Introduction

# Bacterial contamination of drinking water

In spite of the efforts put into maintaining drinking water safe, we still face many bacterial water contaminations, from bacteria such as Legionella, Salmonella, Escherichia that cause gastroenteritis, skin disease and other ailments (Leclerc 2002). In the US alone, ~40,000 people are hospitalized due to microbial contamination of water, resulting in 970,000 dollars of damages annually (collier 2012). The source of these contaminations may be direct contamination of water sources, pipelines or home and institutional taps or failures in the sanitation system in bottling and water processing plants.

Water treatment plants use different methods to prevent microbial contamination of drinking water. The water may be filtered by several types of filters, chemically sanitized (by chlorination or ozonation) and sometimes even reverse osmosis is used to insure water safety. Further down the supply chain the water is sometimes re-filtered to get rid of any contaminants that arrive from piping contaminators and in some homes and institutions an in-house treatment is done using micro-filters, active carbon and UV light sanitation. In recent years we have seen more of these in-house sanitation approaches in home “water bars”. In spite of all these efforts, pathogenic bacteria sometimes reach consumers. In these cases, the local authorities would often redirect water distribution until the contamination is either disinfected or naturally subside (Ashbolt 2015, MinOH 2013, MinOH 2016, MinOH 2017).

Due to the dangers of water contamination, constant monitoring of the microbial quality of water is required. Timely identification and treatment of contaminations can prevent pathogens from reaching consumers, saving a lot of money and improving public health. The water industry uses many different methods to detect contaminations, such as turbidity measurement and culture-based selective assays that detect coliforms, fecal coliforms and fecal streptococci. These tests are done throughout the water treatment process and in many control points in pipelines. This main issue with these tests is that the time to get results is long, which means that in practice – contaminated water reaches consumers. Furthermore, the standard tests are usually not specific, or are specific only to indicators (such as coliforms) and cannot distinctly detect pathogens such as *Legionella* and *Campylobacter.* Another standard test is the heterotrophic plate count test, which quantifies the general microbial load in the water without discriminating harmful or harmless bacteria (Rompre 2002, Edberg 2000).

Common microbiological lab methods for the detection of bacteria generally include culturing samples in selective media, in optimal conditions, followed by colony counts or turbidity measurements (Edberg 2000). Colony counts are an indication of the number of bacteria that is measured by the normalized amount of bacterial colonies on an agar plate. The number is usually normalized to colony-forming-units per ml (CFU/ml). These methods are problematic because the formation of colonies requires a long time, between 12 hours and a week. Turbidity measurements can be done without culturing but they allow rough estimates, efficient only in exceptional cases where the microbial load is as high as 106 CFU/ml or more. These disadvantages are one of the greatest challenges in the food and water industries, and cause a threat to public health (Leclerc 2002, Hennekinne 2012). Moreover, current methods require a large team of trained personnel and lots of resources (growth media, petri dishes etc.) which makes testing expensive (Rompré 2002). Not only that, but testing is only done on small sample sizes of between 0.1-1 liter out of hundreds of thousands that pass through the system. The tests are done periodically and the majority of water is not tested. Since water contamination is unpredictable and sporadic (Cabral 2010, Frolich 2017), a failure in testing in the water industry can be identified which may undermine public health. This failure is well known but presently no high quality and affordable alternative is commonly used (Rompre 2002, Willemse-Erix 2009).

Several modern methods have been suggested to replace the standard. Molecular methods, based on specific DNA amplification have been shown to be highly specific to pathogens (romper 2002) but they still require a long culturing stage. More recent methods, based on nucleic acid microarray or Enzyme-Linked Immunosorbent Assay (ELISA) technology can provide highly specific results faster (within hours). However, these methods rely on expensive reagents, instrumentation and highly trained personnel and are thus not applicable to most industrial uses. Spectral methods, based on mass spectrometry have also been proposed, especially the Matrix-assisted laser desorption/ionization (MALDI-TOF) approach. These utilize high energy lasers to ionize samples, causing molecular structures to breakdown and then measure the products using mass spectrometry (ref). While this approach requires fewer expensive reagents, it remains an expensive alternative due to the cost of the instruments. Further, the method requires culturing (albeit shorter than standard) which takes time and is limited to specific types of bacteria (16). There still remains a need for a rapid, accurate and quantitative method for the detection of bacteria in drinking water that is cheap and reliable.

