A Sheffield Hallam University thesis

Human identification through advanced forensic mass spectrometry of blood and fingermarks

Cameron Edward Glenn Heaton

A thesis submitted in partial fulfilment of the requirements of
Sheffield Hallam University
for the degree of Doctor of Philosophy

In collaboration with the Defence, Science & Technology Laboratory
(DSTL), Porton Down, UK

August 2022

Dedications

I would like extend my gratitude to my Director of Studies, Professor Simona Francese. Simona, thank you for fostering my passion for research which commenced with my undergraduate degree project, and securing a BMSS funded studentship for me to continue forensic research following my BSc. Thank you also for all of the guidance you have provided in many forms over the course of my PhD programme. Thank you also to Dr Laura Cole and Richard McColm for the supervision over the duration of the project.

Also thank you to my PhD cohort, Rob Tempest, Jack Slater, Lucy Flint and Oana Voloaca, who all provided much needed laughs and motivation both in and out of the labs.

Thank you to the Fingermark Research Group (past and present) for providing patient training, inspiration and successful on-going research collaborations: Katie Kennedy, Dr. Cristina Russo, Dr. Ekta Patel, Dr. Robert Bradshaw. I wish you continued success in your respective fields.

Thank you also to Jason Eyre and colleagues from Sheffield Royal Hallamshire Hospital Haematology Department for being particularly accommodating providing haemoglobin samples and consultancy for the project.

Finally I am most grateful to Chloe for keeping me motivated and inspired to finish research and thesis, without you it's doubtful I would have completed.

Candidate Declaration

I hereby declare that:

I have not been enrolled for another award of the University, or other academic or professional organisation, whilst undertaking my research degree.

None of the material contained in the thesis has been used in any other submission for an academic award.

I am aware of and understand the University's policy on plagiarism and certify that this thesis is my own work. The use of all published or other sources of material consulted have been properly and fully acknowledged.

The work undertaken towards the thesis has been conducted in accordance with the SHU Principles of Integrity in Research and the SHU Research Ethics Policy.

The word count of the thesis, excluding references and title section, is:

64,291

Name	Cameron Edward Glenn Heaton
Award	Doctor of Philosophy
Date of Submission	1st August 2022
Faculty	Health & Wellbeing
Director of Studies	Professor Simona Francese
Signature	▉

Abstract

Matrix assisted laser desorption ionisation (MALDI) mass spectrometry (MS) has been demonstrated in recent years to be an effective toolbox in the detection and identification of substances of forensic relevance. The goal of this PhD programme was to develop analytical protocols to obtain novel molecular information from two biological sample types; fingermarks and bloodstains. MALDI MS was employed to acquire mass spectrometry profiling (MSP) and mass spectrometry imaging (MSI) data from these two common evidence types to generate biological and chemical information towards a molecular 'suspect profile'.

The use of MALDI MS to acquire data on the peptide and protein composition of fingermarks, in combination with advanced statistical processing was developed to investigate the determination of sex from fingermarks.

A blind validation study was conducted for the robust, multifaceted identification of human and animal blood, other human biofluids and interferent substances from a large sample set of stains and fingermarks. MALDI MS was employed in combination with a 'bottom-up' proteomic approach for the determination of biological matrices through the identification of proteotypic peptides.

MALDI MS was also utilised for the determination of a subset of haemoglobin variants in human blood, through the detection of proteotypic peptides, employing a 'bottom-up' proteomic approach. The detection of haemoglobin variants in blood encountered at a crime scene has implications towards associative evidence of the presence of an individual, which grows in significance the rarer (and less prevalent) the variant.

The application of these three techniques towards these two biological matrices, and the validation data acquired for each, demonstrates their analytical capabilities. In all cases they have been demonstrated to be compatible with visualisation techniques (both fingermark enhancement techniques and blood enhancement techniques, respectively), showing potential for their integration into the current operational workflows for both fingermarks and blood. In conclusion, although only demonstrated under laboratory controlled conditions, MALDI MS shows promise towards the future analysis of these biological matrices within evidence recovery and investigative workflows, when specific molecular information may be sought.

Table of Contents

Abbreviations

α-CHCA – α-cyano-4-hydroxycinnamic acid

µL – Microlitre

µm – Micrometre

AB1 – Acid Black 1 (see also: amido black)

ACN – Acetonitrile

API – Atmospheric Pressure Interface

Aq – Aqueous

ATR-FT-IR – Attenuated Total Reflectance Fourier Transform Infrared Spectroscopy

AY7 – Acid Yellow

BET – Blood Enhancement Technique

BMRC – Biomolecular Research Centre

BV3 – Basic Violet 3

BY40 – Basic Yellow 40

CAF – Cyanoacrylate fuming

CAST – Centre for Applied Science and Technology

CFC – chlorofluorocarbons

CI – Chemical ionisation

CID – Collision induced dissociation

CPCD – Coupled Photophysical and Chemical Dynamics model

CSI – Crime Scene Investigator

Da – Dalton

DART – Direct analysis in real time

DC – Direct Current

DESI – Desorption Electrospray Ionisation

DFO – 1,8-Diazafluoren-9-one

DHB – 2,5-dihydroxybenzoic acid

DSTL – Defence, Science and Technology Laboratory

EI – Electron Ionisation

ESI – Electrospray Ionisation

eV – Electron Volt

FET – Fingermark Enhancement Technique

Fmol – Femtomole

FRG – Fingermark research group

FTIR – Fourier Transform Infrared

GC – Gas Chromatography

Hb - Haemoglobin

HDMS – High Definition Mass Spectrometry

HFE – hydrofluoroether

Hz – Hertz

IEF – Isoelectric Focussing

IMS – Ion Mobility Separation

IR – Infrared

ISD – In-Source Decay

kDA – Kilodalton

KeV – Kiloelectron Volt

KHz – Kilohertz

LCV – Leucocrystal Violet

LDA – Linear Discriminant Analysis

LDI – Laser Desorption Ionisation

MALDI – Matrix Assisted Laser Desorption Ionisation

MeV – Mega Electron Volt

mL – Millilitre

mm – Millimetre

MS – Mass Spectrometry

MSI – Mass Spectrometry Imaging

MSP – Mass Spectrometry Profiling

MS/MS – Tandem Mass Spectrometry

m/z – Mass to charge ratio

Nd-YAG – Neodynium-Yttrium Aluminium Garnet

ng – Nanogram

nm – Nanometre

NMR – Nuclear Magnetic Resonance

ns – nanoseconds

ORO – Oil Red O

PCA – Principle Component Analysis

PD – Physical Developer

PEG – Polyethylene Glycol

pg – Picogram

PI – Photochemical Ionisation Model

QTOF – Quadrupole time-of-flight

ROI – Region of Interest

S/N – Signal to Noise ratio

SALDI – Surface Assisted Laser Desorption Ionisation

SHU – Sheffield Hallam University

SIMS – Secondary Ion Mass Spectrometry

SOCO – Scene of Crime Officer

TATP – Triacetone Triperoxide

TIC – Total Ion Count

TiO_2 – Titanium dioxide

TFA – Trifluoroacetic acid

TOF – Time of Flight

TPA – Two Photon Absorption

TQ – Triple Quadrupole

UK – United Kingdom

UV – Ultra violet

VIP - Variable Importance in Projection

VMD – Vacuum Metal Deposition

XGBOOST - eXtreme Gradient Boosting

Chapter 1: Introduction

Introduction

Forensic science is the implementation of scientific practices towards solving crimes as part of a criminal investigation. This involves the collection, preservation and analysis of exhibits relevant to a crime. This PhD project focuses on two types of commonly encountered biological evidence, fingermarks and blood, separately, and in combination, as bloody marks. The aim of this PhD is to demonstrate the use of Matrix Assisted Laser Desorption Ionisation (MALDI) to reveal novel molecular information that could be used to aid investigations in the future, by obtaining a molecular 'suspect profile' from these two biological evidence types. MALDI is a mass spectrometric analytical technique used to detect the molecular species within an analyte. Chapter one will provide an introduction to the biological matrices of fingermarks and blood, and a background into their use as evidence, including current analysis techniques. Chapter two will focus on experimental methods for the determination of an individual's sex through analysis of the peptide and protein composition of fingermarks using MALDI, in combination with advanced statistical processing to aid with classification. Chapter three will focus on a pre-validation study for the development of a method for the reliable detection of blood, and the discrimination of human blood from animal blood, other human biofluids and several interferents using MALDI. Chapter four will focus on the novel detection and identification of haemoglobin variants within human blood in a forensic context, to aid with suspect or victim identification, using MALDI. Chapter five will present a conclusion of the results obtained in this PhD programme, and a review of the techniques potential readiness for operational deployment at crime scenes or in forensic laboratories.

Fingermarks

History

Fingerprints have been used as personally identifiable features since the 19th Century. Henry Faulds was first to suggest the use of fingermarks for criminal identification, in a paper submitted to *Nature* in 1880. This submission included the salient quote "*When bloody finger marks or impressions on clay, glass, etc., exist, they may lead to the scientific identification of criminals*" (**Lee & Gaensslen, 2001**). Around the same time, Sir Francis Galton is credited with realising that an individual's prints do not change over their life time and that each individual's prints are unique. Juan Vucetich, a Croatian working as a Chief of Police in Argentina, was the first to record individual's fingerprints to keep on file and founded the first fingerprint bureau in 1891. Less than a year later he was credited with the first conviction resulting from fingermark evidence, when he matched a bloody fingermark found at a murder crime scene to a suspect, who subsequently confessed. In 1902, the first conviction using fingerprint evidence in the UK was made from a burglary crime scene (**McRoberts, 2011**). The first scientific research reporting that third level detail was permanent and unique was made in 1918 by Wilder and Wentworth (**Wilder & Wentworth, 1918**).

To this day, fingermarks are still considered one of the most useful types of physical evidence for identification purposes, and are responsible for more criminal identifications globally than DNA (**Francese, 2019**). The first trial in the UK that relied on fingerprint evidence was to secure a conviction for a burglary in 1902, which was presented in court by Inspector Charles Collins of Scotland Yard. The first murder trial in the UK was the Deptford Murder Trial in 1905, with

3

the fingerprint evidence also being presented by Inspector Collins. A fingermark on the victim's cash box was matched to one of the defendants, one of two brothers on trial. The Inspector claimed that he had previously never found two prints to have more than three characteristics in common; the cash box mark and the defendants print had 11 characteristics in common, and as a result the brothers were found guilty and sentenced to death (**McRoberts, 2011**).

Fingermarks Background

Fingermarks are the result of incidental contact between the fingertips and a surface, upon which a molecular residue may be left. This residue comprises of sweat from the eccrine and sebaceous glands, and is often barely visible or not at all hence the term latent marks. Eccrine secretions are mostly aqueous, with around 98% being water, and the rest being made up of inorganic and water soluble organic molecules such as urea, amino acids and proteins. Sebaceous secretions are made up of fat soluble organic compounds including fatty acids, sterols, glycerides and wax esters (**Knowles, 1978**). Despite sebaceous glands not being present on the fingers or palms, fingermarks can contain a high proportion of lipids, which are secreted from sebaceous glands, and this is hypothesised to be transferred to fingertips from contact with sebum-producing regions such as the face, neck and hair (**Ramotowski, 2012**).

Fingermarks can also contain endogenous, exogenous and semi-exogenous substances. Endogenous substances are those that originate in the body and include organic species such as lipids and amino acids, and inorganic species that can include metal ions. Exogenous substances are external contaminants that the fingertips may come into contact with, before being transferred onto another surface (touch chemistry). In a forensic context, these can include

drugs of abuse, medicines, toiletries and cosmetics, accelerants, explosives and blood, amongst others. The chemical composition of fingermark residue has recently been shown to change with time. Gorka *et al.* conducted an assessment of intravariabilty over the course of a year using MALDI MSI. They found that on average, between 25-45% of compounds are common to all of the marks from a single donor analysed during the study (**Gorka *et al.,* 2022**).

Whilst fingermarks are normally left incidentally, fingerprints are purposely collected to record biometric information. To date, no two fingerprints have been found to be the same, not even in identical twins; this phenomenon gives an advantage over DNA, which can produce identical profiles among twins. For example, in 2008, Police obtained a DNA profile from a discarded cigarette at the scene of a fatal carjacking. This led them to a man named Donald Smith, who matched eye witness accounts, but who claimed his innocence and implicated his identical twin brother, Ronald. Fingerprints obtained from the scene matched with Ronald, and not Donald, who was exonerated (**abcnews.go.com, www.cbsnews.com**).

Fingerprint patterns remain constant throughout an adult individual's life, and the pattern will even regenerate after attempts have been made to remove them via burning or acid. However, studies have shown that the pattern can change in some cases of severe scarring, or the fading of the ridge pattern. This can be as a result medical conditions or therapeutic treatment, for example chemotherapy-induced acral erythema in patients undergoing chemotherapy (**Wong *et al.,* 2009**). In addition, there are three known genetic conditions that have been found to result in a person being born without ridges and valleys on the surface of a fingertip. These are Naegeli-Franceschetti-Jadassohn Syndrome (NFJS) and Dermatopathia Pigmentosa Reticularis (DPR), (both

keratin-related gene mutations) and Adermatoglyphia, which unlike the other two conditions, presents no other symptoms.

A fingerprint ridge pattern can be categorised into three levels based on the complexity of the features; most broadly is the ridge flow type; which can be categorised into arches, loops and whorls. These categories were identified by Sir Francis Galton in his 1892 book *Finger Prints* (**Galton, 1892**). These have since been refined into 5 major classes of fingermarks to include tented arches, and distinguishing between right and left loops by the Galton-Henry classification system. The ridge flow represents level 1 detail, and whilst it is not sufficiently unique to use as a method of individual identification, it is useful for preliminary classification. Level 2 covers the individual features of the ridge detail, termed *minutiae* (a term coined by Galton). Gutiérrez-Redomero *et al.,* illustrated 18 such features used by the Spanish Scientific Police in a paper in 2011 **(Fig. 1)** (**Gutiérrez-Redomero *et al.,* 2011**).
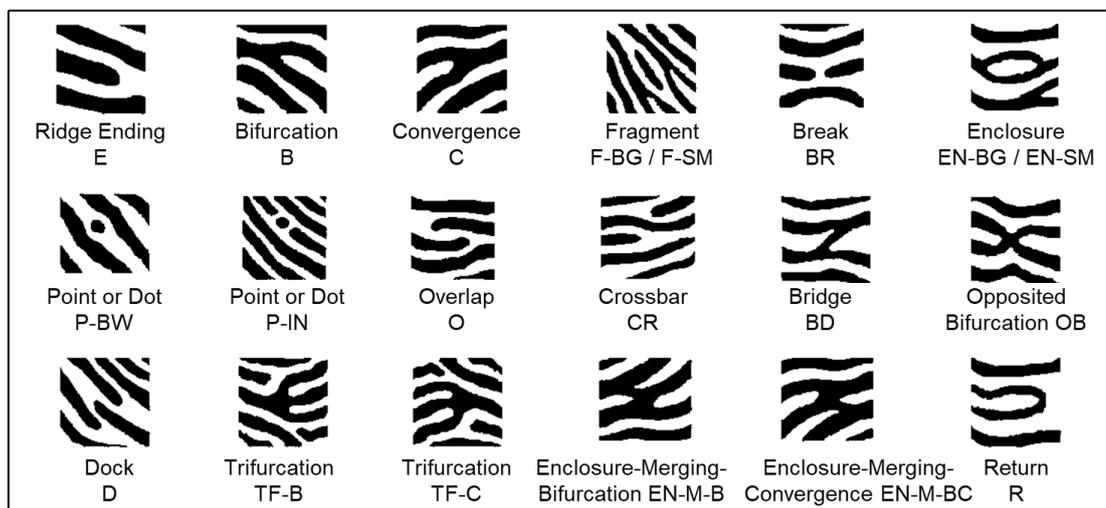
Figure 1. The 18 features of level 2 ridge detail used by the Spanish Scientific Police [*Adapted from: Gutiérrez-Redomero et al., 2011*].

Level 3 detail concerns even more minute features such as sweat pores, creases, scars and dimensional attributes like the width, shape or contour of a ridge.

Fingerprint Classification

The most popular ten-print classification systems are the Vucetich System, the Henry Classification System and the Roscher System, which are based on the different general shapes and patterns on the ridges of the fingers.

Vucetich System: Juan Vucetich, whilst implementing the French Bertillon Anthropometric Identification System, became particularly interested in fingerprints. He implemented his own system for classifying them, using four fingerprint patterns in his book *Dactilospia Comparada*. In 1893 the famous Rojas murder was solved, cementing the usefulness of his fingerprint system within the Bureau, which is the system still used in South America to this day (**Lee & Gaensslen, 2001**).

Henry Classification System: The Henry Classification system, developed as the result of collaboration between Henry Faulds, William Herschel and Edward Henry around 1901, was adopted by Scotland Yard in 1903. The FBI is another prolific user of the Henry system, along with most of the Commonwealth (**Lee & Gaensslen, 2001**).

Roscher System: Henry Roscher was a disciple of Henry Faulds, and developed the Roscher system in Germany, which also gained popularity in Japan (**McElreath, 2013**)**.**

ACE-V System: Currently, the most commonly used examination process is the ACE-V system, an acronym for Analysis, Comparison, Evaluation and

Verification. Analysis requires examiners to determine the quality of the features, and the quantity of first, second and third level detail. These criteria can be affected by the deposition surface, the detection/enhancement techniques used and the environmental conditions as well as the pressure with which the marks were left. A combination of these factors would determine the suitability of marks for further examination. Comparison involves examiners studying marks and prints side-by-side and using all levels of detail to determine the extent of the similarity. Evaluation allows examiners to take the quality and quantity of agreement/disagreement of features into account and come to a conclusion on the origin of the marks; match or no match. Verification would require another examiner to perform the ACE workflow and support or refute the initial examiner's conclusion (**Ramotwoski, 2012**).

There are typically three types of fingermarks that might be found at a crime scene; visible (patent marks), invisible (latent marks) and impressions (plastic marks) (**Lee & Gaensslen, 2001**).

Latent marks are the most common type of fingerprint evidence, but also the most challenging to utilise, requiring specialist treatment to effectively visualise (**Lennard, 2001**). Identification or elimination of a suspect based upon the detection of a fingermark requires the successful visualisation or 'enhancement' of a latent mark followed by the identification (or elimination) of individuals based on comparison with reference prints (**Lee & Gaensslen, 2001**).

Fingermark Enhancement Techniques

    Background

Before the second half of the 20th Century, fingermark evidence revolved around patent marks. However, being visible did not automatically mean that they were easy to photograph. Specialist lighting was sometimes required, presenting a challenge to illuminate marks *in situ*. Initially, enhancement reagents were developed to aid with the photography of patent marks, and this technique was later extended to visualising latent marks (**Bleay *et al.,* 2017**).

Surfaces are broadly separated into two types, porous and non-porous. Porous substrates include materials such as paper and cardboard, wood and other cellulose-based items. These items tend to absorb some of the fingermark constituents; therefore techniques which target amino acids can be effective on these surfaces, as amino acids can be absorbed into the pores and be protected from the external environment. For non-porous substrates, such as glass, metal and plastics, fuming, dyes/stains, powders or vacuum metal deposition (VMD) is usually applied. A third type of surface sometimes encountered is semi-porous, for example glossy cardboard, some wood finishes and Clingfilm (**McRoberts, 2011**).

Fingermark development techniques are normally intended to be applied in a specific sequential order in order to obtain the best enhancement and contrast against their background surface. The general workflow involves visual inspection with a specialist light source, sequential mark processing and documentation of developed marks following each step (**Bleay *et al.,* 2012, Bandey *et al.,* 2014**).

Categories of Latent Fingermark Enhancement Techniques

Fingermark enhancement techniques can broadly be separated by their mechanism of action. The Home Office categorise these processes as optical,

physical or chemical enhancement. Optical processes exploit the properties of light to create contrast between a mark and a surface. These techniques include colour separation processes, such as multispectral imaging, monochromatic illumination and colour filtration. Optical processes can also utilise fluorescence, bright visible light or wavelengths outside the visible spectrum, such as infrared (IR), ultraviolet (UV) or X-ray imaging to enhance visualisation of marks (**Bleay *et al.,* 2017**). Optical processes are routinely used for an initial speculative examination of an exhibit, and are usually applied first due to their non-destructive nature.

Physical Enhancement

Where optical processes have not obtained optimum visualisation, physical or chemical methods may be used to 'develop' marks to improve contrast against the background surface. The method selected depends on variables including the substrate that the mark has been deposited on and the suspected composition of the mark, which may be mixed with a contaminant, such as blood. Physical enhancement works due to the mechanical adherence of a reagent to the greasy components of latent marks. These include powders, powder suspensions, and vacuum metal deposition.

Fingerprint Powders

Powder dusting is the most common technique for latent mark enhancement. Application is usually with a brush, but can have a destructive effect if used by an inexperienced Crime Scene Investigator (CSI) (**James *et al.,* 1991**). There are four types of enhancement powder; regular, metallic, luminescent and thermoplastic.

Regular fingerprint powders comprise of two parts; a resinous polymer for adhesion and a colorant for contrast, and the most common colours are black, white and grey but they are available in a variety of colours and are selected depending on the colour of the background surface (**Lee & Gaensslen, 2001**). Magnetic fingerprint powders are fine ferromagnetic powders, applied with a special magnetic applicator, and work particularly well on leather, plastic, walls and human skin. They are widely used for visualising prints on vertical surfaces (**Lee & Gaensslen, 2001**). Luminescent fingerprint powders contain a fluorophore, which can re-emit light upon excitation. Excitation can take place under UV light, laser light, or other specialist light sources. These types of powders are particularly useful for visualisation of marks on multi-coloured backgrounds (**Lee & Gaensslen, 2001**). Thermoplastic fingerprint powders are those such as photo-copier toner or dry inks, and development with heat fuses marks to the surface (**Lee & Gaensslen, 2001**).

Powder Suspensions

Powder suspensions are fine powders dispersed through a mixture of concentrated detergent and wetting agent. The dissolved surfactant species accumulate on the surface of the powder particles, which encapsulates them and prevents aggregation. Initially developed for treating adhesive surfaces, such as tapes, their mechanism of action with latent marks is not fully known, but it is believed that particulate preferentially deposits on the constituents of eccrine sweat. As such it is debatable as to whether this is a physical or chemical enhancement process (**Bandey *et al.,* 2017**).

Vacuum Metal Deposition

Vacuum Metal Deposition (VMD) can operate in two distinct ways; with the metal adhering to the passive substrate and leaving the ridge detail as semi-transparent, or enhancing the ridge detail and leaving the surface untouched. The latter is known as a reverse VMD development (**Steiner & Bécue, 2018**).

VMD involves the deposition of metals, typically gold and then zinc (or silver and then cadmium) particles onto the constituents of latent marks. Evaporation of the metal powders occurs under high vacuum. Latent marks will be revealed if the composition of the mark differs significantly from the substrate, like breathing on a mirror, where the size of the water droplets forming on a latent mark are different from the fine mist left on the glass. A small quantity of gold powder is evaporated within the VMD chamber which forms a very thin layer (nanometers) across the substrate. The size and distribution of the particles depends on the different surface interactions with the fingermark or substrate. If the fingermark has a high fatty composition, gold particles will diffuse into the mark. Zinc powder is then evaporated within the chamber, which will preferentially deposit on the gold particles, but not the fingermark regions, which have absorbed the initial gold particulate (**Bleay _et al.,_ 2017**). This mechanism of action, where the zinc layer forms on the substrate, is termed 'normal' (or regular/positive) development, illustrated in **Fig. 2a**. Normal VMD development is effective at obscuring complex backgrounds such as text.

Occasionally, performing VMD will result in reverse (negative) development, where the secondary metal (zinc) particulate deposits on the fingermark residue rather than the substrate. There are several theories as to why this happens (discussed by Ramotowski, 2012 and Bleay _et al.,_ 2017), but essentially the gold particles are not absorbed by the fingermark residue, allowing the zinc particles to preferentially cluster on the gold particles adhered to the surface of

the fingermark, rather than the substrate, illustrated in **Fig. 2b** (**Bleay _et al.,_ 2017**). In some instances, positive and negative development may occur on the same substrate.
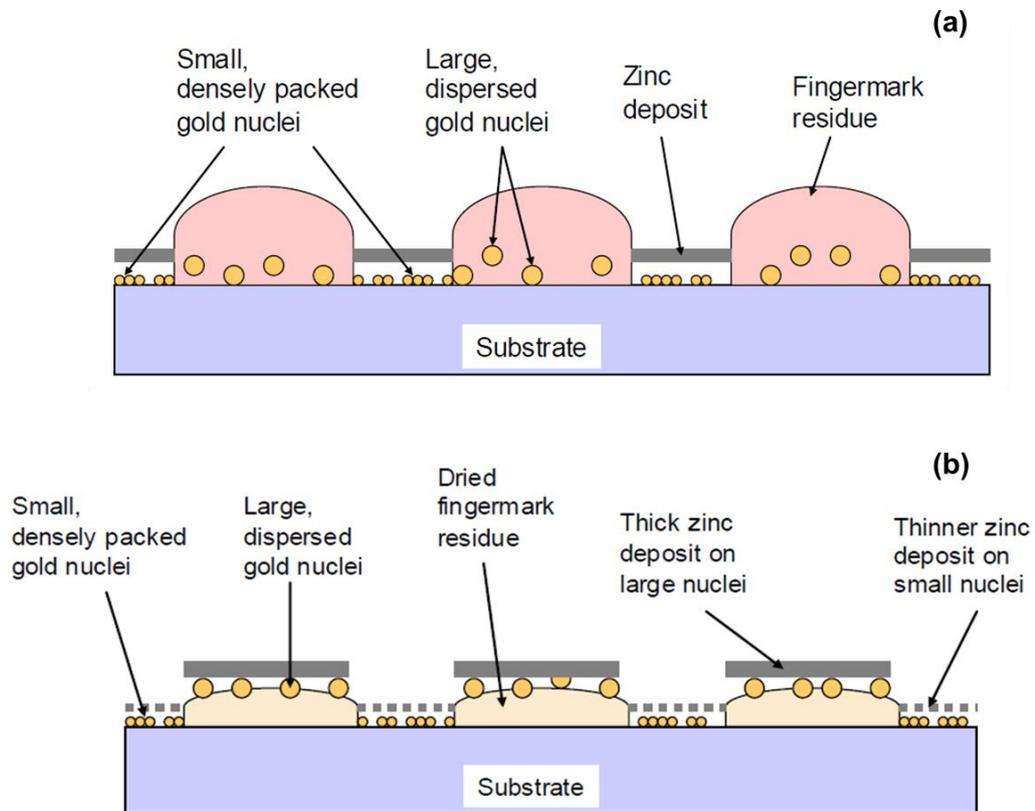
Figure 2. Schematic of (a) regular VMD, with zinc depositing on the available gold particles on the substrate, and (b) reverse VMD, with some zinc deposition on the gold particles on the substrate, but the majority adhering to the large dispersed gold particles on the residue. _[Adapted from Bleay et al., 2017]._

Chemical Enhancement

Amino acids constitute the most abundant class of chemical compounds in a latent fingermark (**Bossers, 2011**).

Fingermark Stains/Dyes

Fingermark stains or dyes are a class of chemical called lysochromes (lyso = fat, chrome = colour) that stain fatty acids, lipoproteins and triglycerides. Initially used to stain lipoproteins for microscopy visualisation or for use in

electrophoretic separation, they found a use as fingermark visualisation agents. Most fingermark stains are azo dyes, with an azo functional group attached to one or more aromatic rings (**Guigui & Beaudoin, 2007**). They are soluble in non-polar compounds such as triglycerides and operate by dissolving into fat content. For application, they are dispersed in polar solvents, but upon contact with non-polar substances such as the lipid content of fingermarks, they migrate out of the solvent and into the fat, preferentially colouring the mark (**Bandey et al., 2014**). Common azo-based lysochromes used for fingermark staining include Oil Red O (ORO a.k.a. solvent red 27 a.k.a. Sudan Red 5B), Sudan III (a.k.a. solvent red 23), Sudan IV (a.k.a. solvent red 24) and Sudan Black B (a.k.a. solvent black 3). They produce red, orange and blue-black stains respectively (**Guigui & Beaudoin, 2007**).

Ninhydrin

Synthesised by Ruhemann in 1910, it was realised that the compound ninhydrin (or 2,2-dihydroxy-1,3-indanedione) produced a coloured product when reacting with amino acids in peptides and proteins. It was proposed as a fingermark development reagent by Oden and von Hofsten in 1954 after accidentally developing a latent mark on paper (**Oden & Hofsten, 1954**). The solvent formulation has been refined many times since, moving away from using prohibited chlorofluorocarbons-based (CFC) solvents, towards being hydrofluoroether-based (HFE) (**Bleay et al., 2012, Bleay et al., 2017**).

In particular, two HFEs were trialled (2,3-dihydrodecafluoropentane (HFC4310mee) and 1-methoxynonafluorobutane (HFE7100)) by Hewlett et al. (**Hewlett et al., 1997**), both outperforming CFC based formulations, and the

superior HFE7100 based solvent being the formulation currently recommended by the Centre for Applied Science and Technology (CAST) (**Bleay *et al.,* 2017**).

Ninhydrin is applied by spraying, swabbing or immersing evidential items (**Lee & Gaensslen, 2001**). The generally accepted mechanism of how it reacts with amino acids was proposed by Friedman and Williams (**Friedman and Williams, 1974**), and involves a Strekker degradation of indanetrione to form an intermediate that reacts with another indanetrione molecule to form the brightly coloured Ruhemann's purple (**Fig. 3**) The full reaction can take days or weeks (**Lee & Gaensslen, 2001**). Its action can be accelerated using humidity and/or heat, although this is not recommended as ninhydrin can react with components of paper substrates, and heating speeds up both the required and unrequired reactions (**Jelly *et al.*, 2009**).
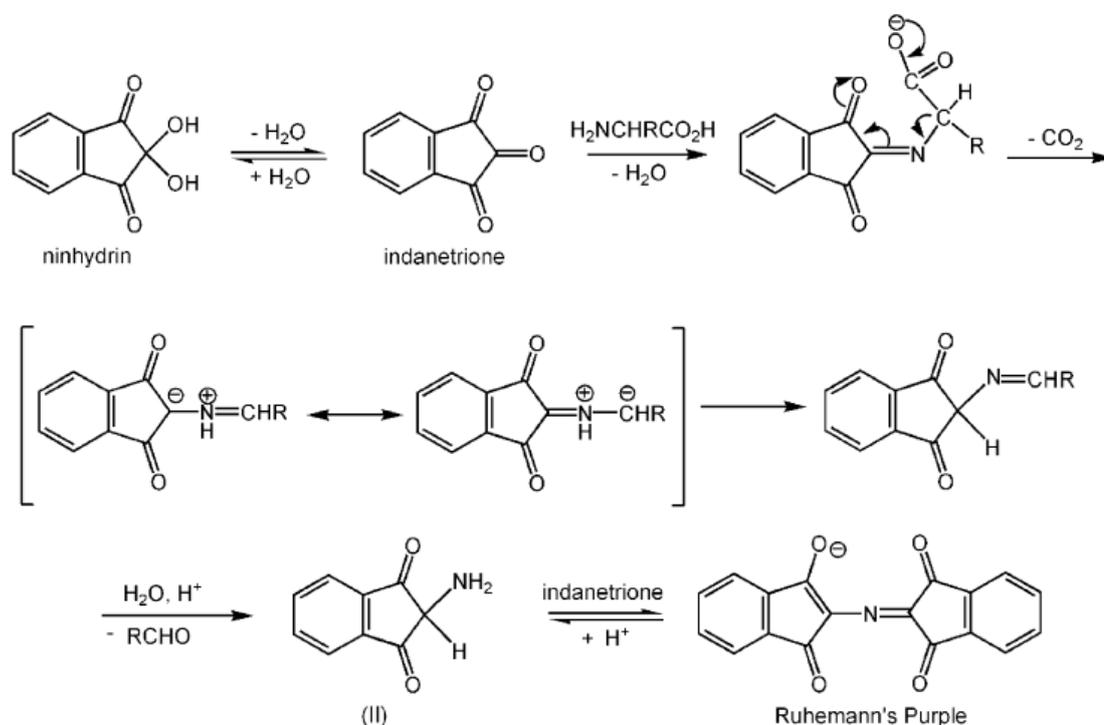
Figure 3. Mechanism for the formation of Ruhemann's Purple [*Adapted from Jelly et al., 2009*].

Ninhydrin is the predominant reagent for use on porous surfaces such as paper. Researchers have experimented with ninhydrin analogues but none have replaced ninhydrin as the most preeminent amino acid enhancement reagent in operational use (**Bleay *et al.,* 2017**). In advanced forensic laboratories, ninhydrin is used in sequence with other enhancement techniques; usually after 1,8-Diazafluoren-9-one (DFO) and before Physical Devloper (PD) (**Lee & Gaensslen, 2001**).

DFO

The second most common amino acid enhancing reagent is DFO. It was investigated in 1990 by Grigg and co-workers, maintaining the essential functional group (the five membered ring with a carbonyl in the middle and two adjacent dipoles on each side) (**Grigg *et al.,* 1990**). It was found to produce a red pigment upon reaction with amino acids, with a structure not dissimilar to Ruhemann's purple. It also fluoresced (**Lee & Gaensslen, 2001**).

Physiochemical Enhancement

Cyanoacrylate Fuming

The cyanoacrylate fuming technique was the result of another accidental discovery; L.W. Wood, working at the Police Headquarters in Northampton, noticed that his fingermarks developed on a film tank that he had repaired with super glue (**Ramotowski, 2012**). First researched as method of fingermark enhancement in the early 80's, cyanoacrylate or 'superglue fuming' was found to have limited success by the Home Office Central Research Establishment (HOCRE), and research was handed over to the Police Scientific Development Branch (PSDB), who determined humidity played a key role in the speed and

sensitivity of the technique. Humidity levels of 80% were found to be the optimum point between slow and ineffective development, and over-development of background surfaces (**Bleay *et al.,* 2012, Bleay *et al.,* 2017**). This technique operates by heating ethyl cyanoacrylate below a hanging exhibit in a specially designed chamber, with the preferential attraction of the cyanoacrylate monomer forming polymeric chains on the fingermark ridges. The polymerisation mechanism is illustrated in **Fig. 4**.
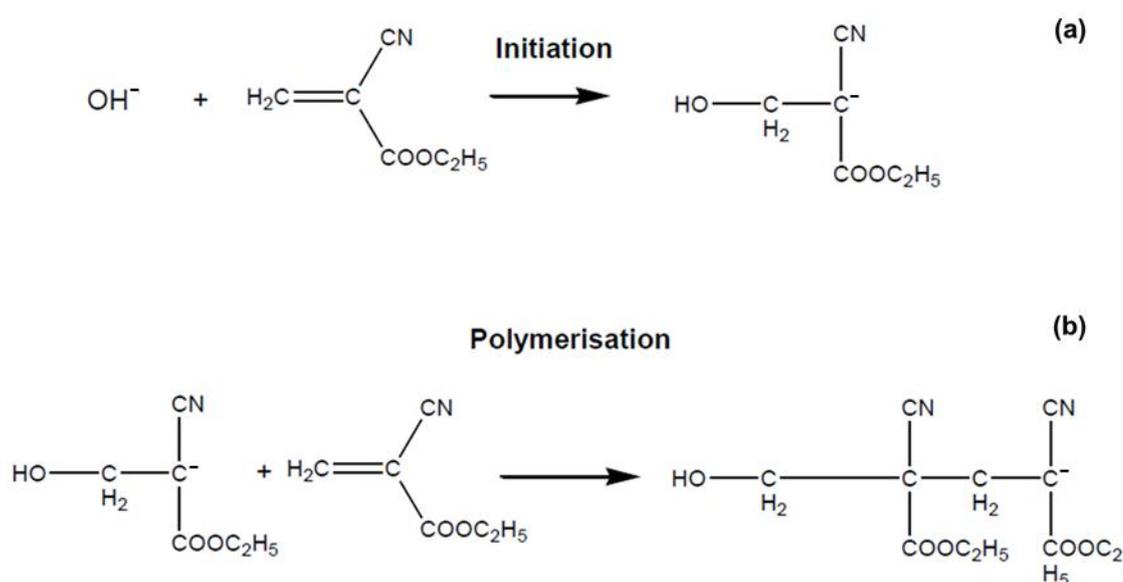
Figure 4. The polymerisation reaction of ethyl cyanoacrylate (superglue). *[Adapted from Bleay et al., 2017].*

The precise mechanism of action is not clear, however it is believed that the humidity in the chamber causes sodium chloride crystals in the mark to take up water, and cyanoacrylate polymerisation is initiated by weak bases, which might be initiated by NaCl as a result of this interaction (**Bleay *et al.,* 2012**). The Fingermark Visualisation Manual even postulates that as polymerisation is Lewis base initiated, and that water can act as a nucleophile, the aqueous content of latent marks can be enough to instigate polymerisation. Once one

cyanoacrylate molecule has undergone nucleophillic attack, it will readily react with further molecules and undergo propagation (**Bandey *et al.,* 2014**).

The PSDB recommended heating at 120°C, a process which took around 30 minutes. Its use is recommended for most non-porous surfaces, but has successfully been reported for use on plastics, tapes, metals, treated and untreated wood and even smooth rocks (**Lee & Gaensslen, 2001**). It is rarely recommended as a stand-alone visualisation technique, as the white superglue deposits can provide poor contrast, especially against light surfaces. For this reason cyanoacrylate fuming (CAF) is usually used in combination with other techniques, prior to powdering or fluorescent dyes. Rhodamine 6G (basic red 1) is a common staining dye used for such a purpose, which is still used despite being a suspected carcinogen, and needing an argon ion laser light source to aid visualisation (**Bleay *et al.,* 2012, Bleay *et al.,* 2017**). More recently, basic yellow 40 (BY40) has been investigated to mitigate the use of hazardous dyes used previously.

In more recent years, a number of methods and techniques have focused on detecting and identifying evidence within fingermarks, expanding the focus from only considering traditional enhancement and recovery techniques, to utilising molecular analysis.

Blood

Blood Background

Blood is a forensically relevant biological matrix that can be visible to the naked eye, or latent, both in stains and in fingermarks. Its visualisation is

accomplished through a different range of enhancement techniques according to whether its presence is within (or suspected to be in) stains of fingermarks.

Present in the vascular system of all vertebrates, blood has three essential functions. Transportation, mainly of oxygen and carbon dioxide between the lungs and cells; protection, providing antibodies and clotting factors; and regulation, mainly of body temperature and pH (**Ramotowski, 2012**). In humans, blood accounts for 7% of the total body weight and its components are plasma (~54.3%), erythrocytes (red blood cells) (~45%), leukocytes (white blood cells) (<1%) and thrombocytes (platelets) (<1%) by volume.

The plasma, comprising of roughly 90% water, contains dissolved solutes, with the majority of these being proteins. The most prominent of these are albumins (54%), globulins (38%) and fibrinogen (7%) (**Ramotowski, 2012**).

Platelets are a constituent of blood that in conjunction with coagulation factors, respond to bleeding by adhering together at the site of the injury, forming a blood clot.

White blood cells (WBCs) are the only blood cells to have nuclei and are responsible for protecting the body from infectious diseases. However in healthy adults they make up <1% total blood volume.

By contrast, approximately 84% (by number) of the cells in the human body are red blood cells (RBCs) (**Sender, 2016**). RBCs are composed of 95% haemoglobin (by dry weight, or 35% if you include water), making it a very abundant mammalian protein. Haemoglobin is a heterotetramer metalloprotein, meaning each haemoglobin protein is made up of four subunits; usually two alpha and two beta globin chains. The word haemoglobin is a portmanteau

combining *haem* and *globin*, reflecting that each subunit of haemoglobin has a haem molecule encapsulated by a globular protein. The haem molecule is a porphyrin ring structure centred around an iron atom. The iron atom has a vital role in oxygen transport, with each iron atom able to bind a single oxygen molecule through ion-induced dipole forces. Haemoglobin is present in all vertebrates (with the exception of crocodile icefish) and its detection via a selection of analytical techniques can confirm the presence of blood in an unknown sample.
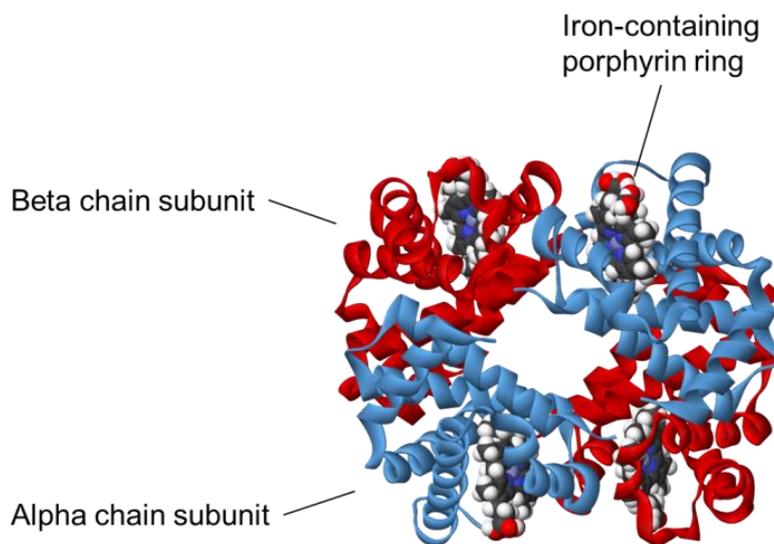
Figure 5. Haemoglobin molecules with the opposing alpha and beta chains highlighted in blue and red and the porphyin ring labelled. [*Adapted from: https://www.ck12.org/c/physical-science/protein*].

Forensic Use of Blood

Blood is commonly found at the scenes of some of the most serious violent crimes, such as murders or physical or sexual assaults, or where some level of force is used, such as breaking-and-entering scenarios. This type of evidence can originate from either a victim or a perpetrator. Its detection and visualisation can provide information around the dynamics of the bloodshed (from blood pattern analysis) but it can also yield specific chemical and biological

information about the individual who sustained an injury. This can come from exogenous compounds, such as drugs and alcohol, or endogenous molecules to reveal physiological information.

Bloody Fingermarks

Bloody fingermarks are a specific type of blood evidence that can be particularly useful, as they can provide the biometric information for suspect identification along with physical evidence of bloodshed and associative evidence.

Blood Enhancement Techniques

Background

CSIs use presumptive colorimetric tests to quickly ascertain if a stain is blood or not. They are considered presumptive because they target molecules not specific to blood, such as proteins or amino acids, and therefore can yield false positive results with other substances containing these species. Whilst not as reliable as confirmatory techniques such as liquid chromatography-mass spectrometry (LC-MS), they are able to give a rapid, presumptive, identification based on a chemical reaction eliciting a colour change in the reagents used. There are three common types of preliminary colorimetric test employed, targeting increasingly specific components of the blood namely amino acids, proteins and the haem molecule (**Bossers, 2011**). Like with fingermarks, the type of enhancement employed can depend on the surface they are discovered on.

Blood Enhancement Technique Categories

There are three categories of blood enhancement techniques, amino-reactive compounds, protein dyes and haem-reacted compounds, listed in order of increasing specificity.

Amino-reactive compounds

Amino acids are the single repeating building blocks of all proteins. Amino-reactive dyes interact with the globular proteins of the haemoglobin tetramer, as well as the fibrous and membrane proteins present in the blood, (**Ramatowski, 2012**) and free amino acids present in the plasma (**Bossers, 2011**). However, due to their non-specific nature, this is only a presumptive test.

Ninhydrin was previously mentioned for the detection of latent marks due to their amino acid content. However, ninhydrin can also be a very effective enhancement reagent for marks suspected of being contaminated with blood, due to the high concentration of amino acids (**Bleay** *et al.,* **2017**). Its reaction mechanism is show above in **Fig. 3**.

DFO produces a fluorescent product in the presence of amino acids (**Fig. 6**). 1,2-indandione also reacts with amino acids to form Joullié's Pink. Mechanisms for ninhydrin and DFO are illustrated by Wilkinson (**Wilkinson, 2000**).
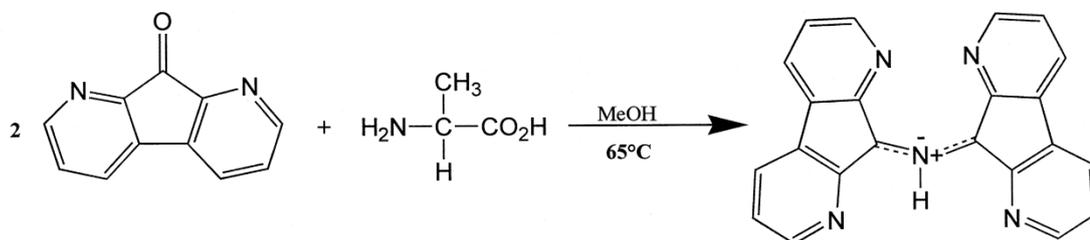
Figure 6. Reaction mechanism of DFO and amino acids to form a fluorescent product. *[Wilkinson, 2000].*

Protein dyes

Protein dyes, as the name suggests, target proteins. These reagents contain anionic (-) sulphonate groups that bind to the cationic (+) functional groups of the protein (**Fig. 7**). This reaction needs to take place in mildly acidic conditions, enabling the presence of cationic groups on the protein, whilst leaving the sulphonate groups deprotonated (**Bossers, 2011**). Common examples of these dyes are Acid Black 1 (AB1), Acid Violet 17 (AV17), Acid Yellow 7 (AY7), which yield different coloured dyes, and can be used to contrast with different coloured backgrounds. The application of excess dye has the capacity to remove significant quantities of blood, so a fixative step should be applied to samples prior to dye enhancement. Heat treatment was used until 1981, when it was superseded by immersion in methanol. In 2004, the Home Office switched to a water-based formulation over concerns of the hazards of methanol-based solvents (**Bleay *et al.,* 2017**).
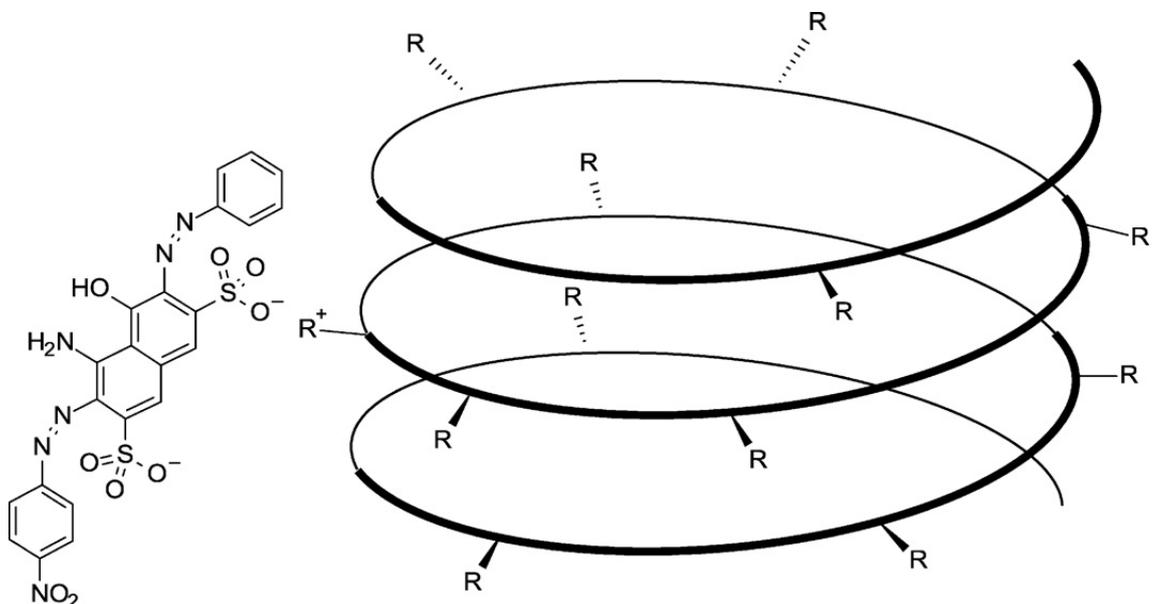
Figure 7. Schematic showing the interaction between Acid Black 1 and a protein. The 'R's represent cationic side groups that attract the anionic sulfonate groups in the dye structure. *[Bossers et al., 2011].*

Haem-reactive compounds (Peroxidase reagents)

Haem-reactive compounds utilise the presence of the haem moiety in the haemoglobin proteins in the blood. They are considered more specific blood detection reagents, than protein dyes, as they specifically target haem (only found in blood) rather than proteins which can be present in many other non-blood substances. Ramotwoski summarises current examples of these in the most recent edition of Lee and Gaensslen's Advances in Fingerprint Technology (3rd edition) (**Ramotowski, 2012**). Haem-reactive compounds can be divided into 3 groups; crystal test, catalytic tests and antibody tests.

Catalytic tests

Catalytic tests are presumptive, as they do not 'detect' haem, but use it to catalyse a colorimetric reaction. Haem catalyses an oxidation reaction, which can be exploited for the enhancement of blood stains by eliciting a colour change in the production of an oxidised product. Sometimes these reagents are known as peroxidase reagents, as hydrogen peroxidase is often used to facilitate these reactions (**Bossers, 2011**). However, this mechanism makes them susceptible to both false-positive and false-negative results, as other compounds can assist in the catalysation of these reactions.

A catalytic test developed by Kastle & Shedd in 1901 was modified to detect blood by Meyer in 1903, and the Kastle-Meyer test was devised. It was demonstrated that the oxidative properties of haem in the presence of hydrogen peroxide catalysed the oxidation of phenolphthalin into phenolphthalein, producing a vivid pink colour (**Kastle & Shedd, 1901, Meyer, 1903**).

Leuco-dyes are an alternative catalytic test to the Kastle-Meyer test. They work in the same way as the Kastle-Meyer test, with hydrogen peroxide initiating the

reaction, producing different colours using reagents such as leucomalachite green (LMG) and leucocrystal violet (LCV) (**Bleay et al., 2017**).

Another class of reagents form unstable intermediates that fluoresce in reaction to haem in the presence of a catalyst. Luminol and Fluorescein are two common fluorescent reagents, having two popular commercial formulations; Bluestar® and Hemascein®, respectively. These are evaluated and compared in a presentation by Bilous *et al.* and a publication by Seashols *et al.* (**Bilous et al., 2011**) (**Seashols *et al.,* 2013**).

Haemoglobin Variants

Haemoglobin Variants Background

Max Perutz determined the molecular structure of haemoglobin using x-ray crystallography in 1959, for which he was awarded the Nobel Prize for chemistry in 1962 (**Perutz, 1962**). He discovered that human haemoglobin is a tetramer containing four protein subunits. Each subunit is made up of alpha-helix structures, connected in a globin fold arrangement. Each subunit contains a heme group, a heterocyclic porphyrin ring with an iron ion at the centre. The heme found in haemoglobin is heme b, although other forms with slightly different functional groups bound to the porphyrin ring can exist, such as heme a and heme c, which are found in different pathways.
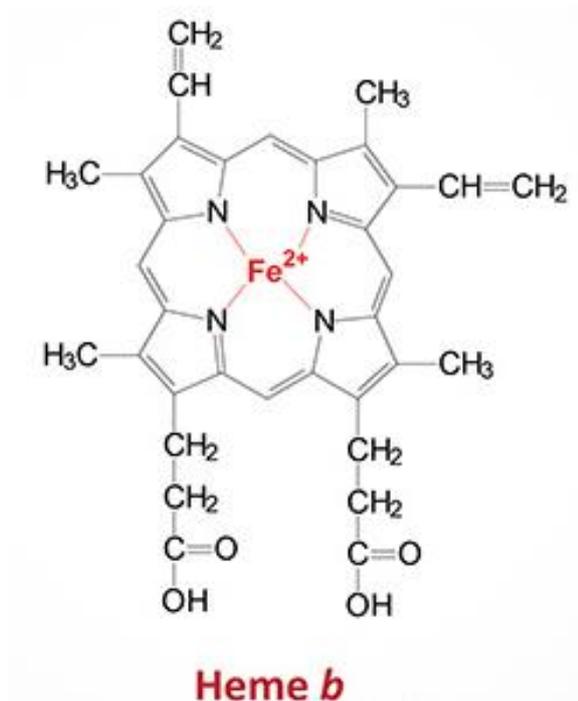
Heme *b*

Figure 8. Molecular structures of heme b. [*Adapted from: https://themedicalbiochemistrypage.org/heme-and-bilirubin-metabolism/*].

In adult humans, haemoglobin is usually comprised of two alpha (α) and two beta (β) subunits, designated as $\alpha_2\beta_2$. Infants are born with two α and two gamma (γ) subunits, which are gradually replaced by the body through the child's development. This is known at foetal haemoglobin (HbF), designated as $\alpha_2\gamma_2$. The different subunits vary by length and sequence. α subunits are 141 amino acids in length and β 146 amino acids. γ haemoglobin is 147 amino acids in length, but can exist in two forms (HbG1 and HbG2), with a single substitution mutation at position 136. At this position, where an alanine is found in the HbG1 form and a glycine is found in the HbG2 form. Foetal haemoglobin usually persists in small quantities in the body into adulthood (around 0.2 - 1.0%) (**Wild & Bain, 2016**). Certain disorders can cause gamma chain production to continue at an elevated level into adulthood. These are known as Hereditary Persistence of Foetal Haemoglobin (HPFH) diseases.

Structural variants can be caused by mutations in either the α or β chains. The chains are either denoted in Greek notation (such as with α, β, δ or γ), or if a mutation of one of these chains occurs, by the designated variant in superscript (such as β^C_2 or $\beta^{D\text{-}Iran}_2$) with the number of chains in subscript. A structural variant of HbA^0 can also persist into adult life; HbA^2 is comprised of two alpha and two delta chains ($\alpha_2\delta_2$) and normal quantities are 2 – 3.3% in the adult blood (**Wild & Bain, 2016**). But, like with foetal haemoglobin, disorders can cause elevated levels into adulthood. There are small margins between healthy and pathological HbA^2 values, so accurate quantification is important for diagnosis (**Mahmud *et al.*, 2021**).

Haemoglobinopathies

Other disorders can cause anomalies in the production of $\alpha_2\beta_2$ haemoglobin, causing mutations to occur in the α or β chains. These are termed haemoglobinopathies and can be categorised into two groups. Those that cause quantitative abnormalities, i.e. affect the rate of production of haemoglobin, are termed thalassemias. Those that cause qualitative abnormalities, i.e. structural modifications, are termed variants. These can be insertion, deletion or substitution mutation/s to the amino acid sequences in either the α or β chain, compared to $\alpha_2\beta_2$ or 'normal' haemoglobin, which can also be termed HbA^0.

'Normal' $\alpha_2\beta_2$ haemoglobin subunit chains	
Alpha chain	MVLSPADKTN VKAAWGKVGA HAGEYGAEAL ERMFLSFPTT KTYFPHFDLS HGSAQVKGHG KKVADALTNA VAHVDDMPNA LSALSDLHAH KLRVDPVNFK LLSHCLLVTL AAHLPAEFTP AVHASLDKFL ASVSTVLTSK YR
Beta chain	MVHLTPEEKS AVTALWGKVN VDEVGGEALG RLLVVYPWTQ RFFESFGDLS TPDAVMGNPK VKAHGKKVLG AFSDGLAHLD NLKGTFATLS ELHCDKLHVD PENFRLLGNV LVCVLAHHFG KEFTPPVQAA YQKVVAGVAN ALAHKYH

Table 1. HbA^0 haemoglobin α and β chain amino acid sequences.

The most common and well known haemoglobin variant is sickle cell anaemia, also denoted as HbS, but other variants have a (relatively) high incidence in the UK. In addition to the sickle cell variant, HbC, HbD, HbE and HbJ variants are encountered in Haematology Clinics in England.

