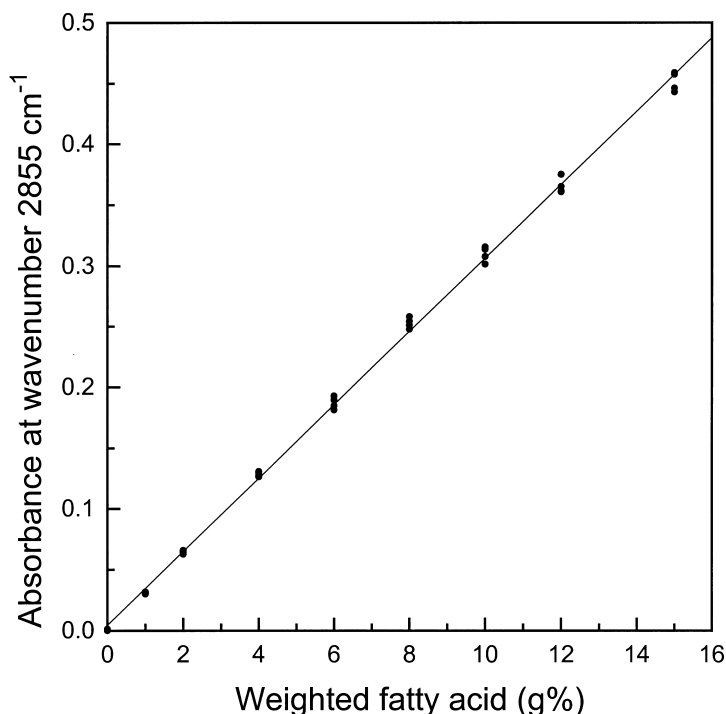


Figure 2. Calibration curve for the Fourier transform infrared absorbance values at 2855  $\text{cm}^{-1}$  using weighted stearic–palmitic acid standards (65:35). Absorbance values were baseline corrected. The curve can be described by  $y=0.0302x + 0.005$  ( $r=0.99$ ; Standard error of calibration=0.006 g% fat); slope= 0.0302 (standard deviation: SD=0.0002), intercept=0.0045 (SD=0.0016).



Previously, Franck et al. (10) reported on the use of a horizontal attenuated total reflection crystal for mid-infrared spectroscopy on stools. We tested their method, but could not reproduce their results. Close inspection of their results revealed that they used a selection of stools, excluding those that were too liquid, too solid, or which contained visible food particles or other fragments. Moreover, water appeared to cause major interference in their spectra as its absorption in the mid-infrared is very strong and tends to override the spectral characteristics of the other compounds of interest. Because water is the major component in stools (amounting to 60–90%), this makes the method of Franck et al. less appreciable. Our approach of using a simple extraction step of fatty acids from stool prevents the interfering effect of water. Another advantage of this separation of fatty acids from stools is the exclusion of water-soluble interfering substances, and the opportunity to use a primary standard (stearic–palmitic acids) for calibration.

We used the absorptions of the CH symmetric ( $2855\text{ cm}^{-1}$ ) or the C=O stretch vibrations ( $1709\text{ cm}^{-1}$ ) of the primary fatty acid standard for quantification. Other methods based on NIR or FT-IR, without sample purification, use complex mathematical models, like MLR and PLS for quantification and require the analysis of at least 30 samples by the reference method (Van de Kamer) to obtain a reliable mathematical model (7-9).

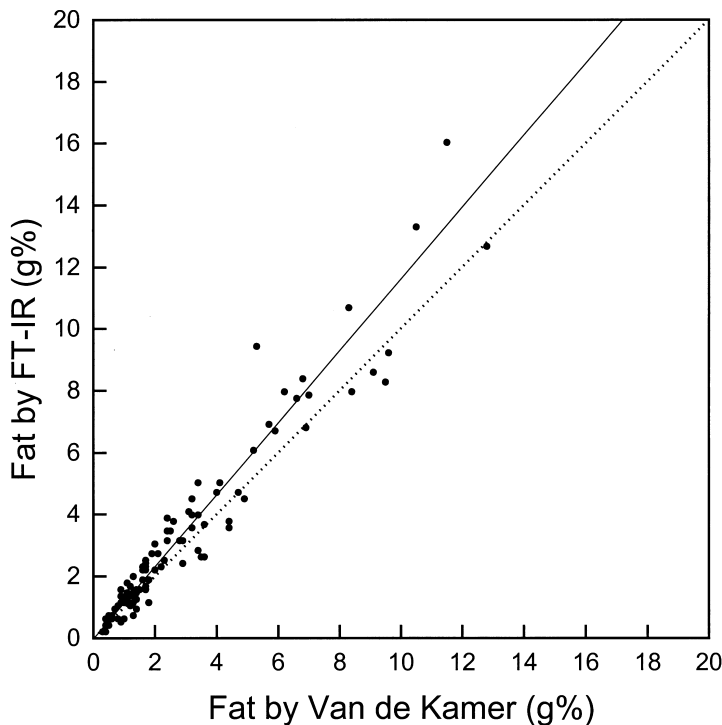


Figure 3. Relationship of fat estimated by Fourier transform infrared (FT-IR) and Van de Kamer methods. The solid line is the Passing & Bablok regression line ( $y=1.164x - 0.023$ ;  $r=0.96$ ; standard Error of Prediction=0.89 g% fat;  $n=97$ ). The dotted line denotes the line of identity. The difference of 4.5% for the amount of fat (g%) determined by FT-IR, which results from the use of a mixture of free fatty acids instead of triglycerides as a calibrator, is taken into account (see text).

Van de Kamer (16) has very comprehensively tested the analytical recovery of the extraction procedure. After a single extraction with acidified petroleum ether–ethanol, a recovery of 98% from stool was obtained for fatty acids with a chain length greater than or equal to C<sub>14</sub> (3). This method, however, gave significantly different results between samples analysed in the three university hospital laboratories, whereas our FT-IR method, calibrated with a primary standard, gave interchangeable results for the fat content in stools measured in two different university hospital laboratories.

Furthermore, within one laboratory the fat content measured by the Van de Kamer method differed significantly from that found by our standardized FT-IR method (Fig. 3). Both these differences (the Van de Kamer method between laboratories and between Van de Kamer and our FT-IR method) can be attributed to the lack of standardization of the conventional titrimetric method, which could result in deviation from the true value and decreased precision. Improved of the inter-laboratory variation of the Van de Kamer method may be achieved by distributing a calibrator. This might of course also further improve the inter-laboratory variation of the new FT-IR method. However, optimization of the Van de Kamer method does not seem to be useful, as the new FT-IR method has a total analysis time at least 50% shorter than the Van de Kamer method and the results of the FT-IR method agree well with the Van de Kamer method (Fig 3).

The use of two fatty acids standards enables the calibration of both the CH<sub>2</sub> and the C=O bands. Determination of the relative quantity of the CH symmetric stretch vibration of the CH<sub>2</sub> residues, the C=O stretch vibration of the carboxylic acid functionality of free fatty acids (1709 cm<sup>-1</sup>) and the C=O vibration of glycerol esters (1746 cm<sup>-1</sup>) allows the identification of abnormal fatty acids:triglycerides ratios in faeces by calculation of the absorbance ratio of CH<sub>2</sub> (2855 cm<sup>-1</sup>):C=O (1746 cm<sup>-1</sup> or 1709 cm<sup>-1</sup>). These spectral fingerprints can be of clinical use (17). Normally about 95% of faecal fat is excreted as soaps of free fatty acids (16). However, theoretically, in faecal samples of patients with maldigestion, due to cystic fibrosis, chronic pancreatitis (diminished amount of lipase), or a lack of bile acid secretion in the gut, increased amounts of glycerol esters can be found in stools.

### *Conclusion*

The new FT-IR method is simple and enables rapid diagnosis and monitoring of steatorrhoea. A further decrease in total analysis time could be achieved by using automatic sampling devices for FT-IR spectrometers. We have also shown that this method is more precise with respect to inter-laboratory variation than the Van de Kamer method. The use of primary standards for calibration facilitates the introduction of this method in other laboratories using mid-infrared spectroscopy. For routine application we recommend inclusion of a fatty acid standard mixture in the extraction procedure to check extraction efficiency and reproducibility.

### **Acknowledgements**

We thank A.W. Kingma and P. Koops for their skillful technical assistance. We thank N. Boer and S. van der Pouw from Perkin-Elmer for their technical support. We are grateful to Prof. Dr. J.H. van der Maas and Dr. E. Lutz for their stimulating discussions.

**REFERENCES**

1. Rosenberg IH, Sitrin MD. Screening for fat malabsorption. *Ann Intern Med* 1981; 95: 776-7.
2. Hamilton I. Small intestinal mucosal permeability. In: Pounder RE, ed. *Recent advances in gastroenterology* No 6. Edinburgh: Churchill Livingstone, 1986; 73-91.
3. Van de Kamer JH, Ten Bokkel Huinink H, Weyers HA. Rapid determination for the determination of fat in feces. *J Biol Chem* 1949; 177: 349-55.
4. Wybenga DR, Inkpen JA. Lipids. In: Henry RJ, Cannon DC, Winkelman JW, eds. *Clinical Chemistry: principles and techniques*. New York: Harper and Row 1974; 1421-93.
5. Phuapradit P, Narang A, Mendonca P, Harris DA, Baum JD. The steatocrit: a simple method for estimating stool fat content in newborn infants. *Arch Dis Child* 1981; 56: 725-7.
6. Duncan A, Hill PG. A UK survey of laboratory-based gastrointestinal investigations. *Ann Clin Biochem* 1998; 35: 492-503.
7. Peuchant E, Salles C, Jensen R. Value of a spectroscopic "fecalogram" in determining the etiology of steatorrhea. *Clin Chem* 1988; 34/1: 5-8.
8. Benini L, Caliani S, Vaona B, Talamini G, Vantini I, Guidi GC, et al. Near infrared spectrometry for fecal fat measurement: comparison with conventional gravimetric and titrimetric methods. *Gut* 1989; 30: 1344-7.
9. Bekers O, Postma C, Lombarts AJPF. Determination of faecal fat by Near-Infrared spectroscopy. *Eur J Clin Chem Clin Biochem* 1995; 33: 83-6.
10. Franck P, Sallerin JL, Schroeder H, Gelot MA, Nabet P. Rapid determination of fecal fat by Fourier transform infrared analysis (FT-IR) with partial least-squares regression and an attenuated total reflectance accessory. *Clin Chem* 1996; 12: 2015-20.
11. Stevens J. Multiple regression. In: *Applied multivariate statistics for the social sciences*. Hillsdale (NJ): Lawrence Erlbaum Associates, Publishers, 1986; 51-112.
12. Muskiet FAJ, Van Doormaal JJ, Martini IA, Wolthers BG, Van der Slik W. Capillary gaschromatographic profiling of total long-chain fatty acids and cholesterol in biological materials. *J Chromatogr Biomed Appl* 1983; 278: 231-44.
13. Siegel S, Castellan Jr. NJ. *Non-parametric statistics for the behavioral sciences*. (2nd edition). New York: McGraw-Hill, 1988.
14. Passing H, Bablok W. A new biometrical procedure for testing the equality of measurements from two different analytical methods. *J Clin Chem Clin Biochem* 1983; 21: 709-20.
15. Volmer M, Wolthers BG, Metting HJ, Haan THY, Coenegracht PJM, Van der Slik W. Artificial neural network predictions of urinary calculus compositions analyzed with infrared spectroscopy. *Clin Chem* 1994; 40: 1692-7.
16. Van de Kamer JH. *Vet in faeces*. Utrecht, The Netherlands: Dissertation in Dutch, 1948.
17. Nakamura T, Takebe K, Tando Y, Arai Y, Yamada N, Ishii M, et al. Faecal triglycerides and fatty acids in the differential diagnosis of pancreatic insufficiency and intestinal malabsorption in patients with low fat intakes. *J Int Med Res* 1995; 23: 48-55.



## Chapter 2

# **Enhanced time-saving extraction procedure for the analysis of fecal fat by Fourier Transform Infrared spectroscopy**

Bernadette S. Jakobs<sup>1</sup>, Marcel Volmer<sup>2</sup>, Marc T.W. Hofs<sup>1</sup>, Dorine W. Swinkels<sup>1</sup>

<sup>1</sup>*Department of Clinical Chemistry, University Hospital Nijmegen, St Radboud, PO Box 9101, 6500 HB Nijmegen,*

<sup>2</sup>*Department of Clinical Chemistry, University Hospital Groningen, PO Box 30001, 9700 RB Groningen, The Netherlands.*

*Clin Chem 2000;46(7):1019-1020*

## INTRODUCTION

Recently, we presented an improved procedure for the determination of fecal fat by means of Fourier transform infrared spectroscopy (1). This method can be used in laboratories equipped with a mid-infrared spectrometer. With this method, fecal fat was extracted from stool samples with petroleum ether–ethanol. After extraction, the petroleum ether was dried, and the fatty acids were redissolved in chloroform before measurement. Because the extraction procedure in the previously described analytical method (1) was still rather time-consuming, we replaced the petroleum ether–ethanol extraction with a single chloroform extraction.

## MATERIALS AND METHODS

In the new extraction procedure, 1 gram of homogenized stool sample was suspended in 2.5 mL water. After the suspension was mixed, 0.5 mL of 12 mol/L HCl and 2.0 mL of chloroform (gradient grade) were added, and the sample was shaken vigorously for 10 min. At this stage, the samples were either stored at  $-20^{\circ}\text{C}$  or analyzed immediately. The extract was centrifuged for 5 minutes at 3000g at room temperature, after which the organic layer was transferred to a transmission cell (path length, 0.05 mm) with calcium fluoride crystals.

Spectra ( $n=111$ ) were scanned in the mid-infrared region from 4000 to  $650\text{ cm}^{-1}$  with a Perkin-Elmer Spectrum 2000 spectrometer (Perkin-Elmer). Calibration was performed with a mixture of stearic and palmitic acid (65:35, by weight) ranging from 0 to 150 g/kg (1). For both the 'old' and the 'new' extraction procedures, the spectroscopic band at  $2855\text{ cm}^{-1}$  (C–H symmetric stretch vibration) was used to calculate the amount of fat (g/kg). Passing & Bablok regression was performed for method agreement.

## RESULTS

The results obtained by the two methods showed good agreement ( $r = 0.991$ ; Fig. 1). By Passing & Bablok regression, the slope was 1.055 (range, 1.026–1.088), the intercept was 0.241 (range, 0.181–0.296), and the standard deviation of residuals ( $S_{y|x}$ ) was 0.365. For the new extraction procedure, the intra- and interassay imprecision (as the CVs;  $n=10$ ) was 4.0% and 5.0% respectively, for a stool sample containing 43 g/kg fat. For a sample with 26 g/kg fat, the intra- and interassay CVs were 3.9% and 10.0%, respectively. Recovery of the stearic–palmitic acid (65:35, by weight) calibrator added to stool was  $>95\%$  with the new extraction procedure. The majority of fecal fat is mainly composed of  $\text{C}_{16:0}$  and  $\text{C}_{18:0}$  free fatty acids which can be extracted easily from the stool with the new extraction method. The reduction in analysis time gained in this way is approximately  $\sim 2.5$  h for a series of 10 stools.

## DISCUSSION

We conclude, that the new simplified extraction procedure for fecal fat determination gives comparable results to the old extraction procedure and allows considerable reduction in analysis time, use of chemicals, and technical equipment.

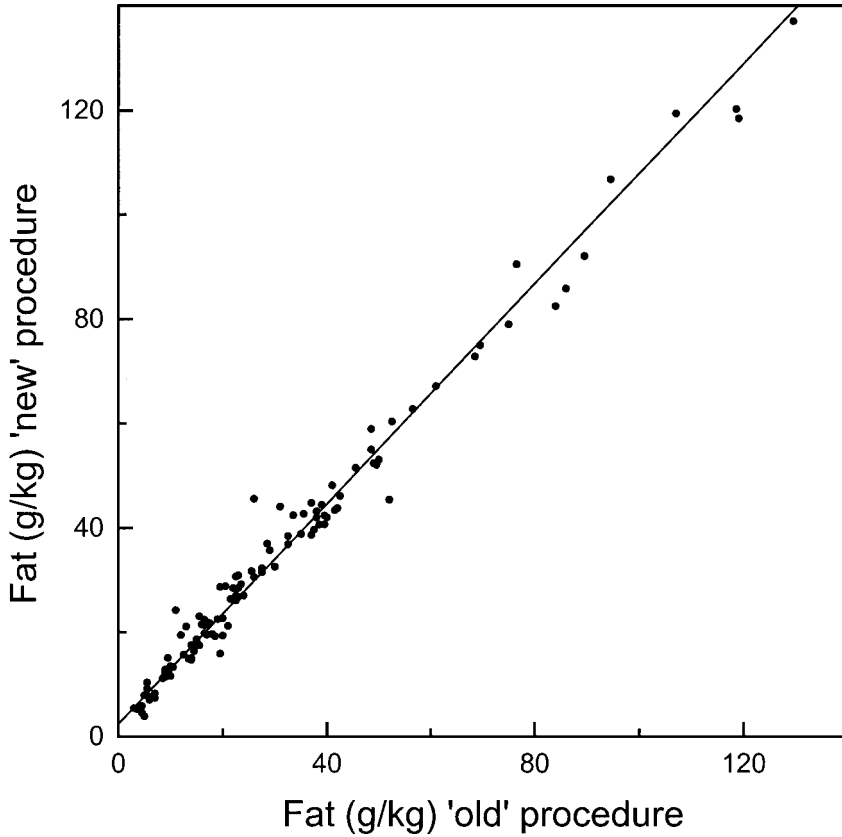


Figure 1. Relationship of fecal fat (n=111) determined with the old and new extraction procedure. Absorbance values at  $2855\text{ cm}^{-1}$ , obtained with Fourier transform infrared spectroscopy, were used for calibration.

## REFERENCES

1. 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.





## Chapter 3

# **Investigation of the applicability of a mid-infrared spectroscopic method using an Attenuated Total Reflection accessory and a new near-infrared transmission method for the determination of faecal fat**

Marcel Volmer<sup>1</sup>, Anneke W. Kingma<sup>1</sup>, Peter C.F. Borsboom<sup>2</sup>,  
Bert G. Wolthers<sup>1</sup>, and Ido P. Kema<sup>1</sup>

<sup>1</sup>*Department Pathology and Laboratory Medicine, University Hospital Groningen, PO Box 30001, 9700 RB Groningen, The Netherlands, and* <sup>2</sup>*PB Sensortechnology & Consultancy bv, Westeremden, The Netherlands*

*Ann Clin Biochem 2001;38:256-263*

## ABSTRACT

In many laboratories, the titrimetric method of Van de Kamer is used for the analysis of faecal fat content of patients suspected of steatorrhoea. We investigated the applicability of a mid-infrared (MIR) spectroscopic method using an Attenuated Total Reflection (ATR) accessory and a new near-infrared (NIR) spectroscopic method. For the NIR method sealed plastic bags, containing the stool samples, were used as transmission cells. Standardization was obtained using a previously described MIR method, with a NaCl flow-cell, as reference method. Partial least-squares regression was used for the calibration of each method. Full cross-validation of the calibration set was used for internal validation of each method. Fifteen per cent of the stool samples could not be estimated with the ATR method within reasonable accuracy limits compared with the reference. The standard error of prediction of the NIR method was 1.1 g/dL. We conclude that the new NIR method is a promising technique for routine use. However, some further experiments need to be done with triplicate measurements of each sample and the use of an external validation set.

---

## INTRODUCTION

The titrimetric method described by Van the Kamer et al. (1) is still the most important direct assay of faecal lipid content for diagnosis and monitoring of steatorrhoea. However, various other traditional laboratory assays, such as gravimetric (2) and acid steatocrit (3) methods, are also used for the determination of faecal fat. Unfortunately, most of these methods are timeconsuming and cumbersome for laboratory technicians.

Other assays, based on infrared (IR) measurements have been developed. Several authors reported on the use of a near-infrared (NIR) reflectance method (4-7), some of which used a gold-plated integrating sphere as sampling device. One of these publications (7) described NIR reflectance measurement through a polyethylene– polyamide film. Recently, Franck et al. described a method based on mid-infrared spectroscopy (MIR) using an attenuated total reflectance (ATR) sampling device (8). They noted that some preselection of samples was necessary. The advantage of their method is that it makes use of standard Fourier transform infrared (FT-IR) equipment that can also be employed in other clinical chemical applications, such as the analysis of the composition of urinary stones (9). Both IR methods are based on reflectance measurements. The major advantage of these infrared spectroscopic methods for faecal fat determination is that the measurements can be performed on untreated stool samples. This prevents prolonged handling of stools using extraction procedures.

In case of conventional assays, simple calibration and calculation methods can be used, often based on primary standards mixtures. However, measurements of untreated stool samples containing a large number of interfering substances result in complex IR spectra. Calibration and prediction of the component concentrations of interest, in the presence of many interfering substances, require the use of more sophisticated multivariate calibration methods such as multiple linear regression (MLR) and partial least-squares (PLS) regression (10). Consequently, measurements of untreated stool samples require secondary reference samples for calibration. These are biological samples from which the sample compositions have been determined by means of a generally accepted method, such as that of Van de Kamer. However, Bekers et al. (5) highlighted the need to create a new calibration curve in every clinical chemical laboratory, when using secondary reference samples. A paper describing a multicentre trial showed that the results of the Van de Kamer method were poorly exchangeable (11). The authors of this paper developed a method,

based on a simple extraction of faecal fat from stool samples and determination by MIR spectroscopy using a NaCl transmission cell, which resulted in improved standardization. The aim of the current study was to investigate the applicability of IR spectroscopic methods for the determination of faecal fat, without the necessity of any sample pre-treatment. We developed a new NIR spectroscopic method using plastic bags as transmission cells. We also reinvestigated the applicability of the MIR spectroscopic method using the ATR sampling device (8). Standardization was obtained with the previously described MIR technique as a reference method (11).

## MATERIALS AND METHODS

### *Sample collection*

One hundred and nineteen stool samples were collected from patients suspected of malabsorption and maldigestion. The stools were collected over three successive days, stored at  $-20^{\circ}\text{C}$  until analysis and carefully homogenized before measurement. Homogeneity was obtained by manually blending the complete stool sample with a stirring rod for at least 4 min. From each stool, the relative water content was determined by weighing the sample before and after freeze-drying. A visual estimation of the homogeneity was also recorded.

### *Determination of faecal fat using the flow-cell reference method*

The fat content of the 119 stool samples was estimated with the MIR cell method (11). Briefly, after addition of HCl and ethanol, all samples were extracted once with petroleum ether. After evaporation of the lipid extracts, the samples were dissolved in chloroform and measured in a NaCl transmission flow-cell (path length 0.1 mm; Specac Ltd, Kent, UK). Primary standard mixtures of stearic-palmitic acid (65:35) were used for calibration. Chloroform was used for background measurements. The spectra were recorded in the MIR region from  $3500\text{--}1500\text{ cm}^{-1}$ , at  $4\text{ cm}^{-1}$  wavenumber intervals, using a Bio-Rad Excalibur FTS 3000MX spectrometer, with Merlin software (Bio-Rad, Cambridge, MA, USA). Calculation was based on the peak height of the  $\text{CH}_2$  band at  $2855\text{ cm}^{-1}$ . Each stool sample was analysed in duplicate. The measured faecal lipid contents of the stools were used as secondary reference standards.

### *Determination of faecal fat using the ATR method*

The homogenized samples ( $n=119$ ) were spread out on a horizontal attenuated total reflectance (h-ATR) sampling device and measured with the FTS 3000MX spectrometer. Each sample was analysed in duplicate. Between each measurement the ATR crystal was successively cleaned with water and 96% alcohol and then carefully dried under a steam of nitrogen. The h-ATR trough plate assembly was supplied with a nine-reflection ZnSe crystal (Pike Technologies, Madison, WI, USA). Background measurements were performed with water on the crystal. The spectra were acquired in the MIR region from  $4000\text{--}750\text{ cm}^{-1}$  at  $4\text{ cm}^{-1}$  wavenumber intervals.

The recorded absorbance spectra were transferred to a program for quantitative analysis (The Unscrambler 6.11; Camo ASA, Oslo, Norway). With this program the fingerprint area of the spectra ( $1800\text{--}900\text{ cm}^{-1}$ ) was selected. The variances of the spectra were unified by normalization (10) of every spectrum with respect to its area under the curve of the selected wavenumbers. PLS regression was used for calibration, using the results of the flow cell method as reference data.

*Determination of faecal fat by means of the new 'NIR' transmission method*

For 12 samples we did not have sufficient material to perform the analysis. From the remaining 107 samples, about 2–3 g of each stool sample was transferred to a transparent plastic bag of polyethylene film, 25  $\mu\text{m}$  thick. After vacuum-sealing the bags, the samples were measured with a new transmission device (PB Sentechnology & Consultancy bv, Westeremden, The Netherlands). Using this device, the path length of the 'plastic bag' transmission cells was set at exactly 5.0 mm. The optical beam diameter was 1 cm. Each sample was measured twice, using a different spot of the sample in the plastic bag. A plastic bag filled with air was used as reference measurement (blank or 100%). The spectra were acquired in the NIR region from 850–1400 nm, at 5 nm wavelength intervals, using a scanning spectrometer (Fig. 1). This spectrometer was made up of the following commercially available parts: tungsten halogen light source (30 W), SpectraPro Monochromator (150 mm), grating (bandpass performance 5 nm), InGaAs detector (5 mm), quartz lightfiber (6 mm) and a Spectracard for data acquisition (all from Acton-Research, Massachusetts, USA). The recorded absorbance spectra were transferred to Unscrambler for quantitative analysis. The wavelength range of 900–1375 nm was selected. The spectra were normalized on their area under the curve. After this, each spectrum was smoothed with the Stavitzky-Golay method (12) using a polynomial order of 5. Partial least squares (PLS) regression was used for calibration, using the results of the flow cell method as reference data.

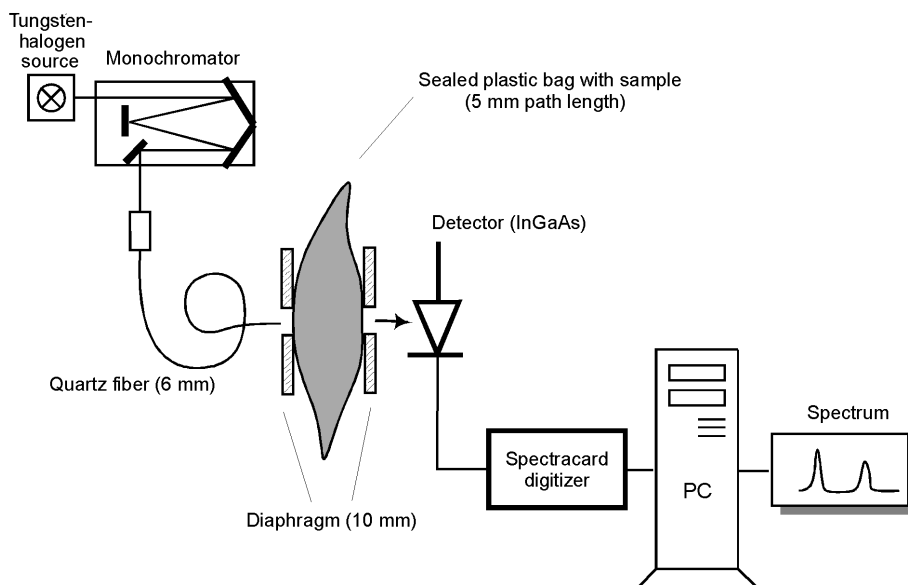


Figure 1. Scanning spectrometer and transmission device.

*PLS regression of the ATR and NIR measurements*

To prevent the occurrence of over-optimistic results derived from the PLS regression, full cross-validation was used, in which each calibration sample was successively removed from the calibration set and considered as an unknown. The result of this 'unknown sample' was then predicted based on the remaining  $n-1$  calibration samples. In all cases, the results

from this so-called internal ‘cross-validation’ method, based on patient data, were used for method comparison. The internal cross-validation results were also used for estimation of the optimum number of PLS factors (10), to prevent over- or under-fitting of the calibration model. The optimum number of PLS factors accounts for most of the predictive ability of the PLS calibration model. Using PLS regression, an extensive outlier- and influence detection was carried out (10).

#### *Method comparison*

The NCCLS EP9A guidelines were followed for method comparison using patient samples. Passing and Bablok regression (13) was used for the determination of the agreement between the reference method (flow-cell) and the ATR, and NIR methods. Samples with high within-method or between-method duplicate outliers were removed, according to the EP9A guidelines. In addition, we also calculated the within-method duplicate error, expressed as a standard deviation. The duplicate results of each sample were averaged, before performing the Passing and Bablok regression. In addition to the slope, intercept and correlation coefficient, the standard error of prediction (SEP, also expressed as  $S_{Y|X}$ ) was calculated for the estimation of the prediction bias.

#### *Precision and linearity of the methods*

The analytical precision and linearity of the ATR and NIR methods were tested using Lipofundin (Braun Melsungen AG, Melsungen, Germany), a lipid-containing intravenous fluid with 20% medium-chain/long-chain triglycerides. Two concentrations were used: 5% and 10% lipid. We also performed precision measurements using two stool samples, with 6.4 g/dL and 10.0 g/dL faecal fat, respectively.

The repeatability (intra-assay) was tested by means of an independent six-fold (Lipofundin) or ten-fold (stool samples) measurement at each level. The reproducibility (inter-assay) was obtained by measuring each level in duplicate on six (Lipofundin) or ten (stool samples) consecutive days. From these data, the intra-, and inter-assay coefficient of variation (CVs) were calculated. The linearity of the methods was tested with a Lipofundin dilution series, ranging from 2.5% to 20% lipid at 2.5% intervals. Each concentration level was measured five times. The Lipofundin dilution series was also used as a calibration model for the Lipofundin precision measurements, in accordance with the calculation methods as described for the ATR and NIR methods.

## **RESULTS**

The majority (77%) of the stool samples had normal water contents (70–80%), 10% had a water content less than 70% and 13% of the stools contains more than 80% water; 92% of the samples had a normal visually estimated homogeneity (no visible food, or other fragments). Faecal fat analyses with the flow-cell method showed that 117 out of 119 stool samples had fat contents in the range 0.2–18.9 g/dL [g fat/100 g faeces wet weight (w/w-%)]. These fat contents were equally distributed among the entire concentration range. Two of the 119 samples had an extremely high fat content (30.1 and 31.1 g/dL, respectively) and were only used as external validation samples for the ATR and NIR methods. Selection of the spectral regions for PLS regression was based on inspection of the PLS loading weights and regression coefficients for both the ATR and NIR methods. For the ATR method a continuous wavenumber range from 1800–900  $\text{cm}^{-1}$  (fingerprint area) was selected, whereas for the NIR method a wavelength range of 900–1375 nm was selected. Figure 2a shows ten NIR spectra for the selected wavelength range (stool samples ranging from 1.0 to

4.6 g/dL fat). Figure 2b shows the PLS regression coefficients of the NIR spectra of the final calibration model, showing the most important spectral phenomena. Figure 2c shows a spectrum of sunflower oil.

After PLS regression applied on the 117 ATR spectra, 16 samples proved to be outliers with great influence on the PLS regression analysis compared to the flow-cell method. After removing these 16 samples, 14 PLS factors were necessary to obtain a stable PLS regression using full cross-validation. For these 101 remaining calibration samples neither within- nor between-method outliers were found according to the EP9A guidelines. The Passing & Bablok comparison method applied on the 101 ATR samples demonstrated a slope of 0.935 (0.865 – 1.006), an intercept of 0.433 (-0.152 – 0.824), a correlation coefficient of 0.9411, and a SEP of 1.1 g/dL, compared with the flow-cell method.

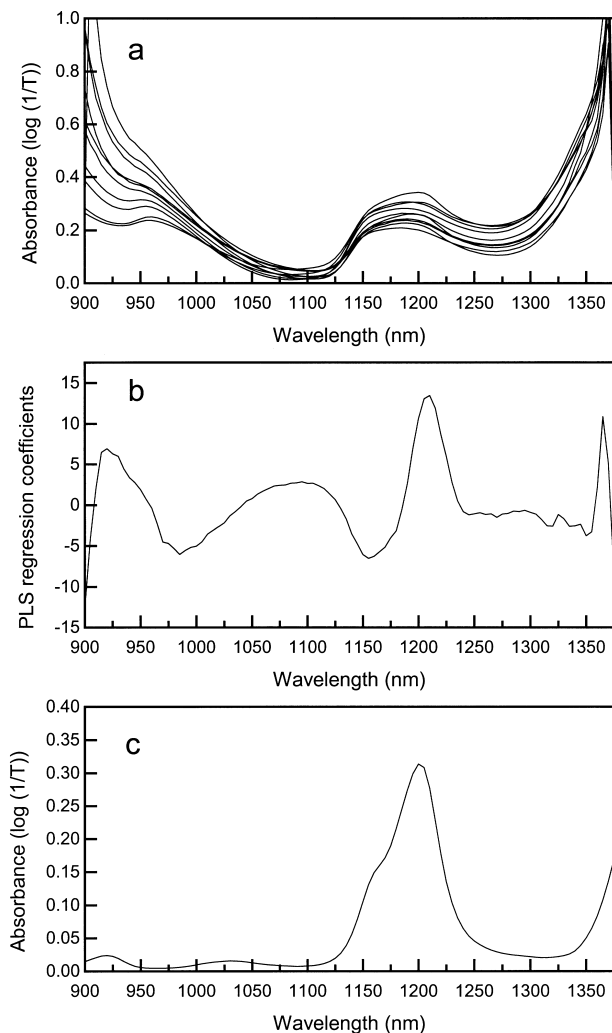


Figure 2. (a) Selected part of ten NIR spectra containing 1.0–4.6 g/dL fat. (b) PLS regression coefficients obtained with PLS regression analysis from the NIR calibration spectra. (c) Spectrum of sunflower oil.