# Light Spectroscopy

Light spectroscopy approaches, such as Raman, infra-red and fluorescence spectroscopy have been suggested by numerous experts as alternatives to the standard methods (16, 17). These different approaches are based on the interaction of light with the chemicals that comprise bacteria, and cause the light to change in measureable ways. While these interactions are relatively rare and complex, modern instruments with sensitive spectrometers can detect the subtle changes to the light's wavelength and characterize the light after the interaction with a sample; this is called the sample's spectrum. Furthermore, theoretically there is a correlation between altered light intensity and bacterial concentration, according to Beer's law (21, phil Williams), since the more bacteria are in a sample, more photons will interact with them. However, in order to create a reliable method for detecting bacteria in the food and water industries, a large and robust dataset of spectra must first be collected (16). Once the dataset is created, a statistical model can be designed for rapid detection of bacteria according to their spectra. The main advantages of light spectroscopy are that it requires very little to no samples preparation, no expensive reagents and the instrumentation can be relatively cheap (depending on the quality of the instruments). Further, once a dependable method is designed, no skilled personnel are required and in many cases the sample scanning can be done automatically and even on-site, thus saving on operational costs and reducing human error. In this work we endeavor to apply two different spectroscopy approaches, Raman and fluorescence spectroscopies, for the detection of bacteria in water.

# Raman spectroscopy

## Scientific background

Raman spectroscopy is a modern analytical tool, which can deliver rapid results at a relatively low cost. The method is based on Raman Shift, a phenomenon first described by CV Raman in 1928. When photons of a specific wavelength interact with a molecule some of the photons go through a Raman Shift, where the photons’ energy (and wavelength accordingly) are altered. This is because some of the photons wave energy is transformed into vibrational energy in the molecular structure (Bernhard S:1995). The shift is different between molecular structures and bonds, since each vibrational mode can transfer a different amount of energy. Consequently, the Raman effect is specific and a molecular “fingerprint” can be assembled (Bernhard, Kaiser). Raman shift is dependent upon the Polarizability of a substance, i.e. the flexibility of its electron cloud. This is because the energy is transferred via the electron cloud, and in a low polarizability substance the photons cannot transfer the energy. Water molecules (H2O) have very low polarizability, and thus a minute Raman signal, making them an excellent background for Raman spectroscopy (Stockel 2015).

Raman shift is measured using a light spectrometer which measures the light’s wavelength after it has gone through a sample and converts it into a digital signal for computer processing (Stockel 2015). The digital spectrum displays the intensity of light at each wavelength that is longer than the original transmitted light (which is cutoff from the signal). This light had undergone Raman Shift (fig. XXX). The wavelength can be displayed as absolute wavelength in nanometers, but is usually converted into energy shift in units of cm-1.

Figure 1 Spectral fingerprint of ethanol

## Uses of Raman spectroscopy for the detection of bacteria

Different studies focused on different aspects of applying Raman spectroscopy for bacteria detection. In the 1990s, some of the first attempts to describe bacteria and their Raman fingerprints have managed to accurately differentiate different strains of bacteria (Mannoharan 1990, William 1994, Naumann 1995, Fehrmann 1995) such as *Clostridia* spp., *staphylococcus* *aureus*, *Escherichia coli*, *Bacillus subtilis* and others using long laser exposure times and very high concentrations of bacteria. Later, Zeiri et al (18, 19) have shown that most of the spectral fingerprint of bacteria is due to DNA moieties such as adenine and Flavin adenine dinucleotide (FAD). They found that certain phosphorous and carboxylic bonds can also be detected using Raman. In another paper, Premasiri et al (20) have shown that the Raman spectra of bacteria is also related to guanine (which is another DNA moiety).

These findings have encouraged research into the application of Raman spectroscopy for detection of pathogenic and spoilage causing bacteria in the food industry. Meisel et al (26) have demonstrated detection of *Brucella, Escherichia and Yersinia* spp*.* in milk samples. A database of over 2,000 spectra had to be collected, and each sample was scanned using a Raman microscope to improve the signal. A machine learning process enabled accurate differentiation between species with >95% accuracy. It is important to note that while the bacteria were grown in milk, scanning was done after washing with water. A similar study (28) demonstrated an ability to differentiate meat associated *Salmonella, Listeria and E. coli* species with >85% specificity. Both studies required overnight culturing to achieve meaningful results and the use of microscopes, which are both expensive and time consuming. Similar works by Wang (29), Nicolau (21) and others have demonstrated the possibility of detecting and differentiating bacteria relevant for the food industry.