HbS has a single substitution mutation at position 6 on the beta chain, where the glutamic acid has swapped to a valine. HbC also has a single substitution mutation at position 6 on the beta chain, where the glutamic acid has swapped to a lysine. HbE has a single substitution mutation at position 26 on the beta chain, where the glutamic acid is swapped to a lysine. Hb J-Baltimore has a single substitution mutation at position 16 on the beta chain. Hb D-Iran has a single substitution mutation at position 22 on the beta chain. HbA_2 has 10 substitution modifications on the beta chain, in positions 9 (S>T), 12 (T>N), 22 (E>A), 50 (T>S), 86 (A>S), 87 (T>Q), 116 (H>R), 117 (H>N), 125 (P>Q) and 126 (V>M). These mutations have been summarised in **Table 2**.

Each amino acid has a different molecular mass, due to their individual atomic structure. Therefore each haemoglobin variant subunit chain containing a mutation will have a different mass. These discrepancies in molecular mass could be used to distinguish each variant, but the intact proteins have a high molecular weight; theoretical *m/z* 15,742 for αHb, theoretical *m/z* 16,483 for βHb chains (**Bradshaw *et al.,* 2014**). Mass spectrometric techniques have low mass accuracy at very high molecular weights, and as some variant chains only differ by 1 amino acid, or a few Daltons, high mass accuracy would be required. 'Bottom-up' proteomics, a method where the long protein chains are cleaved into smaller peptide chains (<1,000 Da), is a better approach for this application. If the peptide containing the modification region could be detected

(and was unique to that variant), haemoglobin variants could be identified using this approach.

Variant	Chains	Mutation position	Beta chain amino acid sequence
HbA⁰	$\alpha_2\beta_2$	n/a	MVHLTPEEKS AVTALWGKVN VDEVGGEALG RLLVVYPWTQ RFFESFGDLS TPDAVMGNPK VKAHGKKVLG AFSDGLAHLD NLKGTFATLS ELHCDKLHVD PENFRLLGNV LVCVLAHHFG KEFTPPVQAA YQKVVAGVAN ALAHKYH
HbA²	$\alpha_2\delta_2$	β9, β12, β22, β50, β86, β87, β116, β117, β125, β126	MVHLTPEEK**T** AV**N**ALWGKVN VD**A**VGGEALG RLLVVYPWTQ RFFESFGDLS **S**PDAVMGNPK VKAHGKKVLG AFSDGLAHLD NLKGTF**SQ**LS ELHCDKLHVD PENFRLLGNV LVCVLA**RN**FG KEFTP**QM**QAA YQKVVAGVAN ALAHKYH
HbC	$\alpha_2\beta^C_2$	β6	MVHLTP**K**EKS AVTALWGKVN VDEVGGEALG RLLVVYPWTQ RFFESFGDLS TPDAVMGNPK VKAHGKKVLG AFSDGLAHLD NLKGTFATLS ELHCDKLHVD PENFRLLGNV LVCVLAHHFG KEFTPPVQAA YQKVVAGVAN ALAHKYH
Hb D-Iran	$\alpha_2\beta^{D\text{-}Iran}_2$	β22	MVHLTPEEKS AVTALWGKVN VD**Q**VGGEALG RLLVVYPWTQ RFFESFGDLS TPDAVMGNPK VKAHGKKVLG AFSDGLAHLD NLKGTFATLS ELHCDKLHVD PENFRLLGNV LVCVLAHHFG KEFTPPVQAA YQKVVAGVAN ALAHKYH
Hb D-Punjab	$\alpha_2\beta^{D\text{-}Punjab}_2$	β121	MVHLTPEEKS AVTALWGKVN VDEVGGEALG RLLVVYPWTQ RFFESFGDLS TPDAVMGNPK VKAHGKKVLG AFSDGLAHLD NLKGTFATLS ELHCDKLHVD PENFRLLGNV LVCVLAHHFG K**Q**FTPPVQAA YQKVVAGVAN ALAHKYH
HbE	$\alpha_2\beta^E_2$	β26	MVHLTPEEKS AVTALWGKVN VDEVGG**K**ALG RLLVVYPWTQ RFFESFGDLS TPDAVMGNPK VKAHGKKVLG AFSDGLAHLD NLKGTFATLS ELHCDKLHVD PENFRLLGNV LVCVLAHHFG KEFTPPVQAA YQKVVAGVAN ALAHKYH
Hb J-Baltimore	$\alpha_2\beta^{J\text{-}Baltimore}_2$	β16	MVHLTPEEKS AVTALW**D**KVN VDEVGGEALG RLLVVYPWTQ RFFESFGDLS TPDAVMGNPK VKAHGKKVLG AFSDGLAHLD NLKGTFATLS ELHCDKLHVD PENFRLLGNV LVCVLAHHFG KEFTPPVQAA YQKVVAGVAN ALAHKYH
HbS	$\alpha_2\beta^S_2$	β6	MVHLTP**V**EKS AVTALWGKVN VDEVGGEALG RLLVVYPWTQ RFFESFGDLS TPDAVMGNPK VKAHGKKVLG AFSDGLAHLD NLKGTFATLS ELHCDKLHVD PENFRLLGNV LVCVLAHHFG KEFTPPVQAA YQKVVAGVAN ALAHKYH

Table 2. Each Hb variant investigated in the study, with their denoted alpha and beta chain amino acid sequence, with mutations highlighted in red.

Mass Spectrometry

History of Matrix Assisted Laser Desorption Ionisation

Matrix assisted laser desorption ionisation (MALDI) was conceived by Karas, Hillenkamp and colleagues in 1985 (**Karas *et al.*, 1985**). This involved blasting a sample with a laser which simultaneously ionised the ablated molecules, which could then be qualified based on their mass-to-charge (*m/z*) ratio. MALDI was a refinement of the more rudimentary technique of laser desorption ionisation (LDI), but added a reagent that contained a *conjugated* ring structure to dissipate the laser's energy and reduce over-fragmentation of the analyte.

Around the same time, Tanaka *et al.* were working on a similar technique that involved the sample being mixed with cobalt powder and glycerol which allowed ionisation of larger species (**Tanaka *et al.,* 1988**). Tanaka later received one quarter of the 2002 Nobel Prize for Chemistry for his contribution to the MALDI ionisation of proteins.

Initially, MALDI was used to investigate the molecular composition of single points on the surface of an analyte, by firing the laser at a single 'pixel'. This is known as mass spectrometry profiling (MSP). This technique was modified by Spengler *et al.* (**Spengler *et al.,* 1994**) to incorporate the laser spot moving in a regular raster pattern over an area of the sample. This enabled spatial molecular information to be extracted and was defined as mass spectrometry imaging (MSI).

A group in Texas headed by Caprioli are credited with being the first to extend MSI to study biological samples, using MALDI Imaging to study proteins and peptides in tissue sections (**Caprioli *et al.,* 1997**).

A benefit of MALDI over other mass spectrometric techniques (such as LC-MS) is that MALDI produces singly charged ions. This has the benefit of making the spectra easy to interpret, as they are not populated with multiply charged ions. It is also a rapid analysis method. A full well plate could be profiled manually in less than one hour. The profiling process can also be automated, so each well can be analysed in as little as a few seconds. Critically, however, the main benefit of MALDI MS over conventional mass spectrometric techniques is the ability to map 2D spatial information of analytes on the surface of a sample. This application has been widely used in the clinical research sector, but also has applications in forensic science. For example, MALDI MS Imaging (MSI) has been demonstrated to locate drugs and medications within fingermark residues, that are absent in the background, indicating the individual that deposited the mark came into direct contact with these substances, rather than them being incidentally present on a surface that a mark was later deposited on (**Kaplan-Sandquist *et al.,* 2014, Sundar & Rowell, 2014, Groeneveld *et al.,* 2015**). MALDI MSI has also been shown to detect alcohol and drug metabolites exclusively in marks (**Bradshaw *et al.,* 2017**), the use of which can be linked back to an individual through conventional biometric identification. This is a critical advantage of using MALDI MSI over traditional 'gold standard' forensic analytical techniques, such as LC-MS (or LC-MS/MS). Very recently this molecular imaging capability has even been extended to offer 3 dimensions of spatial information (**Vos *et al.,* 2021**).

History of Electrospray Ionisation

Electrospray as a method of mass spectrometry ionisation was first reported by Dole in 1968 (**Dole *et al.* 1968**). It was combined with a quadrupole mass analyser by Yamashita and Fenn in 1984 (**Yamashita & Fenn, 1984**), and Fenn

was awarded one quarter of the 2002 Nobel Prize for Chemistry for his contribution to the development of soft ionisation methods in the late 1980s.

Initially, electrospray ionisation (ESI) was reserved for protein analysis, after Fenn *et al.* demonstrated its success at generating multiply charged ions from proteins up to 2000 Th. It is now used for complex mixtures and can be coupled to high-performance liquid chromatography (HPLC). Ionisation with an electrospray source is achieved by applying a voltage to a liquid forced through a capillary tube. This is done under ambient conditions and the high voltage, in the region of 3 – 6 kV, causes the liquid to form a fine spray of charged droplets in the form of a Taylor cone (**Hoffman & Stroobant, 2007**).

Fundamentals of Mass Spectrometry

Mass Spectrometry (MS) is an analytical technique enabling the measurement of the mass-to-charge ratio (*m/z*) of an ion, from which the atomic or molecular weight can be inferred. Mass spectrometry instrumentation can be categorised based on the method of ionisation, which is somewhat dependant on the phase and chemical nature of the analyte. It can also be dependent on the degree of fragmentation desired. 'Soft' ionisation techniques transfer small amounts of energy to the analyte, which causes little to no fragmentation of the analyte molecules. 'Hard' ionisation techniques deliver large amounts of energy to the analyte, causing significant fragmentation of the molecular structure. **Table 3** summarises these techniques.

Hard ionisation techniques	Soft ionisation techniques
• Electron Ionisation (EI) • Secondary Ion Mass Spectrometry (SIMS) • Laser Ablation Inductively Coupled Plasma Mass Spectrometry (LA-ICP-MS)	• Chemical Ionisation (CI) • Fast Atom Bombardment (FAB) • Atmospheric-Pressure Chemical Ionisation (APCI) • Electrospray Ionisation (ESI) -found in Desorption Electrospray Ionisation (DESI) and Laser Ablation Electrospray Ionisation (LAESI) • Matrix Assisted Laser Desorption Ionisation (MALDI)

Table 3. A summary of hard and soft mass spectrometric ionisation techniques and their acronyms.

Ionisation Source

The ionisation source is what defines the type of mass spectrometry instrumentation. The ionisation sources utilised during this PhD programme are ESI and MALDI. Both are considered 'soft' ionisation, as the energy transfer results in little to no fragmentation of the analyte molecules.

Whilst MALDI is widely used in both academic and commercial spheres, the exact mechanisms around the desorption and ionisation mechanics are contested. This has been the subject of debate and reviews elucidate the mechanisms in great detail, which have been revised in recent years (**Knochenmuss, 2014, Knochenmuss, 2016**). The most established of these will be summarised below.

Matrix assisted laser desorption ionisation (MALDI) is a method of irradiating an analyte with a laser in a way that releases them from the sample surface. This is more correctly termed ablation, but the term adsorption carries over from field desorption (FD), somewhat a predecessor of MALDI. The process of material ablation and the subsequent ionisation are interconnected, but its helps to consider them separate processes, not least because a large majority of the matrix-analyte molecules are released as neutral species (**Hillenkamp, 2007**). The intricacy around this complex process has led to several desorption/ionisation theories to be hypothesised, although these mainly focus on the ionisation stage.

Ionisation Models

The complex mechanism behind MALDI ionisation is contested, but it is generally accepted that the ionisation process takes place in two steps; a primary ionisation step concurrently with or immediately after the laser pulse, and a secondary ionisation process in the gaseous expansion of ablated material. The near-instantaneous primary ionisation process is the more disputed area of the mechanism. The four proposed mechanisms currently at the forefront of the research literature are the Photoionisation Model, the Lucky Survivor Model, the Thermal Proton Transfer Model, and most recently the Coupled Photophysical and Chemical Dynamics (CPCD) Model.

The interactions necessary to produce charged matrix-analyte molecules are even less well understood than the physics behind the sample ablation. For some of the proposed models it helps to differentiate the ionisation of the matrix and the analyte molecules into two separate stages.

Photoionisation Model (a.k.a. Gas Phase Proton Transfer Model)

Knochenmuss' first, widely accepted Photoionisation mechanism hypothesis (formerly the Energy Pooling Model) was published in 2002. It was proposed that MALDI ionisation was more dependent on fluence (the energy of a laser divided by the area of illumination) than peak power (the energy of the pulse divided by duration). This led to the description of the 'two-pulse effect' when it was observed that two sub-threshold laser pulses, offset by a slight time delay, individually giving a negligible signal, together resulted in a strong signal. It built on the pooling model outlined in previous experiments from 1998 (**Karbach & Knochenmuss, 1998**).

In the older photoionisation model, the matrix and analyte molecules are assumed to be initially neutral species in the crystalline structure. Photoionisation of the matrix molecules then initiates the process, with charge transfer to the analyte molecules occurring within the gaseous plume (**Knochenmuss, 2002**).

Considering a MALDI laser wavelength of 337 or 355 nm, resulting in photon energies of 3.6 and 3.3 eV respectively, photoionisation of a single matrix molecule would require energy transfer of at least 2 photons. MALDI lasers at the above wavelengths provide insufficient photon flux (number of photons per area per time) for resonant two-photon absorption (TPA) to occur. (Non-resonant TPA is where 2 photons combine to bridge an energy threshold larger than their individual capacity. Resonant TPA takes 2 single-photon transitions, via an intermediate state). In the gas phase, the ionisation threshold is too high to be overcome even by resonant TPA. However in solid crystalline form, due to a combination of lower ionisation potential and/or additional thermal energy, ionisation of matrix molecules can occur (**Hillenkamp & Peter-Katalinić, 2007**).
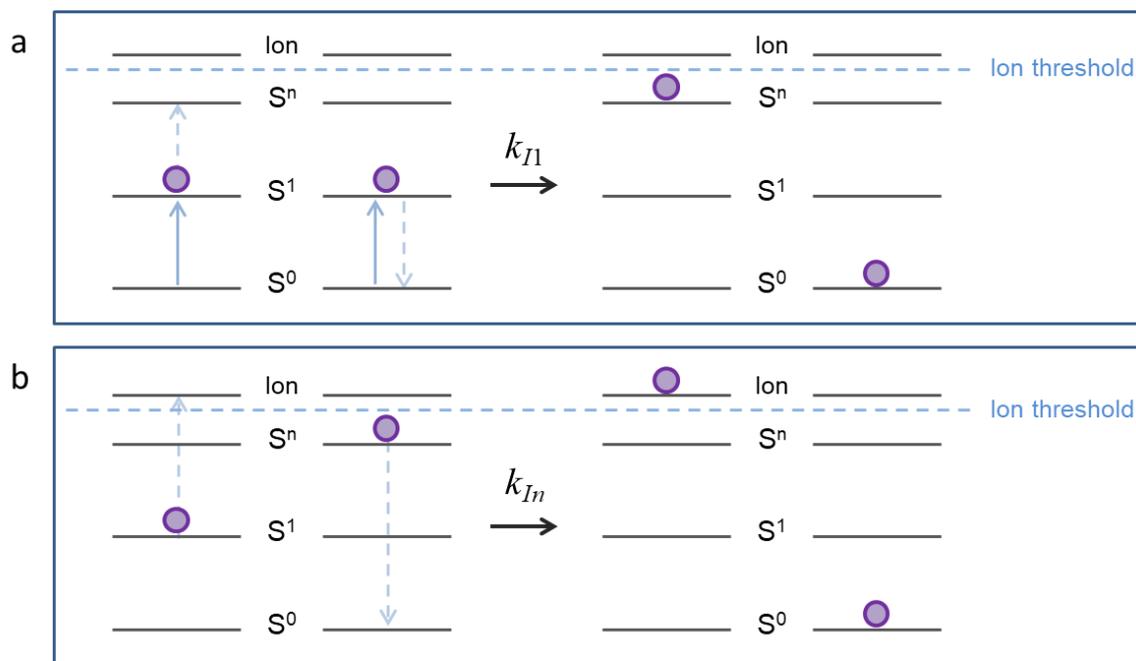
Figure 9. Representation of ion formation in the photoionisation model. (a) The 'two-pulse effect' shows the energy pooling of 2 photons in an S^1 excited state, to yield a photon in a higher excited state, S^n, and in the ground state S^0. (b) Pooling of S^1 and S^n photons to yield an ion plus photon in the ground state.

Molecular dynamics simulations, such as the pooling model, accurately account for the parameters observed experimentally, and consistently predict that "while clusters are formed early during the ablation process, their internal energy does not seem to suffice for a decay by matrix evaporation", which is one of the assumptions made by the Lucky Survivor model for ionisation, although the author notes that these prediction models contain oversimplifications and are restricted to small volumes and time restrictions by computer capacity (**Hillenkamp & Peter-Katalinić, 2007**).

However, one of the limitations of the photoionisation model was that no analyte was included in the modelling, and so more recent studies have tried to include this parameter in their hypotheses for ionisation mechanisms.

Lucky Survivor Model

Jaskolla & Karas (**Jaskolla & Karas, 2011**) provide "compelling evidence" to support the Lucky Survivor Model, previously postulated by Karas, Glückmann & Schäfer in 2000 (**Karas, Glückmann & Schäfer, 2000**). The 2011 model borrows elements of previous photoionisation and photochemical approaches, but tries to account for why only singly charged ions are detected, despite multiply charged clusters forming in the desorption event. This model also proposes gas phase protonation. In the Lucky Survivor Model, the assumption is made that sample proteins already exist as charged species in the matrix-analyte crystal structure (**Hillenkamp and Peter-Katalinić, 2007**).

The Lucky Survivor Model theorises that analyte molecules retain their charge states from being in solution. Peptides and proteins, under normal preparation conditions, are protonated species, and would be crystallised as positively precharged analytes, present opposite their counterions. The original Lucky Survivors Model from 2000, proposed that either ablated clusters had equal numbers of ions and counterions and were without net charge, or those clusters carried a net charge, initiated by charge separation upon dissociation of the sample.

Thermal Proton Transfer Model

The modern Thermal Proton Transfer Model is built upon the Polar Fluid Model proposed in 1998 by the groups lead by Beavis and Chait, which was the first to suggest that thermal interactions, rather than electronic, were (at least partially) responsible for ion formation. The Polar Fluid Model was also the first to decouple the matrix and analyte ionisation processes (**Niu *et al.*, 1998 & Chen *et al.*, 1998**).

Next, Wang's group were the first to publish a 'quantitative interpretation of ion desorption' based on the transfer of thermal energy processes. They used experiments to demonstrate that thermal proton transfer reactions could occur on a solid state surface, using the matrix ion-to-neutral ratio of 10^{-5}, reported in previous literature. This model assumed chemical and thermal equilibrium in a solid state, contrary to prior models that assumed chemical equilibrium was achieved in the gas phase (**Lai *et al.*, 2010**).

Wang's thermodynamic theory was backed up by Kim's group, who agreed that thermal reactions cause ion formation, but argued that matrix ion-to-neutral ratios were likely to be in the region of 10^{-7} (**Bae *et al.*, 2012**). Previously, thermally induced reactions were not popularly believed to contribute significantly to the MALDI ionisation process, due to low ion-to-neutral ratios (10^{-4}) providing negligible thermal contribution (**Chu *et al.*, 2014**). Kim's group followed up their 2012 study the following year with a study claiming quantitative reproducibility and a proposed gas-phase mechanism for the ionisation process, involving two thermal reactions, the 'autoprotolysis of matrix molecules' and 'matrix-to-analyte proton transfer' (**Ahn *et al.*, 2013**).

A study in 2014 by Ni's group was the first to publish the 'thermal proton transfer mechanism', which they claimed was responsible for MALDI ionisation (**Chu *et al.*, 2014**). Recent studies had shown that under certain conditions the ion-to-neutral ratio can be 10^{-7}, and therefore thermally induced reactions contributed more than first thought. Ni's group used a polar fluid model to calculate the ion-to-neutral ratios, and discuss that although the calculated values of this model were one order of magnitude out from those obtained experimentally, this is significantly closer than the values predicted by other models such as the photoionisation model, which differ by four orders of

magnitude. This led them to conclude that thermally induced reactions played a key role in MALDI ionisation.

Ni's group again re-evaluated the ion-to-neutral ratio in another study in 2015. They noted that the matrix and analyte ion-to-neutral ratios were not separately defined in previous models, and their experiments revealed that they vary considerably. They summarised that matrix ion-to-neutral ratios are in the range of $10^{-9} - 10^{-6}$, and analyte ion-to-neutral ratios are in the region of 10^{-3} to 10^{-8} depending on the proton affinity of the analyte (**Lu *et al.*, 2015**).

Coupled Photophysical and Chemical Dynamics Model

By 2014, Knochenmuss had revised his Photoionisation Model hypothesis. Following an extensive investigation of ion yields using various matrices (applying different wavelength and fluence parameters) by Soltwisch *et al.* (**Soltwisch, 2012**), Knochenmuss used the dataset to develop his Coupled Photophysical and Chemical Dynamics Model (CPCD). Knochenmuss stated that unlike thermal models, the CPCD model was consistent with the empirical data provided by Soltwisch *et al.* (**Knochenmuss, 2014**).

In Knochenmuss' next paper on the new model in 2016, he defines the boundaries of the primary and secondary ionisation processes. Primary processes concern the initial charge separation and involve matrix molecules. Secondary processes involve the ion-molecule reactions, which leads to the generation of the ions that are detected (**Knochenmuss, 2016**).

Mass Analyser

Mass analysers separate ions based on their mass-to-charge ratio. As with the selection of ionisation sources available, there are several types of mass

analysers based on slightly different principles. Although they are all either static or dynamic electric or magnetic fields (or combinations of both), they achieve separation in different ways. The types of mass analyser are summarised in Table **4**.

Mass Analyser	Acronym	Principle of separation
Electric Sector	E/ESA	Kinetic energy
Magnetic Sector	B	Momentum
Quadrupole	Q	m/z (trajectory stability)
Ion Trap	IT	m/z (resonance frequency)
Time of Flight	TOF	Velocity (flight time)
Fourier Transform Ion Cyclotron Resonance	FT-ICR	m/z (resonance frequency)
Fourier Transform Orbitrap	FT-OT	m/z (resonance frequency)

Table 4. The different types of mass analyser, their acronym and their principle of separation [*Adapted from Hoffmann & Stroobant, 2007*].

Time of Flight (TOF)

The first types of mass analysers were time of flight (TOF) tubes, patented by Stephens in 1952. Following the generation of ions, these charged species are accelerated by an electric field into a long flight tube under vacuum where the initial separation is amplified. The electric field imparts equal kinetic energy to all ions of the same charge. Because kinetic energy is equal to $\frac{1}{2}mv^2$, (where m = mass, v = velocity) for species of the same charge, ions with a higher mass will have a lower velocity through the flight tube and vice versa. In a flight tube of known length, the time for each ion to reach the detector is related to the velocity and therefore the *m/z*. Generally, users simply calibrate the mass range rather than calculating the *m/z* of the detected ions.

Early TOF flight tubes were linear, but later incorporated reflection technology to increase the path length, and to time-focus the ions onto the detector, correcting for small differences in the kinetic energy of the ions, therefore increasing separation and resolution as a result (**Mamyrin, 2001**).
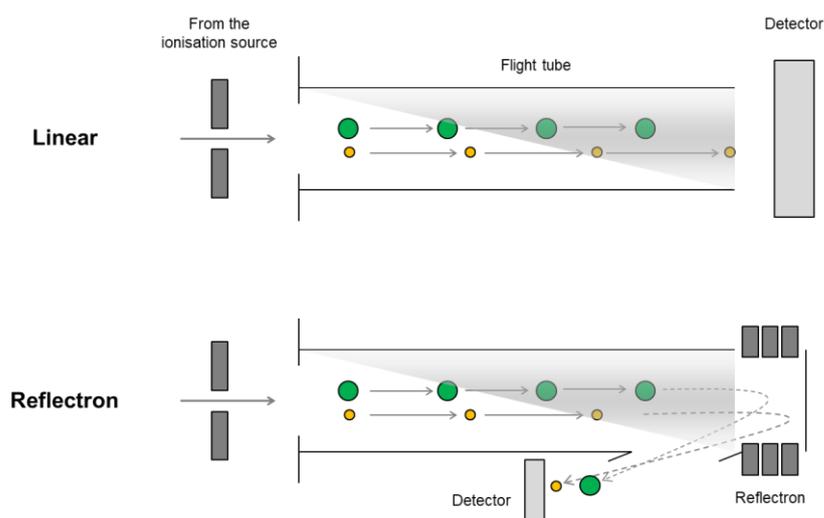
Figure 10. Flight path of ions in a linear TOF vs. in a reflectron TOF.

Quadrupole (Q)

Quadrupoles separate ions by exploiting their charged properties when exposed to an electric field. Quadrupole analysers oscillate ions down a channel between four parallel rods. These rods can be cylindrical or hyperbolic, connected in opposite pairs. Direct current (DC) and radio frequency (RF) potentials are applied to the pairs of rods; either 'V', a radio frequency or 'U', a DC offset, both measured in volts.

(**Hoffman & Stroobant, 2007**, **McGuinness & Giannoukos, 2019**).

This interchanging electric field configuration generated by the rods alternates between attracting and repelling charged molecules, allows ions with a specific mass-to-charge ratio to pass in a spiral through the channel between the rods and to reach the detector at the other end; these ions have a 'bounded oscillation' trajectory. Ions falling outside of the specified mass-to-charge ratio end up having an unstable trajectory, or unbounded oscillation, and the unstable spiral causes them to collide with the rods, resulting in neutralisation of the ions, which then do not reach the detector.
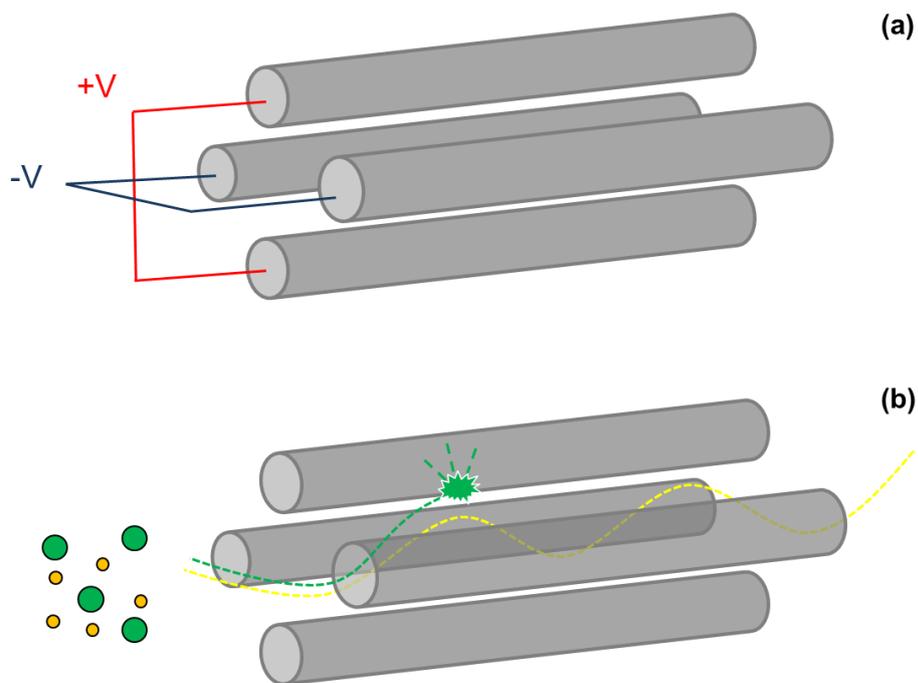
Figure 11. A typical quadrupole configuration with electric potentials indicated **(a)**, and ions of a selected *m/z* passing through the channel to the detector (yellow). Ions outside the selected range have an unstable trajectory, and collide with the rods resulting in discharge of the ions, meaning they don't reach the detector (green) **(b)**.

A single quadrupole can be operated in two filtering modes; in 'full scan' or 'selected ion monitoring'. In full scan mode, the DC and RF voltages are ramped whilst keeping their ratio the same, to allow sequential scanning of ions with different mass-to-charge ratios. In selected ion monitoring mode, the quadrupole has a specific voltage applied to maintain a constant electric field and only allow ions with specific mass-to-charge ratios to exit the quadrupole and reach the detector. This mode can be used in conjunction with further analysers to perform MS/MS analysis by allowing fragmentation of specific parent ions.

Mathieu Stability Diagrams

Émile Léonard Mathieu worked on various areas of mathematical physics in the 19th Century, but his work on the vibrations of elliptical surfaces is relevant here. His work focused on drum skins, but translates to the motions of ions in

quadrupoles. The electric fields applied to the pairs of rods relate to stability of ion trajectories in the x and y dimensions, represented by the x and y axes of the Mathieu stability diagram. Where they combine is a region of combined stability (**Fig. 12a**). When only an RF voltage is applied, the quadrupole acts as an ion guide, and ions are focused through the quadrupole. When a DC voltage is applied, ions are forced from the x axis at a gradient, represented by the scan line (**Fig. 12b**). The gradient is proportional to the ratio between the DC voltage (U) and the RF voltage (V). The scan line (U:V ratio) is set to pass through the tip of the stability diagram. The closer the scan line is to the apex, the fewer the stable ions and the more selective the ion filter (**Fig. 12c**).
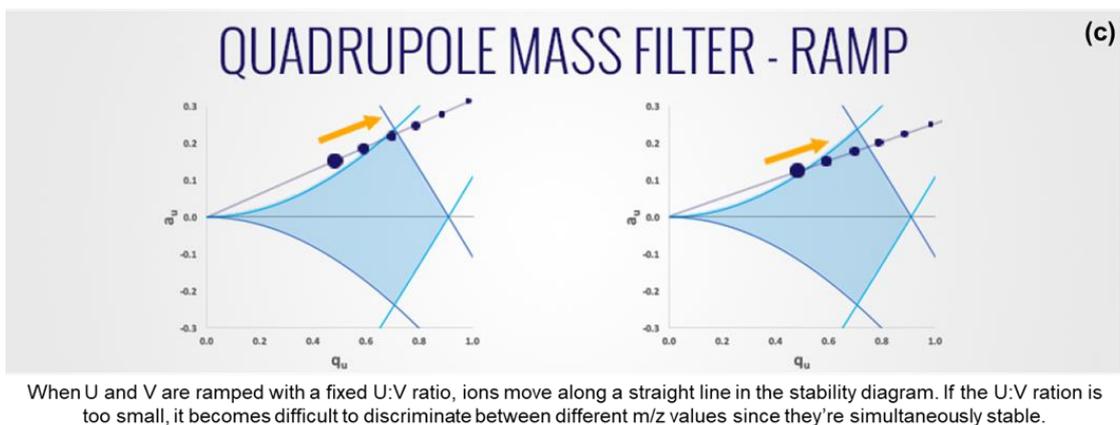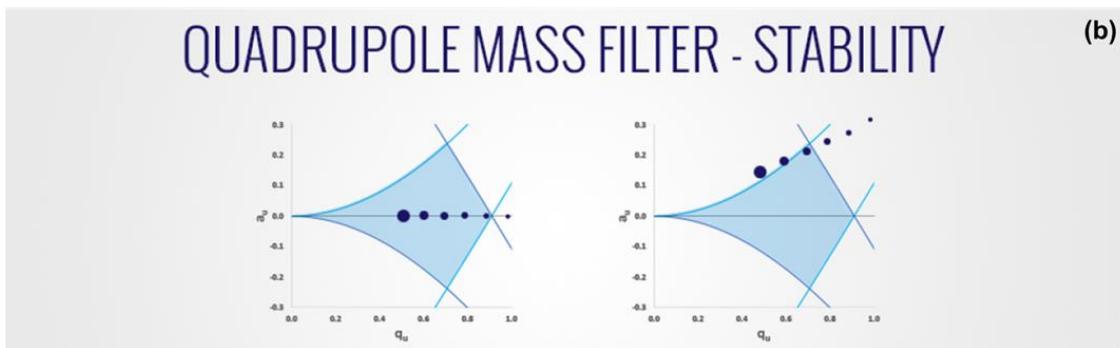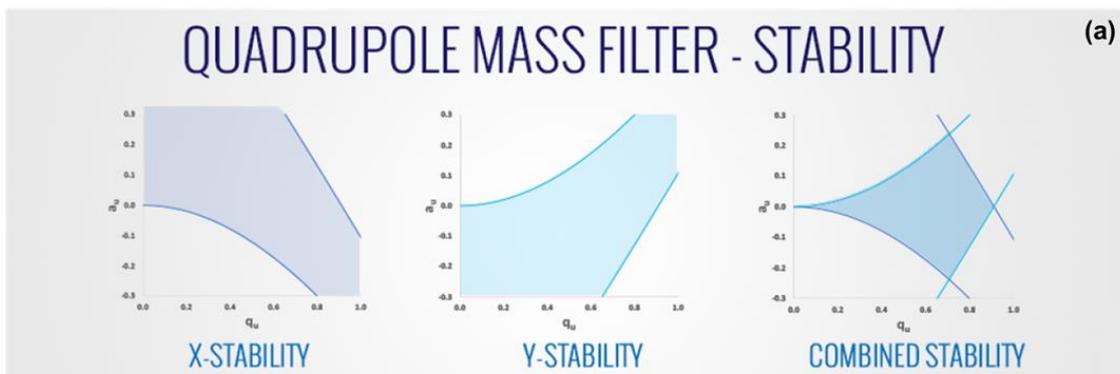
Figure 12. (a) the regions of x axis stability, y axis stability and the region of combined stability on the Mathieu stability diagram. **(b)** a quadrupole with only an RF voltage applied, in ion guide mode vs. a quadrupole with a DC voltage appied to create a gradient scan line. **(c)** ramping at a fixed U:V ratio can allow quadrupoles to perform in ion selective mode, allowing the quadrupole to transmit ions of a specific *m/z. [adapted from: http://www.massspecpro.com/technology/mass-analyzers/quadrupole-mass-filter]*

Multi-analyser systems

Quadrupole analysers can also be used in tandem. Each quadrupole can

operate in two modes; mass selective (mass filter) mode or as an ion guide (in

RF only mode). They are typically arranged as a triple quadrupole (QQQ), with Q1 and Q3 functioning as mass spectrometers. Q1 and Q3 operate in mass selective modes, which can filter ions by mass-to-charge. The voltages applied to the rods can be tuned to let ions of a certain mass-to-charge travel in a stable trajectory down the centre of the quadrupole. Ions not matching the selected m/z will travel down the quadrupole in unstable oscilations and discharge against the rods. The central quadrupole, Q2, functions as a collision cell, and ion guide, which is not mass selective, but produces a narrow beam of low energy, focused ions.

Fragment ion or product ion scan

MS1 →	CID	MS2
Precursor selection	Fragmentation	Fragment detection

Parent ion or precursor ion scan

MS1	CID	MS2 →
Precursor scan	Fragmentation	Fragment detection

Neutral loss scan

MS1	CID	MS2
Precursor *m* scan	Fragmentation	Fragment *m-a* scan
Q1	**Q2**	**Q3**
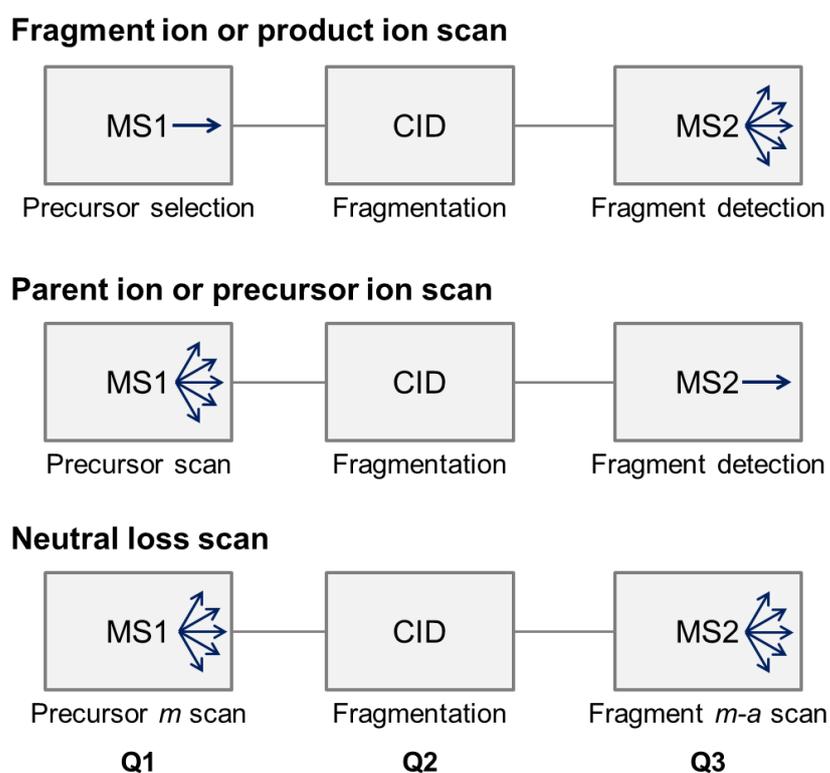
Figure 13. Scan mode configurations for triple-Q mass spectrometers.

The importance of fingermarks as forensic evidence

As discussed in their introduction, fingermarks have long been considered physical evidence; biometric information capable of identifying an individual from their unique ridge pattern. However, relatively recent analytical techniques can be employed to reveal a snapshot of molecular information. This has the potential to provide intelligence on a person's lifestyle and activities prior to leaving a mark. This has useful operational applications concerning crime scene marks, especially if traditional database matching was unable to provide a suspect. This can be the case when the original marks are smudged, partial, empty or overlapping. This could also be down to sub-optimal enhancement, which is dependent on the application of the fingermark enhancement technique (FET) and the correct FET selection. Other variables affecting the quality of a recovered mark include the deposition surface, environmental factors and time between deposition and recovery. Finally, even if the marks recovered are of the highest quality, if the suspect's mark is not present in a fingerprint database, this evidence results in a dead end (unless the suspect subsequently ends up in custody).

Fingermarks by their nature incorporate sweat from the pores in the fingertips. This sweat acts a solution for endogenous substances originating inside the body, for example lipids and proteins, which can give an insight into the physiological state of a person. These biomolecules have the potential to act as pathological, pharmacological or biological biomarkers. Fingermarks can also incorporate traces of external contaminants that the fingers or hands may have come into contact with; these are known as exogenous substances. Examples of these sorts of compounds which could offer forensic intelligence include illicit

drugs, explosive powders and condom lubricants. Additionally, compounds that would not necessarily appear to provide information pertaining to a crime may give CSIs an advantage when trying to identify individuals. These compounds can include medications, paint, make up and blood. There remains a third distinct class of substances; semi-exogenous species, that do not originate in the body, but may be present as a result of ingestion, inhalation, absorption or injection of foreign substances, and then excreted in one form or another. These can be present 'intact' or broken down as a result of the bodies many metabolic pathways. Examples of this class of substances would include food and drink, drugs, medications and cigarettes. Therefore, molecular analysis of fingermarks can attempt to provide further forensic evidence pertaining to a suspect, in the form of lifestyle or physiological information if the biometric information from a fingermark cannot be matched; for example that they are a smoker if nicotine is detected in their fingermarks (**Benton *et al.,* 2010**).

Furthermore, if an identification is made through matching a recovered fingermark to a print on record, chemical analysis can provide information on previous activities, and can link that information to the suspect. For example, if drug metabolites are detected, this could give an insight into the frame of mind of an individual at the time the fingermarks were deposited (**Bradshaw *et al.,* 2017**).

Prior Fingermark Analysis

Prior research in the molecular analysis of fingermarks

The techniques employed for the molecular analysis of fingermark residue can broadly be classified as spectroscopic or spectrometric techniques. Of the spectrometric methods, desorption electrospray ionisation (DESI) was the first

technique reported in the literature for fingermark analysis in 2008 (**Ifa *et al.,* 2008**). Cooks' group were able to map several exogenous compounds within the ridges of the fingermark using DESI MSI, after spiking participant's fingertips, from a variety of surfaces including glass, paper and plastic. These compounds comprised of the drugs cocaine, tetrahydrocannabinol (THC) and cannabidol (Cbd), and components of the plastic explosives Semtex and C4. The group also detected and identified endogenous constituents such as fatty acids and triglycerides. The identified compounds were confirmed with MS/MS. A benefit of DESI is the lack of sample preparation required and that the electrospray process is minimally destructive. Even in 2008, the group was able to achieve a spatial resolution of 150 x 150 μm, which was a lower spatial resolution than spectroscopic techniques such as FTIR and Raman, but conversely offered better chemical specificity.

Following this initial pioneering concept of the molecular fingermark analysis, other groups used alternative mass spectrometric instrumentation to investigate the content of fingermark residue; in addition to DESI these included MALDI, direct analysis in real time (DART), surface assisted laser desorption ionisation (SALDI) and secondary ion mass spectrometry (SIMS), each with their own benefits and drawbacks. Based on a broad foundation of research, MALDI was identified as a useful process (in specific circumstances) in the Home Office Fingermark Visualisation Manual in 2014. The authors of the Fingermark Visualisation Manual had worked closely with academic partners to evaluate some pioneering non-chemical fingermark visualisation techniques, of which MALDI and SIMS were the only mass spectrometric techniques included (**Table 5**) (**Bandey *et al.,* 2014**).

Process	Category	Institution
Scanning Electron Microscopy	B	[not specified]
Scanning Kelvin Probe	B	Swansea University
ATR-FTIR	C	Imperial College London
MALDI MSI	C	Sheffield Hallam University
SIMS	C	University of Surrey

Table 5. A list of novel non-chemical imaging fingermark enhancement techniques included in the Home Office Fingermark Visualisation Manual (2014).

Any category classification that isn't Category A is not used routinely for operational work, but might be recommended in certain circumstances - 'no other options available; all Category A options have been exhausted; niche application; or lack of equipment for other processes.' Niche applications suggested for SIMS include 'filling in' ridge detail that was not successfully enhanced with traditional physical or chemical FETs, or to provide depth information, for example to indicate whether a mark was deposited before or after writing. Niche applications identified for MALDI MSI also include 'filling in' missing ridge detail, but also for 'identifying and mapping the distribution of unusual and/or significant contaminants (**Bandey *et al.,* 2014**).

Prior Fingermark Research: MALDI Profiling and Imaging (Francese's Fingermark Research Group)

Prior to its inclusion in the Fingermark Visualisation Manual, much of the proof of concept research involving MALDI MSP and MSI of fingermarks in the UK had been carried out by Francese's group, the Fingermark Research Group (FRG). Early work began in 2009 with a publication reporting the imaging of endogenous lipids. Detection and mapping of oleic acid and its degradation

products within groomed and ungroomed marks was carried out. A decrease in the relative intensity of oleic acid species was found to give an indication of the age of the marks over a 7 day aging period. In addition, mapping of specific *m/z* values of endogenous species was found to provide more complete ridge detail that an optical image of the fingermark prior to analysis, leading the group to conclude that MALDI MSI was effective as both an enhancement and species detection technique (**Wolstenholme *et al.,* 2009**).

The next paper from this lab was the development of the dry-wet method of MALDI α-cyano-4-hydroxycinnamic acid (α-CHCA) matrix application, and proposed the integration of dry matrix powdering with current forensic visualisation workflows, if it was anticipated that latent marks would subsequently be sent for MALDI analysis (**Ferguson *et al.,* 2011**).

The same group then undertook a proof-of-concept study into the determination of different condom brands, from analysing the different lubricant formulations present in the fingermarks of donors who had handled condoms, with a view to applying this to identifying sexual offenders. Sexual offenders are becoming increasingly 'forensically aware' and many are using condoms in an attempt to limit the transfer of DNA evidence when carrying out such crimes. This study was able to resolve the condom brands Condomi max love and Trojan-Enz through the detection of polyethylene glycol (PEG) and nonoxynol-9 (and adducts), respectively in the marks after handling the condoms. MALDI MSI analysis again has the benefit of linking an individual through biometric identification to an act (i.e. that of handling a condom) that would be an unusual activity, and providing investigators with additional circumstantial evidence against a suspect. If that suspect was to be in possession of condoms of the same brand/type as those identified through molecular analysis of fingermarks

at the scene of a sexual assault or from a rape swab kit, this would provide more significant corroborative evidence (**Bradshaw *et al.,* 2011**).

A following study, combining the improved enhancement and novel species detection methods noted in the first paper, sought to resolve overlapping fingermarks. Overlapping fingermarks make it very difficult for examiners and automated systems to match against a fingerprint database, as it is not always clear which ridge detail belongs to which mark, because conventional physical and chemical FETs visualise all marks present. Bradshaw *et al.* reported on the use of MALDI MSI to identify characteristic ion signals, pertaining to compounds found uniquely within individual fingermarks in sets of overlapping marks, to resolve them. The group first targeted condom lubricants as an exogenous contaminant, with signals coming from a spermicidal lubricant, and were able to separate overlapping marks from those that had handled the condom from those that hadn't. The group were also able to detect and identify caffeine metabolites as a semi-exogenous compound, in the marks of a donor who had previously drunk coffee, and resolve them from the marks of a donor who had not. This was the first time that consumption and excretion of semi-exogenous substances had been reported to have been detected in fingermark by MALDI MSI. This demonstrated an additional forensic application, as this route mimicked the consumption of drugs, and revealed it might be possible to distinguish between fingermarks of suspects that had consumed drugs versus those that had only handled them through the presence of metabolites (**Bradshaw *et al.,* 2012**).

The resolving of overlapping marks had been attempted before by other groups such as Ifa *et al.,* Chen *et al.,* and Tang *et al.,* but by using DESI, IR spectroscopy, and laser desorption ionisation (LDI) respectively. In addition,

these studies had spiked the donors fingermarks with artificially elevated quantities of the analyte or lipid material; the Bradshaw *et al.* study investigated groomed, ungroomed and natural marks, better representing crime scene marks (**Ifa *et al.,* 2008**) (**Chen *et al.,* 2009**) (**Tang *et al.,* 2010**).

Further research from the Francese group was the refinement of the dry-wet method of MALDI matrix application, specifically for analysing latent marks, and the comparison of this application technique against convention spray coat methods. This validation study, building on the advent of the method from their previous paper, found that dusting latent marks with fine α-CHCA matrix powder prior to solvent spray coating resulted in better clarity ridge detail and enhanced ion signal intensity (**Ferguson *et al.,* 2013**).

The group summarised their previous work in a review article in 2014. They reported on the capability of MALDI MSI to achieve sufficient spatial resolution for the visualisation of level 1 and level 2 fingermark detail in molecular images, but noted that, for level 3 detail, with state-of-the-art technology of the time, may approximately 100 hours acquisition time as a result of the enhanced spatial resolution. The authors concluded that the ability to detect hundreds of species within a single mark had the potential for enhancing the quality of physical evidence through the stitching together of multiple molecular images. This is advantageous over conventional FETs, which usually only target one species. The authors summarised their research towards resolving overlapping marks. In addition to enhancing the physical information exploited through MALDI MSI analysis, the authors also reviewed their research involving the chemical information that can be acquired from marks, including drugs, condom lubricants, and beverages (prior consumption of coffee through the detection of caffeine) (**Francese *et al.,* 2013**).

The group then worked in collaboration with the research arm of the Home Office, the Centre for Applied Science and Technology (CAST) (the authors of the Fingermark Visualisation Manual, the definitive manual for the UK enforcement authorities) to carry out the validation study required to have the technique listed in the Fingermark Visualisation Manual. The team demonstrated that chemical information could be obtained from marks via MALDI MSI following most common FETs. In addition, they were able to show that under certain conditions, greater and better clarity ridge detail could be visualised compared to conventional visualisation techniques, for example to overcome the over-enhancement of a mark on a PET bottle visualised with VMD. This marked the transition of MALDI for fingermark analysis from the confines of academic research labs towards operational deployment for real crime scene marks (**Bradshaw *et al.,* 2013a**).

Bradshaw *et al.* built upon their previous proof-of-concept condom identification study from 2011, expanding on this work to encompass a multi-disciplinary approach to identify a larger variety of condom brands and types. The group utilised MALDI MSI, MS/MS Ramen microscopy and attenuated total reflectance-fourier transform infrared (ATR-FTIR) spectroscopy, in a multimodal approach. They expanded the capability of the technique from being able to differentiate two condom brands, to differentiate 6 condom types, by detecting discriminate signals. These were tentatively identified as series of various polymer ions. As ATR-FTIR was only employed for one type of condom, it is feasible that advances in instrumentation regarding sensitivity and resolution since this study would enable the entire analysis to be done by MALDI MS, streamlining the process, and utilising one analytical technique for all samples. This would make the technique more attractive for processing case work, where

a comprehensive database of condom molecular profiles could be used to identify condom brands/types and link the biometric information potentially obtained from fingermarks. This would provide supporting circumstantial evidence if the suspect was found to be in possession of the same brand of prophylactic (**Bradshaw *et al.,* 2013b**).

Addressing the issue of the potential detection of isobaric species, reported elsewhere in the literature, the FRG reported the MS/MS analysis of endogenous oleic acid, confirming its presence in a mark through the detection of its product ions. Whilst the presence of additional isobaric species cannot be excluded, MS/MS offers a valuable confirmation of the identity of an analyte through the fragmentation of the parent ion and resulting product ion fragments.

In addition, the FRG reports their first analysis of a real crime scene mark. The mark provided CSIs with insufficient ridge detail for an identification, but preliminary MALDI MS imaging analysis indicated the presence of cocaine, with a strong ion signal at *m/z* 304. Again, MALDI MS/MS was carried out and cocaine product ions were detected, confirming the suspected presence of cocaine in the recovered mark (**Bradshaw & Francese, 2014**).

Following the success of the detection of the first semi-exogenous substances by MALDI MSI in 2012, the group next looked to investigate the detection of a wide range of illicit drugs in fingermarks. 17 compounds were investigated, comprising of parent drugs and metabolites from 5 illicit drug classes: amphetamines, alkaloids, opioids, cannabinoids and designer drugs. The group also hypothesised that the detection of metabolites would indicate drug use, whereas lack of metabolites would indicate handling, in a user vs. dealer scenario. Ungroomed marks were prepared by spiking 50 µL of the compounds

individually onto glass slides and allowing the solvent to evaporate before rubbing clean fingertips in the substance and depositing marks on aluminium TLC sheets. A further set of marks was prepared using the same method, but with quantities derived from preliminary limit of detection (LOD) experiments. The marks were then analysed by MALDI MS and the identity of the compounds confirmed with MS/MS. The group also demonstrated the compatibility of the technique with common fingermark enhancement techniques (FETs), cyanoacrylate fuming (CAF) and vacuum metal deposition (VMD), using the Fingermark Visualisation Manual procedure for the VMD enhancement. Data showed that whilst CAF negatively affected the ion signal of the analytes, the ion signals were 'as strong or even stronger', following VMD enhancement. CAF and VMD were also used in combination with each other and in sequence with Basic Yellow. Compatibility studies are crucial to undertake during research to ensure that these techniques could be viable operationally, where it is likely that marks may have been previously enhanced to visualise them at a crime scene before being made available for mass spectrometric analysis.

Furthermore, this was the first time that illicit drugs and metabolites had been mapped with MALDI MS imaging within fingermarks, critically, in tandem with FETs. Clear ridge detail was visible when selecting ion signals from cocaine, amphetamine, heroin, and respective metabolites. THC and tetrahydrocannabinolic acid (THCA) resulted in poor image quality without ridge detail, but still demonstrated it was possible to detect THC and at least one metabolite. This preliminary data demonstrated the ability to detect a range of illicit drugs within artificially spiked fingermarks, in "physiological" quantities, using MALDI MS imaging, without encountering ion suppression resulting from the fingermark components (**Groeneveld *et al.,* 2015**).

In 2016, Francese's research group commenced MALDI MS analysis of fingermark lifts, i.e. latent marks that had been enhanced and then 'recovered' with lifting tape. Prior to this, with a few exceptions, experiments had been carried out with artificially deposited marks on surfaces under laboratory conditions, where the deposition surface had always been appropriate to insert into the mass spectrometer sample chamber. However, this would not always be practical for operational crime scene marks, as marks may be discovered on surfaces that are not appropriate to be inserted into a sample chamber, being too large (in the case of doors for example), curved surfaces (such as bullet casings) or being impractical to be removed from their original location (such as surfaces on a train or plane). In these instances, it would be beneficial to use fingermark lifts to recover the marks from their original locations. This is a technique already employed by crime scene investigators, and as such, it may not be possible for 'primary' lifts to be obtained for mass spectrometric analysis; therefore this work investigated both 'primary' lifts, and also 'secondary' lifts, where subsequent lifts were recovered a second time when a 'primary' lift had already been taken (**Bradshaw *et al.,* 2016**).

The group worked with West Yorkshire Police to develop a method to process these lifts to produce 'optimum results'; either enhanced ridge detail or chemical information pertaining to the suspects lifestyle prior to deposition, both through MS analysis. They examined various scenarios with the lifts, including no further enhancement, re-enhancement using the same FET, and further enhancement using the dry-wet method (**Ferguson *et al.,* 2011**). The scope of the study covered powder FETs, namely: aluminium, TiO_2 and carbon black. They found that the dry-wet method was less successful on lifted marks, possibly due to "sub-optimal matrix-to-analyte ratio" as a result of the decreasing quantity of

analyte on a primary or secondary lift, compared to applying matrix directly to deposited marks. However, they reported that whilst aluminium and carbon powdered marks yielded similar ridge clarity across the remaining three methods (no further enhancement, re-enhancement using original FET, re-enhancement using α-CHCA matrix), TiO_2 powdered marks yielded the best results following re-enhancement of post-primary-lift marks with further TiO_2. They hypothesised that this was because TiO_2 nanoparticles can be used as a MALDI matrix, and that additional powdering of the marks, prior to a secondary lift, allowed for additional particle adherence to the fingermark substrate, enabling the recovery of higher quantities of endogenous material. This work proposes the viability of successful MALDI analysis of secondary lifts when primary lifts are unavailable. This is an important finding, contributing to the operational viability of this technique for fingermark analysis, as fingermark deposition surfaces are rarely suitable to be inserted into a sample chamber, being either too large, impractical or relating to surfaces that are uneven or curved (**Bradshaw *et al.,* 2016**).

Francese reviewed these recent developments and publications in a review chapter in 2016, covering sample preparation, matrix deposition, and detection of medium molecular weight species such as peptides and proteins and small molecular weight molecules such as condom lubricants and illicit drugs (**Francese, 2016**).

In 2016, the group also started exploring a new avenue of fingermark research. Deininger *et al.* report for the first time the mapping of blood signals within the ridge detail of bloody fingermarks using MALDI imaging. They used a bottom-up proteomic approach following an *in situ* digestion protocol to identify peptides originating from blood proteins such as haemoglobin, hemopexin and

serotransferrin. This built on previous research by the group into the detection and identification of blood derived proteins by MALDI MSP (**Patel *et al.,* 2016**).

CSIs have presumptive colorimetric tests at their disposal to use at crime scenes to give an indication of blood presence, and there are a host of laboratory tests that can confirm the presence of blood, but all of these traditional techniques are destructive in nature, typically requiring swabbing and extraction of analytes for analysis. Mass spectrometric imaging offers a method to retain the 2D spatial information of the mark, while at the same time allowing detailed chemical analysis to take place, providing the potential to link a suspect through biometric identification to the events of the bloodshed. The detection and identification of blood derived peptides is a much more specific technique for the confirmation of the presence of blood than the colorimetric tests available, especially when MS/MS is carried out. Here, the authors performed MS/MS analysis on the two most prevalent α and β chain haemoglobin peptides at *m/z* 1529.7 and 1274.7 respectively, to sequence these peptides amino acid sequences and confirm the identity of the signals. This work presents the first protocol for digesting and imaging blood contaminated fingermarks with MALDI, using trypsin and a 3 hour incubation and was instrumental in informing the current project (**Deininger *et al,* 2016**).