PLS regression analysis of the 105 patient samples analysed with the NIR method did not show influencing outliers. However, 1 sample had to be removed from the calibration data, because it turned out to be a between-method outlier based on the EP9A guidelines. After removing this sample, 8 PLS factors were necessary to obtain a stable PLS regression, using full cross-validation. Passing & Bablok comparison applied on the PLS regression results from the remaining 104 NIR patient samples demonstrated a slope of 0.964 (0.902 – 1.025), an intercept of 0.218 (-0.141 – 0.576), a correlation coefficient of 0.9570 and a SEP of 1.1 g/dL fat, compared to the flow-cell method.

Figure 3 shows the Passing & Bablok method comparison charts for the ATR (3a) and NIR (3b) methods, both against the estimated fat content obtained with flow-cell method. All outliers, including the two samples with more than 30 g/dL fat, were analysed as external validation samples. The PLS predicted data from the 18 ATR outlier spectra are shown in Fig. 4. Seven out of the 18 outliers were stool samples with low or high water contents, or had abnormal visual homogeneity.

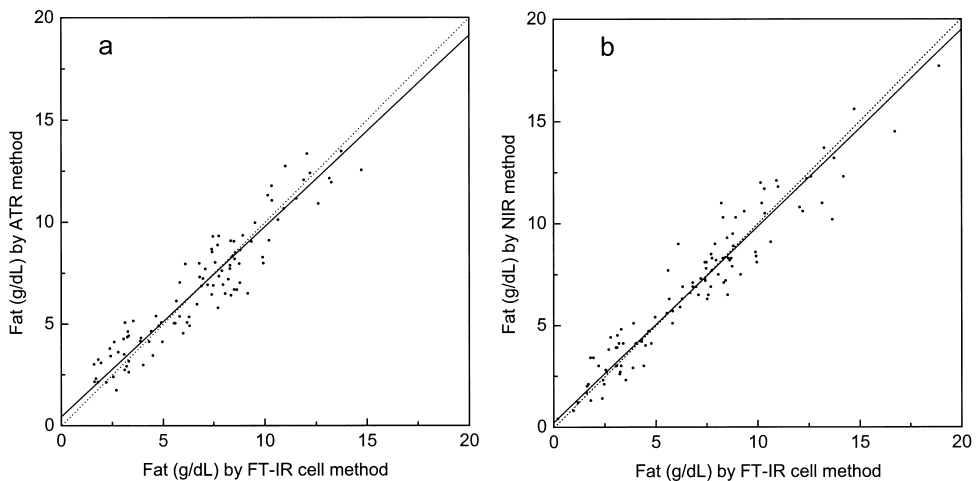


Figure 3. Passing & Bablok method comparison charts of the ATR method against the flow-cell method (a) and of the NIR method against the flow-cell method (b). The results from the ATR and NIR methods were obtained from the cross-validation results of the PLS analysis. The continuous line denotes the Passing & Bablok regression line and the dotted line denotes the line  $Y=X$ .

Only two of the ATR outlier samples corresponded to the 12 samples that were unavailable for NIR analysis because of lack of sample material. The predicted outcome (in g/dL) of the three samples (one between-method outlier and two extreme fat contents), which were left out of the NIR method, were: 11.9/15.1, 30.1/32.3 and 30.1/35.6 ('flow-cell'/ 'NIR', respectively).

The within-method duplicate error (expressed as standard deviation) obtained by the PLS cross-validation results of the ATR method was 0.55 g/dL, and 0.74 g/dL for the NIR method. The results from the precision and linearity measurements of the ATR and NIR methods are shown in Table 1.

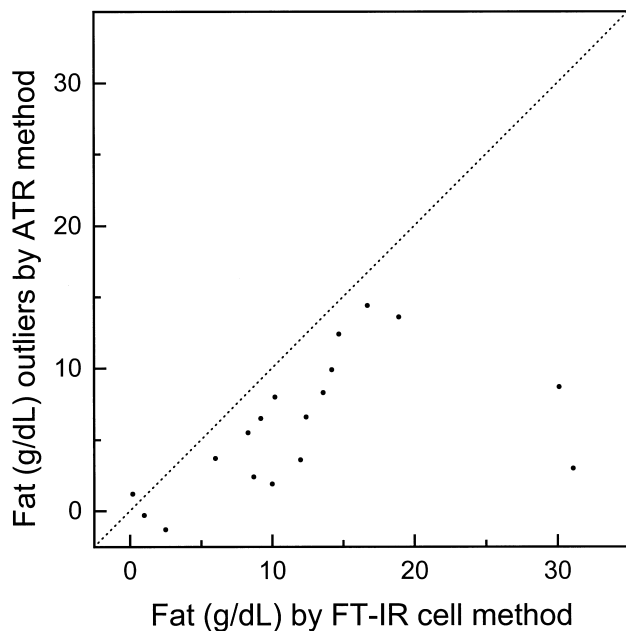


Figure 4. Scatter plot of 18 PLS-predicted outlier samples obtained with the ATR method against the results obtained with the flow-cell method. The dotted line denotes the line  $Y=X$ .

## DISCUSSION

FT-IR measurements of untreated stool samples can be performed rapidly, without prolonged handling of stool samples. Several authors reported on the use of a NIR reflectance method (4-7), some of them using a gold-plated integrating sphere as sampling device in combination with dedicated equipment. However, in clinical chemistry, the applicability of this rather expensive equipment is restricted to the application of fat, sugar, nitrogen and water contents of faeces. Therefore, the applicability of this equipment is hampered by the necessity of substantial numbers of samples to be analysed in order to justify the purchase of the apparatus (5).

Franck et al. (8) described the use of a FT-IR method, employing an ATR sampling device, for the prediction of faecal fat. They made preliminary sample selections, based on the faecal water content and the sample homogeneity. Only samples with normal water contents (called type P1 and P2) and normal homogeneity are allowed to enter the PLS regression analysis. Using these criteria, about 65% of all stool samples from patients suspected of steatorrhoea could be analysed by IR spectroscopy.

We also examined the applicability of this ATR method, but did not use such preliminary sample selections. Fifteen percent (18 out of 119) of our samples proved to be serious outliers using PLS regression with full cross-validation. Only 39% (seven out of 18) of these outliers turned out to be abnormal with respect to water contents or visual homogeneity, suggesting that there must be additional factors causing these outliers. ATR is a reflectance measurement with an extremely small penetration depth (maximally 5  $\mu\text{m}$  in the MIR region) in a sample spread on the ATR crystal (14). Fat is highly insoluble in



water, and even small departures from sample homogeneity could cause errors using this method. Using the ATR method, it was necessary to use 14 PLS factors to obtain a stable PLS regression equation. From this high number of factors it can be concluded that the ATR spectra (in the region 1800–900  $\text{cm}^{-1}$ ) have a too complex matrix structure to be solved with PLS regression.

Table 1. Precision and linearity

Method	Intra-assay % CV		Inter-assay % CV		Linear range* (2.5 g/dL intervals)
	5 g/dL fat	10 g/dL fat	5 g/dL fat	10 g/dL fat	
Lipofundin (n=6)	5 g/dL fat	10 g/dL fat	5 g/dL fat	10 g/dL fat	
ATR	2.6	1.1	2.8	1.6	2.5–17.5 g/dL
NIR	2.0	2.7	4.5	3.8	2.5–20.0 g/dL
Stool sample (n=10)	6.4 g/dL fat	10 g/dL fat	6.4 g/dL fat	10 g/dL fat	
ATR	2.9	2.0	3.5	2.2	
NIR	4.1	3.9	4.7	4.5	

CV = coefficient of variation; ATR = attenuated total reflection; NIR = near-infrared.

\* Repeated five times

We developed a new NIR method based on traditional transmission measurement. Unfortunately, water causes very high absorbance levels in the NIR region. To minimize these absorbance effects, we used only the lower part of the NIR infrared region (850–1400 nm) to retain sufficient path length for our transmission measurements. Another cause of the loss of light is the degree of light scattering caused by the sample. Using the NIR region from 850–1400 nm it was possible to use a path length of 5 mm. We also used an optical beam diameter of 1 cm, which consequently covered a large part of the sample. The thickness of the polyethylene film of plastic bag amounted to only 1% of the total path length. To compensate for possible side effects of the plastic film, we also used an empty plastic bag for blank measurements. Using a continuous spectral range from 900 to 1375 nm proved to be the best for PLS regression. In this spectral region, the  $\text{CH}_2$  second overtone band (C–H stretch) at 1215 nm (15) was demonstrated to be the most important band for lipid calibration (Fig. 2b). Only one sample was detected as an outlier, using this method.

After removing the outliers from the calibration datasets, the slopes of the two methods (ATR and NIR) did not substantially differ from 1.0 and the intercepts were not significantly different from 0.0 using Passing and Bablok regression analysis. The SEP of both methods was 1.1 g/dL using full cross-validation. We used a reasonable large number of calibration samples covering the actual concentration range. We also performed careful

PLS modelling, aiming at reduction prediction error. This was done by data pre-processing, optimising the number of PLS factors to prevent under and over-fitting of the model, and by looking for possible outliers. Therefore, we expect that the errors obtained with the full cross-validation procedure will be similar to the errors that will be obtained with an external validation set (10). Franck et al. (8) obtained a similar SEP (1.06 g/dL), using 91 external validation samples with concentrations ranging from 0.5 to 13.5 g/dL fat.

A large number of the 18 samples determined as outlier with the ATR method had severe deviations from the flow-cell method when their fat contents were determined as validation samples (see Fig. 4). No such severe deviations were found for the three validation samples of the NIR method: 11.9/15.1, 30.1/32.3 and 30.1/35.6 g/dL lipid for the flow-cell /NIR methods, respectively.

The linear ranges of the ATR and NIR calibration methods were large enough to cover the faecal fat concentrations of patient samples suspected of steatorrhoea. The analytical precision of the NIR method is slightly less than that of the ATR method, but good enough for routine practice, because the sampling errors of faecal fat determination are probably much larger. The within-method error of the NIR method (0.74 g/dL) is greater than that obtained by the ATR method (0.55 g/dL). Inspection of the EP9A scatter plot for the flow-cell and NIR methods of all the results (contrary to the mean of the replicates in Fig. 3b) showed that many of the replicate results from the NIR method deviate to some extent from each other. However, these differences were all within the acceptability levels of the within-method duplicate test, according to the EP9A guidelines. From this finding, we conclude that it would be advisable to perform at least a triplicate measurement of each stool sample, using the NIR method.

Based on our findings and those of Franck et al. – namely, that the fat content of a certain number of stool samples could not be estimated with the ATR method within reasonable accuracy limits compared with the reference method – we do not recommend the use of this method. As mentioned earlier, we think that this result is probably caused by the very small penetration depth in the sample material.

Finally, we conclude that the NIR method, using sealed plastic bags as transmission cells, makes handling of the stool samples less cumbersome for laboratory technicians. The linearity, precision and SEP of the NIR method seem to be adequate for routine practice of the calibrated and evaluated concentration range. However, some further experiments should be done with triplicate measurements of the stool samples and using an external validation set. Based on our findings we think that the new NIR method, based on transmission measurements, is a promising technique for routine analysis of faecal fat.

### **Acknowledgements**

We thank F. Hoolhorst and P. Koops for their skilful technical assistance. We thank Dr. B. van Haeringen and Dr. P. Schilder from Bio-Rad for their technical support.

### **REFERENCES**

1. Van de Kamer JH, Ten Bokkel Huinink H, Weijers HA. Rapid method for the determination of fat in feces. *J. Biol Chem* 1949; 177: 347–55.
2. 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.

3. Phuapradit P, Narang A, Mendonca P, Harris DA, Baum JD. The steatocrit; a simple method for estimating stool fat content in newborn infants. *Arch Dis Child* 1981; 56: 725–7.
4. Stein J, Purschian B, Bieniek U, Caspary WF, Lembcke B. Near-infrared reflectance analysis: a new dimension in the investigation of malabsorption syndromes. *Eur J Gastroenterol Hepatol* 1994; 6: 889–94.
5. Bekers O, Postma C, Lombarts AJ. Determination of faecal fat by near-infrared spectroscopy. *Eur J Clin Chem Clin Biochem* 1995; 33: 83–6.
6. Neumeister V, Henker J, Kaltenborn G, Sprössig C, Jaross W. Simultaneous determination of fecal fat, nitrogen, and water by near-infrared reflectance spectroscopy. *J Pediatr Gastroenterol Nutr* 1997; 25: 388–93.
7. 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.
8. 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.
9. Volmer M, Wolthers BG, Metting HJ, Haan THY, Coenegracht PMJ, van der Slik W. Artificial neural network predictions of urinary calculus compositions analyzed with infrared spectroscopy. *Clin Chem* 1994; 40: 1692–7.
10. Martens H, Naes T. *Multivariate calibration*. Chichester, UK: John Wiley & Sons, 1989.
11. Jakobs BS, Volmer M, Swinkels DW, Hofstede 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: 343–49.
12. Stavitzky A, Golay MJE. Smoothing and differentiation of data by simplified least squares procedures. *Anal Chem* 1964; 36: 1627–39.
13. Passing H, Bablok W. A new biometrical procedure for testing the equality of measurements from two different analytical methods. *J Clin Chem Clin Biochem* 1983; 21: 709–20.
14. Harrick NJ. *Internal reflection spectroscopy*. Interscience Publishers, New York, 1967.
15. Osborne BG, Fearn T, Hindle PH. *Practical NIR spectroscopy: with applications in food and beverage analysis*. Second edition. Longman Scientific & Technical, Harlow, England, 1988, p 30.



## Part II

### Urinary calculus analysis



## Chapter 4

# **Partial least-squares regression for routine analysis of urinary calculus composition with Fourier transform infrared analysis**

M. Volmer<sup>1</sup>, A. Bolck<sup>2</sup>, B.G. Wolthers<sup>1</sup>, A.J. de Ruiter<sup>3</sup>,  
D.A. Doornbos<sup>2</sup>, W. van der Slik<sup>1</sup>

<sup>1</sup> *Central Laboratory for Clinical Chemistry, University Hospital Groningen  
PO Box 30001, 9700 RB Groningen, <sup>2</sup> Chemometric Research Group,  
University Centre for Pharmacy, A. Deusinglaan 2, 29713 AW Groningen,  
<sup>3</sup> The Netherlands Department of Urology, University Hospital Groningen, PO Box 30001,  
9700 RB Groningen, The Netherlands*

*Clin Chem 1993;39(6):948-954*

### ABSTRACT

Quantitative assessment of urinary calculus (renal stone) constituents by infrared analysis (IR) is hampered by the need of expert knowledge for spectrum interpretation. Our laboratory performed a computerized search of several libraries, containing 235 reference spectra from various mixtures with different proportions was performed. Library search was followed by visual interpretation of band intensities for more precise semi-quantitative determination of the composition. We tested partial least-squares (PLS) regression for the most frequently occurring compositions of urinary calculi. Using a constrained mixture design, we prepared various samples containing whewellite, weddellite and carbonate apatite and used these as calibration set for PLS regression. The value of PLS analysis was investigated by the assay of known artificial mixtures, and selected patients' samples from which the semiquantitative compositions were determined by computerized library search followed by visual interpretation. Compared with this method, PLS analysis was superior with respect to accuracy and necessity of expert knowledge. Apart from some practical limitations in data-handling facilities, we believe that PLS regression offers a promising tool for routine quantification, not only for whewellite, weddellite and carbonate apatite, but also for other compositions of the urinary calculus.

---

### INTRODUCTION

At present, >95% of all patients with urolithiasis are successfully treated with extracorporeal shock wave lithotripsy [ESWL] in combination with intracoporal methods (1) for in situ calculi fragmentation. For patients with urinary stones in the ureter the percentage of success with ESWL treatment is <95%. Because of the high frequency of stone recurrence, which amounts to ~50% after 10 years (1), and limitations that exist for manipulative surgical intervention, there is increasing need for causal therapy. It is therefore important to be informed about the composition of the urinary calculi. In contrast to classical surgery for extracting renal calculi from the urinary tract, ESWL is hindered by the fact that often many calculi fragments are lost for further analysis. Since the introduction of ESWL in our hospital, ~3 years ago, we have found that, because of insufficient material, the composition of 34% of the urinary calculi presented to our laboratory could not, or only partly be determined with the classical wet chemical procedure. Infrared analysis (IR) of urinary calculi does not have this disadvantage. Using the KBr disk technique, samples as small as 1 mg can be analyzed (2). Other advantages of IR analysis over wet chemical analysis are higher speed, better reproducibility, and uniform sensitivities for all components (3). For these reasons the IR method was introduced in our laboratory.

The primary objective of urinary calculus analysis is to determine the qualitative composition. However, quantitative assessment of constituents, with a precision of 5-10% for individual components, is also important (4). Routine usage of IR is hampered by the need to interpret IR spectra from >20 components in different combinations and proportions. Systematic schemes have been developed for objective qualitative analysis of IR spectra of calculi (5). Hesse and Sanders (6) issued an atlas containing a collection of about 225 IR reference spectra from the most commonly encountered urinary calculus components and their binary and ternary mixtures. At present most IR spectrophotometers are accompanied by software packages that offer possibilities to produce libraries of digitized spectra and of searching for unknown compositions by matching the unknown spectra with those within the library. Expert systems



should also be suitable for pattern recognition tasks. Although described for x-ray diffraction analysis of urinary calculi (7), no expert system has never been applied to the IR technique as far as we know. Unfortunately all these methods for the interpretation of IR spectra are only semiquantitative.

Urinary calculi are multicomponent mixtures, most being composed of two or three components. Simple linear regression is not appropriate for quantification in these cases. Multivariate calibration methods, such as classical multiple linear regression, inverse linear regression, principal component regression and partial least-squares (PLS) regression, have been developed for quantitative spectral analysis (8). The most robust alternative of these methods (9), the PLS method, is used for some spectral applications (10,11). If the Lambert–Beer law holds spectroscopic data (X) can be linearly related to the concentrations (Y), after factorial decomposition with PLS. The reference samples, also called training set, or calibration matrix, are used to estimate the regression coefficients. A test set consisting of independent reference samples is used to validate the quality (stability) of the PLS model. This test set can also be used to estimate the composition of unknown samples.

The aim of our work was to develop a library with IR reference spectra of authentic components and mixtures and to test a PLS model for the most frequently occurring compositions of urinary calculi.

## MATERIALS AND METHODS

### *Samples*

Library search and PLS were performed on mixtures of calcium oxalate dihydrate (whewellite; BDH, Brunschwig Chemie, The Netherlands), calcium oxalate monohydrate (weddellite), and carbonate apatite. Because weddellite and carbonate apatite were commercially unavailable, we synthesized the weddellite (12) and obtained the carbonate apatite from a collection of carefully selected patients' samples. The purity of carbonate apatite was checked by comparison of the IR spectra with the spectrum from the Hesse atlas (6) and by means of wet chemical analysis.

Because the component fractions of the mixtures must add to unity (100%), the constrained factor space of a ternary mixture can be mapped to an equilateral triangle (Figure 1). The angular points of the triangle mark the single components, the edges the binary and the inside points the ternary mixtures. The components of each individual point in the triangle sum up to 100%. The design for the calibrating (training) of the PLS model consisted of 25 mixture points. This mixture design had, except whewellite / weddellite mixtures, a uniform distribution for all calibration mixtures over the triangle (Figure 1A). To obtain a better differentiation between both calcium oxalates, we used mixtures in steps of 10% for whewellite and weddellite. The mixtures for library search, consisting of 46 samples, were prepared according to the design depicted in Figure 1B.

An independent test set was formed for validation of the quality of the PLS model. This test set contained 9 artificial mixtures of whewellite and weddellite; 8 artificial mixtures of whewellite, weddellite and carbonate apatite; and 20 selected samples of urinary calculi from patients concerning various compositions of whewellite, weddellite and carbonate apatite. Two of these patients' samples contained calcium hydrogen phosphate dihydrate (brushite) or magnesium ammonium phosphate hexahydrate (struvite) as an extra component.

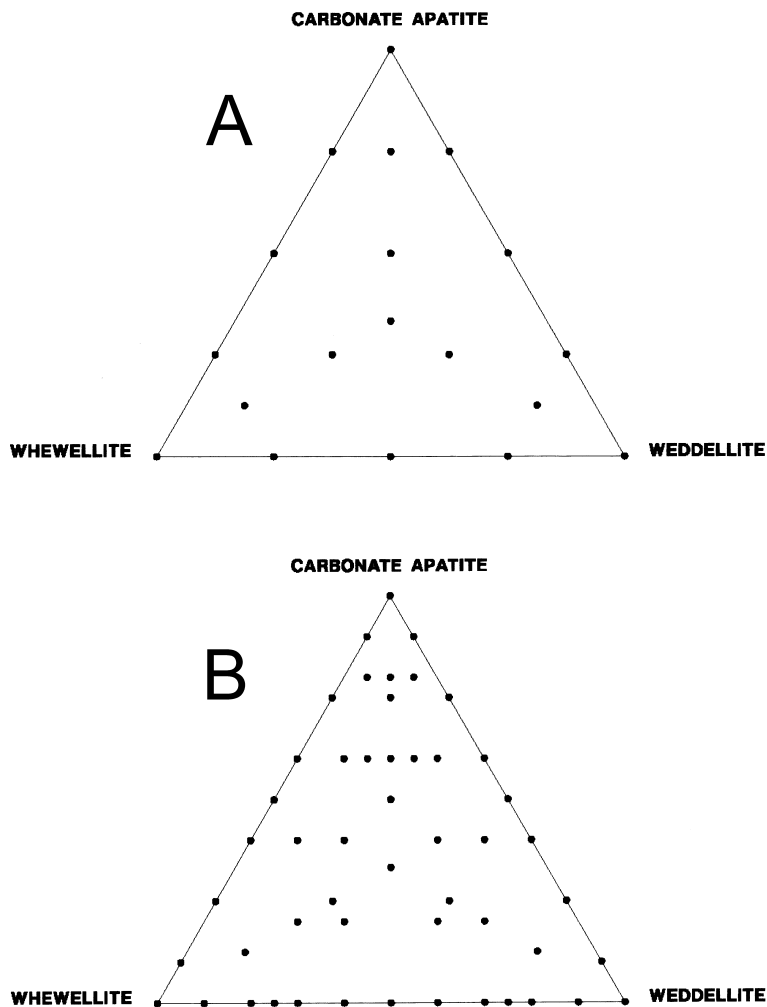


Figure 1. Ternary mixture designs of whewellite, weddellite, and carbonate apatite, as used for the PLS training set (A) and spectral library search (B). The designs are constrained factor spaces from which all component fractions add up to 100%.

### *Analytical procedures*

Patients' samples were totally ground in an agate mortar. Potassium bromide (KBr Uvasol; Merck, The Netherlands) was pulverized, dried at 100 °C for at least 24 h and stored at 37 °C until use. Only 1 mg of grounded calculi material was mixed with 180 mg KBr. Pellets of mixtures for library search and PLS were prepared by weighing a minimum of 10 mg for the minor component of the mixture. We placed 100 mg of these mixtures in a 13-mm-diameter evacuable pellet die (Specac: Kent, England), evacuated air from the pellet die for 2 minutes, then formed pellets by applying a pressure of 750 MPa. The vacuum was maintained for 2 min after release of pressure. We used 100-mg pellets of KBr for preparation of the background spectra.

*IR analysis*

Transmission spectra were recorded after automatic subtraction of background spectra with a Mattson Galaxy FT-IR 3020 spectrophotometer (Mattson Instruments; WI) for wavenumbers 4000 to 400  $\text{cm}^{-1}$  and spectral resolution of 4  $\text{cm}^{-1}$ .

*Data-handling procedures*

Recorded spectra were converted to absorption spectra with the First v1.52 program (Mattson Instruments). Firstbase v1.52 (Mattson Instruments) was used to store data from mixtures needed for the library. The Pearson linear or product-moment correlation coefficient ( $r$ ), obtained from the First v1.52 program, was used as the measure of agreement between the unknown spectrum and the reference spectra from the library.

To prevent storage problems, we used Firstbase to reduce recorded spectra for PLS regression to a spectral resolution of 16  $\text{cm}^{-1}$  before storing them as ASCII data. The resulting 225 data points for each spectrum were collected in a matrix by means of a spreadsheet program (Quattro Pro v3.0, Borland International Inc., CA). This matrix of spectra, with each column of the spreadsheet containing the absorptions of a spectrum, was transposed to an orientation in rows (each spectrum in a row). This spectral matrix was imported into Unscrambler II v3.0 (Camo A/S; Trondheim, Norway) for PLS regression, after which the concentration data of the calibration matrix (Y) were entered. After dividing the spectral data into a training set and a test set, we unified the total variance of all spectra (X) by normalizing (13) every spectrum with respect to its area under the curve for the selected wavenumbers. The absorbancies at each wavenumber of the absorption matrix X were mean centered. No scaling was carried out on these variables. Regression with the PLS-2 algorithm was carried out for wavenumbers 4000 to 400  $\text{cm}^{-1}$  and wavenumbers 1392 to 496  $\text{cm}^{-1}$ , respectively. The PLS calibration model relates the frequency data (X) to the concentration data (Y) through a smaller set of variables, the so-called latent factors, or components. Cross-validation, for estimating the number of these factors for an optimal PLS model with maximum explained variance, was carried out by inspecting the variance plot for the first local minimum (9,14). All further calculations for calibration and test set data were carried out with the selected number of factors.

*Statistical and validation methods.*

Calibration lines were calculated for validation of the quality of the PLS model. The predictive value of the PLS model was validated with the independent test set. Regression lines, as well as lines deviating plus or minus 10% from the regression lines, were calculated, based on actual and predicted values of the artificial validation samples. Uncertainty limits were calculated by the PLS regression program for each predicted value, as measure of the probability of a correct response. Compositions of patients' samples were estimated by comparison of unknown spectra with reference spectra from the library, followed by an independent visual inspection of the intensities of the bands by two specialized technicians. Validation was also performed by visual comparing of the spectrum from a patient's sample with the spectrum of an artificial sample that corresponds as well as possible to the PLS-predicted composition of the patient's sample.

**RESULTS**

From the retrospective data of the past 3 years of wet chemical analysis of urinary calculi in our laboratory and a pilot study on IR analysis of the calculi composition, we found that 81%

of all urinary calculi contained calcium. The vast majority of these calculi were binary and ternary mixtures composed of both calcium oxalates and (or) carbonate apatite. For this reason the PLS model was tested with these three components.

IR transmission spectra of whewellite, weddellite, and carbonate apatite showed different bands and intensities at different wavenumbers (Figure 2). However, intensities of characteristic bands of the three components in the spectral range 4000–400  $\text{cm}^{-1}$  were found to be variable not only as to their specific composition, but also as to other factors. Consequently, we selected only wavenumbers 496 to 1392  $\text{cm}^{-1}$  (see frame in Figure 2F) were selected for analysis of the three components with library search. The symmetrical CO stretch vibration band at 1328  $\text{cm}^{-1}$  is of great significance for identification of both calcium oxalates. Whewellite can be distinguished from weddellite by the ratio of the CO deformation bands at 512 and 784  $\text{cm}^{-1}$ . The broad bands at 592 and 626  $\text{cm}^{-1}$  can also be used to distinguish whewellite from weddellite. The phosphate stretch vibration band at 1048  $\text{cm}^{-1}$  and its shape is important for the identification of carbonate apatite; in addition the deformation vibration bands at 574 and 606  $\text{cm}^{-1}$  are more or less characteristic of this compound. For quantitative determination of mixtures of calcium oxalate and carbonate apatite, the ratio of the characteristic bands at 1328 (oxalate) and 1048 (phosphate)  $\text{cm}^{-1}$  is important. For mixtures containing equal weights (50/50) of whewellite and carbonate apatite (Figure 2D) or weddellite and carbonate apatite (Figure 2E), the ratios of the bands are clearly different. Figure 2 F is an example of a spectrum from a patient's sample composed of whewellite, weddellite, and Carbonate apatite (25/20/55 by wt). This composition was derived as follows: Library search was used to determine the global composition of this sample, and then interpolation by means of visual inspection of the mutual band intensities and comparison with reference spectra established a more precise composition.

Principal Component Analysis (PCA) (15), as part of the Unscrambler program, was carried out on the selected wavenumber range of the spectral matrix to select only those wavenumbers that vary as consequence of the components. This analysis selected 33 wavenumbers from wavenumbers 496 to 1392  $\text{cm}^{-1}$  for further study. Because whewellite and weddellite have strongly overlapping bands and can only be distinguished by the forms and intensities of the bands (Figure 2, A and B), we used PLS regression to solve this multicomponent problem.

Cross-validation carried out on the calibration set (see Figure 1A) with PLS determined that the first 2 factors explained 98.9% of the variation of the concentrations. These two factors, computed by a decomposition of the spectral matrix, represent the main, but hidden, systematic variation in the calibration data. Further calibration and prediction of the concentrations of the unknown samples were carried out with these two factors to prevent overdetermination of the outcome. Regression lines, based on the compositions of the PLS mixture design (Figure 1A) were calculated with the PLS-2 algorithm for whewellite, weddellite, and carbonate apatite (Figure 3C). The x-axis depicts the actual compositions of the components in the calibration set in percentage by weight, whereas the y-axis depicts the compositions predicted by the PLS model.

Table 1 demonstrates a direct correspondence between the actual and PLS-predicted compositions for several artificial mixtures and patients' samples, for validation of the PLS model. Correspondence between actual and PLS-predicted composition of these independent

test samples is graphically depicted in Figure 4. Because compound values  $<10$  g% ( $<100$   $\mu\text{g/g}$ ) were designated as "trace amounts" for actual compositions of some patients' samples (Table 1), we used 5 g% as the mean actual value for these components in Figure 4. The vast majority of the predicted values of the artificial mixtures show  $<5\%$  deviation from the actual composition. Two patients' validation samples deviated slightly  $>10\%$  from the calculated regression lines for whewellite (Figure 4A) and weddellite (Figure 4B), both depicted as a triangle outside the dotted lines. The "erroneously matched" samples containing brushite (Stone19) and struvite (Stone20) were added to the validation set to test the robustness of the PLS model for compositions deviating from the calibration set. The brushite-containing calculus deviated  $>10\%$  for whewellite (Figure 4A), weddellite (Figure 4B), and carbonate apatite (Figure 4C), whereas the struvite-containing sample did not.

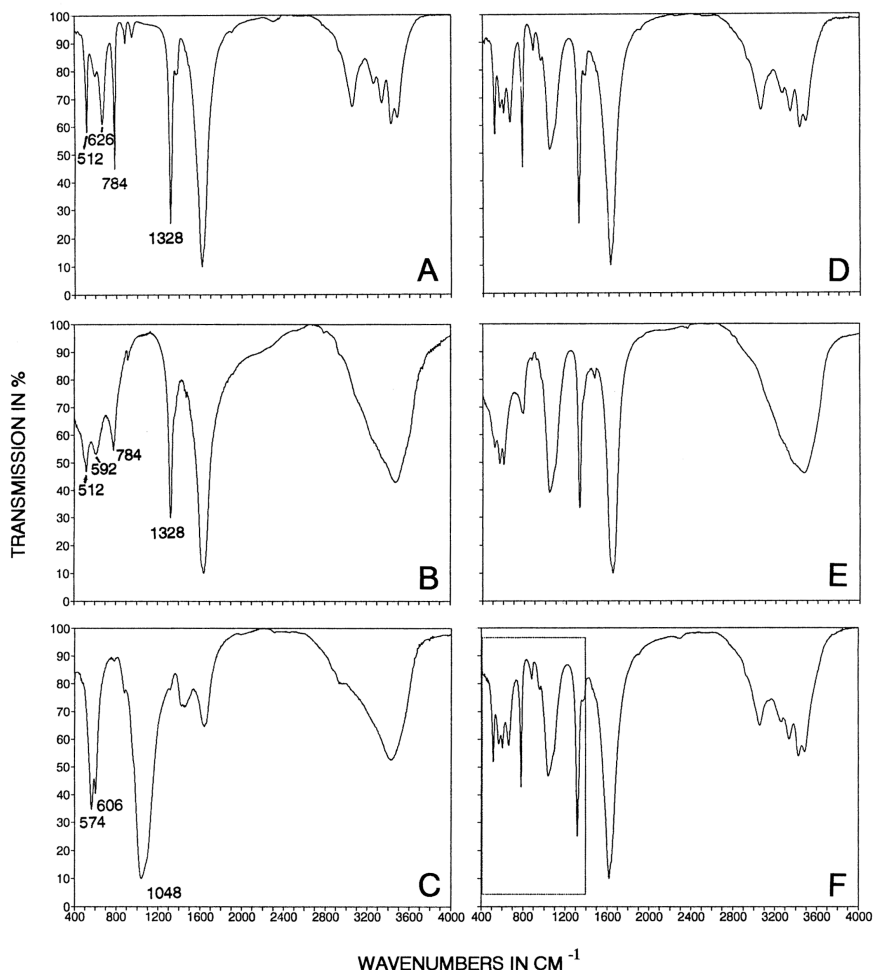


Figure 2. Infrared transmission spectra of whewellite (A), weddellite (B), carbonate apatite (C) and equal weight mixtures of whewellite/carbonate apatite (D) and weddellite/carbonate apatite (E). Panel F, representing a spectrum of a patient's sample containing whewellite, weddellite and carbonate apatite, depicts the selected wavenumber range used for PLS and library search.