A different approach was to try and quantify the bacteria in a sample. The challenge with detecting and quantifying bacteria is that the Raman fingerprint of bacteria is very weak (popp). This has led to the use of surface enhanced Raman spectroscopy (SERS) techniques. Zhou et al (36) have managed, using gold particles and SERS, to detect *E. coli* cells at less than 100 CFU/ml in a 3 µL sample. The gold particles were added to the sample, causing a localized surface Plasmon resonance reaction, which traps light and increases Raman effect. The same approach was used by Sundaram et al (32) to increase their signal by a factor of 106-1010. While highly attractive, the use of SERS is difficult to scale up to the food and water industries. The high cost of gold and silver, the expertise required to construct them and the need to replace them regularly make this method too expensive for industrial scale testing.

## Low resolution Raman spectroscopy

While most studies have applied microscopy based approaches or surface enhancement using gold and silver nanoparticles, few have tried to detect bacteria using low resolution tools. Schmilovitch et al (38) detected *Erwinia* and *Clavibacter* species (known plant pathogens) using a simple 785 nm laser and a spectrometer, without any light enhancement techniques. They managed to detect the bacteria at concentrations as low as 100 cells/ml, using a Partial Least Squares (PLS) prediction model, with an accuracy >90%. This work can be regarded as a proof of concept for detection of bacteria using low resolution Raman spectroscopy that can be applied for the food and water industries. Other works by Mello (39) and Mizrah (Mizrah 2007) have shown similar results in enteric bacteria and yeast respectively. All these discoveries were done using low resolution Raman, with advanced algorithms to prepare and analyze large datasets. These algorithms (such as PLS, Support Vector Machine and others) elucidate the spectra and consequently improve detection limits and resolution, without the need for expensive instrumentation or sample preparation. Such an approach, when properly calibrated, could be adopted in the food and water industry.

# Fluorescence spectroscopy

## Scientific background

Fluorescence is a natural phenomenon that results from the absorption of a photon by a molecule or ion, which causes electron excitation from the ground state to higher energetic levels (excited state). When excited, the electron may lose some energy due to molecular bond vibrations, collisions and other non-radiative interactions. This causes the electron to reach a lower-energy excited state. The electron returns to the ground state by naturally emitting a photon. However, the emitted photon has a lower energy level than that of the exciting photon due to the loss of energy by internal interaction (41) (fig. YYY).

A measurement of the fluorescence effect (i.e. of the difference between the excitation and emission wavelengths of light) can be applied for describing the chemical nature of a sample.



Figure 2. Jablonski diagram illustrating fluorescence

This approach is widely used for the analysis of soluble organic matters in natural water (Ishii 2012, Borisover 2009, Guo 2011), sea water (Yamashita 2011, Stedmon 2007), wastewater (Cohen 2014, Yang 2015) and drinking water (Bieroza 2009, Baghoth 2011, Sorensen 2018). Certain organic compounds, particularly those with aromatic ring structures, will fluoresce when a specific light source is applied to them. This has been widely described in the literature specifically as a method for the quantification and identification of humic and proteinaceous substances. In proteinaceous substances, the major fluorescence effect is derived from the aromatic amino acids: Tyrosine, Tryptophan and Phenyl-alanine (Fig. 3). These proteinaceous substances can be free amino acids, proteins (folded or denatured), partially digested proteins, short peptides and even some different organic molecules which contain the aromatic amino acids, or similar structures such as indole groups. The fluorescence of these structures is shifted according to the different combinations and structures in which the amino acids are organized (Lakowitcz: Protein fluorescence).

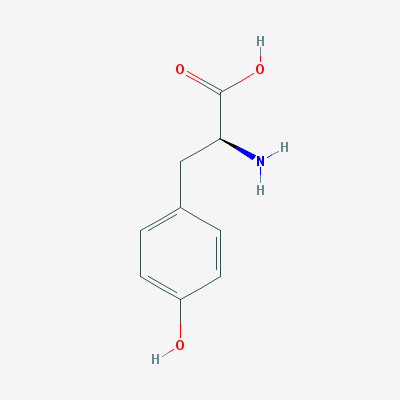
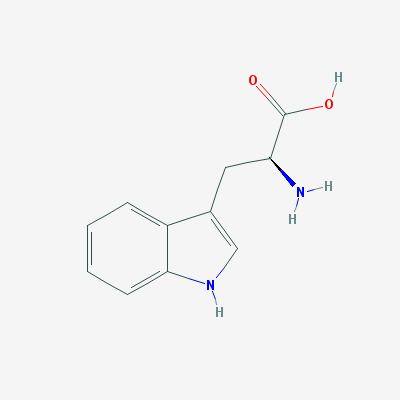
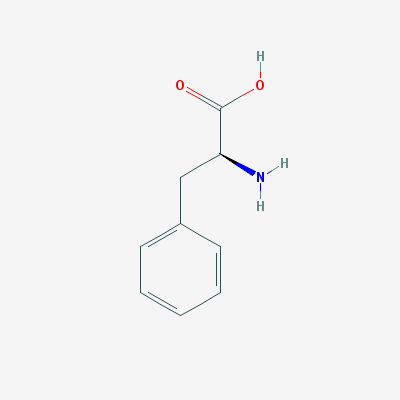