Expanding on their use of real crime scene marks from their research in 2014, Bradshaw *et al.* report further MALDI imaging methods for four real crime scene fingermarks, sourced from collaborators at West Yorkshire Police, from 'high profile' crime scenes. One mark was analysed directed on the deposition surface; on a drug packet that had subsequently been treated with cyanoacrylate fuming and BY40. Two separate marks had been found on light fittings in a cannabis farm and one enhanced with TiO2 and one with Al powder.

A fourth mark was recovered from a window frame following enhancement with carbon black powder; these three marks were available to researchers at primary lifts, and were the first real crime scene lifts the group had analysed, which was made possible by the validation work the group had completed in 2016 (**Bradshaw *et al.,* 2016**), which indicated the best methods to be employed when concerning marks that had been previously enhanced with different FET powders. The researchers ensured compliance with CSI requirements, so that any findings would be admissible in court (**Bradshaw *et al., 2017***).

The subsequent lift from the mark on the light fitting gave a very poor transfer of material and neither ridge detail nor molecular information could be retrieved. The authors hypothesised that this may be because the mark had been exposed to high temperatures of the light fitting, which would lead to evaporation of the volatile content. The second lift from a plug socket switch at the same crime scene also yielded poor ridge detail, wherein the CSI concluded it was insufficient for matching purposes. Whilst no further ridge detail was obtained, molecular information was revealed; among the usual endogenous compounds were signals at *m/z* 304.2 and 332.2, which were suspected to relate to *n*-alkyl dimethylbenzylammonium (DMA), an antibacterial agent. The parent ion at 304.2 was selected for MS/MS analysis and its structure confirmed through the ion fragment at *m/z* 212.3. However, ion fragment signals at *m/z* 182.2 also revealed the presence of cocaine, which coincidentally also has a monoisotopic mass of *m/z* 304.2. Cocaine is a forensically relevant substance, and MALDI analysis was able to offer chemical information not previously at the disposal of investigations. This was detected in marks where lifts recovered such small quantities of material that visual database matching was deemed

futile, demonstrating the sensitivity capabilities of the technique on fingermarks obtained through casework rather than in laboratory conditions, and highlighting its potential to be utilised following visualisation but unsuccessful suspect identification through biometric information. As the third mark was discovered on a drugs packet, the deposition surface could be inserted directly into the sample chamber. This was following cyanoacrylate fuming and BY40 visualisation, with the total ion current (TIC) selection providing comparable ridge detail to the visual enhancement, even though BY40 had been shown previously by the group to supress ionisation (**Groenveld *et al.,* 2015**). The fourth mark again contained a signal at *m/z* 304.2, this time relating to cocaine, as revealed through MS/MS analysis and the detection of the fragment ion signal at *m/z* 182.2. The detection of cocaine's parent ion is, in itself, not necessarily an indication of cocaine use, as it does not distinguish between consumption and handling scenarios. However, the additional detection of several cocaine metabolites by MALDI IMS within the mark (benzoylecgonine (BZE), ecgonine methyl ester (EME) and norcocaine (NCOC)) did indicate cocaine consumption. This was subsequently confirmed by submitting the suspect to a drugs test, in which cocaine was detected in their system. In addition to the cocaine metabolites, cocaethylene (CE) was detected in the mark, a metabolite that only forms when cocaine is consumed at the same time as alcohol. Whilst having drugs in one's system is not illegal in the UK, it can give an insight into a suspects altered state of mind at the time when they deposited a mark. Unbeknown to the researchers, the suspect had firmly denied alcohol consumption, but later recanted this when presented with the evidence (**Bradshaw *et al.,* 2017**).

This study represents an important advancement of MALDI as a forensic chemical profiling technique for fingermark analysis, transitioning from academia into "real world" investigations, using marks that had been recovered from the 'field' in less than ideal laboratory conditions, and having had FETs applied. Molecules detected through MALDI MSP, MSI and IMS were useful in further informing investigations, above and beyond the information that is available from conventional processing of fingermark evidence. The authors acknowledge that this is study is a proof of concept, and a large study would need to be undertaken, taking into account surface of deposition, time since deposition, environmental exposure (such as the possibility of high temperature experienced by the mark on the light fitting) and previous FETs employed for MALDI analysis to be feasibly integrated into an operational workflow (**Bradshaw *et al.,* 2017**).

In 2017, Francese *et al.* listed all of the analytical techniques capable of chemical profiling of fingermarks; FTIR, Raman, GC/MS, DESI, DART, SIMS, SALDI and MALDI and briefly reviewed several techniques under the "other" umbrella term. All of the reported "other techniques" had not reported significant operational capabilities and few covered the compatibility of the techniques with previously applied FETs, which must be considered when evaluating analytical techniques, as marks are likely to have been exposed to visualisation reagents before being made available for chemical analysis. They reviewed their own research which included the use of powders, powder suspensions, fuming, staining and VMD on a range of porous, semi-porous and non-porous surfaces. In some instances, such as the VMD-MALDI MSI workflow, MALDI MSI was able to reveal additional ridge detail not previously visible following visualisation alone. The authors go on to discuss the different methods of matrix application

and their success with various FETs. They evaluated the dry-wet method and the use of the method with both α-CHCA and curcumin, with the latter being discontinued due to the difficultly of cleaning deposition surfaces after visualisation (**Francese *et al.,* 2017**).

The review then goes on to cover the classes of molecules that have been investigated within fingermarks, including lipids, pharmaceutical and illicit drugs, explosives, polymers and endogenous species such as lipids, peptides and proteins. Of the endogenous substances, the group noted that their research focusing on lipids has centred on reconstructing ridge detail, but since 2009 noted the possibility of using lipids for fingermark dating purposes. Other groups had studies lipid compositional changes over time using techniques such as GC/MS and FTIR, but are not currently operationally viable due to donor variability. However, the authors suggested using lipids and their degradation products and relative intensities to predict the aging of marks. They also cover the ability of the peptide and protein composition of marks to distinguish between male and female donors, with some limitations. The study employed restriction criteria on the donors and used ungroomed marks which are lacking the presence of external contaminants, for simplicity. In addition, these marks were stored for a period of time before analysis, increasing the potential for degradation of the analytes, and were also not enhanced marks. Due to the inadvertent detection of certain proteins linked with tumour onset in breast cancer, the authors discuss the possibility of using fingermark analysis as a non-invasive alternative to breast cancer screening (**Francese *et al.,* 2017**).

In terms of exogenous substances, the authors discuss the success and challenges of the analysis of condom lubricants. Whilst distinguishing between several condom brands has been successful, showing a lubricant-contaminated

mark was left as a result of handling a condom and not a mark left in lubricant already present on a surface remained a challenge. The authors proposed future work on investigating whether lubricant signals were present in both the ridges and the valleys of the mark to differentiate between these two scenarios. They also acknowledged that a much larger database of the molecular signatures of virtually every condom brand and type would be necessary for this to be a useful investigative tool. The authors also reviewed recent work on the detection and visualisation of blood on the ridge pattern, utilising haem and intact haemoglobin. They were able to distinguish between human, equine and bovine blood through differences in the primary sequences of the haemoglobin chain, and even detect haem and haemoglobin at 1000 and 250,000 times lower concentrations that physiologically present. Although this analytical approach is more specific and less likely to result in false positives than the presumptive tests available to CSIs, the authors noted that a bottom-up proteomic approach would not only provide greater certainty through improved mass accuracy of smaller species than intact proteins, but could also generate many blood-specific proteins, increasing the certainty in the detection of blood, and elevating the forensic value. Finally, the authors discussed illicit drugs in both an exogenous and semi-exogenous context. In the former case, where the fingertips had been artificially contamination with known concentrations, molecular images relating to each drug could be obtained for cocaine, heroin, morphine, 3,4-methylenedioxyamphetamine (MDA) and THC. Whilst the MALDI imaging analysis was less successful for the drug consumption scenario, the presence of the drugs and their metabolites could still be detected via MALDI MSP and MS/MS. However, upon evaluation of patient marks from a drug rehabilitation clinic, where the study was expecting to find the prescribed

methadone, MALDI and DESI were also able to detect the presence of cocaine and BZE (with DESI additionally detecting EME) while SIMS was not able to provide any meaningful data. DESI has the advantage of being portable, as well as no necessary sample preparation, whereas it required analysis parameters to be optimised each time. MALDI requires matrix application and is not portable (owing in large part to the vacuum necessary for its function) but the parameters are more robust and it offers MS/MS and IMS features (**Francese et al., 2017**).

Following the release of the UK's new polymer bank notes (£5 in 2016, £10 in 2017), Bradshaw identified that some of the new security features employed had been found to effect the application of FETs. While the notes were made from non-porous polypropylene, certain special inks and security features also exhibited semi-porous features, which can make selecting an appropriate FET challenging. Traditionally, DFO and ninhydrin were used on the previous paper bank notes, but as these visualisation agents are suited for porous surfaces, they are less effective on the new notes. Research had shown, from places that have had polymer banknotes for some time, such as Australia, that VMD and CAF are much more effective for visualisation of marks, depending on the location and material of the part of the note where they are found. However, variation in different countries currencies makes it difficult to recommend one FET over another even, when they are based on the same polymer, because of different transparent windows and coatings. The authors trialled the use of MALDI imaging of marks deposited on various areas of the polymer notes, to develop a protocol for the sample preparation and analysis of this challenging surface. The authors were able to generate clear molecular images of ridge detail while reducing (and totally eliminating in some cases) background

interference from the multi-coloured surface and features, and in combination with FETs such as CAF and powder suspension. The study also investigated the imaging of cocaine in artificially spiked marks on polymer notes and refined a method for its identification through the presence of the precursor ion at *m/z* 304.2 and a product ion at *m/z* 182.1, in concentration s as low as 10 ng/mL. Due to research showing background cocaine contamination on a large proportion of bank notes, imaging, rather than profiling, is particularly useful in this instance, as the presence of cocaine exclusively on the ridges of fingermarks is an indication of it originating from contact with the individual, rather than background contamination. However, a limitation of this study was that both the notes and marks were cleaned (the use of a washing protocol and groomed marks) prior to analysis, potentially cleaning the notes of any background levels of illicit drugs, so it cannot be concluded that MALDI imaging would not detect some level of background contamination, and not only find cocaine present ion the marks. In addition, only 2 donors were utilised in this study, with one donor depositing the majority of marks, so this technique's effectiveness would need to be evaluated over a larger sample population before being proposed for operational casework. In addition, only 2 FET-MALDI workflows were investigated, so compatibility with all FETs would need to be evaluated for consideration in the Fingermark Visualisation Manual (**Scotcher & Bradshaw, 2018**).

Most recently, the group evaluated the suitability of MALDI imaging in the sequential workflow of latent fingermarks following conventional enhancement techniques, for marks on a range of tapes. In fingermark enhancement manuals and handbooks, for sequential processing, an order of application is recommended from least to most destructive. This is to ensure maximum

visualisation, while preserving as much ridge detail for further enhancement. It has been demonstrated through the previous work by the group that MALDI MS is compatible with singular FET processes, and in four previous instances with sequential FET processing, in some cases yielding further ridge detail than revealed by visualisation techniques alone (**Bradshaw *et al.,* 2013b**). However, the evaluation of MS techniques following sequential visualisation workflows is incomplete, and this technical note covers MALDI MSI results following up to four sequential visualisation processes on two surfaces. Marks were deposited on the adhesive side of clear tape and brown parcel tape, a challenging substrate to enhance due to its adhesive properties. For this reason, cyanoacrylate fuming (CAF) was the primary process employed in each case, followed by powders, powder suspensions and VMD in a range of combinations. Here the authors demonstrated that not only was MALDI able to generate ion maps from the marks following up to four sequential visualisation processes, MALDI MSI was also able to "provide additional ride detail in areas that are undeveloped or overdeveloped following the application of FET". This was due to the amalgamation of molecular images of (potentially) hundreds of molecules present in natural marks. As shown previously, lipid species (diacylglycerols, triacylglycerols and fatty acids) generally provided the best clarity images. In addition, better contrast was observed in the MSI images due to being able to totally subtract the background, which is particularly useful in cases where the deposition substrate is multi-coloured. Ion suppression effects were observed with the use of BY40, supporting previous findings by the group (**Groenveld *et al.,* 2015**). The ability of MALDI MSI was best demonstrated when powder suspension (PS) was used as the third sequential FET. CAF > BY40 > PS followed by MALDI yielded ridge detail not visible following the FETs

alone. The molecular images were also successful following the dual sequential use of CAF > BY40, and the quadruple sequential application of those FETs in addition to Basic Violet 3 (BV3), in the sequence CAF > BY40 > PS > BV3 (**Fig. 14**). This is due to the ability of MALDI imaging to detect species not targeted by the enhancement reagents.
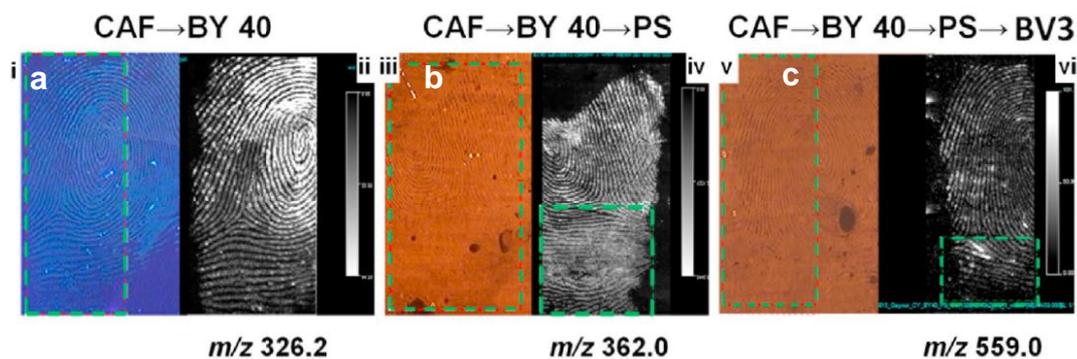
Figure 14. optical and MALDI MSI images of marks following the application of cyanoacrylate fuming and Basic Yellow 40 (a), with the addition of the sequential application of powder suspension (b) and Basic Violet 3 (c). *[Adapted from Bradshaw et al. 2020].*

In addition, because MALDI is able to simultaneously detect thousands of molecules in an imaging acquisition, individual signals occurring in different area of the mark can be overlaid to combine their molecular map detail, further enhancing the ridge detail visualisation, as show here in **Figure 15**.
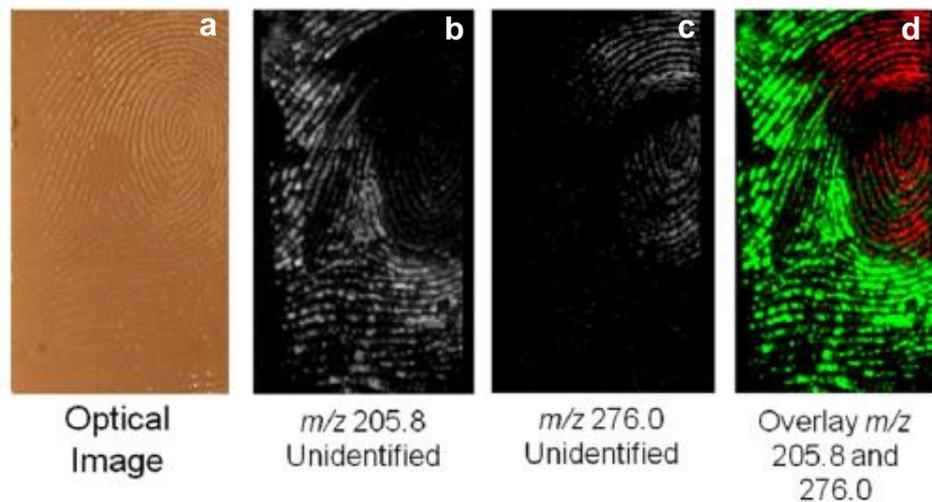
Figure 15. Optical image (a) molecular image showing unidentified species at *m/z* 205.8 (b), *m/z* 276.0 (c) and overlaid and alternately coloured (d) detected in a mark sequentially enhanced with CAF > BY40. *[Adapted from Bradshaw et al. 2020].*

This study demonstrated successful MALDI MS imaging of latent marks with up to four sequential FET processes applied. In contrast to similar research in the area with different MS techniques, this study used natural marks, to better replicate crime scene marks, as opposed to groomed and ungroomed marks used elsewhere. This research shows promising results for the two substrates investigated, but further work is needed to cover a wider range of deposition surfaces, and all possible combinations of FETs routinely applied to latent marks (**Bradshaw *et al.,* 2021**).

Prior Fingermark Research: MALDI Profiling and Imaging (Other Groups)

Simultaneously, while MALDI fingermark research was being undertaken by the Fingermark Research Group in Francese's lab, other research groups were utilising MALDI to obtain chemical information from fingermarks.

In 2009, Benton's group were investigating if they could detect the semi-exogenous substances nicotine and cotinine in smokers fingermarks. They compared employing MALDI (using 2,5-dihydroxybenzoic acid (DHB) matrix)

and SALDI (using hydrophobic silica particles doped with carbon black) to analyse groomed fingermarks of smokers on metal surfaces and lifts from glass slides. The marks were the prepared either as they were, dusted with DHB matrix, or dusted with a previously developed magnetic hydrophobic black powder, comprising of carbon black, phenyltriethoxysilane (PTEOS), and iron powder. The samples then underwent laser desorption ionisation (LDI) analysis (MALDI for DHB dusted marks, SALDI for hydrophobic black powdered marks) to detect the parent ions for both nicotine and cotinine. Signals were identified at nominal *m/z* 163 and 177 respectively. The presence of these molecules was confirmed by MS/MS of these parent ions undergoing collision induced dissociation (CID) fragmentation to elucidate their structure. They noted that dusting with the hydrophobic black powder resulted in spectra with signals at least as intense as those obtained with conventional MALDI matrix DHB, however the black powder had the added advantage of acting as an effective visualisation agent. Images of these ion signals were also obtained, and although some ridge detail is visible in the ion maps, optical images are not presented for a comparison to the original marks. In addition, control marks are not presented and there are no comments towards these signals not appearing in non-smokers marks following imaging experiments, despite the authors acknowledging background levels of (at least) signals at *m/z* 163 in profiling experiments. The clarity of the images is also reasonable but not excellent for a 100 µm^2 pixel size, and would be unlikely to result in an identity match based on these ion map images alone. The authors reported that although there did not appear to be a correlation between the number of cigarettes consumed and the peak intensities, they concluded that the detection of nicotine and cotinine in fingermarks via LDI methods was a good potential marker for indicating that an

individual was a smoker, although a limit of detection study was not attempted to discover the minimum number of cigarettes that would need to be consumed on a daily basis for these compounds to be detectable (**Benton *et al.,* 2010**).

Yagnik *et al.* proposed a method for the simultaneous MALDI imaging of exogenous and endogenous substances and MALDI MS/MS imaging of targeted compounds in fingermarks at the same time. They used multiplex MALDI mass spectrometry imaging, which simultaneously offers the high spatial resolution of the linear ion trap and the high mass resolution of the orbitrap to acquire MS, MS/MS and MSI data in one acquisition. The parallel nature of this technique offers 'significant time savings' compared to only using orbitrap data. In addition, acquiring MS/MS data in parallel to the orbitrap scans can distinguish between structural isomers, a feature that would normally require complex ion mobility spectrometry (IMS). Finally the dual methods of acquisition can be obtained in a single imaging run, which is useful when the sample is scarce, requiring further processing in a forensic workflow (for example further enhancement) or when the sample must be kept for posterity as evidence. They applied this technique to latent fingermarks, by setting their raster pixel sizes to 100 µm^2, containing four 50 µm^2 'spiral steps', one of which was used for the orbitrap scan and the remaining three being used for MS/MS ion trap scans (**Yagnik *et al.,* 2013**).

Miller's group in the U.S. were using MALDI MS to investigate "touch chemistry"; analysis of exogenous contaminants present in marks resulting from contact with the fingertips. In 2014 they published on the detection of pharmaceuticals, drugs and explosives in groomed and natural marks. Specifically, they analysed the drugs ibuprofen, aspirin and acetaminophen from tablets, procaine and pseudoephedrine (to mimic cocaine and

methamphetamine), and the explosive powders trinitrotoluene (TNT) and cyclotrimethylenetrinitramine, known colloquially for security reasons as Research Department Explosive (RDX). Groomed marks were contaminated by handling powders of the four analytes, and by the additional scenarios of handling whole or broken aspirin, ibuprofen and acetaminophen tablets, to more closely replicate the process of ingestion. Natural marks were contaminated by rubbing the fingertips together, then rubbing them across glass slides upon which solutions of varying concentrations had been allowed to evaporate on, before being deposited on aluminium slides. In addition to the detection of these compounds, the group wanted to test the compatibility of the technique with common FETs, so were developed with either only black powder, CAF and black powder, or black powder followed by hinge lifting tape. Excess powder was removed by tapping the developed slides on a hard surface, although the authors noted that this resulted in contamination of the MALDI vacuum chamber, and that a more effective method of removing excess powder was needed. They concluded that, at the time, whilst the 'handling of tablets' scenario did not result in sufficient analyte transfer to be detectable via MALDI MSP/MSI, the handling of these compounds in powdered form and in the form of residue left behind after evaporating the solvent from glass slides was sufficient for detection through MALDI MS profiling and imaging, which is useful to demonstrate that the compounds reside exclusively in the ridges and not from background contamination already present on the deposition surface. However, the authors noted that "analyte powders were easily spread throughout the fingermark" when applying development powders with a brush. Disappointingly, from the incomplete set of images provided, the target molecules appear to be present in some of the blank controls, and as no

MS/MS elucidation was performed on any of the analytes, identification of these molecules could be disputed. Finally, although they also state that MALDI is a "non-destructive method" and "can be used without impeding the latent print process", the ablation process does remove a miniscule quantity of substance from the surface of the mark, and the application of the MALDI matrix is an irreversible addition to the sample, which has not yet been shown to have no effect on subsequent processing (**Kaplan-Sandquist *et al.,* 2014**).

At the same time, Rowell's group were investigating the detection of drugs in lifted enhanced marks using both MALDI MS and SALDI MS approaches. They attempted the detection of five drugs; the pharmaceuticals aspirin, paracetamol and caffeine, and the drugs of abuse cocaine and methadone in ungroomed marks. The fingers were externally contaminated with 1 mg of each analyte before depositing groomed marks on clean glass slides or directly onto the MALDI target plate. Half of the deposited marks underwent CAF before being dusted with either DHB matrix or magnetic black powder, while the other half just underwent powdering with no CAF. The marks were then lifted from the glass slides with lifting tape. Unprepared lifts proved unsuccessful, so the group experimented with exposing the marks to solvent vapour prior to lifting which improved the transfer. Of the marks deposited directly on the MALDI target plate, those that were only dusted with magnetic black powder produced signals for each of the drugs, in contrast to those that were dusted with DHB matrix, which produced no drug signals. However, the lifted marks that had been exposed to acetone vapour produced peaks that the authors attributed to the drugs. They claimed that they had detected aspirin adducts ([M + Na]$^+$ monoisotopic *m/z* 203.04, experimental *m/z* 203.186, error 719 ppm) ([M + K]$^+$ monoisotopic *m/z* 219.04, experimental *m/z* 219.163, error 562 ppm)

paracetamol ([M+H]$^+$ monoisotopic *m/z* 152.06, experimental *m/z* 152.158, error

644 ppm) ([M + Na]$^+$ monoisotopic *m/z* 174.06, experimental *m/z* 174.185, error

718 ppm) ([M + K]$^+$ monoisotopic *m/z* 190.06, experimental *m/z* 190.173, error

595 ppm) (fragment, monoisotopic *m/z* 109.06, experimental *m/z* 109.105, error

413 ppm), and caffeine ([M+H]$^+$ monoisotopic *m/z* 195.08, experimental *m/z*

195.186, error 543 ppm), although for the protonated caffeine parent ion signal

the intensity is less than convincing. In addition, the mass errors between the

monoisotopic mass-to-charge ratios and the experimental values are very large

(413 - 719) and bring into question the veracity of the identifications (**Table 6**).

	[M+H]$^+$		[M + Na]$^+$		[M + K]$^+$		Fragments	
Drug	mono	exp.	mono	exp.	mono	exp.	mono	exp.
Aspirin	/	/	203.04	203.186	219.04	219.163	137.04	137.022
Paracetamol	152.06	152.158	174.06	174.185	190.06	190.173	109.06	109.105
Caffeine	195.08	195.186	/	/	/	/	/	/

Table 6. The monoisotopic mass-to-charge ratio and the experimental value of the protonated parent ion, sodiated and potassiated adducts and fragments of the three pharmaceutical drugs.

Interestingly, the results showed that magnetic black powder produced spectra

an order of magnitude more intense than DHB matrix, favouring the SALDI

rather than MALDI approach to the analysis, which is surprising as DHB is

usually effective at assisting the ionisation of low molecular weight (LMW)

molecules. Also surprising are the results that showed that the fingermark lifts

(from the target plate and the glass slide) produced considerably more intense

spectra than the marks analysed directly on a MALDI target plate, despite the

authors claiming "peak intensities were lower in lifted marks" (**Sundar &**

Rowell, 2014).

When examining the marks contaminated with illicit drugs the lifted mark spectra looked almost identical to the blank mark spectra (Fig. 7 in their publication). The authors attributed the signals at *m/z* 304.07 to cocaine and *m/z* 310.21 to methadone, despite both these signals appearing to be present in the blank sample, and despite the fact that a known fingermark contaminant benzalkonium chloride (BZK) also has a nominal *m/z* of 304.1. The authors did not perform MS/MS analysis to confirm the detection of the species they had preliminarily identified. However, using MALDI/SALDI imaging, they were able to detect the presence of signals at *m/z* 304.16 and 310.21, across the whole of the marks, again attributing these ions to cocaine and methadone (**Sundar & Rowell, 2014**).

Prior Blood Analysis

Prior Blood Research Using MALDI MS at Sheffield Hallam University

There has been a host of research in recent years devoted to the chemical analysis of blood and bloody fingermarks in a forensic context. Non-destructive methods are favoured when evidence must be retained for further analysis or posterity. Mass spectrometry techniques can be minimally destructive, with the MALDI laser ablating just a few micrometres from the surface of a sample, leaving the spatial arrangements of the ridge intact. This is in contrast to the 'gold standard' of forensic chemical analysis, LC-MS, which requires removal or destruction of at least part of a sample and often an extraction step, leaving a partial sample for any further enhancement or analysis.

Bradshaw *et al.* used MALDI MSP and MSI to specifically detect haem and intact haemoglobin in fresh and aged blood mark samples in 2013 (**Bradshaw et al., 2014**). Patel *et al.* built on this research to develop a shotgun 'bottom-up'

proteomic approach using MALDI MSP for the detection of blood with both an *in-solution* and also an *in-situ* digestion method for the analysis of previously enhanced blood marks, demonstrating compatibility with CSI visualisation techniques. This technique was able to distinguish between human, bovine and equine blood, based on differing haemoglobin peptide chains (**Patel *et al.*, 2016**). Deininger *et al.* refined this work to enable the MALDI MSI mapping of blood peptides on the ridges of bloody fingermarks, generating chemical information whilst retaining the complete ridge pattern (**Deininger *et al.*, 2016**). Whilst these studies offered a proof-of-concept into the reliable detection of blood and discrimination of the blood of three species, validation data is required to demonstrate robustness and reliability.

Proteomics

Proteomics is the study of the proteome, which comprises of the proteins contained within a biological system. The state of the proteome is not constant and may react to external influences. Proteins are macromolecules which can have large molecular weights. They are made up of extended chains of amino acids, which act as the building blocks for all proteins. There are around 500 known amino acids, twenty of which occur naturally in the human body, which play an essential biochemical role.
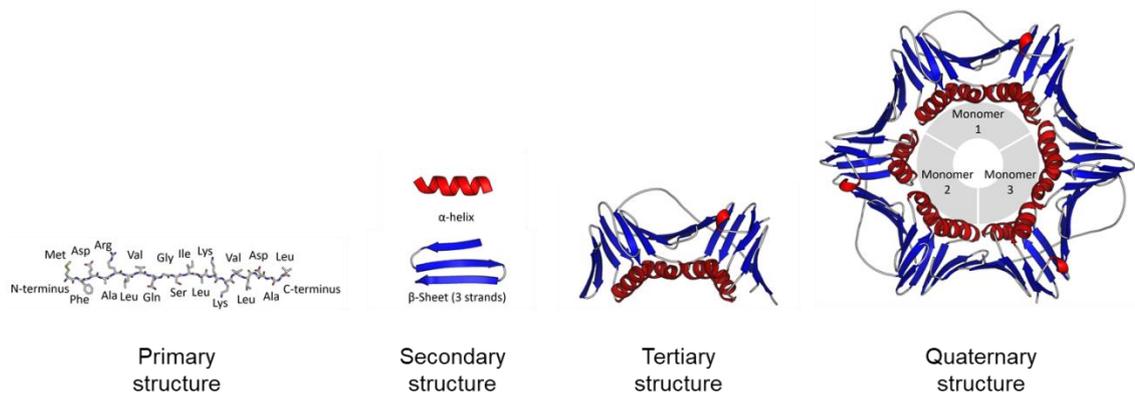
Figure 16. Diagrams of proteins primary, secondary, tertiary and quaternary structure. *[Adapted from: https://www.rcsb.org/structure/1AXC]*

The state of the proteome can be investigated using mass spectrometry. Current mass spectrometric instrumentation can detect and characterize individual proteins in complex mixtures with high levels of specificity, sensitivity and mass accuracy. However, usually some degree of separation precedes the analysis. One- or two-dimensional separation can be performed offline, or, more recently, individual peptide separation can be performed online, directly before ionization for mass spectrometric analysis. Typically this would be MALDI or ESI.

Protein detection can fall into two categories; analysis of the whole protein, which is termed top-down, and analysis of shorter segments of the proteins, or peptides, termed bottom-up. Whole proteins can be tens of thousands of amino acids in length, resulting in very high molecular weight molecules. MALDI MS instruments can struggle with high mass accuracy for molecules over 30 kDa which means accurately identifying large proteins becomes difficult. For this reason, bottom-up proteomics is commonly used for mass spectrometric analysis of the proteome.

Top-down Proteomics

Top-down proteomics is a method of MS/MS analysis that introduces intact protein ions to the mass analyser, typically by ESI, following a pre-purification step. This can involve using an ion trap mass spectrometer to isolate the required protein, or it can utilise a chromatographic step such as two-dimensional gel electrophoresis. Fragmentation of the intact protein ions occurs in the gas phase and utilises electron-capture dissociation or electron-transfer dissociation. The whole protein structure is then inferred from the partial protein fragments. Interpretation of top-down MS/MS spectra can be complex due to the generation of multiply charged product ions.

Bottom-up Proteomics

Bottom-up proteomics is the more common of the two methods. The bottom-up strategy can be further stratified into two workflows; separation followed by digestion or digestion followed by separation. The prior method first employs a separation stage, typically chromatography or gel electrophoresis to isolate proteins of interest from a complex sample. This fraction would then be analysed by peptide mass fingerprinting (PMF) against a peptide database or via MS/MS fragmentation, with the resulting spectrum being matched against a MS/MS database. For the latter method, the whole sample undergoes a proteolytic digestion followed by mass spectrometric analysis. This method is also known as 'shotgun' proteomics, due to its untargeted nature. This technique requires less sample preparation, but the resulting abundance of peptides from all existing proteins can make for a complicated spectrum to interpret. Sensitive instrumentation is required to detect low abundance peptides, or ones which ionise less readily.

In-source Decay

In-source Decay (ISD) is an approach that can be used with MALDI, in order to record a mass spectrum containing many protein fragments. With in-source decay, fragmentation occurs immediately after the laser shot and before ion extraction (**Demeure *et al.,* 2010**). However, this is technically not MS/MS as it does not contain an initial ion isolation step. Importantly, for in-source decay it is crucial to purify the proteins before the analysis so that the fragments contained in the mass spectrum all originate from the same protein.

Aims and Objectives

Fingermarks are the oldest and most widely used trace evidence type for identification purposes; they still account for more criminal identifications worldwide than DNA (**Francese, 2019**). It is common knowledge that their unique pattern constitutes biometric information that can be either matched against a database or against potential suspects. However, in addition to physical evidence, they can provide chemical evidence through the analysis of their molecular composition. Over the last 13 years, the Fingermark Research Group at Sheffield Hallam University has explored the detection of exogenous, endogenous and semi-endogenous substances within fingermark residue, employing Matrix Assisted Laser Desorption Ionisation Mass Spectrometry. A aim of this PhD programme is to build on previous pioneering research by the group concerning the determination of sex from the analysis of peptides and proteins within fingermarks. The previous work, published in 2012 (**Ferguson *et al.,* 2012**) reported 68.9% and 74.4% prediction accuracy for female and male marks respectively (if each spectrum was processed individually). This study

was limited in terms of sample size (80 participants), had exclusion criteria (participants must be between 18-45 years old, not have used drugs or medication in the 2 weeks prior), and collected ungroomed marks (i.e. the hands had been washed before deposition to remove external contaminants). More modern instrumentation has also surpassed the technical specifications of those employed previously.

The present work aims to investigate a method for the determination of sex from fingermarks, minimising some of the constraints previously reported in the literature. The present study will increase the sample size and employ a new data processing approach to increase the statistical validity of the method. Exclusion criteria will also be removed to ensure a more diverse sample population, and natural marks will be collected, where the hands have not been washed prior to deposition, to assess the robustness of the method without the artificial removal of potential contaminants. Compatibility of the method with a subset of fingermark enhancement techniques will also be assessed, as latent deposits are routinely visualised at the crime scene. The method development is designed specifically to advance the applicability of the results obtained towards the proposal for the potential use of this MALDI-based approach in an operation context for crime scene marks.

Blood is another biological matrix often encountered at the scenes of violent crimes. However, the definitive identification of bloodstains remains a challenge. Presumptive colorimetric tests are used for the initial screening of suspected blood deposits, but their application can dilute and contaminate evidence, may degrade latent marks and DNA (depending on their formulation) and are susceptible to false positives. Substances indicated by a presumptive test also require subsequent confirmatory analysis, which can be laborious and time

consuming and are usually destructive, consuming valuable (and potentially limited) material which may be required for further analysis, such as DNA profiling for example. An aim of this PhD programme is the development of a MALDI-based method for the confirmatory detection of blood using a proteomic approach. Proteins are more stable and persistent than traces of DNA, particularly in harsher environmental conditions. A 'bottom-up' proteomic approach will be evaluated, not only for the definitive detection of blood, but also for distinguishing between human blood and animal blood (from a small subset of domestic animals), through the detection and identification of blood-specific peptides and proteins. Furthermore, to evaluate the specificity of the method, a blind validation study will be undertaken comprising other human biofluids, (semen, saliva) and other visually similar non-biofluid substances, particularly those which can result in false-positives with common blood enhancement reagents. This technique has the potential to offer more reliable identification than presumptive tests, in a faster and less destructive manner than conventional confirmatory approaches.

A final aim of this PhD programme is to investigate whether a MALDI MS approach can yield pathological information from human blood in a forensic context. The analysis of semi-endogenous (ingested) substances is well established in forensic science; the application of toxicological screening is well known for the detection of alcohol and drugs, but it remains largely an untapped resource of endogenous physiological information (despite established serology screening in the healthcare sector). It is hypothesised that pathological information could be used to help rule in/out a victim/suspect as the source of blood encountered at a crime scene. For example, as haemoglobin variants are rare in the UK, the discovery of blood containing a specific variant may be

strong associative evidence if it was known to match a victim or suspect. This would prove particularly useful in scenarios where DNA profiling may not yield a match (trace quantities may not be sufficient for amplification and sparse or degraded samples may yield a partial profile or mixed profile. Even if a complete profile is generated, a match relies on a corresponding sample being present in the database).

The final aim of this PhD programme is to evaluate whether MALDI MS and MSI can be employed successfully for the detection of haemoglobin variants within human blood. The application of 'bottom-up' proteomics will be employed both *in solution* (in extracts) and *in situ*, when applied directly to exhibits to preserve the 2D confirmation of a sample, a critical benefit when analysing certain evidence types such as blood-contaminated fingermarks. MALDI tandem mass spectrometry (MS/MS) will be applied for the confirmation of variant identification. It is the first time in the author's knowledge that human haemoglobin variant analysis has been considered in a forensic context.

In summary, the overall objective of this PhD programme is to evaluate novel methods of molecular analysis for two commonly encountered biological evidence types, utilising MALDI MS, MS/MS and MSI, in combination with a 'bottom-up' proteomic approach. This is with a view to demonstrating proof-of-concept and initial validation data for each of the aims discussed. The results of this research will be presented in the following chapters.

References

Ahn, S. H., Park, K. M., Bae, Y. J., & Kim, M. S. (2013). Quantitative reproducibility of mass spectra in matrix-assisted laser desorption ionization and unraveling of the mechanism for gas-phase peptide ion formation. *Journal of Mass Spectrometry.*, *48*(3), 299–305.

Bae, Y. J., Shin, Y. S., Moon, J. H., & Kim, M. S. (2012). Degree of Ionization in MALDI of Peptides: Thermal Explanation for the Gas-Phase Ion Formation. *Journal of the American Society for Mass Spectrometry*, *23*(8), 1326–1335.

Bandey, H. L., Bleay, S. M., Bowman, V. J., Downham, R. P., Sears, V. G., (2014). Fingermark Visualisation Manual, *The Home Office*, 1-932.

Beadoin, A., (2004), New technique for revealing latent fingerprints on wet, porous surfaces: Oil Red O, *Journal of Forensic Identification,* 54 (4) 413–421.

Benton, M., Rowell, F., Sundar, L., & Jan, M. (2010). Direct detection of nicotine and cotinine in dusted latent fingermarks of smokers by using hydrophobic silica particles and MS. *Surface and Interface Analysis*, *42*(5), 378–385.

Bilous, P., Fossum, M., Hallmark, C., (2011) Blood Enhancement Reagents, Luminol, Bluestar®, Fluorescein, and Hemascein™ : a quantitative comparison of properties essential for crime scene investigations. Presented at IAFS Meeting, Portugal.

Bleay, S. M., Sears V. G., Bandey, H. L., Gibson, A. P., Bowman V. J., Downham, R., Fitzgerald, L., Ciuksza, T., Ramadani, J. & Selway, C., (2012) Fingerprint Source Book v1.0 (first edition), *CAST Publication*, 1-502

Bleay, S., Sears, V., Downham, R., Bandey, H., Gibson, A., Bowman, V., Fitzgerald, L., Ciuksza, T., Ramadani, J., Selway, C., (2017). Fingerprint Source Book v2.0 (second edition), *CAST Publication 081/17*. 1-666.

Bossers, L. C. A. ., Roux, C., Bell, M., & McDonagh, A. M. (2011). Methods for the enhancement of fingermarks in blood. *Forensic Science International (Online)*, *210*(1), 1–11.

Bradshaw, R., Wolstenholme, R., Blackledge, R. D., Clench, M. R., Ferguson, L. S., & Francese, S. (2011). A novel matrix-assisted laser desorption/ionisation mass spectrometry imaging based methodology for the identification of sexual assault suspects. *Rapid Communications in Mass Spectrometry*, *25*(3), 415–422.

Bradshaw, R., Rao, W., Wolstenholme, R., Clench, M. ., Bleay, S., & Francese, S. (2012). Separation of overlapping fingermarks by Matrix Assisted Laser Desorption Ionisation Mass Spectrometry Imaging. *Forensic Science International*, *222*(1), 318–326.

Bradshaw, R., Bleay, S., Wolstenholme, R., Clench, M. R., & Francese, S. (2013a). Towards the integration of matrix assisted laser desorption ionisation mass spectrometry imaging into the current fingermark examination workflow. *Forensic Science International*, *232*(1), 111–124.

Bradshaw, R., Wolstenholme, R., Ferguson, L. S., Sammon, C., Mader, K., Claude, E., … Francese, S. (2013b). Spectroscopic imaging based approach for condom identification in condom contaminated fingermarks. *Analyst (London)*, *138*(9), 2546–2557.

Bradshaw, R., Bleay, S., Clench, M. ., & Francese, S. (2014). Direct detection of blood in fingermarks by MALDI MS profiling and Imaging. *Science & Justice*, *54*(2), 110–117.

Bradshaw, R., Francese, S., (2014). Matrix assisted laser desorption ionisation tandem mass spectrometry imaging of small molecules from latent fingermarks, *Spectroscopy Europe,* 26(4), 6-8. https://www.spectroscopyeurope.com/system/files/pdf/MS_26_4_0.pdf *[last accessed: 05/06/22].*

Bradshaw, R., Denison, N., & Francese, S. (2016). Development of operational protocols for the analysis of primary and secondary fingermark lifts by MALDI-MS imaging. *Analytical Methods*, *8*(37), 6795–6804.

Bradshaw, R., Denison, N., & Francese, S. (2017). Implementation of MALDI MS profiling and imaging methods for the analysis of real crime scene fingermarks. *Analyst (London), 142*(9), 1581–159.

Bradshaw, R., Wilson, G., Denison, N., & Francese, S. (2021). Application of MALDI MS imaging after sequential processing of latent fingermarks. *Forensic Science International*, *319*, 110643–110643.

Breuker, K., Knochenmuss, R., Zhang, J., Stortelder, A., & Zenobi, R. (2003). Thermodynamic control of final ion distributions in MALDI: in-plume proton transfer reactions. *International Journal of Mass Spectrometry*, *226*(1), 211–222.

Caprioli, R. M., Farmer, T. B., & Gile, J. (1997). Molecular Imaging of Biological Samples: Localization of Peptides and Proteins Using MALDI-TOF MS. *Analytical Chemistry (Washington)*, *69*(23), 4751–4760.

Chaleckis, R., Murakami, I., Takada, J., Kondoh, H., & Yanagida, M. (2016). Individual variability in human blood metabolites identifies age-related differences. *Proceedings of the National Academy of Sciences - PNAS*, *113*(16), 4252–4259.

Chen, X., Carroll, J. A., & Beavis, R. C. (1998). Near-ultraviolet-induced matrix-assisted laser desorption/ionization as a function of wavelength. *Journal of the American Society for Mass Spectrometry*, *9*(9), 885–891.

Chen, T., Schultz, Z. D., & Levin, I. W. (2009). Infrared spectroscopic imaging of latent fingerprints and associated forensic evidence. *Analyst (London), 134*(9), 1902–1904.

Cho, S., Jung, S.-E., Hong, S. R., Lee, E. H., Lee, J. H., Lee, S. D., & Lee, H. Y. (2017). Independent validation of DNA-based approaches for age prediction in blood. *Forensic Science International : Genetics*, *29*, 250–256.

Chu, K. Y., Lee, S., Tsai, M.-T., Lu, I.-C., Dyakov, Y. A., Lai, Y. H., … Ni, C.-K. (2014). Thermal Proton Transfer Reactions in Ultraviolet Matrix-Assisted Laser Desorption/Ionization. *Journal of the American Society for Mass Spectrometry*, *25*(3), 310–318.

Deininger, L., Patel, E., Clench, M. R., Sears, V., Sammon, C., & Francese, S. (2016). Proteomics goes forensic: Detection and mapping of blood signatures in fingermarks. *Proteomics (Weinheim), 16*(11-12), 1707–1717.

Demeure, K., Gabelica. V., & De Pauw, E. A. (2010). New Advances in the Understanding of the In-Source Decay Fragmentation of Peptides in MALDI-TOF-MS. *Journal of the American Society for Mass Spectrometry*, *21*(11), 1906–1917.

Dhakal, C., 2013, Comparison of different latent fingermark development techniques on wet paper surfaces, Journal article submitted in part fulfilment of MSc, Kings College, London, September

Dole, M., (1968), Molecular beams of macro-ions, *The Journal of Chemical Physics,* 49(5). 2240

Doty, K. C., & Lednev, I. K. (2018). Differentiating Donor Age Groups Based on Raman Spectroscopy of Bloodstains for Forensic Purposes. *ACS Central Science*, *4*(7), 862–867.

Downham, R. P., Sears, V. G., Hussey, L., Chu, B.-S., & Jones, B. J. (2018). Fingermark visualisation with iron oxide powder suspension: The variable effectiveness of iron (II/III) oxide powders, and Tween® 20 as an alternative to Triton™ X-100. *Forensic Science International*, *292*, 190–203.

Ferguson, L. S., Creasey, S., Wolstenholme, R., Clench, M. R., & Francese, S. (2013). Efficiency of the dry-wet method for the MALDI-MSI analysis of latent fingermarks. *Journal of Mass Spectrometry.*, *48*(6), 677–684.

Francese, S., Bradshaw, R., Ferguson, L. S., Wolstenholme, R., Clench, M. R., & Bleay, S. (2013). Beyond the ridge pattern: Multi-informative analysis of latent fingermarks by MALDI mass spectrometry. *Analyst (London)*, *138*(15), 4215–4228.

Francese, S. (2016). Techniques for fingermark analysis using MALDI MS: A practical overview. *In: Advances in MALDI and Laser-Induced Soft Ionization Mass Spectrometry* (pp. 93–128).

Francese, S., Bradshaw, R., & Denison, N. (2017). An update on MALDI mass spectrometry based technology for the analysis of fingermarks-stepping into operational deployment. *Analyst (London), 142*(14), 2518–2546.

Fraser, J., Deacon, P., Bleay, S., & Bremner, D. H. (2013). A comparison of the use of vacuum metal deposition versus cyanoacrylate fuming for visualisation of fingermarks and grab impressions on fabrics. *Science & Justice, 54*(2), 133–140.

Friedman, M., David W.L., (1974), Stoichiometry of formation of Ruhemann's purple in the ninhydrin reaction, *Bioorganic Chemistry*, Vol.3(3), pp.267-280

Galton. F., (1892). Finger Prints. *American Anthropologist, 6*(3), 341–342.

Gorka, M., Thomas, A. & Bécue, A., (2022) Chemical composition of the fingermark residue: Assessment of the intravariability over one year using MALDI-MSI, Forensic Science International, **338**, [pre-print].

Grigg, R., Mongkolaussavaratana, T., Pounds, C.A., Sivagnanam, S., (1990), 1,8-Diazafluoren-9-one and related compounds. A new reagent for the detection of alpha-amino acids and latent fingerprints, *Tetrahedron Lett*, 31 (49) 7215–7218.

Groeneveld, G., De Puit, M., Bleay, S., Bradshaw, R., & Francese, S. (2015). Detection and mapping of illicit drugs and their metabolites in fingermarks by MALDI MS and compatibility with forensic techniques. *Scientific Reports, 5*(1), 11716–11716.

Guigui, K., & Beaudoin, A. (2007). The use of Oil Red O in sequence with other methods of fingerprint development. *Journal of Forensic Identification, 57*(4), 550–581.

Gutiérrez, E., Galera, V., Martínez, J. M., & Alonso, C. (2007). Biological variability of the minutiae in the fingerprints of a sample of the Spanish population. *Forensic Science International*, *172*(2), 98–105.

Gutiérrez-Redomero, E., Alonso-Rodríguez, C., Hernández-Hurtado, L. E., & Rodríguez-Villalba, J. L. (2010). Distribution of the minutiae in the fingerprints of a sample of the Spanish population. *Forensic Science International*, *208*(1), 79–90.

Hewlett, D. F., Sears, V. G. and Suzuki, S. (1997) 'Replacements for CFC113 in the Ninhydrin Process: Part 2', *J. Forens. Ident.*, vol. 47 (3), pp 300–306.

Hoffmann, E, Stroobant, V., (2007), Mass Spectrometry: Principles and Applications by Edmond de Hoffmann, 3rd ed. (third edition), *John Wiley & Sons Inc*

Huang, Y., Yan, J., Hou, J., Fu, X., Li, L., & Hou, Y. (2015). Developing a DNA methylation assay for human age prediction in blood and bloodstain. *Forensic Science International : Genetics*, *17*, 129–136.

Ifa, D. R., Manicke, N. E., Dill, A. L., & Cooks, R. G. (2008). Latent Fingerprint Chemical Imaging by Mass Spectrometry. *Science (American Association for the Advancement of Science)*, *321*(5890), 805–805.

James, J., Pounds, C., & Wilshire, B. (1991). Obliteration of Latent Fingerprints. *Journal of Forensic Sciences*, *36*(5), 1376–1386.

Kaplan-Sandquist, K., LeBeau, M. A., & Miller, M. L. (2013). Chemical analysis of pharmaceuticals and explosives in fingermarks using matrix-assisted laser desorption ionization/time-of-flight mass spectrometry. *Forensic Science International*, *235*(C), 68–77.

Karas, M., Bachmann, D., & Hillenkamp, F. (1985). Influence of the wavelength in high-irradiance ultraviolet laser desorption mass spectrometry of organic molecules. *Analytical Chemistry (Washington)*, *57*(14), 2935–2939.

Karbach, V., & Knochenmuss, R. (1998). Do single matrix molecules generate primary ions in ultraviolet matrix-assisted laser desorption/ionization. *Rapid Communications in Mass Spectrometry*, *12*(14), 968–974.

Kastle, J. H., Shedd O. M., (1901) 'Phenolphthalin as a reagent for the oxidizing ferments', Am. Chem. J., vol. 26, p 526.

Knochenmuss, R., Stortelder, A., Breuker, K., & Zenobi, R. (2000). Secondary ion-molecule reactions in matrix-assisted laser desorption/ionization. *Journal of Mass Spectrometry.*, *35*(11), 1237–1245.

Knochenmuss, R., & Zenobi, R. (2003). MALDI Ionization: The Role of In-Plume Processes. *Chemical Reviews*, *103*(2), 441–452

Knochenmuss, R. (2003). A Quantitative Model of Ultraviolet Matrix-Assisted Laser Desorption/Ionization Including Analyte Ion Generation. *Analytical Chemistry (Washington)*, *75*(10), 2199–2207.

Knochenmuss, R. (2006). Ion formation mechanisms in UV-MALDI. *Analyst (London)*, *131*(9), 966–986.

Knochenmuss, R. (2009). A bipolar rate equation model of MALDI primary and secondary ionization processes, with application to positive/negative analyte ion ratios and suppression effects. *International Journal of Mass Spectrometry*, *285*(3), 105–113.

Knochenmuss, R. (2014). MALDI mechanisms: Wavelength and matrix dependence of the coupled photophysical and chemical dynamics model. *Analyst (London)*, *139*(1), 147–156.

Knochenmuss, R. (2016). The Coupled Chemical and Physical Dynamics Model of MALDI. *Annual Review of Analytical Chemistry (Palo Alto, Calif.)*, *9*(1), 365–385.

Knowles, A. M. (1978). Aspects of physicochemical methods for the detection of latent fingerprints. Journal of Physics. E, Scientific Instruments, 11(8), 713–721.

Lai, Y.-H., Wang, C.-C., Lin, S.-H., Lee, Y. T., & Wang, Y.-S. (2010). Solid-Phase Thermodynamic Interpretation of Ion Desorption in Matrix-Assisted Laser Desorption/Ionization. *The Journal of Physical Chemistry. B*, *114*(43), 13847–13852.

Lee, H. C., & Gaensslen, R. E., (2001). *Advances in fingerprint technology* (2nd ed.). Boca Raton, Fla. CRC Press.

Lennard, C., (2001). The detection and enhancement of latent fingerprints, 13th INTERPOL Forensic Science Symposium, Lyon, France

Lin, H.-Y., Dyakov, Y. A., Lee, Y. T., & Ni, C.-K. (2021). Temperature Dependence of Desorbed Ions and Neutrals and Ionization Mechanism of Matrix-Assisted Laser Desorption/Ionization. *Journal of the American Society for Mass Spectrometry*, *32*(1), 95–105.

Lu, I.-C., Lee, C., Lee, Y.-T., & Ni, C.-K. (2015). Ionization Mechanism of Matrix-Assisted Laser Desorption/Ionization. *Annual Review of Analytical Chemistry (Palo Alto, Calif.)*, *8*(1), 21–39.

Mahmud, N., Maffei, M., Mogni, M., Forni, G. L., Pinto, V. M., Barberio, G., … Coviello, D. (2021). Hemoglobin A2 and Heterogeneous Diagnostic Relevance Observed in Eight New Variants of the Delta Globin Gene. *Genes*, *12*(11), 1821–.

Maltoni, D., Maio, D., Jain, A. K. & Prabhakar, S. (2009). Handbook of Fingerprint Recognition, 2nd ed. (second edition), Springer, 1-494

Mamyrin, B. A., (2001). Time-of-flight mass spectrometry (concepts, achievements, and prospects). *International Journal of Mass Spectrometry*, *206*(3), 251–266.

McElreath, D. H., Doss, D. A., Jensen III, C. J., Wigginton Jr., M., Kennedy, R., Winter, K. R., Mongue, R. E., Bounds, J., Estis-Sumerel, J. M., (2013) Introduction to Law Enforcement, *Routledge (New York),* 1-479.

McRoberts, A., (2011) The Fingerprint Source Book, U.S. Department of Justice, Office of Justice Programs, National Institute of Justice

Meyer, E., (1903) 'Beiträge zur Leukocytenfrage', Muench Med. Wochenshr, vol. 50 (35), p 1489.

Niu, S., Zhang, W., & Chait, B. T. (1998). Direct comparison of infrared and ultraviolet wavelength matrix-assisted laser desorption/ ionization mass spectrometry of proteins. *Journal of the American Society for Mass Spectrometry*, *9*(1), 1–7.

Odén, S., & Hofsten, B. V. (1954). Detection of Fingerprints by the Ninhydrin Reaction. *Nature (London)*, *173*(4401), 449–450.

Park, J.-L., Kim, J. H., Seo, E., Bae, D. H., Kim, S.-Y., Lee, H.-C., … Kim, Y. S. (2016). Identification and evaluation of age-correlated DNA methylation markers for forensic use. *Forensic Science International : Genetics*, *23*, 64–70.

Patel, E., Cicatiello, P., Deininger, L., Clench, M. R., Marino, G., Giardina, P., … Francese, S. (2016). A proteomic approach for the rapid, multi-informative and reliable identification of blood. *Analyst (London)*, *141*(1), 191–198.

Perutz, M. F., (1962) X-ray analysis of haemoglobin, *Nobel Lecture*, 1-21.

Ramotowski R, Cantu AA, Joullié MM, Petrovskaia O., (1997), 1,2-Indanediones: a preliminary evaluation of a new class of amino acid visualizing compounds. *Fingerprint Whorld* 23 (90) 131–140.

Ramotowski, R. S. (2012). Lee and Gaensslen's Advances in fingerprint technology, 3rd ed. (third edition). Boca Raton, FL: CRC Press.

Scotcher, K., & Bradshaw, R. (2018). The analysis of latent fingermarks on polymer banknotes using MALDI-MS. *Scientific Reports*, *8*(1), 8765–13.

Seashols, S. J., Cross, H. D., Shrader, D. L., & Rief, A. (2013). A Comparison of Chemical Enhancements for the Detection of Latent Blood. *Journal of Forensic Sciences*, *58*(1), 130–133.

Soltwisch, J., Jaskolla, T. W., Hillenkamp, F., Karas, M., & Dreisewerd, K. (2012). Ion Yields in UV-MALDI Mass Spectrometry As a Function of Excitation Laser Wavelength and Optical and Physico-Chemical Properties of Classical and Halogen-Substituted MALDI Matrixes. *Analytical Chemistry (Washington)*, *84*(15), 6567–6576.