The lowest and highest uncertainty limits of the predicted values ranged from 0.8–4.0% for the artificial samples, 2.2–8.8% for the patients' samples, and 6.0–20.0% for the "erroneously matched" patients' validation samples, for all three components.

Comparison of the transmission spectrum of the patient sample, indicated as Stone14 in Table 1 and an artificial mixture with nearly the same composition as predicted by PLS for the patient's sample (23/19/58 and 25/16/59% by wt. for whewellite/weddellite/ carbonate apatite, respectively) demonstrates good correspondence for wavenumber range 496–1392  $\text{cm}^{-1}$  (Figure 5). The correlation coefficient between the two spectra was 0.994.

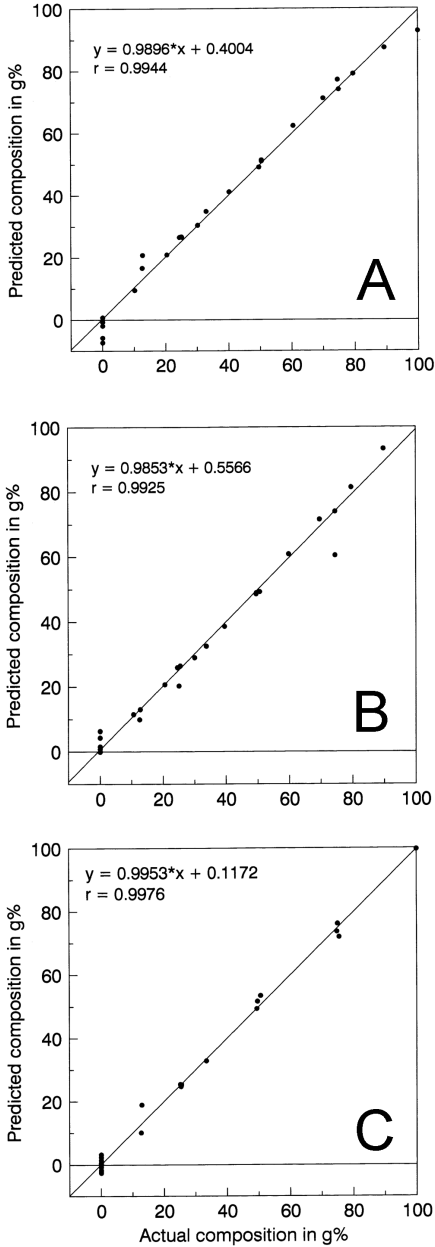


Figure 3. Calibration lines from whewellite (A), weddellite (B), and carbonate apatite (C). Actual values are the known mass percentages of the respective components of the calibration mixtures, whereas the predicted values are derived from PLS regression analysis.

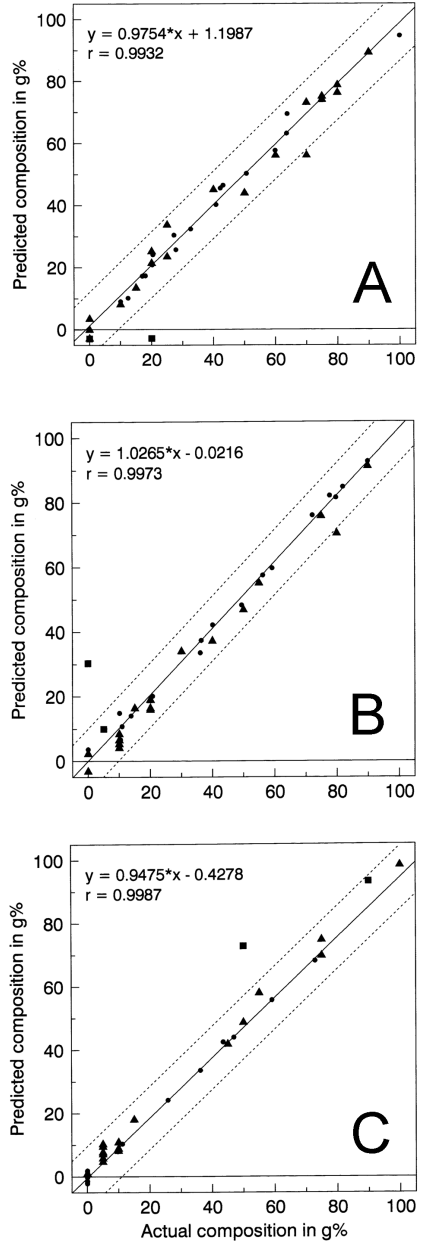


Figure 4. PLS predicted compositions (y) compared with actual compositions (x) for artificial (●), patients' (▲) and "erroneously matched" patients' (■) validation samples for whewellite (A), weddellite (B), and carbonate apatite (C). Regression lines (—) and lines deviating plus or minus 10% (---) from the calculated regression lines are based on the artificial (●) validation mixtures

Table 1. Actual and PLS-predicted compositions (g%) of artificial mixtures and urinary calculi, for validation of the PLS-model<sup>a</sup>.

Sample	Whewellite		Weddellite		Carbonate apatite		Brushite	Struvite
	Actual	Predicted	Actual	Predicted	Actual	Predicted	Actual	
Mix1	18.0	17.3	82.0	84.7	0.0	-2.0		
Mix2	59.9	57.6	40.1	42.1	0.0	0.3		
Mix3	40.8	40.2	59.2	59.6	0.0	0.2		
Mix4	10.0	9.0	90.0	92.6	0.0	-1.6		
Mix5	50.6	50.2	49.4	48.2	0.0	1.6		
Mix6	27.8	25.7	72.2	75.9	0.0	-1.5		
Mix7	100.0	94.6	0.0	3.6	0.0	1.9		
Mix8	20.2	20.8	79.8	81.4	0.0	-2.2		
Mix9	63.6	63.1	36.4	37.3	0.0	-0.3		
Mix10	63.8	69.4	10.4	6.6	25.8	24.1		
Mix11	20.4	24.1	36.1	33.5	43.5	42.4		
Mix12	17.0	17.2	10.1	14.8	72.9	68.0		
Mix13	42.2	45.5	10.9	10.7	46.9	43.9		
Mix14	32.6	32.4	56.2	57.4	11.2	10.3		
Mix15	43.1	46.4	20.7	20.1	36.2	33.5		
Mix16	27.2	30.4	13.8	14.0	59.1	55.6		
Mix17	12.4	10.1	77.8	82.0	9.8	7.9		
Stone1	75.0	75.0	15.0	16.3	10.0	8.8		
Stone2	20.0	21.3	10.0	4.0	75.0	74.6		
Stone3	80.0	78.6	20.0	15.8	Trace <sup>b</sup>	5.7		
Stone4	0.0	3.3	55.0	55.0	45.0	41.7		
Stone5	90.0	89.2	10.0	5.2	Trace	5.7		
Stone6	20.0	25.1	80.0	70.3	Trace	4.6		
Stone7	Trace	-0.2	0.0	2.2	100.0	98.1		
Stone8	25.0	33.7	0.0	-3.3	75.0	69.6		
Stone9	70.0	73.0	20.0	18.9	10.0	8.1		
Stone10	40.0	45.0	10.0	6.4	50.0	48.6		
Stone11	75.0	73.9	10.0	8.3	15.0	17.9		
Stone12	70.0	56.0	30.0	33.9	Trace	10.2		
Stone13	80.0	76.2	20.0	16.3	Trace	7.6		
Stone14	25.0	23.4	20.0	18.8	55.0	57.8		
Stone15	10.0	8.0	90.0	91.1	0.0	0.9		
Stone16	60.0	56.0	40.0	37.1	Trace	7.0		
Stone17	15.0	13.4	75.0	75.7	10.0	10.8		
Stone18	50.0	43.9	50.0	46.7	Trace	9.4		
Stone19	20.0	-2.9	0.0	30.3	50.0	72.7	30.0	
Stone20	0.0	-3.0	Trace	9.9	90.0	93.1		10.0

<sup>a</sup> Actual are the compositions of known artificial mixtures (Mix1–Mix17) and patients' samples (Stone1–Stone20) from which the actual compositions were estimated by library search, followed by visual inspection of the spectral band intensities.

<sup>b</sup> Less than 10%

## DISCUSSION

We found that deviations up to 20% from the actual quantitative compositions of the urinary calculi, as derived from visual inspection of spectral band intensities, can be expected after computerized library search. Library search is useful for the determination of the qualitative and the global quantitative composition of the urinary calculus, but should be followed by a visual inspection of the spectrum for more precise determination of the quantitative composition. If the urinary calculus is composed of more than 2 components, interpretation of the proportions of band intensities by visual inspection of the spectrum becomes rather complex. The ratios of oxalate (1328 cm<sup>-1</sup>) and phosphate (1048 cm<sup>-1</sup>) are quite different for



equal mixtures of whewellite and carbonate apatite (figure 2D) in comparison with weddellite and carbonate apatite (Figure 2E). Expert knowledge is needed to interpret the ratios of the relevant spectral bands for ternary mixtures composed of whewellite, weddellite, and carbonate apatite (Figure 2F). In our opinion it would for this reason be virtually impossible to build an expert system for routine quantification of urinary calculus composition, consisting of many binary and ternary mixtures in various proportions of >20 components.

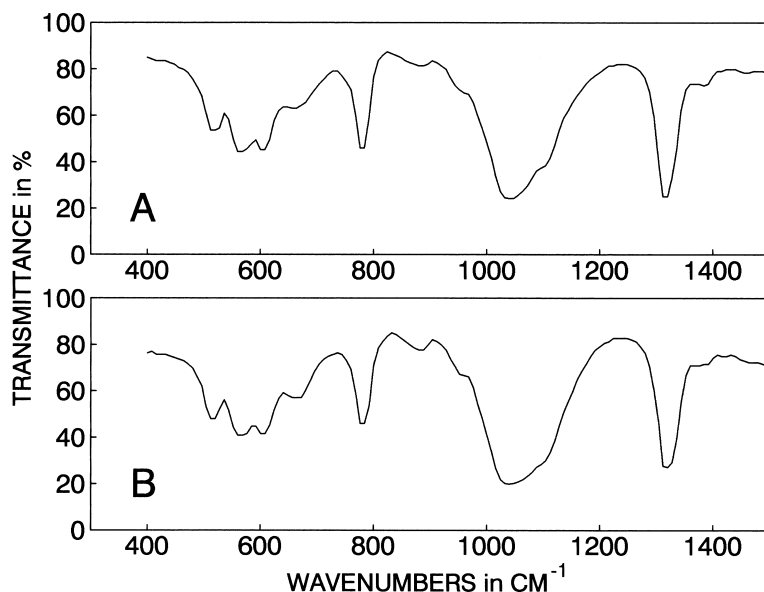


Figure 5. Transmission spectrum of a patient sample (A) containing a PLS-predicted composition of whewellite, weddellite, and carbonate apatite (23/19/58, by wt.) and a spectrum of an artificial mixture (B) with nearly the same composition (25/16/59, respectively) as the patient's sample.

It is important for PLS regression, as well as library search, to remove wavenumbers with no structural information as well as spectral bands that are variable because of factors other than the components themselves. We found that five factors were needed to explain all systematic variation with PLS regression from the fully normalized spectrum from 400 to  $4000\text{ cm}^{-1}$  (data not shown). The calibration lines calculated from the full spectrum were very good. However, with these 5 factors a very disappointing outcome of the validation samples was obtained, as a result of serious over-determination of the PLS model. For reliable results, selection of the appropriate wavenumbers is apparently needed.

The regression lines in Figure 3 demonstrate that PLS regression, with two factors obtained from the selected wavenumber range, gave most satisfactory results for carbonate apatite. An even better impression of the reliability of the PLS model can be obtained from the results of the independent validation samples, both artificial mixtures and renal stones. The results from Table 1 as well as the regression lines with their 10% deviation lines in Figure 4, demonstrate that PLS regression is very useful for the determination of the quantitative composition. Only two artificial mixtures had predicted values of whewellite that deviated >5%, but still <6% from the actual composition (MIX7 and MIX10 in Table 1). Predicted values of all other

components of the artificial mixtures deviated by <5%. The PLS predicted compositions of the patients' samples showed that only two of the samples deviated slightly by >10% from the actual compositions, as determined by visual inspection of the spectrum. The brushite-containing validation sample demonstrates that qualitative compositions dissimilar to the calibration set result in serious deviations from the actual compositions for all components. The struvite-containing calculus did not show these deviations, which can be explained by its low concentration (10% by wt.) and its spectral similarity with carbonate apatite. From these findings it can be concluded that prediction of the composition of unknown samples with PLS regression should be carried out only for qualitative compositions similar to the calibration set. The uncertainty limit of the predicted value, as provided by PLS regression, is a measure of degree of certainty of the match and provides some measure of probability of a correct response. Uncertainty limits <4% for the artificial and 8.8% for the patients' validation samples are good enough for routine analysis of urinary calculus composition, with PLS regression. The brushite- and struvite-containing calculi had a low probability of a correct response, as found by their high uncertainty limits.

Practical use of PLS regression in the laboratory is still hampered by the fact that many steps – e.g., data reduction by decreasing spectral resolution, wavenumber selection and normalization of the selected wavenumbers – should be carried out before the actual PLS regression, if one is to obtain reliable results for the unknown urinary calculus composition. At present many different programs must be used to achieve that purpose. Manufacturers of spectroscopic software may want to reduce these manipulations by making software available in one program, including the PLS-1 and PLS-2 algorithm.

#### **Acknowledgements**

We are indebted to T.H.Y de Haan and J.H. Verwoerd for technical assistance. We thank A. Hesse for providing information about the synthesis of some components.

#### **REFERENCES**

1. Lingeman JE, Smith LH, Woods JR, Newman DM. Urinary calculi. ESWL, endourology and medical therapy. Philadelphia-London, Lea & Febiger, 1989.
2. Gault MH, Ahmed M, Kalra J, Senciall I, Cohen W, Churchill D. Comparison of infrared and wet chemical analysis of urinary tract calculi. *Clin Chim Acta* 1980;104:349-359.
3. Röhle G, Voigt U, Hesse A, Breuer H. Ergebnisse aus Ringversuchen für Harnsteinanalysen. *J Clin Chem Clin Biochem* 1982;20:851-859.
4. Hesse A, Sanders G, Döring R, Oelichmann J. Infrarotspektroskopische Harnsteinanalyse: automatisierte Spectrenauswertung mit Hilfe der Factoranalyse. *Fresenius Z Anal Chem* 1988;330:372-373.
5. Oliver LK, Sweet RV. A system of interpretation of infrared spectra of calculi for routine use in the clinical laboratory. *Clin Chim Acta* 1976;72:17-32.
6. Hesse A, Sanders G. Atlas of infrared spectra for the analysis of urinary concrements. Georg Thieme Verlag Stuttgart, 1988.
7. Blijenberg BG, Wulkan RW, Zwang L, Liem TL, Leijnse B. Computergestützte Auswertung von Röntgendiffractometrie-Analysen. *J Clin Chem Clin Biochem* 1987;25:719-722.

8. Haaland DM, Thomas EV. Partial least-squares methods for spectral analysis. 1. Relation to other quantitative calibration methods and the extraction of qualitative information. *Anal Chem* 1988;80:1193-1202.
9. Geladi P, Kowalski BR. Partial least-squares regression: a tutorial. *Anal Chim Acta* 1986;185:1-17.
10. Lindberg W, Persson JÅ. Partial least-squares method for spectrofluorimetric analysis of mixtures of humic acid and ligninsulfonate. *Anal Chem* 1983;55:643-648.
11. Kruse-Jarres JD, Janatsch G, Gless U. Reagentless determination of glucose and other constituents in blood by ATR-FT-IR-spectroscopy. *Clin Chem* 1989;35:1854-1856.
12. Berg W, Hesse A, Schneider HJ. A contribution to the formation mechanism of calcium oxalate. III. On the role of magnesium in the formation of oxalate calculi. *Urol Res* 1976;4:161-167.
13. Martens H, Naes T. *Multivariate calibration*. John Wiley and Sons, Chichester, England, 1991.
14. Unscrambler II version 3.0. User's guide. CAMO A/S, Trondheim, Norway, 1990.
15. Stevens J. *Applied multivariate statistics for the social sciences*. Lawrence Erlbaum Associates, Hillsdale, New Jersey, USA, 1986.



## Chapter 5

# **Artificial neural network predictions of urinary calculus compositions analyzed with infrared spectroscopy**

Marcel Volmer<sup>1</sup>, Bert G. Wolthers<sup>1</sup>, Harm J. Metting<sup>2</sup>, Thijs H.Y. de Haan<sup>1</sup>,  
Pierre M.J. Coenegracht<sup>2</sup>, and Wim van der Slik<sup>1</sup>

<sup>1</sup> *Central Laboratory for Clinical Chemistry, University Hospital Groningen, P.O. Box 30001, Groningen,* <sup>3</sup> *Chemometric Research Group, University Centre for Pharmacy, A. Deusinglaan 2, 9713 AW Groningen, The Netherlands.*

*Clin Chem 1994;40(9):1692-1697*

## ABSTRACT

Infrared (IR) spectroscopy is used to analyze urinary calculus (renal stone) constituents. However, interpretation of IR spectra for quantifying urinary calculus constituents in mixtures is difficult, requiring expert knowledge by trained technicians. In our laboratory IR spectra of unknown calculi are compared with reference spectra by means of a computerized library search of 235 reference spectra from various mixtures of constituents in different proportions, followed by visual interpretation of band intensities for more precise semiquantitative determination of the composition. To minimize the need for this last step, we tested artificial neural network models for detecting the most frequently occurring compositions of urinary calculi. Using constrained mixture designs, we prepared various samples containing ammonium hydrogen urate, brushite, carbonate apatite, cystine, struvite, uric acid, weddellite, and whewellite for use as a training set. We assayed known artificial mixtures as well as selected patients' samples from which the semiquantitative compositions were determined by computerized library search followed by visual interpretation. Neural network analysis was more accurate than the library search and required less expert knowledge because careful visual inspection of the band intensities could be omitted. We conclude that neural networks are promising tools for routine quantification of urinary calculus compositions and for other related types of analyses in the clinical laboratory.

---

## INTRODUCTION

Fourier-transform infrared (IR) analysis is a reliable method for qualitative and quantitative assessment of urinary calculus composition of patients with urolithiasis (1). However, complex IR spectra from stone components in different combinations and proportions require expert interpretation. Different schemes have been developed for objective qualitative analysis of spectra of calculi (2;3). For semiquantitative prediction of components analyzed with IR, most spectrophotometers are accompanied by software packages for computer library searches. Quantification of calculus compositions analyzed by IR can be performed with methods such as factor analysis (4) and partial least-squares (PLS) regression (5). When the Lambert–Beer Law holds, PLS is a reliable method for analyzing spectroscopic data from the urinary calculus (5). However, PLS regression can be subject to serious nonlinearities, such as are sometimes observed in spectroscopic data from multicomponent mixtures. Lack of on-line data handling-facilities before the data are quantified by PLS regression restricts the practical use of this approach in routine laboratories.

Artificial neural networks, which are less subject to these limitations, are simple computer programs, for which the working concept is partly derived from the mechanism of action of the brain. The principal property of artificial neural networks is association and learning from examples with linear and nonlinear data structures; their power is pattern recognition and prediction. Recently, neural networks have been applied to evaluations of clinical laboratory data (6-9). Artificial neural networks consist of one input layer, one or more hidden layers, and one output layer (Fig. 1). Each layer consists of one or more elements (“neurons”) that are connected to all elements of the next layer. These neurons are simple computational devices that receive one or more input signals. Most neural network programs can change the network structure (topology) by varying the number of neurons in all or some of the defined layers, if desired, before training.

The network is trained by presenting it with an extensive set of training samples that contain both input values and their actual outcomes. Given a certain network topology, the learning or

training phase involves the modification of the connection strengths (weights) between the neurons of the succeeding layers. These weights are values representing the stimulating or inhibiting influence of an input signal. Initially the weight factors are randomly chosen. The weight adaptations during training are intended to reduce the error between the network outcome and the required outcome of the presented training signal. To obtain this error reduction, the artificial neurons of the hidden and the output layer compute a weighted sum of all their inputs; the outcomes of the weighted additions are passed to a differentiable transfer function, which is mostly sigmoid or linear.

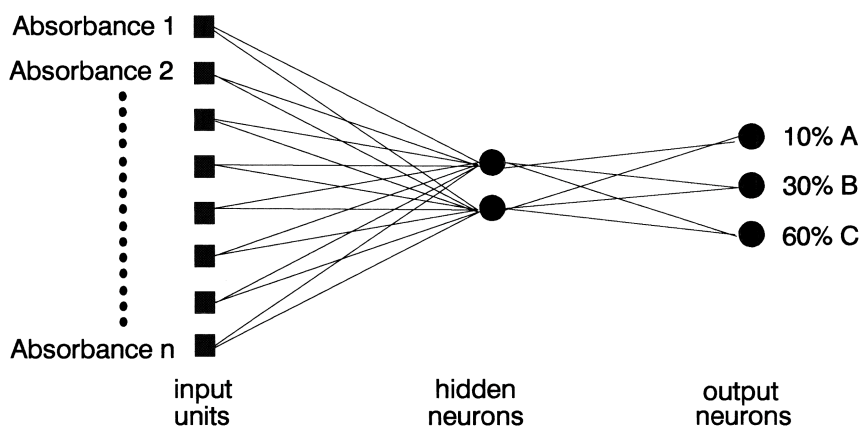


Figure 1. Simplified scheme of an artificial neural network with three layers for estimating the composition of a mixture containing components A, B, and C, from e.g. an absorption spectrum; in this case the absorbances are presented to the input layer.

After feeding the input data forward through the network during training, the results of all calculations at the output neurons are compared with the required outcome. If the difference between the required and calculated outcome is greater than a predefined tolerance, the weight factors are adjusted by means of the back-propagation learning rule (10). After all training data have been presented to the network in one pass, the process (feed forward, error calculation, and back propagation), is repeated until use of all the input data produces the desired output within the predefined tolerance. The procedure results in a combination of weight factors, which are used for a fast calculation of the unknown outcome. Before the network is used for prediction of unknown sample compositions, the network performance should be validated by analyzing a separate group of samples with known compositions that were not used for training.

In a sense, artificial neural networks extract their own models from the data. IR spectra from urinary calculi composed of several constituents sometimes have data that depart from the Lambert–Beer Law, which is based on a linear relation of the absorption with the concentration. Artificial neural networks are especially useful for the extraction of quantitative information from data with nonlinear structures. Therefore, we evaluated the use of an artificial neural network for quantifying eight commonly occurring components of urinary calculi.

## MATERIALS AND METHODS

### *Samples*

The network was trained on mixtures of ammonium hydrogen urate, calcium hydrogen phosphate dihydrate (brushite; Riedel-de-Haën, Hannover, Germany), carbonate apatite, cystine (Sigma, Brunswick Chemie, Amsterdam, The Netherlands), magnesium ammonium phosphate hexahydrate (struvite; Riedel-de-Haën), uric acid (Sigma), calcium oxalate monohydrate (weddellite) and calcium oxalate dihydrate (whewellite; BDH, Brunswick Chemie, Amsterdam, The Netherlands). Ammonium hydrogen urate, weddellite, and carbonate apatite were not commercially available, so we synthesized ammonium hydrogen urate (according to a protocol from Hesse A. Hesse, Germany) and weddellite (11) and obtained carbonate apatite from a collection of carefully selected patients' samples. The purity of the carbonate apatite was examined by comparing the IR spectrum with the spectrum from the Hesse atlas (3) and by standard chemical analysis.

For network training, we prepared 160 binary and ternary mixtures containing the eight most commonly occurring urinary calculus components. To test the quality of the neural network during training, we prepared an independent test set of 57 artificial mixtures of the same components as the training set.

For use in validating the quality of the network after training, an independent validation set was formed from 36 selected samples of patients' urinary calculi containing various compositions of the described components. The compositions of the patients' samples were estimated by comparing the spectra of the unknown calculi with reference spectra from a computer library, and independent visual inspection of the intensities of the bands by two specialized technicians. To test the deportment of the neural network for components not available in the training set, we included in the 36 patients' samples 3 that contained uric acid dihydrate, amorphous calcium phosphate, and chloramphenicol as an additional, or sole component.

Details of preparation of the mixtures, analytical procedures, and IR analysis are described elsewhere (5).

### *Data Preprocessing*

The recorded IR spectra, with wavenumbers of 4000–400  $\text{cm}^{-1}$  and a spectral resolution of 16  $\text{cm}^{-1}$ , were stored as ASCII data in a personal computer. Matrices were formed from all training, test, and patients' spectral data. Every row of a matrix contained the normalized spectral data for a continuous set of selected wavenumbers. Concentration data (percent of total weight) for the training and artificial test sets were added to the end of each row. Normalization (12) was done by unifying the total variance of every spectrum for the selected wavenumbers with regard to its area under the curve.

### *Neural Network Topology*

The neural network software package was written in Turbo C (Borland International, Scotts Valley, CA) with reference to a program listing (13). For quantitative analysis, the Chemometric Research Group of the University Centre of Pharmacy in Groningen extensively modified this program. An IBM compatible 80486DX (33 MHz) personal computer was used to run the program.

The training set was used to obtain a trained network, which has a mapping (topology) suitable for analysis of as many as eight components in unknown samples. The number of units of the input layer was equal to the number of selected wavenumbers ( $n = 91$ ). Eight output neurons



were used, one for each component used in the training set. Sigmoid neurons were used for the hidden layers, linear neurons for the output layer (13). There were no direct connections between the input and output layer. Input patterns were scaled between 0 and 1 for the absorbance data and between -1 and +1 for the concentration data.

If too much noise is present in the patterns, the model might overfit the data as a consequence of the nonlinear behavior of the sigmoid transfer function. In that case, the network will store the individual patterns instead of the important underlying features. With oversized networks, excessive training also will lead to overfitting of the training data. Therefore, to optimize (minimize) the number of hidden layers and their neurons, we monitored the outcome of the independent test set.

The absorbances of each spectral wavenumber were simultaneously presented to the units of the input layer, one spectrum at a time. At the same time, the corresponding target values were presented to the neurons of the output layer. After each training pattern was presented to the network in the feed-forward step this way, the differences between the predicted output and the correct or real values of the output were calculated. These differences were then used to update the weights in the back-propagation step of the training process. Differences between the required and predicted output were expressed as mean square error for the back-propagation learning rule (10;14;15). The whole process, in which all training patterns are presented to the network for updating the weights, is called one epoch or iteration. To allow better comparison with actual component concentrations, we used the square root of the mean square error (RMSE) as program output. The learning rates for back-propagation were chosen so as to provide the shortest training time, thereby preventing any oscillation (15). The momentum, another convergence-improving back-propagation parameter, was always set to zero.

#### *Statistical and Validation Methods*

A major goal of artificial neural network development is a trained network that generalizes well. Generalization is the capability of the neural network to recognize unknown patterns that are similar, but not identical, to the training input patterns. Current literature presents many modifications to back-propagation algorithms to enhance the generalizing capabilities of neural networks (8;15). We used plots with RMSE as a function of the number of epochs for both training and test data for monitoring the network performance during training. The test set was not exposed to the training process itself (adjustment of the weights), but rather was used to monitor training performance during training. The neural network has good generalizing capabilities when the RMSE decreases monotonically with the number of epochs, for both training and artificial test data. If the convergence of both curves was poor, a different network mapping was used. The choice of the final network topology was based on monitoring the network performance with the test patterns during training. The number of epochs that produced the minimum RMSE value of the artificial test data was used to predict the composition of the unknown samples. Generalization properties of the artificial neural network were also validated by means of scatter plots and correlation lines, based on actual and predicted values of the artificial test samples.

The predictive value of the network, after final training, was tested by comparative evaluation of the 36 patients' samples whose compositions had been estimated with library search. The predictive value of the network was also evaluated by visual comparison of the spectrum from a patient's sample with the spectrum of an artificial sample that had been prepared according to the neural network-predicted composition of the patient's sample.

## RESULTS

In a previous study (5) we found that intensities of characteristic bands from 4000–400  $\text{cm}^{-1}$  varied not only according to the specific composition but also to other factors. Consequently, we used only a continuous range of wavenumbers between 1936 and 400  $\text{cm}^{-1}$ , which contained the bands most important for component identification. Normalization was carried with this wavenumber range also.

The minimum network size was figured out first: Network reduction to one hidden layer gave no loss of learning ability. The number of neurons of the hidden layer was then reduced. Once the optimum was found, we varied the scaling factors for in- and output and the learning rate. Selection of the optimal topology was based on monitoring plots, with RMSE as a function of the number of epochs, for both training and test data during training. Criteria for selection were: a minimum divergence of both curves, and the lowest value for RMSE of the test data. The optimal network topology we obtained had eight hidden neurons, input scaling between 1.0 and 0.0, output scaling between 0.6 and 0.4, and a learning rate of 0.1. With this final network topology, 15000 epochs were needed to reach an RMSE of 1.842 for the training data and 3.471 for the artificial test data (Fig. 2).

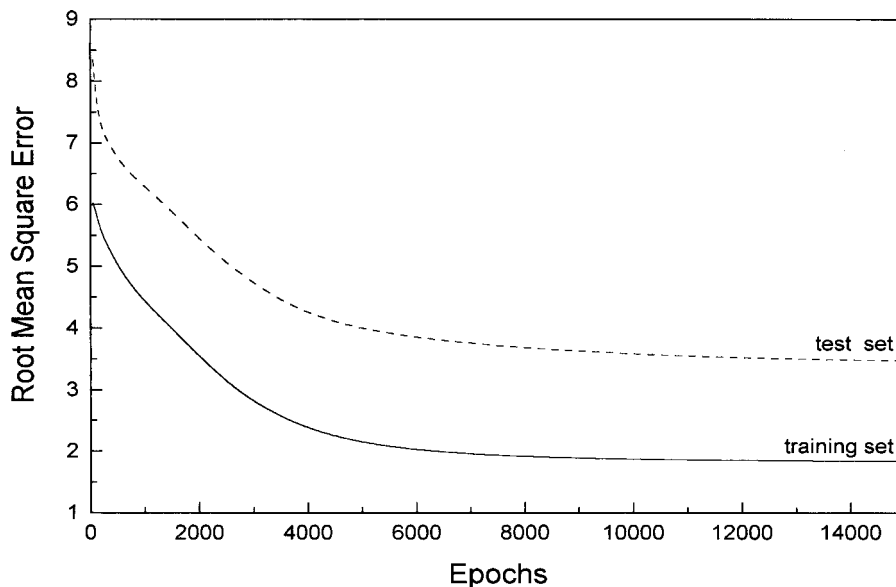


Figure 2. Determination of the minimum number of epochs for the training and test sets, predicted with the network.

Correlation between the actual and network-predicted composition of the artificial test samples used for monitoring the network performance during training is depicted in Fig. 3. The eight components in these samples were: ammonium hydrogen urate, brushite, carbonate apatite, cystine, struvite, uric acid, weddellite, and whewellite.

The majority of the predicted values of the artificial mixtures deviated <5% from the actual composition. Of 57 artificial test samples, 7 had predicted values deviating slightly >10% from the actual composition: 5 for weddellite and 2 for whewellite.

As shown in Fig. 4 the absorption spectra of a patient's sample (sample 33 in Table 1) and an artificial mixture with nearly the same composition as predicted by the network for the patient's sample (65:10:20:5 by wt. for ammonium hydrogen urate:uric acid:weddellite:whewellite) demonstrates good correspondence for wavenumber range 400–2000  $\text{cm}^{-1}$  (Fig. 4). The correlation coefficient between the two spectra was 0.991.

Table 1 shows close correspondence between the actual (estimated) and neural network-predicted compositions for most of the patients' samples used for validation of the network model after training. All samples had compositions that are commonly found in biological samples, although not at the frequency of occurrence listed in the Table. Retrospective data on IR analyses of urinary calculi compositions in our laboratory indicated that ~80% of all urinary calculi contained calcium. The vast majority of these calculi were binary and ternary mixtures composed of both calcium oxalates and (or) carbonate apatite.

The majority of the predicted values of the patients' mixtures show  $\leq 10\%$  deviation from the library search estimated composition. Sample 15 had different estimates for carbonate apatite and struvite, and sample 31 had different estimates for the calcium oxalates weddellite and whewellite. The “erroneously matched” specimen samples 34, 35, and 36 were added to the validation set to test the robustness of the network model for compositions that are wholly or partly absent from the training set. The neural network calculated 80% uric acid and 10% ammonium hydrogen urate for the uric acid/uric acid dihydrate containing calculus, and 90% carbonate apatite for the amorphous calcium phosphate/carbonate apatite-containing calculus. A serious mismatch was proposed for the chloramphenicol-containing calculus (sample 36).