Figure 3. Molecular structures of (from left to right) tyrosine, tryptophan and phenylalanine (cite: Pubchem X3)

## Uses of fluorescence for the detection of bacteria

The presence of these proteinaceous substances has been suggested as a proxy to the presence of bacteria. Previous studies have shown that bacteria have spectral fingerprint within the proteinaceous substance region of fluorescence (roughly 200-400 nm excitation and emission) (Determan, Elliot 2006). Furthermore, different studies have shown an ability to differentiate between different bacterial species, at a high concentration, in colonies (Belal 2011) and in enriched blood samples (Walsh 2013) using this method (or a larger range of excitation and emission wavelengths).

This has led to studies for the development of a method to detect bacteria in drinking water. Two major approaches are used in the literature for measuring protein-like fluorescence in drinking water: The "Peak-Picking" approach and multi-spectral approaches based on Excitation-Emission Maps (EEMs). In the peak picking or single wavelength approach, a sample is illuminated by monochromatic light at a specific wavelength, and the emission of light at a different wavelength is measured. For example, in order to measure tryptophan fluorescence Tedetti et al (Tedetti 2013) used excitation at 280 nm and measured emission at 340 nm. Conversely, the EEM based approach uses a range of different excitation and emission wavelengths in order to study a sample. By measuring the emission of light at each pair of wavelengths this method can establish a clear picture of the different organic elements in the sample (fig. 4).

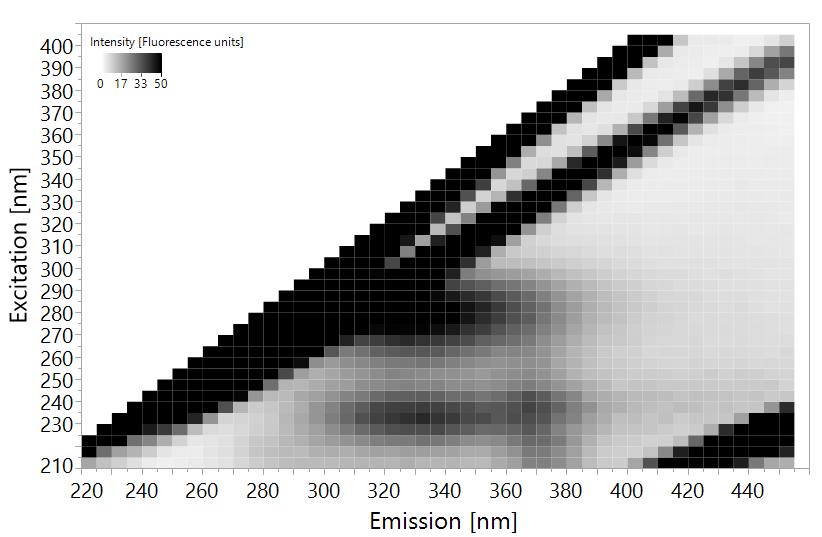


Figure 4. EEM of E. coli suspended in distilled water at a concentration of 105 CFU/ml. The intensity has been cutoff at 50 fluorescence units to enhance visibility.

The clear tradeoff between the different approaches is in resolution over costs. While the single wavelength approach is cheaper and easier to analyze, EEMs yield more information at higher operational and computational costs.