Spengler, B., Hubert, M., Kaufmann, R., (1994), MALDI ion imaging and biological ion imaging with a new scanning UV-laser microprobe, Poster presented at the 42nd Annual Conference on Mass Spectrometry and Allied Topics, Chicago, USA. [*https://www.uni-giessen.de/fbz/fb08/Inst/iaac/spengler/forschung/dateien/poster_maldi_anwendung*]

Steiner, R., & Bécue, A. (2018). Effect of water immersion on multi- and mono-metallic VMD. *Forensic Science International*, *283*, 118–127.

Sundar, L., & Rowell, F. (2014). Detection of drugs in lifted cyanoacrylate-developed latent fingermarks using two laser desorption/ionisation mass spectrometric methods. *Analyst (London)*, *139*(3), 633–642.

Sundar, L., & Rowell, F. (2015). Drug cross-contamination of latent fingermarks during routine powder dusting detected by SALDI TOF MS. *Analytical Methods*, *7*(9), 3757–3763.

Tanaka, K., Waki, H., Ido, Y., Akita, S., Yoshida, Y., Yoshida, T., & Matsuo, T. (1988). Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry*, *2*(8), 151–153.

Tang, H.-W., Lu, W., Che, C.-M., & Ng, K.-M. (2010). Gold Nanoparticles and Imaging Mass Spectrometry: Double Imaging of Latent Fingerprints. *Analytical Chemistry (Washington)*, *82*(5), 1589–1593.

Vidaki, A., Ballard, D., Aliferi, A., Miller, T. H., Barron, L. P., & Syndercombe Court, D. (2017). DNA methylation-based forensic age prediction using artificial

neural networks and next generation sequencing. *Forensic Science International : Genetics*, *28*, 225–236.

Vos, D. R. N., Ellis, S. R., Balluff, B., & Heeren, R. M. A. (2021). Experimental and Data Analysis Considerations for Three-Dimensional Mass Spectrometry Imaging in Biomedical Research. *Molecular Imaging and Biology*, *23*(2), 149–159.

Wilder, H. H., & Wentworth, B., (1918). Personal Identification: Methods for the Identification of Individuals, Living Or Dead, *Anthropometry*,

Wilkinson, D. (2000). Study of the reaction mechanism of 1,8-diazafluoren-9-one with the amino acid, l-alanine. *Forensic Science International*, *109*(2), 87–103.

Wolstenholme, R., Bradshaw, R., Clench, M. R., & Francese, S. (2009). Study of latent fingermarks by matrix-assisted laser desorption/ionisation mass spectrometry imaging of endogenous lipids. *Rapid Communications in Mass Spectrometry*, *23*(19), 3031–3039.

Wong, M., Choo, S.-P., & Tan, E.-H. (2009). Travel warning with capecitabine. Annals of Oncology, 20(7), 1281–1281.

Wood, M. A., & James, T. (2009). ORO. The Physical Developer replacement? *Science & Justice*, *49*(4), 272–276.

Xiong, J., Jiang, H.-P., Peng, C.-Y., Deng, Q.-Y., Lan, M.-D., Zeng, H., … Yuan, B.-F. (2015). DNA hydroxymethylation age of human blood determined by capillary hydrophilic-interaction liquid chromatography/mass spectrometry. *Clinical Epigenetics*, *7*(1), 72–72.

Yagnik, G. B., Korte, A. R., & Lee, Y. J. (2013). Multiplex mass spectrometry imaging for latent fingerprints. *Journal of Mass Spectrometry.*, *48*(1), 100–104.

Yamashita, M., & Fenn, J. B. (1984). Electrospray ion source. Another variation on the free-jet theme. *J. Phys. Chem.; (United States)*, *88*(20), 4451–4459.

Yi, S. H., Xu, L. C., Mei, K., Yang, R. Z., & Huang, D. X. (2014). Isolation and identification of age-related DNA methylation markers for forensic age-prediction. *Forensic Science International : Genetics*, *11*(1), 117–125.

Zbieć-Piekarska, R., Spólnicka, M., Kupiec, T., Makowska, Ż., Spas, A., Parys-Proszek, A., … Branicki, W. (2015a). Examination of DNA methylation status of the ELOVL2 marker may be useful for human age prediction in forensic science. *Forensic Science International : Genetics*, *14*, 161–167.

Zbieć-Piekarska, R., Spólnicka, M., Kupiec, T., Parys-Proszek, A., Makowska, Ż., Pałeczka, A., … Branicki, W. (2015b). Development of a forensically useful age prediction method based on DNA methylation analysis. *Forensic Science International : Genetics*, *17*, 173–179.

Zubakov, D., Liu, F., Kokmeijer, I., Choi, Y., van Meurs, J. B. ., van IJcken, W. F. ., … Kayser, M. (2016). Human age estimation from blood using mRNA, DNA methylation, DNA rearrangement, and telomere length. *Forensic Science International : Genetics*, *24*, 33–43.

https://www.cbsnews.com/news/identical-dna-murder-has-georgia-police-seeing-double-which-twin-did-it/ [*last accessed: 05/02/22*]

https://abcnews.go.com/TheLaw/atlanta-twin-murder-case-echoes-fingerprint-origins/story?id=9909586 [*last accessed: 05/02/22*]

Chapter 2: Determination of sex from fingermarks through the analysis of endogenous peptides and proteins using MALDI MS

Introduction

Fingermarks are currently the most commonly collected evidence type from crime scenes, and, despite DNA profiling, the majority of suspect identifications results from fingerprinting (**Bleay, 2019**). Whilst fingermarks are used to match a suspect to a fingerprint record in a National Database (in the UK named IDENT1), fingermarks can yield more than just the biometric information contained through the corresponding ridge flow and *minutiae*. As they are comprised of eccrine and sebaceous secretions, they can reveal biological and chemical information about their donors, giving an insight into an individual's physiological, pathological and pharmacological state (**Francese *et al.,* 2013**). Part of this biochemical profile that would be advantageous to exploit is the determination of the sex of an individual in order to contribute to their identification or to reduce the pool of suspects. Currently, *in lieu* of any witness statements, it is only possible to identify the sex of an individual who was present at a crime scene through the recovery of a DNA sample, which must then undergo DNA profiling. This is an expensive and time consuming process. In addition, DNA is sensitive to environmental conditions, susceptible to degradation and often found in such low abundance that a full profile cannot be generated. Furthermore, if the sample was recovered from a frequently touched surface, mixed DNA profiles are often discovered. If the sex of an individual could be obtained through the chemical content of their fingermark, a valid alternative would be available.

By 2012, a few groups had attempted the determination of the sex of an individual by mass spectrometry based techniques to exploit the lipid content of fingermarks using Laser Desorption Ionisation (LDI) Mass Spectrometry (**Asano**

et al., **2002 & Emerson** *et al.,* **2011**). However, these authors reported failure of the methods tested for sex classification. Other groups tried a different approach, utilising a dual enzyme cascade assay to measure the absorbance of the complex formed between the L-amino acid oxidase type IV (L-AAO) and the horseradish peroxidase type VI (HRP) bioassay (**Huynh** *et al.,* **2015**). They reported that this resulted in a 99% correct predictive value for the discrimination between male and female donor's marks. However, these promising claims were based upon an artificial "fingermark mixture" and the scope did not extend to trialling actual groomed, ungroomed or natural fingermarks. Furthermore, the study did not investigate the compatibility of the method with enhanced marks, which, in the course of an operational workflow, would likely have been visualised *in situ*, to aid with visualisation and recovery, and only made available for mass spectrometric analysis post-enhancement. This was considered in a follow up publication by the same group, using 1,2-Indanedione enhanced marks (**Mekkaoui Alaou & Halámek, 2019**). 1,2-Indanedione is a reagent that targets amino acids, producing a fluorescence response, usually recommended for enhancing latent marks on paper (**Bandey** *et al.,* **2014 & Lee & Joullié, 2015**). The Halámek study found that the fluorescence intensity of the resulting amino acid-1,2-Indanedione complexes *in solution* was twice that in the female samples compared to their male counterparts. They reported that this was a consequence of "the concentration of amino acids in fingerprint residue in females is almost the double of their concentration in males", although this claim was not supported with empirical evidence in the 2019 Halámek publication. This was in contrast to a claim made by Croxton *et al.* in 2010, who had observed that although mean amino acid levels in sweat within latent marks were higher in females than males, in only

one instance for asparagine was this difference significant (P < 0.05) (**Croxton et al., 2010**). However, the Halámek group had previously reported values for average amino acid concentrations derived from sweat from male and female donors to be approximately double for females compared to males, although this was concentrations in sweat and not fingermarks (**Huynh et al., 2015**). This claim was supported by an earlier study by Mark and Harding, who also analysed amino acid composition in eccrine sweat and reported concentrations to be approximately twice that for females compared to males across each amino acid (**Mark & Harding, 2013**).

This proposed method has the advantage that, in addition to being a quick and simple assay, visualisation of the latent mark is obtained simultaneously, and in a non-destructive manner. However, while promising, the study is limited by a small sample population size of only four donors (2 male, 2 female) which is not accompanied by any statistical interpretation of the results Furthermore, the authors noted that while the *in solution* assay yielded fluorescence intensity of twice that of the male samples for female samples, on latent marks on paper, the difference was less pronounced. They ascribed this observation to the pressure applied by the donors at the point of deposition, bringing into question the operational robustness of this method, due to the variability in the pressure applied by individuals when leaving marks at the scene of a crime.

More recently, the application of Desorption Electrospray Ionisation (DESI) Mass Spectrometry Imaging (MSI) yielded the discrimination of female and male donors fingermarks (12 samples from 2 female and 2 male donors), by exploiting, again, their lipid composition (**Bardin et al., 2018**). DESI offers a non-destructive technique with minimal sample preparation, under ambient conditions, and is able to acquire a large molecular profile in a single

experiment. Principle component analysis (PCA) was applied to discriminate between donors. Negative ionisation DESI yielded a 96% sex prediction, whilst positive mode yielded a 100% correct sex prediction. While, in this instance, this publication shows promising findings, only a sample size of 4 donors was used, the type of marks employed was not reported and compatibility with fingermark enhancement techniques was not attempted. Statistically, only the cross validation data was reported.

Subsequently, to date, there are no analytical methods that can be proposed to determine sex from fingermarks in an operational context. A study, published in 2012 by Ferguson *et al.*, provided a proof-of-concept for a proteomics-based approach to solving this problem (**Ferguson *et al.*, 2012**). This study employed Matrix Assisted Laser Desorption Ionisation Mass Spectrometry (MALDI MS) to detect endogenous peptides and small proteins present in the fingermarks. These species are excreted in eccrine secretions, meaning proteins and peptides are present in an aqueous environment made up of ~98% water and ~2% organic and inorganic species; amino acids and urea, and electrolytes (**Knowles, 1978**). The quantity of peptidic material present in a fingermark was recently empirically determined to be between 0.2 – 51. µg (**Oonk *et al.*, 2018**). Despite the relatively low abundance, especially compared to that of lipids, Ferguson *et al.* reported that, by detecting fingermark peptide and protein profiles via MALDI MS and processing the resulting spectral datasets with Partial Least Squares Discriminant Analysis (PLS-DA), sex discrimination of the donor was possible with a predictive accuracy of between 67.5 and 74.4%. This increased to 85% if less stringent classification criteria were applied. However, in this preliminary study, only 80 participants aged between 18-45 years were included. Those that had used drugs or medication within the last 2 weeks were

excluded, and importantly, ungroomed marks were collected. Removing external contaminants is useful for preliminary proof-of-concept studies under laboratory conditions, but this does not translate to crime scene marks, which could be contaminated with a near-infinite number of compounds (**Francese *et al.,* 2013**). In contrast to many possible external contaminants, one would expect to find similar groups of endogenous species present within fingermarks; comprising of peptides and proteins from eccrine sweat, and lipids and fatty acids from contact with areas of the body with sebaceous glands, such as the face (**Francese *et al.,* 2017**). While there may be some intra-individual variation qualitatively and quantitatively, it is likely that these groups of species would be present in all natural marks. Excreted peptides and proteins are good targets for interrogation; they are more persistent than DNA, which can degrade when exposed to harsh environmental conditions and can yield amplification issues. The lipid content of marks has been studied extensively, but their composition diminishes significantly over time and varies depending on time of day, complicating establishing baseline levels. Furthermore, their validity has not been demonstrated (**Francese *et al.,* 2017**). Amino acids make an ideal target for analysis as they are both more robust than DNA and more consistently expressed than lipids.

One of the aims of this PhD programme was to build upon the knowledge gained from this previous study, and try to assimilate the technique with what would be encountered operationally with crime scene marks. This was proposed by implementing a number of features to the study: (i) increasing participant sample size (from 80 to 172 donors), to increase and validate statistical findings; (ii) the restriction criteria were removed, although samples were not collected from individuals younger than 18 years of age over ethical

and privacy concerns; (iii) Natural marks were collected, to more realistically mimic what would be encountered at crime scenes. As latent marks are routinely visualised at crime scenes, to reveal their presence and ultimately to match it to a database fingerprint record, it was important in this study to investigate, for the first time, compatibility with previously enhanced marks and determine whether the enhancement technique interfered with the mass spectrometric detection. Finally, different statistical approaches were employed for the processing of the data obtained. In particular, an XGBOOST method was used which was released in 2014, 2 years after the conclusion of the initial study, as opposed to the Partial Least Squares – Discriminant Analysis (PLS-DA) and Variable Importance in Projection (VIP) scoring used previously.

The work described in this chapter was the subject of an article published in Forensic Chemistry, entitled 'Investigating sex determination through MALDI MS analysis of peptides and proteins in natural fingermarks through comprehensive statistical modelling' (Heaton *et al.,* 2020).

Experimental

Materials

ALUGRAM® SIL G/UV254 pre-coated aluminium sheets, α-cyano-4-hydroxycinnamic acid (α-CHCA) and trifluoroacetic acid (TFA) were purchased from Sigma-Aldrich (Gillingham, UK). Acetone, acetonitrile and methanol were purchased from Fisher Scientific (Loughborough, UK). Doubled sided conductive copper foil shielding tape was purchased from 3M (St. Paul, MN, USA). Indium tin oxide (ITO)-coated slides were purchased from Delta Technologies Ltd. (Loveland, CO, USA). Sirchie Indestructible White "Hi-Fi" Volcano, WA White and WA Aluminium latent fingerprint enhancement powders

were purchased from WA Products (Essex, UK). Sirchie Silk Black "Hi-Fi" Volcano latent fingerprint enhancement powder was purchased from Sirchie (Youngsville, NC, USA). The Breeze™ single-use fibreglass zephyr brushes were purchased from SceneSafe (Essex, UK).

Instrumentation and instrumental parameters

For this study, MALDI mass spectrometric analyses were conducted using two Bruker rapifleX MALDI TOF/TOF mass spectrometers (Bruker Daltonic GmbH, Bremen, Germany) equipped with a neodymium-doped yttrium aluminium garnet (Nd:YAG) 355 nm SmartBeam™ 3D laser operating at a repetition rate of 10kHz, with a laser spot size of 95 x 95 μm. Calibration was performed between each set of donor samples using a Bruker Protein Calibration Standard I mixed with adrenocorticotropic hormone corticotropin-like intermediate peptide (ACTH CLIP) fragment [18-39] to provide a calibration range from 2,465 - 12,361 Da. Calibration was performed in quadratic mode with at least four calibration points. Data acquisition was carried out in positive linear mode between 2,000 and 12,600 Da. Each acquisition contained 10,000 shots from each of the matrix spots across the mark, to correct for matrix 'sweet spots', giving a sum total of 30,000 shots per mark.

Methods

Fingermark collection

In preparation for this study, 6 natural fingermarks were collected from each participant, from staff at West Yorkshire Police and Sheffield Hallam University. 172 individuals provided samples, comprising of 90 female and 82 male donors. All donors were at least 18 years of age and samples were collected in accordance with approved SHU ethics application ER17244422. Natural marks

were used with no pre-cleaning or enriching stage, but fingertips were rubbed together to evenly distribute fingertip material, before depositing three marks from the left and right hands onto 2 separate pre-cleaned aluminium TLC sheets, which had had the silica removed.

These were immediately placed inside plastic microscope slide holders and kept at ambient conditions until analysis. All participants were provided with a participant information form and signed a participant consent form before collection, and filled out a participant questionnaire form immediately after fingermark deposition (**Appendix 1, (i)-(iii)**).

Fingermark enhancement

Of the two sets of three marks, one set was enhanced with Sirchie Indestructible White powder, while the other set remained unenhanced. Each set of marks on the TLC sheets were enhanced with a separate disposable zephyr brush, to avoid cross-contamination of material between sample sets.

Sample preparation for MALDI analysis

Fingermarks were stored in plastic microscope slide holders at ambient conditions before analysis. Once removed, the aluminium sheets were mounted on conductive ITO slides with double-sided copper tape. 0.5 µL of calibration standard was pipette mixed with 0.5 µL α-CHCA matrix at 5 mg/mL in 70:30 ACN:0.5% TFA$_{aq}$ near the top edge of the slide. For both the enhanced and non-enhanced sets of marks, 6 matrix spots were applied down the middle of the marks (0.5 µL, same formulation as with the standard), of each individual set, directly before MALDI analysis (**Fig. 17**). Only 3 were used for sample acquisition, but the additional spots were to account for instances where the

matrix-analyte crystals had not formed homogenously, which could be determined visually when the plate was viewed under magnification in the instrument, before acquisition. All samples were analysed within 5 weeks of collection.
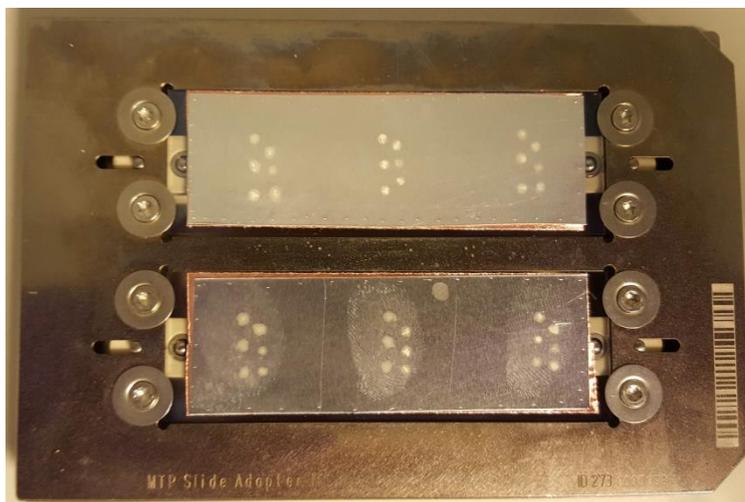
Figure 17. Aluminium sheets mounted on ITO slides with conductive copper tape and secured in a Bruker plate adapter. Matrix spots can be seen within the marks on the surface, and the calibrant spot can be seen in the top right of the middle mark section.

Statistical analysis

Pre-processing

During the acquisition, the instrument obtained 10,000 shots on each of the 3 matrix spots spotted onto the mark, summing these into a single 30,000 shot acquisition for each individual mark. Statistical models were devised where each mark was treated as an individual sample, and also which combined and averaged the spectra from the 3 marks from the same donor into one file for processing.

The raw mass spectra were converted into imzML file formats using R (v 3.6.1, R Foundation for Statistical Computing, Vienna, Austria) with the MALDIquant

package. The spectra were then imported into SCiLS Lab software (v 5.12.0, Bruker Daltonics, GmbH, Bremen, Germany). Baseline correction was applied using the tophat parameters and a 200 interval width. For both the enhanced and non-enhanced data sets, the whole data sets were averaged to produce a single spectrum, which was imported into mMass (v 5.5.0) to identify potential signals of interest that could be used for discrimination. Peak picking, smoothing and baseline correction were applied (**Fig. 18**).
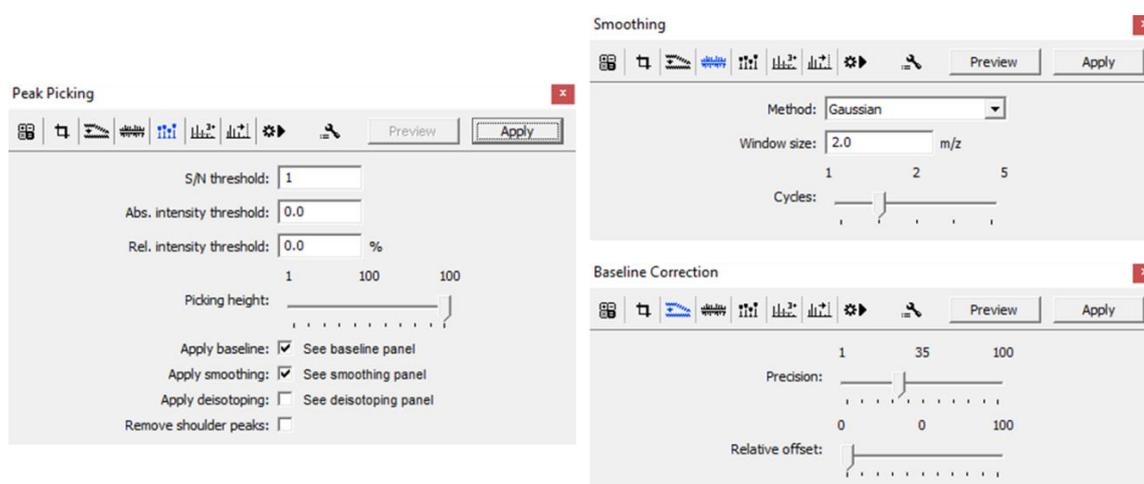
Figure 18. mMass peak picking, smoothing and baseline correction parameters used. A S/N threshold of 1:1 was used, with 100% peak picking height. Gaussian smoothing was used with a window size of 2 m/z through 2 cycles. Bassline precision was set to 35%.

A signal-to-noise threshold of 1:1 was selected in mMass in order to generate a sufficiently large peak list. An S/N threshold of 3:1 was trialled, but generated a peak list of only 29 signals, which was not enough to attempt a discriminant analysis (LDA). Lowering the S/N threshold to 2:1 generated a peak list of 48 signals, which was still low to attempt an LDA. Lowering the S/N threshold again to 1:1 increased the number of picked peaks to 139, which allowed for a more robust LDA. This peak list was then imported into SCiLS Lab and used as a reference mass list for Linear Discriminant Analysis (LDA) classification.

As part of a separate approach, an R script was written to automatically perform initial processing of both the male and female samples. It applied baseline correction, TIC normalisation and spectral smoothing using the MALDIquant package. In this approach, peak picking was also performed using the MALDIquant package. The signal-to-noise threshold and the required minimum frequency rate were varied to identify the most effective parameters. The results of altering the S/N threshold and frequency rate are shown in **Table 7**.

S/N	Required fraction of spectra needed for a peak to be included in the analyses (minFreq)			
	1%	**10%**	**50%**	**90%**
2	2867	2764	141	6
3	2570	1773	61	-
5	869	428	27	-
10	335	90	6	-
20	168	19	-	-

Table 7. The results of altering the signal-to-noise threshold and frequency rate values on the number of peaks obtained in the peak list in the R MALDIquant package.

Statistical processing

A 10-fold supervised Linear Discriminant Analysis classification was performed in SCiLS Lab on both the enhanced and non-enhanced data sets. The model distribution is based on the assumption that all data is sampled from a multivariate normal distribution, with different means in each group and identical covariance across the groups. In the training phase, the group means and dataset covariance are calculated. For classification, the spectra are assigned to the group that maximises the probability.

In a parallel approach, classification model training was instead conducted externally to SCiLS Lab. All machine learning has been performed in Python, utilising the scikit-machine learning learn package. A range of classification models and feature selection methods have been systematically compared. k-fold cross validation has been repeatedly used to provide measures of model accuracy (here with k set to 5, 10, 25, 50) to determine the robustness of the accuracy metrics for different k *values*. A full list of models is supplied in **Table 8**.

Model name	Reason for inclusion
LDA	Standard classification model available in the SCiLS Lab software
Random forest	Decision tree-based ensemble learning classification method
Naïve Bayes	A standard baseline model, typically used for benchmarking purposes
XGBOOST	A tree-based ensemble method using gradient boosting that has recently gained large popularity
PLS-DA	Classification method used in the previous study (Ferguson *et al.*, 2012)

Table 8. Overview of classification models used in the current analysis *[adapted from Heaton et al., 2020].*

Within the current cross-validation protocol, during each k of the k-fold train/test splits, all three marks (technical replicates) from a given donor have been explicitly assigned to either the train or test set. In addition to using all picked *m/z* positions as inputs to the classifiers in **Table 8**, a range of distinct feature selection strategies have also been tested (**Table 9**) including PLS-DA (Partial Least Squares-Discriminant Analysis) VIP (Variable Importance in Projection) score, random forest feature importance (based on the decrease in Gini impurity associated with each feature across trees) and univariate feature

selection via a chi-squared test to identify *m/z* peaks that are most likely to be dependent on sex. Importantly these feature selection steps are a precursor to the full classification model training, and any of the model schemes in **Table 8** can be trained using the outputted subset of *m/z* positions from the chosen feature selection strategy. Furthermore, to ensure that each test set of samples remains hidden during cross validation, only the training subset of samples is used during the feature selection strategy, separately for each distinct k-fold cross validation split instance.

Method	Description
None	The baseline method is to use all peak picked m/z positions as features in both model training and validation.
PLS-DA VIP score	In line with the Ferguson *et al*., (2012) methodology, a custom version of the Variable Importance in Projection (VIP) scoring scheme has been implemented in the current analysis in python. Through an initial round of PLS-DA model training, VIP scores per feature (m/z position) can be calculated, in order to identify the subset of m/z positions that were most predictive of sex. This subset of m/z positions was then used for a further round of model training (using any of the chosen models from Table 8).
Random forest feature importance	Similarly to the VIP score above, other classifiers provide alternative means are identifying features (m/z positions) that are deemed most important in the model decision making process. A standard, commonly used approach is to use the random forest feature importance. These are derived based on the extent at which each feature contributes to the decrease in weighted Gini impurity (averaged across decision trees in the forest), during training. Again, the subset of identified m/z positions was then used for a further round of model training (using any of the chosen models from Table 8).

	An additional strategy for identification of 'best' features for model training utilises a univariate statistical test (here a Chi-squared test) to identify the features that are most likely to be dependent on sex, and therefore relevant for classification. Again, the subset of identified m/z positions was then used for a further round of model training (using any of the chosen models from Table 8).
Univariate feature selection	

Table 9. Summary of the statistical methods employed in the current analysis *[adapted from Heaton et al., 2020]*.

Neural network implementation for multi-class classification – The Keras package (with Tensorflow backend) was used to run a dense neural network architecture for joint prediction of sex and contaminants for the unenhanced fingermark dataset. Keras is an open source software library that acts as an Application Programming Interface (API) for artificial neural networks within Python. Python is a programming language and open source software that can run many scientific and mathematical software packages. The architecture consisted of two fully connected dense layers (of 100 and 10 nodes respectively), incorporating 20% dropout and Rectified Linear Unit (RelU) Activation Function layers between each dense layer. An Activation Function is a mathematical function that transforms a certain input to return a required output. They can apply non-linearity to the output, allowing the network to learn complex patterns within the data, which would otherwise behave like a linear regression model, with limited learning capacity. A ReLU Activation Function is one that will output the input directly if positive, or output zero if negative, and can be expressed as $f(x) = \max(0, x)$. (**https://iq.opengenus.org/relu-activation**).

Two contrasting output layer designs have been compared: (i) an output layer consisted of two nodes with softmax activation for sex prediction only, and (ii) an output layer consisted of two pairs of two nodes, with each pair having an

independent softmax normalisation for sex and contamination state prediction. In both cases, the model was trained for 100 epochs with a categorical cross-entropy loss function (one separate loss function for each pair in (ii)). The softmax Activation Function takes a vector of numbers and converts it into a vector of probabilities, relative to the scale of each value in the vector. To understand the softmax function, 2 other functions should be understood. The 'max' function returns the maximum numerical value from a list. The 'arg max' returns the position in a list that matches the maximum numerical value. The 'soft max' function is a "softer" version of the 'arg max' function that returns probabilistic values. The 'arg max' function would return a value of 1 for the list index (position) of the maximum numerical value, and a value of 0 for everything else. Whereas the 'soft max' function would return probabilistic values based on the values in the list, rather than discounting everything that wasn't the maximum value index (**Bishop, 1995**).

Results and Discussion

The proof-of-concept work published by Ferguson *et al.* demonstrated that peptides and proteins could be used to a certain extent to differentiate between fingermarks deposited by male and female individuals (**Ferguson *et al.,* 2012**). However, this study fell short of recommending the technique as part of the operational workflow due to limitations of the study design. The small participant sample size of 80 donors, the use of ungroomed marks (when the hands are washed before deposition, and therefore not representative of crime scene marks), only including participants between 18 and 45 years old and excluding participants who had taken drugs or medication within 2 weeks prior meant the

sample population was greatly restricted and not representative of a real environment. Furthermore, the absence of investigation into the compatibility of the method with prior enhancement with any CSI latent mark visualisation techniques meant that conclusions made were limited and not representative of marks potentially encountered during investigations, and hence, it was not possible to evaluate the translation of the method to 'real' crime scene marks within casework scenarios. However, the method design used previously was useful in informing the current study, and an increased sample size was used. A power calculation was performed as part of the study design to ensure sufficient samples would be collected for a valid statistical conclusion. A Multiple Regression: Omnibus statistical test was performed, using an A Priori power analysis, with an α err prob of 0.05, Power (1-β err prob) of 0.95 and Number of predictors of 1, which indicated that 180 participants would be enough to draw valid statistical conclusions. The researchers attempted to include 200 participants but were limited to 172 by the number of willing volunteers during the sample collection stage.

In further, previously unpublished research by the group (undertaken by Dr. Robert Bradshaw & Dr. Ekta Patel), they concluded that the long time between sample collection and analysis (4 months), and the freezing/thawing of the samples might have contributed to degradation of the integrity of the fingermarks, therefore negatively impacting the protein and peptide signals that could be detected. This research did investigate compatibility of protein and peptide detection via MALDI with an enhancement technique, namely vacuum metal deposition (VMD) using gold nano-particles and was included as part of the Heaton *et al.,* 2020 publication. This research found that the addition of gold

nano-particles caused signal suppression of the peptide and protein spectral peaks compared to those with no enhancement technique applied.

Compared to unenhanced marks, around 20% of the marks showed the peptide and protein spectral profile expected, but at reduced signal intensity, 30% showed peptide and protein signals, but were dominated by gold nano-particle clusters, characterised by a 187 Da mass unit difference and the remaining 50% were devoid of any peptide/protein signals, just yielding gold nano-particle clusters. (**Fig. 19**). These differences were attributed to the quality of the donor ('good', 'bad' or 'average' donors) in relation to the amount of secretion material deposited by an individual within a fingermark.
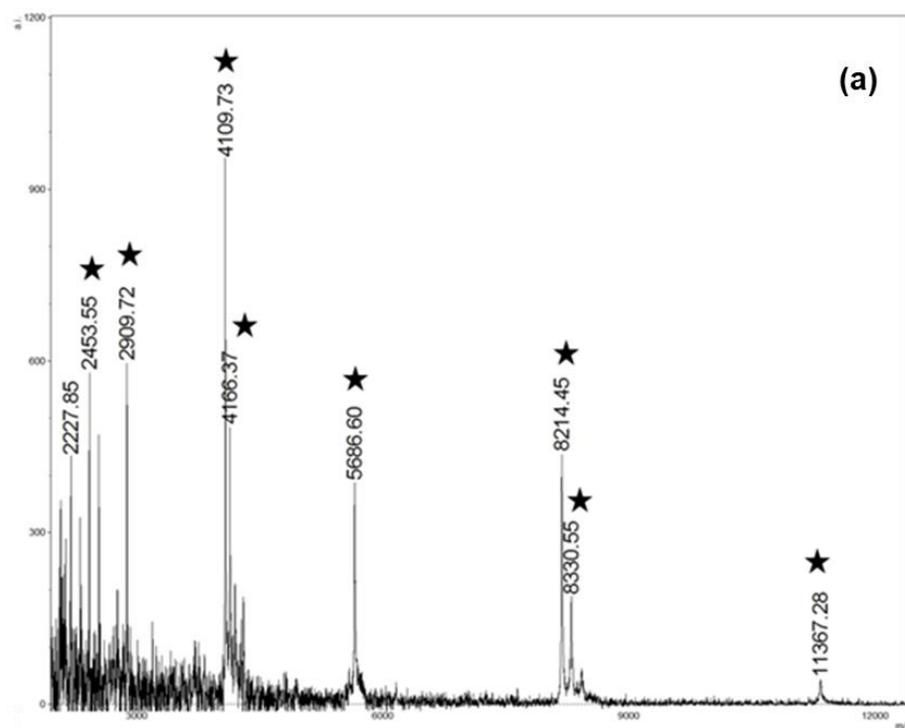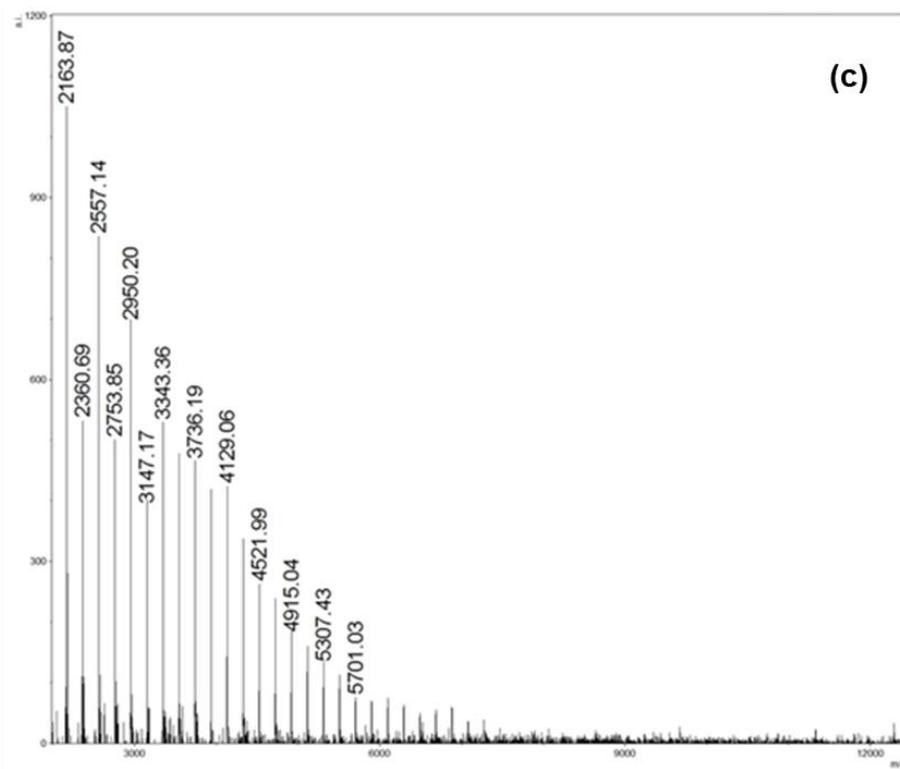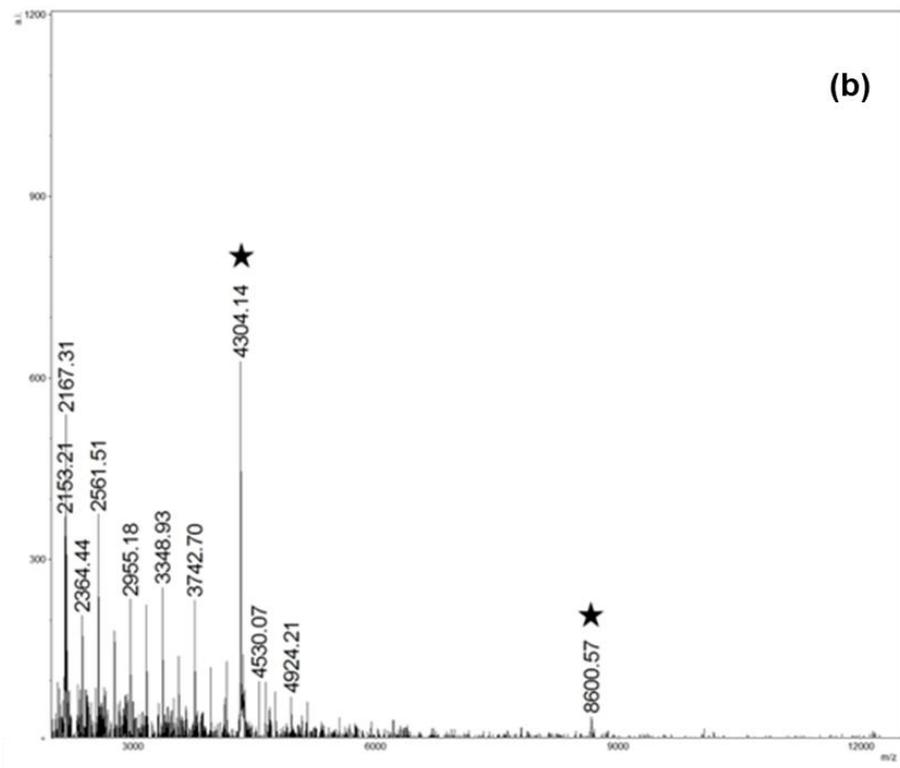
Figure 19. Example MALDI MS spectra of natural marks that had been enhanced with VMD. (a) Example spectra with comprehensive peptide/protein profile. (b) Example spectra with a supressed peptide/protein profile. (c) Example spectra devoid of any discernable peptide/protein signals. Star labels denote identified peptide/protein signals. *[Adapted from Heaton et al., 2020].*

In response to this data, VMD was discounted as an enhancement technique for the present study. Powders are the most commonly used enhancement reagents by CSIs in the UK, and here 4 of the most commonly used CSI powders were trialled to test for possible signal suppression on pre-enhanced marks; Sirchie Indestructible White, Sirchie Silk Black, WA White and WA Aluminium powders (**Fig. 20**).
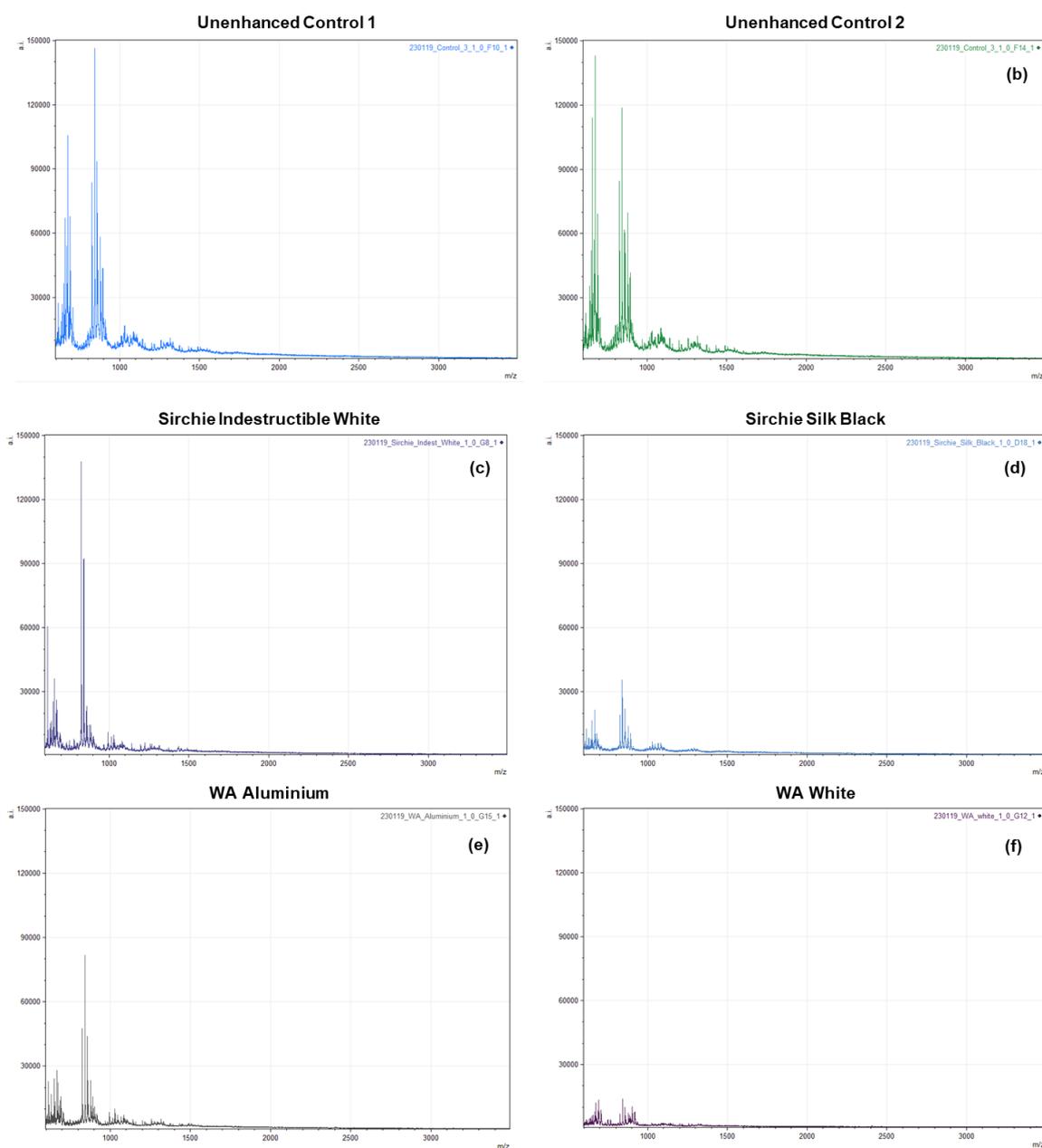
Figure 20. Comparison of spectra analysed on a Bruker Autoflex of unenhanced control fingermarks (a & b) and fingermarks enhanced with (c) Sirchie White, (d) Sirchie Black, (e) WA Aluminium and (f) WA White enhancement powders.

The unenhanced control fingermarks can be seen in panels (**Fig. 20a**) and (**Fig. 2b**), with a maximum signal intensity of ~145,000 a.u. Sirchie Indestructible White enhancement powder caused significantly less analyte signal suppression than each of the other three enhancement powders, as shown in panel (**Fig. 20c**). Each spectra has been scaled to 150,000 a.u. on the y axis, with the most intense signal in the Sirchie Indestructible White enhanced spectra at 140,000 a.u, (**Fig. 20c**), Sirchie Silk Black at 35,000 a.u. (**Fig. 20d**), WA Aluminium at 80,000 a.u. (**Fig. 20e**) and WA White at 15,000 a.u. (**Fig. 20f**). Consequently, Sirchie Indestructible White powder was selected as the enhancement powder to be used in this study to demonstrate compatibility with the method.

The analysis in this preliminary study was carried out on a Bruker autoflex MALDI mass spectrometer. Since then, there have been advances in MALDI instrumentation and Bruker have released a more recent model, the rapifleX MALDI instrument, with improved sensitivity and resolution, which was used in the study presented in this chapter.

In contrast to the preliminary study, here all 1,032 marks were stored under ambient conditions and analysed within 5 weeks of collection in order to minimise any potential sample degradation associated with freeze/thawing and prolonged storage. The author of this thesis analysed all of the samples, thus maintaining consistency and reducing operator variability. Due the improved sensitivity, the rapifleX instrument provided signals up to two orders of magnitude more intense compared to the autoflex instrument (**Fig. 21**).
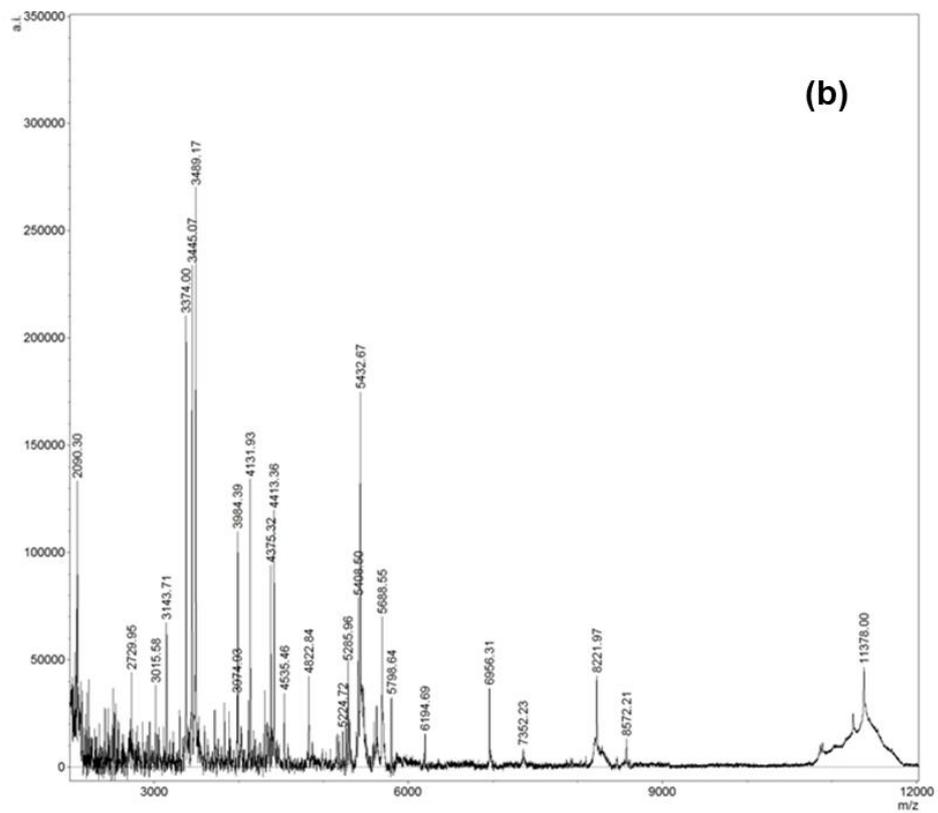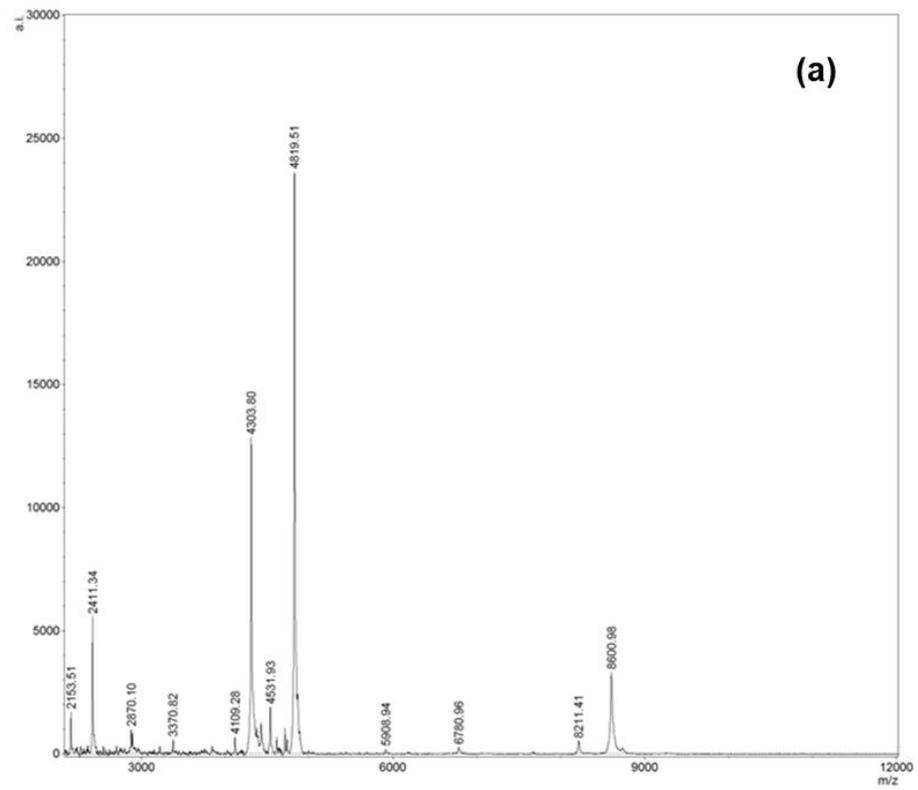
Figure 21. Comparison of signal intensity between Bruker autofleX MALDI MS spectra (a) and rapifleX MALDI MS spectra (b) for non-enhanced fingermarks. (*From Heaton et al., 2020).*

120

Non-enhanced natural marks – the spectra obtained from each mark was imported to mMass where baseline subtraction, smoothing and peak picking were applied. A signal-to-noise threshold of 3:1 resulted in a peak list of 29 signals, which was not a sufficiently large signal dataset to attempt LDA. A signal-to-noise threshold of 2:1 resulted in a peak list of 48 signals, which was sufficient for LDA classification in SCiLS Lab. The 10-fold LDA classification yielded an accuracy score of ~60%, when each fingermark was treated as an individual and unrelated sample (regardless of if they were one of three from a single donor). SCiLS Lab auto-predicts the labels of each training sample during the classification protocol. This auto-prediction feature and the separate treatment of the marks from each donor set presented a limitation in the validity of the SCiLS Lab cross validation calculation (**Fig. 22**).
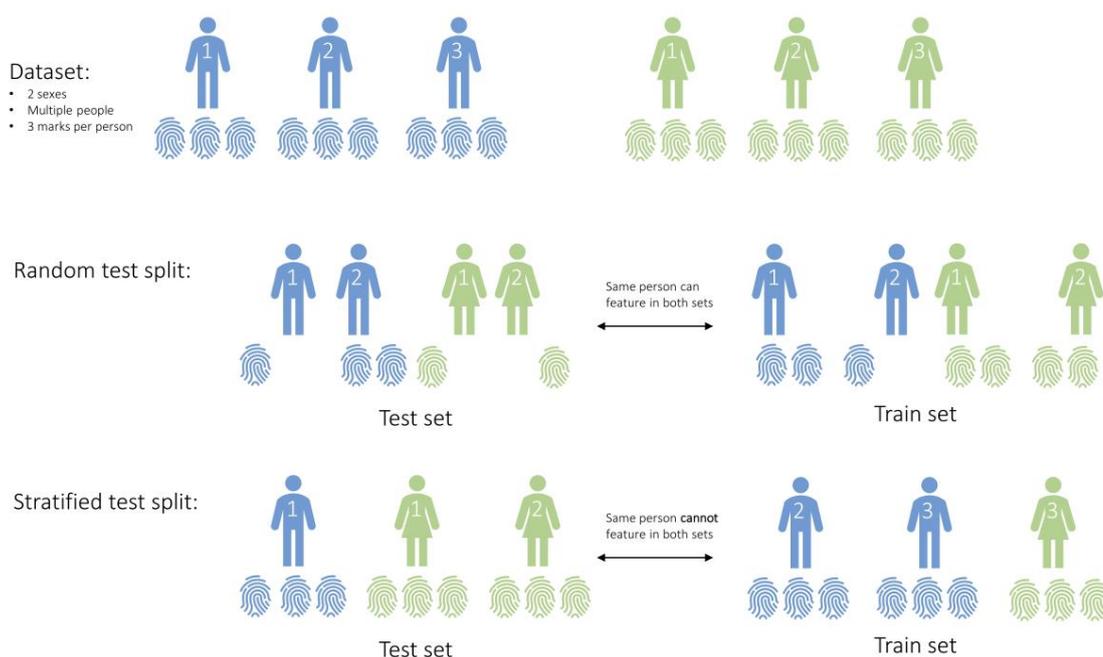
Figure 22. Dataset stratification methods. Top row: the full dataset comprises of 3 fingermarks per person. Middle row: a completely random test/train split may split a donor's 3 marks between the test set and the training set. Bottom row: a stratified test/train split can be performed to ensure that all 3 marks from a single donor are assigned to the same set (either test set or train set) during a particular splitting operation.

SCiLS Lab treats the data as unsupervised when performing each of the k-fold cross validations; it does not know that each of the three technical replicate samples came from the same donor. As a result, a single donor's marks can be split between the training and test sets. This gives the model an advantage due to the similarity of the technical replicates to more easily correctly assign samples to the correct group if they have a replicate in the training set, and is not representative of the model to classify unseen datasets. To address this, classification model training was conducted externally in Python using the scikit-machine learning learn package. A range of distinct feature selection strategies were trialled to evaluate their effectiveness on model predictive performance **(Table 3)**.

This approach was included to allow a direct comparison with the previously published sex determination study (**Ferguson *et al.,* 2012**). The Ferguson *et al.* study used Variable Importance in Projection (VIP) scores to estimate the importance of each variable in the projection in a Partial Least Squares-Discriminant Analysis (PLS-DA) classification model, to identify the most relevant *m/z* peak signals and create a subset to use during model training. K-fold cross validation has been used to determine the accuracy of a model on random 1/k test fractions of the samples.

The accuracy measures for two classification strategies were calculated: (i) *Separate mark scoring:* every one of the 516 marks were treated as individual samples, and the ability of the model to correctly label each separate mark was assessed, (ii) *Majority vote scoring:* the model uses each of the 3 technical replicates in each donor subset and assigns sex based on the majority of the marks. For example, if a male donor provided 3 marks {M1, M2, M3} which the model assigns {Male, Male, Female} then the overall prediction of the donor

would be 'Male', in agreement with the majority of individual mark assignments. However, with this strategy only 172 predictions were made in contrast to strategy (i) where 516 predictions were made.