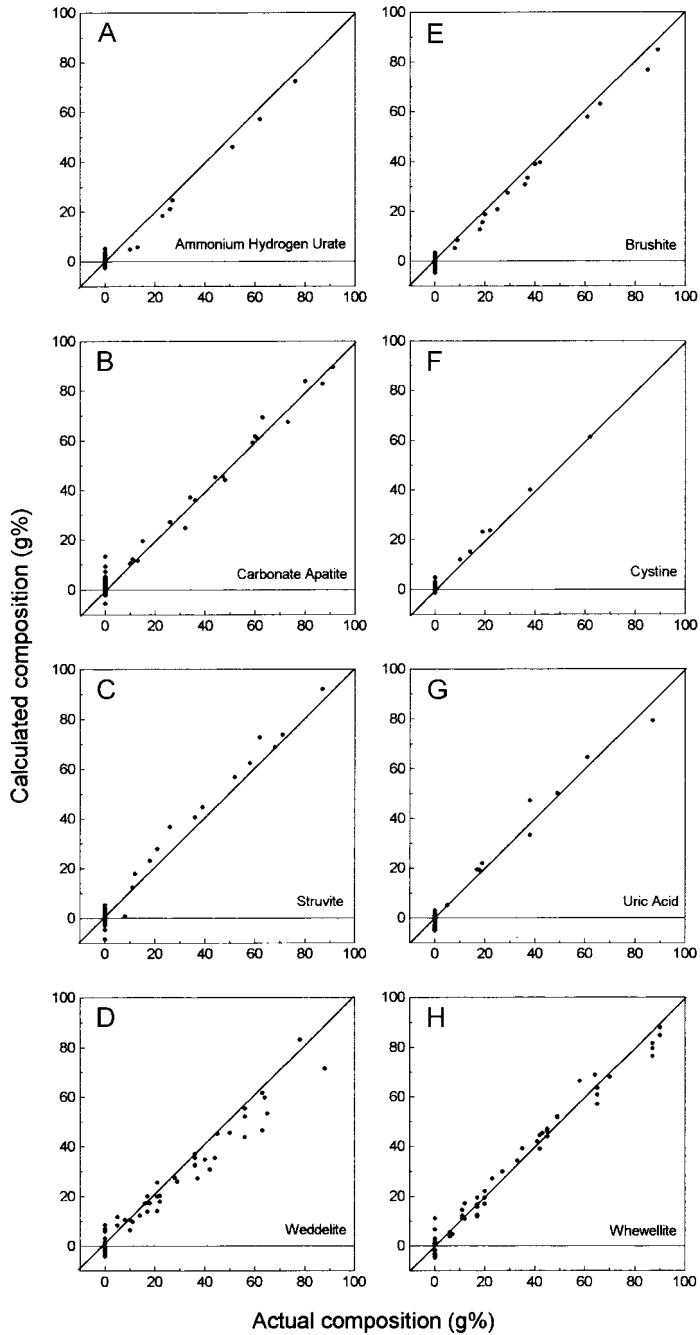


Figure 3. Validation plots for the test set, predicted by the network: A, ammonium hydrogen urate; B carbonate apatite; C, struvite; D, weddellite; E, brushite; F, cystine; G, uric acid; H, whewellite.

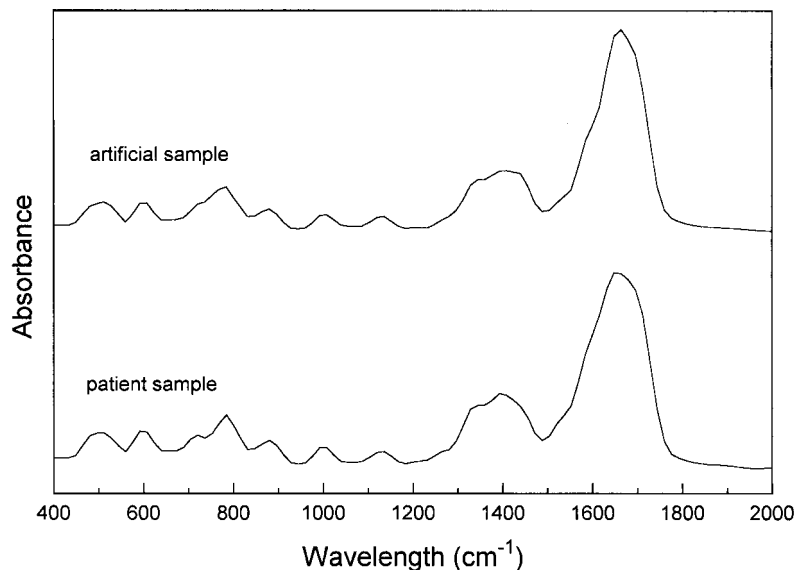


Figure 4. Absorption spectrum of a patient's sample (no. 33 in Table 1) and a spectrum of an artificial mixture with the same composition as predicted for the patient's sample.

The network prediction of the patient's sample was ammonium hydrogen urate:uric acid:weddellite:whewellite (65:10:20:5, by wt.).

## DISCUSSION

Computerized library searches are useful for routine determinations of the qualitative and global quantitative composition of urinary calculi. However, a visual inspection of the spectrum for more precise determination of the quantitative composition is also required. If the urinary calculus is composed of more than two components, interpretation of the proportions of band intensities by visual inspection of the spectrum becomes more complex. In a previous study (5) we found PLS regression to be reliable for quantifying a mixture of carbonate apatite, weddellite, and whewellite. For these components, almost equal errors were obtained with the PLS (RMSE = 1.7) and neural network model (RMSE = 1.6) for the same artificial test data.

Removal of wavenumbers that contributed no structural information was important for obtaining trained networks that could produce generalizable results for unknown samples, after neural network training. The criterion for a correct mapping is thought to be a similar performance on training and test data, i.e., the same order of RMSE and the same trend. If the resulting errors follow the same trend, one can safely assume that the underlying features of the patterns are learned and not merely specific features of the training set. Therefore, one can expect good predictions for new patterns. We found monotonically decreasing curves and minimum divergence between the curves of training and test data (Fig. 2). Consequently, we obtained generalizable results for the patients' validation data (Table 1) with the final network topology that was sufficiently trained after 15000 epochs.

The scatter plots in Fig. 3, representing the correlations between actual compositions and network-calculated compositions of the test samples, show that the network model yielded reliable results for these samples. Only 7 of 57 artificial mixtures had network-predicted values for whewellite or weddellite that deviated slightly >10% than the actual composition.

As shown in Table 1, the network-predicted compositions for the patients' samples deviated by slightly >10% from the actual composition, found by visual inspection of the spectrum, for only two samples (samples 15 and 31). Of course one should keep in mind that the "actual" compositions of patients' samples used for network validation are estimations also.

Table 1. Comparison of actual and network-predicted urinary calculus compositions for validation of the neural network model<sup>a</sup>

Sample	AMUR	BRUS	CARB	CYST	STRU	URIC	WEDD	WHEW
1	20, 10	80, 90						
2							20, 20	80, 80
3							85, 75	15, 25
4			100, 95				0, 5	
5			40, 40				60, 60	
6			70, 75				10, 10	20, 15
7			30, 30				40, 40	30, 30
8			65, 70				5, 5	30, 25
9			15, 15				20, 15	65, 75
10			5, 5				30, 20	65, 75
11			5, 5				15, 15	80, 80
12			60, 60				10, 10	30, 30
13			90, 95		5, 5			5, 0
14			80, 80		15, 15		5, 5	
15			85, 70		10, 25		5, 5	
16			80, 90		20, 10			
17			95, 95		5, 5			
18			50, 55		50, 45			
19			65, 75		35, 25			
20			55, 60		45, 40			
21			5, 5			80, 80	5, 10	10, 5
22						50, 65	20, 10	30, 25
23						30, 30	10, 5	60, 55
24		20, 20	80, 80					
25		90, 90	10, 10					
26		50, 45	50, 55					
27		30, 20	70, 80					
28		40, 40	50, 50				10, 10	
29		55, 55	35, 35				10, 10	
30		40, 40	45, 45				15, 15	
31		70, 65					0, 20	30, 15
32		40, 45		30, 40		30, 15		
33	65, 65					10, 10	15, 20	10, 5
34 <sup>b</sup>	10					80	10	
35			90				5	5
36			40	40			140	-120

<sup>a</sup> When two results are listed, the first is the composition estimated by library search plus visual inspection of the band intensities; the second is the network-predicted value.

<sup>b</sup> Sample 34 was composed of uric acid and uric acid dihydrate (50:50 by wt.), sample 35 of amorphous calcium phosphate and carbonate apatite (80:20 by wt.) and sample 36 of chloramphenicol (100%).

AMUR, ammonium hydrogen urate; BRUS, brushite; CARB, carbonate apatite; CYST, cystine; STRUV, struvite; URIC, uric acid; WEDD, weddellite; WHEW, whewellite.

The chloramphenicol-containing sample (no. 36) demonstrates that qualitative compositions unrelated to those in the training set can result in serious deviations from the actual compositions for some or all components. Its calculated contents — carbonate apatite, cystine, weddellite, and whewellite, 40:40:140(!): -120(!) by wt. — were nonsensical, despite the fact that they sum to 100%. For the patient's sample containing uric acid dihydrate (no. 34) the results were not so disparate (calculated contents of uric acid and ammonium hydrogen urate 80:20 by wt.), because of the spectral similarity of uric acid dihydrate to these components. The same is true for the amorphous calcium phosphate-containing sample (no. 35) because of its spectral similarity to carbonate apatite. We therefore conclude that prediction of compositions of unknown samples with artificial neural networks should be considered with great care when compositions are greatly dissimilar from the training set. Apparently, it is necessary to add some simple expert rules after completion of the neural network calculation of the unknown samples.

The results in Table 1 and the scatter plots in Fig. 3 show that artificial neural networks can be useful for routine determination of quantitative urinary calculus compositions. However, the training of artificial neural networks is hampered by the fact that many variables must be optimized manually to obtain networks that generalize well. The time for one complete training session can also be considerable. Training sessions for the network (i.e., 15000 epochs) were ~24 h. Some batch-type system for unattended operation, by which processing parameters are changed automatically after each training session, is strongly recommended. However, once reliable networks have been obtained, prediction of unknown sample compositions takes only a few seconds.

Almost all urinary calculi contain no more than three components. If the identity of the urinary calculus components is known beforehand, slightly better results can be obtained by using a network that has been trained only with the components of interest. However, for analysis of urinary calculi, the gain in accuracy obtained by using such networks over that for the network trained with the eight components is negligible (data not shown).

We conclude that prediction of quantitative urinary calculus compositions by artificial neural networks is of practical use in clinical laboratories. Other multivariate calibration methods, e.g., PLS, can be successfully applied (5), but these require different steps of data preprocessing. Consequently, these multivariate calibration methods are mostly available only in commercial programs with limited or no preprocessing possibilities. Therefore, we are now engaged in preparing software that will make available in one program data preprocessing, neural network processing, and some simple expert rules for the prediction of urinary calculus compositions analyzed with IR.

### Acknowledgements

We thank A. Hesse for providing information about the synthesis of some components.

### REFERENCES

1. Gault MH, Ahmed M, Kalra J, Senciall I, Cohen W, Churchill D. Comparison of infrared and wet chemical analysis of urinary tract calculi. *Clin Chim Acta* 1980;104:349-59.
2. Oliver LK, Sweet RV. A system of interpretation of infrared spectra of calculi for routine use in the clinical laboratory. *Clin Chim Acta* 1976;72:17-32.
3. Hesse A, Sanders G. Atlas of infrared spectra for the analysis of urinary concrements. Stuttgart: George Thieme Verlag, 1988.

4. Hesse A, Sanders G, Döring R, Oelichmann J. Infrarotspektroskopische Harnsteinanalyse: automatisierte Spectrenauswertung mit Hilfe der Factoranalyse. *Fresenius Z Anal Chem* 1988;330:372-3.
5. Volmer M, Bolck A, Wolthers BG, de Ruiter AJ, Doornbos DA, van der Slik W. Partial leastsquares regression for routine analysis of urinary calculus composition with Fourier transform analysis. *Clin Chem* 1993;39:948-54.
6. Schweiger CR, Soeregi G, Spitzauer S, Maenner G, Pohl AL. Evaluation of laboratory data by conventional statistics and by three types of neural networks. *Clin Chem* 1993;39:1966-71.
7. Schweiger CR, Söregi G, Pohl AL. Detection of monoclonal gammopathies in serum electrophoresis by neural networks. *Clin Chem* 1993;39:1984-5.
8. Astion ML, Wener MH, Thomas RG, Hunder GG, Bloch DA. Overtraining in neural networks that interpret clinical data. *Clin Chem* 1993;39:1998-2004.
9. Sharpe PK, Solberg HE, Rootwelt K, Yearworth M. Artificial neural networks in diagnosis of thyroid function from vitro laboratory tests. *Clin Chem* 1993;39:2248-53.
10. Rumelhart DE, Hinton GE, Williams RJ. Learning representation by backpropagating errors. *Nature* 1986;323:533-6.
11. Berg W, Hesse A, Schneider HJ. A contribution to the formation mechanism of calcium oxalate. III. On the role of magnesium in the formation of oxalate calculi. *Urol Res* 1976;4:161-7.
12. Martens H, Naes T. *Multivariate calibration*. Chichester, UK: John Wiley, 1991:377-41.
13. Wolf T. Neuronen im Computer. Perceptronnetze erkennen Muster. *Die Microcomputer Zeitschrift* 1990;4:92-108.
14. 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.
15. Bos A, Bos M, van der Linden WE. Artificial neural networks as tool for softmodelling in quantitative analytical chemistry: the prediction of the water content of cheese. *Anal Chim Acta* 1992;256:133-44.



## Chapter 6

### **Infrared analysis of urinary calculi, applying a single reflection accessory and a neural network interpretation algorithm**

Marcel Volmer<sup>1</sup>, Jules C.M. de Vries,<sup>2</sup> Henk M.J. Goldschmidt<sup>2</sup>

<sup>1</sup> *Department of Pathology and Laboratory Medicine, University Hospital Groningen (UHG),  
P.O. Box 30001, 9700 RB Groningen, The Netherlands,*

<sup>2</sup> *Department of Clinical Chemistry and Hematology (CKCHL),  
St. Elisabeth Hospital, P.O. Box 10111, 5000 JC Tilburg, The Netherlands.*

## ABSTRACT

*Background:* Preparation of KBr tablets, used for Fourier infrared (FT-IR) analysis of urinary calculus composition, is time-consuming and often hampered by pellet breakage. Therefore, we developed a new FT-IR method for urinary calculus analysis. This method makes use of a Golden Gate Single Reflection Diamond Attenuated Total Reflection sample holder, a computer library, and an artificial neural network (ANN) for spectral interpretation.

*Methods:* The library was prepared from 25 pure components and 236 binary and ternary mixtures of the 8 most commonly occurring components. The ANN was trained and validated with 248 similar mixtures and tested with 92 patient samples, respectively.

*Results:* The optimum ANN model yielded root mean square errors of 1.5% and 2.3% for the training and validation sets, respectively. Fourteen simple expert rules were added to correct systematic network inaccuracies. Results of 92 consecutive patient samples were compared with those of a FT-IR method with KBr tablets based on an initial computerized library search followed by visual inspection. The bias was significantly different from zero for brushite (-0.8%) and the concomitantly occurring whewellite (-2.8%) and weddellite (3.8%), but not for ammonium hydrogen urate (-0.1%), carbonate apatite (0.5%), cystine (0.0%), struvite (0.4%), and uric acid (-0.1%). The 95% level of agreement of all results amounted to 9%.

*Conclusions:* The new Golden Gate method is superior because of its smaller sample size, user-friendliness, robustness and speed. Expert knowledge for spectral interpretation is minimized by the combination of library search and ANN prediction, but visual inspection remains necessary.

---

## INTRODUCTION

Therapy to prevent urinary calculi recurrence requires quantitative estimates of the composition of urinary calculi. Extracorporeal shock wave lithotripsy, now widely used for removal of urinary calculi, necessitates the use of laboratory techniques that allow component identification in minute amounts of material. Traditional wet chemistry techniques, x-ray diffraction and infrared (IR) spectroscopy are the current analytical methods. Of these, wet chemical analysis is rather inaccurate and imprecise (1) and requires relatively large amounts of sample. X-ray diffraction is suitable for quantification of mineral containing samples, such as urinary calculi (2), but it cannot adequately detect amorphous substances, such as carbonate apatite or dahlite (3). Infrared (IR) spectroscopy has been applied in clinical chemistry for analyses of biofluids and solid biosamples (4). This technique often produces complex spectra with contributions from a sizeable number of unknown interfering substances when applied to authentic biological material. Analyses of these complex spectra is facilitated by the use of chemometrics (5), which is a generic term for the application of expert systems, neural networks, and other mathematical and statistical methods.

Fourier transform infrared (FT-IR) spectroscopy has become a standard technique for urinary calculus analysis. FT-IR makes use of a diversity of sample holders, such as photoacoustic detection (6), diffuse reflectance FT-IR (DRIFT) (7), and KBr tablet transmission (8;9). For the routine visual interpretation of urinary calculus IR spectra, Hesse et al. (10) have issued an atlas with IR spectra from pure urinary calculus

components and their mixtures, all embedded in KBr tablets. Another, less time-consuming option is computerized analyses based on library search [e.g. SEARCH (11) and LITHOS (2)], expert rules [CIRCOM (12), STONES (9)] or other chemometrical techniques like partial least-squares (PLS) regression (13) and artificial neural networks (ANNs) (14).

IR spectroscopy using KBr tablets is the current method for analysis of urinary calculus compositions in our laboratory. The preparation of KBr tablets is time-consuming and often hampered by pellet breakage. To overcome these drawbacks, we developed a new IR method, using a Golden Gate Single Reflection Diamond Attenuated Total Reflection (ATR) device. This method makes use of authentic sample material without any sample pretreatment. The results of the Golden Gate assay were quantified by a program dedicated for the prediction of the outcome of urinary calculus composition analyses. The new method was validated by comparing the results with those obtained with the IR assay with KBr tablets. The quantitative results from this KBr method were estimated from the IR spectra by the use of an initial computerized library search and followed by visual inspection of the spectra.

## MATERIALS AND METHODS

### *Samples*

#### Samples for the library of the Golden Gate NEURANET (GGN) method

For the construction of a library for the GGN method, we prepared IR spectra of 25, mostly commercially available, components (Table 1) and 236 mixtures. Usually, no more than three components can be detected in one patient sample. The majority of all urinary calculi contain one or more (maximum, three) of the eight most commonly occurring components. Therefore, the 236 mixtures were prepared from these commonly occurring urinary calculi components. The components are as follows: ammonium hydrogen urate (AMUR), brushite (BRUSH), carbonate apatite (CARB), cystine (CYST), stuvite (STRUV), uric acid (URIC), weddellite (WEDD), and whewellite (WHEW; Table 1). The mixtures were restricted to binary and ternary mixture designs. These, so-called constrained and balanced, mixtures were prepared in linear ranges of 0–100% with step sizes of 10%. A more detailed description of the preparation of these mixtures can be found elsewhere (13). AMUR was synthesized according to the following brief instructions: 1.68 g of uric acid was suspended in 500 mL water of 37 °C. A 30 g/L ammonia solution was added drop by drop (2 drops/s) until the solution was completely clear. After the solution cooled, the water layer was aspirated. The crystals were subsequently rinsed with water and diethyl ether and dried at 100 °C. With this synthesis, ~0.84 g AMUR can be obtained, which remains stable for ~6 months.

CARB and WEDD were obtained from patient samples by a selection based on purity. Purity was established by comparison of IR spectra with those in the Hesse atlas (10) and by standard wet chemical analysis. All mixtures were carefully mixed using a pestle and mortar.

#### Samples for training and validation of the neural network of the GGN method

For the development of the neural network, the previously mentioned 236 library mixtures and pure samples of the 8 commonly occurring components were used. Two additional mixtures were added to the set, giving a total number of 248 mixtures. The two extra mixtures were added to make the validation set more representative.

Patient samples for comparison of the KBR and GGN methods

One hundred consecutively collected urinary calculus samples from 70 males (median age, 56.5 years; range 5–75 years) and 30 females (median age, 49 years; range 21–74 years) served for testing the predictive performance of the new GGN method. The majority of them (>95%) were derived from patients treated with extracorporeal shock wave lithotripsy. Before analysis, the whole patient sample was carefully ground using a pestle and mortar. The quantitative compositions each sample was also obtained from the KBR method, using computerized library search followed by visual inspection of the spectrum. The samples were considered a representative selection of urinary calculi in our routine practice. The composition of the calculi will be described in greater detail in the *Results* section.

Table 1. Compounds used for the development of the GGN method

Component name	Abbreviation <sup>a,b</sup>	Source <sup>c</sup>
2,8-dihydroxy adenine		Sigma, Sigma-Aldrich Chemie (Zwijndrecht)
Albumin (human)		Patient
Aammonium hydrogen urate	AMUR	Synthesized
Amorphous calcium phosphate		Patient
Bilirubin		Merck, Fisher Scientific (Den Bosch)
Blood clot		Patient
Bovine albumin		Sigma, Sigma-Aldrich Chemie (Zwijndrecht)
Brushite	BRUS	Fluka Chemika, Sigma-Aldrich Chemie (Zwijndrecht)
Calcite		Fluka Chemika, Sigma-Aldrich Chemie (Zwijndrecht)
Carbonate apatite	CARB	Patient
Cholesterol		Merck, Fisher Scientific (Den Bosch)
Cystine	CYST	Sigma, Sigma-Aldrich Chemie (Zwijndrecht)
Fatty substance		Patient
Gypsum		Merck, Fisher Scientific (Den Bosch)
Hydroxyl apatite		Fluka Chemika, Sigma-Aldrich Chemie (Zwijndrecht)
Monetite		Merck, Fisher Scientific (Den Bosch)
Newberyite		Fluka Chemika, Sigma-Aldrich Chemie (Zwijndrecht)
Palmitic acid		Fisher Scientific (Den Bosch)
Quartz		Fluka Chemika, Sigma-Aldrich Chemie, Zwijndrecht)
Sodium-hydrogen urate		Sigma, Sigma-Aldrich Chemie (Zwijndrecht)
Struvite	STRUV	Riedel de Haën, Sigma-Aldrich Chemie (Zwijndrecht)
Uric acid	URIC	Fluka Chemika, Sigma-Aldrich Chemie (Zwijndrecht)
Weddellite	WEDD	Patient
Whewellite	WHEW	Fluka Chemika, Sigma-Aldrich Chemie (Zwijndrecht)
Xantine		Fluka Chemika, Sigma-Aldrich Chemie (Zwijndrecht)

<sup>a</sup> Components with an abbreviation are used for neural network processing.

<sup>b</sup> The eight most commonly occurring components

<sup>c</sup> All source located in the Netherlands

## *Analytical Methods*

### Standardization

Before each series, we validated the FT-IR instrument by measurement of a polystyrene transmission standard. Validation comprised wave number positions and absorbancies of known IR bands. The linearity of IR analyses of the KBr and Golden Gate assays was tested using a dilution series of URIC at concentrations of 0–100% with step sizes of 10%. URIC was diluted with WHEW, and the area of the URIC band at  $1120\text{ cm}^{-1}$  served for establishment of test linearity. The “runs-test” was used for establishment of significant deviations from a straight-line (15).  $P \leq 0.05$  was considered statistically different.

### KBr method

Pulverized urinary calculus (1.5 mg) was mixed with 180 mg KBr with a pestle and mortar. From this mixture, 100 mg was taken for the preparation of a urinary calculus KBr-tablet at  $10^9$  Pa pressure under vacuum for 2 minutes. A more extensive description can be found in the Hesse atlas (10). The spectra were scanned in the mid-infrared region from  $4000\text{--}400\text{ cm}^{-1}$  at  $4\text{ cm}^{-1}$  wave number intervals in a Bio-Rad FTS 135 spectrometer equipped with a cooled DTGS detector and Win-IR (Ver. 3.04) software (both from Bio-Rad Laboratories Inc., Spectroscopic Division). A 100- mg KBr-tablet was used as a blank for background subtraction. Samples producing weak spectra (absolute difference between absorbance maximum and minimum less than  $A=0.25$ ) were reanalyzed using tablets with higher sample:KBr ratios.

The quantitative composition of each sample was estimated by comparison of the recorded spectra with KBr reference spectra that were stored in a computer library (LITHOS; Bio-Rad). This library contains data of pure components of urinary calculi and 227 mixtures. Its content was similar to a LITHOS library that is used for x-ray diffraction. Win-IR search (Ver. 1.03; Bio-Rad Laboratories Inc., Sadtler Division) served as search engine. This search engine applies the Euclidean distance-matching algorithm to the finger print area ( $2000\text{--}400\text{ cm}^{-1}$ ) of the absorbance spectra to obtain a spectral hit list. Additional evaluation and interpolation led to an estimate of the quantitative composition of a sample, since the first hit is not necessarily the correct one and because even large libraries cannot contain full detail. After the library search, the final composition was obtained by visual inspection of spectral band intensities by two experienced technicians blinded to the results of the Golden Gate method.

### General outline of the GGN method

The so-called Golden Gate is a sample-holding device equipped with a Single Reflection Diamond ATR crystal (Graseby Specac) for measurement of micro samples. The standard ZnSe lens was replaced with a KSR5 lens to enable measurements between  $600\text{--}250\text{ cm}^{-1}$ . Carefully pulverized material (1-2 mg) was applied to the flat surface of the diamond crystal and pressurized at  $3 \times 10^8$  Pa. The reproducibility of this pressure was guaranteed by using the build-in pressure restraint of the pressure applicator of the Golden Gate device. The active sampling area of the crystal was  $1.13\text{ mm}^2$  (diameter 0.6 mm). The uniformity of the crystal spreading on the sensing area was controlled by viewing through the looking glass of the pressure applicator of the Golden Gate device. The samples were always measured at room temperature. A Bio-Rad FTS 135 spectrometer, equipped with a cooled DTGS detector and Win-IR software, was used for scanning in the mid-infrared region from  $4000\text{--}400\text{ cm}^{-1}$  at  $4\text{ cm}^{-1}$  wave number intervals. An empty crystal served for background measurement and blank subtraction. The background spectra were always

collected before a series of 10 sample spectra. All training, validation, and test samples were measured in a more or less random order over ~6 months. Each spectrum was acquired by coaddition and averaging of 16 interferograms. The Golden Gate crystal was cleaned with water and 960 mL/L alcohol after each measurement.

The NEURANET program (Ver. 3.0; Bio-Rad Laboratories Inc., Spectroscopic Division) was used for quantification of urinary calculi, whose compositions are expressed as mass percentage. The selection of this program was based on earlier studies (13;14). This program contains two supplementary quantification methods and was particularly developed for interpretation of IR spectra of urinary calculi in the range of 4000–400  $\text{cm}^{-1}$ . The first method is computerized library search, and the second is based on artificial neural network prediction. Library search can be used for quantification of any composition of a calculus, assuming that the components are available at the library. The neural network may be used for more accurate predictions of compositions of urinary calculi, but is restricted to process the outcomes of a maximum of 10 components simultaneously. Therefore, the neural network can only be used for quantification of the most commonly occurring components of urinary calculi. For calculi composed of these commonly occurring components, both quantification methods should provide almost the same outcome. In this case, the library search serves as a verification method of the ANN because it can depict the unknown spectrum graphically together with a number of the library spectra (stacked, or overlaid). For rarely occurring components, the results obtained with library search must be used. The availability of both quantification methods facilitates the interpretation process. Additionally, the program offers the possibility of adding some simple expert rules to the network-predicted results. These rules may be added to solve problems caused by small but systematic inaccuracies in the network outcome of patient samples. Furthermore, they are used to give an indication that the results from library search should be used in case of rare components unavailable to the network model and round the network outcome of each component to the nearest 5%. The expert rules aim a generalization of the quantitative results of future patient samples. The rules may be defined as simple “Basic” like “IF ... THEN ... ELSE ...” statements.

The neural network engine of the NEURANET program is based on a back-propagation neural network (16). This program contains a three-layer network, consisting of an input layer with a number of nodes (neurons) equal to the number of input variables (absorbances at different wave numbers), a single hidden layer with a variable number of nodes, and an output layer with a number of nodes equal to the number of components (maximum, 10). The input nodes are connected to the output nodes via the nodes of the hidden layer. All nodes of the hidden layer have every possible connection with the input and the output nodes. Each connection carries the signal and an individual weight. The final weights in all connections reflect the knowledge of the underlying spectral patterns. The complex of weights may be interpreted as the regression coefficient in a regression analysis. The multiple inputs of a spectrum are converted to a single concentration of a single component. The final set of weights is found by learning by back-propagation. With this method, the neural network is provided with a set of training spectra (samples) with known concentrations and iterates around a loop in which it predicts for each sample the analyte concentrations and compares these to the known concentrations (forward step). Depending on the differences between the calculated and the known outcome concentrations, the weight values will be readjusted (backward step). This happens for each sample of the training set in turn and is repeated many times over the complete data set. The number of

iterations (epochs) is usually very large. Before training, the starting weights are randomized between  $-0.1$  and  $0.1$ .

The performance of the network is monitored by looking at the root mean square error (RMSE). The RMSE is calculated by first taking the sum of squared differences between the desired and obtained output values of the training set. The square root is taken from the average of the sum of squares, which are averaged by the number of outcomes (maximum of 10) and the number of training samples. NEURANET contains several parameters for data preprocessing (e.g., selection of wave number ranges and scaling) and network design (topology). The neural network parameters have to be tuned by means of an independent validation set. This validation set is a representative set of spectra from samples of known composition and is used for testing the performance of the network but not for training. The training behavior of the network is monitored by looking at a graph depicting the decrease and convergences of the RMSE of both the training and the validation sets against the number of epochs. In worse cases, both RMSE curves will diverge instead of converge. To prevent overfitting of the neural network model, the training process is stopped when the RMSE of the validation set is at its minimum value (early stopping rule). In addition to training and validation (used for tuning), the performance of the network should be tested with a test set. More information about neural network processing can be found elsewhere (17;18).

The NEURANET program enables building of one or more named methods, based on spectroscopic absorbance data. Each method contains a combination of standard information (e.g., description of the components), a spectral library, a trained neural network model, and some expert rules. After selection of a method and of a spectrum from the file list, the program automatically performs a library search, network predictions, and expert rule filtering for prediction of the outcome of a sample with an unknown composition. The whole combination of the analysis with the Golden Gate sampling device and the final NEURANET model is called the GGN method.

#### *Development of the library and neural network of the GGN method*

The 261 samples, composed of 25 pure components and 236 mixtures, were analyzed with the Golden Gate sampling device, and their spectra were added to the library of the NEURANET program. They were stored at  $16\text{ cm}^{-1}$  resolution intervals of the  $4000\text{--}400\text{ cm}^{-1}$  analysis range. A spectral range was defined for searching in the  $3700\text{--}450\text{ cm}^{-1}$  range with the correlation-matching algorithm. The spectra of 248 pure components and mixtures of AMUR, BRUSH, CARB, CYST, STRUV, URIC, WEDD, and WHEW were recorded with the Golden Gate device. Of these, 199 were used as a training set for the neural network. This training set served for the construction of a network model that has a mapping (topology) suitable for the analysis these eight components (eight output neurons) in the unknown samples. The remaining 49 spectra were used as a validation set. A spectral range from  $1840$  to  $448\text{ cm}^{-1}$  with  $16\text{ cm}^{-1}$  resolution intervals was selected for neural network processing, giving rise to 88 input neurons. The network topology parameters for training the neural network were optimized by monitoring the RMSE of the validation set. The RMSE values for both the training and the validation sets were graphically depicted to check potential overfitting of the neural network model. A more extensive description of network processing in relation to urinary calculus analysis can be found elsewhere (14). A few expert rules were added to the network-predicted data for further optimization of the results of unknown samples. These rules were added as a result of small but structural differences in composition found between visual inspection of the Golden Gate spectra by

two technicians and the network predicted-results of the patient samples (test set). The rules we added without any foreknowledge of the results from the KBr method.

#### *Data processing and statistics*

The results of the KBr and GGN methods were compared using Altman-Bland agreement plots (19) for the 8 commonly occurring components and a combination of these. With these agreement plots, the individual differences ( $AMUR\%[KBr]_n - AMUR\%[GGN]_n$ ,  $CYST\%[KBr]_n - CYST\%[GGN]_n$  of sample  $n$ , and so forth) of both methods are calculated and plotted against the individual mean results (e.g., mean of  $AMUR\%[KBr]_n$  and  $AMUR\%[GGN]_n$  of sample  $n$ ) of both methods. The bias (mean of the individual differences between the GGN and KBr methods) and the 95% agreement limits (1.96 SD of the differences between both methods) are summarized in Table 3. The bias and 95% agreement level were calculated for the eight components separately and for a combination of them by taking the individual calculated differences of the eight components together. The bias and 95% agreement limits of the combination of the eight components are shown in an agreement plot in Fig.3.

## RESULTS

A KBr transmission spectrum of a patient sample consisting of 60% BRUS and 40% WEDD is shown in Fig 1A. Fig. 1B shows a transmission spectrum of the same sample obtained with the Golden Gate device. Analysis of the URIC dilution series with the run-test showed that the IR intensities of both the KBr and GGN assays were linear at 0–100%.