Sorensen et al, in 2015, suggested the application of simple single-wavelength measurements of tryptophan fluorescence for the detection of microbial contamination of water. They used a portable fluorometer with excitation at 280 nm which measures emission at 360 nm on water taken from boreholes and shallow wells in Zambia. This study has shown the ability of this fairly simple instrument to detect an *E. coli* contamination of >1 CFU/100ml of water, which was attributed to ~3.5 ppb of tryptophan. A follow up study, by Baker et al (Baker 2015) used a similar instrument to detect *E. coli* contaminations in low quality catchment water in South Africa rural areas. This experiment managed to accurately detect *E. coli* contamination at a concentration of >100 CFU/100ml with a tryptophan equivalence of 3 ppb. Further studies by Sorensen et al in 2016 and 2017 show similar results, displaying an ability to detect low levels of *E. coli* contamination in low quality drinking water in 3rd world countries such as India, Malawi and Zambia, with the tryptophan equivalence ranging between 1.5-3.5 ppb of tryptophan as a measure of contamination. These studies suggest that the application of single-wavelength fluorescence spectroscopy, at the tryptophan peak of excitation-emission 280nm-360nm (+-10nm) can be used for the detection of bacteria in low quality water. It is Important however to note that water with >1 CFU/100ml of *E. coli* is considered very low quality in most western countries. The Israeli, European Union and United State regulations all prohibit consumption of water with this level of contamination, and the phenomena is relatively rare (ref).

A similar study to the one done in the UK, where the water quality is regarded very good, used a lab-constructed fluorometer with a LED excitation light source at 280 nm and an ability to measure emission at 350 nm. In this study, measurements were taken throughout the supply chain (at the source, in checkpoint in the pipeline and at home taps) of both fluorescence and standard microbial indictors such as *E. coli* contamination and heterotrophic plate counts (HPC). No correlation was found between the microbial indicators and the fluorescence intensity at the tryptophan peak. This finding implies that the single-wavelength method may not be sensitive enough for high quality water.

In 2018, Heibati et al investigated a multispectral approach to detecting bacteria in drinking water in Sweden. While Sorensen, Baker and Bridgman all used only the tryptophan peak of excitation-emission 280-360, Heibati et al used the full spectra at excitation 220-600 nm and emissions at 240-800 nm (3 nm resolution) to construct high resolution EEMs. The EEMs were analyzed by complex multivariate statistical methods in order to analyze the large dataset of EEMs, this included Partial Least Squares (PLS) and Parallel Factor Analysis (PARAFAC). No severe microbial contamination, and had no *E. coli* or coliforms were found throughout the study's measuring period. The only indicator which was applicable for analysis was HPC, and only weak correlations were observed between the EEM data and HPC bacterial abundances. This may serve as an indication that at very low concentrations (<150 CFU/ml of heterotrophic bacteria) it is difficult to detect bacteria using the fluorescence measurement approach.

These studies demonstrate a possibility of using fluorescence spectroscopy for the detection of bacteria in water. This method, like Raman spectroscopy is relatively cheap to apply and requires no sample preparation. It appears though, that simple single wavelength measurements are not sensitive enough to enable detection of low levels of bacteria, as required in modern industrialized countries such as Israel. For this purpose, perhaps the multispectral EEM approach will be better suited. Such an approach, when properly calibrated, could be adopted in the food and water industry.

# Multivariate data analysis

Because of the complex nature of EEMs and Raman spectra, they are always analysed using multivariate statistics. Partial Least Squares (PLS) regression or discriminant analysis is a particularly suitable method for predicting bacterial concentration using spectral data since it performs well even when the number of predictor variables is high and some variables correlate with each other (Heibati?). In the context of drinking water monitoring, it would be desirable to predict bacteria concentration (Y) from easily-obtained measurements of Raman spectra or fluorescence EEMs, which are comprised of hundreds of wavelengths (X­n). PLS algorithms find the correlation between these different variables (Xn), and then correlates those to the dependant variable (Y). Using an iterative process, this enables optimization of the importance or weight given to each variable and a prediction formula can be calculated (Equation 1) (Geladi). Because of the nature of the analysis it will almost always manage to properly "predict" Y. However, prediction models might have a bias called overfitting, where your model is only relevant to your dataset. In order to overcome this, the data is split into calibration and validation sets. The model is first trained on the calibration or training set, and is then tested for accuracy and validity on the validation set. This approach is called cross validation and enables calibration of robust prediction models (Geladi, GHOLIZADEH, Schmilovitch).

Y = Predicted Bacterial concentration

*i* = the wavelength vector (all wavelengths analysed)

ai= coefficient at wavelength *i*

X = light intensity at wavelength *i*

e = the calculated error

Equation 1 - PLS prediction formula

MISSING – WRAP UP PARAGRAPH