The majority vote strategy was to allow a comparison to the modelling applied by Ferguson *et al.* (**Ferguson *et al.,* 2012**), although the current study acquired 3 spectra per donor in contrast to the previous 9 spectra per donor, so the strategy is not directly comparable. **Figure 23** represents visually the effect of feature selection for the VIP scoring scheme.
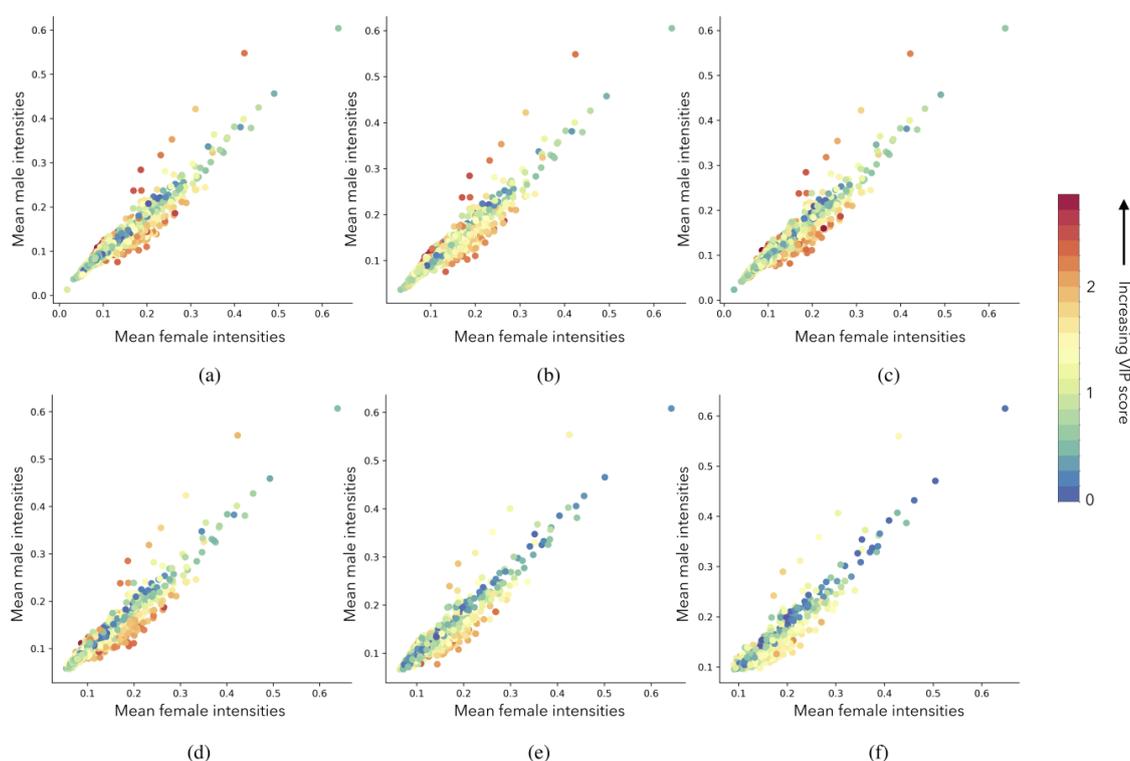
Figure 23. The relationship between the average male and female samples for a subset of selected peak picking strategies. An S/N of 2:1 (a & b), 3:1 (c & d) and 5:1 (e & f) was selected, with a cross-spectra minimum peak occurrence rate of 1% (a, c, e) and 10% (b, d, f). Each point represents a peak picked position and the x and y axis values indicate the average TIC normalised intensity at this peak picked *m/z* position for the female and male samples, respectively. The scatter points have been colour scaled by the calculated PLS-DA VIP score. *[From Heaton et al., 2020].*

For each of the peak picking strategies, the *m/z* signals that were assigned the highest VIP scores (in red) are those that are least correlated between the male samples (x-axis) and female samples (y-axis), on average. **Figure 24** illustrates the overall k-fold accuracy scoring results.
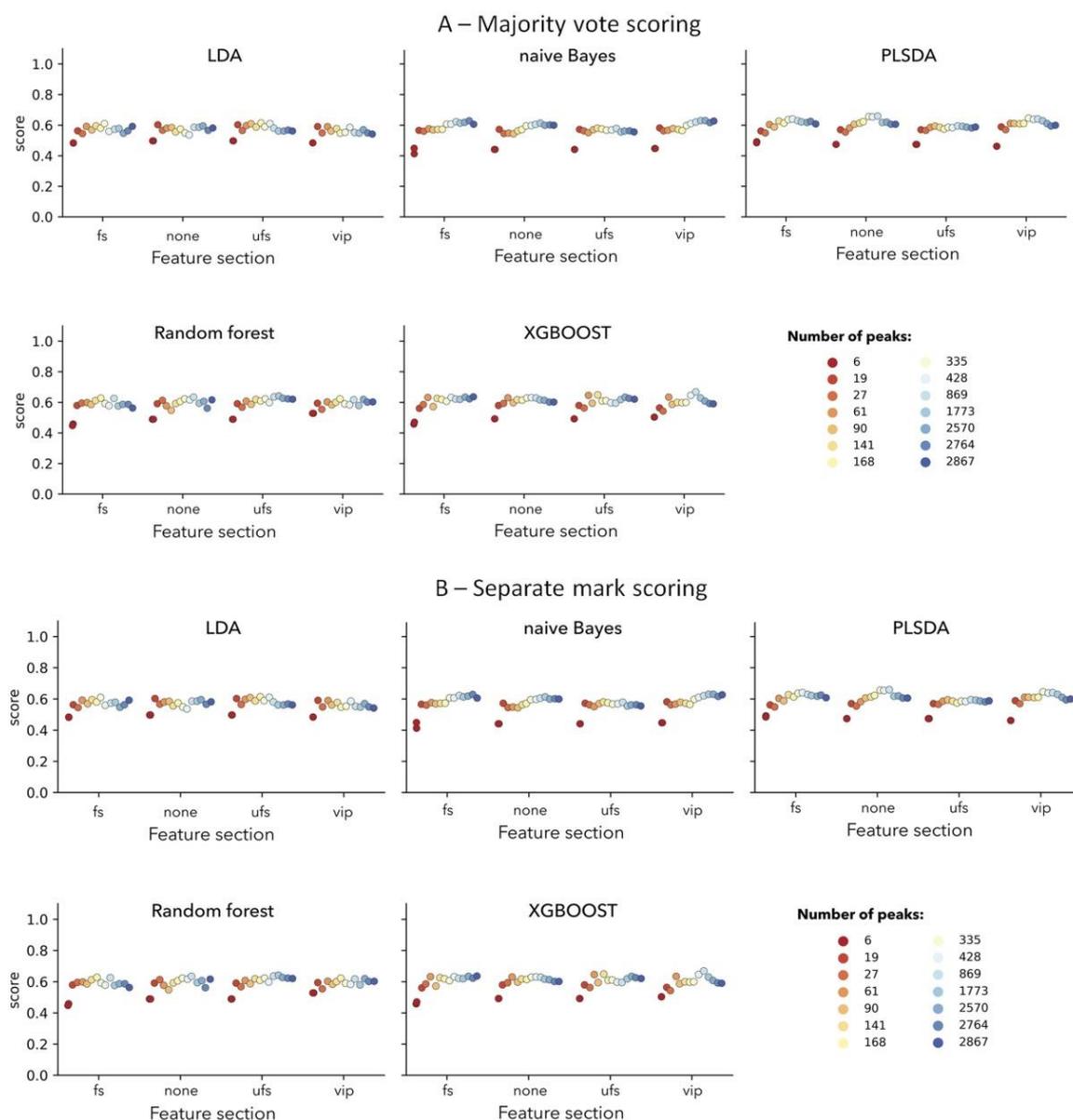
Figure 24. 5-fold cross validation accuracy scores for five model schemes: LDA, random forest, naïve Bayes, XGBOOST and PLS-DA for the (A) majority vote scoring and (B) separate mark scoring schemes. In each subplot, the four different feature selection schemes presented in Table S1 have been displayed separately ('vip' = VIP scoring, 'none' = all m/z peaks (i.e. non-feature selection step), 'fs' = random forest feature selection, and 'ufs' = univariate feature selection). In addition, the accuracy scores have been separated by distinct peak picking strategies, and ordered by the number of m/z positions associated with each strategy (see colour scales). *[From Heaton et al., 2020]*.

A: k=10, Majority vote scoring
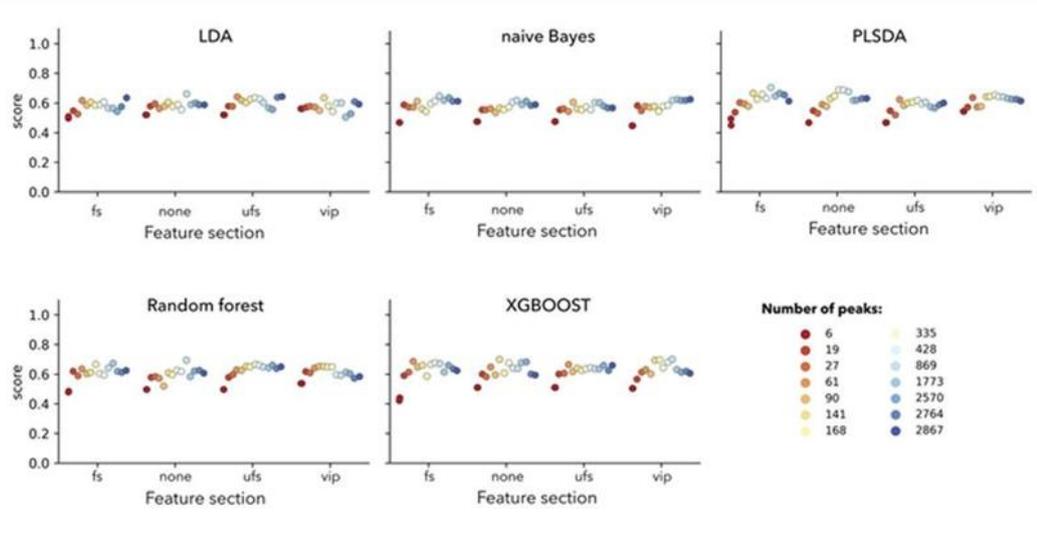
B: k=10, Separate mark scoring

125
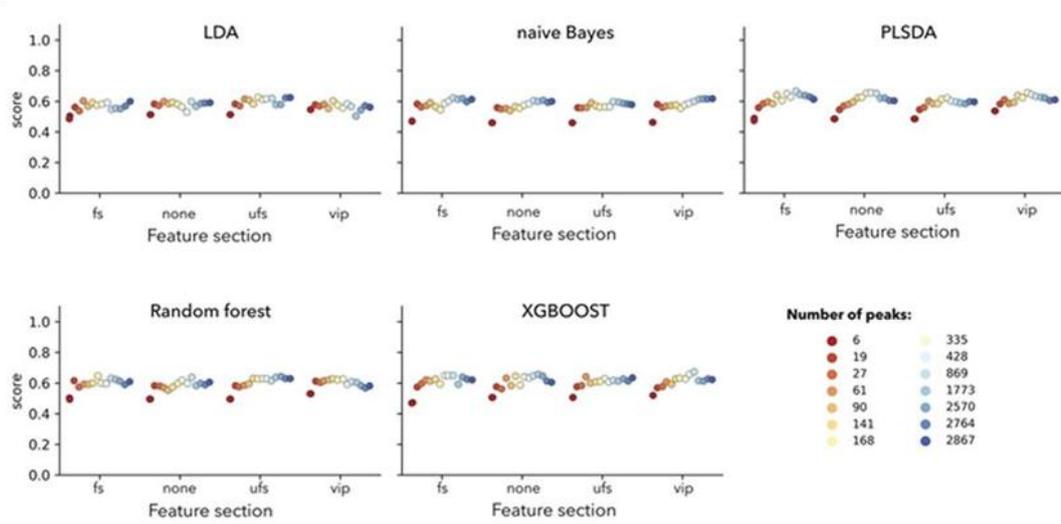
C: k=25, Majority vote scoring
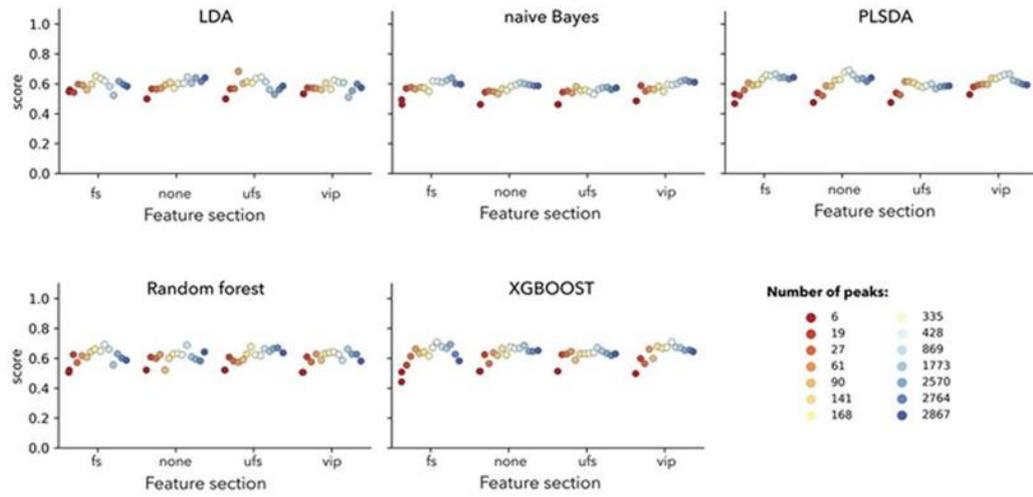
D: k=25, Separate mark scoring

126
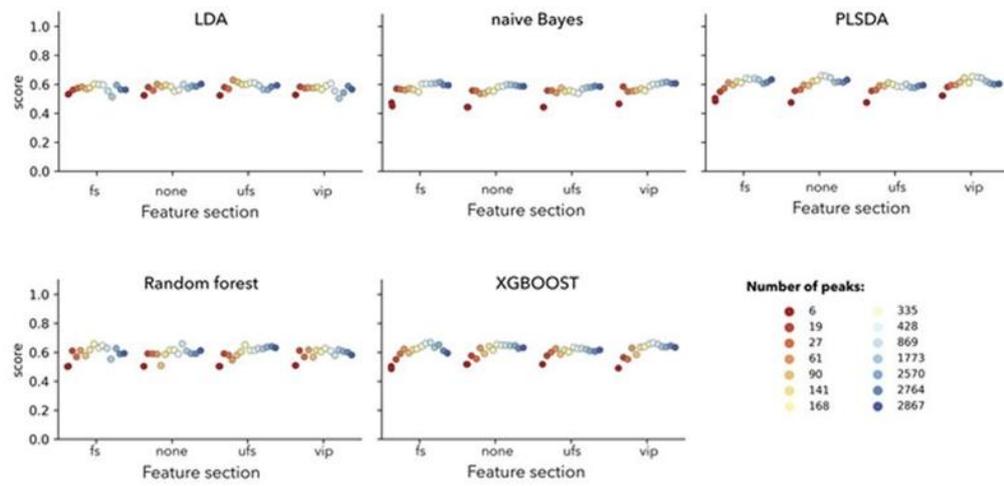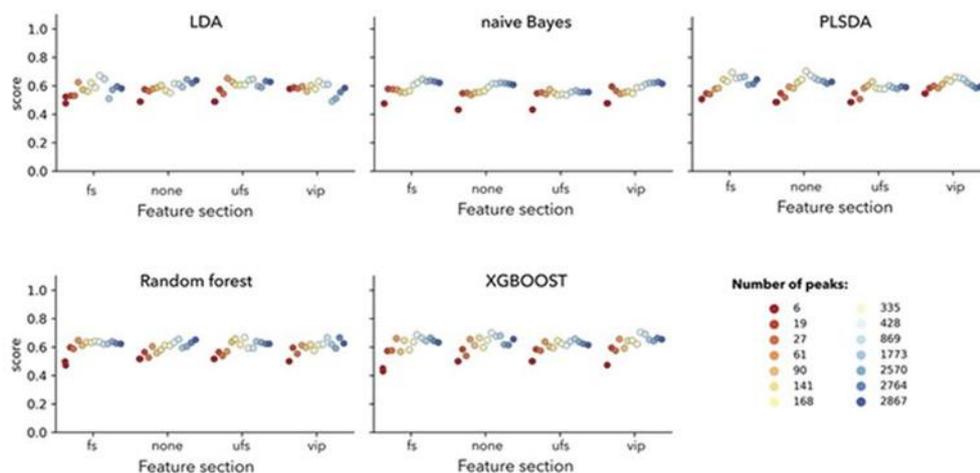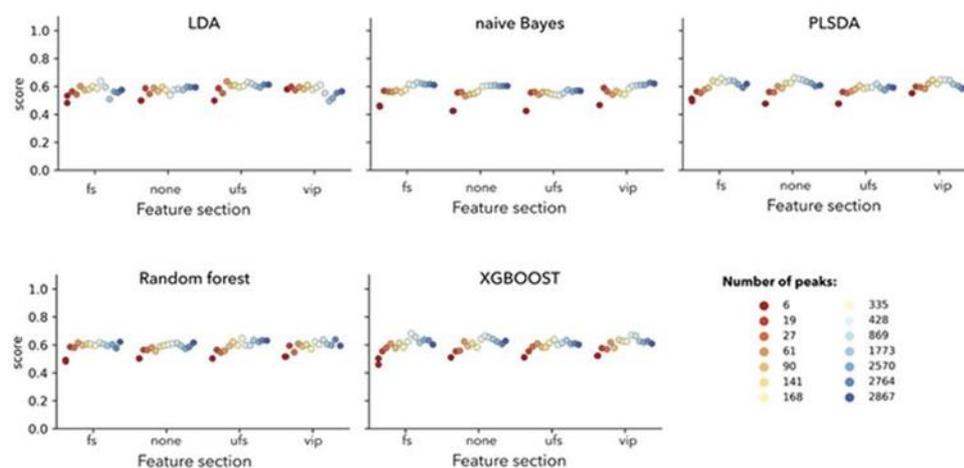
Figure 25. k-fold cross validation accuracy scores for five model schemes: LDA, random forest, naïve Bayes, XGBOOST and PLS-DA for the (A, C, E) majority vote scoring and (B ,D, F) separate mark scoring schemes. Cross validation k values are (A & B) 10, (C & D) 25 and (E & F) 50. In each subplot, the four different feature selection schemes presented in Table S1 have been displayed separately ('vip' = VIP scoring, 'none' = all *m/z* peaks (i.e. non-feature selection step), 'fs' = random forest feature selection, and 'ufs' = univariate feature selection). In addition, the accuracy scores have been separated by distinct peak picking strategies, and ordered by the number of *m/z* positions associated with each strategy (see colour scales).

The results have been split by model type and then within each model type,

each subplot has been further divided into the distinct peak picking strategies

(**Table 7**).

As shown in **Figures 24 and 25**, the computed k-fold cross validation accuracy scores appear to be relatively robust to (i) the choice of k (5, 10, 25, or 50) for cross-validation, (ii) the choice of feature selection strategy and (iii) the choice of using either the separate mark scoring or majority vote scoring strategy to account for the technical replicates not being split between the training and test sets. Conversely, the choice of peak picking parameters does appear to significantly affect the resulting k-fold cross-validation scoring, with a trend between numbers of included *m/z* peak signals and mean accuracy being observed across models. Overall, XGBOOST performed better than other trialled methods. However the performance boost is minimal with a k-fold accuracy of 60-70% attainable by all model types for particular peak picking strategies.

The maximum reported k-fold cross validation score was attained by the XGBOOST model at 70.9%, with k=25 and peak picking parameters of {S:N=5, minFreq=0.1}, under the majority scoring strategy and VIP feature importance selection strategy, as in the Ferguson *et al.* study (**Ferguson *et al.,* 2012**). The top 20 (averaged over k-fold repeats) scoring models are reported in **Table 10**.

| Model | Peak picking strategy | | Number of k-folds | Feature selection strategy | Accuracy score (mean calculated over k-fold repeats) |
	Minimum Frequency	Minimum signal-to-noise			
XGBOOST	minFreq = 0.1	S/N = 5	25	PLS-DA VIP score	0.709
XGBOOST	minFreq = 0.1	S/N = 5	50	PLS-DA VIP score	0.705
PLSDA	minFreq = 0.01	S/N = 10	50	None (use all peak-picked peaks)	0.703
XGBOOST	minFreq = 0.01	S/N = 5	10	PLS-DA VIP score	0.7
XGBOOST	minFreq = 0.1	S/N = 5	50	None (use all peak-picked peaks)	0.7
XGBOOST	minFreq = 0.5	S/N = 2	10	None (use all peak-picked peaks)	0.699
XGBOOST	minFreq = 0.01	S/N = 20	10	PLS-DA VIP score	0.696
Random Forest	minFreq = 0.01	S/N = 5	10	None (use all peak-picked peaks)	0.695
XGBOOST	minFreq = 0.5	S/N = 2	10	PLS-DA VIP score	0.694
XGBOOST	minFreq = 0.01	S/N = 5	50	PLS-DA VIP score	0.692
PLSDA	minFreq = 0.1	S/N = 5	25	None (use all peak-picked peaks)	0.690
PLSDA	minFreq = 0.01	S/N = 10	10	None (use all peak-picked peaks)	0.688
PLSDA	minFreq = 0.01	S/N = 5	50	Random forest feature importance	0.688
PLSDA	minFreq = 0.1	S/N = 5	10	None (use all peak-picked peaks)	0.688
Random Forest	minFreq = 0.01	S/N = 5	25	None (use all peak-picked peaks)	0.688
XGBOOST	minFreq = 0.1	S/N = 3	25	None (use all peak-picked peaks)	0.686
XGBOOST	minFreq = 0.01	S/N = 10	5	None (use all peak-picked peaks)	0.683
LDA	minFreq = 0.5	S/N = 3	25	Univariate feature selection	0.683
XGBOOST	minFreq = 0.01	S/N = 3	10	None (use all peak-picked peaks)	0.683
XGBOOST	minFreq = 0.1	S/N = 5	10	PLS-DA VIP score	0.682

Table 10. Top 20 scoring models (average score calculated over k-fold repeats, with k specified in the 'Number of k-folds' column). The scoring scheme was the 'majority' for all models.

Within the current study, using 3 spectra per donor, the maximal XGBOOST accuracy (~71%) roughly corresponds to the 2/3 majority vote scoring employed as one of the harsher classification criterion by Ferguson *et al.* (**Ferguson *et al.,* 2012**). In the 2012 study, this harsher classification yielded a score of 67.5% for the PLS-DA classifier employing 9 spectra per donor, although the model additionally defined some samples as 'undecided' in addition of the 'male' and 'female' classifiers.

A preliminary PCA analysis identified 4 samples within the dataset as outliers on the plot. Upon inspection, the corresponding mass spectra displayed severe polymer contamination, identifiable by the regular repeating mass units between neighbouring peaks. This particular polymer was concluded to be polyethylene glycol (PEG) -based, based on the 44 Da difference between signals, which completely dominated the spectra and supressed protein and peptide signals in the range m/z 2,000 –5,000 (**Figure 26**).
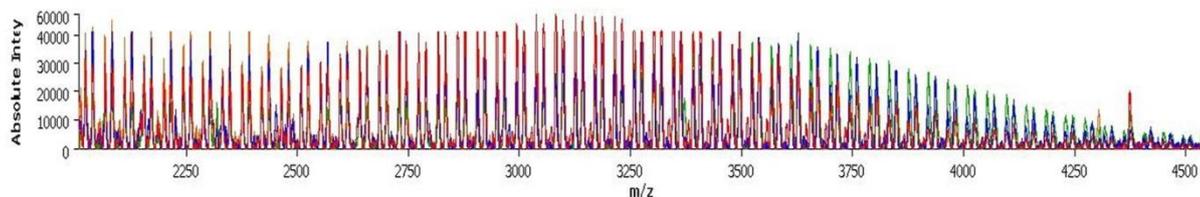
Figure 26. The four spectra that were identified as outliers on the PCA plot, exhibiting severe polymer contamination, supressing the protein and peptide signals.

These 4 spectra originated from the marks of female donors. Upon examination of the corresponding participant questionnaires, each donor had reported using toiletry products likely to contain PEG-based compounds, including, but not limited to, shower gel, shampoo, soap, moisturiser, foundation, body lotions and hair products. 78 out of 90 female participants reported using some type of cosmetic/toiletry products. As these 4 spectra were unclassified by the model due to their overwhelming contamination, and lack of relevant signals, they were removed from the model, and the model training/testing repeated. Omission of these spectra did not lead to any clear improvement in model sex classification performance.

In an alternative strategy, as opposed to training the above sex classification model using each donor's three fingermark samples individually (3 spectra per donor), the model was re-run, but summing the 3 individual spectra per donor into 1 sum total spectra per donor. This was to negate intra-sample variability from each MALDI MS acquisition due to localised matrix crystallisation across the 3 sample spots. However, this alternative approach did not yield better model performance compared to treating each sample as an individual technical replicate.

Further visual examination of the dataset revealed roughly half of all spectra contained some level of polymer contamination, from mild to severe, although

all the remaining spectra (apart from the 4 that were removed) had some protein and peptide signals that could be used for classification purposes. Consequently it was investigated whether this contamination was having an effect on the models predictive accuracy. Previously the data had been classified into two discrete categories; 'male' or 'female'. The model was assessed for when 4 discrete categories were available for predictive classification, (i) male non-contaminated, (ii) female non-contaminated, (iii) male contaminated, (iv) female contaminated. The model was subsequently presented with each of the 4 distinct new classification groups, rather than learning the distinction itself.

Figures 27 and 28 display the results of the XGBOOST and LDA classification models respectively (with other classifiers showing similar results). For each peak picking strategy (i.e. selected S:N ratio), it was not evident that the 4-output state model performed better than the 2-output state model.
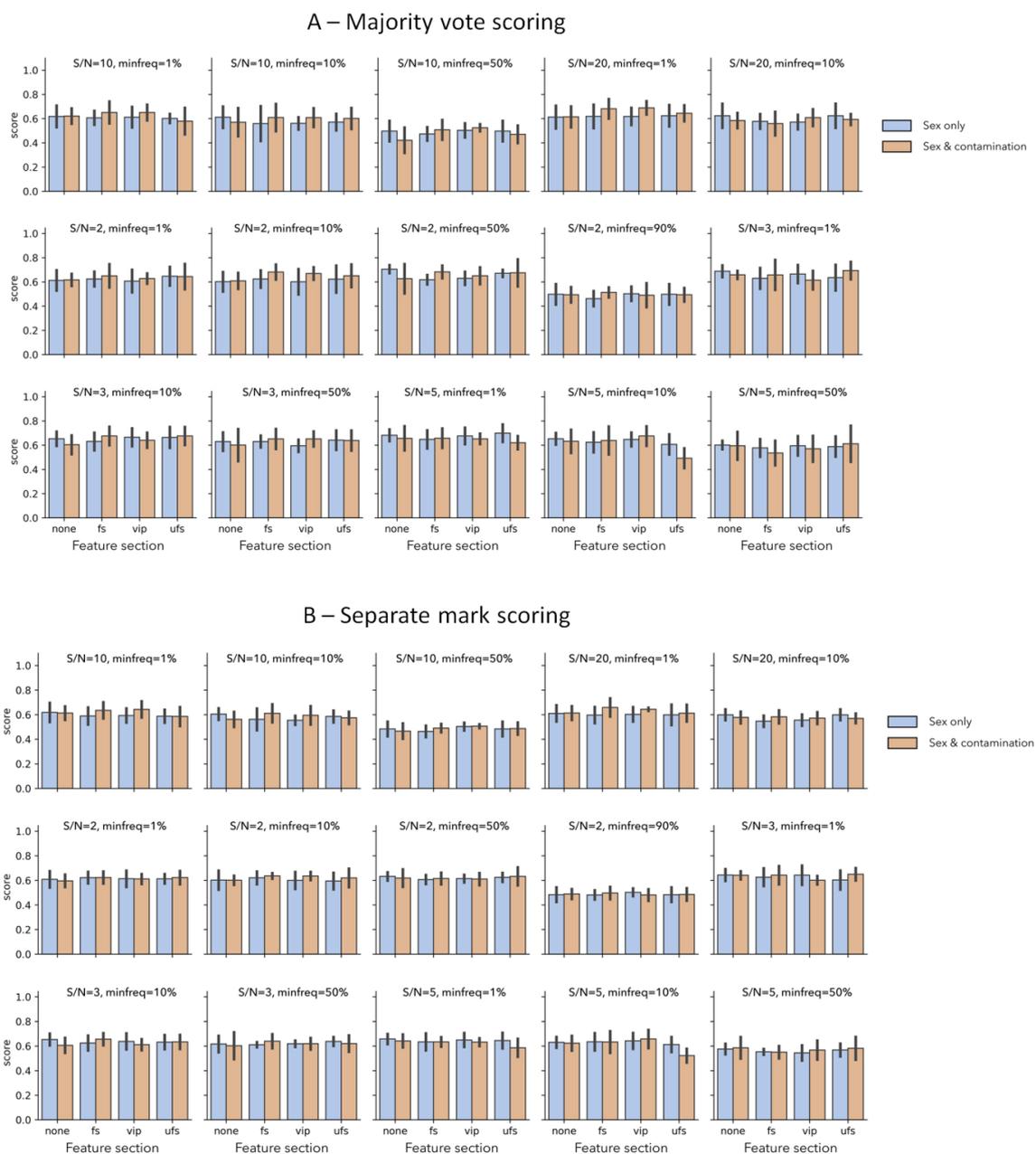
Figure 27. Majority-vote (a) and separate-sample scoring (b) performances reported for the XGBOOST classification scheme for different peak-peaking strategies (see subplot grids). In each subplot, the blue bars indicate the baseline XGBOOST model performances from 2-output classification and the orange bars indicate the performances of the modified strategy that incorporates explicit separation of the contaminant and non-contaminant samples. Error bars indicate 1 standard deviation of the k accuracy scores around the mean for each k-fold cross validation. The displayed data are the results from k=5 cross validation; other k-fold results were qualitatively similar.
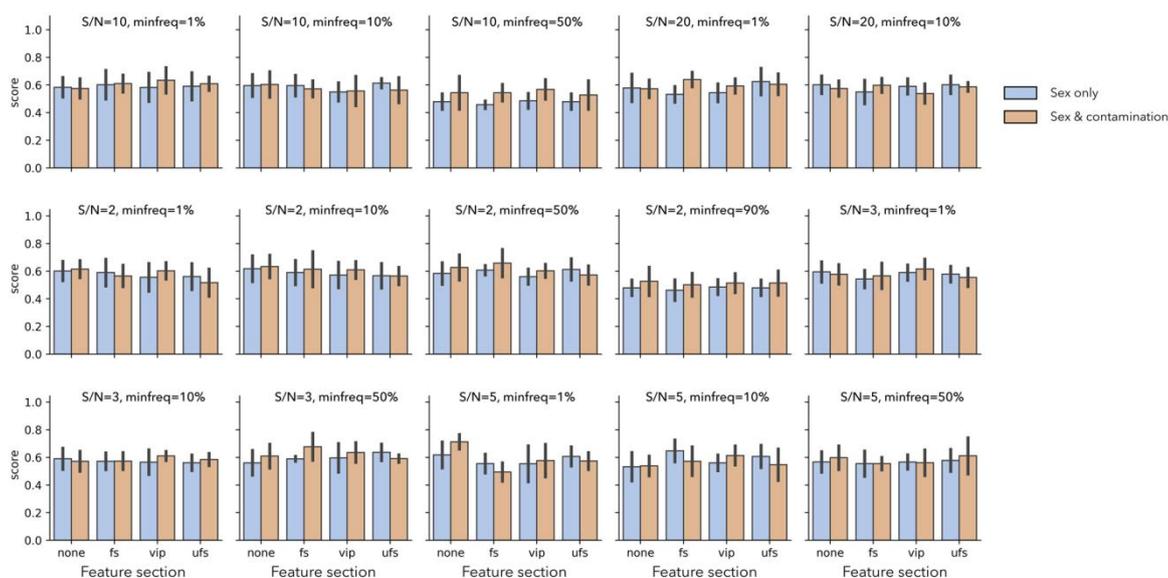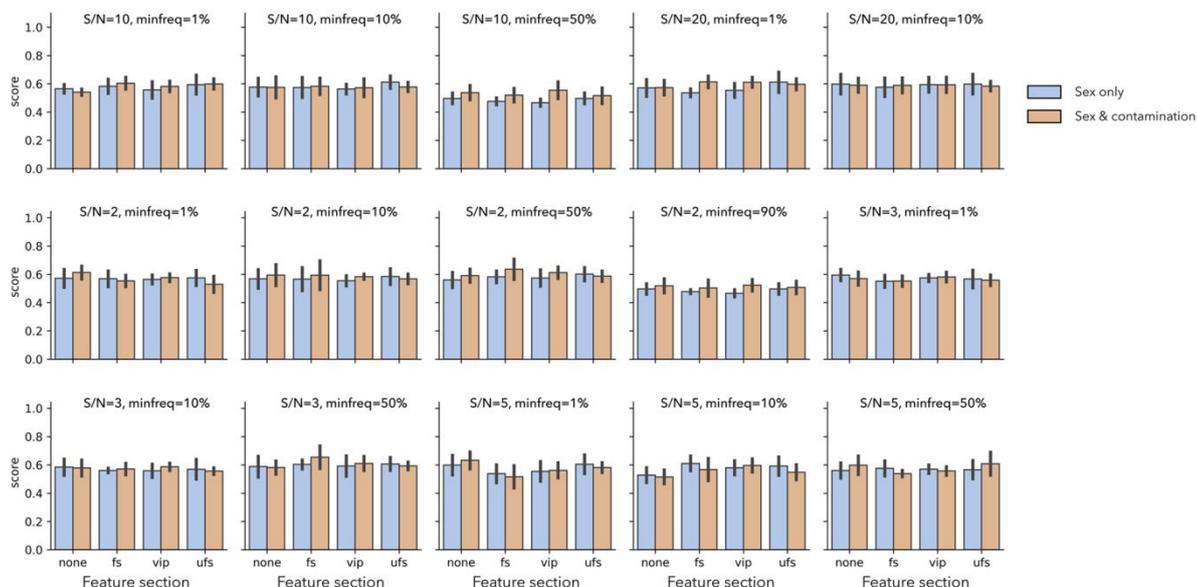
Figure 28. Majority-vote (a) and separate-sample scoring (b) performances reported for the LDA classification scheme for different peak-peaking strategies (see subplot grids). In each subplot, the blue bars indicate the baseline LDA model performances from 2-output classification and the orange bars indicate the performances of the modified strategy that incorporates explicit separation of the polymer-contaminated and non-contaminated samples. Error bars indicate 1 standard deviation of the k accuracy scores around the mean for each k-fold cross validation. The displayed data are the results from k=5 cross validation; other k-fold results were qualitatively similar.

Multi-class classification using a neural network architecture – As the presented

data shows, no significant improvement was evident in the performance of the

model's sex prediction capability, using either XGBOOST, random forest, LDA,

PLS-DA or naïve Bayes modelling. One limitation of these classification schemes is the fact that they rely on the output class labels being disjoint (i.e. having no elements in common) (as show in **Figure 20**), which often hinders the ability to train these models sufficiently in instances where one class may be underrepresented.

(a) 2 output states   (b) 4 output states
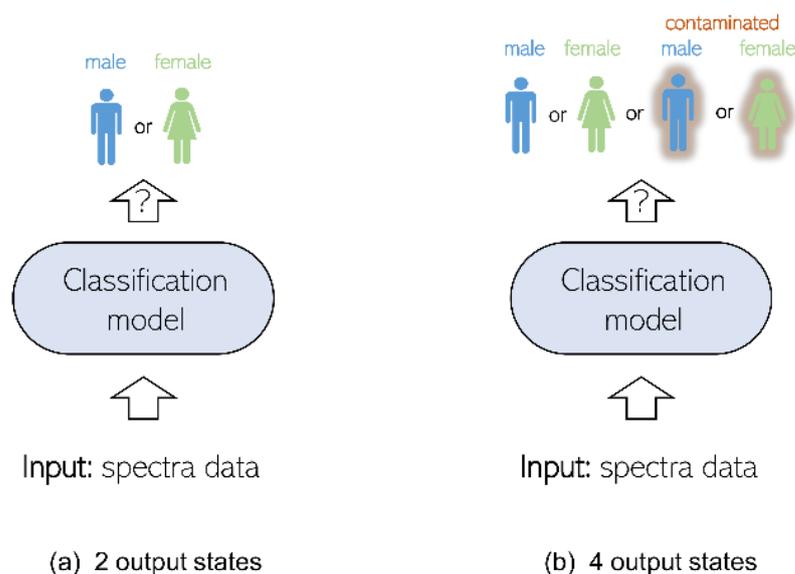
Figure 29. Classification models employed for the non- enhanced dataset. (a) Previous 2-output state classification scheme, in which a classification model is trained to label input spectra data as either 'male' or 'female'. (b) Modified 4-output state classification scheme, in which the classifier now must label each input spectra by both sex and contamination state.

For example, if the number of male contaminated samples was much lower than the number of female contaminated samples, or if two distinct classes are actually very similar to each other, without many discriminating factors. These potential issues may be mitigated by the application of a neural network-type architecture (**Figure 21**), which can be trained to make predictions against multiple, non-disjoint properties simultaneously. It should be noted that **Figure 30a** is analogous to **Figure 29a**, but that the 4 output states in **Figure 29b** are not identical to **Figure 30b**. **Figure 30** shows a direct comparison of neural

network model performances for the model architectures presented in **Figure 29a** (in blue) and **Figure 29b** (in orange).

(a) (b)
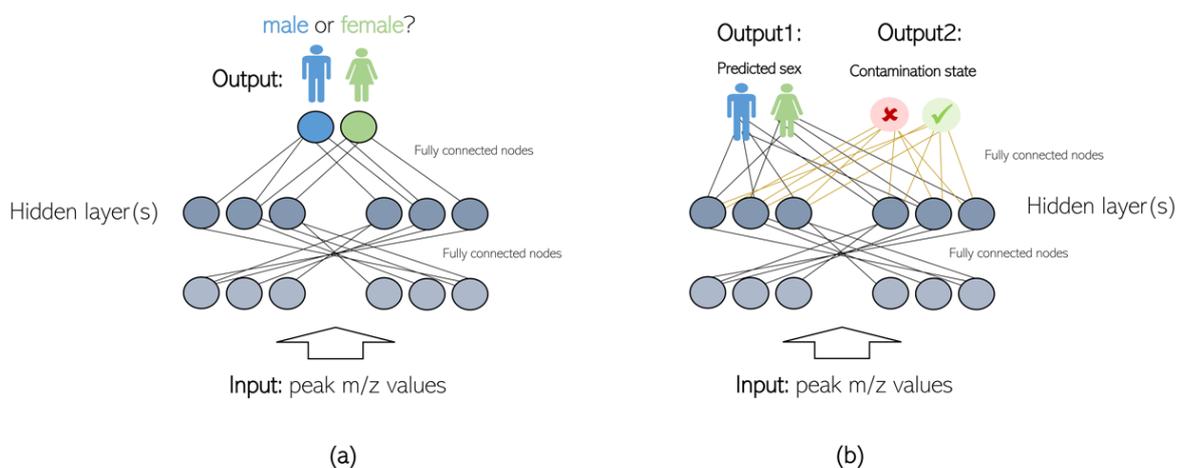
Figure 30. Analysis of unenhanced fingerprint dataset via neural network architecture. (a) Previous 2-output state classification scheme (using a fully-connected neural network architecture), with an output layer consisting of two nodes with softmax activation for male/female sex classification. (b) Multi-class output neural network architecture with an output layer consisting of two pairs of two nodes, with each pair having an independent softmax normalisation, such that each input sample is simultaneously classified by sex and contamination level separately. *[From Heaton et al., 2020]*.

The data shown in **Figure 31** are the result of k=5 cross validation; other k-fold results were qualitatively similar. It is not evident that a neural network architecture approach that takes into account sample contamination performs any better than the model schemes presented in **Table 8**.
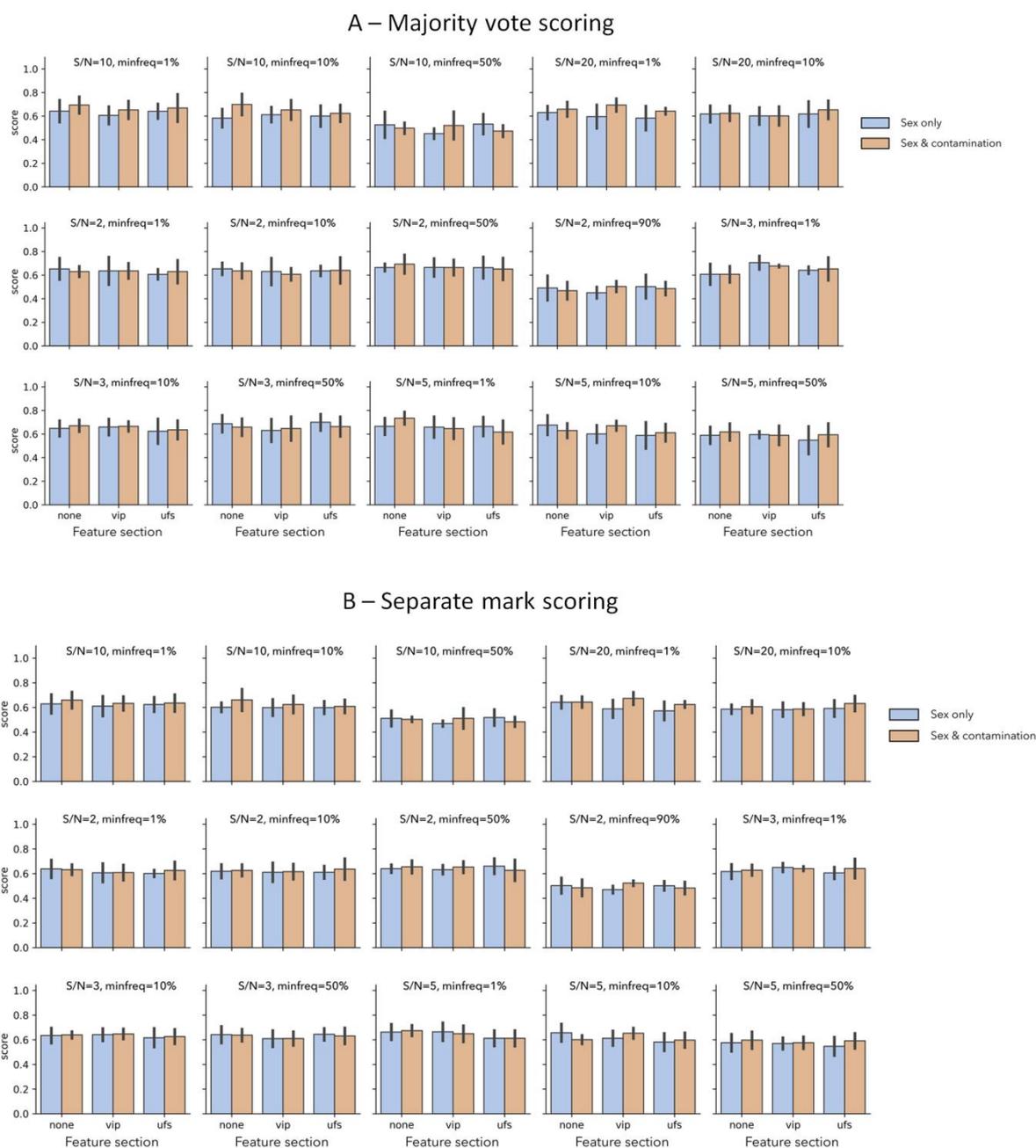
Figure 31. Majority-vote (a) and separate-sample scoring (b) performances for the neural network classification scheme for different peak-peaking strategies. In each subplot, the blue bars indicate the baseline model performances when contamination state is ignored and the orange bars indicate the performances of the modified strategy that incorporates explicit separation of the contaminant and non-contaminant samples. Error bars indicate 1 standard deviation of the k accuracy scores around the mean for each k-fold cross validation. The displayed data are the results from k=5 cross validation; other k-fold results were qualitatively similar and are not shown *[From Heaton et al., 2020]*.

Enhanced natural marks – the statistical modelling processes applied to the

non-enhanced set of fingermarks was then applied to the complementary set of

marks that had been enhanced with white FM powder. However, this was not including the 4 output classification and multi-class classification using the neural network architecture. Importantly, each model was trained using each of the 3 fingermarks as individual samples (516 samples in total). Each model was trained and run using k-fold cross validation, with variable k values. The range of peak picking parameters summarised in **Table 3** were trialled, varying the S/N ratio threshold and minimum number of peaks required, using the MALDIquant R package. Interestingly, the four donor's marks which had exhibited the severe polymer contamination in the non-enhanced samples did not show the same overwhelming contamination in the enhanced set of marks. A possible explanation for this could be that as the non-enhanced marks were the first set that the donors deposited with one hand, it could be speculated that donors might use their dominant hand for the first set of marks, and this would likely be the hand they would apply cosmetics and toiletries with due to the better dexterity. Another possibility may be that the presence of the fine enhancement powder supressed the polymer signals in the spectra, allowing the protein and peptide signals to be detected. All of the samples from this dataset were subjected to the same statistical modelling as their non-enhanced counterparts. **Figure 32** appears to demonstrate that the enhancement of the fingermarks with this particular white powder did not generally affect the performance of the models ability to predict the sex of the donor, when trialling a range of peak picking parameters and model types.
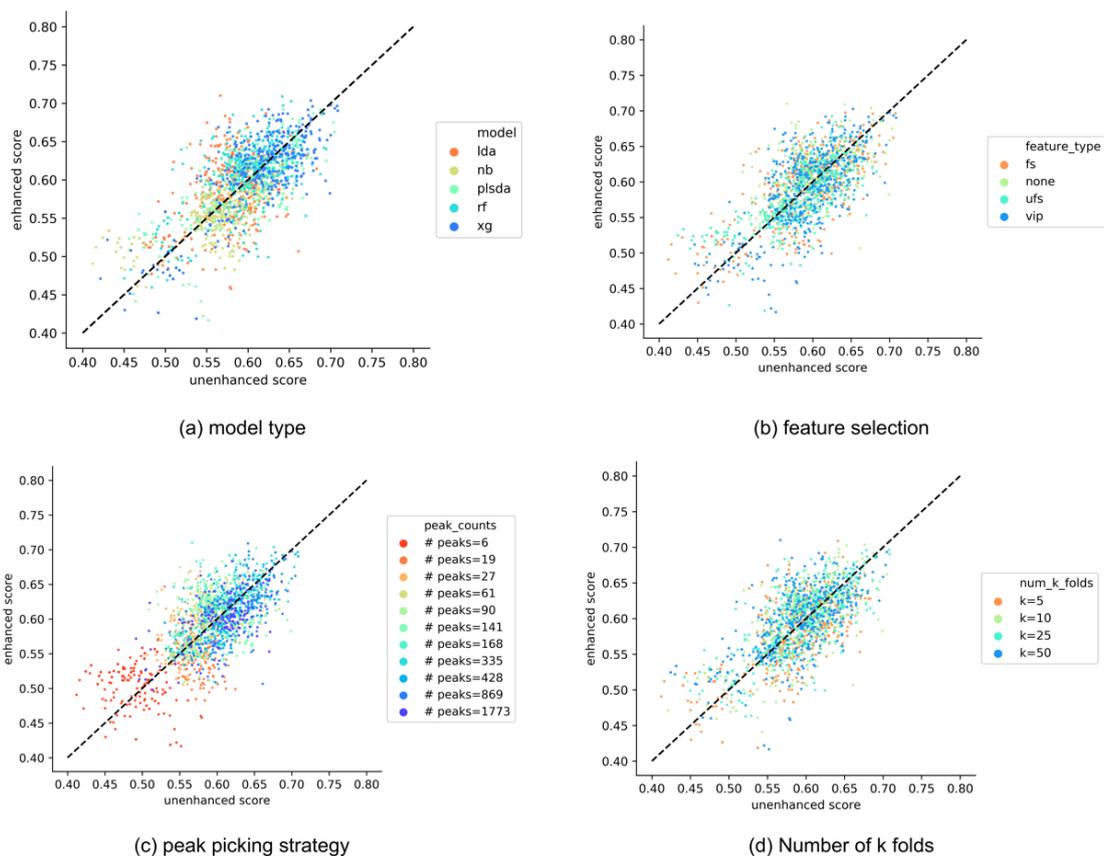
Figure 32. Scatter plots indicating the strong positive correlation between the mean cross validation scores derived from models trained on unenhanced samples (x-axes), and the corresponding mean cross validation scores for models instead trained on enhanced samples (y-axes). Each dashed black line indicates the diagonal y=x function. In each subplot, the scatter point distribution is identical: each scatter point is the mean cross validation score calculated over the k-fold repeats for a particular {model, feature selection, peak picking strategy} combination. However in each subplot the data has been coloured by a different parameter: (a) model type, (b) feature selection strategy, (c) peak picking strategy, (d) number of k folds. *[From Heaton et al., 2020]*.

Ferguson *et al.* (**Ferguson *et al.,* 2012**) acquired 9 spectra per fingermark compared to the 3 in this study, and those 9 technical replicates were treated as individual samples, as opposed to averaging the 3 fingermarks acquired in this study. The prediction accuracy achieved by Ferguson *et al.* was 85% in relation to the cross validation value, using a majority vote model. However, many of the assignments disagreed with each other (i.e. a single donor had samples assigned both male and female by the model). Consequently, if each individual spectrum was treated as a discrete dataset, the models prediction accuracy

dropped to 68.9% for the female samples and 74.4% for the male samples. As such, the 70.9% prediction accuracy reported for the current study is similar to the average of 71.7% reported by Ferguson *et al.* (**Ferguson *et al.*, 2012**), despite the changes to the current method outlined above to use natural fingermarks that are more comparable to 'operational crime scene marks'. However, the reported results are not directly comparable due to the variables between the studies' sample collection, pre-processing, and statistical processing stages.

In addition, the use of more advanced MALDI instrumentation with significantly improved sensitivity and speed of acquisition whilst using the Bruker rapifleX vs. the Bruker autofleX (Bruker Daltonik, Bremen, Germany) allowed for the acquisition of spectra at least 2 orders of magnitudes more intense, containing a greater number of protein and peptide signals. A contributing factor to this increased signal intensity was the laser repetition rate of 10,000 Hz on the rapifleX instrument, enabling the accumulation of 30,000 shots per spectra in this study compared with just 300 shots in the previous Ferguson *et al.* study with the autofleX instrument (**Ferguson *et al.*, 2012**). This allowed for a better S/N ratio (**Figure 21**), and as a result, it was hypothesised that the newer instrument would detect a greater number of signals useful for sex prediction. This was not found to be the case, although the increased sensitivity may have aided with minimising the effects of intra-sample variability encountered with natural marks. The model did not allow for a statistical improvement on the prediction of sex compared to previous work. However, it does more closely replicate the sort of marks encountered operationally, using natural marks as samples and trialling the effectiveness of the method on enhanced marks, than any previous work.

The technical improvement of the instrumentation used presents a double-edged sword; the increased sensitivity allows for the detection of molecules in far smaller concentrations than before, but also amplifies the signals of contaminants. In this study, polymer contamination, likely originating from the application of make-up/toiletries, were present in the majority of spectra, in some cases completely obscuring all useful protein and peptide signals. It was hypothesised that the prevalence of polymer contamination within the marks was a major factor in the models low predictive power. Around 40-45% of the polymer-contaminated species were given an incorrect sex assignment by model. This value was calculated for the XGBOOST model (k=5 folds), using the median values of the incorrectly assigned contaminated sample rates for each of the varying S/N values for the different peak picking parameters. Within the non-contaminated subset, the median percentage of those incorrectly classified was also around 40-45%, depending on the pick picking parameters, suggesting that the issue cannot solely be attributed to the presence of the polymer within the spectra. Even if machine learning algorithms could be written to exclude polymer signals from the classification procedure, the model cannot correctly classify the samples if the relevant signals are too low or too few in number, or relevant for discrimination.

A third scoring strategy, *full consensus scoring*, was also attempted in addition to the separate mark scoring and majority scoring strategies. This strategy required all sex predictions made by the model, for a single donor, to agree for the assignment to be considered; i.e. for a single donor, all predicted assignments must be in agreement - either 'male' or 'female' - for the result to be incorporated. This approach allowed for an increase in the predictive power of the strategy to 86.1% (using XGBOOST) (**Figure 33a, Table 5**). However, it

lead to around 50% of the donors samples being discarded from the dataset (for a range of peak picking parameters and feature selection strategies), for each k-fold cross validation strategy (**Figure 33b**). The other model types also displayed similar trends, as shown in **Figure 33a**, but with lower maximum full consensus scores (**Table 5**).
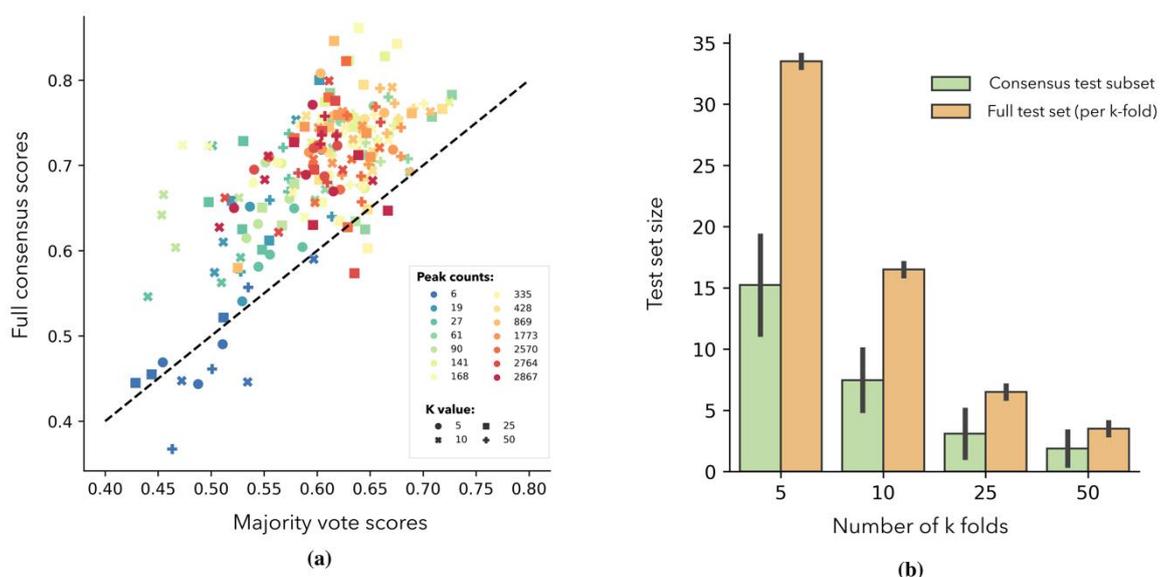
Figure 33. Application of 'full consensus scoring' strategy. (a) Correlation between the previously calculated majority vote scores (x-axis) and full consensus scoring schemes (y-axis), for the example case of the XGBOOST classifier. Each scatter point corresponds to the average cross-validation accuracy score for a specific peak picking strategy (see colour legend), and k-fold (see legend), with each of the four feature selection strategies in **Table 9** treated as separate scatter points. The diagonal line y=x is shown. (b) For each k-fold k value (x-axis), the total number of test set individuals included within the full consensus scoring scheme is indicated (green bars) and compared to the corresponding recorded full test sizes per individual k-fold (orange bars). Error bars illustrate 1 standard deviation over each m/z peak-picking strategy (**Table 7)** and the four distinct feature selection strategies (**Table 9**).

Model	Peak picking strategy		Number of k-folds	Feature selection strategy	Scoring scheme (mean calculated over k-fold repeats)		
	Minimum Frequency	Minimum Signal-to-noise			Full consensus	Majority vote	Separate mark
XGBOOST	minfreq = 0.01	S/N = 20	25	PLS-DA VIP score	0.861	0.639	0.628
XGBOOST	minfreq = 0.01	S/N = 5	25	PLS-DA VIP score	0.846	0.616	0.638
XGBOOST	minfreq = 0.01	S/N = 10	25	Random forest feature importance	0.843	0.675	0.666
Random forest	minfreq = 0.01	S/N = 20	25	Random forest feature importance	0.839	0.631	0.622
Random forest	minfreq = 0.01	S/N = 10	25	Univariate feature selection	0.835	0.621	0.628
Random forest	minfreq = 0.01	S/N = 20	25	PLS-DA VIP score	0.834	0.616	0.627
XGBOOST	minfreq = 0.5	S/N = 2	25	PLS-DA VIP score	0.828	0.664	0.622
XGBOOST	minfreq = 0.5	S/N = 2	25	None (use all peak-picked peaks)	0.823	0.630	0.598
XGBOOST	minfreq = 0.01	S/N = 3	25	None (use all peak-picked peaks)	0.822	0.627	0.615
Random forest	minfreq = 0.01	S/N = 5	50	Univariate feature selection	0.817	0.673	0.652
LDA	minfreq = 0.5	S/N = 3	25	None (use all peak-picked peaks)	0.813	0.561	0.584
Random forest	minfreq = 0.01	S/N = 5	10	Random forest feature importance	0.810	0.611	0.594
PLSDA	minfreq = 0.5	S/N = 3	25	PLS-DA VIP score	0.809	0.593	0.596
XGBOOST	minfreq = 0.01	S/N = 5	5	Random forest feature importance	0.808	0.603	0.633
LDA	minfreq = 0.01	S/N = 20	10	Univariate feature selection	0.804	0.565	0.564
XGBOOST	minfreq = 0.1	S/N = 2	25	None (use all peak-picked peaks)	0.800	0.602	0.579
XGBOOST	minfreq = 0.1	S/N = 20	25	Univariate feature selection	0.800	0.602	0.579
PLSDA	minfreq = 0.01	S/N = 20	25	PLS-DA VIP score	0.800	0.623	0.645
Random forest	minfreq = 0.01	S/N = 5	25	PLS-DA VIP score	0.799	0.660	0.632
XGBOOST	minfreq = 0.1	S/N = 2	10	None (use all peak-picked peaks)	0.799	0.611	0.600

Table 11. Top 20 scoring models based on the full consensus scoring scheme. Here, average scores have been calculated over k-fold repeats, with k specified in the 'Number of k-folds' column.