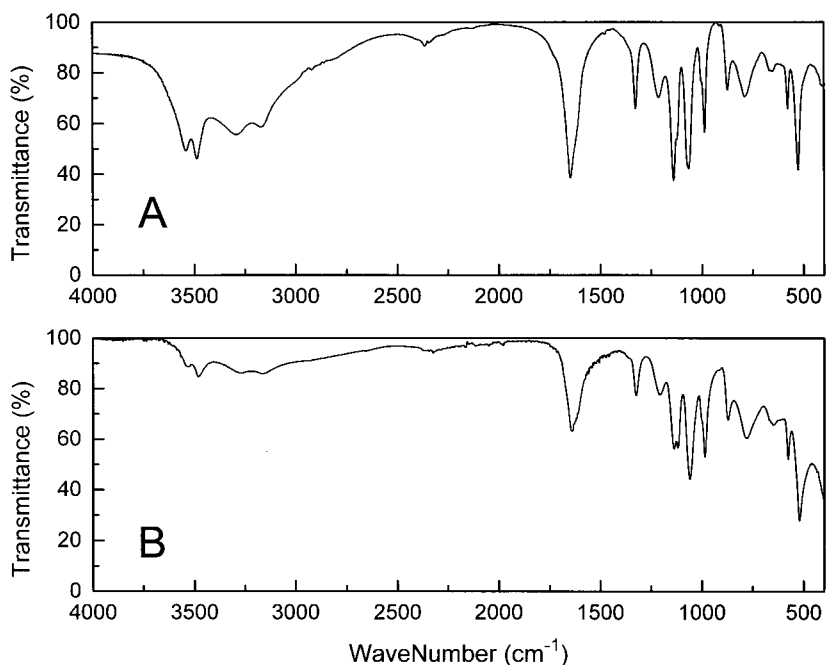


Figure 1. IR transmission spectra of a urinary calculus containing 60% BRUS and 40% WEDD obtained by the KBr method (A) and the Golden Gate device (B).



### *Compositions of urinary calculi with the KBr method*

Two of the 100 analyzed patient samples produced weak spectra. They were removed from the data set because of insufficient sample material for reanalysis. The quantitative compositions of the remaining 98 patient samples, as analyzed with the KBr method, revealed that 92 of them contained at least one of the 8 commonly occurring components. All 92 patient samples were single components or binary or ternary mixtures of one of these 8 components. The majority of them contained calcium oxalate (WHEW and WEDD) and/or CARB. The detection frequency of each of the eight components in the 92 samples was: AMUR, 1.1%; CYST, 1.1%; URIC, 3.3%; STRUV, 4.3%; BRUS, 13%; CARB, 48%; WEDD, 70.7%; and WHEW, 75%. The percentage urinary calculi that contained one, two, or three of these components were 14%, 54%, and 32%, respectively. Six of the 98 urinary calculi contained less frequently occurring components. One consisted of quartz, two of uric acid dihydrate and three of a fatty substance. Two of the latter were highly similar to feces, while the other was similar to palmitic acid.

### *Development of the GGN method*

#### Neural network

After repeated, batch-automated training of the neural network with different topologies, a final topology was found. Each training session took ~20 min for each topology. The final topology had eight hidden neurons. With this topology the RMSE steadily decreased to a minimum value of 2.3% for the validation-set (Fig. 2). This minimum value was reached after a training of 54000 epochs (cycles). At this number of epochs, the error of the training set was 1.5%.

#### Intermediate analysis of composition of urinary calculi

The 98 patient samples, which were also analyzed by the KBr method, were analyzed with the Golden Gate device. The composition of each calculus was estimated with library searches and neural network prediction of the intermediate GGN method. Computerized estimation of the composition of a single sample with the GGN method was obtained within ~1 s by means of simultaneous library search and network prediction followed by expert-rule filtering. The compositions of six of the 98 patient samples could not be estimated by the ANN because these samples did not contain any of the eight commonly occurring components available in the network model. Four of these samples could be detected with library searches of the GGN method in the first hit (searched from 3700 to 450  $\text{cm}^{-1}$ ). As with the KBr method, one was found to contain quartz, two contained feces and one contained a component similar to palmitic acid. The composition of the remaining two, which contained uric acid dihydrate according to the KBr method, could not be resolved by library search with the GGN method, since the spectrum of uric acid dihydrate was not available in this library.

#### Addition of expert rules

As a result of visual inspection of the Golden Gate spectra and the outcome of the neural network predictions of the 92 patient samples (test set), 14 simple expert rules were added to the GGN method (Table 2). After addition of all expert rules to the GGN method, the composition of the 92 patient samples, as analyzed with the Golden Gate device, was reestimated with the final GGN method. The results of the final GGN method were used for comparison with those of the KBr method.

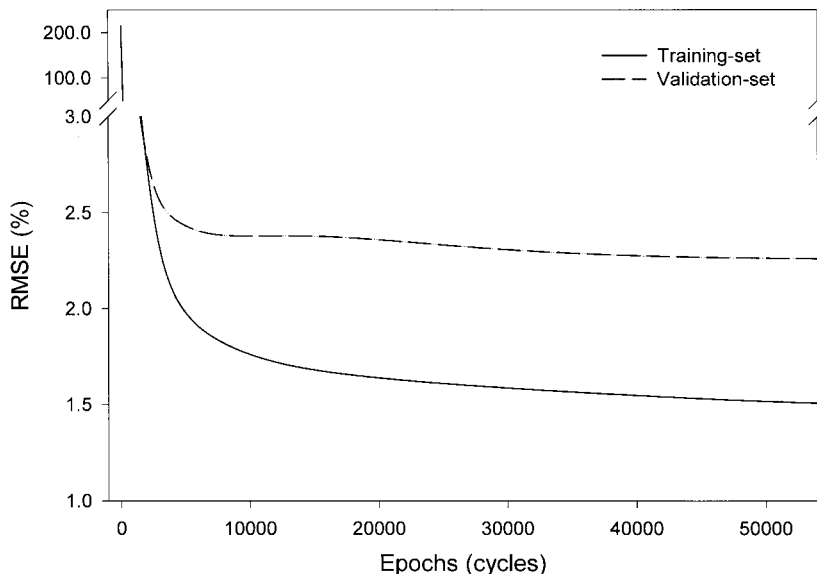


Figure 2. RMSE of both the training and the validation sets plotted against the number of network iterations (epochs). The minimum RMSE (2.3%) for the validation set was reached at 54000 epochs. At this number of epochs, a RMSE of 1.5% was found for the training set.

#### *Comparison of the KBr method and the GGN methods*

The agreement between the KBr and GGN methods, as obtained from the Altman-Bland plots of the results of 92 patient samples is shown in Table 3. The Altman-Bland plot of all patient results is shown in Fig. 3. The plot compares the results of all eight components obtained from the 92 patient samples, analyzed with the KBr and the GGN methods. The dashed lines express the 95% confidence interval of the differences between both methods. Of the 92 samples, 2 consisting of WHEW + WEDD and 1 consisting of CARB + STRU showed 20% difference between both methods (see Fig. 3). Because each of the three samples were mixed stones composed of two concomitantly occurring components, an increased amount (percentage) of one component relative to the other sample produced an equal decrease of the amount of the other component. Therefore, the Altman-Bland plot shows six data points at 20 % difference between the methods. For example, the sample containing CARB and STRU was composed of 60% CARB and 40% STRU measured with the KBr method, whereas it contained 80% CARB and 20% STRU measured with the Golden Gate method. The resulting data points (x,y) of the Altman-Bland plot are (70,20) and (30,20).

## **DISCUSSION**

We describe a new IR method for the analysis of the composition of urinary calculi. This method makes use of a Golden Gate Single Reflection Diamond ATR device and a newly developed library and neural network model for quantification. The outcome was compared with that of an IR method with KBr tablets.

Visual inspection of the IR spectra made clear that KBr spectra (Fig. 1A) have more definite bands than those recorded with the Golden Gate ATR device (Fig. 1B). The Golden

Gate ATR device also yields spectra with less absorbance intensities at higher wave numbers when compared to the traditional transmission spectra (Fig. 1A). The underlying cause is different sample radiation penetration depths at different wave numbers. It did not influence the interpretation of the spectra obtained with the Golden Gate assay, probably because of the predominant use of the bands with sufficient spectral definition in the 2000–400  $\text{cm}^{-1}$  region (fingerprint area).

Table 2. Expert rules added to the GGN method.

Name	Expert rule
cystcheck1	IF Cyst<10 THEN Cyst=0;
uriccheck	IF [AmUr+ Uric]<7.5 THEN AmUr=0 AND Uric=0 AND Normalize,;
amurcheck1	IF (Uric < 7.5) AND (AmUr > 12.5) THEN AmUr=[Uric + AmUr] AND Uric=0 AND Normalize;
amurcheck2	IF (AmUr < 7.5)AND (Uric > 12.5) THEN Uric=[Uric + AmUr]AND AmUr=0 AND Normalize;
whewwed1	IF Wedd<0 THEN WEDD=0 AND Whew=[Wedd+Whew];
whewwed2	IF Whew<0 THEN Whew=0 AND Wedd=[Wedd+Whew];
whewwed3	IF ((Wedd<3.5) AND (Whew<3.5)) AND ((Wedd>0) AND (Whew>0)) THEN IF Wedd>Whew THEN Wedd=[Wedd+Whew] AND Whew=0 ELSE ,Whew=[Whew+Wedd] AND Wedd=0 AND Normalize,;
carbcheck1	IF (Carb>35) AND ((Brus<10) AND (Stuv<10)) THEN Carb=[Brus+Carb+Stuv] AND Brus=0 AND Stuv=0;
carbcheck2	IF (Carb>35) AND ((Brus>10) AND (Stuv<10)) THEN Carb=[Carb+Stuv] AND Stuv=0 AND Normalize;
amurcheck3	IF (([AmUr+Uric]>4) AND ([AmUr+Uric]<30)) AND ([Brus+Carb+Stuv]>50) THEN AmUr=0 AND Uric=0;
amurcheck4	IF (AmUr>25) AND ([Brus+Carb+Stuv]<30) THEN AmUr=[AmUr+Brus+Carb+Stuv] AND Brus=0 AND Carb=0 AND Stuv=0;
amurcheck5	IF ([AmUr+Uric+Stuv]<10) AND (Carb>10) THEN Carb=[Carb+AmUr+Uric+Stuv] AND AmUr=0 AND Uric=0 AND Stuv=0;
amurcheck6	IF ([AmUr+Uric+Stuv]<10) AND (wedd>10) THEN Wedd=[Wedd+AmUr+Uric+Stuv] AND AmUr=0 AND Uric=0 AND Stuv=0;
amurcheck7	IF ([AmUr+Uric+Stuv]<10) AND (Whew>10) THEN Whew=[Whew+AmUr+Uric+Stuv] AND AmUr=0 AND Uric=0 AND Stuv=0;

Table 3. Comparison of the results of the KBr and GGN methods for 92 patient samples.

	Number	Bias (95% confidence Interval), %	95% Level of agreement, %
AMUR	92	-0.1% (-0.3 to 0.1)	2.0%
BRUS	92	-0.8% (-1.5 to -0.2) <sup>a</sup>	6.2%
CARB	92	0.5% (-0.7 to 1.6)	10.8%
CYST	92	0.0% (0.0 to 0.0)	0.0%
STRU	92	0.4% (-0.9 to 0.1)	4.7%
URIC	92	-0.1% (-0.6 to 0.4)	4.4%
WEDD	92	3.8% (2.5 to 5.2) <sup>a</sup>	12.8%
WHEW	92	-2.8% (-4.3 to -1.3) <sup>a</sup>	14.2%
ALL	736	0.0% (-0.3 to 0.3)	9.0%

<sup>a</sup> Significant bias

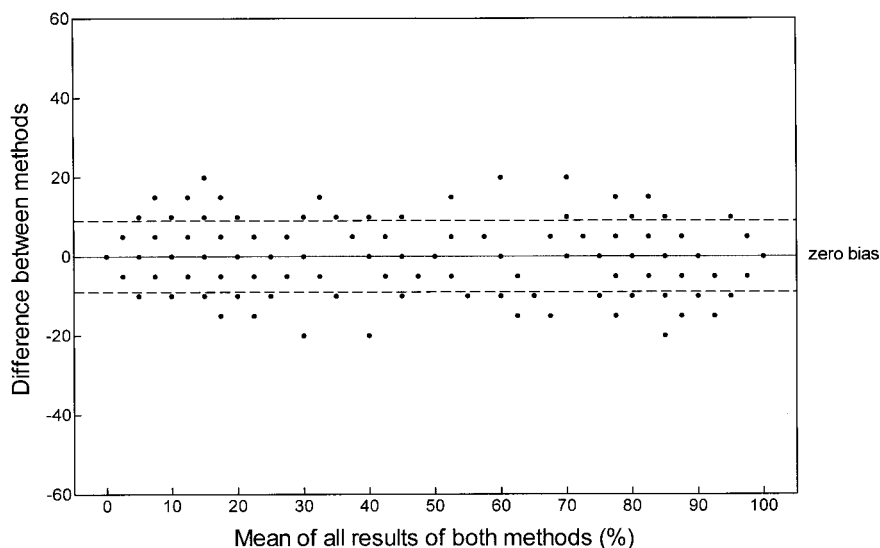


Figure 3. Bland-Altman method agreement plot of the results of eight components in 92 urinary calculi as analyzed with the KBr method and the GGN method. The dashed lines are the 95% confidence interval of the differences between both methods.

#### *Development of the neural network model*

From a previous study (14), we found that the compositions of urinary calculi using neural network prediction were similar to the results obtained with PLS regression. In both cases, the urinary calculi were analyzed with IR spectroscopy, using KBr tablets. The PLS regression model and the neural network model were developed for quantification of

mixtures of WHEW, WEDD and CARB, whose incidence rate in urinary stones is ~80% in Western countries (20). The RMSE values of the validation sets were 1.7% for PLS regression and 1.6% for network prediction, respectively. The previous study also described the development of a neural network model trained with the eight most commonly occurring components. This neural network model has been used successfully for several years in our laboratory (UHG). On the basis of these findings, we concluded that for quantification of the composition of urinary calculi, neural networks would be useful as or better than linear models such as PLS regression. We therefore developed a new neural network model with the NEURANET program after replacing the KBr sampling device by the Golden Gate accessory.

Training of the ANN went remarkably well, despite the previously mentioned relatively poor spectral band definition in the fingerprint area compared with the KBr method. Although the NEURANET program contains several facilities to make network training rather simple, there are several difficulties in applying ANN models. The ANN parameters (topology) are often difficult to estimate, and large training sets are often needed. The number of training samples should be at least more than the number of input units (88 in our case), but the required number of training samples also strongly depends on the noise level in the targets and the complexity of the adaptation of the network to the target function. The absorbance values of IR spectra are linearly related to the concentration (Lambert–Beer law). Therefore, a relatively small number of learning samples ( $n=199$ ) was needed for training. We gave special attention to the risk of overtraining (overfitting). In this case the network loses generalization (robustness) and will adapt (learn) to unimportant spectral features, such as noise. Overfitting (and underfitting) was monitored by looking at the RMSE errors of the validation and training sets (Fig. 2). If the validation error became much higher than training error, the network was probably overfitted and another topology was applied. In addition, the robustness of the neural network model is important and can be tested by retraining the neural network with different initial weights each training session, providing that the other conditions are left unchanged (e.g., topology settings, training and validation sets). If the RMSE values at a certain number of epochs show large differences for the different training sessions, the network model can be considered to be unstable. The final criterion for assessing the network model was the comparison of the results from the independent test set with the results obtained with the reference method.

Several heuristics exist for the choice of the starting values of many of the topology parameters. However, since the optimal parameter settings strongly depend on the nature of the problem and on the chosen representation of the input and output objects, it is not safe to rely exclusively on heuristics. Therefore, an operator must have sufficient knowledge of network training to select the topology parameters (e.g., number of hidden neurons) and must interpret the numerical and graphical network outcome. Using NEURANET, consecutive unattended training with different topologies was possible when a number of topologies were set out in advance. After training, estimation of compositions of patient samples is very fast when the stored ANN method is used.

More detailed information about the theoretical background of ANNs is out of the scope of this report, but can be found elsewhere (17;18).

#### *Development of the expert rules*

Urinary calculi are always composed of pure components or binary or ternary mixtures. For those components absent in the sample (e.g., five absent in case of a three-component

calculus), small positive or negative numbers may occur in the network outcome of patient samples on a regular base (Table 4). This is a consequence of network training, which will always predict the outcome of the eight components simultaneously, whether or not they are present in the sample. Because the total outcome of any sample is always 100% (Table 4), a small outcome (percentage) may occur for those components absent in the sample. These inaccuracies may occur because no prediction is perfect, the samples may contain trace amounts of impurities caused by their passage through the urinary tract, and some samples produce rather noisy spectra. In this last case, the network assumes detection of small amounts of a component characterized by a great number of spectral bands (e.g., AMUR). Therefore, a few expert rules were defined (see Table 2). In essence, the expert rules can be considered as automated corrections of the network outcome, which otherwise would have been made manually by expert technicians after visual inspection of the IR spectra. Some of these rules are counterparts of each other, describing almost the same type of correction (e.g., amurcheck1 and amurcheck2). Another rule, named whewwedd3, seems to be rather complex, but only assigns the smallest oxalate outcome (Whew or Wedd) to the largest one, providing that both oxalate outcomes are positive and <3.5%. The rationale is not a physical/chemical one, but only a small correction. If this rule was not applied, both oxalate outcomes would be forced to zero by rounding and normalization, in spite of the fact that a small amount of oxalate is present in the sample.

Table 4. Network- and expert system-predicted outcomes for samples A<sup>a</sup> and B<sup>b,c</sup>.

	AMUR	BRUS	CARB	CYST	STRU	URIC	WEDD	WHEW
<b>Patient A<sup>d</sup></b>								
Network	-3%	46%	44%	-2%	0%	-0%	13%	2%
Expert		45%	40%				15%	
<b>Patient B<sup>e</sup></b>								
Network	0%	2%	94%	0%	1%	-1%	-5%	9%
Expert			95%					5%

<sup>a</sup> Sample A contains BRUS, CARB, and WEDD.

<sup>b</sup> Sample B contains WHEW and CARB.

<sup>c</sup> Both samples were analyzed with the Golden Gate assay.

<sup>d</sup> Library first hit: WEDD-BRUS 10%-90%; second hit: BRUS-CARB 50%-50%.

<sup>e</sup> Library first hit: WHEW-CARB 10%-90%; second hit: WEDD-CARB 10%-90%.

Except for normalization of the network outcome to the nearest 5%, no special expert rule was applied to patient sample A (Table 4). Sample B shows somewhat inaccurate results of both calcium oxalates (WHEW and WEDD). WHEW turned out to be predominant relative to WEDD. Because this has happened several times, an expert rule was defined by simply adding the values of both calcium oxalates and forcing the value of WEDD to zero. This rule was defined as follows:

IF wedd<0 THEN whew=[wedd+whew] AND wedd=0;

This rule caused 0% WEDD and 4% WHEW. After normalization of sample B the final composition of the expert system was 5% WHEW and 95% CARB (Table 4). This outcome reflected the real composition of this sample, which was based on careful visual inspection of the band intensities of the KBr spectrum by a trained technician.

#### *Method comparison*

The 92 consecutive samples used for method comparison were regarded to be a representative selection of urinary calculi in our daily practice. They had similar frequency distribution of components and number of components per sample, compared with historical data (not shown). X-ray diffraction is occasionally recommended as a reference method for urinary calculus analyses. However, x-ray diffraction cannot adequately detect amorphous substances (3). CARB is, for example, sometimes overlooked, but can be detected by a simple CO<sub>2</sub> test following acidification with HCl. Quantitative analysis of CARB may, however, be difficult. We therefore decided to compare the GGN results in both an analytical and a managerial sense with those obtained by an IR method with KBr tablets. This method was routinely used at the time of the study and was to be substituted with a less time-consuming and more robust analytical method.

The bias of the outcome of the KBr and GGN methods of the 92 patients' samples was significantly different from zero for BRUS, WHEW, and WEDD (Table 3). The small bias of BRUS (-0.8%) seems irrelevant. The biases for WHEW (-2.8%) and WEDD (3.8%) are small and carry different signs, probably related to their concomitant occurrence in urinary calculi. The 95% levels of agreement of WHEW, WEDD, and CARB were >10% (Table 3). These components often occur concomitantly in a single sample, causing complex spectral patterns. The 95% level of agreement of all results was 9%. This value should be taken as an indication, since it is statistically not correct to base such calculations on mutually dependent variables (each sample occurs 8 times). Only 3 of 92 patient samples exhibited maximum differences of 20%. These differences occurred consistently in samples that contained two rather similar components (WHEW + WEDD, and CARB + STRU). It is not known what analytical precision and bias are relevant in terms of the prevention of urinary calculus recurrence. We nevertheless consider the encountered differences minor and possibly irrelevant with respect to the ultimate (dietary) advice.

Apart from adequate quantification of the eight commonly occurring components, the library search in the GGN method enabled detection and quantification of rarely occurring components in four samples. This feature may be further developed by the addition of other components to the library, like uric acid dihydrate, in the near future. On the other hand, library search may be used for verification of network results. However, it may sometimes be somewhat difficult to establish an accurate quantitative composition of a sample in this way. This is illustrated by the results from the first and second hits obtained with a library search of patient A in Table 4.

In conclusion, the Golden Gate assay seems superior to the KBr assay because of its smaller sample size, because there is no need for sample pretreatment except for grinding, the turnaround time is shorter, and no time is lost because of KBr tablet breakage (Table 5). The GGN method does, however, require higher initial investment because of the Golden Gate ATR sampling device.

Table 5. Managerial summary of the KBr and Golden Gate assays

	KBr assay	Golden Gate assay
Sample size	1.5 mg	≤ 1 mg
Sample pretreatment	Yes	No
TAT	30 min.	15 min.
MTBF	1 day (pressing)	
Price <sup>a</sup> (US \$)	27600	34800
Required Knowledge	experienced	experienced

TAT, turnaround time (analysis and interpretation); MTBF, mean time between failures.

<sup>a</sup> FTIR + sampling device investments.

Because no sample pretreatment is needed, different brands of FT-IR spectrometers give similar spectra under equal local conditions (e.g., temperature and sample pressure), and the chemical composition of urinary calculi is similar in most developed countries, it would be interesting to investigate whether the GGN method could be transferred to other laboratories, without retraining the neural network with local data. This, however, awaits confirmation. The required expert knowledge for spectral interpretation is minimized by use of the ANN and library, but visual inspection remains necessary.

### Acknowledgements

We thank Wil Wennekes (CKCHL) and Theo de Haan (UHG) for excellent technical assistance. We also thank Bio-Rad for technical support.

### REFERENCES

1. Rebentisch G, Berg W. Comparison of methods and guaranteeing quality of analysis of urinary calculi—5<sup>th</sup> international ring test. *Urol Int* 1989;44:298–302.
2. Blijenberg BG, Wulkan RW, Zwang L, Liem TL, Leijnse B. Computergestützte Auswertung von Röntgendiffractometrie-Analysen. *J Clin Chem Clin Biochem* 1987;25:719–22.
3. Rebentisch G, Berg W, Pirlich W, Hommann D. Assessment and maintenance of the quality of urolith analyses in a comparison of methods. 4<sup>th</sup> International Ring Test to check quality. *Int Urol Nephrol* 1988;20:35–45.
4. Ng LM, Simmons R. Infrared spectroscopy. *Anal Chem* 1999;71:343–50.
5. Brown SD, Blank TB, Sum ST, Weyer LG. Chemometrics. *Anal Chem* 1994;66:315R–59R.
6. Gould N, Hallson PC, Kasidas GP, Samuell CT, Weir TB. Rapid computer-assisted infrared analysis of urinary calculi using photoacoustic detection. *Urol Res* 1995;23:63–9.
7. Berthelot M, Cornu G, Daudin M, Helbert M, Laurance C. Diffuse reflectance technique for infrared analysis of urinary calculi. *Clin Chem* 1987;33:780–3.
8. Oliver LK, Sweet RV. A system of interpretation of infrared spectra of calculi for routine use in the clinical laboratory. *Clin Chim Acta* 1976;72:17–32.



9. Hesse A, Gergeleit M, Schuller P, Moller K. Analysis of urinary stones by computerized infrared spectroscopy. *J Clin Chem Clin Biochem* 1989;27:639–42.
10. Hesse A, Sanders G. Atlas of infrared spectra for the analysis of urinary concrements. Stuttgart: George Thieme Verlag, 1988.
11. Lehmann CA, Mc Clure GL, Smolen I. Identification of renal calculi by computerized infrared spectroscopy. *Clin Chim Acta* 1988;173:107–16.
12. Hesse A, Sanders G, Döring R, Oelichmann J. Infrarotspektroskopische Harnsteinanalyse: Automatisierte Spektreuswertung mit Hilfe de Factoranalyse. *Frensenius Z Anal Chem* 1988;330:372–3.
13. Volmer M, Bolck A, Wolthers BG, de Ruyter de AJ, Doornbos AH, van der Slik W. Partial Least-Squares regression for routine analysis of urinary calculus composition with Fourier Transform infrared analysis. *Clin Chem* 1993;39:948–54.
14. Volmer M, Wolthers BG, Metting HJ, de Haan THY, Coenegracht PMJ and van der Slik W. Artificial neural network predictions of urinary calculus compositions analyzed with infrared spectroscopy. *Clin. Chem.* 1994;40:1692–1697.
15. Siegel SN, Castellan Jr NJ. Nonparametric statistics for the behavioral sciences (2nd Edition). New York: McGraw-Hill 1988;58-64.
16. Rumelhart DE, Hinton GE, Williams RJ. Learning representation by backpropagation errors. *Nature* 1986;323:533–6.
17. Bos A, Bos M, van der Linden WE. Artificial neural networks as tool for soft-modeling in quantitative analytical chemistry: the prediction of water content of cheese. *Anal Chim Acta* 1992;256:133–44.
18. Smits JRM, Melssen WJ, Buydens LMC, Kateman G. Using artificial neural networks for solving chemical problems. *Chemometr Intell Lab* 1994;22:165–89.
19. Bland JM, Altman DG. Statistical methods for assessing agreements between two methods of clinical measurement. *The Lancet* 1986;1:307-10.
20. 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.



## Chapter 7

### **Neuranet: The computer application for analyzing IR spectra of urinary calculi using an artificial neural network and library search**

Marcel Volmer<sup>1</sup>, Bert G. Wolthers<sup>1</sup>, Peter Koops<sup>2</sup>, and Ido P. Kema<sup>1</sup>

*<sup>1</sup>Department Pathology and Laboratory Medicine, University Hospital Groningen,  
PO Box 30001, 9700 RB Groningen,  
The Netherlands, and <sup>2</sup>Tauw, Deventer, The Netherlands*

## INTRODUCTION

Patients with urinary calculi are frequently treated with extracorporeal shock wave lithotripsy (ESWL). The chemical composition of these calculi is often analyzed with Fourier transform infrared spectroscopy (FT-IR), using different sampling devices, such as tablet holders for KBr tablets, diffuse reflectance, and the Golden Gate Single Reflection Diamond ATR sampling device. These FT-IR sampling methods make use of only minute amounts of authentic sample material, requiring no, or limited sample pre-treatment. The analysis is carried out in order to investigate the underlying disorder that causes the formation of the urinary calculus. With this information, the clinician can prescribe medicine, or diets to prevent recurrence of the urinary calculus.

Spectra of urinary calculi are often rather complex, requiring a substantial amount of expert knowledge of lab-technicians to quantify the calculus compositions by means of visual interpretation of the spectra. In this case the use of computerized library search may support this interpretation process. The computer compares the spectral characteristics of the unknown sample with all spectra available in the library and calculates which library spectrum is the most similar to the patient spectrum. This method may be limited by the number of spectra in the library. Normally huge numbers of library spectra are needed to obtain reliable results. It is virtually impossible to build a library, containing all possible combinations of components. Therefore, it is often necessary to combine and interpolate the best hits of a 'library' search, to obtain a reliable estimation of the composition of a patient sample. This interpolation process still has to be done by visual inspection of the spectra by trained technicians, after initial library search.

In order to improve the prediction accuracy of the compositions of urinary calculi we developed a computer program, based on an artificial neural network (ANN) tool and computerized library search. This program was especially developed for interpretation of spectra from urinary calculi, measured in the mid-infrared region between 4000–400  $\text{cm}^{-1}$ . The network device of this program can predict the relative composition of maximally ten components of a single sample (1). The neural network device is based on the back-propagation principle (2) (see also chapter 3.2.2), being a three-layer network, consisting of an input layer with a number of nodes (neurons) equal to the number of input variables (absorbances of the selected wave numbers), a single hidden layer with a variable number of nodes, and an output layer with a number of nodes equal to the number of components to predict. The input nodes are connected to the output nodes via the nodes of the hidden layer. All nodes of the hidden layer have every possible connection with the input and the output nodes (see Fig. 30, chapter 3.2.2). Each connection carries the product of the absorbance value belonging to a certain wave number and an individual weight. To obtain a network model, appropriate for the prediction of compositions of patient samples, the ANN should be presented with a sufficient number of samples with a known composition. These samples are called the training (calibration) samples. During training, the obtained values of the output neurons are compared to the real outcomes of the respective components of the training samples. In the next step the weight values are readjusted (backward step), depending on the differences between the calculated and real outcome of the components. After readjustment of the weights, the whole calculation process is repeated again. This process of calculation of the outcome, computation of the differences between the calculated and real outcome of the components, and readjustment of the weights, will be

repeated (iterated) until a pre-fixed number of iterations is reached, or the difference between the calculated and real outcome is between a acceptable limiting value (error value). The differences between the calculated and real outcome of the components is expressed as root mean square error (RMSE). The final weights in all connections are stored in the network model and reflect the knowledge of the underlying spectral patterns. In a certain sense the weights may be compared to the regression coefficients of e.g. linear regression. In contrast to library search, an ANN is a real quantitative method, capable of interpolation between quantitative compositions unavailable to the trained neural network model.

For further improvement of the accuracy of the neural network predicted compositions, we developed a subroutine in the program for the definition of a number of user definable expert rules. These user-defined rules have a syntax corresponding to simple 'Basic' language IF ... THEN ... ELSE ... statements. The syntax of these rules may be found in appendix A. In addition to the user definable rules, we also added a number of static 'expert' rules that should alert a technician to use the results of the library search in case of rarely occurring components, which were not included in the training set to the network model. Furthermore we added static rules for rounding the network outcome of each component to the nearest 5%. All rules aim at a generalization of the quantitative results of patient samples.

This designed computer program, which we called Neuranet (1) was developed because no specialized software was available for the prediction of the composition of urinary calculi, except for standard software for spectral processing and the availability of some libraries containing spectra of urinary calculi (3;4). The program enables building of one or more named methods, based on spectroscopic absorbance data of urinary calculi, in the range of 4000–400  $\text{cm}^{-1}$ . Each method contains a combination of: standard information (e.g. description of the components); a spectral library; a trained neural network model; some expert rules and properties for tuning and measurement of the performance of the different technologies included in the method. After selection of a method and a previously recorded spectrum from a file list, the program automatically performs data pre-processing, computerized library search, neural network predictions and expert rule filtering, for prediction of the outcome of a sample with an unknown composition. This manuscript explains the major functionality of the program. More in depth information with respect to the theoretical background of neural network processing is given in chapter 3.2.2 and can be found elsewhere (2;5;6).

#### *General description of the 'Neuranet' program*

Neuranet is a Windows based program running on a Windows 95/98 and NT platform. The program can read any IR absorbance or transmission file, provided that the spectra are measured from 4000–400  $\text{cm}^{-1}$ . Transmission spectra will automatically be converted to absorbance spectra by the program. The spectrum files should be stored as comma separated value (.CSV) files, or files with the spectrum file format (.SPC). Because FT-IR spectra obtained from urinary calculi, may differ as a result of different sampling devices, users are enabled to build their own library and neural network model.

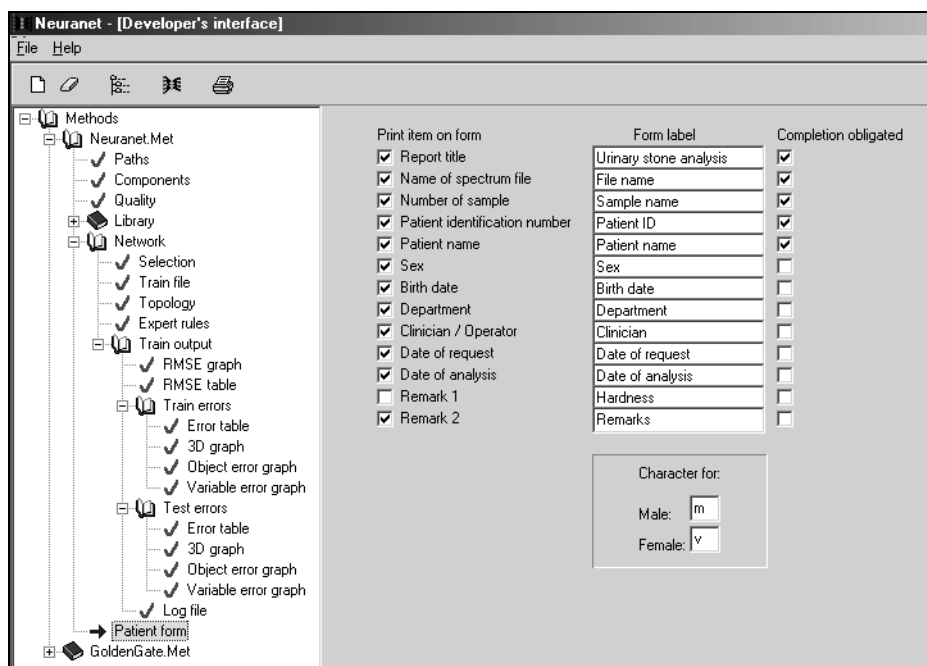


Figure 1. Screen capture of the developers part of Neuranet. The method explorer (left panel) shows an unfolded tree with the major sections (books, or folders) of the Neuranet method (Neuranet.Met). Each section contains one or more items whose properties have to be set on the workspace. The workspace (right panel) shows the properties of the patient form of this method.

Therefore we developed two separate computer applications in one program namely: the 'developers part' and the 'user part'. With the developers part of the program, the user can build one or more so-called methods. These methods contain all settings to perform spectral interpretation and quantification. A method can be selected and used in the user part of the program. Each method contains the following sections: a general section, a library section, a network section and a patient report section. Figure 1 shows the outline of the developers part of the Neuranet method.

Each section of a method is described by a number of items, whose properties have to be established before the method can be used in the user part of the program. An extensive list with functional specifications of all sections of the developers part is given in appendix B. Methods may be created from scratch, modified, deleted or copied. After making several copies of an initial method, one can make small changes to e.g. the topology (parameter settings) of the network section of a method. Activating the 'train' button in the developers part of the program gives a user the opportunity to select one, or several methods for unattended batch-wise training. Because training of neural networks may be a time consuming process, comparison of several network outcomes will be much easier after this unattended batch-wise training. Each method, including all settings will be stored on disk. Only methods with a fully customized network model and library can be used in the user part of the program. Sections, or items of methods that are not properly filled-in for use in the user part of Neuranet, will be flagged red. The developer part of the program may be

password protected to prevent accidental changed of fully customized methods. An example of method development is given in appendix D. This development is illustrated by a method intended for the prediction of 3 components (whewellite, weddellite and carbonate apatite).

For prediction of the composition of patient samples, the user part of Neuranet must be used. After selection of a method, the settings will be loaded. This method contains all relevant settings, such as the spectrum file location and a trained neural network model. The user part can be divided in two major sections: the spectrum explorer and the workspace of the program (Figure 2). The spectrum explorer contains a list with files of all recorded spectra. After selection of a spectrum, library search and network prediction followed by filtering by expert rules will be performed automatically, within approximately one second. The results from these processes are made visible on a form of the workspace. The workspace of the user part contains 5 tab-sheet forms that can be selected by pressing a tab-label on the workspace. The workspace contains the following forms: Patient entry, Library search, Network results, Spectral view and All results. All fields on these forms will be automatically filled in with results obtained from the currently selected patient sample, or can be entered by the user. A more extensive description of the functional specifications of spectrum explorer and the five tab-sheet forms of the user part of Neuranet is given in appendix C.

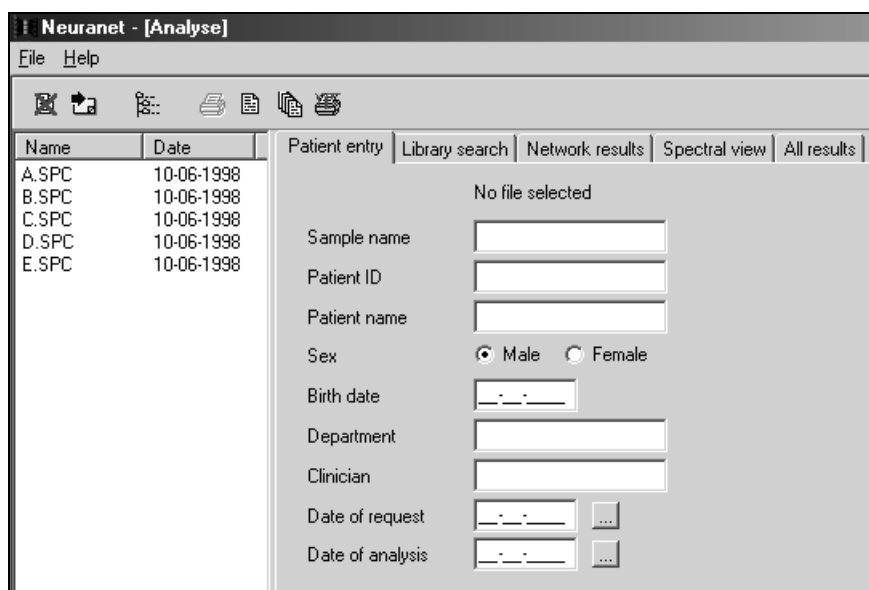


Figure 2. User part of Neuranet with the spectrum explorer (left panel) and workspace (right panel). The workspace is subdivided over 5 tab-sheets. In this case the 'patient entry' is active.

After processing of a number of patient spectra in the same run, the final compositions of the patient samples are stored on disk as spectrum file (.SPC). The contents of this file may be imported in a Laboratory Information Management System (LIMS). A more detailed description of the results of some selected patient samples will be given in the next section.

*Processing of patient samples with the 'user part' of Neuronet*

For the prediction of the composition of urinary calculi, the network model of the Neuronet method (1) was developed for the prediction of the following eight components: ammonium hydrogen urate (AMUR), brushite (BRUSH), carbonate apatite (CARB), cystine (CYST), struvite (STRUUV), uric acid (URIC), weddellite (WEDD) and whewellite (WHEW). In addition to mixtures similar to those used for network training, the library of the Neuronet method contained several special components. After network prediction the results were filtered by some selected expert rules. Like any Neuronet method, this method contains 3 standard expert filters, namely: extreme value limit, truncating to zero and normalization. The extreme value limit was set to minus 15% to notify the user with a message to ignore the network results and use the results from the library search in case of patient samples with qualitative compositions unavailable to the network model. The truncating to zero rule was set to 3.5, meaning that all network outcome  $\leq 3.5$  will be forced to zero, before normalization. The normalize rule effects rounding the network outcome to the nearest 5 fold and normalization of the component concentrations to 100% (see also appendix D). In addition to the three standard rules, the Neuronet method contained 22 dynamic expert rules. These rules were added to correct small, but systematic differences between the network outcome and the real composition of a patient sample. The real compositions of the patient samples were estimated by a trained technician by means of visual inspection of the spectrum and was based on his knowledge about the consistence / inconsistency of particular components in certain mixtures (e.g. a small amount of uric acid in a cystine stone would be highly unlikely).

In addition to the prediction of the compositions of urinary calculi with Neuronet (network results, expert system filtering and library search), the final results were always validated by visual inspection of the spectra by a trained technician.

To demonstrate the use of the described Neuronet program, we subsequently processed (as an example) 5 IR spectra from patient samples A–E with unknown compositions, after selection of the Neuronet method (Neuronet.Met) in the user part of the program. The network-, and the expert system predicted compositions and the final results of the five patient samples are shown in Table 1. The results of each sample will be described in more detail.

Results from sample A

After selection of the corresponding prerecorded SPC file of sample A in the spectrum explorer of the user part, library search and network prediction is performed almost instantaneously. Figure 3 shows a list with the twelve most similar spectra of the Neuronet library compared to sample A. In this case, the spectral bands (absorbances) of the fingerprint area ( $2112\text{--}450\text{ cm}^{-1}$ ) of the patient spectrum were compared to the absorbances of the corresponding wave numbers of all available library spectra. The correlation coefficient was used as search algorithm. Because the correlation coefficient is a similarity measurement, the most similar library spectrum is placed on top of the list.

The four-finger band around  $1000\text{ cm}^{-1}$  (square bracket in Figure 4) indicates that this sample contains brushite. However, brushite does not seem to be the only calcium phosphate material in this patient sample. The spectral definition of the phosphate band around  $1000\text{ cm}^{-1}$  is more or less comparable to the same band(s) in the library spectrum containing brushite / carbonate apatite in a relative abundance of 50%/50% (w/w).



Patient entry		Library search	Network results	Spectral view	All results
Fingerprint range (2112-450 cm <sup>-1</sup> )		Correlation coefficient			
No.	Composition			Distance	
» 1	Whewellite-Brushite 10%-90%			0.9466	
» 2	Brushite-Carbonate Apatite 50%-50%			0.9450	
» 3	Weddellite-Brushite 10%-90%			0.9428	
4	Brushite-Carbonate Apatite 40%-60%			0.9387	
5	Brushite-Carbonate Apatite 75%-25%			0.9352	
6	Whe-Wed-Brushite 12.5%-12.5%-75%			0.9344	
7	Brushite-Carbonate Apatite 91%-9%			0.9338	
8	Brushite-Carbonate Apatite 36%-63%			0.9303	
9	Weddellite-Brushite 25%-75%			0.9233	
10	Struvite-Brushite 25%-75%			0.9156	
11	Whewellite (Roth)-Brushite 30%-70%			0.9080	
12	* Brushite (Merck 2146)			0.9077	

Figure 3. Library search list of sample A.

Table 1. Network- and expert system predicted and final compositions.

	AMUR	BRUSH	CARB	CYST	STRU	URIC	WEDD	WHEW
Sample A:								
network	-3	46	44	-2	0	0	13	2
expert		45	40				15	
final		45	40				15	
Sample B:								
network	-1	3	98	-3	1	0	3	-1
expert			100					
final			100					
Sample C:								
network	0	23	78	-1	1	-0	2	-3
expert		25	75					
final		25	75					
Sample D:								
network	7	-3	68	3	22	1	0	2
expert			80		20			
final			80		20			
Sample E:								
network	-3	-76	1	87	39	-15	126	-59
expert	Alert message: The expert system could not interpret this composition							
final	Cholesterol 100%							

See text for the abbreviation of the component names.

From these findings we may conclude that the sample contains small amounts of calcium oxalate and a mixture of brushite and carbonate apatite. Obtaining the quantitative composition from these search results may be rather difficult. Therefore we will have a look at the results obtained with the neural network (Figure 5). From the network results it can

be seen that it predicts the sample to contain equal amounts of brushite (46%) and carbonate apatite (44%) and a small amount of weddellite (13%). The truncate to zero rule of the expert system forced the network outcome of ammonium hydrogen urate, cystine and whewellite to zero (all less than 3.5%).

In theory, a patient spectrum with a correlation coefficient of 1.000 would be identical to the library spectrum. On the other hand, it is unlikely that dissimilar components in the library will display correlation coefficients higher than a generally accepted value of 0.95. In case of sample A, none of the library spectra had a correlation coefficient higher than 0.95. Therefore, they are rather dissimilar to the patient sample. From the first nine hits one could suppose that sample A should contain calcium oxalate (Whew, or Wedd) and/or calcium phosphate (brushite and carbonate apatite). Figure 4 shows the spectra of the 3 best hits, on top of the spectrum of sample A. The spectral band at  $1300\text{ cm}^{-1}$  (arrow in figure 4) shows evidence that the sample contains calcium oxalate (whewellite, or weddellite).

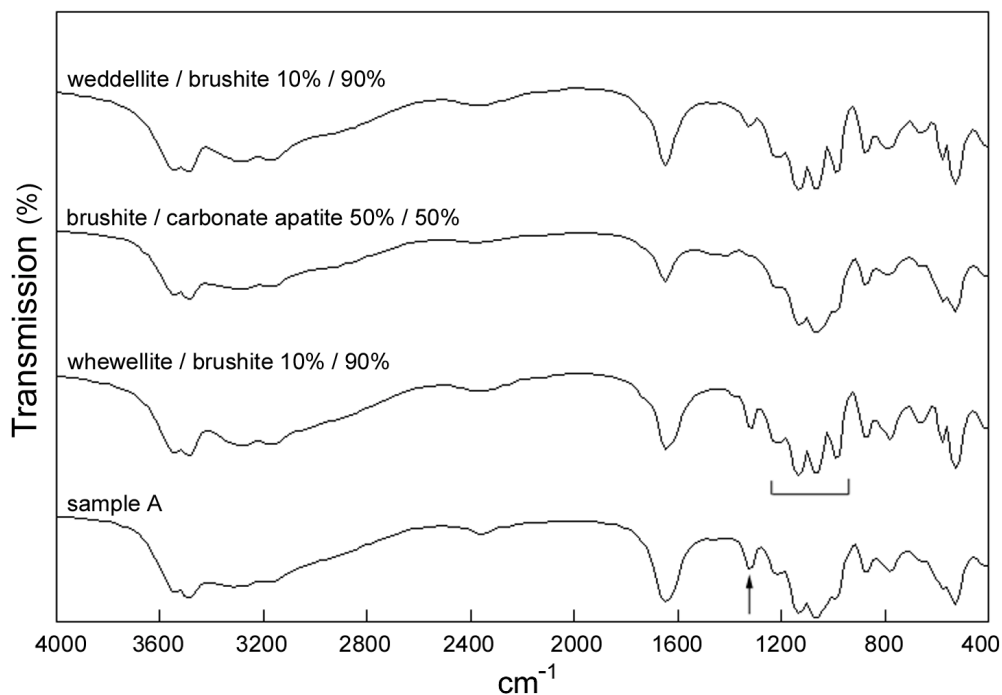


Figure 4. Spectral view of sample A and de 3 best hits from library search.


Patient entry		Library search		Network results		Spectral view		All res	
E2F 									
Results									
Component names	Network	Expert	Final						
Am. H. Urate	-3								
Brushite	46	45							
Carbonate Apatite	44	40							
Cystine	-2								
Struvite	0								
Uric acid	0								
Weddellite	13	15							
Whewellite	2								

Figure 5. Network predicted and expert system results of sample A.

To preserve normalization of the sample outcome, the results of brushite, carbonate apatite and weddellite were rounded to the nearest 5 fold and normalized to 100% by means of the normalize rule of the expert system. Visual inspection of the spectrum by a trained technician revealed that the final composition of the sample was similar to the one obtained by the expert system (Table 1).

### Results from sample B

Figure 6 shows the results from library search of sample B. For library search, the fingerprint area of the spectrum was used ( $2112\text{--}450\text{ cm}^{-1}$ ). The first 3 hits were very similar to the patient spectrum. They had correlation coefficients of 0.9915, 0.9899 and 0.9897, respectively. At first sight, the fingerprint areas of the 3 library spectra and of patient sample B look very similar. However, a closer look will show the differences. There are more striking differences at higher wave numbers of the spectra. The area between  $3700\text{--}3200\text{ cm}^{-1}$  shows that the first hit of library search (brushite / carbonate apatite 15% / 85%) is not very similar to the patient spectrum. The same is more or less true for the second hit of library search (struvite / carbonate apatite 10% / 90%). The patient sample has a somewhat bell shaped band in the area between  $3700\text{--}3200\text{ cm}^{-1}$ , whereas the shape of the band of the second hit is more flattened. At first sight, the third hit from library search shows to be rather similar with sample B (Figure 6). Neural network prediction and expert system filtering revealed that the composition of sample B was 100% carbonate apatite. This result corresponded to the third hit obtained from library search. From this sample it is shown that the first hit from library search is not always the correct one.

### Results from sample C

The determination of the composition of sample C did not meet with special difficulties. Both library search, and network prediction showed the same outcome. This result was adapted as final result, after validation of the spectrum.

Results from D

According to library search, sample D was composed of a mixture of struvite and carbonate apatite. The first hit from library search showed struvite / carbonate apatite in relative abundance of 30%/70%. The second and third hit showed relative compositions of 20%/80% and 40%/60%, respectively.

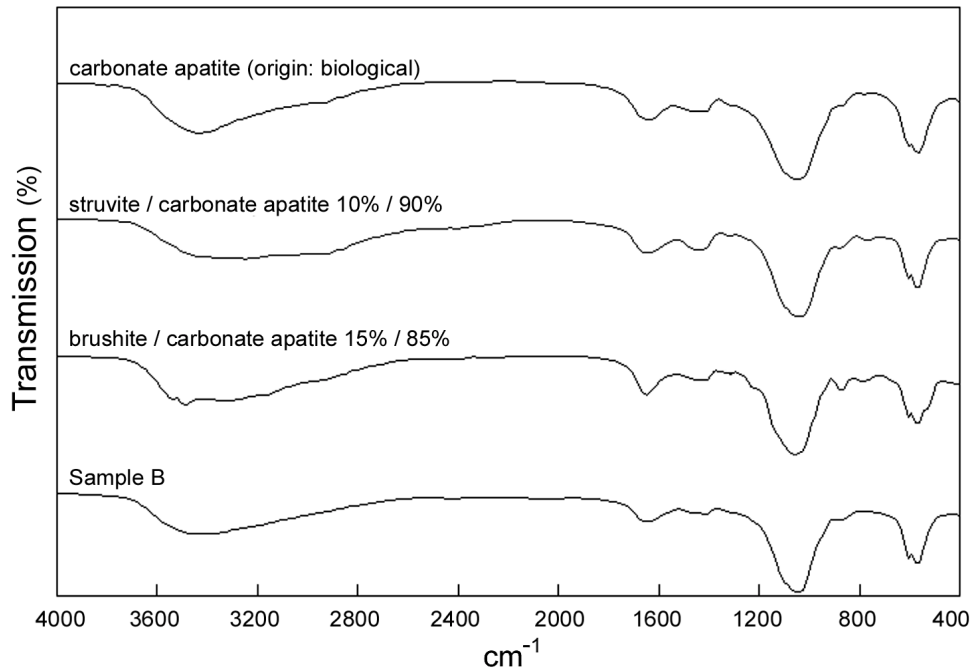


Figure 6. Transmission spectrum of sample B and the first three hits from library search.


Patient entry   Library search   Network results   Spectral view   All res				
E2F 				
Results				
Component names	Network	Expert	Final	
Amm. H. Urate	7			
Brushite	-3			
Carbonate Apatite	68	80		
Cystine	3			
Struvite	22	20		
Uric acid	1			
Weddellite	0			
Whewellite	2			

Figure 7. Network and expert system results of sample D.

In addition to library search, the network model had predicted the presence of 7% ammonium hydrogen urate (see Figure 7). Although not impossible, the presence of such a small amount of ammonium hydrogen urate in a sample containing struvite (infection stone) and carbonate apatite is rather unlikely. Figure 8 shows the absorbance spectrum of sample D. The spectrum of this sample does not show the presence of any ammonium hydrogen urate, which normally demonstrates a great number of sharp spectral bands in the fingerprint area of the spectrum. However, the spectrum contains a little noise that may be interpreted by the network as a small amount of ammonium hydrogen urate. The presence of ammonium hydrogen urate is more often mistakenly identified in noisy spectra. For correction of this systematically occurring inaccuracy an expert rule was added to the Neuranet method. After application of this rule, followed by the truncating to zero rule and normalization rule, the composition of the sample was struvite / carbonate apatite 20%/80%. This composition was similar to the final composition after assessment of the spectrum by a trained technician (Table 1).

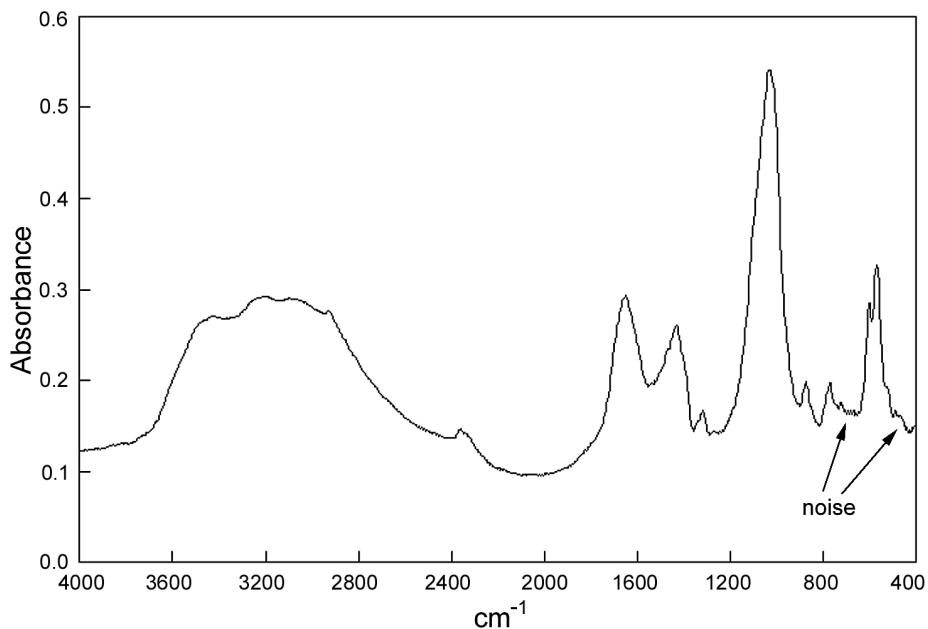


Figure 8. Absorbance spectrum of sample D.

#### Result from sample E

The network predicted results of sample E showed very large positive and negative numbers for the components (Figure 9).

The results from library search showed that the calculus was composed of 100% cholesterol (correlation coefficient 0.9779). The first three hits obtained from library search are also normally reported on the workspace of the network results (Figure 9). Because the calculus is composed of cholesterol, it must be a bladder stone. The extreme value rule of the expert system ignores the network results and gives the user a hint to make use of the results from library search (Figure 9).

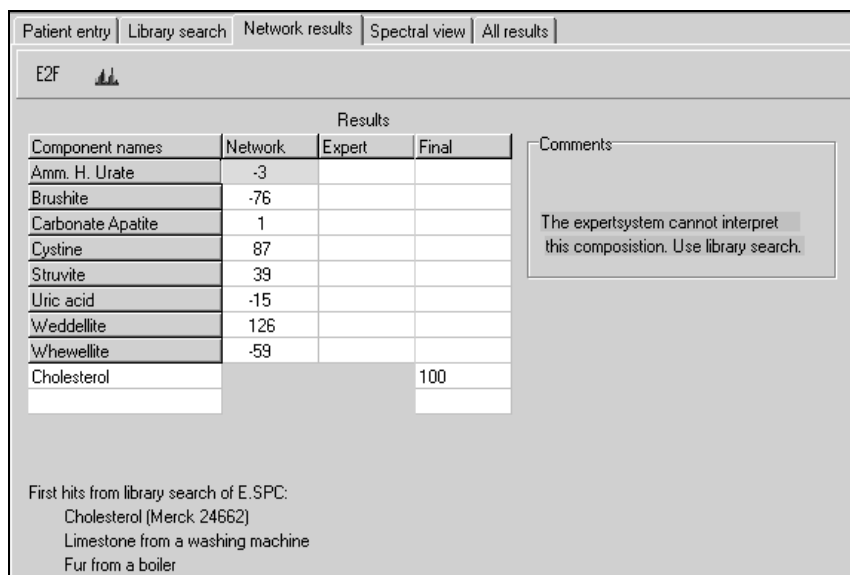


Figure 9. Network results of sample E.

These results illustrate that the Neuronet program can be very helpful in the interpretation process, needed for prediction of compositions of urinary calculi, analysed with IR spectroscopy. The program may be used for prediction of quantitative compositions of calculi, using any kind of IR sampling device, assuming measurement of spectra from 4000–400  $\text{cm}^{-1}$  in the mid-infrared region. The operation of the program is simple. However, notwithstanding the applicability of the program, visual inspection of the spectra by trained technicians remains necessary.

## REFERENCES

1. Volmer M, Wolthers BG, Metting HJ, de Haan THY, Coenegracht PMJ and van der Slik W. Artificial neural network predictions of urinary calculus compositions analyzed with infrared spectroscopy. *Clin. Chem.* 1994;40:1692–1697.
2. Rumelhart DE, Hinton GE, Williams RJ. Learning representation by backpropagation errors. *Nature* 1986;323:533–6.
3. Lehmann CA, Mc Clure GL, Smolen I. Identification of renal calculi by computerized infrared spectroscopy. *Clin Chim Acta* 1988;173:107–16.
4. Blijenberg BG, Wulkan RW, Zwang L, Liem TL, Leijnse B. Computergestützte Auswertung von Röntgendiffractometrie-Analysen. *J Clin Chem Clin Biochem* 1987;25:719–22.
5. Bos A, Bos M, van der Linden WE. Artificial neural networks as tool for soft-modeling in quantitative analytical chemistry: the prediction of water content of cheese. *Anal Chim Acta* 1992;256:133–44.
6. Smits JRM, Melssen WJ, Buydens LMC, Kateman G. Using artificial neural networks for solving chemical problems. *Chemometr Intell Lab* 1994;22:165–89.

## Appendix A: Syntax of the expert rules

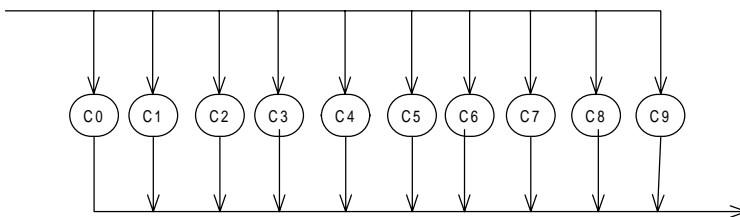
In this appendix, the syntax diagrams describe the exact definitions of expert rules. These diagrams are similar to syntax diagrams used for the description of the syntax of programming languages. To read a syntax diagram, follow the arrows. Frequently, more than one path is possible. In a syntax diagram, parallel rectangular and circular boxes depict the alternative pathways. Rectangular boxes are drawn for constructions that can be specified subsequently, whereas text in circular boxes cannot be specified any further. Therefore the text inside the circular boxes must be handled unchanged.

The expert system makes use of the following terms:

- Components, numbers and logical operators
- Arithmetic operations between brackets
- Factors
- Expressions
- Statements
- Rules

The terms are ordered from a low to a higher aggregation level. Except for components, numbers and logical operators, each term may be composed of one, or more of the previous terms. Rules have the highest aggregation level and are used as final expert rule. The next part contains a description of the respective terms.

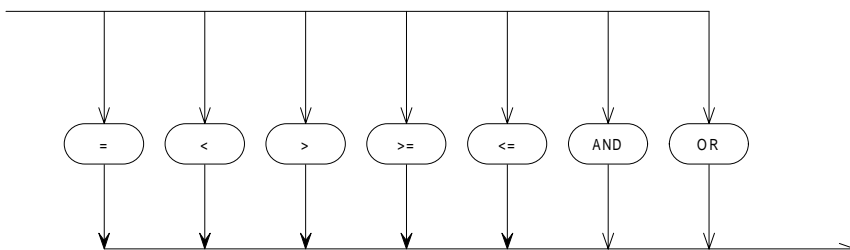
- *Components.* A name of a component is represented with the character 'c' and a number (0 – 9). In this way maximum 10 components can be defined, depending on the component definitions in the general section (appendix B) of the method in the developers part. The components (variables) c0 – c9 can be seen as abbreviated components names, which are used to store a value of the respective component. The syntax diagram is as follows:



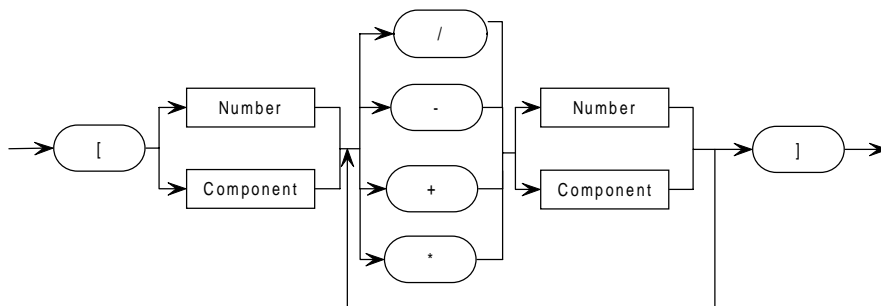
Example: Suppose c7 represents the component whewellite.

If whewellite is 10% the result of the  $c7 = 0$  (statement) means that whewellite will be set to zero (instead of 10%).

- *Logical operators.* The expert system makes use of the following logical operators:

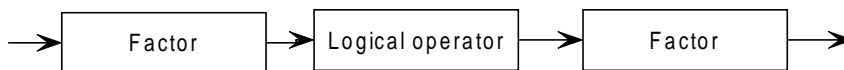


- *Numbers.* Any kind, positive, or negative.
- *Arithmetic operations between brackets.* A text between brackets (...[...]) is an arithmetic expression. In the syntax diagram, the - sign is entered as ctrl -. The expert system does not evaluate the multiply, or divide sign before the plus or minus sign. The result of [8+3\*2] will be evaluated as 22 and not as 14!

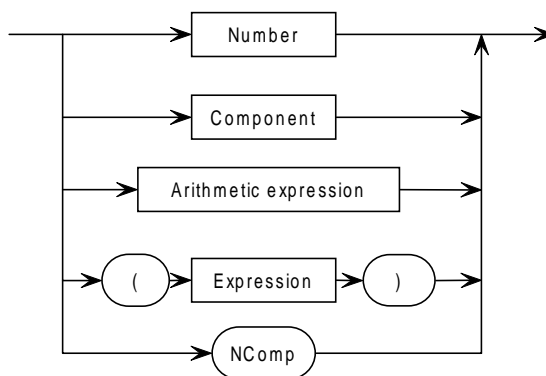


Example: The result of [c0 + c4] is the sum of the concentrations of the components c0 and c4. The result of [c4 - 10] is the difference between the component c4 and the number 10.  
Another arithmetic operation could be e.g. [c7 / c2 - 2].

- *Expressions.* Expression are build from a factor, a logical operator and another factor:



In which a factor is:



NComp is a variable that contains the number of components with a concentration greater than zero. A single expression contains no parentheses. Only when expressions are nested, parentheses are required.



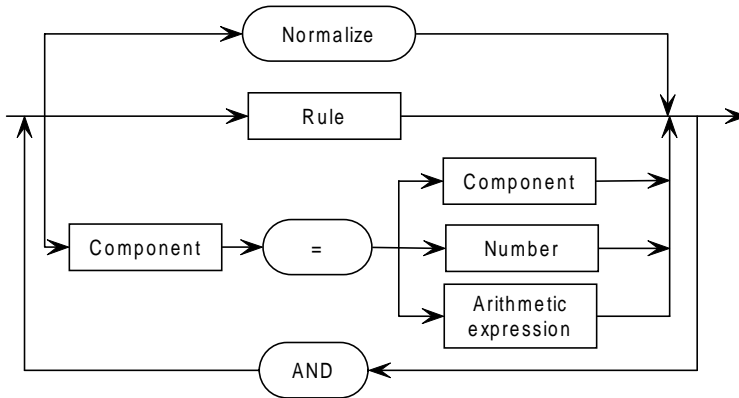
Examples:

$c0 < 10$ , is true when the composition of  $c0$  is lower then 10.

$(c0 < 10) \text{ AND } (c0 > 5)$ , is true when the composition of  $c0$  lies between 5 and 10.

$[c4 + c5] > 60$ , is true when sum of the composition of  $c4$  and  $c5$  is greater then 60.

- *Statements.* Statements are the commands that will be executed by the expert system. They are described in the following syntax diagram:  
Normalize will cause normalization of the component concentrations to 100%

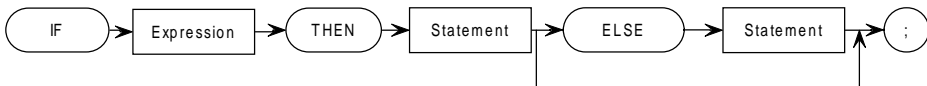


Examples:

$c0=[c0+c4]$ , The composition of  $c4$  is added to component  $c0$ .

$c0=10 \text{ AND } \text{Normalize}$ , The composition of  $c0$  is set to 10 and the data will be normalized.

- *Rules.* A rule is the final syntax of the expert system to interpret. Each rule must be described by the following syntax and must be terminated with a semicolon:



Examples:

IF  $(c0 < 10) \text{ AND } (c5 > 60)$  THEN  $c5=[c5+c0]$  AND  $c0=0$ ;

If the result of  $c0$  is less then 10 and  $c5$  is greater than 60, this rule adds the result of  $c0$  to  $c5$  and makes  $c0$  zero.

## Appendix B: Specifications of the Neurinet developer part

### General section:

- Path: defines the location of the spectrum files on disk.
- Components: names, abbreviations and descriptions of up to ten components that can be predicted by the neural network simultaneously.
- Quality: some limits for triggering a warning to the user in case of overloaded-, or weak spectra, or in case of possible background problems.

### Library section:

- Library settings: for setting of the permitted search algorithms (absolute-, or squared difference, absolute-, or squared derivative, Euclidean distance and the correlation coefficient), the default search algorithm, the default number of spectral hits for printing and the default number of library spectra for viewing.
- Search area: the library resolution (4, 8, or 16  $\text{cm}^{-1}$ ) and one, or more named areas for searching must be specified. These areas can be selected by drawing rectangles over the chosen wave number regions.
- Library spectra: one or more library spectra can be added to the library. For each spectrum a description must be defined. These spectra can be added one by one, or by a reading the file names and their descriptions from a spreadsheet file (Lotus 123 version 1).