Whilst having to discard around half of the samples encountered is undesirable, it would still be beneficial to have developed a method that could successfully predict the sex of those samples exhibiting sufficient relevant information within their spectra. Normally during crime scene investigation, smudged, partial or overlapping marks may be discovered but discarded due to them not providing sufficient ridge detail to attempt a match for identification purposes. It is rare to encounter a perfect 'typical' fingermark in an operational context. However, molecular analysis of the marks chemical constituents does not require ridge detail, therefore the proteomic method proposed in this study could be attempted on these marks which may previously have been discarded for identification, opening new avenues for potential evidential information.

This study makes use of three marks per donor, but the number encountered at crime scenes may vary greatly, and also may be as low as one recoverable

mark. Unless arranged linearly, it would be problematic to identify if multiple marks had even come from the same donor. Indeed, Lauzon & Chaurand and Gorka *et al.* both propose methods for linking unidentifiable marks found at the same crime scene to the same individual, using MALDI MS and Ag LDI MS respectively (**Lauzon & Chaurand, 2018, Gorka *et al.,* 2019**). In the absence of these techniques, a single crime scene mark would be insufficient for the majority vote or full consensus scoring systems. Furthermore, the statistical performance of the models when based on three samples cannot be assumed to transfer to scenarios where fewer marks are available.

To determine how the technique would perform with access to only a single mark, instead of three per donor, a further k-fold cross validation was run for each model type, but with 1 mark randomly selected from the 3 marks deposited by each donor. Model training was performed as previously described, using all marks. As shown in **Figure 34**, there is a strong positive correlation between the mean cross validation scores when one mark per donor is used and the three different scoring strategies that utilised 3 samples per individual. This demonstrates the capability of the method when only one fingermark is available. Furthermore, using only one mark per test set individual allowed for a classification accuracy of 70.9%, equal to that obtained from the test sets with three marks per donor (when using XGBOOST, k=25 cross validation, peak picking parameters of S/N > 3 and minimum peak frequency = 50%). These results indicate that the model can function as effectively when presented with only 1 mark per donor to work with, but might suggest that the benefits attributed to utilising 3 marks per donor is only apparent when the full consensus strategy is employed.
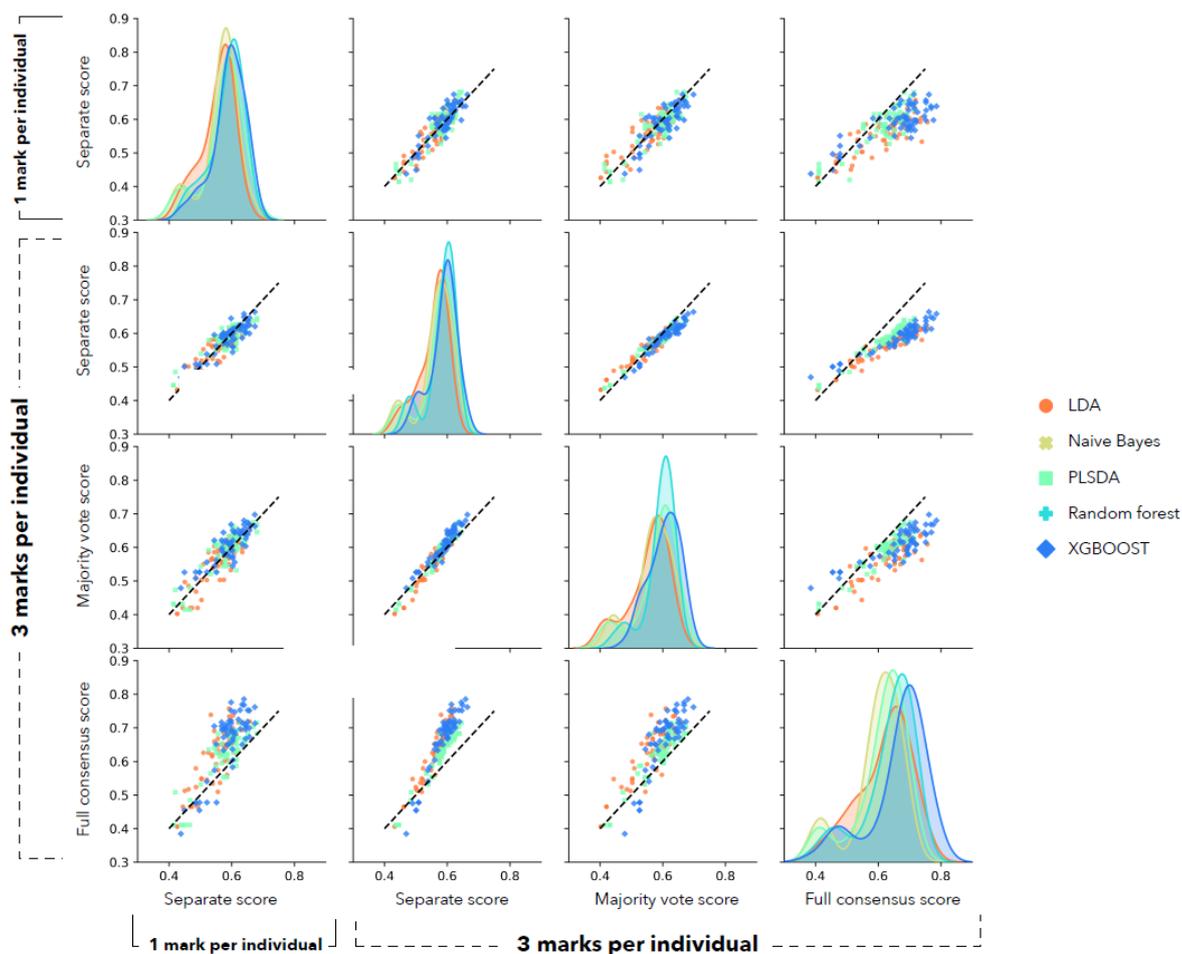
Figure 34. Scatter plots indicating (for k=5 cross validation performance assessment) a strong positive correlation between the mean cross validation scores derived from test sets comprised of a random 1/3 fingerprint marks per individual (top row, simulating the practical scenario in which only one fingerprint is available for testing), and the other 3 presented scoring schemes which are all derived using 3 marks per individual (the separate mark, majority vote, and full consensus scoring schemes). In each scatter subplot, each scatter point is the mean cross validation score calculated over the k-fold repeats for a particular {model, feature selection, peak picking strategy} combination. The main diagonal illustrates the distribution of each performance metric values across the parameter space of feature selection and peak picking strategies for a model type, here using a Gaussian kernel density estimate (KDE) to permit efficient comparison between the 5 model types.

Conclusion

The present study provides an analytical method for sex classification using samples most closely resembling the type of fingermarks that may be encountered at a crime scene. The use of natural fingermarks (not employing any artificial cleaning or enrichment of the fingertips) and the trialling of

compatibility of the method with one commonly used fingermark enhancement powder advances the present approach towards utilisation in operational investigations. Although previous studies have used groomed and ungroomed marks as a proof of concept to artificially remove contaminants and enrich the quantity of substance available for analysis, these marks obtained under 'laboratory conditions' are potentially unlike those which CSIs may encounter. However, it should be noted that 'natural' marks can vary greatly from a compositional point of view, and that marks encountered at crime scenes may in fact have originated from individuals who had recently washed their hands, or had been touching other parts of their body and enriching their fingertips with sebaceous material, therefore effectively mimicking ungroomed and groomed marks, respectively. Natural marks were used in this study to determine if the proposed proteomic method was robust enough to incorporate marks where the hands had not been cleaned or artificially enriched with sebaceous material, and more closely replicate the majority of marks that would be encountered operationally, but this decision led to the detection of polymer signals in a significant number of the samples analysed. Whilst this finding makes sex determination more difficult, it is more analogous to what 'real life' marks would contain. Unless a method of either boosting the protein and peptide signals or supressing the polymer signals was utilised, many of these marks would not be suitable to attempt sex determination.

Indeed, the lack of fundamental improvement in the predictive accuracy scores of the various models has been attributed to PEG-based polymer contamination obscuring the signals relevant for discriminating between the sexes in almost half of the samples analysed. In an operational context, such samples would have to be discarded from the sex determination analysis, due to suppression of

the protein and peptide signals required for determination of sex, but could offer other molecular information pertaining to the identity of an individual in other ways. For example, detection of a specific polymer present in the spectra could lead to the identification of the type of toiletry product that it appears in. Whilst the argument can be made that almost any toiletry or cosmetic product is not used exclusively by either one of the sexes, unique spectral profiles of recovered marks could be matched against the spectral profiles of toiletries used by a suspect to see if they contained the same types of compounds, in a similar approach to that used previously by the group (**Bradshaw *et al.,* 2011, Bradshaw *et al.,* 2013**).

Comprehensive statistical modelling was used in order to trial sex classification while changing several variables such as the S/N threshold, peak picking parameters, and the number of samples used, in addition to various statistical models. The full consensus scoring strategy is analogous to the approach previously used in the proof of concept study by Ferguson *et al.* (**Ferguson *et al.,* 2012**). Given the best prediction accuracy score of 86.1%, (while comparable to the previous proof of concept study, with its exclusion criteria) this would not allow definitive identification of a suspect's sex through analysis of their fingermarks. However, this research does represent a protocol to approach further work, addressing advances in experiment design including sampling procedure, data acquisition, processing and statistical modelling. As an extension of previous research by the Fingermark Research Group and others, this work crucially investigated the feasibility of using a single mark for sex classification, which had not been attempted before.

Future work

To propose for this technique to be used operationally would require more research. This would include generating 'cleaner' spectra to present the models with, either through amplifying the signals of the proteins and peptides required for sex discrimination, or supressing the intensity of the polymer signals within the spectra. This study also only assessed the compatibility of the technique with one fingermark enhancement powder, but more would be needed to make this technique operationally viable. In addition, this research used the same deposition surface for all the marks analysed; aluminium slides as they are conductive and also flat enough to fit inside a source chamber. Further deposition surfaces would need to be investigated, in addition to fingermark lifts (with tape or lifters) for marks deposited on surfaces not suitable for applying mass spectrometry to *in situ*.

Code availability

Python and R scripts written for all sex classification model training and validation are available in Appendix 2.

Acknowledgements

sample collection. Charles Bury is gratefully acknowledged for his contribution to this research performing all of the statistical modelling and generating all of the R and Python scripts discussed in the present study.

This research was the subject of a publication in Forensic Chemistry, entitled 'Investigating sex determination through MALDI MS analysis of peptides and proteins in natural fingermarks through comprehensive statistical modelling' which in addition to the research discussed above included previous research with data acquisition carried out by the Fingermark Research Group, by Dr. Robert Bradshaw and Dr. Ekta Patel (**Heaton *et al.,* 2020**)**.**

References

Asano, K.G., Bayne, C.K. Horsman, K.M. and Buchanan, M.V., (2002) Chemical Composition of Fingerprints for Gender Determination, *J. Forensic Sci.,* **46**, 1-3

Bandey, H.L., Bleay, S.M., Bowman, V.J., Downham, R.P., and V. G. Sears, V.G., (2014) *Fingermark Visualisation Manual,* Home Office, 1-932

Bardin, E., Claude, E. and Takatz, Z., (2018) Analysis of Fingerprints by Desorption Electrospray Ionization Mass Spectrometry Imaging, *Waters Application Note,* 1-3

Bishop, M., (1995) Neural Networks for Pattern Recognition, *Advanced Texts in Econometrics,* 1st Ed.

Bleay. S. M., End User Commentary on Mass Spectrometry Methods for the Recovery of Forensic Intelligence from Fingermarks *IN* Francese. (2019). Emerging technologies for the analysis of forensic traces (Francese, Ed.). Springer. 29-31.

Bradshaw, R., Bleay, S., Wolstenholme, R., Clench. M.R., and Francese, S., (2013) Towards the integration of matrix assisted laser desorption ionisation mass spectrometry imaging into the current fingermark examination workflow, *Forensic Sci. Int.,* **232**, 1-3, 111-24

Bradshaw, R., Wolstenholme, R., Ferguson, L.S., Sammon, C., Mader, K., Claude, E., Blackledge, R., Clench, M.R., Francese, S., (2013) Spectroscopic imaging based approach for condom identification in condom contaminated fingermarks, *Analyst*, **138**, 2546-2557

Bradshaw, R., Wolstenholme, R., Blackledge, R. D., Clench, M. R., Ferguson, L. S. and Francese, S. (2011). A novel matrix-assisted laser desorption/ionisation mass spectrometry imaging based methodology for the identification of sexual assault suspects. *Rapid Communications in Mass Spectrometry*, **25** (3), 415-422

Croxton, R. S., Baron, M. G., Butler, D., Kent, T., & Sears, V. G. (2010). Variation in amino acid and lipid composition of latent fingerprints. Forensic Science International, 199(1), 93–102

Emerson, B., Gidden, J., Lay, Jr. J.O. and Durham, B., (2011) Laser Desorption Ionization Time-of-Flight Mass Spectrometry of Triacylglycerols and Other Components in Fingermark Samples, *J. Forensic Sci.,* **56** 381-389

Ferguson, L.S., Wulfert, F., Wolstenholme, R., Fonville, J.M., Clench, M.R., Carolan, V.A., and Francese, S., (2012) Direct detection of peptides and small proteins in fingermarks and determination of sex by MALDI mass spectrometry profiling, *The Analyst,* **137**, 4686-4692

Francese, S., Bradshaw, R., Ferguson, L., Wolstenholme, R., Clench, M.R., and Bleay, S., (2013) Beyond the ridge pattern: multi-informative analysis of latent fingermarks by MALDI mass spectrometry, *The Analyst,* **138**, 4215-4228

Francese, S., Bradshaw, R., & Denison, N. (2017). An update on MALDI mass spectrometry based technology for the analysis of fingermarks - stepping into operational deployment. Analyst (London), 142(14), 2518–2546.Gorka, M., Augsburger, M., Thomas, A., and Bécue, A., (2019) Molecular composition of fingermarks: assessment of the intra-and inter-variability in a small group of donors using MALDI-MSI, *Forensic Chemistry,* **12**, 99-106

Heaton, C., Bury, C., Patel, E., Bradshaw, R., Wulfert, F., Heeren, R.M., Cole, L., Marchant, L., Denison, N., McColm, R., and Francese, S., (2020) Investigating sex determination through MALDI MS analysis of peptides and proteins in natural fingermarks through comprehensive statistical modelling, *Forensic Chemistry*, **20**, 100271

Huynh, C., Brunelle, E., Halámková, L., Agudelo, J., and Halámek, J., (2015) Forensic Identification of Gender from Fingerprints Anal. Chem., **87**, 11531-11536

Knowles, A.M., (1978) Aspects of physicochemical methods for the detection of latent fingerprints, *J. Phys. E: Sci. Instrum.,* **11**, 713–721

Lauzon, N., Chaurand, P., (2018) Detection of exogenous substances in latent fingermarks by silver-assisted LDI imaging MS: perspectives in forensic sciences, *Analyst* , **143**, 3586-3594

Lee, J., and Joullié, M., (2015) Fine-tuning latent fingerprint detection on paper using 1,2-indanedione bi-functional reagents, *Tetrahedron,* **71**, 7620-7629

Mark, H. & Harding, C. R. (2013). Amino acid composition, including key derivatives of eccrine sweat: potential biomarkers of certain atopic skin conditions. International Journal of Cosmetic Science, 35(2), 163–168.

McDonnell, L.A., Heeren, R.M.A., de Lange, R.P.J., and Fletcher, I., (2006) Higher sensitivity secondary ion mass spectrometry of biological molecules for high resolution, chemically specific imaging, *J. Am. Soc. Mass. Spectrom.* **17**, 1195-1202

Mekkaoui Alaou, I., and Halamek, J., (2019) Fluorescence of 1,2-Indanedione with Amino Acids Present in the Fingerprint Residue: Application in Gender Determination, *J. Forensic Sci.,* **64**, 1495-1499

Oonk, S., Schuurmans, T., Pabst, M., de Smet, L.C.P.M., and de Puit, M., (2018) Proteomics as a new tool to study fingermark ageing in forensics, *Scientific Reports,* **8**, 2045-2322

Strohalm, M., Hassman, M., Košata, B., and Kodíček, M., (2008) mMass data miner: an open source alternative for mass spectrometric data analysis, Rapid Commun Mass Spec, **22** (6), 905-908

Wolstenholme, R., Bradshaw, R., Clench, M.R., and Francese, S., (2009) Study of latent fingermarks by matrix-assisted laser desorption/ionisation mass spectrometry imaging of endogenous lipids, *Rapid Commun. Mass Spectrom.,* **23**, 3031-3039

https://iq.opengenus.org/relu-activation/ [*last accessed: 13/03/22*]

Chapter 3: A tiered approach for the bottom-up proteomic detection of biofluids and species identification from blood using MALDI MS, MS/MS and MSI

Introduction

Blood is an extremely valuable type of evidence when encountered at the scene of violent crimes. Its location, quantity, distribution and even composition can give an insight into the dynamics surrounding the bloodshed. The discipline of Blood Pattern Analysis (BPA) is solely dedicated to the reconstruction of events. Intelligence can be recovered from serology, BPA, DNA profiling, or chemical analysis of the molecular composition of the blood, including the detection of drugs. Even the physical state of the blood can be informative; wet drops/pools of blood must have been recently shed, whereas dried blood spots take a little time to form. Small speckles of blood, unrelated stains resembling blood or invisible traces are particularly challenging for crime scene investigators (CSIs) who often rely on rapid colorimetric tests to apply to large areas suspected to contain biofluids to locate latent stains, or perform quick swab tests of suspected biofluids. These are considered presumptive and subsequent confirmatory tests are required. Furthermore, not all blood stains may be relevant to the crime being investigated. For example, blood found in a kitchen may have originated from a butchered animal and not a human victim or suspect. Therefore specific, robust and sensitive tests are also needed to ascertain the source of blood. In 1996, in the first US criminal trial to use animal DNA, prosecutors presented two jackets and a pair of pants, worn by two suspects accused of a double murder of a couple and their dog, as evidence. The items of clothing bore bloodstains from the dog, and both suspects were convicted of the double murder (**www.cbsnews.com**). The company that carried out the analysis, PE AgGen, claimed there was a 1 in 350 million chance that the blood did not belong to that dog (**archive.seattletimes.com**). Confirming not only that a stain is blood, but the origin species of that blood

could be of critical importance in certain circumstances, such as that case study, or suspected animal abuse or wildlife crime cases.

There are a range of current blood enhancement techniques available to both CSIs to assist with the detection of latent bloodstains at crime scenes, and to forensic scientists in crime laboratories. Typical blood enhancement techniques (BETs) are rapid colorimetric tests, exploiting a chemical interaction resulting in a colour change to indicate the presence of component/s within the blood. Their selection depends on whether the evidence is a bloodstain or a blood mark and also on the surface where this biofluid is present. However, depending on their method of action, they may exhibit poor specificity and give rise to false a positive result, hence why these tests are called presumptive. A common group of enhancement reagents are acid dyes which interact or react with the protein content of a blood stain, but can elicit a positive response with other protein-containing substances such as semen and saliva (**Bradshaw *et al.,* 2013, Francese, 2019**). Haem-reactive enhancement reagents are more specific for blood, but can also have their colorimetric reactions catalysed by other substances, such as in the case of Luminol, which has been shown to false-positively react with bleach (**Francese, 2019**), plant peroxidases (**Lee & Pagliaro, 2000**), enamel paint (**Quickenden & Creamer, 2001**) furniture polishes and interior fabrics in motor vehicles (**Creamer *et al.,* 2003**) and some metal ions such as copper (**Ramotowski, 2013**). Furthermore, some BETs must be viewed under certain specialist lighting conditions, for example Luminol, which is best visualised in a dark environment, and hence not always practical to achieve in the field. Additionally, adding large quantities of reagents to suspected blood stains not only dilutes the available sample with chemicals, but also contaminates them with the constituent chemicals.

Another potential concern with the use of BETs is their compatibility with further sequential downstream analysis during the investigation. Haem-reactive presumptive tests, such as Luminol and BlueStar, as well as confirmatory tests such as the Takayama Test, have been demonstrated to degrade DNA up to 30 days after application (**de Almeida, 2011**). DNA is a valuable biometric evidence type, which can be used to identify (or exclude) a suspect through matching with a possible database DNA record or using samples obtained from potential offenders. Whilst considered the 'gold standard' of biometric identification when successful, DNA analysis does not reveal the biofluid or tissue type of the sample, which might be critical to the investigation (**Duong *et al.,* 2021**). DNA can also degrade depending on the local environmental conditions, yielding either a partial or even an insufficient profile to match against the database. Furthermore, mixed donor samples, may yield complex profiles which are not easily resolved.

In contrast, mRNA, is more resilient than DNA and can readily distinguish between biofluids (**Jakubowska *et al.,* 2013**). Other groups have reported the detection of specific biomarkers in other biofluids using messenger RNA (mRNA) -based techniques. mRNA has the advantage over DNA in that it can determine the biofluid of origin and is more persistent then DNA. Jakubowska *et al.* reported on the determination of vaginal fluid and menstrual blood by profiling 5 mRNA and bacterial markers, the detection of which have applications in the investigation of sexual crimes (**Jakubowska *et al.,* 2013**). Roeder & Haas report on the determination of saliva, cervicovaginal fluid and peripheral and menstrual blood through the detection of at least 5 mRNA markers per fluid. However, markers were not specific to each body fluid and required some numerical weighting to determine the fluid type (**Roeder & Haas,**

2013). Li *et al.* trialled a similar approach, this time investigating the use of microRNA (miRNA) markers to distinguish between menstrual and peripheral blood. They state that miRNA is less susceptible to environmental degradation that mRNA. However, this process is time consuming and the authors conclude that the technique might not be resistant to intra-individual variability, stage of the menstrual cycle, or the age of the individual (**Li *et al.,* 2017**). Although mRNA profiling is still considered somewhat of an emerging technique, Gannicliffe notes in summary that of all the recent advancements in body fluid identification, this is one of the most mature, as there are examples of the technique being incorporated into routine casework (**Gannicliffe, 2019**). However, the author notes the barriers to more widespread implementation, namely with the interpretation of data. Further validation would be required to establish operational robustness, particularly concerning cold cases where any collected evidence is especially valuable.

In terms of alternative techniques, spectroscopic methods have recently been employed as non-destructive alternatives for blood detection. Spectroscopic methods are not currently featured as Category A methods in the Fingermark Visualisation Manual, meaning they are not recommended for routine casework (although this may change following further research in the most recent edition of the Manual, currently in publication). Raman spectroscopy exploits light scattering phenomena to generate a measurable spectral shift. Here, a sample is irradiated with a specific wavelength of monochromatic light (usually a laser) which interacts with functional groups causing molecular vibrations within the molecule. This interaction changes the energy of photons either up or down, initiating scattering of light. This can take the form of Stokes Raman scattering if the wavelength of the scattered light is greater than the wavelength of the laser

157

($\lambda_{scatter} > \lambda_{laser}$), Anti-Stokes Raman scattering if the wavelength of the scattered light is less than the wavelength of the laser light ($\lambda_{scatter} < \lambda_{laser}$) or Rayleigh scattering if the wavelength of the scattered light is equal to the wavelength of the laser light ($\lambda_{scatter} = \lambda_{laser}$). The vast majority of scattering events are Rayleigh scattering, where the energy of the molecule is unchanged. One out of each ten million interactions results in the molecule gaining or losing energy and yielding Stokes or Anti-Stokes scattering respectively. Because the majority of molecules will be in the ground state, Stokes scattering is the more probable process, therefore Raman almost always measures Stokes scatter (**Nascimento, 2018**). Characteristic spectra can be produced and compared to known spectra to identify molecules. Raman has been used to irradiate blood at 752 (**Boyd *et al.,* 2011**) and 785 nm (**Lemler *et al.,* 2014**), which was reported to have detected haemoglobin in addition to other blood-derived proteins. However, it has been reported by Lemler *et al.* that denaturation of the blood can cause spectral shifts as a function of time since deposition (**Lemler *et al.,* 2014**).

Hyperspectral Imaging (HSI) uses visible light to measure the reflectance spectrum of a sample. Islam's group reported on the application of HSI for the detection and identification of blood stains (**Li *et al.,* 2014**). Blood absorbs in a narrow band in the region of 415 nm known as the Soret band, with a λmax of 412 nm (**Hanson & Ballantyne, 2010**). Similarly to Raman, spectra are compared against reference spectra. However, HSI is not free from limitations. A background reference spectrum must first be generated using an uncontaminated part of the surface, and at a crime scene, it would be difficult to guarantee that this was free from any latent blood. In addition, HSI is not effective on dark surfaces, as they absorb much of the incident light, so the

reflectance spectrum is susceptible to noise (**Li** *et al.,* **2014**). Furthermore, the Soret band shifts towards shorter wavelengths as the time since deposition increases. Nevertheless, Islam's group reported HSI to be effective on red substrates, which would usually present a challenge due to the lack of contrast against blood (**Li** *et al.,* **2014**). The group also reported on the ability of HSI to detect and identify blood in blood-contaminated fingermarks. They reported the non-contact, non-destructive identification of haemoglobin between 400-500 nm, and distinguished blood-contaminated marks from those deposited with other red/brown contaminants. Additionally, blood ridge detail was observed up to the 9th mark in a depletion series, in blood that had been diluted 1:20, and even in bloody marks up to 6 months post deposition. Furthermore, the group claimed to have detected latent blood stains with blood that had been diluted 1:15,000 (**Cadd** *et al.,* **2016a**). However, this work was all carried out on white tile, where blood would already have a good contrast against the background surface. The group therefore followed up this study by assessing the feasibility of the technique against darker tiles in addition to a range of porous and non-porous substrates from glass and plastic to wood and cotton, including pig skin to replicate human skin. On the darker tiles, ridge detail was only detected to the 2nd depletion marks although blood detection was possible to the 7th depletion, compared to ridge detail up to the 3rd depletion and blood detection on the 12th depletion on the light coloured tiles. The technique was also successful on the other substrates (**Cadd** *et al.,* **2016b**). More recently, the Islam group has attempted to address the challenging subject of time since deposition of blood-contaminated fingermarks, comparing the HSI spectra of 1 day and 30 day aged blood marks (**Cadd** *et al.,* **2018**). However, this was only demonstrated on one substrate, white tile, and required a false colour aging

scale, which would be specific to the deposition surface and so would have limited applicability to crime scene marks that can be found of a wide variety of surfaces, having undergone drastically varying environmental conditions. Due to the rapid and non-destructive nature of HSI, it offers a viable alternative for the detection of blood, compared to the application of separate, destructive presumptive and confirmatory testing. Although theoretically portable, there are no known reports in the literature regarding its use in the field. This might be due to the difficulties of ensuring that whilst taking background reference spectra, no latent blood is present, given that HSI can detect blood that has been diluted 15 fold and would therefore be invisible to the naked eye. However, the authors note the challenges of utilising HSI for forensic science casework; unlike the pure samples analysed under laboratory conditions, casework samples can be contaminated or exist as complicated mixtures. There are also a plethora of diverse surfaces that evidence may be encountered on which can complicate measurements. The many environmental variables cannot accurately be modelled without precise data, requiring extensive validation.

Additionally, Attenuated Total Reflection Fourier-Transform Infrared Spectroscopy (ATR FT-IR) has been proposed. FT-IR measures the absorbance of vibrational bonds in a compound over a range of wavelengths. Instead of using a monochromatic light source, as with Raman, Fourier Transform spectroscopy utilises a light source containing many frequencies simultaneously. Mistek and Lednev attempted identification of different species from blood samples using ATR FT-IR. They reported 100% prediction accuracy in distinguishing between human and animal blood, and that PLS-DA was able to demonstrate complete separation between human, canine, feline and chicken

blood (**Mistek & Lednev, 2015**). More recently, they demonstrated the same technique, in a study comprising of 12 mammalian species (**Mistek & Lednev, 2020**). However, this technique requires the scraping of dried blood from the substrate onto the ATR crystal for analysis, which although not destructive in that no sample is consumed, does necessitate the destruction of the originally enhanced sample, which may not always be desirable.

Despite spectroscopic techniques presenting viable approaches for some forensic evidence types, particularly in regard to the (on some levels) non-destructive nature and lack of contact required for analysis, their underpinning working principles mean they are not highly specific techniques as they only enable the detection of the functional groups within a molecule rather than the molecule itself (unless a spectral database is used for matching). Mass spectrometric techniques, on the other hand, measure an intrinsic property of the molecule, which is their molecular weight through their mass-to-charge data, and hence they are more specific than spectroscopic methods. Putative assignments can be confirmed through fragmentation of the parent ion through MS/MS analysis. The generated ion fragments can either be matched via large compound libraries, or manually sequenced to confirm the structural identity of the compound.

Despite the novel proof-of-concept, Liquid Chromatography-tandem Mass Spectrometry (LC-MS/MS) remains the gold standard for confirmatory MS/MS analysis, particularly for biological samples. Proteomic-based mass spectrometry approaches have been previously reported by Illiano *et al.,* among other groups. The authors reported the detection and identification of proteotypic peptides for blood, semen and saliva using LC-MS/MS (**Illiano *et al.,* 2018**). Furthermore, the Kirkbride group propose a toolbox of mass

spectrometry based techniques for forensic applications; in a first paper, they utilised LC-MS/MS for the detection and discrimination of vaginal fluid and blood from samples trapped under the fingernails, through the detection of proteotypic proteins including cornulin, cystatin, peroxiredoxin-1 and glutathione S-transferase P, which they report are characteristic of vaginal fluid. Furthermore, material lodged under the nails has been shown to persist even after hand washing, and proteins specific to vaginal fluid can be determined through analysis via HP LC-MS/MS (**Kamanna *et al.,* 2017**). For all of its advantages, such as high sensitivity and better fragmentation, MS/MS confirmation of analytes and high numbers of protonated species, (particularly for large biological molecules such as proteins), LC-MS/MS is a destructive technique, rendering (possibly limited) material unavailable for further processes. LC-MS and MS/MS also cannot run samples concurrently, and generates large quantities of data per run, the majority of which may not be forensically relevant for the investigation, although LC MS^e analysis can be performed in data-independent analysis mode to allow for simultaneous protein identification (**Souza *et al.,* 2016**).

Yang *et al.* first proposed the use of LC interfaced with Matrix Assisted Laser Desorption Ionisation (MALDI) Mass Spectrometry (MS) for the analysis of fluid-specific proteins for body fluid identification in 2013. The method enabled the detection of marker proteins for blood (including haemoglobin), semen (including semenogelin) and saliva (including α-amylase 1) down to sub-nanoliter quantities and was able to identify multiple biofluids in mixed samples. The group were able to detect haemoglobin in blood and confirm the presence of haemoglobin β peptides through *De Novo* sequencing (**Yang *et al.,* 2013**).

However, the sample preparation stage included separation by reversed phase liquid chromatography, resulting in a time-intensive and convoluted workflow.

Whilst Yang *et al.* were the first to report the use of LC MALDI for blood detection, its hyphenation with the chromatographic step made it a sub-optimal application. Seraglia *et al.* were the first (in the author's knowledge) to use MALDI MS in a forensic context, for the detection of blood. In 2004 the group utilised MALDI MS for the identification of blood on car carpet through the detection of intact α and β haemoglobins (**Seraglia *et al.,* 2004**).

In the previously referenced Kirkbride group paper, the group report the MALDI Mass Spectrometry Imaging (MSI) of fingermarks contaminated with blood and vaginal fluid, mapping haemoglobin and cornulin in the fluids, respectively. This evidence would have obvious implications in the investigations of sexual assault crimes, with the potential of identifying individuals through the biometric information yielded by the fingermark ridge details provided by the detection of chemical constituents within the mark itself (indicating substances that their fingertips might have come into contact with prior to the accidental mark deposition) (**Kamanna *et al.,* 2017**).

Following the earlier tandem LC MALDI work of Yang *et al.*, Bradshaw *et al.* were the first to report detection of blood in fingermarks and generation of blood specific molecular images of the associated ridge pattern in fresh and seven days old marks using MALDI (without the chromatographic separation step previously reported). Bradshaw *et al.* directly detected haem and haemoglobin molecules within fingermarks in 2014. The authors demonstrated that they were able to perform a confirmatory test for the presence of blood by detecting intact haem at nominal *m/z* 616, with an in source ion fragment at *m/z* 557, and intact

haemoglobin molecules (αHb at *m/z* 15127, βHb at *m/z* 15868) through MALDI MS profiling, and in a depletion series even when blood was no longer visible with the naked eye. Furthermore, the authors demonstrated that they were able to map these molecules within a fingermark contaminated with blood, using MALDI Mass Spectrometry Imaging (MSI). Haem was imaged with good ridge detail overlay in fresh marks, and was even possible, although with reduced clarity, in marks aged up to 7 days. Also, uniquely from any prior published work, the authors even provided molecular ion images of intact haemoglobin in marks that had been visualised with Acid Black 1 enhancement reagent, demonstrating the potential for compatibility within existing enhancement workflows (**Bradshaw *et al.,* 2014**).

Patel *et al.* reported a more specific method for blood detection in 2015. The authors employed a bottom-up proteomic approach for the detection of a number of blood-derived peptides, namely haemoglobin, myoglobin complement C3, apolipoprotein A-1, alpha-1-antitryspin, haemopexin, serotransferrin, EPB 4.2, EPB 3, and alpha 2-macroglobulin, in both *in solution* and *in situ* samples of stains and an aged palm print. This proteomic-based approach for the detection and identification of blood is more specific than the intact method reported by Bradshaw *et al.* due to the detection of multiple markers across a range of blood-derived peptides, and more accurate due to the lower molecular weights of the peptides in comparison to intact proteins. Whilst the *in solution* analysis provided confirmation of blood, the *in situ* approach was particularly novel allowing for the confirmation of blood stains or marks with only a 5 minute digest. However, the trade off against speed was the quantity of identifiable proteins detected. Nonetheless, the technique was successfully demonstrated in confirming the presence of blood on a 9-year-old

bloodstained handprint, which had previously been enhanced with Acid Black 1. This circumstance is promising for the pursuit of cold case investigations where blood is involved. In addition to the detection of blood, the authors were also able to differentiate blood from human, equine and bovine sources, through the analysis of heterogeneous haemoglobin peptide chains (**Patel *et al.,* 2016**).

Deininger *et al.* further advanced this preliminary research in 2016 by detecting blood and mapping blood-specific proteins within the ridge pattern of blood-contaminated fingermarks, combining the *in situ* digestion approach with MALDI MSI. The authors were able to detect multiple blood-derived proteins, including α haemoglobin, complement C3, haemopexin and serotransferrin, and produce molecular ion images, reconstructing the ridge detail of the blood-contaminated mark (**Fig. 35**). This technique illustrates benefits twofold; first it provides a confirmatory analytical test for the presence of blood. Second, there is the potential to link external contaminants (such as blood) to an individual through the physical biometric information available through conventional fingerprint identification methods. This could mean an individual was present at, or very soon after, a blood shed event, as dried blood would not transfer to the fingertips in this way (**Deininger *et al.,* 2016**).
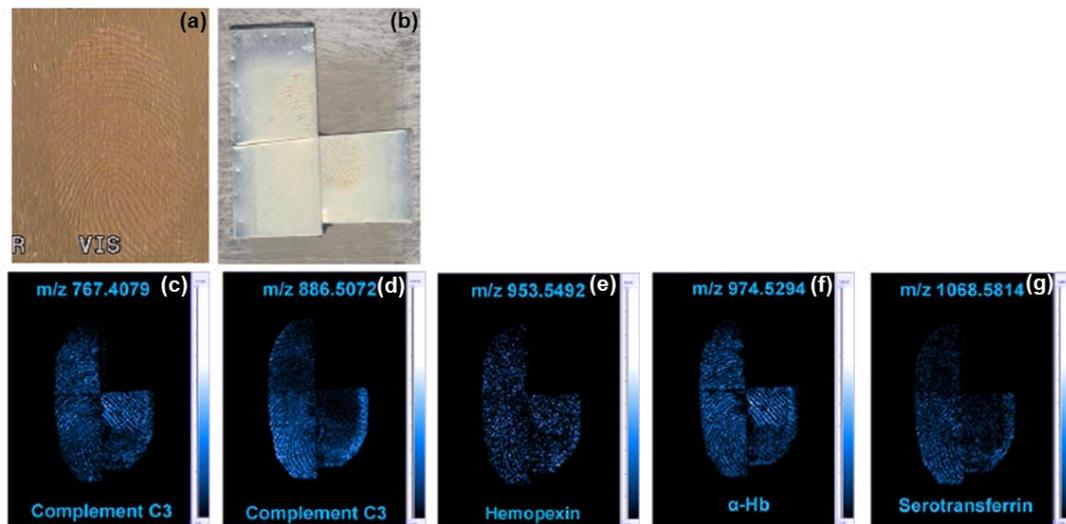
Figure 35. MALDI MSI images of blood-specific peptides detected in bloody fingermarks following the digesting protocol described in Deininger *et al.,* 2016. Optical image prior to trypsin application (a). Optical image post-trypsin application (b). Molecular ion images of signals at: *m/z* 767 attributed to Complement C3 (c), *m/z* 887; Complement C3 (d), *m/z* 954; Hemopexin (e), *m/z* 975; αHb (f) and *m/z* 1069; Serotransferrin. *[Adapted from Deininger et al., 2016].*

In further work reported in the author's thesis, Deininger presented a number of further tentative protein identifications following *in solution* digestion of human blood. The author reported the detection of blood specific peptides deriving from the proteins haemoglobin (α and β chains), haemopexin, serotransferrin, complement C3, alpha-1 antitrypsin, apolipoprotein A1, alpha-2-macroglobulin, erythrocyte membrane protein band (EPB) 3 and 4.2, reporting that some of the EPB 4.2 peptides detected are specific to human blood (**Deininger, 2018**).

The previously discussed literature from the Kirkbride group, utilising MALDI MSI to provide spatial information on blood- and vaginal fluid-contaminated fingermarks, both pre- and post-enhancement, is very promising. However, the spatial resolution of the molecular ion images of haemoglobin and vaginal fluid-derived peptides is suboptimal for the clarity required by dactyloscopy (**Kamanna *et al.,* 2017**). In a further paper, the Kirkbride group describe the compatibility of DNA profiling with MALDI MS/MS in combination with bottom-up

166

proteomics for the comprehensive examination of biological fluid trace evidence from a single swab. Again investigating material sampled from under fingernails, the group report on the identification of an individual through PCR DNA profiling, in parallel to determination of the biological fluid, blood or vaginal fluid, through bottom-up proteomics, from separate groups of fibres plucked from a single swab for each analysis. The authors elucidate the complementary approach of DNA profiling and MALDI MS profiling to generate information towards the identification of an individual and the bodyfluid in question, and, subsequently piece together the circumstances around a crime. However, the authors found that the clean-up procedure necessary for PCR, which involves cell lysis and magnetic bead clean-up of nucleic acids, meant that haemoglobin could not be detected via MALDI MS post-PCR due to the effectiveness of the clean-up procedure. They also noted that the presence of haemoglobin within a sample does not necessarily mean that sample is blood. MALDI MS is so sensitive and haemoglobin peptides preferentially ionise compared to some of the other biofluid proteins, so the presence of haemoglobin could not rule out that there were trace levels of blood present in another biofluid such as semen, saliva, vaginal fluid or menstrual fluid, depending on the circumstances. Despite these issues, their approach shows promise towards using these techniques in tandem without compromising the evidence, which may be scarce (**Kamanna *et al.,* 2018**).

The author of this chapter also investigated the compatibility of MALDI MSP and MSI prior to DNA profiling. DNA typing is capable of identifying an individual by creating a profile, but not of determining which biofluid/s the sample originated from. The authors employed MALDI MS for the confirmation of human blood through a bottom-up proteomic approach, followed by DNA profiling to

investigate the prior technique degraded the available sample. It was a concern that the short wavelengths employed by UV MALDI lasers of 337 or 355 nm may be detrimental to DNA typing, as short wavelength UV light is purposely used for degrading DNA and RNA for sterilising items. In fact this knowledge has been exploited to utilise UV-C (100 – 280 nm) light for sterilisation purposes, particularly around 254 nm (**Reed, 2010**). The study generated blood fingermarks on a painted substrate, enhanced with either ninhydrin or Acid Black 1, simulating bloody prints on painted walls. In both cases, MALDI MSP successfully detected haem and human-specific haemoglobin peptides. In addition, it was possible to map the location of these analytes within the marks using MALDI MSI. Following MALDI analysis, the study found that the majority of samples (73%) were still able to provide a profile suitable for comparison, demonstrating a level of compatibility between the two analytical techniques (**Kennedy et al., 2021**).In 2019, Jiang et al. utilised MALDI MS to differentiate blood, semen, urine, sweat and saliva. Their method involved intact detection of metabolites as biomarkers, in partnership with multivariate statistical analysis, and required minimal sample preparation, and took less than 10 minutes from sampling to data acquisition (**Jiang et al., 2019**). The study relied on the detection of small metabolic molecules such as haem, haemin, urea, creatinine, phosphate and uric, citric and lactic acids. However, while useful for distinguishing biofluids in combination with the whole metabolite profile, not all metabolites were specific to each fluid, for example, lactic acid was used to determine blood and sweat; phosphatidyl ethanolamine for blood and saliva; citric acid for semen and urine; glutamic acid for semen and sweat; uric acid for urine and saliva; phosphoric acid for sweat and saliva. Therefore the full metabolic profile should be considered and PCA performed, rather than

detection of characteristic markers specific to each biofluid. While they claim that 9 out of 10 blind samples were correctly identified through Principal Component Analysis (PCA), all but 3 samples fall outside the clusters of each biofluid (**Jiang *et al.,* 2019**).

In contrast to MALDI, other groups have been focussing on methods of ambient sample introduction. In a step away from low pressure vacuum sample chambers, Rankin-Turner *et al.* proposed a novel atmospheric pressure sample introduction method utilising a prototype Desorption off Surface (DoS) probe, based on the principle of thermal desorption, to introduce volatile organic compounds (VOCs) to an ambient mass spectrometer detector (an Advion ExpressIon compact mass spectrometer). The group used discriminant analysis to resolve blood from other biofluids in situ, with minimal sample disturbance (**Rankin-Turner *et al.,* 2019**). This is advantageous in terms of speed (no pump down required) and sample size is not limited to the dimensions of a sample chamber.

An advantage of using MALDI MS or MSI over the 'gold standard' LC-MS approach for identification is that while MALDI MS cannot be considered completely 'non-destructive', due to it ablating several micrometres of material from the surface of a sample, it is most certainly not as destructive as the swabbing and extraction sample preparation procedures necessary for chromatography-based techniques, which would destroy at least some of the ridge detail. MALDI MS also offers the analyst the chance to acquire data from many areas within the sample plate in a matter of seconds, once the sample has been prepared, whilst LC-MS runs may take in the order of an hour for a single sample.

Finally, MALDI MSI offers chemical information relating to specific coordinates within a sample, and is able to generate molecular ion images which can be overlaid on top of an optical image to distinguish the location of analytes. The current capabilities and advancements in the technology which have been discussed present a current and viable alternative to false-positive colorimetric tests currently used at the crime scene, and the time consuming and complex LC-MS confirmatory laboratory testing required to confirm the identity of a suspected blood sample. The technique has successfully been demonstrated on marks which have been previously enhanced, and even on samples which have been aged for up to 9 years on a ceramic tile (enhanced with Acid Black 1) (**Patel *et al.,* 2016**), and up to 37 years on fabric (enhanced with ninhydin) (**Francese, 2019**) proving its capability to integrate into current investigation workflows. As a result of some of this literature and their own investigations, the Home Office have recognised its capabilities and have listed MALDI currently as a Category C process in the 2014 edition of the Home Office Fingermark Visualisation Manual (**Bandey *et al.,* 2014**). This is ranked on the A-F scale of processes where those in Category A are defined as "standard processes for routing operational use", moving down to Category F, where processes are not recommended for health and safety reasons. Category B and C processes are those described as "optional processes for occasional operational use" and are recommended if Category A processes have been exhausted or for niche applications. However, it is anticipated that following further research and validation in the intervening years, MALDI will be elevated to a Category B process in the most recent edition of the Fingermark Visualisation Manual, due to be published in 2022 (**Bandey *et al., [in press]***).

The Home Office also listed MALDI as a 'specialist imaging technique' in the most recent edition of the Fingerprint Source Book (v2.0, 2017), adding "capable of providing additional information about a fingermark and/or previously undetected ridge detail and its use should not be discounted". The Centre for Applied Science and Technology (CAST) revealed why the decision was made not to designate MALDI as a Category A technique, citing "the additional time taken to perform the analysis; and the number of scenarios where additional contextual information is necessary may be limited" (**Bleay *et al.,* 2017**).

Over the past 8 years, MALDI MS has been proven as a viable alternative, mitigating, and in some cases resolving, the challenges exhibited by the techniques previously discussed and acting as a confirmatory test. Continued validation studies would contribute to demonstrating the potential of the mass spectrometric techniques of MALDI MS and MSI to elevate the confidence of end users of the ability of MALDI to reveal potentially crucial molecular information. It would be desirable for the completion of full validation studies to enable the official deployment of these MALDI instrumentation, either mobile units deployed at crime scenes, or within the arsenal of forensic laboratories for recovered evidence when the presence of blood is suspected. The next achievement would be the elevation of mass spectrometric techniques from Category C and B towards Category A in the Home Office Fingermark Visualisation Manual.

Towards this goal, this chapter proposes an alternative method for the reliable identification of blood, that overcomes the shortcomings of colorimetric presumptive testing, and also demonstrates the capability of the technique to offer information of the species of origin of blood in stains and bloody

fingermarks. This work builds on a foundation of research, undertaken by the research group and others, utilising a MALDI MS and proteomics approach for the identification of common blood (and species-specific) biomarkers for species determination. This study comprised of blood samples from human and 4 other domestic animal species. The specificity of the method was investigated using interferents that have been shown in the literature to yield false positive results with BETs, as well as other human biofluids. The method was designed to have a greater specificity that these conventional presumptive tests, and to be faster and easier to perform than the current 'gold standard' LC MS/MS. This work should be considered a pre-validation study, investigating a small range of domestic animal species, the blood of which may be encountered incidentally in a domestic crime scene through the preparation of meat. A limited number of common blood enhancement techniques were also assessed for compatibility with the proposed method. The method was successful in the identification of proteotypic peptide biomarkers and the discrimination of human blood from animal blood and from other human biofluids and interferents. Iterative method development allowed for the correct identification of a final batch of 13 blind samples, all correctly assigned as blood/non blood, with the correct species identification in all but once instance, where the species assignment was inconclusive. This pre-validation study contributes to the body of work regarding MALDI as a forensic analytical technique and demonstrates its capabilities to end users for use in specific situations. It also contributes some validation data required for the recommendation of MALDI MS for potential use in casework and the elevation of MALDI into higher categories technological readiness in publications such as the Home Office Fingermark Visualisation Manual.

Materials and Methods

Materials

Acetonitrile and formic acid were purchased from Fisher Scientific (Loughborough, UK). Trifluoroacetic acid (TFA), α-cyano-4-hydroxycinnamic acid (α-CHCA) and Millipore ZipTips containing C18 stationary phase were purchased from Sigma Aldrich (Poole, UK). Trypsin (sequencing grade, modified, lyophilized) was purchased in 20 μg vials from Promega (Southampton, UK). RapiGest™ SF surfactant was purchased in 1 mg vials from Waters Corporation (Wilmslow, UK). Sigma dry tubed swabs were purchased from Medical Wire (Wiltshire, UK). Intravenous blood samples (bovine, porcine and chicken) were purchased in aliquots of 1 mL defibrinated and 1 mL with EDTA for each species from TCS Biosciences (Buckingham, UK). All stain and fingermark samples were donated in a blind fashion by Dr Glenn Langenburg, Elite Forensic Services (Minnesota, USA), which included human and animal blood, (bovine, porcine, chicken, wild boar), other human biofluids (semen, sweat, saliva) and non-biofluid samples (beetroot juice, egg white, ketchup, steak juice). The human blood was collected through phlebotomy. The animal blood was collected by syringing blood from the chest cavity of freshly slaughtered animals by a butcher (one per animal). Samples were provided in duplicate as non-enhanced and enhanced exhibits on aluminium TLC plates which had had the silica removed. Enhancement techniques applied were Acid Black 1, Acid Yellow 7, and Leucocrystal violet (LCV). The samples were then stored in plastic slide holders at room temperature and shipped on dry ice from the US to Sheffield Hallam University. Additional commercial blood samples were obtained intravenously from bovine, porcine and chicken sources. The study was conducted in compliance with the

Ethics Application (HWB-BRERG23-13-14), approved by Sheffield Hallam University.

Instruments and Instrumental Conditions

For sample preparation prior to MALDI MS experiments, a SunCollect automatic matrix application system was used (SunChrom, Germany). The trypsin solution was applied with one system and the MALDI matrix solution with another identical system to avoid cross-contamination. Nine layers of the proteolytic and 4 layers of the matrix solution were sprayed at a flow rate of 2 µL and 4 µL/min respectively, at a pressure of 2 bar using nitrogen as the flow gas. Layers were sprayed constantly without allowing for a drying period in between and on a slow raster speed.

All MALDI analysis was carried out using a Synapt G2 HDMS MALDI-QTOF (Waters Corporation, Wilmslow, UK). The MALDI instrument was fitted with an Nd:YAG laser which was set to a repetition rate of 1 kHz. Before each acquisition, the instrument was calibrated with phosphorus red across a mass range of m/z 600 - 2500, using at least 21 m/z values. A small quantity of phosphorus red was saturated in approximately 1 mL of acetonitrile and 0.5 µL was pipetted into the lock mass well for each sample as an internal standard. A profile for phosphorus red was acquired at the start of each sample acquisition and used as a lock mass against its theoretical reference mass list.

MALDI MS data were acquired in positive reflectron mode between the mass range m/z 600 - 2000. MALDI MSI data were acquired in sensitivity mode, at a mass resolution of 10,000 FWHM. MALDI MS images were acquired at a spatial resolution of 100 x 100 µm. Ion Mobility Spectrometry (IMS) was enabled over the acquisition mass range, applying a variable wave velocity of initial velocity

(900 m/s) to end velocity (100 m/s). MALDI MS/MS data were acquired using argon as the collision gas, with a cooling gas flow of 10 a.u. and a trap gas flow of 2.0 mL/min. Trap collision energy was set at 100 a.u., laser power at 300 a.u., with a low mass resolution of 14.6 a.u., and a high mass resolution of 11.0 a.u. The trap collision energy was set at 100 a.u.

Additional LC-MS/MS analyses were performed on a Xevo G2-XS LC-MS QTOF instrument using an Acquity UPLC system (Waters Corporation, Wilmslow, UK). The mobile phase comprised of A: 0.1% formic acid in ultrapure H_2O and B: 0.1% formic acid in acetonitrile (HPLC grade). The column used was an Acquity UPLC HSS T3 utilising high strength silica (HSS) C18 particle substrate chemistry, with T3 bonding technology. The column contained particles of size 1.8 µm and 100 Å (Å = 1 x 10^{-10}m = 0.1 nm) pore size. The internal diameter was 2.1 mm and the length was 100 mm (**www.waters.com**). Samples of injection volume 5 µL at a flow rate of 0.2 mL/min were injected from an initial 3% mobile phase B. The mobile phase was ramped linearly from 3% B to 95% B over 49 minutes and ramped to 98% in the last 60 seconds, which was maintained for 2 minutes. The mobile phase was then ramped down to 3% B over 4 minutes, with a combined total run time of 56 minutes. The column temperature was set at 45°C. MS/MS acquisitions were carried out in data dependent scan mode in the mass range *m/z* 100 - 1800. The cone voltage was 40 V and the collision voltage was 30 V.

Data Processing

An in-house protein database was generated for all of the species in the study (human, bovine, porcine, chicken, wild boar) with selected peptides originating from the most abundant blood proteins, namely: haemoglobin (alpha and beta

chains), haptoglobin, myoglobin, apolipoprotein, glycophorin A, albumin and erythrocyte membrane protein band 4 (EPB4.2) the sequences of which were retrieved from UniProt (www.uniprot.org). The database was populated using theoretical *m/z* values generated from *in silico* digestion using trypsin as the proteolytic enzyme. Using the 'peptide mass' feature, the parameters allowed for up to 2 missed cleavages, and calculated the peptide mass using monoisotopic values, displaying the 'M+H$^+$' ion mass up to 3000 Da. Values were then filtered to select those in the *m/z* range 600 - 2000.

Spectral processing was performed both in the Waters proprietary MassLynx software (Waters Corporation, Wilmslow, UK) and mMass, an open source spectral processing software (**Stroham *et al.,* 2008, Stroham *et al.,* 2010**). Peak centroiding was performed in MassLynx, then the resulting spectrum mass list was converted to a plaintext text file to enable it to be compatible with mMass. A minimum signal-to-noise (S/N) threshold of 10:1 was applied in mMass. Internal mass libraries were generated within mMass to account for known matrix (matrix cluster and adduct) and trypsin peaks. These could then be background subtracted from the sample spectra to simplify the mass list. Macros were generated to automatically highlight and match peptides in the experimental mass lists, viewed in mMass, which were within ±30 ppm error of the values in the theoretical database. Conditional formatting could then be applied to highlight peptides in the experimental mass lists that were proteotypic or those that were shared between multiple species. Blood proteins were investigated from human, bovine, porcine and chicken theoretical reference lists. The UniProt database does not contain wild boar specific blood protein sequences and therefore, pig (sus scrofa) was selected as taxnomic species best approximation.

The peptides matched in mMass were then checked in MassLynx and confirmed if they were within ±15 ppm of the theoretical values.

Raw MS/MS spectra were converted into plaintext text files to be processed in mMass, where peak smoothing and labelling were applied. The peak labels were compiled into a mass list and automatically searched against the MASCOT database (**www.matrixscience.com**) using the mMass plugin to compare against *in silico* generated MS/MS spectra. The *Chordata* taxonomy was selected when chicken blood was putatively identified, the *Mammalia* taxonomy when bovine or porcine blood was putatively identified. Within MASCOT, the search parameters allowed for up to 2 missed cleavages, with the ion fragments displayed as monoisotopic values, displaying the 'M+H$^+$' ion mass, with a precursor ion tolerance of ±40 ppm and a fragment ion tolerance of ±60 ppm. The search was set to allow for variable methionine oxidation and for detection of the *y* and *b* ions and *a* and *c* immonium ions.

Methods

Extraction

The *in solution* samples were provided in duplicate with one set unenhanced and one set enhanced with either AB1, AY7 or LCV. A 10 µL aliquot was withdrawn from the samples, which was added to 40 µL of 40 mM ammonium bicarbonate to adjust the pH so as not to degrade the proteins.

Each biofluid stain and fingermark was swabbed with a swab that had been preliminarily dipped into a 70:30 ACN:H_2O solution. The swab was then cut to fit comfortably in an Eppendorf tube containing 1 mL 70:30 ACN:H_2O and

sonicated for 10 minutes to extract the proteins that had adsorbed to the swab head.