### Network section:

- Selection: the resolution (4, 8, or 16  $\text{cm}^{-1}$ ) of the spectra and one, or more named areas for network processing must be defined. These areas can be selected by drawing rectangles over the chosen wave number regions.
- Train file: one, or more spectra can be added to the network. For each spectrum the known composition of the defined components (expressed as weight %) must be entered. The training set can be divided in a training-set for calibration of the network model and a validation-set for tuning the parameters of the network model. A checkbox must be set to define whether the spectrum should be used as a training, or validation object. The spectra can be added one by one, or by a reading their file names, training- or validation-status and quantitative compositions from a spreadsheet file (Lotus 123 version 1).
- Topology: these settings are needed for each training session (Figure 2). The number of input neurons are automatically set by the program and is based on the resolution and the selected wavenumber range. The number of output neurons is similar to the number of defined components, as set in the “General section”. The other settings can be changed by the user. The user can select from 3 transfer functions (sigmoid, symmetric sigmoid and linear). The kind of transfer functions that can be selected by the user depends on the type of connection (between the input–hidden, hidden–output, or input–output neurons). Fully connected means, that the input neurons can be directly coupled to the output neurons. To prevent adaptation of the network to fixed patterns in the training data, the shuffle training-fact checkbox can be set, causing randomized offering of the training- and validation-samples every training cycle (epoch). The input, as well as the output may be scaled between -1 and +1, or auto-scaled. The “training stopping rule” is based on a maximum number of cycles, or a minimum Root Mean Square Error (RMSE). Logging of the RMSE errors against the number of epochs will

automatically occur. To prevent long error files, an epoch ‘step size’ may be set. Setting the check-box Initialize weights (randomized between  $-0.1$  and  $+0.1$ ) starts training from the scratch, otherwise the training will continue from the point it stopped before. The properties of the topology are shown in Figure 13.

- **Expert rules:** simple BASIC like expert rules can be added, modified, or removed. Each expert rule has its own name. Every rule is composed of simple IF ... THEN ... ELSE statements. Arithmetic expressions with different operators (+, -, x, /) are allowed, as well as logical operators (=, >, <, >=, <=, AND, OR). More complex rules can be build by writing the expressions between parentheses. A “Normalize” function takes care of rounding values to the nearest 5 fold and takes care that all components add up to 100%. An extreme value can be defined to alert a user to use the results from library search in case of rare occurring components, that were unavailable in the network model. More information about the syntax of the rules can be found in appendix A. An expression evaluator is available, as well as a syntax checker.
- **RMSE graph and table:** contains a graph and table showing the number of epochs against the RMSE of both training- and validation-sets respectively. To prevent long error lists, the RMSE values are only shown at epochs equal to the step size setting (see the topology settings). This table and graph (Figure 16) are important for validation of the performance of the network model. Careful inspection is needed to prevent under-, or over-fitting of the network model.
- **Train errors:** this consists of a table containing the absolute differences (errors) between the predicted and the expected outcome of each training object (sample), which are shown for each component separately. A 3-D graph shows the contents of this table (x-axis = sample number, y-axis = component number, and z-axis = error). This 3-D graph can be rotated for a better perspective view. A second graph (2D) shows the mean absolute difference of all components for each training object (mean component error against the object number). A third graph (2D) shows the mean absolute difference of all objects for each component (mean object error against the component number). The table and graphs can be used for the detection of outlier objects, or the assessment of problems with certain components. Large errors of certain objects (spectra, or samples), found after training with a small number of epochs (100 – 200), may be indicative for typing errors of the composition (%), interchanging of samples, or just outliers.
- **Test errors:** the same tables and graphs as described at the ‘train errors’, but with the results of the validation set.
- **Log file:** a form containing the training start and stop time and all possible errors that did occur during training.

*A patient form section:*

- With this section the developer of the method can select the items and define their descriptions that will be printed on the report (Figure 1). This report will be printed in the ‘user part’ of the program.

**Appendix C: Specifications of the Neuronet user part**

- **Patient entry:** with this form the user can enter some patient identification information, such as patient id, patient name, sample name, and the date of request. This information will be printed on the patient report. The patient identification entry will also be stored on disk for later use.
- **Library search:** this form contains an ordered list with the results obtained from library search. The list contains the descriptions of the library spectra and the results from the default search algorithm (e.g. the correlation coefficient). The order of the list is based on the results from the search algorithm. In case of the correlation search algorithm, spectra from the library, most similar to the patient spectrum will be placed on top of the list. The first  $x$  number of library spectra from the list are flagged for viewing on the **Spectral view** page. However, the user can select any other spectrum from the list for viewing. After selection of another search algorithm, or after selection of another search area from a drop down list, a new library search will be performed instantaneously.
- **Network results:** this form contains the results from the neural network prediction of a limited number of components (maximum 10). These components were specified in the **General** section of the Developers part of the program. For each component, the network results are shown in the first column of a table. The predicted data of the neural network, which were processed by the expert system, are shown in the second column of this table. If the expert system could not interpret the network results, a hint is given to use the results from the library search. The final composition of the urinary calculus can be entered in the last column of this table. The description of the first hit from the library search, shown on the previous '**Library search** tab-sheet form', is shown for direct comparison with the network and expert system results. A user can add two extra component names with their quantitative compositions to the table on the **Network results** form. This may be necessary for components unavailable in the neural network model. The user can also enter the hardness of the sample and a general remark on this form. If the recorded spectrum is not within the defined quality level, as specified in the **Quality** entry of the General section of the network model, a message will be shown to check the reliability of the results. The form contains a button for viewing the spectrum. Several properties for viewing this spectrum can be changed. These properties will be described at the next tab-sheet form. As soon as the results in the 'final results' column are entered, they will be automatically transported to the table of the '**All results** tab-sheet form'.
- **Spectral view:** On this form the results of the patient spectrum is shown, together with a number of library spectra. The number of library spectra and the kind of spectrum depends on whether these spectra are selected (flagged) or not on the **Library search** form. Several options may be set for viewing of the spectra: the
  - wavelength set to  $\text{cm}^{-1}$ , or nm (x-axis)
  - spectra may be shown in transmission (%), or absorbance (y-axis)
  - spectra may be stacked, or overlaid
  - a crosshair may be shown
  - the spectra may be zoomed and dragged
  - horizontal and vertical grids may be shown

In the stacked mode, the patient spectrum is always the lowest spectrum in the graph, followed with library spectrum with the best hit directly on top of it and the rest of the spectra in their respective hit order. The purpose of the spectral view is a verification of numerical presentation of the library search from the first tab sheet and the results obtained with the network prediction. The graph may be printed in its actual state (e.g. zoomed).

- All results: This form contains a table (list) with all results (network, expert system and final) of all patient spectra that were processed in the current program session. This list can be printed or saved on disk.

## Appendix D: Development of a method file

As an example, the most important parts of method development will be demonstrated by creation of a potential new method intended for prediction of the composition of 3 of the most commonly occurring components of urinary calculi. This appendix only describes the development of a method. More detailed information and theoretical background in relation to library search and neural network processing may be found in paragraph 3. The components of this new method are weddellite (We), whewellite (Wh) and carbonate apatite (Cap). In a sample, a combination of the three components always add up to 100%. For the development of this method primary (synthetic) standards were used whenever possible. These synthetic samples always must reflect their biological counterparts as much as possible. If synthetic components are unavailable one has to make sure that the biological standards used for the development of the method are actually pure. All spectra were measured from 4000–400  $\text{cm}^{-1}$ , using a resolution of 4  $\text{cm}^{-1}$ .

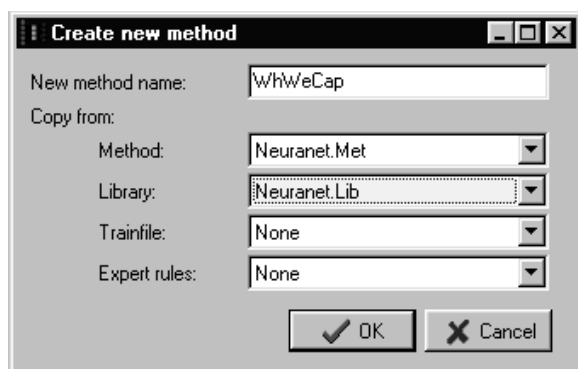


Figure 10. Dialog for the creation of a new method.

New methods may be created from scratch, or by copying parts (sections) of existing methods. Figure 10 shows the creation of a new method, named WhWeCap, whose settings (properties) are partly copied from an existing method named Neuranet. After creation, double clicking the method name WhWeCap in the method explorer of the program will unfold the properties of the method. An example of such a method tree containing the properties of the Neuranet method is given in Figure 1. Before specification of any network settings, the component names and their descriptions have to be entered (see Figure 11).

After specification of the component names, the library settings must be defined. At first the default and permitted library search algorithms have to be set. The following algorithms may be set: absolute-, or squared-difference, absolute-, or squared- derivative, Euclidean difference and the correlation coefficient. A more detailed description of these algorithms may be found in paragraph 3c. The user may select a search algorithm from a drop-down list (containing the permitted algorithms) in the user part of the program. Additionally, one or more search areas have to be defined. Each search area is given a unique name and is defined by one, or more non-continuous wave number ranges. For instance, if one is interested in searching the whole spectrum range from 4000–400  $\text{cm}^{-1}$ , the well known  $\text{CO}_2$  band (between 2250–2450  $\text{cm}^{-1}$ ) may be excluded by selection of wave number areas

before and after that band. Dragging with the mouse over the spectrum will carry out the wave number selections. Except for wave number selection, the resolution of the library has to be set at 16, 8, or 4  $\text{cm}^{-1}$  (Figure 12). Each method may contain one or more of these uniquely named wave number groups, which can be selected by name from a drop-down list, in the user part of the program.

Component	Name	Description	Abbreviation
1	Whewellite	[CaC#2#O#4#1H#2#O]	Whew
2	Weddellite	[CaC#2#O#4#2H#2#O]	Wedd
3	Carbonate Apatite	[Ca#10#(HPO#4#),(CO#3#)#6#(OH,CC Carb	
4			
5			
6			
7			
8			
9			
10			

Figure 11. Definition of the component names.

Two surrounding '#' characters define a subscript layout of the enclosed character.

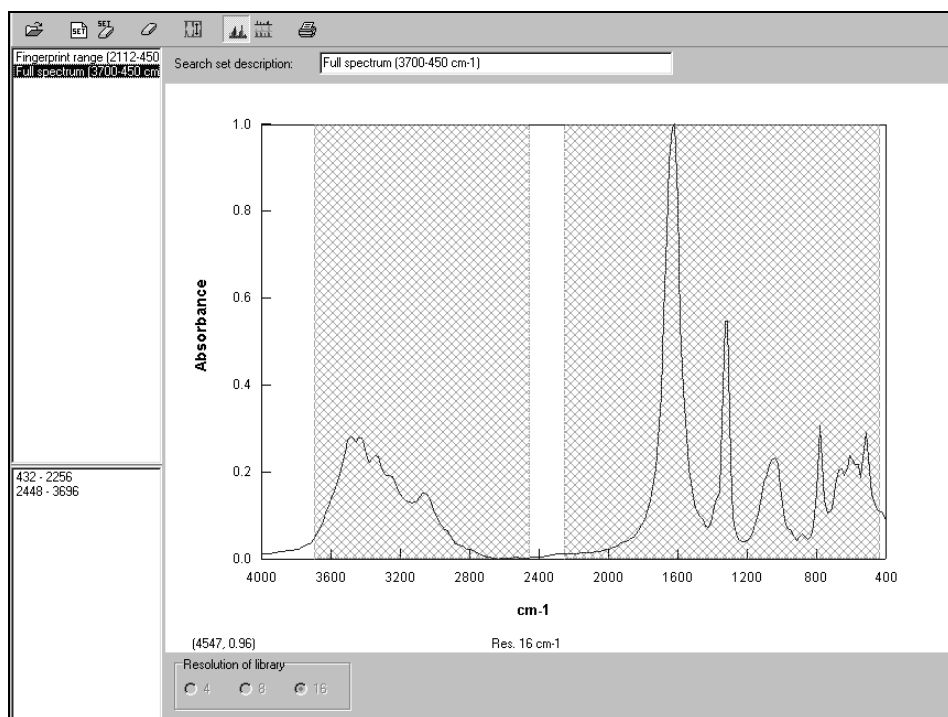


Figure 12. Search area of a wave number group, named 'Full spectrum', defining a non-continuous wave number range, by excluding the CO<sub>2</sub> band between 2250–2450  $\text{cm}^{-1}$  and the high end of the spectrum.

For library search applied on IR spectra of urinary calculi, a resolution of  $16\text{ cm}^{-1}$ , a wave number range limited to the fingerprint area of the spectrum ( $2112\text{--}450\text{ cm}^{-1}$ ), and the correlation search algorithm proved to be sufficient. After definition of the library settings, the spectra for the library have to be entered. They may be entered one by one, or imported from a spreadsheet containing a list of file names and descriptions. The descriptions should express the quantitative compositions of the spectra. An example of such a list may be found in appendix E. After filling-in the library definitions, the neural network pre-processing parameters had to be defined.

The resolution of the spectra to be processed by the neural network was set to  $16\text{ cm}^{-1}$  (selected out of 16, 8, or  $4\text{ cm}^{-1}$ ) and the spectral wave number range was selected from  $1424\text{--}400\text{ cm}^{-1}$ , by dragging with the mouse over the spectrum. Both resolution and wave number range settings proved to be sufficient to obtain accurate results from neural network processing. Then the spectra for training- and validation (testing) of the neural network model were imported by reading their file names and compositions from a spreadsheet file (see appendix F). The samples from the training-set are needed for finding the final weights that reflect the knowledge of the underlying spectral patterns, whereas the samples of the validation-set are used for testing of the performance of the network, as we will see later on. After filling-in the preceding network settings, the network topology parameters must be entered (Figure 13). As a result of the selected wave number range and resolution, 65 input neurons will be calculated and entered by the program ( $((1424-400)/16)+1$ ), whereas the number of output neurons equals to the number of components (3).

	Neurons	Function	Learning rate	Momentum
Input:	65			
Hidden:	8	Symm. sigmoid	0.05	0
Output:	3	Linear	0.05	0
<input type="checkbox"/> Fully connected		Linear	0.15	0
<input type="checkbox"/> Shuffle training-fact.				
Scale input between	0	and	1	
Scale output between	0.4	and	0.6	
Maximum number of epochs	20000	Stop error	0.100	%
Stepsize for error-file	100	<input checked="" type="checkbox"/> Initialize weights		

Figure 13. Initial network topology.



The number of hidden neurons determines the total number of connections (weights) of the network. Many heuristic guidelines are available for the choice of the number hidden neurons. However, since the optimal number of hidden units depends strongly on the nature of the problem and on the chosen representation of the input and output objects it is not save to rely exclusively on heuristics. Nevertheless we still used a simple heuristic, by taking the square root of the number of input neurons (65), which resulted in 8 hidden neurons. For the hidden units a symmetric sigmoid activation function was used, to transform the outcome of the hidden neurons. This function is a squashing S-shaped function, resulting in values between  $-1$  and  $+1$ . A linear activation function was used for transformation of the outcome of the output neurons. In order to achieve faster convergence with fewer oscillations during training, the learning rate was set to 0.05 (range 0–1) for both hidden and output neurons. This learning rate is a reduction factor for the adaptations of the weights, during back-propagation. The momentum, another factor to avoid getting trapped in local minima during training was set to zero (see paragraph 3b-II). The scaling of the input variables (absorbances) was set to a range between 0 and 1. Rescaling the input values of the spectra between 0 and 1 was performed to take care that the variability of the input values reflect their importance relative to the input values of other spectra (otherwise spectra with absorbances ranged between e.g. 0.8 and 1.3 may predominate absorbances ranged between 0.4 and 0.8). The output values were range scaled between 0.4 and 0.6 as starting points. Output scaling is necessary to prevent arithmetical overflow errors as a consequence of the use of high output values (between 0 and 100% in our case). The predicted output values are re-scaled to their normal numerical value between 0 and 100%, before presenting them as program output. The same re-scaling takes place to present the calculated RMSE values (error between the calculated and known outcome). The point at which the training should stop was set to 20000 epochs (training cycles), or when the RMSE of the training-set reached a value  $\leq 0.1\%$ . To prevent storage of the RMSE errors of every single training cycle, the step size for storage of the RMSE values of both training- and validation-set was set to 100 epochs. Because we performed a new training session, the weights were randomized with values between  $-0.1$  and  $+0.1$ , before training.

After starting training, the training process stopped after iteration with 20000 epochs. The result of this training is shown in Figure 14. From this graph it can be seen that the training error gradually decreases, but that the validation error increases after a steep descent, after a few epochs. This divergence could indicate the occurrence of outliers.

Therefore the network was retrained with the same topology settings as before, but terminated after 5000 epochs. The mean component error of each training object after this short training is shown in Figure 15. This error is expressed as the mean of the absolute differences between the expected and predicted outcome of the 3 components for each object (sample). In this graph two samples, namely object number 27 and 29, have extremely large mean component errors ( $>30\%$ ). A closer inspection of the results of this graph revealed that the quantitative compositions of two samples of the training-set were interchanged (<sup>x</sup> annotated in appendix F). After fixing this problem and re-training with 20000 epochs the training showed good convergence and a RMSE of 3.3% for the validation-set.

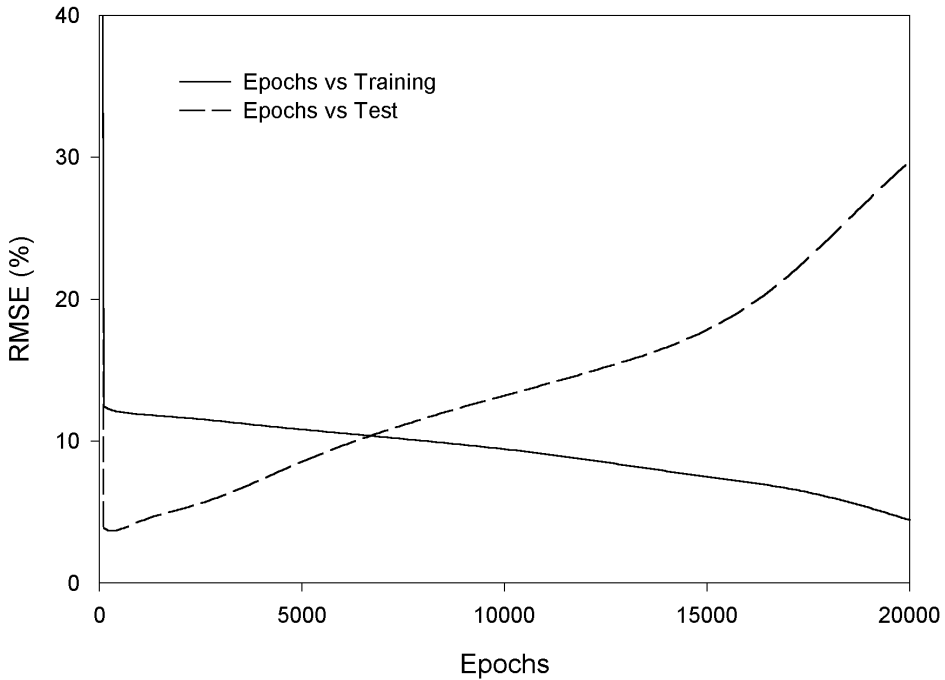


Figure 14. RMSE graph of the initial training process (test).

For further optimization of this network model different topologies were developed, which were executed by batch-automated training. The content of the initial topology was varied, by decreasing the number of hidden neurons from 8 to 2 with steps of 2. To prevent premature stopping of the training process, the stopping error was always set to 0.1%. The number of epochs for termination of the training was always set to 200000. In all cases the assessment of the quality of the network model was based on minimum RMSE value of the validation set that was reached at a certain number of epochs.

The topology with 2 hidden neurons showed to be the best. We also modified the learning rate (0.03, 0.07 and 0.1) and the range of the output scaling (0–1, 0.1–0.9, 0.2–0.8 and 0.3–0.7), but the initial setting of 0.4–0.6 showed to be most optimal. Each training session of 200000 epochs took about 5 minutes for each topology. The final topology had the following settings: 65 input neurons, 2 hidden neurons, 3 output neurons, a symmetric sigmoid transfer function for the hidden layer, a linear transfer function for the output layer, a learning rate of 0.05 between the input- and hidden-, as well as between the hidden- and output-layers, scaling of the input values between 0 and 1, and scaling of the output values between 0.4 and 0.6. The momentum was set to 0 in all cases. With this topology the RMSE steadily decreased to a minimum value of 0.25% and 2.83% for the training- and validation-sets, respectively. These minimum values were reached after a training with 101556 epochs (cycles). With this final network model the total object error of all samples of the validation-set were  $\leq 4\%$ . The RMSE graph of the final training process is shown in Figure 16.

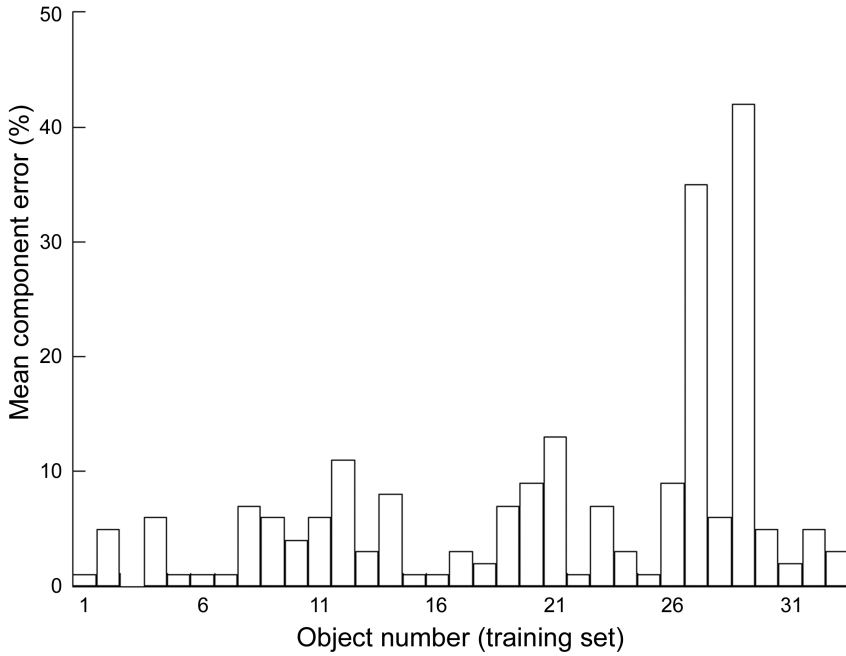


Figure 15. Mean component error of each sample (object) of the training-set.

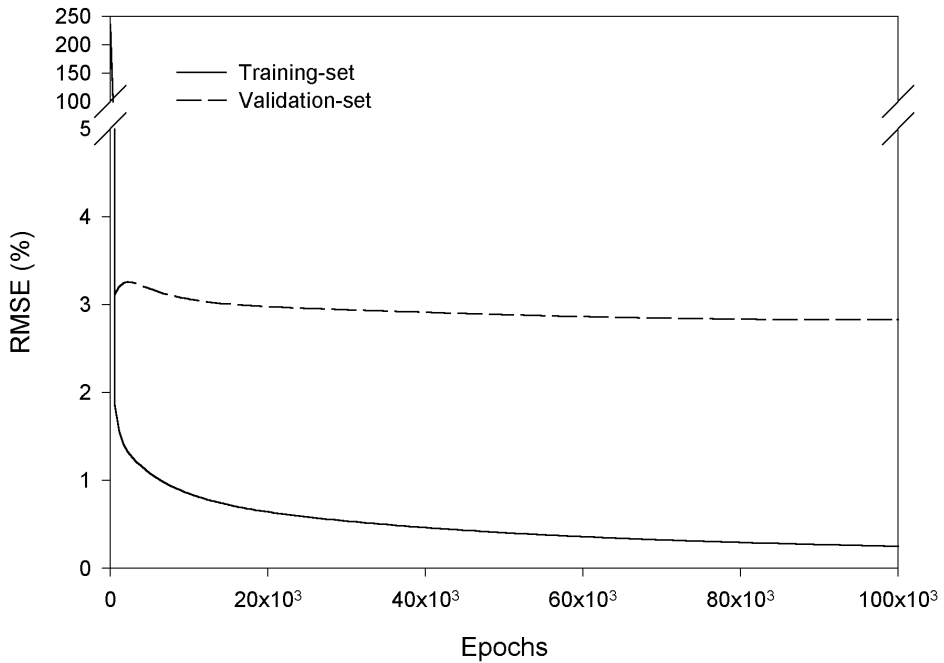


Figure 16. RMSE graph of the final training process.

To complete the method, some expert rules have to be specified (Figure 17).

Normalize	Do not use expertsystem for unknown components with values less than:	-10.00
Normalize	At normalisation, round data to zero if value is less than:	3.50
Normalize	IF c0 < 0 THEN Normalize ELSE Normalize;	

Figure 17. Expert rules.

One of the rules is intended to give users a hint to ignore network results and to use the results obtained with library search, as large negative numbers may occur when the network tries to predict samples with qualitative composition unavailable to the network model. We have set this limit to minus 10%. If one or more network results (predicted component outcome) meet the preceding criterion ( $\leq -10\%$ ), no further processing will be performed, except for showing a hint on screen. Another rule that may be set is a directive that forces the network outcome of a component to zero if the value of the component is less than a limiting value. This value was set to 3.5%. Besides the two static rules, one final rule has to be set, named 'Normalize'. The function of the rule is to round the outcome of the network predictions to the nearest 5% and to take care that the sum of the outcome is always 100%. Forcing values to zero will be performed before any data is normalized. After these settings, the final WhWeCap method may be used in the user part of the program.

After running the user part of Neuranet (data not shown), a few samples showed to have small inaccuracies<sup>1</sup> in the network outcome of calcium oxalate (whewellite and weddellite), as for samples composed of 100% carbonate apatite. The results of whewellite and weddellite were almost equal, but had opposite positive and negative values, such as 8% / -7% / 99% for whewellite/weddellite/carbonate apatite. If such network inaccuracies occur on a regular base, an expert rule should be developed. Therefore we added the following rule, named WhWeCap, to the WhWeCap method:

IF [Whew+Wedd]<5 THEN Whew=0 AND Wedd=0;

The outcome of a sample, subject to these systematic inaccuracies of network results, is shown in table 2. This table shows the outcome of the network results before and after reprocessing with the WhWeCap and Normalize expert rules.

<sup>1</sup> Inaccuracies may be caused by the following: To obtain reliable mixtures, huge amounts of pure patient material would be needed for preparation of the library and training- and validation-sets of the ANN. This would be virtually impossible. Therefore synthetic, commercially available, components were used whenever possible. Of course, the quality and spectral properties of these synthetic components were comprehensively compared to their biological counterparts in urinary calculi. On the other hand, urinary calculi may contain very small amounts of impurities (e.g. chromogens), as a consequence of their passage through the urinary tract. This may lead to very small differences in the spectra from the synthetic and biological pure components. These differences are far less than the predominant spectral features of the components. Unfortunately the neural network model does not contain this information and sometimes calculates very small amounts (%) of components, whose presence would be very unlikely in certain compositions.

Table 2. The results of a sample with small, but systematically errors before and after the application of an expert rule.

Sample results after:	whewellite	weddelite	carbonate apatite
Network processing	8%	-7%	99%
WhWeCap expert rule	0%	0%	99%
Normalize rule			100%

After addition this rule to the WhWeCap method, the method may be applied for routine use in the user part of the program. Usually a few more expert rules have to be added after a while. This is because the user will acquire better understanding of possible systematic inaccuracies of network results in the course of time. Of course this three component network model and library is insufficient for normal routine use. Therefore, normally an ANN method should be used, which is trained with the most commonly occurring components in urinary calculi (about 8 components). In addition to these commonly occurring components the library should also be provided with a number of more or less rarely occurring components.

## Appendix E: Specification of the spectra of the library

File name	Description of the composition
C:\WhWeCap\WEDDSYNT.SPC	* Weddellite (origin: synthesized)
C:\WhWeCap\WHEWELIT.SPC	* Whewellite (BDH U.N. 2811)
C:\WhWeCap\CARBONAP.SPC	* Carbonate Apatite (origin: synthesized)
C:\WhWeCap\WHEWED19.SPC	Whewellite - Weddellite 10%-90%
C:\WhWeCap\WHEWED28.SPC	Whewellite - Weddellite 20%-80%
C:\WhWeCap\WHEWED37.SPC	Whewellite - Weddellite 30%-70%
C:\WhWeCap\WHEWED46.SPC	Whewellite - Weddellite 40%-60%
C:\WhWeCap\WHEWED55.SPC	Whewellite - Weddellite 50%-50%
C:\WhWeCap\WHEWED64.SPC	Whewellite - Weddellite 60%-40%
C:\WhWeCap\WHEWED73.SPC	Whewellite - Weddellite 70%-30%
C:\WhWeCap\WHEWED82.SPC	Whewellite - Weddellite 80%-20%
C:\WhWeCap\WHEWED91.SPC	Whewellite - Weddellite 90%-10%
C:\WhWeCap\WECA1090.SPC	Weddellite - Carb Apatite 10%-90%
C:\WhWeCap\WECA2575.SPC	Weddellite - Carb Apatite 25%-75%
C:\WhWeCap\WECA4060.SPC	Weddellite - Carb Apatite 40%-60%
C:\WhWeCap\WECA5050.SPC	Weddellite - Carb Apatite 50%-50%
C:\WhWeCap\WECA6040.SPC	Weddellite - Carb Apatite 60%-40%
C:\WhWeCap\WECA7525.SPC	Weddellite - Carb Apatite 75%-25%
C:\WhWeCap\WECA9010.SPC	Weddellite - Carb Apatite 90%-10%
C:\WhWeCap\WHCA1090.SPC	Whewellite - Carb Apatite 10%-90%
C:\WhWeCap\WHCA2575.SPC	Whewellite - Carb Apatite 25%-75%
C:\WhWeCap\WHCA4060.SPC	Whewellite - Carb Apatite 40%-60%
C:\WhWeCap\WHCA5050.SPC	Whewellite - Carb Apatite 50%-50%
C:\WhWeCap\WHCA6040.SPC	Whewellite - Carb Apatite 60%-40%
C:\WhWeCap\WHCA7525.SPC	Whewellite - Carb Apatite 75%-25%
C:\WhWeCap\WHCA9010.SPC	Whewellite - Carb Apatite 90%-10%
C:\WhWeCap\WWC12127.SPC	Whe-Wed-Carb Apatite 12.5%-12.5%-75%
C:\WhWeCap\WWC12712.SPC	Whe-Wed-Carb Apatite 12.5%-75%-12.5%
C:\WhWeCap\WWC71212.SPC	Whe-Wed-Carb Apatite 75%-12.5%-12.5%
C:\WhWeCap\WWC25255.SPC	Whe-Wed-Carb Apatite 25%-25%-50%
C:\WhWeCap\WWC25525.SPC	Whe-Wed-Carb Apatite 25%-50%-25%
C:\WhWeCap\WWC52525.SPC	Whe-Wed-Carb Apatite 50%-25%-25%
C:\WhWeCap\WWC333.SPC	Whe-Wed-Carb Apatite 33%-33%-33%
C:\WhWeCap\WW82AP82.SPC	Whe-Wed-Carb Apatite 62%-16%-22%
C:\WhWeCap\WW64AP82.SPC	Whe-Wed-Carb Apatite 49%-32%-19%
C:\WhWeCap\WW46AP82.SPC	Whe-Wed-Carb Apatite 31%-46%-23%
C:\WhWeCap\WW28AP82.SPC	Whe-Wed-Carb Apatite 17%-63%-20%
C:\WhWeCap\WW82AP64.SPC	Whe-Wed-Carb Apatite 48%-12%-40%
C:\WhWeCap\WW64AP64.SPC	Whe-Wed-Carb Apatite 37%-24%-39%
C:\WhWeCap\WW46AP64.SPC	Whe-Wed-Carb Apatite 24%-36%-40%

File name	Description of the composition
C:\WhWeCap\WW28AP64.SPC	Whe-Wed-Carb Apatite 13%-47%-38%
C:\WhWeCap\WW82AP46.SPC	Whe-Wed-Carb Apatite 32%-8%-60%
C:\WhWeCap\WW64AP46.SPC	Whe-Wed-Carb Apatite 25%-16%-59%
C:\WhWeCap\WW46AP46.SPC	Whe-Wed-Carb Apatite 16%-24%-60%
C:\WhWeCap\WW28AP46.SPC	Whe-Wed-Carb Apatite 9%-32%-60%
C:\WhWeCap\WW82AP28.SPC	Whe-Wed-Carb Apatite 16%-4%-80%
C:\WhWeCap\WW64AP28.SPC	Whe-Wed-Carb Apatite 13%-8%-79%
C:\WhWeCap\WW46AP28.SPC	Whe-Wed-Carb Apatite 8%-12%-80%
C:\WhWeCap\WW28AP28.SPC	Whe-Wed-Carb Apatite 5%-17%-78%
C:\WhWeCap\0514-2C.SPC	Whewellite-Weddellite 35%-65%

## Appendix F: Specification of spectra for training and validation of the neural network

File name	Train = 0			
	Test = 1	whew	wedd	carb
C:\WhWeCap\WEDDSYNT.SPC	0	0.0	100.0	0.0
C:\WhWeCap\WHEWELIT.SPC	0	100.0	0.0	0.0
C:\WhWeCap\CARBONAP.SPC	0	0.0	0.0	100.0
C:\WhWeCap\WHEWED19.SPC	0	11.0	89.0	0.0
C:\WhWeCap\WHEWED28.SPC	0	20.0	80.0	0.0
C:\WhWeCap\WHEWED37.SPC	0	30.0	70.0	0.0
C:\WhWeCap\WHEWED46.SPC	0	40.0	60.0	0.0
C:\WhWeCap\WHEWED55.SPC	0	50.0	50.0	0.0
C:\WhWeCap\WHEWED64.SPC	0	60.0	40.0	0.0
C:\WhWeCap\WHEWED73.SPC	0	70.0	30.0	0.0
C:\WhWeCap\WHEWED82.SPC	0	80.0	20.0	0.0
C:\WhWeCap\WHEWED91.SPC	0	90.0	10.0	0.0
C:\WhWeCap\WECA1090.SPC	0	0.0	10.0	90.0
C:\WhWeCap\WECA2575.SPC	0	0.0	24.0	76.0
C:\WhWeCap\WECA4060.SPC	0	0.0	40.0	60.0
C:\WhWeCap\WECA5050.SPC	0	0.0	51.0	49.0
C:\WhWeCap\WECA6040.SPC	0	0.0	60.0	40.0
C:\WhWeCap\WECA7525.SPC	0	0.0	75.0	25.0
C:\WhWeCap\WECA9010.SPC	0	0.0	90.0	10.0
C:\WhWeCap\WHCA1090.SPC	0	10.0	0.0	90.0
C:\WhWeCap\WHCA2575.SPC	0	25.0	0.0	75.0
C:\WhWeCap\WHCA4060.SPC	0	40.0	0.0	60.0
C:\WhWeCap\WHCA5050.SPC	0	50.0	0.0	50.0
C:\WhWeCap\WHCA6040.SPC	0	60.0	0.0	40.0
C:\WhWeCap\WHCA7525.SPC	0	75.0	0.0	25.0
C:\WhWeCap\WHCA9010.SPC	0	90.0	0.0	10.0
C:\WhWeCap\WWC12127.SPC <sup>x</sup>	0	74.0	13.0	13.0
C:\WhWeCap\WWC12712.SPC	0	13.0	74.0	13.0
C:\WhWeCap\WWC71212.SPC <sup>x</sup>	0	13.0	12.0	75.0
C:\WhWeCap\WWC25255.SPC	0	24.0	25.0	51.0
C:\WhWeCap\WWC25525.SPC	0	25.0	50.0	25.0
C:\WhWeCap\WWC52525.SPC	0	50.0	25.0	25.0
C:\WhWeCap\WWC333.SPC	0	33.0	34.0	33.0
C:\WhWeCap\MONSTER1.SPC	1	64.0	10.0	26.0
C:\WhWeCap\MONSTER2.SPC	1	20.0	36.0	44.0
C:\WhWeCap\MONSTER3.SPC	1	17.0	10.0	73.0
C:\WhWeCap\MONSTER4.SPC	1	42.0	11.0	47.0
C:\WhWeCap\MONSTER5.SPC	1	33.0	56.0	11.0
C:\WhWeCap\MONSTER6.SPC	1	43.0	21.0	36.0
C:\WhWeCap\MONSTER7.SPC	1	27.0	14.0	59.0
C:\WhWeCap\WW82AP82.SPC	1	62.0	16.0	22.0
C:\WhWeCap\WW64AP82.SPC	1	49.0	32.0	19.0
C:\WhWeCap\WW46AP82.SPC	1	31.0	46.0	23.0
C:\WhWeCap\WW28AP82.SPC	1	17.0	63.0	20.0
Etc.				

<sup>x</sup> compositions interchanged



## Summary

The **scope** of this thesis was to further develop the assays of fecal fat and urinary calculi. The developments mainly consisted of the investigation of the applicability of new infrared spectroscopic methods and the application of chemometrical methods for quantification. Furthermore, we gave information about the pathophysiology background of our studies.

The quality of some traditional clinical chemical methods needs improvement. Therefore, studies were undertaken to develop new methods for the determination of fecal fat and the composition of urinary calculi. For these new methods, infrared spectroscopy has been used for the determination of the components of interest. To keep the complexity of the analytical methods as low as possible, the measurements of the components were performed on authentic sample material whenever possible. A drawback of this approach was that prediction of the quantity of the sample components was more difficult due to the presence of interfering substances in the sample material. Chemometric techniques were applied to solve difficulties with the prediction of the chosen component concentrations caused by interfering substances. In the majority of the developed methods, the analyte concentrations were predicted by means of multivariate calibration methods such as partial least-squares (PLS) regression and artificial neural networks (ANN).

The **Introduction** describes the pathophysiological background, the analytical chemical techniques and some post-analytical quantification methods (chemometrics), which may provide background information about the analysis of fecal fat and the prediction of the composition the urinary calculus. The introduction does not only describe the analytical methods that were subject of the studies, but also some related reference methods. The interpretation of the results obtained from some analytical methods used in clinical laboratories has become more and more complex. This is e.g. caused by the increased demand for measurement of analytes in un-pretreated sample material. This has e.g. led to the development of the so-called chemometrical methods in the past decades. In our studies we applied chemometrics for multivariate calibration of spectral data. For the multivariate calibration, we especially made use of artificial neural networks and partial least-squares regression.

**Part I** describes the development of modern analytical methods for the determination of fecal fat. The determination of fecal fat is important in the diagnosis of steatorrhea. Steatorrhea is a malabsorption syndrome in which the digestion or absorption of dietary fat in the gastro-intestinal tract is impaired and which subsequently will lead to increased amounts of fat in stool. Emphasis was given to the use of infrared spectroscopy, the development of a new reference method and a fast method for the analysis of fecal fat.

*Chapter 1* describes the development of a new method for the determination of fecal fat. Immediate cause of this study was the finding of a multi-center study that the fecal fat results from 28 patient samples obtained with the traditional titrimetric Van de Kamer method were different among the hospital laboratories. With the new method, the fatty acids were extracted from the stool samples with acidified petroleum ether-ethanol and the extracts were dried and dissolved in chloroform. An infrared spectrum was recorded in the range from 4000–650  $\text{cm}^{-1}$  using a transmission cell. Standard mixtures of stearic and palmitic acids (65:35) were used for calibration. Quantification was based on the absorbance band of the  $\text{CH}_2$  group at 2855  $\text{cm}^{-1}$ . The correlation coefficient between the

Van de Kamer method and the infrared spectroscopic method was 0.96 ( $y=1.12x - 0.02$ , standard error of prediction= 0.89 g%). No significant differences were found when the infrared results of the 28 fecal samples from patients were compared between three different university hospital laboratories. The new infrared method was simple and rapid, and provided sufficient intra- and inter-laboratory precision for the diagnosis of steatorrhea.

*Chapter 2* describes a further improvement of the method as described in chapter 1. The method in chapter 1 described that fecal fat was extracted from stool samples with petroleum ether-ethanol. After extraction, the petroleum ether was dried, and the fatty acids were redissolved in chloroform before measurement. Because this extraction procedure was still rather time-consuming, it was replaced with a single chloroform extraction. The results of 111 patient samples were analyzed with the simplified extraction method and evaluated by comparing the results with the results of the same samples obtained by the petroleum ether-ethanol extraction method. The results of both methods showed good agreement ( $r=0.991$ ,  $y=1.055+0.24$  and a standard error of prediction 0.365 g%). It was concluded that the new simplified extraction procedure for fecal fat determination gives similar results to the earlier described extraction procedure and allows considerable reduction in analysis time, use of chemicals and technical equipment.

In *chapter 3* the applicability of a mid-infrared (MIR) spectroscopic method, using an attenuated total reflection (ATR) accessory and a new near-infrared (NIR) spectroscopic method for the analysis of fecal fat, were investigated. For the NIR method, sealed plastic bags containing the stool samples were used as transmission cells. Standardization was obtained by using the infrared method described in chapter 2 as a reference procedure. Partial least squares regression was used for calibration of the new methods. Fifteen per cent of the stool samples could not be estimated with the ATR method within reasonable accuracy limits compared to the reference method. The standard error of prediction of the NIR method was 1.1 g%. We concluded that the new NIR method is a promising technique for routine use. It was further concluded that further experiments have to be done with triplicate measurements of each sample and the use of an external validation set.

**Part II** describes the development of methods for the determination of the composition of urinary calculi. Urinary calculi (renal stones) have plagued man over the centuries. A serious problem of renal stones is that they tend to recur. Therefore, the analysis of the composition of the calculi is important for proper treatment of patients with urolithiasis, especially to prevent the recurrence of stones. In the past, the analysis of urinary stones has been performed with tradition wet chemical analysis. This method, however, is inaccurate and imprecise. In the past few years, the wet chemical analysis has been replaced by infrared spectroscopic methods. Unfortunately, the quantitative assessment of urinary calculus constituents by infrared analysis (IR) is hampered by the need of expert knowledge for spectrum interpretation. Therefore, we developed post-analytical methods by application of some chemometric techniques in order to simplify the determination of the quantitative composition of an urinary calculus. We also developed a new infrared spectroscopic method.

*Chapter 4* describes the development of a post-analytical method for the prediction of the three most frequently occurring components of urinary calculi. The method was based on PLS regression, which was applied to infrared KBr transmission spectra. The spectra were scanned

in the mid-infrared region from 4000–400  $\text{cm}^{-1}$ . The calibration set for PLS regression was composed of 25 artificial mixtures of whewellite, weddellite and carbonate apatite in different proportions. The value of PLS analysis was investigated by the assay of a different set of known artificial mixtures, and selected patient samples from which the composition was determined by computerized library search followed by visual interpretation of the band intensities for a more precise determination of the sample composition. Library search was applied by computerized search in several libraries, containing 235 reference spectra from various mixtures with different proportions. Compared to this method, PLS analysis was found to be superior with respect to accuracy and necessity of expert knowledge. From these results, it was concluded that PLS regression is a promising tool for routine quantification, not only for whewellite, weddellite and carbonate apatite, but also for other compositions of the urinary calculus.

*Chapter 5.* In the previous chapter it was described that library search still must be followed by visual interpretation of the band intensities for a more precise determination of the urinary calculus composition. To minimize the need for this last step, an artificial neural network model was developed and tested for the quantification of the eight most frequently occurring components of urinary calculi. Various samples were used as a training set, by the preparation of several binary and ternary mixtures of ammonium hydrogen urate, brushite, carbonate apatite, cystine, struvite, uric acid, weddellite, and whewellite. In addition, a number of known artificial mixtures were assayed for evaluation, as well as a number of selected patients' samples from which the compositions were determined by computerized library search followed by visual interpretation. Neural network analysis was found to be more accurate than library search and required less expert knowledge because careful visual inspection of the band intensities could be omitted. It was concluded that neural networks are promising tools for routine quantification of urinary calculus compositions.

*Chapter 6* describes the development of a new infrared method for urinary calculus analysis that made use of a so-called, Golden Gate sample holding device, and which was equipped with a Single Reflection Diamond attenuated total reflection (ATR) crystal for measurement of micro samples. An amount of 1–2 mg of carefully pulverized material was applied onto the flat surface of the diamond crystal and pressurized before measurement. This method was used as a replacement of the traditional infrared method making use of KBr tablets. The reason for this was the fact that the preparation of KBr tablets, as applied in chapter 4 and 5, is still time-consuming and often hampered by pellet breakage. A combination of computerized library search and an artificial neural network (ANN) for spectral interpretation was used for the prediction of the sample composition. The library was prepared from 25 pure components and 236 binary and ternary mixtures of the eight most commonly occurring components. The ANN was trained and validated with 248 similar mixtures and tested with 92 patient samples, respectively. The optimum ANN model yielded Root Mean Square Errors of 1.5% and 2.3% for the training and validation set, respectively. Fourteen simple expert rules were added to correct small systematic network inaccuracies. Results of 92 consecutive patient samples were compared with those of an infrared method with KBr tablets based on initial computerized library search, followed by visual inspection. Using Altman and Bland plots, the bias proved significantly different from zero for brushite (–0.8%) and the concomitantly occurring whewellite (–2.8%) and weddellite (3.8%), but not for the other components. The 95% level of

agreement of all results amounted to 9%. It was concluded that the new Golden Gate method is superior because of its smaller sample size, user-friendliness, robustness and speed. Expert knowledge for spectral interpretation is minimized by the combination of library search and ANN prediction, but visual inspection remains necessary.

*Chapter 7* describes a computer program, called Neuranet, which was especially developed for the determination of the quantitative composition of urinary calculi from infrared spectra. The spectra had to be measured in the mid-infrared region between 4000–400  $\text{cm}^{-1}$ . The program was equipped with tools for computerized library search and artificial neural network (ANN) prediction. In addition the program contains a subroutine for definition of a number of user definable expert rules intended to improve the accuracy of the neural network predicted compositions. The network part of this program was a three-layer so-called back-propagation network that can predict the relative composition of maximally ten components in a single sample. Besides a general description of the 'Neuranet' program, the chapter contains examples of the prediction process of a number of patient samples. In addition, the chapter also contains a number of appendices describing the program specifications, the development of a Neuranet method (e.g. network training) and a syntax description of the expert rules.



## Samenvatting

Opzet van dit proefschrift was het verder ontwikkelen van methodes voor de bepaling van vet in feces en het vaststellen van de samenstelling van urinewegstenen. De ontwikkeling bestond voornamelijk uit het onderzoeken van de toepasbaarheid van nieuwe infrarood spectroscopische methodes en het gebruiken van chemometrische technieken voor het kwantificeren van de resultaten. Verder is er informatie gegeven over de pathofysiologische achtergrond van de onderzochte studies.

De kwaliteit van enkele traditionele klinisch chemische analyse methoden laat te wensen over. Om deze reden hebben wij ons bezig gehouden met de ontwikkeling van nieuwe methodes voor de analyse van vet in feces en het vaststellen van de samenstelling van nierstenen. Voor de in dit proefschrift beschreven methodes werd gebruik gemaakt van infrarood spectroscopie voor het bepalen van de componenten in het patiëntenmateriaal. Om de bewerkelijkheid van de methode zo klein mogelijk te houden zijn de metingen zo veel mogelijk direct uitgevoerd in het onbewerkte patiënten materiaal. Het nadeel van deze aanpak is dat de vaststelling van de gehalten van de componenten bemoeilijkt wordt door de aanwezigheid van storende (interfererende) componenten in het materiaal. Om het probleem van de interfererende componenten op te lossen is gebruik gemaakt van chemometrische technieken voor het vaststellen van de gehalten van de gewenste componenten. Hierbij is meestal gebruik gemaakt van multivariate calibratie methoden, zoals partial least-squares (PLS) regressie en kunstmatige neurale netwerken.

De **Introductie** beschrijft de pathofysiologische achtergrond, de analytisch chemische methodieken en enkele postanalytische kwantificering methoden, met als doel om meer inzicht te verschaffen in de analyse van vet in feces en het vaststellen van de samenstelling van nierstenen. De introductie beschrijft naast de gebruikte analytische methodes ook enkele daaraan gerelateerde referentie methodes. De interpretatie van de resultaten van sommige analytische methoden in het klinisch chemische laboratorium is in veel gevallen complexer geworden. Dit wordt bijvoorbeeld veroorzaakt door het feit dat er een toenemende behoefte bestaat om de gehalten van de stoffen in het onbewerkte monstermateriaal te meten. Dit heeft onder andere weer geleid tot de ontwikkeling van de zogenaamde chemometrische methoden in de afgelopen decennia. Voor onze studies hebben wij gebruik gemaakt van chemometrische technieken en wel in het bijzonder voor de multivariate kalibratie van onze spectrale meetgegevens. Voor de multivariate kalibratie maakten wij gebruik van kunstmatige neurale netwerken en 'partial least-squares' regressie.

**Deel 1** beschrijft de ontwikkeling van enkele moderne analyse methodes voor de bepaling van vet in feces. De analyse van vet in feces is van belang voor de vaststelling van de diagnose van steatorrhoea. Steatorrhoea is een aandoening waarbij het via de voeding genuttigde vet slecht wordt verteerd, of slecht wordt geabsorbeerd in de dunne darm, wat vervolgens leidt tot een verhoogde uitscheiding van vet in de feces. Er werd gebruik gemaakt van infrarood spectroscopie en er werd een nieuwe referentie methode ontwikkeld. Verder werd er een snelle methode voor de analyse van fecaal vet bestudeerd.

*Hoofdstuk 1* beschrijft de ontwikkeling van een nieuwe methode voor de analyse van vet in feces. Aanleiding van dit onderzoek was de bevinding dat de resultaten van 28 patiënten monsters, waarvan het vet in feces was gemeten met de traditionele titrimetrische Van de Kamer methode in een drietal Academische Ziekenhuizen, niet goed overeenkwamen. Met de nieuwe methode werden de vetzuren geëxtraheerd uit de feces met behulp van een



aangezuurd mengsel van petroleum ether-ethanol. Na extractie werd de petroleum ether laag gedroogd en werden de vetzuren opgelost in chloroform. Na overbrengen van de chloroform in een transmissie cel, werd een infrarood spectrum opgenomen in het golflengte gebied van 4000–650  $\text{cm}^{-1}$ . Voor de calibratie werd gebruik gemaakt van standaard mengsels met oplopende concentraties en bestaande uit de vetzuren stearinezuur en palmitinezuur (65:35). De  $\text{CH}_2$  infrarood absorptie band bij 2855  $\text{cm}^{-1}$  werd gebruikt voor de kwantificering. De vergelijking tussen de met de Van de Kamer en de middels infrarood spectroscopie verkregen resultaten van 97 patiënten in één van de academische centra gaf een correlatie te zien van  $r=0.96$  ( $y=1.12x - 0.02$ , standaard schattingsfout 0.89 g%). Er werd geen significant verschil gevonden wanneer de infrarood resultaten van de 28 patiënten samples werd vergeleken tussen drie academische ziekenhuis laboratoria. De nieuwe methode is eenvoudig en snel uit te voeren en heeft voldoende intra- en interlaboratorium precisie om gebruikt te kunnen worden voor de diagnose van steatorrhoea.

*Hoofdstuk 2* beschrijft een verdere verbetering van de in hoofdstuk 1 beschreven methode. Deze methode beschreef de extractie met petroleum ether-ethanol, waarna de vetzuren werden heropgelost in chloroform na drogen. Deze methode is nog steeds tamelijk tijdrovend en werd vervangen door een enkelvoudige chloroform extractie van het feces monster. De resultaten van 111 patiënten monsters werden geanalyseerd met de verbeterde extractie methode en geëvalueerd door deze te vergelijken met de resultaten van dezelfde monsters die geëxtraheerd waren met de petroleum ether-ethanol methode. De resultaten van beide methodes vertoonden een goede mate van overeenstemming ( $r=0.991$ ,  $y=1.055+0.24$  en een standaard schattingsfout van 0.365 g%). De nieuwe vereenvoudigde extractie procedure voor de analyse van vet in feces geeft een aanzienlijke besparing in tijd, chemicaliën en te gebruiken apparatuur.

In *hoofdstuk 3* werd de toepasbaarheid van de analyse van vet in feces middels een mid-infrarood (MIR) spectroscopische methode bestudeerd, waarbij gebruik gemaakt werd van een zogenaamd ‘attenuated total reflection’ (ATR) accessoire. Verder werd de toepassing van een nieuwe ‘nabij-infrarood’ (NIR) spectroscopische methode onderzocht. Bij de NIR methode werden dichtgelaste plastic zakjes, met daarin het patiënten materiaal, gebruikt als transmissie cellen. Voor de standaardisatie werd gebruik gemaakt van de in hoofdstuk 2 beschreven infrarood methode als referentie methode. Bij beide nieuwe methodes werd gebruik gemaakt van partial least-squares regressie voor de calibratie. Helaas vertoonde 15% van de patiënten monsters, geanalyseerd met de ATR methode, te grote afwijkingen als deze vergeleken werden met de resultaten van de referentie analyse. De standaard schattingsfout van de NIR methode was 1.1. g%. Geconcludeerd werd dat de nieuwe NIR methode een veelbelovende techniek is voor het routinematig analyseren van vet in onbewerkte feces. Aanvullende experimenten zijn echter nodig en de methode moet nog gevalideerd worden met een externe (onafhankelijke) validatie set, met bij voorkeur triplo metingen van ieder monster.

**Deel II** beschrijft de ontwikkeling van postanalytische methoden voor de analyse van de samenstelling van urineweg stenen. De mensheid wordt al eeuwenlang geteisterd door urinewegstenen. Een bekend probleem van urinewegstenen is dat ze de neiging hebben om

terug te komen na verloop van tijd. Daarom is het van belang om de juiste samenstelling van de urinewegstenen vast te kunnen stellen voor een adequate behandeling van de patiënt en wel in het bijzonder om herhaling van urinewegsteenvorming te voorkomen. In het verleden werd de analyse veelal uitgevoerd door middel van natchemische analyse. Deze methode is echter niet erg juist en reproduceerbaar. De laatste jaren is deze methode in veel laboratoria vervangen door een infrarood spectroscopische analyse. De interpretatie van de spectra van urinewegstenen is echter niet altijd even eenvoudig. Daarom hebben wij postanalytische methodes ontwikkeld met behulp van enkele chemometrische technieken om het vaststellen van de kwantitatieve samenstelling van de urinewegstenen te vergemakkelijken. Verder hebben wij gekeken naar de mogelijkheid van een verdere optimalisatie van de infrarood spectroscopische analyse.

*Hoofdstuk 4* beschrijft de ontwikkeling van een postanalytische methode voor het vaststellen van de samenstelling van de drie meest voorkomende componenten, whewelliet, weddelliet en carbonaat apatiet, in urinewegstenen. De methode was gebaseerd op PLS regressie van spectra die waren opgenomen met behulp van KBr transmissie tabletten in het midden-infrarood gebied ( $4000\text{--}400\text{ cm}^{-1}$ ). De calibratie set van de PLS regressie bestond uit 25 synthetische mengsels van whewelliet, weddelliet en carbonaat apatiet, ieder met een verschillende relatieve samenstelling. De waarde van de PLS methode werd geëvalueerd door het analyseren van een afzonderlijke set met synthetische mengsels en met een set geselecteerde patiënten monsters waarvan de samenstelling werd vastgesteld via computer 'library search' gevolgd door een visuele beoordeling van de spectrale band intensiteiten voor een meer exacte vaststelling van de samenstelling. De library search werd uitgevoerd door verschillende bibliotheken te doorzoeken, met daarin 235 referentie spectra met verschillende kwalitatieve samenstellingen en in verschillende verhoudingen. Vergeleken met deze methode werd PLS regressie beter bevonden met betrekking tot zowel de juistheid van de uitslag als de vereiste noodzaak van expert kennis. Er werd geconcludeerd dat PLS regressie goed bruikbaar was als postanalytische methode voor routinematige kwantificering van de samenstelling van urinewegstenen, niet alleen voor whewelliet, weddelliet en carbonaat apatiet, maar ook voor andere samenstellingen van urinewegstenen.

*Hoofdstuk 5.* In het voorgaande hoofdstuk was reeds beschreven dat library search doorgaans gevolgd moet worden door een visuele interpretatie van de band intensiteiten voor het vaststellen van de juiste samenstelling van de urinewegstenen. Om de noodzaak van deze laatste stap te minimaliseren werd een kunstmatig neurale netwerk model ontwikkeld en uitgetest voor het vaststellen van de samenstelling van de urinewegstenen. Er werden verschillende monsters gemaakt voor de training set door het bereiden van binaire en ternaire mengsels bestaande uit ammonium waterstofuraat, brushiet, carbonaat apatiet, cystine, struviet, urinezuur, weddelliet en whewelliet. Daarnaast werd er een aantal kunstmatige mengsels geanalyseerd voor evaluatie, evenals een aantal geselecteerde patiënten monsters, waarvan de samenstelling vooraf was vastgesteld met behulp van library search gevolgd door een visuele interpretatie van de spectra. De resultaten van de neurale netwerk analyse werd meer juist bevonden dan die van library search en bovendien bleek de expertkennis, nodig voor het beoordelen van de spectrale bandintensiteiten, vrijwel nooit noodzakelijk. Er werd geconcludeerd dat neurale netwerken goed gebruikt kunnen worden voor het routinematige analyseren van de samenstelling van urinewegstenen.

*Hoofdstuk 6* beschrijft de ontwikkeling van een nieuwe infrarood methode voor de analyse van urinewegstenen. Hierbij werd gebruik gemaakt van een zogenaamde ‘Golden Gate monsterhouder die uitgevoerd was met enkelvoudige reflectie ATR en geschikt is voor het analyseren van zeer kleine hoeveelheden monstermateriaal. De infrarood meting werd uitgevoerd na het aanbrengen en onder druk brengen van 1 tot 2 mg zorgvuldig vermalen monster op het vlakke oppervlak van het diamant kristal. Deze methode was een vervanging van de gebruikelijke methode, die gebruikt maakt van KBr tabletten. Aanleiding voor de ontwikkeling van de nieuwe methode was het feit dat de bereiding van KBr tabletten, zoals beschreven in hoofdstuk 4 en 5, tamelijk tijdrovend was en bovendien bleken de tabletten tamelijk gevoelig voor breuk. Voor het vaststellen van de samenstelling van de urinewegstenen werd gebruik gemaakt van achtereenvolgens library search en een kunstmatig neuraal netwerk model. De bibliotheek bestond uit referentiespectra van 25 zuivere stoffen en 236 binaire en ternaire mengsels van de acht meest voorkomende componenten in nierstenen. Het netwerk werd getraind en gevalideerd met 248 soortgelijke mengsels en vervolgens getest met 92 patiënten monsters. De totale fout van het optimaal getrainde netwerk model was respectievelijk 1.5% en 2.3% voor de training en de validatie set. Verder werden nog 14 expert regels beschreven om kleine afwijkingen van de neurale netwerk uitkomst te corrigeren. De resultaten van de 92 patiënten monsters werden vergeleken met de resultaten van spectra geanalyseerd met de KBr methode. De samenstelling van de KBr spectra werd vastgesteld aan de hand van library search, gevolgd door een visuele beoordeling van de spectra. Door gebruik te maken van Altman en Bland grafieken werd alleen een significante afwijking tussen de resultaten van beide methodes gevonden voor brushiet (-0.8%) en het gelijktijdig voorkomende whewelliet (-2.8%) en weddeliet (3.8%). De gemiddelde afwijking van alle resultaten was 9%. Geconcludeerd werd, dat de Golden Gate methode de voorkeur verdiende boven de KBr methode, vanwege de kleine hoeveelheid benodigd monster, de gebruikersvriendelijkheid, de robuustheid en de snelheid van de analyse. Hoewel de benodigde hoeveelheid expert kennis kleiner was dan bij gebruik van uitsluitend library search, werd toch geconcludeerd dat enige visuele interpretatie noodzakelijk blijft bij gebruik van de nieuwe methode.

*Hoofdstuk 7* beschrijft een programma, genaamd Neuranet, dat speciaal is ontwikkeld voor het kwantificeren van de samenstelling van urinewegstenen aan de hand van infrarood spectra. Deze spectra moeten gemeten zijn in het midden-infrarood gebied van 4000–400  $\text{cm}^{-1}$ . Het programma is uitgerust met een aantal gereedschappen, zoals library search en een neuraal netwerk. Daarnaast bevat het programma een faciliteit waarmee de gebruiker een aantal expert regels kan definiëren om de nauwkeurigheid van de netwerkuitkomsten te verbeteren. Het gebruikte netwerk bestaat uit drie lagen en is gebaseerd op het zogenaamde ‘back-propagatie’ principe. Het netwerk model kan in een urinewegsteen de relatieve samenstelling bepalen van maximaal tien componenten. Het hoofdstuk bevat naast een algemene beschrijving van het programma ook voorbeelden van het vaststellen van de samenstelling van een aantal van patiënten afkomstige urinewegstenen. Daarnaast bevat het hoofdstuk enkele appendices die de specificaties van het programma beschrijven, zoals de ontwikkeling van een Neuranet methode (b.v. netwerk training) en een syntax beschrijving van de expertregels.



Bij het schrijven van dit hoofdstuk dacht ik al vrij snel aan de geveleugelde uitspraak: 't Wordt nooit 'n tien. Ook mij schieten soms woorden tekort om een buitengewone bewondering uit te spreken over zaken waar superlatieven zeker op hun plaats zijn. Ik hoop dan ook dat mogelijke misconcepties mij in dit verband vergeven zullen worden.

Dit proefschrift zou niet tot stand zijn gekomen zonder de medewerking van vele personen. Ik wil graag iedereen bedanken die, direct of indirect, bij de totstandkoming betrokken zijn geweest. Enkelen wil ik hier met name noemen.

Mijn promotor Prof. Dr. S. Poppema wil ik bedanken voor de ruimte die ik kreeg voor mijn onderzoek ten tijde van mijn overgang naar de afdeling Pathologie en zijn begeleiding bij het afronden van dit proefschrift. De leden van de beoordelingscommissie, Prof. Dr. F.A.J. Muskiet, Prof. Dr. A.K. Smilde en Prof. Dr. J.L. Willems ben ik erkentelijk voor hun voortvarendheid met het doorlezen van het manuscript.

Mijn referent en begeleider Dr. B.G. Wolthers wil ik bedanken voor zijn onafgebroken inzet gedurende mijn promotieonderzoek en ook daarbuiten. Bert, jij bent de grondlegger geweest van dit onderzoek. Zonder jouw enthousiasme en kennis op het gebied van de analytische chemie was dit proefschrift absoluut niet tot stand gekomen. Ditzelfde geldt ook voor mijn tweede referent en begeleider Dr. I. P. Kema. Ido, jouw inzet was met name groot tijdens de latere fase van mijn onderzoek. Zonder jouw stimulans was dit proefschrift er mogelijk niet gekomen. Aan mensen als jij is altijd gebrek.

Prof. Dr. Frits A.J. Muskiet, reeds eerder genoemd, wil ik bedanken voor zijn scherpzinnige en vakkundige inbreng gedurende mijn gehele onderzoeksloopbaan. Frits, jouw wetenschappelijke kennis, gedrevenheid en kritische benadering van onderzoek zijn zeer leerzaam voor mij geweest.

Prof. Dr. W. van der Slik en Dr. F.R. Hindriks hebben aan de bakermat van mijn onderzoek gestaan. Beide hebben zij veel betekend voor de chemometrie in het toenmalige Centraal Klinisch Chemische Laboratorium. Ook de kortstondige, maar intensieve en inspirerende samenwerking met de leden van de toenmalige vakgroep Chemometrie (o.a. Dr. Pierre M.J. Coenegracht, Dr. Annabel Bolck en Harm Metting) van de farmaceutische faculteit van de RUG heb ik als bijzonder plezierig en leerzaam ervaren.

Gedurende de diverse stadia van het onderzoek was de steun van met name Theo de Haan (urineuwegsteen analyse) en Anneke Kingma (vetanalyse) onontbeerlijk. Jullie waren altijd bereid tot het enthousiast analyseren van eindeloos grote series, ook als het weer eens nodig werd gevonden om de zoveelste analysevariant uit te proberen. Peter Koops ben ik dankbaar voor zijn geweldige ondersteuning bij het ontwikkelen van het Neuronet programma.

Voor Peter Borsboom heb ik een grote bewondering. Jouw creativiteit beperkt zich niet alleen tot het theoretisch uitdenken van nieuwe analyseapparaten, maar ook tot het praktisch tot stand komen daarvan daarvan. De technische vooruitgang is gebaat bij creatieve geesten zoals jij.

Dr. Bernadette S. Jakobs, Dr. Dorien W. Swinkels en Marc Hofs (allen van het klinisch chemisch laboratorium van het St. Radboud ziekenhuis in Nijmegen), evenals Dr. Ron A.M. Voorbij, Stella Donkervoort en Marjan Joosting (allen van het klinisch chemisch laboratorium van het Academisch Ziekenhuis in Utrecht) wil ik bedanken voor de samenwerking op het gebied van de vetanalyse. Deze samenwerking heb ik als zeer plezierig ervaren en toont des te meer aan dat onderzoek tussen diverse onderzoekscentra heel goed mogelijk is.

Ook Dr. Henk M.J. Goldschmidt en Jules de Vries (beide St. Elisabeth Hospital in Tilburg) wil ik bedanken voor de buitengewoon plezierige coöperatie op het gebied van de urinewegsteen analyse. Henk, ik heb bewondering voor je inzet op het gebied van de automatisering in de klinische chemie en de chemometrie in het bijzonder.

Dr. Joost C.J.M. Swaanenburg, ex kamergenoot, wil ik bedanken voor de vele genoeglijke momenten. Joost je was me toch nog te vlug af. Ook Eddie Ligeon, paranimf en tevens kamergenoot in lang vervlogen tijden wil ik hartelijk bedanken. Eddie je onverwoestbare enthousiasme is haast onvervangbaar.

Het onderzoek maakte slechts een beperkt onderdeel uit van mijn dagelijkse beslommingen. Een veelheid aan mensen heeft het daarom onwillekeurig toch mogelijk gemaakt om dit onderzoek tot een goed einde te brengen. Zonder al deze personen bij naam te noemen wil ik toch een ieder bedanken voor zijn of haar al dan niet directe bijdrage aan mijn onderzoek. Zonder anderen tekort te willen doen geldt dit toch wel in het bijzonder voor de medewerkers en oud medewerkers van het laboratorium voor bijzondere bepalingen (KCSB). Mijn dank gaat vanzelfsprekend ook uit naar het thuisfront.

Marcel Volmer