Enzymatic Digestion

A trypsin Gold and RapiGest™ SF solution was prepared fresh before each analysis by reconstituting 150 µg/mL stock trypsin in 50 mM ammonium bicarbonate (AmBic) solution, with RapiGest added at 0.1% (v/v) (stock trypsin reconstituted to 40 µg/mL, stock RapiGest reconstituted to 2 mg/mL; when combined 50:50 giving final concentrations of 20 µg/mL trypsin + 1 mg/mL RapiGest (0.1%). Nine µL of this trypsin working solution, at a final concentration of 20 µg/mL, was added to the 10 µL of sample and further 40 µL AmBic. The sample solution was then incubated for 1 hr at 37 °C in a sealed Eppendorf vial. After 1 hr, the sample was removed from the incubation chamber and 2 µL 5% TFA was added to stop the digestion. Sample aliquots of 5 µL were purified using C18 ZipTips with 5 µL 50:50 ACN:0.1% TFA as the elution buffer. Remaining digested samples were stored in sealed Eppendorf vials at -80 °C for further analysis, if required.

Commercially available intravenous blood from bovine, porcine and chicken sources (obtained in duplicate both with and without EDTA) were subjected to the same preparation procedures to generate reference spectra. Each blood reference sample was diluted 1:200 in ultra-pure H_2O, and a 10 µL aliquot was withdrawn for proteolytic digestion and purification.

The stain and fingermark samples were subjected to *in situ* sample preparation. A 150 µg/mL trypsin solution was prepared, again containing 0.1% RapiGest (v/v) in 50 mM ammonium bicarbonate. Samples were fixed to a sample tray with double sided tape, and 9 layers of the trypsin working solution were

sprayed at a flow rate of 2 µL/min using a SunCollect sprayer. The samples were then mounted on a polystyrene float in a sealed container containing 80 mL 50% MeOH$_{(aq)}$ and incubated at 37 °C for 1 hr.

Matrix Application

For the *in solution* samples, a 10 mg/mL solution of α-CHCA was prepared in 70:30 ACN:0.5% TFA. A quantity of 0.5 µL was pipette-mixed with 0.5 µL of sample solution onto the MALDI target plates by aspirating with a pipette.

For the *in situ* samples, a 5 mg/mL solution of α-CHCA in 70:30 ACN:0.5% TFA was loaded into the syringe of the SunCollect sprayer and 4 layers were sprayed at a flow rate of 4 µL/min at a slow raster speed.

Results and Discussion

The stain and contaminated-fingermark samples provided by Elite Forensic Services were provided in a blind fashion to the analyst. These samples were comprised of: human biofluids (blood, semen, saliva and sweat), animal bloods (bovine, porcine (domestic pig and wild boar) and avian (chicken)). In addition, some contaminants that either resembled blood due to their colouration, or that yielded false-positive results to common presumptive enhancements (because of their proteinous content) were included, namely egg yolk, beetroot juice, tomato ketchup, steak sauce (**Fig. 36a**). Acid Black 1 (AB1), Acid Yellow 7 (AY7) or Leucocrystal Violet (LCV) were employed as blood enhancement techniques (BETs), using standard operating procedures applied by Elite Forensic Services. Part of the samples were used to build a refined

identification strategy, through an iterative process, which was validated on a final sample subset of 13 samples (**Fig. 36b**).

The spectral data generated from the MALDI analysis were interrogated to address different levels of sample identification (**Fig. 36c**). For the suspected blood samples, 3 tiers of identification were possible; (1) 'was it blood?' (Y/N), (2) 'if it was blood, was it of human or animal origin?' (3) 'if the blood is from an animal, what species was it from?'.
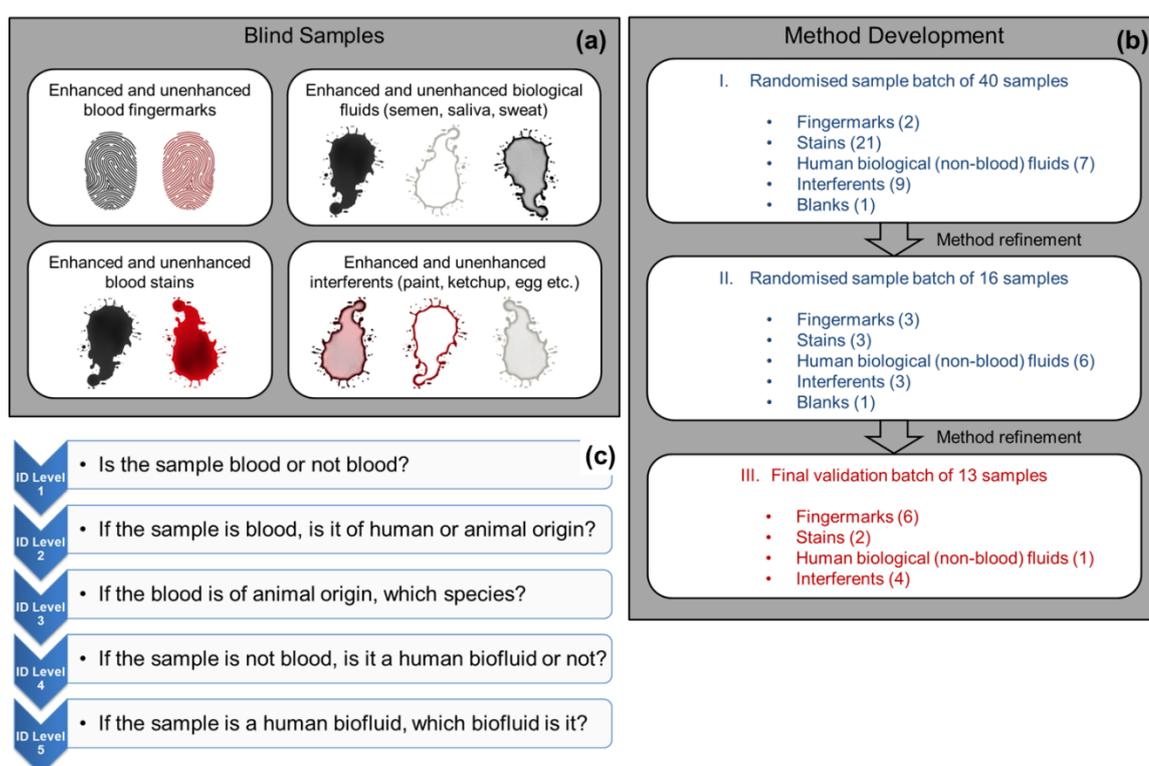
Figure 36. Schematics of samples, analysis rationale and method development pursued in this study. (a) blind sample subsets; (b) Stepwise method development for detection of blood and its provenance; (c) tiers of sample identification theoretically possible.

The primary focus of the study was on the 3 tiers of blood identification. However, it would have been theoretically possible to answer the question of whether the analytical specimen was a different biofluid or a biofluid-unrelated substance (ID level 4) and, if it was another human biofluid, which one (ID level 5). The study permitted an incidental finding on two semen biomarkers and this

was built into the final data processing strategy. The data processing refinements included the selection of the most appropriate signal-to-noise ratio values and of the mass error tolerances (mass accuracy). The mass accuracy was a direct result of the instrument calibration accuracy. Method optimisation resulted in using an S/N of 10:1 and mass error tolerance of ± 15 ppm.

In the initial stages of the study, the MALDI MS spectra acquired were obtained directly after proteolysis to minimise sample preparation. However the spectra exhibited considerable background peaks arising from non-protein signals particularly from non-commercial blood (the source of the blind samples). This background contamination issue was mitigated through the use of C18 ZipTips to pre-concentrate the proteins and purify the sample. This purification step resulted in a greater peak population and greater S/N and intensity, aiding identification.

The experiment design involved a tiered approach for the detection of unknown samples. The first three levels formed a progressive workflow if the sample was found to be blood from level 1, increasing with specificity and narrowing down the species of origin of the blood. Levels 4 and 5 were reserved for samples found to be non-blood at tier 1. They were to determine if the sample was not blood, was it still a human biofluid (level 4) and if it was, could this biofluid be identified (level 5).

ID level 1: The first level of identification was to determine if a suspected stain was blood or not. A positive identification of blood was based on the successful detection of at least one of two blood peptide peaks, present at nominal *m/z* 1275 and 1530, which consistently yielded the most intense signals. These were determined to be peptides resulting from the proteolytic digestion of

haemoglobin using trypsin as the enzyme. These were attributed to a peptide from the beta haemoglobin (βHb) chain, with amino acid sequence LLVVYPWTQR (theoretical *m/z* 1274.726) and from the alpha haemoglobin (αHb) chain, with amino acid sequence VGAHAGEYGAEALER (theoretical *m/z* 1529.734). It was discovered that while these haemoglobin peptides were common to human, bovine and porcine blood, they were absent in chicken blood. Therefore, a positive blood identification at this stage included the caveat that the blood could be either human, bovine or porcine blood, but could not be avian (chicken) blood. Conversely, the absence of these peaks could indicate either a true absence of blood or the presence of chicken blood within the system investigated.

ID level 2: The second level of identification was to determine, if a sample was found to be blood at level one, could it be determined if this was blood from a human or non-human source? As previously mentioned, the most abundant blood peptides originating from the haemoglobin protein were common to all of the mammalian species investigated in the study. The βHb peptide at theoretical *m/z* 1274.726 is present in human, bovine and porcine haemoglobin and the αHb peptide at theoretical *m/z* 1529.734 was present in human and bovine haemoglobin only, but not in porcine haemoglobin. Therefore, although detection of both these αHb and βHb peptides could rule out porcine blood, it could still not help distinguish between human and bovine blood samples though serving to narrow down provenance.

ID level 3: The third level of identification was to determine, if a sample was found to be animal blood at levels one and two, could the animal species be determined? For definitive species identification, the method relied on detection of additional blood proteins that were specific to each species, which could be

identified by detection of their constituent proteotypic peptides. These included peptides from the blood proteins: haemoglobin (additional to those Hb peptides at m/z 1274.726 and 1529.734), haptoglobin, myoglobin, apolipoprotein, glycophorin A, albumin and erythrocyte membrane protein band 4 (EPB4.2). For example, peptides at m/z 932.520 (βHb), m/z 1087.553 (EPB 4.2) and 1378.694 (α Hpt) are all specific to human blood (**Fig. 37**) within the system investigated.

ID level 4: If a sample was found not to be blood at level 1, could it be determined if it was another human biofluid or a non-biofluid substance?

ID level 5: If a sample was found to be a human biofluid at level 4, could it be determined which biofluid? (Within the confines of this study, this was limited to semen, saliva and sweat).
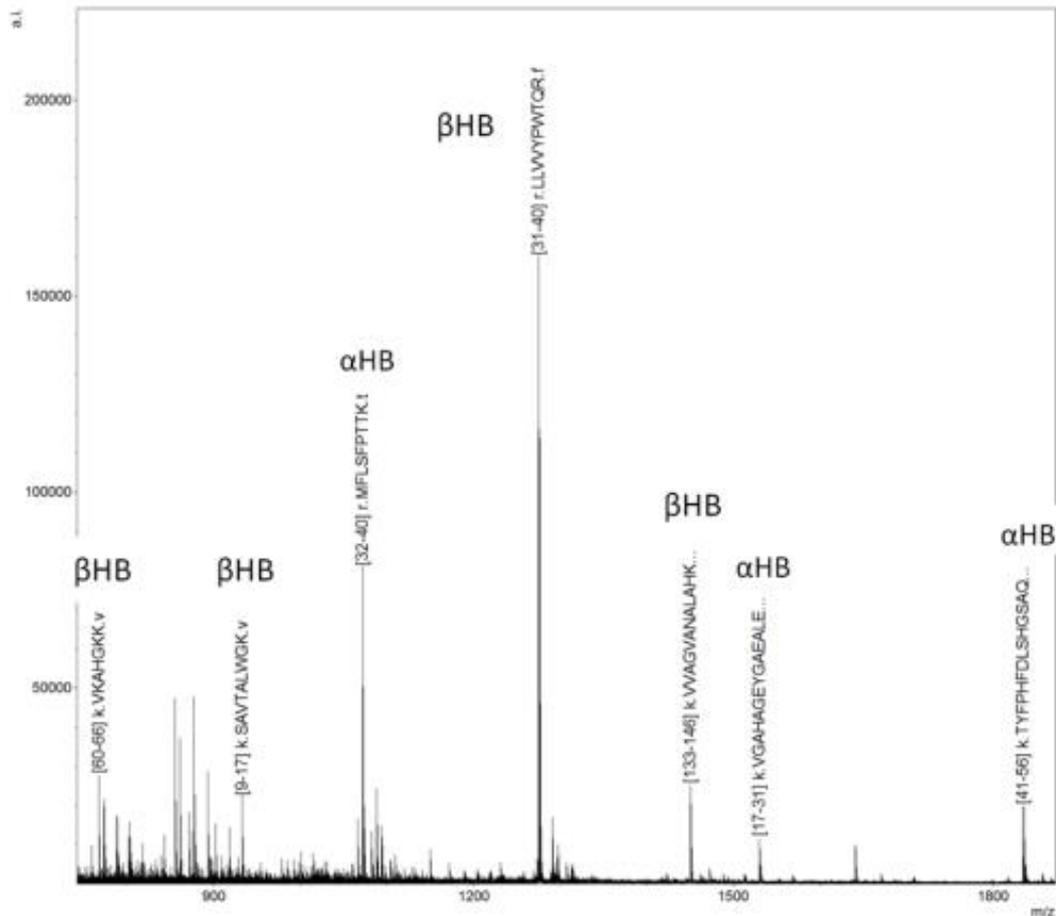
Figure 37. Putative identification of human blood from Sample 59. Haemoglobin peptides are annotated at *m/z* 767.495, 932.520, 1071.554, 1274.723, 1449.801 and 1529.733 and have been assigned as α or β chain peptides.

Whilst most species involved in the study had broadly the same selection of blood proteins, the protein amino acid sequences and therefore the subsequent peptide masses varied slightly. For example for the haptoglobin protein, human haptoglobin is comprised of 406 amino acids and has a mass of 45,205 Da, bovine haptoglobin of 401 amino acids and 44,859 Da, and porcine haptoglobin 347 amino acids and 38,482 Da. As displayed at the bottom of **Table 12**, the proteins have been aligned using the Alignment function in UniProt (www.uniprot.org). This shows the different length and amino acid sequences of the haptoglobin protein across the three species, with similar regions highlighted in grey using the 'similarity' amino acid properties tool. The overall

peptide mass fingerprint (PMF) for each would be unique, and can be used to

help identify the species of origin.

Protein	Number of AA	Mass (Da)	AA Sequence
Human haptoglobin	406	45,205	MSALGAVIAL LLWGQLFAVD SGNDVTDIAD DGCPKPPEIA HGYVEHSVRY QCKNYYKLRT EGDGVYTLND KKQWINKAVG DKLPECEADD GCPKPPEIAH GYVEHSVRYQ CKNYYKLRTE GDGVYTLNNE KQWINKAVGD KLPECEAVCG KPKNPANPVQ RILGGHLDAK GSFPWQAKMV SHHNLTTGAT LINEQWLLTT AKNLFLNHSE NATAKDIAPT LTLYVGKKQL VEIEKVVLHP NYSQVDIGLI LPVADQDQCI RHYEGSTVPE KKTPKSPVGV QPILNEHTFC AGMSKYQEDT CYGDAGSAFA VHDLEEDTWY ATGILSFDKS CAVAEYGVYV KVTSIQDWVQ KTIAEN
Bovine haptoglobin	401	44,859	MSALQAVVTL LLCGQLLAVE TGSEATADSC PKAPEIANSH VEYSVRYQCD KYYKLHAGNG VYTFNNKQWI NKDIGQQLPE CEEDDSCPEP PKIENGYVEY LVRYQCKPYY TLRTCGDGVY TFNSKKQWIN KNIGQKLPEC EAVCGKPKHP VDQVQRIIGG SLDAKGSFPW QAKMVSQHNL ISGATLINER WLLTTAKNLY LGHSSDKKAK DITPTLRLYV GKNQLVEVEK VVLHPDHSKV DIGLIKLRQK VPVNDKVMPI CLPSKDYVKV GRVGYVSGWG RNENFNFTEH LKYVMLPVAD QDKCVKHYEG VDAPKNKTAK SPVGVQPILN ENTFCVGLSK YQDDTCYGDA GSAFVVHDKE DDTWYAAGIL SFDKSCAVAE YGVYVKVTSI LDWVRKTIAN N
Porcine haptoglobin	347	38,482	MRALGAVVAL LLCGQLFAAE TGNEATDATD DSCPKPPEIP KGYVEHMVRY HCQTYYKLRT AGDGVYTLDS NKQWTNKVTG EKLPECEAVC GKPKNPVDQV QRIMGGSLDA KGSFPWQAKM ISHHNLTSGA TLINEQWLLT TAKNLRLGHK NDTKAKDIAP TLRLYVGKKQ EVEIEKVIFH PDNSTVDIGL IKLKQKVPVN ERVMPICLPS KDYVNVGLVG YVSGWGRNAN LNFTEHLKYV MLPVADQEKC VQYYEGSTVP EKKTPKSPVG VQPILNEHTF CAGLSKYQED TCYGDAGSAF AVHDKDDDTW YAAGILSFDK SCRTAEYGVY VRVTSILDWI QTTIADN

```
P00738 HPT_HUMAN      1  MSALGAVIALLLWGQLFAVDSGNDVTDIADDGCPKPPEIAHGYVEHSVRYQCKNYYKLRT
Q2TBU0 HPT_BOVIN      1  MSALQAVVTLLLCGQLLAVETGSEA---TADSCPKAPEIANSHVEYSVRYQCDKYYKLH-
Q8SPS7 HPT_PIG        1  MRALGAVVALLLCGQLFAAETGNEATDATDDSCPKPPEIPKGYVEHMVRYHCQ-------

P00738 HPT_HUMAN     61  EGDGVYTLNDKKQWINKAVGDKLPECEADDGCPKPPEIAHGYVEHSVRYQCKNYYKLRTE  1
Q2TBU0 HPT_BOVIN     57  AGNGVYTFNN-KQWINKDIGQQLPECEEDDSCPEPPKIENGYVEYLVRYQCKPYYTLRTC  1
Q8SPS7 HPT_PIG       54  ----------------------------------------------------TYYKLRTA

P00738 HPT_HUMAN    121  GDGVYTLNNEKQWINKAVGDKLPECEAVCGKPKNPANPVQRILGGHLDAKGSFPWQAKMV  1
Q2TBU0 HPT_BOVIN    116  GDGVYTFNSKKQWINKNIGQKLPECEAVCGKPKHPVDQVQRIIGGSLDAKGSFPWQAKMV  1
Q8SPS7 HPT_PIG       62  GDGVYTLDSNKQWTNKVTGEKLPECEAVCGKPKNPVDQVQRIMGGSLDAKGSFPWQAKMI  1

P00738 HPT_HUMAN    181  SHHNLTTGATLINEQWLLTTAKNLFLNHSENATAKDIAPTLTLYVGKKQLVEIEKVVLHP  2
Q2TBU0 HPT_BOVIN    176  SQHNLISGATLINERWLLTTAKNLYLGHSSDKKAKDITPTLRLYVGKNQLVEVEKVVLHP  2
Q8SPS7 HPT_PIG      122  SHHNLTSGATLINEQWLLTTAKNLRLGHKNDTKAKDIAPTLRLYVGKKQEVEIEKVIFHP  1

P00738 HPT_HUMAN    241  NYSQVDIGLIKLQKVSVNERVMPICLPSKDYAEVGRVGYVSGWGRNANFKFTDHLKYVM   3
Q2TBU0 HPT_BOVIN    236  DHSKVDIGLIKLRQKVPVNDKVMPICLPSKDYVKVGRVGYVSGWGRNENFNFTEHLKYVM  2
Q8SPS7 HPT_PIG      182  DNSTVDIGLIKLKQKVPVNERVMPICLPSKDYVNVGLVGYVSGWGRNANLNFTEHLKYVM  2

P00738 HPT_HUMAN    301  LPVADQDQCIRHYEGSTVPEKKTPKSPVGVQPILNEHTFCAGMSKYQEDTCYGDAGSAFA  3
Q2TBU0 HPT_BOVIN    296  LPVADQDKCVKHYEGVDAPKNKTAKSPVGVQPILNENTFCVGLSKYQDDTCYGDAGSAFV  3
Q8SPS7 HPT_PIG      242  LPVADQEKCVQYYEGSTVPEKKTPKSPVGVQPILNEHTFCAGLSKYQEDTCYGDAGSAFA  3

P00738 HPT_HUMAN    361  VHDLEEDTWYATGILSFDKSCAVAEYGVYVKVTSIQDWVQKTIAEN  4
Q2TBU0 HPT_BOVIN    356  VHDKEDDTWYAAGILSFDKSCAVAEYGVYVKVTSILDWVRKTIANN  4
Q8SPS7 HPT_PIG      302  VHDKDDDTWYAAGILSFDKSCRTAEYGVYVRVTSILDWIQTTIADN  3
```

Table 12. A comparison of the amino acid sequences between human, bovine and porcine haptoglobin, including the number of amino acids in the sequence and the mass of the intact protein. The last row shows the similarities between the three amino acid sequences, with 3 identical amino acids highlighted dark grey and two identical amino acid highlighted light grey, where there has been a substitution mutation in one of the species chains. A dash represents where there has been a deletion mutation in that species, or an insertion mutation in the species it is being compared to.

However, proteotypic peptides were not always detected, and if present usually had weak signal intensity, and poor mass resolution. In some cases, proteotypic peptides from more than one species were detected concurrently, meaning definitive species identification was not possible.

In the initial classification method, a level 2 identification was made and human blood was putatively claimed if the two haemoglobin β and α chain peptides at *m/z* 1274.726 and 1529.734 respectively were detected in the spectra. However, only using these two haemoglobin peptides for human identification meant the method was susceptible to reporting a false positive result for human blood and a false negative for bovine blood.

As method development was an iterative process, the findings from this initial stage were used to inform the subsequent stage and a more robust pre-processing and interpretation methodology was proposed. The hypothesis was made and tested that in the case of 'proteotypic' peptides detected from multiple species, if the shared proteotypic peptides were found in higher numbers in one species, this observation was used to assign the species. This was termed the 'highest frequency of proteotypic peptides' identification strategy'. For example, if 8 peptides were detected that were all present in more than one species, and 8 of them were present in bovine blood, 4 in porcine blood and 4 in chicken blood, could the conclusion be made that it was definitively bovine blood? This approach did not yield a definitive identification as often there was little discrepancy in the number of proteotypic peptides for each species, or there were few proteotypic peptides detected at all. A hypothesis was also tested to determine if the ratio of the intensity of the two most prominent Hb signals could be used to discriminate between human and bovine blood, but this was also

unsuccessful as the most dominant signal varied between samples of the same species.

As mentioned previously, a key aim of the study was to minimise the overall analysis time, including sample preparation, spectral acquisition, data processing and interpretation of the results generated. In relation to the data processing and interpretation, the approach depended on rapid putative identification of peptides based on their *m/z* within a certain mass tolerance. A drawback with this approach was that it was possible for multiple blood protein peptides to be present within each 30 ppm mass window. Consequently, this strategy had the potential to putatively indicate the wrong species, when the correct species identification was revealed. For this reason, this method was subsequently discarded. The results of this approach are displayed in **Table 13**, which shows the putative identifications for tiers 1-3, the enhancement reagent and the correct identification for each sample.

Sample #	Identification			Enhancement reagent	Actual Identification
	Level 1	Level 2	Level 3		
1 S	Blood	Human	n/a	n/a	Bovine blood
2 S	Blood	Human	n/a	n/a	Human blood
3 S	Blood	Human	n/a	n/a	Human blood
5 S	Blood	Animal	Bovine	AB1 +ve	Bovine blood
6 S	Blood	Human	n/a	n/a	Human blood
7 S	Blood	inconclusive	inconclusive	n/a	Porcine blood
12 S	Non-blood	n/a	n/a	n/a	Human sweat
13 S	Non-blood	n/a	n/a	LCV +ve	Porcine blood
14 S	Non-blood	n/a	n/a	AB1 +ve	Human blood
16 S	Blood	Human	n/a	n/a	Human blood
17 S	Blood	Human	n/a	n/a	Human blood
18 S	Blood	Animal	Porcine	n/a	Wild boar blood
26 S	Non-blood	n/a	n/a	LCV -ve	Human saliva

27 S	Non-blood	n/a	n/a	n/a	Human semen
28 S	Non-blood	n/a	n/a	n/a	Bovine blood
29 S	Non-blood	n/a	n/a	AY7 +ve	Chicken blood
30 S	Non-blood	n/a	n/a	AB1 +ve	Egg yolk
31 S	Blood	Human	n/a	LCV +ve	Human blood
34 S	Blood	Human	n/a	LCV +ve	Human blood (+EDTA)
35 S	Non-blood	n/a	n/a	n/a	Chicken blood
36 S	Non-blood	n/a	n/a	n/a	Ketchup
37 S	Blood	Human	n/a	n/a	Human blood
40 S	Non-blood	n/a	n/a	n/a	Human sweat
41 S	Non-blood	n/a	n/a	AB1 +ve	Ketchup
49 S	Non-blood	n/a	n/a	AB1 +ve	Human saliva
53 S	Blood	Human	n/a	n/a	Human blood (+EDTA)
56 S	Non-blood	n/a	n/a	n/a	Porcine blood
57 S	Non-blood	n/a	n/a	n/a	Paint
59 S	Blood	Human	n/a	n/a	Human blood
60 S	Non-blood	n/a	n/a	AY7 +ve	Human saliva
61 S	Blood	Human	n/a	n/a	Human blood
63 S	Non-blood	n/a	n/a	n/a	Porcine blood
78 S	Non-blood	n/a	n/a	n/a	Gold Bond Lotion
79 S	Non-blood	n/a	n/a	n/a	Blank
122 F	Blood	Human	n/a	AY7 +ve	Human blood
141 F	Non-blood	n/a	n/a	AB1 +ve	Ketchup
160 F	Non-blood	n/a	n/a	AY7 -ve	Human saliva
162 F	Blood	Human	n/a	AB1 +ve	Human blood
165 F	Non-blood	n/a	n/a	AY7 +ve	Egg white
175 F	Non-blood	n/a	n/a	LCV -ve	Egg white

Table 13. Initial putative identification of the batch of 40 unknown blind samples analysed using the 'highest frequency of proteotypic peptides' identification strategy. The 'S' or 'F' following the sample number represent either Stain or Fingermark. The enhancement reagents Acid Black 1, Acid Yellow 7 and Leucocrystal Violet are indicated by the acronyms AB1, AY7 and LCV. The '+ve' or '-ve' denotes whether the analytical specimen reacted positively or negatively to the selected BET. The 'actual identification' column is colour-coded green if the blood and species assignment was correct, orange if the blood assignment was correct but the species was incorrect, and red if the blood/non-blood assignment was incorrect.

Classification was considered 'correct' if the level 2 tier (if blood, human or animal) assertion matched the actual sample identity. Of the 40 samples analysed, all of the non-biofluid 'contaminant' samples were correctly recognised as non-blood. These accounted for 9 out of the 40 samples and comprised of four samples that elicited false-positive results with enhancement reagents (ketchup, egg yolk and egg white), one sample that was negative with LCV (egg white), with the remaining four samples (ketchup, paint, lotion and a 'blank') not having a BET applied (100% correct blood exclusion rate). A further 7 samples were comprised of non-blood biofluids (semen, saliva and sweat). Again, all of these were correctly identified as non-blood, exploiting the absence of m/z 1275 and 1530 haemoglobin peptides (100% correct blood exclusion rate). Of the 24 remaining blood samples, 14 were human and 10 were animal blood. Only 1 of the human blood samples (14 S) was not recognised as human blood, as the signature haemoglobin peptides (m/z 1275 and 1530) were not detected, nor were other characteristic blood peptides (m/z 1071, 1449, 1833). Prior enhancement with Acid Black 1 could have not been the reason for this false negative, as another human blood sample (162 F) was enhanced with this BET and correctly assigned to human blood. It is speculated that either this was the case of sample mis-labelling, or the spectrum/sample was of insufficient quality. As the original 14S sample had been used up in the initial swab stage, an aliquot from the initial extract was submitted to proteolysis again yielding the same mass spectral results. Therefore, failure to detect the relevant biomarkers could be ascribed to either an unsuccessful initial extraction or sample mis-labelling and no longer to matrix crystallisation or instrument underperformance. Finally, of the 10 remaining animal blood samples, 3 were bovine, 3 were chicken, 3 were porcine and one was wild boar blood. This was the stage at

which the strategy adopted thus far showed its limitations. Only 4 out of the 10 samples were correctly identified as animal blood, and only 2 of those had assigned the correct species, bovine (5 S) and wild boar (18 S) (40% correctly identified as blood, with only 20% correct species assignment). Only one of these animal blood samples, bovine blood (1 S), resulted as a false positive for human blood, due to the shared presence of the signature *m/z* 1275 and 1530 haemoglobin peptides and other shared proteotypic peptides.

As can be seen from the data displayed in **Table 13**, the MALDI MS analysis was compatible with the prior application of AB1, AY7 and LCV blood enhancement techniques. The correct assignments were made in 14 cases out of a total of 18 samples where BETs were used (~78% success rate). In terms of human blood samples, sample 162 F reacted positively with AB1, sample 122 F reacted positively with AY7 and sample 31 S reacted positively with LCV and all were subsequently correctly confirmed to be human blood through MALDI MS analysis, demonstrating that this technique could successfully be utilised as confirmatory test and compatibly with the application of three of the most common BETs.

The acid dyes (AB1 and AY7) stain the proteinaceous content of the blood and as such are not specific to blood (**Bleay *et al.,* 2017**). The results from this study corroborated this finding. As shown in **Table 13**, the acid dyes reacted with the non-blood samples; human saliva (49 S), ketchup (41 S), egg yolk (30 S) and egg white (165 F), highlighting the limited specificity of protein-reactive acid dyes. Conversely, LCV is a haem-reactive reagent, which has better specificity for blood. It did not positively react with any of the non-biofluid contaminants, such as egg white (175 F) or non-blood human biofluids, such as saliva (26 S). Compatibility between MALDI MS and LCV was also

demonstrated. MALDI MS analysis enables the confirmation of the presence of human blood with (34 S) and without the prior application of LCV (53 S). Sample 34 S contained ethylenediaminetetraacetic acid (EDTA), an anti-coagulant added to blood to stop it from clotting if it is needed for further analytical tests, also demonstrating compatibility with this reagent.

The acid dyes (AB1 and AY7) also reacted with bovine blood (5 S) and chicken blood (29 S), and LCV reacted with porcine blood (13 S) and chicken blood (63 S). AB1 and AY7 bind to the blood proteins cationic groups (**Bossers *et al.,* 2011**), which are also present in the blood of species other than humans. LCV reacts with the haem in the blood in the presence of hydrogen peroxide (**Bleay *et al.,* 2017**), which is also present in the blood of other species. While technically not a false positive, as the BETs were reacting to the presence of blood, they lack the specificity needed to indicate the species or even distinguish between blood of human and animal origin. However, the work presented in this chapter has demonstrated that a bottom-up proteomic strategy utilising MALDI MS for the analysis of blood can not only distinguish between human and animal blood, but when animal blood is present, is also capable of determining the species (from a subset of domestic animals), both with and without prior enhancement.

In summary, while using this data processing strategy, MALDI MS analysis never yielded a false positive for human blood (contrary to some presumptive test application instances), it yielded false negatives for animal blood and as a consequence this also contributed to poor animal species differentiation. In an attempt to evaluate and improve the strategy, fresh blood reference samples were obtained for bovine, porcine and chicken blood, sourced from the jugular of these animals. These were analysed using identical sample preparation,

digestion and analysis conditions as applied to the blind samples. The commercially available reference blood resulted in considerably different spectral profiles to the blind blood samples that had been provided. **Figure 38** compares the difference between the chicken blood sample (10 S), and the 'fresh' chicken blood available from a commercial blood provider.
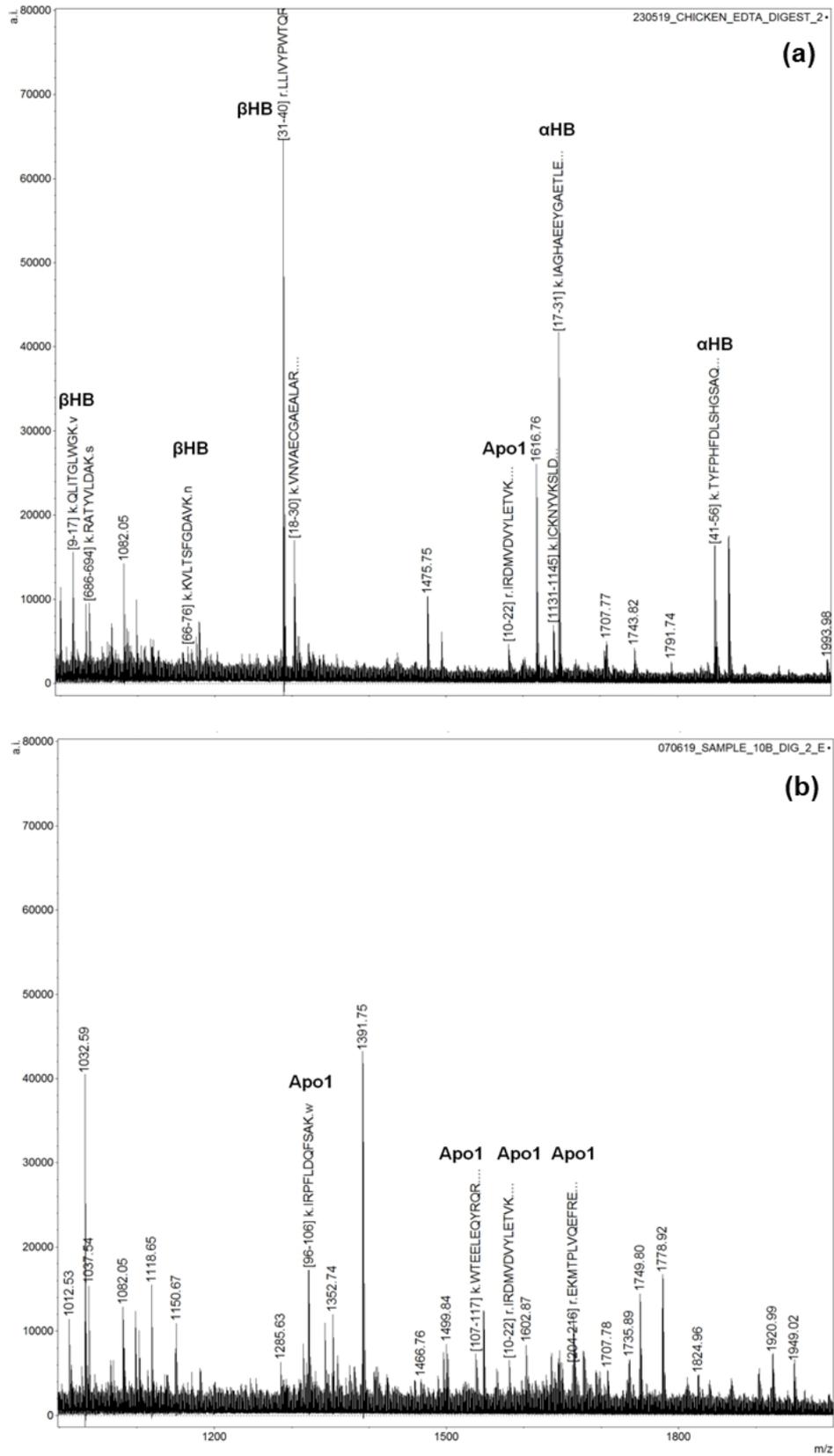
Figure 38. MALDI MS spectra of chicken blood from blind sample (10 S) (a) versus commercially obtained chicken blood, illustrating the differing peptide mass fingerprints (b).

In sample 10, a chicken blood digest sample, the only blood peptides that could be matched against the blood protein database were from apolipoprotein 1 (*m/z* 1321.727, 1537.755, 1580.818, 1662.859), whereas in the purchased reference blood, peptides could be matched against αHb (*m/z* 1645.776, 1847.900, 2121.142 and 2249.224) and βHb (*m/z* 999.487, 1036.561, 1164.648, 1288.736, 1302.645 and 2226.129), in addition to only one apolipoprotein signal in common, at *m/z* 1580.817 (IRDMVDVYLETVK, [10–22]).

When investigating the sampling methods for the procurement of the blood, it transpired that whilst the human blood samples were from venous blood collected by a phlebotomist, and the mammalian blood samples were from a large syringe used to siphon blood from the chest cavity of slaughtered animals as a by-product of the butchering process. The butcher reported that the collection of chicken blood was more challenging due to the reduced quantity of blood in comparison and the dilution of the blood with water as a result of the immediate cleaning procedure associated with the butchering process. It is proposed that these sampling variables were responsible for the prominence of the haemoglobin peptides in the reference blood samples and their absence in sample 10 S and all of the blind chicken blood samples. As previously mentioned, of the four apolipoprotein peptide signals detected in sample 10 S, only one (at *m/z* 1580.818) was also detected in the reference blood sample. This might have been due to suppression effects of the intense haemoglobin peptide signals dominating the reference blood spectra. Interestingly the most intense signals detected in the blind blood samples (*m/z* 1391.746, 1749.798 and 1778.923) could not be assigned to peptides in the in-house custom built protein database or when performing a MASCOT search using the Peptide Mass Fingerprint function (www.matrixscience.com) and it is hypothesised that

the peaks arise from more than one species with the mass resolution being insufficient for separation.

Figure 39 compares the MALDI spectra of a bovine blood reference sample (**39a**) against bovine blood blind sample 36 S (**39b**). Similarly to the chicken blood blind sample 10 S, only the two most prevalent haemoglobin signals could be detected in the bovine blind sample spectra at *m/z* 1274.706 and 1529.743, but they were of considerably lower intensity.
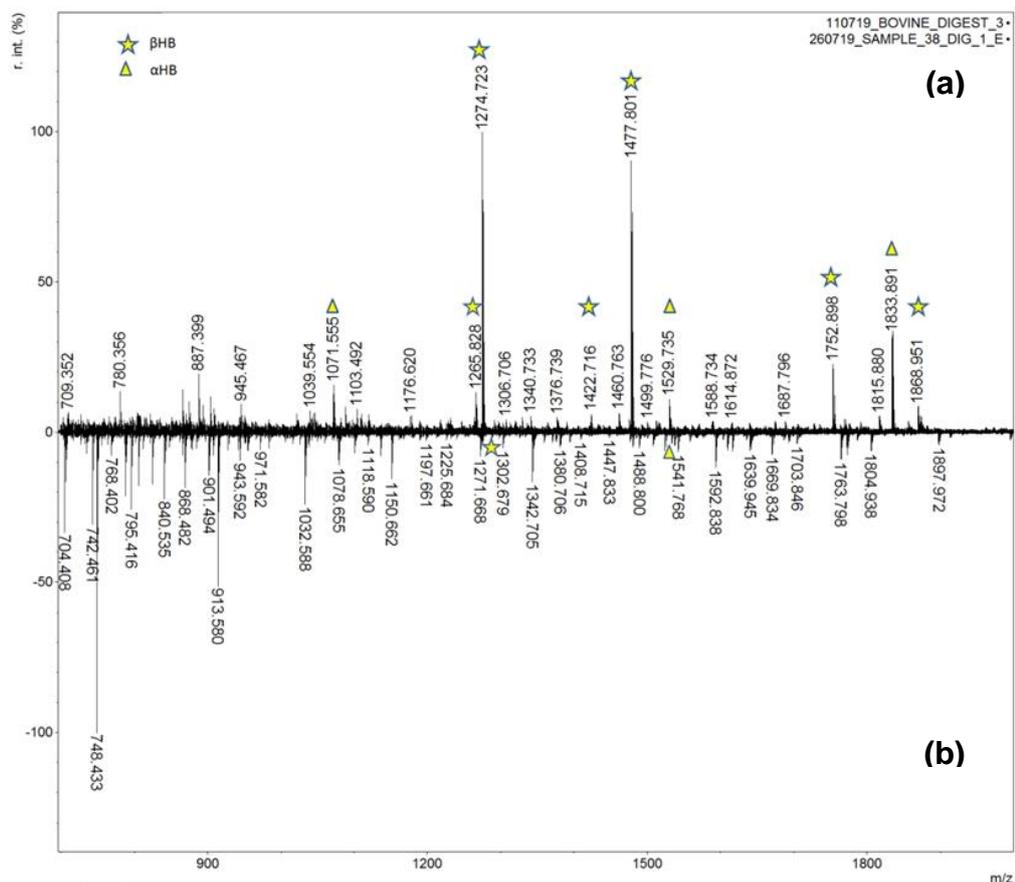
Figure 39. Mirrored MALDI MS spectra illustrating the differences in the peptide mass fingerprint for bovine blood for (a) commercially available bovine blood reference and (b) blind sample 38 S.

Finally, with regards to porcine blood, **Figure 40** compares a porcine blood reference sample (**40a**) against a porcine blood blind sample 7 S (**40b**). Similarly to the bovine and chicken blood instances, haemoglobin peptide

signals were not present in the porcine blind sample spectra but could be detected in the reference blood spectra.
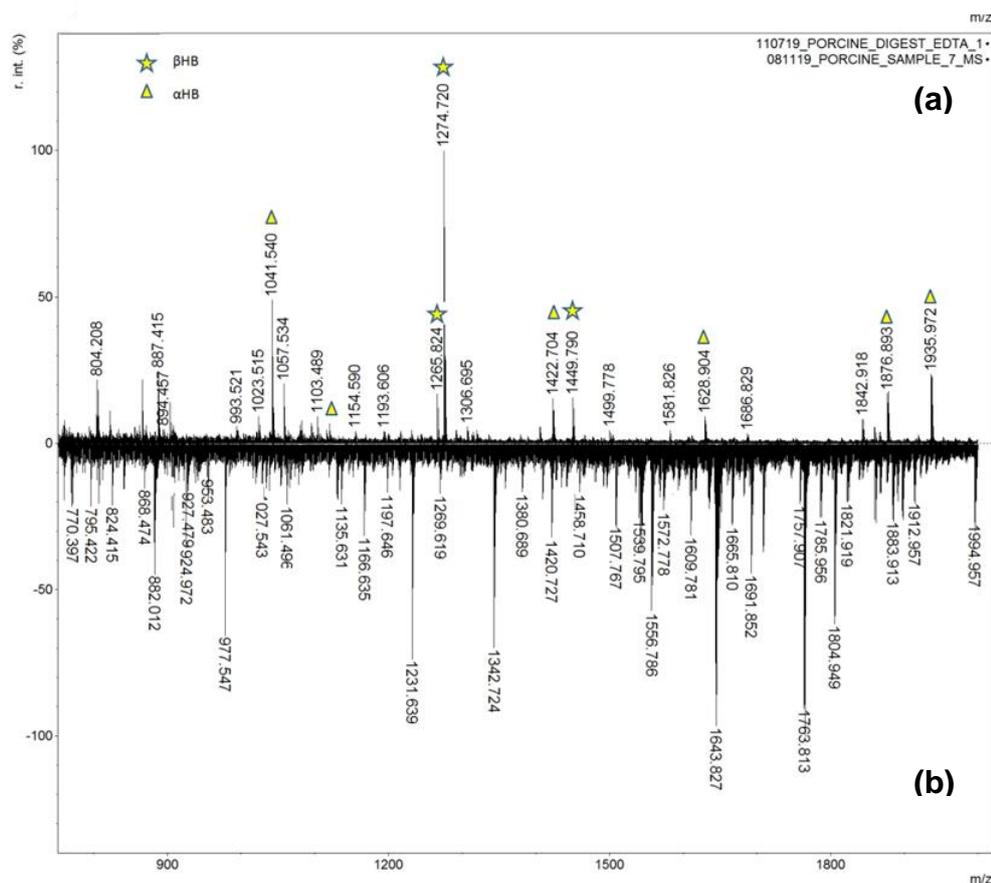
Figure 40. Mirrored MALDI MS spectra illustrating the differences in the porcine blood peptide mass fingerprint for commercially available porcine blood (a) and blind sample 7 S (b).

Overall, it was observed that the animal blood blind sample MALDI MS spectra exhibited much more complex peptide mass fingerprints, with a greater ion population, although a greater number of blood specific protein-deriving peptides could be assigned to the reference blood spectra. It was hypothesised that discrepancies in the blood sampling method might contribute towards the variance between the spectra. The supplier of the reference blood reported that for all species, blood had been obtained by making an incision in the jugular vein of the animal and that EDTA had then been added to the blood at a concentration of 1.8 g/L to prevent coagulation. Together with how the animal

blood for the blind study was collected, these circumstances likely account for the mass spectral differences for the two sample sets.

At this stage it was important to investigate the variation between the spectral profiles. The differences were the location the blood was drawn from (chest cavity vs. jugular), the method (syringing pooled blood vs. draining blood from a vein) and, possibly, the addition of additives or preservatives (such as EDTA).

The unassigned signal detected at nominal *m/z* 1750 was consistently one of the most intense peaks present in the chicken blood spectral profiles. As it was found in all of the chicken blind blood samples, but not in the chicken reference blood samples, it was proposed that it originated from either a protein not included in the custom blood protein database and/or from an exogenous source, for example from an additive added to the blood or from the vessel that the blood was collected in. Additional MALDI MS spectra on the digested fluid sampled from the blood residue present in freshly packaged chicken sold by local butchers and supermarkets also enabled the detection of this unassigned peak at nominal *m/z* 1750. This indicated that additives and preservatives could be ruled out as a source of the 'contaminant' peak, as the butchers had freshly prepared the meat and confirmed that they had not added any additives or preservatives. The MALDI MS/MS analysis of the *m/z* 1749.793 signal produced a suitable number of ion fragments and a subsequent MASCOT search returned the sequence LVSWYDNEFGYSNR of a theoretical *m/z* of 1749.787 and belonging to the protein glyceraldehyde 3-phosphate dehydrogenase (GAPDH) (UniProt accession number: P00356), with a statistically significant probability score of 91%. The MASCOT search results and match to *Gallus Gallus* (the red junglefowl) GAPDH is shown in **Figure 41**. The search parameters were set by using the taxonomy of *Chordata*.
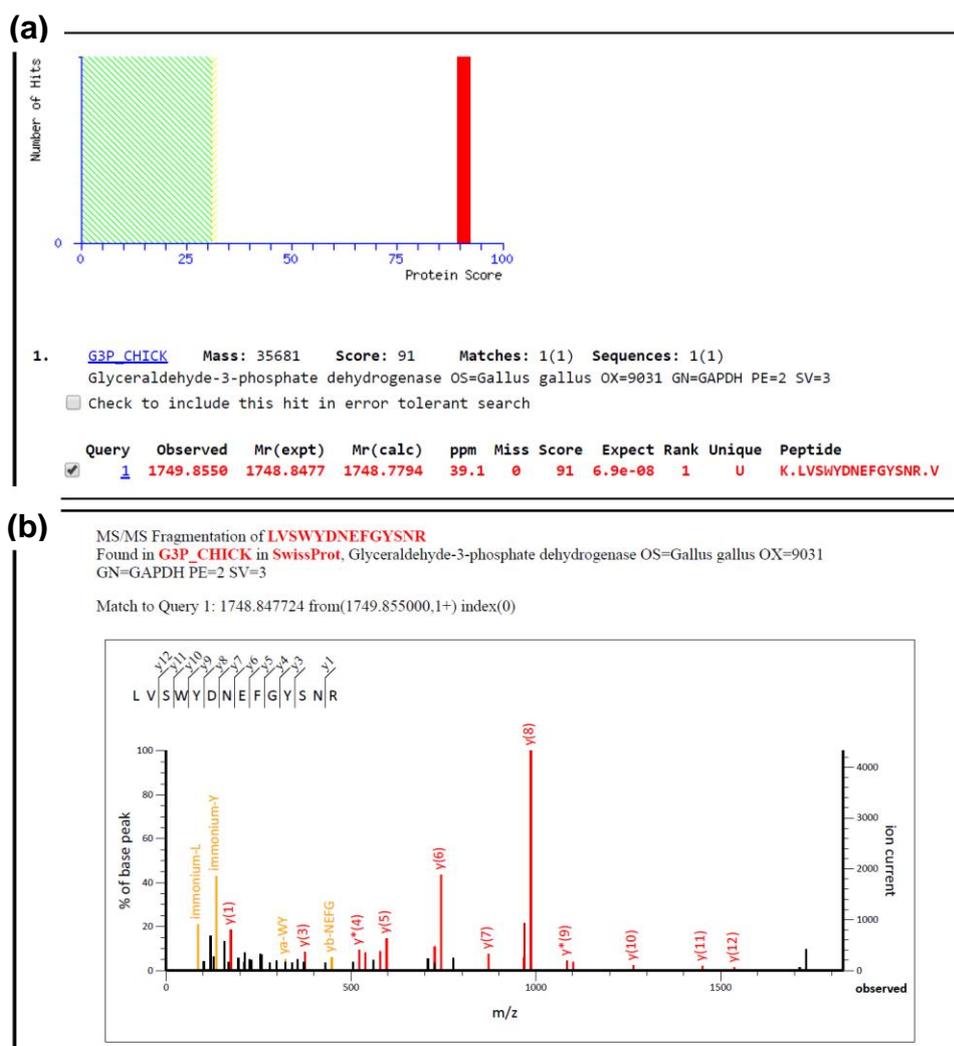
Figure 41. Identification of the ion signal at *m/z* 1749.793 present in chicken blood in blind samples and sourced from a local butcher in freshly prepared meat using MALDI MS/MS. (a) MASCOT database match against chicken GAPDH for the ion signal present at *m/z* 1749.793, with a score of 91%. (b) MASCOT assignment of the ion products resulting from MALDI MS/MS fragmentation of the *m/z* 1749.793 parent ion. *[Adapted from Kennedy et al., 2020].*

MS/MS of an ion present at this *m/z* in the blind chicken blood samples did not yield sufficient b and y ions however to allow for confirmation. LC-MS/MS was therefore employed which did confirm this peptide to belong to *Gallus Gallus* GAPDH through the detection of a signal at *m/z* 835.397, a doubly charged ion, which eluted from the column at 15.06 min. The peptide at nominal *m/z* 1750 was also sequenced by LC-MS/MS in a different blind chicken blood sample (35

S) and this time MASCOT confirmation of GAPDH was achieved with a statistically significant probability score of 92 (**Figure 42**).
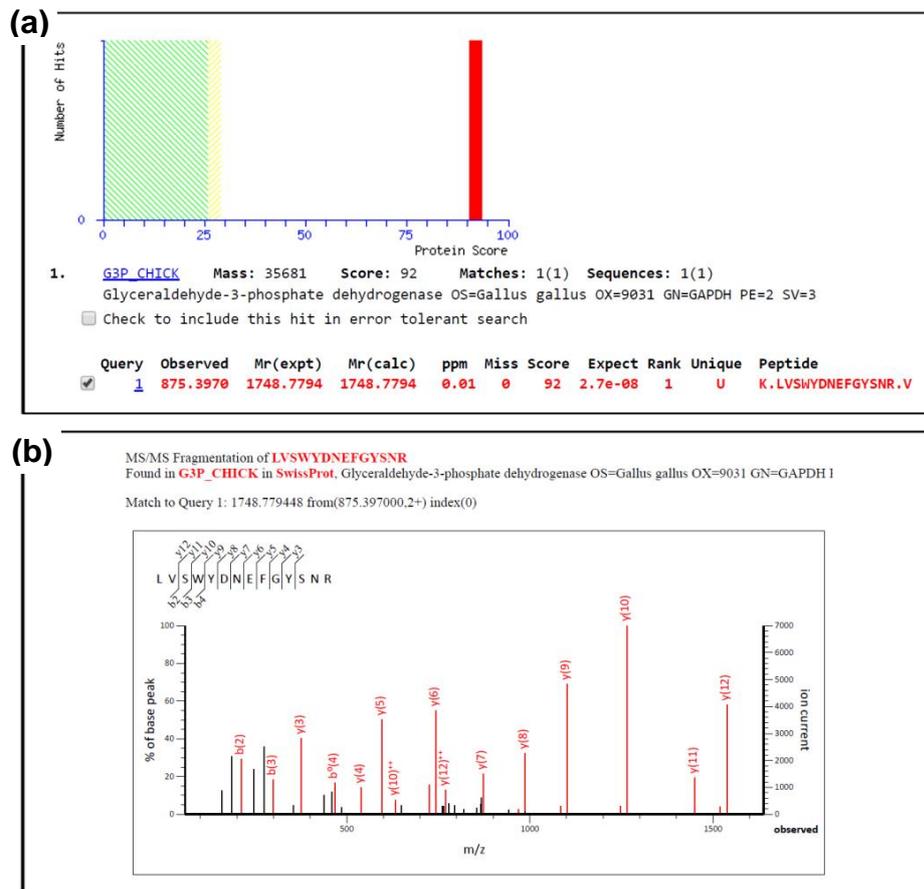
Figure 42. LC-MS/MS confirmation of the amino acid sequence of the ion signal detected at *m/z* 1749.779 in chicken blood (sample 35 S) from the doubly charged ion at *m/z* 875.397. (a) MASCOT database match against chicken GAPDH for the doubly charged ion signal present at *m/z* 875.397, with a score of 92. (b) shows a MASCOT assignment of the ion products resulting from LC-MS/MS fragmentation of the doubly charged *m/z* 875.397 parent ion. *[Adapted from Kennedy et al., 2020].*

The confirmation of the presence of a GAPDH peptide in the blind chicken blood sample resulted in the inclusion of the GAPDH protein in the custom protein database and in this peptide being one of the biomarkers for chicken blood. The inclusion of GAPH as a 'protein of interest' led to the identification of further GAPDH peptides to previously unassigned signals in the same blind chicken blood sample (10 S), at *m/z* 795.415 (mass accuracy -4.4 ppm),

805.431 (mass accuracy -0.1 ppm), 1032.595 (mass accuracy -2.9 ppm), 1358.683 (mass accuracy 2.1 ppm) and 1645.890 (mass accuracy -12.2 ppm).

When investigating the MALDI MS spectra of blood from the other animal species included in the study, the original GAPDH signal at nominal *m/z* 1750 was absent. However, a signal at nominal *m/z* 1764 was present in both the bovine and porcine blind sample spectra, and in the blood residue sampled from butcher-prepared or packaged meat, though absent from the bovine and porcine commercially available intravenous reference blood. MALDI MS/MS analysis and MASCOT search of a bovine blood sample (38 F) confirmed that this signal also originated from at GAPDH protein (**Figure 43**). The amino acid sequence of this peptide was identical to that encountered in the chicken blood, except for the substitution of the second amino acid in the chain, replacing a valine (V) with an isoleucine (I) to give the peptide sequence of L*I*SWYDNEFGYSNR, with a resulting mass difference of 14 Da. LC-MS/MS was also employed to confirm the presence of GAPDH in a porcine blood blind sample through the detection of a signal at *m/z* 885.405, a doubly charged ion, which eluted from the column at 15.65 min. The peptide at nominal *m/z* 1764 was also sequenced by MALDI MS/MS and MASCOT confirmation of GAPDH was achieved with a statistically significant probability score of 96. (Porcine GAPDH contains an identical sequence and was also given as a database match with the same score of 96).
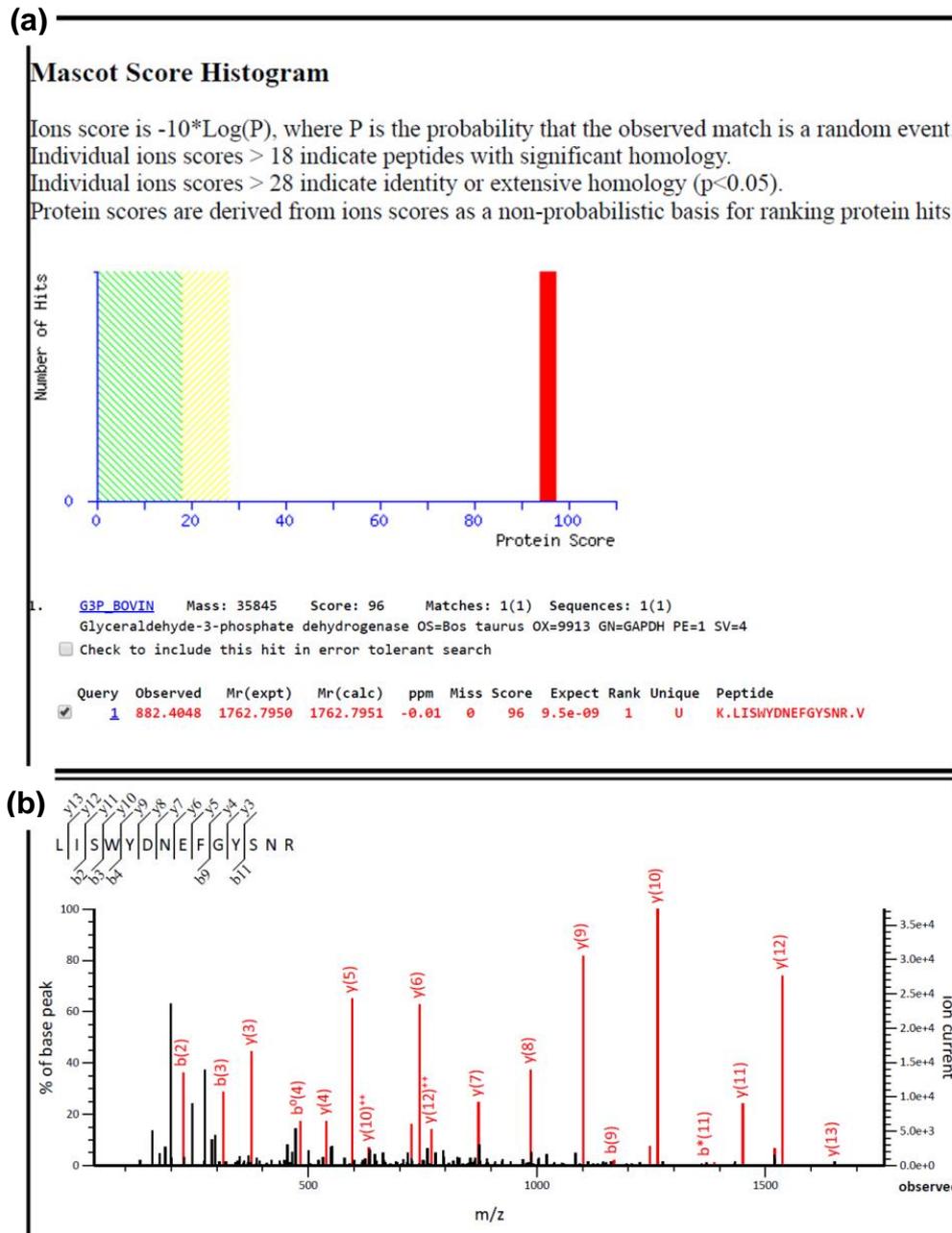
Figure 43. LC-MS/MS confirmation of the signal detected at *m/z* 1763.802 from the doubly charged ion at *m/z* 882.405 in bovine blood (sample 38 F) belonging to GAPDH. (a) MASCOT database match against bovine GAPDH for the doubly charged ion signal present at *m/z* 882.405, with a statistically significant score of 96. (b) MASCOT assignment of the ion fragments resulting from LC-MS/MS fragmentation of the doubly charged *m/z* 882.405 parent ion. *[Adapted from Kennedy et al., 2020]*.

A literature review of glyceraldehyde 3-phosphate dehydrogenase revealed that

GAPDH is generally a housekeeping gene and GAPDH mRNA has also been

used as an "internal control in experiments that measure RNA in skeletal

muscle" due to its relatively constant expression at high levels in muscular

tissue under varying conditions (**Lowe *et al.,* 2000**). GAPDH is reported to have the highest expression of any protein in chicken skeletal muscle. Investigation into the nature of the sampling methods helped to explain why GAPDH was detected in the blind samples and not the intravenous samples. It is hypothesised that for the blind samples, where blood was collected via syringe from the chest cavity of slaughtered animals, the peripheral blood mixed with intramuscular fluid, as a result of the butchering process. In contrast, the commercial reference blood was obtained directly from an incision to the jugular vein, and did not mix with intramuscular fluid. This might also account for the lack of haemoglobin peptides detected in the blind samples due to sample dilution and/or suppression effects during analysis from the intramuscular peptides.

The in-house blood protein database, updated to include GAPDH, allowed for identification of further GAPDH peptides in the bovine and porcine MALDI MS spectra. Multiple GAPDH peptides were detected in the bovine blind sample (38 S) and porcine blind sample (7 S) but not in the corresponding intravenous reference blood samples. An exception was the detection of one signal, at nominal *m/z* 1499, which was present in the blind samples, but also present in the bovine and porcine reference blood spectra, and was subsequently attributed to the GAPDH protein.

Further investigation of previously unassigned signals resulted in the putative identification of the protein pyruvate kinase (PK) in both the blind bovine and porcine MALDI MS spectra, although through the detection of different peptides for each species. The PK protein has a different sequence in bovine and porcine species, although it has a 62% homology across the two mammalian species. Similarly to the GAPDH, no PK peptides were detected in the

intravenous reference blood samples. PK is co-expressed with GAPDH in skeletal muscle tissue, and GAPDH is a critical enzyme in the production of ATP and pyruvate through anaerobic glycolysis (**Nicholls *et al.,* 2012**). This further supports the hypothesis that in the sampling of the blind samples, the blood mixed with intramuscular fluid, but in the sampling of the reference blood from the jugular vein, it did not. A similar finding was observed with myoglobin (MYO), which was detected in both bovine and porcine blind blood samples, but not in the corresponding intravenous blood samples. **Table 14** summarises the results of the putative peptide assignments highlighting in bold those signals verified by MS/MS.

Sample type and number	Protein	Peptide *m/z*	Mass accuracy (ppm)
Intravenous bovine blood	α Hb	1071.554	0.6
		1529.734	0.4
		1833.892	-0.4
	β Hb	1265.828	-2.2
		1274.723	-1.9
		1422.716	-7.6
		1477.801	-0.7
		1752.898	-0.6
		1868.951	-1.6
	GAPDH	1499.776	-8.2
	PK	1340.729	2.8
Bovine blind blood sample (38 S)	β Hb	1274.706	
	GAPDH	795.418	-3.0
		805.432	1.1
		1032.595	-6.6
		1358.681	-9.5
		1369.743	-5.5
		1499.789	-7.2

		1615.880	2.5
		1763.802	2.5
	PK	787.402	-0.6
		840.535	6.0
		868.482	-0.5
		1197.661	11.2
		1302.683	-3.4
		1447.833	-5.3
	MYO	629.343	3.9
		748.433	-3.2
		1271.668	3.8
		1393.821	3.8
		1592.838	-0.7
		1669.834	-1.3
Intravenous porcine blood	α Hb	1041.540	-3.6
		1115.631	-10.1
		1422.704	-3.2
		1628.904	-4.8
		1876.893	-2.3
		1935.972	-3.2
	β Hb	1265.830	-2.2
		1274.726	-1.9
		1449.796	-7.6
		1866.012	-7.2
	GAPDH	1499.778	7.4
Porcine blind blood sample (7 S)	GAPDH	795.422	4.3
		977.547	5.4
		1763.813	6.1
	PK	868.474	-9.6
		905.504	-6.0
		953.483	3.4
		1019.519	3.3

		1197.646	-1.5
		1433.817	-5.4
		1665.810	-3.4
		1821.919	1.5
		1859.906	3.5
		1883.913	4.9
	MYO	1592.820	-11.9

Table 14. Putative and confirmed identification of constituent peptides from the proteins haemoglobin (αHb and βHb), glyceraldehyde 3-phosphate dehydrogenase (GAPDH), pyruvate kinase (PK) and myoglobin (MYO) in bovine and porcine samples obtained from pooling carcass blood for the blind samples and directly from intravenous blood from the jugular.

The Peptide Mass Fingerprint function in MASCOT also returned a statistically significant result for creatine kinase (CK) (UniProt accession number: P12277) when inputting the peptide mass fingerprint obtained from the samples that had undergone MS/MS analysis. CK catalyses the conversion of creatine to phosphocreatine using ATP, and is also prevalent in tissue that rapidly consume ATP, such as skeletal muscle and the heart. These identifications fit the type of samples received as they were prepared from blood pooled in the chest cavity, adjacent to the heart.

The presence of these additional proteins (that were not originally included in the in-house blood protein database) could be utilised to aid species identification once a positive blood indication had been made through the detection of blood-specific peptides such as haemoglobin. They could also potentially serve to indicate the 'type' of blood involved or the manner in which it has been shed. These skeletal muscular proteins differed in their amino acid sequences between species, resulting in unique peptides masses (unique to the species composing the study). For example, the detection of the GAPDH peptide signal at theoretical *m/z* 1763.802 in suspected blood samples would

narrow the animal species down to bovine or porcine, but the detection of the GAPDH peptide at theoretical *m/z* 1749.787 would indicate the sample contained avian blood, as this was a proteotypic peptide specific to chicken. The signals at *m/z* 1592.822 and 1669.837 in the blind bovine blood samples and the locally sourced meat packaging were putatively assigned to myoglobin due to matches against myoglobin peptide *m/z* values in the in-house protein database within ±15 ppm mass tolerance. These identifications were confirmed through MS/MS analysis on both MALDI and LC MS systems. The MALDI MS/MS spectrum interpretation is displayed in **Figure 44**.
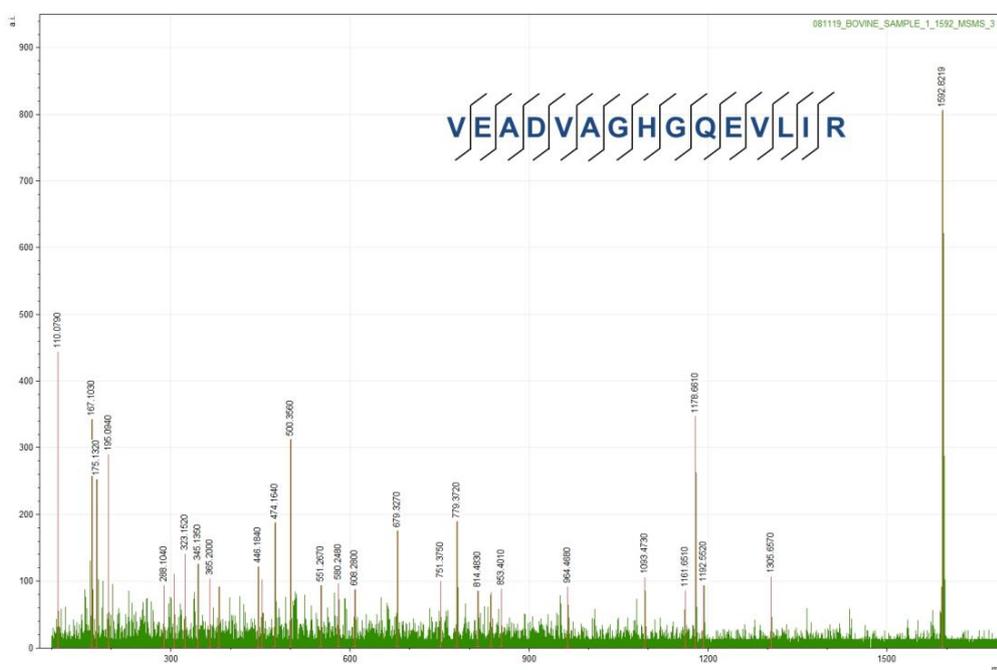
Figure 44. MALDI MS/MS spectrum of myoglobin parent ion *m/z* 1592 from a bovine blood sample. No b & y ions could be assigned through data processing of the MALDI MS/MS analysis, so LC-MS/MS was performed to confirm the sequence.

While it was possible to confirm the identity of the peptide present at *m/z* 1592.837, fragmentation of the suspected myoglobin peptide present at *m/z* 1669.837 using MS/MS did not generate sufficient *y* and *b* ions for sequence confirmation with either MALDI or LC-MS/MS. However, this signal was still

used as a putative source of identification for bovine myoglobin and consequently as a bovine blood marker.

Although the two myoglobin peptide signals (nominal *m/z* 1593 and 1670) were common to both bovine and porcine blood, due to having the same amino acid sequence, a further ion signal attributed to myoglobin was detected only in porcine blood samples at *m/z* 649. As a result, a positive identification for bovine blood was made if the myoglobin peptides appearing at nominal *m/z* 1593, 1670 and 1764 were detected. Conversely, a positive identification for porcine blood was made if myoglobin peptides at nominal *m/z* 649, 1593 and 1764 were detected. It was determined that the *m/z* 1670 signal was proteotypic to bovine blood and the *m/z* 649 signal was proteotypic to porcine blood (of the proteins and species within the scope of this study). The investigation of these additional 'non-blood' proteins allowed for further refinement of the identification strategy, as displayed in **Figure 45**.
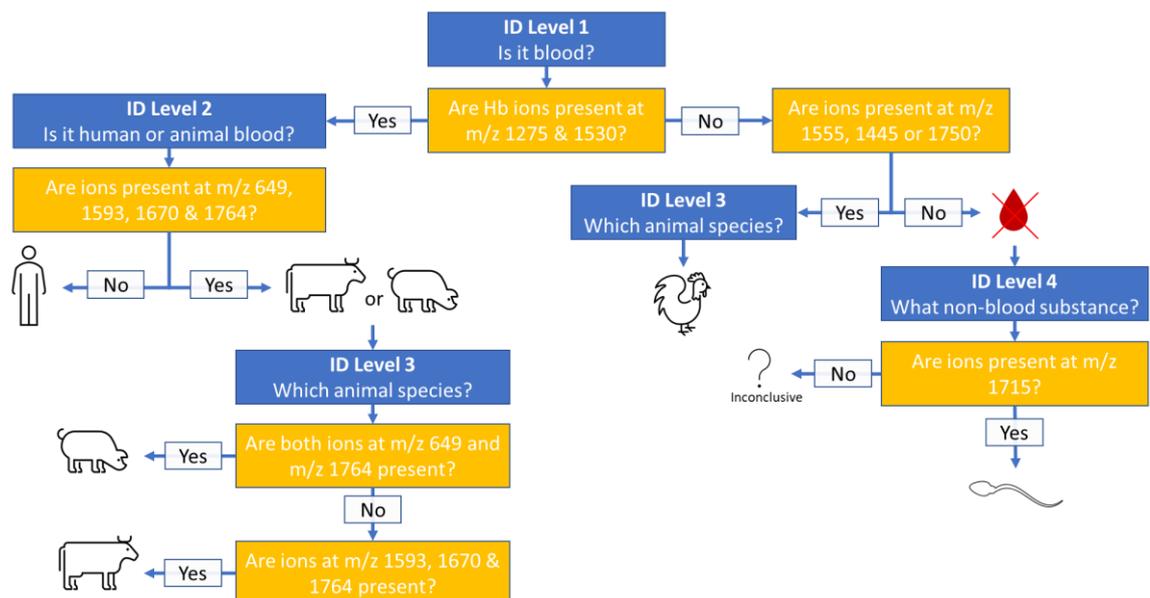
Figure 45. Refined blood and semen identification strategy. This diagram illustrates the workflow applied to the MALDI MS spectra of the blind samples to identify blood and determine provenance of blood as well as identifying semen (putatively). The *m/z*

values given are nominal, but were experimentally verified to 3 d.p. with a mass accuracy of ± 15 ppm.

This iteration allowed for positive or negative identification of human blood, positive or negative identification of animal blood, and the identification of bovine, porcine and chicken blood through the detection of proteotypic peptides, lowering the false-positive and false-negative rates compared to the strategy previously adopted. The strategy also allowed for the reliable identification of semen samples at this point, with or without prior enhancement. Furthermore, sample 32 S, which was a human semen sample, had been false-positively identified as blood, due the detection of a single signal at *m/z* 1529.73 that was putatively assigned to αHb. This was one of the first samples analysed, before the refinement of the interpretation method which later required both α and βHb peptides at nominal *m/z* 1275 and 1530 to be detected to claim the presence of blood.

Once the true identity of the blind sample 32 S had been revealed to be semen, the spectral profile was re-evaluated and a prominent peak at *m/z* 1714.85 was discovered to appear in this and other semen samples. In an attempt to identify this signal, MALDI MS/MS analysis was carried out on this precursor ion. The resulting ion product spectrum was entered into MASCOT protein database, using the 'MS/MS Ion Search' function, yielding a match against the protein semenogelin-1 (SEM-1) (UniProt accession number: P04279), with a statistically significant score of 99 (**Figure 46**).
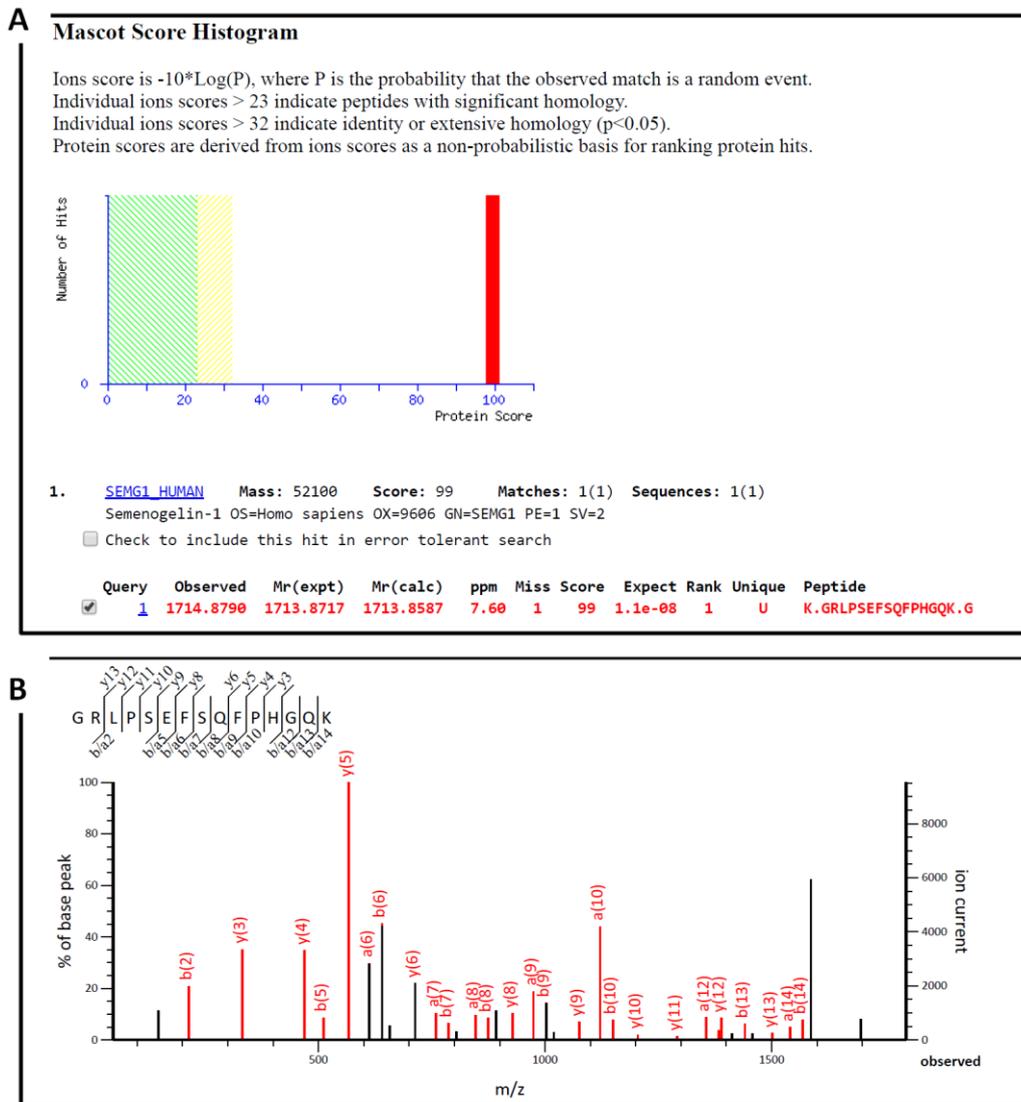
Figure 46. MALDI MS/MS identification of the ion signal at *m/z* 1714.849 present in Sample 32 containing semen. (a) MASCOT match against human semenogelin-1 with a statistically significant score of 99%. (b) MASCOT assignment of the ion product spectrum. *[Adapted from Kennedy et al., 2020]*.

Semenogelin-1 is abundant in semen and is the target of immuno-chromatographic lateral flow test strips in confirmatory semen tests such as colloidal gold-conjugated anti-human semenogelic monoclonal antibody tests, also known as Rapid Stain Identification tests (RSID) (**Old *et al.,* 2012**). The semenogelin protein was also detected by Iliano *et al.* via LC-MS/MS analysis (**Illiano *et al.,* 2018**).

Once the presence of the SEM-1 protein had been confirmed through MS/MS analysis, the protein was added to the in-house protein database after performing an *in silico* digestion in UniProt using identical parameters as for the blood proteins previously described. This then allowed for additional SEM-1 putative assignments against signals detected at *m/z* 1444.764, 1501.744 and 1801.918 (**Figure 47**). As a result, further semen proteins were investigated within the literature and additional semen-specific proteins were added to the in-house database. Those putatively assigned to signals within the spectral profiles consisted of semenogelin-2 (SEM-2) (UniProt accession number: Q02383) and prolactin-induced protein (PIP) (UniProt accession number: P12273) (**Table 15**).
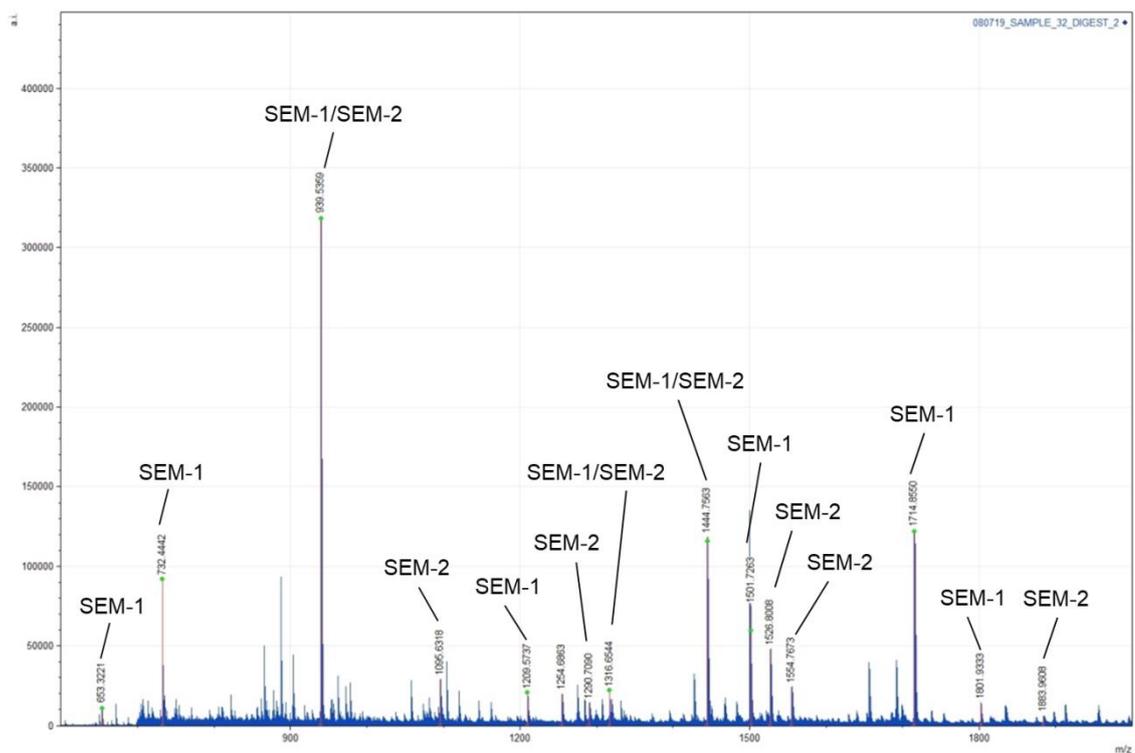
Figure 47. MALDI MS spectrum of sample 32 S, a semen sample with the semen-specific peptides annotated.

Protein	Experimental *m/z*	Theoretical *m/z*			Mass error (ppm)		
SEM-1	653.322	653.327			-7.7		
	732.444	732.452			-10.9		
	939.536	939.541			-5.3		
	1209.574	1209.572			1.7		
	1316.654	1316.659			-3.8		
	1444.756	1444.754			1.4		
	1501.726	1501.743			-11.3		
	1714.855	1714.866			-6.4		
	1801.933	1801.942			-5.0		
SEM-2	939.536	939.541			-5.3		
	1095.632	1095.642			-9.1		
	1290.708	1290.691			13.2		
	1316.654	1316.659			-3.8		
	1444.756	1444.754			1.4		
	1526.801	1526.781	1526.817	1526.817	13.1	-10.5	-10.5
	1554.767	1554.766			0.6		
	1883.961	1883.936			13.3		
PIP	1254.686	1254.698			-9.6		

Table 15. The most abundant semen proteins detected in the blind semen samples, with an example of their experimental *m/z* values compared to their theoretical *m/z* values calculated *in silico*. SEM-2 has 3 peptides within ±15 ppm of one of the experimental values detected, so all three are listed. The first (*m/z* 1526.781) is different from the other two by a single substitution (Q>K) and the second and third (*m/z* 1526.817) are identical in amino acid sequence but appear in different sections of the peptide (SEM-2 [312-325] and SEM-2 [432-445]).

These proteins were reported elsewhere as biomarkers for semen detection via both mass spectrometric approaches (**Thacker *et al.,* 2011, Illiano *et al.,* 2018, Agarwal *et al.,* 2020)** and also using an mRNA based approach (**Juusola & Ballantyne, 2003, Bauer, 2007 and Gomes *et al.,* 2011**). An example spectrum is shown in **Figure 48**, with putatively assigned SEM-1 and SEM-2 signals annotated. Two putative βHb signals are also annotated. The suspected

semen protein signal at nominal *m/z* 1715 underwent MALDI MS/MS analysis for sequencing confirmation (**Figure 48,** *inset*).
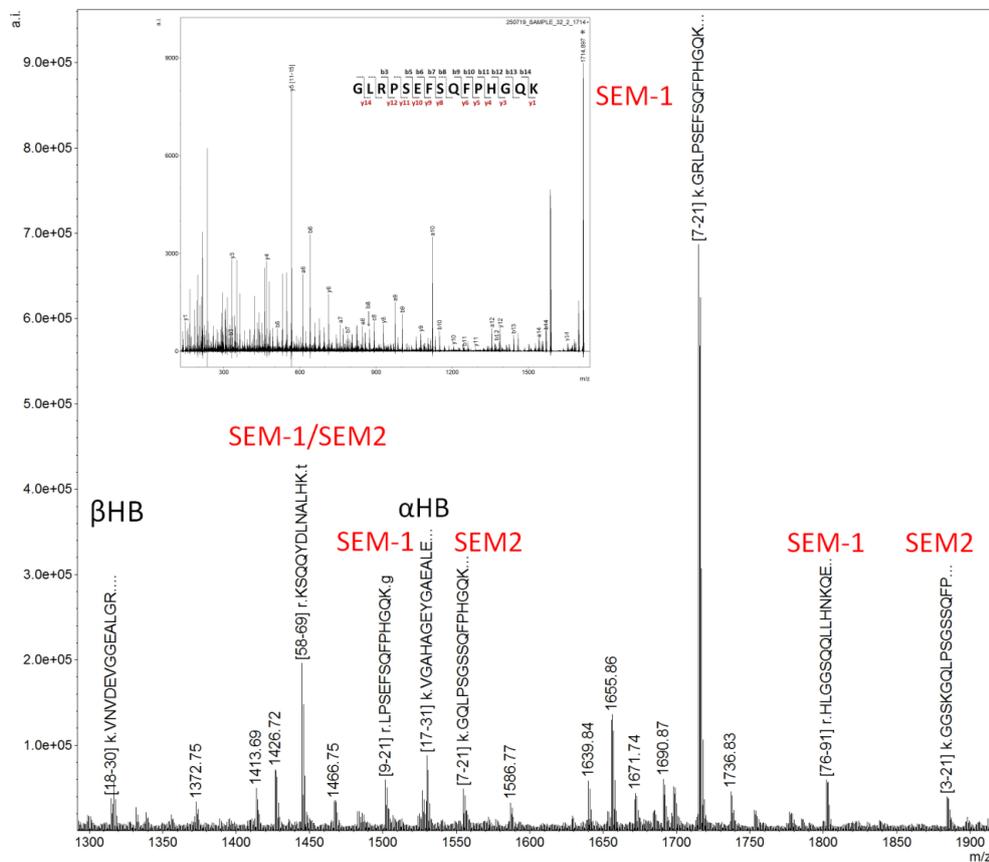
Figure 48. MALDI MS spectrum of blind sample 32 S containing semen. Semenogelin-1 and semenogelin-2 peptides, in addition to βHb were putatively assigned. The ion signal present at *m/z* 1714.849 was confirmed through MALDI MS/MS analysis (inset). *[Adapted from Kennedy et al., 2020]*.

However, the presence of ion signals at *m/z* 1314.682 (βHb) and 1529.734 (αHb) triggered the speculation that blood traces were present as generated by the collection of semen for the preparation of the blind samples. In general terms, haemoglobin may also be detected in semen as a result of certain medical conditions or in saliva as a result of a cut to the mouth or tongue or from bleeding gums. Furthermore, at genuine crime scenes, stains may be a mixture of blood and other biofluids. As such, any proposed method for identification needs to be robust enough to correctly identify a mixture if present

and this should be the subject of further investigation. Furthermore, MS/MS analysis should always be employed to confirm putatively identified biomarkers of blood or other biofluids.

As previously described, Sample 32 was the first blind semen sample to be analysed, but incorrectly classified due to the detection of a single haemoglobin peptide. The iteration of the method utilising the semen-specific proteins led to the correct identification of all four subsequent semen samples as shown in **Table 16**.

Sample #	Identification			Enhancement reagent	Actual identification
	Level 1	Level 2	Level 3		
32 S	Blood	Human	n/a	LCV -ve	Semen
Following discovery of SEM-1 marker					
27 S	Non-blood	n/a	n/a	n/a	Semen
58 S	Non-blood	n/a	n/a	AB1 +ve	Semen
132 F	Non-blood	n/a	n/a	LCV +ve	Semen
158 F	Non-blood	n/a	n/a	AB1 +ve	Semen

Table 16. Putative assignments of blind semen samples before and after the discovery of the SEM-1 marker at *m/z* 1715. S = stain, F = Fingermark, LCV = Leucocrystal Violet, AB1 = Acid Black 1, +ve = BET reaction was positive, -ve = BET reaction was negative.

Importantly, samples 58 S, 132 F and 158 F were previously enhanced using AB1 or LCV which produced a false-positive for blood and this false positive was rectified by the application of MALDI MS as a confirmatory test.

The refinement of the method allowed for re-analysis of samples that had previously been incorrectly classified which were now yielding correct semen identifications. For the re-analysis, samples were re-extracted from the original blind sample where available, under the same enzymatic digestion conditions

previously employed in the study. The identifications for a total of 57 samples, using the refined strategy, are reported in **Table 17**.

Sample #	Identification			Enhancement reagent	Actual identification
	Level 1	Level 2	Level 3		
1 S	Blood	Animal	Bovine	n/a	Bovine blood
2 S	Blood	Human	n/a	n/a	Human blood
3 S	Blood	Human	n/a	n/a	Human blood
4 F	**Blood**	**Animal**	**Chicken**	**n/a**	**Chicken blood**
5 S	Blood	Animal	Bovine	AB1 +ve	Bovine blood
6 S	Blood	Human	n/a	n/a	Human blood
7 S	Blood	inconclusive	inconclusive	n/a	Porcine blood
12 S	Non-blood	n/a	n/a	n/a	Human sweat
13 S	Non-blood	n/a	n/a	LCV +ve	Porcine blood
14 S	Non-blood	n/a	n/a	AB1 +ve	Human blood
16 S	Blood	Human	n/a	n/a	Human blood
17 S	Blood	Human	n/a	n/a	Human blood
18 S	Blood	Animal	Porcine	n/a	Wild boar blood
24 S	**Non-blood**	**n/a**	**n/a**	**n/a**	**Egg**
25 S	**Blood**	**Animal**	**Bovine**	**LCV +ve**	**Bovine blood**
26 S	Non-blood	n/a	n/a	LCV -ve	Human saliva
27 S	Non-blood	n/a	n/a	n/a	Human semen
28 S	Blood	Human	n/a	n/a	Bovine blood
29 S	Non-blood	n/a	n/a	AY7 +ve	Chicken blood
30 S	Non-blood	n/a	n/a	AB1 +ve	Egg yolk
31 S	Blood	Human	n/a	LCV +ve	Human blood
32 S	Non-blood	n/a	n/a	LCV -ve	Human semen
34 S	Blood	Human	n/a	LCV +ve	Human blood (+EDTA)
35 S	Non-blood	n/a	n/a	n/a	Chicken blood

36 S	Non-blood	n/a	n/a	n/a	Ketchup
37 S	Blood	Human	n/a	n/a	Human blood
40 F	Non-blood	n/a	n/a	n/a	Human sweat
41 S	Non-blood	n/a	n/a	AB1 +ve	Ketchup
41 S	**Non-blood**	**n/a**	**n/a**	**AB1 +ve**	**Ketchup**
44 S	**Blood**	**Animal**	**Porcine**	**n/a**	**Wild boar blood**
49 S	Non-blood	n/a	n/a	AB1 +ve	Human saliva
53 S	Blood	Human	n/a	n/a	Human blood
53 S	**Blood**	**Human**	**n/a**	**n/a**	**Human blood (+EDTA)**
56 S	Non-blood	n/a	n/a	n/a	Porcine blood
57 F	Non-blood	n/a	n/a	n/a	Paint
58 S	Non-blood	n/a	n/a	AB1 +ve	Human semen
58 S	**Non-blood**	**n/a**	**n/a**	**AB1 +ve**	**Human semen**
59 S	Blood	Human	n/a	n/a	Human blood
60 S	Non-blood	n/a	n/a	AY7 +ve	Human saliva
61 S	Blood	Human	n/a	n/a	Human blood
63 S	Non-blood	n/a	n/a	LCV +ve	Chicken blood
72 S	**Non-blood**	**n/a**	**n/a**	**n/a**	**Steak sauce**
76 S	**Non-blood**	**n/a**	**n/a**	**AY7 -ve**	**Human semen**
78 S	Non-blood	n/a	n/a	n/a	Lotion Gold Bond
79 S	Non-blood	n/a	n/a	n/a	Blank
107 F	**Blood**	**Animal**	**Porcine**	**n/a**	**Porcine blood**
122 F	Blood	Human	n/a	AY7 -ve	Human blood
132 F	Non-blood	n/a	n/a	LCV -ve	Human semen
141 F	Non-blood	n/a	n/a	AB1 +ve	Ketchup
144 F	**Non-blood**	**n/a**	**n/a**	**n/a**	**Blank**

158 F	Non-blood	n/a	n/a	AB1 +ve	Human semen
160 F	Non-blood	n/a	n/a	AY7 -ve	Human saliva
162 F	Blood	Human	n/a	AB1 +ve	Human blood
165 F	Non-blood	n/a	n/a	AY7 +ve	Egg white
170 F	**Non-blood**	**n/a**	**n/a**	**n/a**	**Human blood**
175 F	Non-blood	n/a	n/a	LCV -ve	Egg white

Table 17. MALDI MS based identification of a batch of 57 blind samples interpreted using the refined proteomic strategy. Forty-five of the samples had been analysed previously but the actual identities not disclosed. Samples 13 and 14 had been completely consumed when previously swabbed and extracted, therefore extracts were digested again; the other samples were prepared from the original material provided. The twelve new samples that had not been analysed previously are highlighted in bold. The final column, 'actual identification' is colour-coded green if the blood and species assignment was correct, orange if the blood assignment was correct but the animal species was incorrect, and red if the blood/non-blood assignment was incorrect. S = stain, F = fingermark. AB1 = Acid Black 1, AY7 = Acid Yellow 7 and LCV = Leucocrystal Violet. +ve/-ve indicates whether the sample reacted positively/negatively with the reagent.

These samples comprise of the 45 samples cumulatively reported in **Tables 13 and 16** with some samples being re-analysed, and 12 samples from the original blind batch not previously analysed, using the refined method.

No false-positives for human or animal blood were reported following data processing and peak assignments using the final iteration of the method. However, a false-negative was reported for sample 170 which was human blood, as it did not exhibit any of the characteristic blood peptides, including haemoglobin signals, despite a complex peptide mass fingerprint. When the sample identity was revealed, details on the sample preparation elucidated that this was the last deposition in a depletion series, where blood was deposited on a fingertip and allowed to almost dry, then a series of marks was made by the same finger. As a result of the depletion, the final mark was not visible to the naked eye, and it is plausible that there was insufficient material to allow for detection above the instrument's limit of sensitivity.

Compared to the initial false-negative rate of 78% for animal blood samples, the false negative rate was reduced to just 7%. This accounted for just a single porcine blood sample where blood-derived peptides were not detected. This is likely a result of a suboptimum digestion and/or clean up procedure, as opposed to the MALDI instrument being unable to detect signals that were present, as they were detected in all further animal blood samples.

No false-positives were reported for the non-biofluid samples either (comprising of beetroot juice, paint, ketchup, egg). No false-positives or false negatives were yielded by the method for the blind human semen samples analysed.

Saliva and sweat identification - In a further iteration of the method, spectral data from those samples correctly identified as non-blood biofluids were interrogated. In samples revealed to contain sweat, the putative presence of proteins clusterin and serum albumin was detected with ±15 ppm mass accuracy. These proteins were previously detected in sweat as reported by the literature (**Espinoza, 1999, Legg, 2014**). However these proteins are not specific to sweat and have also been reportedly been found in other tissues (**Zhang *et al.,* 2003**). Biomarkers specific to sweat are dermcidin- and psoriasin-derived peptides, which have been reported elsewhere in the literature (**Ferguson *et al.,* 2012, Patel *et al.,* 2016**). However, when the spectral peak lists were compared against *in silico* digest mass lists containing peptides originating from these proteins, these signals were not present in the corresponding mass spectra. One sample may have indicated the presence of a single psoriasin-derived peptide, but this was not confirmed through MS/MS analysis, given that it was only detected in one sample.

A similar literature review of proteins detected in saliva revealed α-amylase 1 as a specific marker for saliva (**Kamanna *et al.,* 2016, Illiano *et al.,* 2018, Oonk *et al.,* 2018, Duong *et al.,* 2021**). In fact, α-amylase 1 is the target of commercial immunochromatographic saliva testing kits, such as the RSID™ saliva flow test and ALIgAE® test (**Chatterjee, 2019**) (as opposed to alpha amylase 2, which is found in semen and vaginal secretions) (**Chatterjee, 2019**). An *in silico* digestion was again performed for this protein, yielding only one putative match for the peptide signal at *m/z* 1697.854 (mass accuracy of -1.4 ppm). Whilst this peptide was also detected by Illiano *et al.,* it was not observed in any other saliva blind samples.

In conclusion, at this stage, it is not possible to definitively identify/discriminate saliva and sweat samples utilising the current strategy. These two biofluids (and corresponding samples) were therefore excluded from the final iteration of the method development.

Final validation - A final blind sample batch was analysed to validate the refined interpretative strategy including multiple biomarkers to discriminate human blood, animal blood (down to species level) and human semen samples from non-biofluids. Thirtheen samples were selected blind to the analysist comprising of enhanced and non-enhanced stains and fingermarks. Twelve out of the 13 final validation samples were correctly identified, down to species level when animal blood was present. A summary of the putative identifications of the final validation batch is displayed in **Table 18**.

Sample No.	Enhancement reagent	Putative Identification	Actual Identification
15 S	n/a	Human Blood	Human Blood
22 S	AY7 +ve	Human Blood	Human Blood
104 F	n/a	Chicken Blood	Chicken Blood
121 F	n/a	Not Blood	Saliva
127 F	n/a	Semen	Semen
128 F	n/a	Bovine Blood	Bovine Blood
129F	AY7 +ve	Chicken Blood	Chicken Blood
138 F	AY7 +ve	Animal blood - species inconclusive	Bovine Blood
146 F	n/a	Not Blood	Beetroot juice
147 F	n/a	Semen	Semen
155 F	AY7 +ve	Porcine Blood	Porcine Blood
156 F	n/a	Porcine Blood	Porcine Blood
176 F	AY7 +ve	Not Blood	Semen

Table 18. Validation of the refined interpretative strategy for the identification of further 13 blind samples. The final column, 'actual identification' is colour-coded green if the blood and species assignment was correct or orange if the blood assignment was correct but the species was incorrect or inconclusive. Sample 138 F, which was identified as blood, but could not identify the species, was inconclusive. AY7 = Acid Yellow 7, +ve indicates the BET reaction was positive. [*Adapted from Kennedy et al., 2020*].

One of the samples (138 F) was correctly identified as animal blood through the detection of a myoglobin peptide signal present at *m/z* 1592.825. This myoglobin peptide is common to both bovine and porcine species, so definitive species identification was not possible due to the lack of further proteotypic peptides to match against the protein database. The spectral peak list was checked against the refined in-house protein database but did not match within a mass tolerance of ±15 ppm to any other blood-derived peptides. In the case of sample 176 F, which had been indicated as blood through a positive result with

AY7, the method was able to correctly refute this, and was correctly identified as non-blood by MALDI MS analysis.

Conclusion

The aim of this study was to further develop and validate the MALDI-based proteomic approach for the detection of blood and the determination of blood provenance. The application of MALDI MS in this forensic context carries several benefits over other analytical techniques. For example, whilst colorimetric blood enhancement techniques are rapid, they suffer from a low level of objectivity and scarce robustness. MALDI MS offers a greater level of specificity through the identification of multiple constituent peptides. The more proteotypic peptides detected, the more reliable the assignment. MS/MS analysis can also definitively confirm the identity of a protein through *De Novo* sequencing. MALDI MS is also less destructive than conventional mass spectrometric approaches, such as LC-MS, where a portion of the sample must be swabbed and extracted. Although MALDI MS cannot be considered a non-destructive technique, it is only partially destructive, as a few micrometres of material are ablated from the sample surface. This circumstance allows the retention of the integrity of the sample, which can then undergo further enhancement or analysis, particularly critical for evidence types such as bloody fingermarks, where destruction of the *minutiae* would be undesirable. Compared to LC-MS, data acquisition time is much faster (few seconds versus several minutes) (**Illiano *et al.,* 2018**). Furthermore, MALDI MS can acquire spectra from up to 96 separate positions very rapidly, from each well on a sample plate without ejecting the plate, a process which can even be automated. Finally, the data interpretation can be faster and less complex for

MALDI MS spectral data compared to other types of mass spectrometric analysis without having to account for multiple charge states. MALDI MS predominantly detects singly charged ions, which makes interpretation of spectral data easier. LC-MS is also more susceptible to contamination from the glassware used to prepare the mobile phase and carryover within the system, which is not applicable to MALDI.

The samples selected for this study were chosen to represent evidence types that are encountered at crime scenes but that might be mistaken for human blood. These included animal blood (from bovine, porcine, wild boar and chicken), as well as other human biofluids; semen, saliva and sweat. In addition, some interferents were chosen that would either yield false-positive results to conventional blood enhancement reagents Acid Black 1, Acid Yellow 7 or Leucocrystal Violet, or looked visually similar to blood (egg yolk and egg white, beetroot juice, steak sauce, lotions, ketchup and paint). These samples were comprised of either stains or contaminated fingermarks, where a quantity of the substance was applied to the fingertip before being deposited on cleaned aluminium TLC plates. These were prepared and analysed blind so as not to bias the analytical interpretation. The majority of the experimentation and method development was performed on *in solution* digests that had been swabbed and extracted from the samples. However, the method was also demonstrated to perform effectively on samples *in situ*, by spraying enzyme and matrix directly onto the sample. This is useful for profiling for rapid spot analysis (determining the presence and species of origin of blood or other biofluids at a single point), as established in this study. It also has applications for imaging analysis of small areas of sample, up to a fingerprint in size, where

contaminants present within the mark can be visualised, as demonstrated in Chapter 4.

This work has discussed the ability of a MALDI MS and proteomics-based approach combined with a refined interpretation workflow to discriminate blood from non-blood, discriminate between human and animal blood and identify the animal species from the subset selected for this study. It has also shown the potential for the method to discriminate other biofluids, using semen as an example, in the presence of blood enhancement reagents AB1, AY7 and LCV. In most cases, MALDI MS analysis in combination with bottom up proteomics correctly refuted the results of the application of presumptive BETs yielding false- positives for blood. The technique robustly identified human blood from the outset, using the two prominent αHb and βHb peptides at nominal *m/z* 1275 and 1530.

The initial hypothesis that the blood from animal species could be identified through their corresponding abundant haemoglobin peptide signals was not eventually applicable as mammalian species have similar, and in some cases identical, peptide sequences for haemoglobin and other blood-derived proteins (for example the identical LLVVYPWTQR sequence in human and bovine β haemoglobin).

It was also interesting to discover that the way in which animal blood was sampled had an effect on the peptide mass fingerprint generated. The initial experiment design anticipated blood from the same species to yield almost identical spectra, regardless of the sampling method. However, this study revealed that sampling technique had a significant impact on the peptides and proteins detected in blood samples. The most notable finding was that GAPDH,

not a blood-derived protein but a skeletal muscular protein, was abundant in the blind samples obtained from the chest cavity of slaughtered animals but absent in the samples obtained directly from the vein. This allowed for the sampling methods to be discriminated but might also applications in criminal investigations. The presence of GAPDH was detected in samples where the animals had been slaughtered and butchered so may be an indication that an extremely violent incident has taken place and skeletal muscle has been severed. If it could be confirmed that the blood was of human origin and the presence of GAPDH was also detected, it could be an indication that a person had sustained severe injuries, as opposed to a minor superficial injury. However, as the human blood samples used in this study had been obtained ethically, and no GAPDH was detected, this hypothesis requires further investigation. A possible source of human blood potentially containing GAPDH would be that collected as a result of surgery, where a surgeon has made an incision through the skin and skeletal muscle, which would be required to test this hypothesis. Further work could involve analysing the blood-soaked gauze and bandages used to treat gunshot or knife stabbing victims vs. blood sampled from fingerprick lancets for example. This method combined with a chemometrics approach might discover further differences between the proteinous species generated under different bloodshed conditions.

Another interesting finding was that additional proteins to those identified by Patel *et al.* (**Patel *et al.*, 2016**) were detected in the species within the scope of this study, following *in solution* and *in situ* digestion, although the authors has compared human and equine blood. The species in the present study had complex peptide mass fingerprints comprised of many more blood-derived proteins such as myoglobin, GAPDH and creatine kinase. Further work in this

area could involve assigning all of the protein-derived signals for these species to try and determine additional proteotypic peptides to aid specific species identification.

Finally, all of the samples were generated on the same 'ideal' surface under laboratory conditions and not exposed to difficult environmental conditions. Further work should be undertaken to assess the robustness of the proposed method utilising a range of different deposition surfaces for both stains and fingermarks, a more extensive range of presumptive BETs, and more challenging environmental conditions such as heat and/or humidity, that evidential items might encounter before recovery by CSIs.

The work described in this chapter was the subject of an article published in Scientific Reports, entitled 'Pre-validation of a MALDI MS proteomics-based method for the reliable detection of blood and blood provenance' of which the author of this chapter is a joint first author (**Kennedy *et al.,* 2020**).

References

Agarwal, A., Panner Selvam, M. K., & Baskaran, S. (2020). Proteomic Analyses of Human Sperm Cells: Understanding the Role of Proteins and Molecular Pathways Affecting Male Reproductive Health. *International Journal of Molecular Sciences*, 21(5), 1621–1682.

Bandey, H. L., Bleay, S. M., Bowman, V. J., Downham, R. P., Sears, V. G., (2014). Fingermark Visualisation Manual, *The Home Office*, 1-932

Bauer, M. (2006). RNA in forensic science. *Forensic Science International: Genetics, 1*(1), 69-74.

Bleay, S., Sears, V., Downham, R., Bandey, H., Gibson, A., Bowman, V., Fitzgerald, L., Ciuksza, T., Ramadani, J., Selway, C., (2017). Fingerprint Source Book v2.0 (second edition), *CAST Publication 081/17*. 1-666.

Boyd, S., Bertino, M. F., Seashols, S. J., (2011) Raman spectroscopy of blood samples for forensic applications. Forensic Sci. Int. 208, 124–128.

Bradshaw, R., Bleay, S., Clench, M. ., & Francese, S. (2013). Direct detection of blood in fingermarks by MALDI MS profiling and Imaging. *Science & Justice*, *54*(2), 110–117.

Cadd, S., Li, B., Beveridge, P., O'Hare, W. T., Campbell, A., & Islam, M. (2016a). Non-contact detection and identification of blood stained fingerprints using visible wavelength reflectance hyperspectral imaging: Part 1. *Science & Justice*, *56*(3), 181–190.

Cadd, S., Li, B., Beveridge, P., O'Hare, W. T., Campbell, A., & Islam, M. (2016b). The non-contact detection and identification of blood stained

fingerprints using visible wavelength hyperspectral imaging: Part II effectiveness on a range of substrates. *Science & Justice*, *56*(3), 191–200.

Cadd, S., Li, B., Beveridge, P., O'Hare, W. T., & Islam, M. (2018). Age determination of blood-stained fingerprints using visible wavelength reflectance hyperspectral imaging. *Journal of Imaging*, *4*(12), 141–152.

Chatterjee. S., (2019). Saliva as a forensic tool. *Journal of Forensic Dental Sciences*, 11(1), 1–4.

Creamer, J. I., Quickenden, T. I., Apanah, M. V., Kerr, K. A., & Robertson, P. (2003). A comprehensive experimental study of industrial, domestic and environmental interferences with the forensic luminol test for blood. *Luminescence*, *13*(4), 193–198.

De Almeida, J. P., Glesse, N. & Bonorino, C. (2011) Effect of presumptive tests reagents on human blood confirmatory tests and DNA analysis using real time polymerase chain reaction. *Forensic Sci. Int.* 206, 58–61

Duong, V. A., Park, J. M., Lim, H. J., & Lee, H. (2021). Proteomics in forensic analysis: Applications for human samples. *Applied Sciences*, *11*(8), 3393–3427.

Edelman, G. J., Gaston, E., van Leeuwen, T. ., Cullen, P. ., & Aalders, M. C. . (2012). Hyperspectral imaging for non-contact analysis of forensic traces. *Forensic Science International (Online)*, *223*(1-3), 28–39.

Espinoza, E. O., Lindley, N. C., Gordon, K. M., Ekhoff, J. A. & Kirms, M. A. (1999) Electrospray Ionization Mass Spectrometric Analysis of Blood for Differentiation of Species. *Anal. Biochem.* 268, 252–261.

Ferguson, L.S., Wulfert, F., Wolstenholme, R., Fonville, J. M., Clench, M. R., Carolan, V. A., & Francese, S. (2012). Direct detection of peptides and small proteins in fingermarks and determination of sex by MALDI mass spectrometry profiling. *Analyst* (London), 137(2), 4686–4692.

Francese. S., (2019). Criminal profiling through MALDI MS based technologies - breaking barriers towards border-free forensic science. *Australian Journal of Forensic Sciences*, 51(6), 623–635.

Gannicliffe, C., (2019) Chapter 6 - End User Commentary on Bioanalytical Advancements in the Reliable Visualization and Discrimination of Bodily Fluids. 103-107 IN: *Emerging Technologies for the Analysis of Forensic Traces*, Francese, S., (Ed), Springer Nature Switzerland

Gomes, I., Kohlmeier, F., & Schneider, P. M. (2011). Genetic markers for body fluid and tissue identification in forensics. *Forensic Science International: Genetics Supplement Series, 3*(1), 469-470.

Hanson, E. K. & Ballantyne, J. (2010). A blue spectral shift of the hemoglobin soret band correlates with the age (time since deposition) of dried bloodstains. *PloS One, 5*(9), 12830–12830.

Illiano, A., Arpino, V., Pinto, G., Berti, A., Verdoliva, V., Peluso, G., Pucci, P. and Amoresano, A. (2018). Multiple Reaction Monitoring Tandem Mass Spectrometry Approach for the Identification of Biological Fluids at Crime Scene Investigations. *Analytical Chemistry*, 90(9), 5627-5636.

Jakubowska, J., Maciejewska, A., Pawłowski, R., & Bielawski, K. (2013). mRNA profiling for vaginal fluid and menstrual blood identification. *Forensic Science International: Genetics, 7*(2), 272-278.

Jiang, Y., Sun, J., Huang, X., Shi, H., Xiong, C., & Nie, Z. (2019). Direct identification of forensic body fluids by MALDI-MS. *Analyst (London)*, *144*(23), 717–723.

Juusola, J., & Ballantyne, J. (2003). Messenger RNA profiling: A prototype method to supplant conventional methods for body fluid identification. *Forensic Science International*, 135(2), 85-96.

Kamanna, S., Henry, J., Voelcker, N. H., Linacre, A., & Kirkbride, K. P. (2016). Direct identification of forensic body fluids using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *International Journal of Mass Spectrometry*, *397-398*, 18–26.

Kamanna, S., Henry, J., Voelcker, N. H., Linacre, A., & Paul Kirkbride, K. (2017). A mass spectrometry-based forensic toolbox for imaging and detecting biological fluid evidence in finger marks and fingernail scrapings. *International Journal of Legal Medicine*, *131*(5), 1413–1422.

Kamanna, S., Henry, J., Voelcker, N., Linacre, A., & Paul Kirkbride, K. (2018). A complementary forensic "proteo-genomic" approach for the direct identification of biological fluid traces under fingernails. *Analytical and Bioanalytical Chemistry*, *410*(24), 6165–6175.

Kennedy, K., Heaton, C., Langenburg, G., Cole, L., Clark, T., Clench, M. R., Sears, V., Sealey, M., McColm, R., & Francese, S. (2020). Pre-validation of a MALDI MS proteomics-based method for the reliable detection of blood and blood provenance. *Scientific Reports*, *10*(1), 17087–17087.

Kennedy, K., Bengiat, R., Heaton, C., Herman, Y., Oz, C., Elad, M. L., Cole, L., & Francese, S. (2021). "MALDI-CSI": A proposed method for the tandem

detection of human blood and DNA typing from enhanced fingermarks. *Forensic Science International*, *323*, 110774–110774.

Lee, H. C., & Pagliaro, E. M., (2000) SEROLOGY | Blood Identification *in* Seigel, J. (ed-in-chief) *Encyclopedia of Forensic Sciences*, London: Academic Press. 1331-1338.

Legg, K., Powell, R., Reisdorph, N., Reisdorph, R., & Danielson, P. (2014). Discovery of highly specific protein markers for the identification of biological stains. *ELECTROPHORESIS*, 35(21-22), 3069-3078.

Lemler, P., Premasiri, W. R., DelMonaco, A., Ziegler, L. D., (2014) NIR Raman spectra of whole human blood: Effects of laser-induced and in vitro hemoglobin denaturation. *Anal. Bioanal. Chem.* 406, 193–200.

Li, B., Beveridge, P., O'Hare, W. T., & Islam, M. (2014). The application of visible wavelength reflectance hyperspectral imaging for the detection and identification of blood stains. *Science & Justice*, *54*(6), 432–438.

Li, Z., Bai, P., Peng, D., Wang, H., Guo, Y., Jiang, Y., He, W., Tian, H., Yang, Y., Huang, Y., Long, B., Liang, W., & Zhang, L. (2017). Screening and confirmation of microRNA markers for distinguishing between menstrual and peripheral blood. *Forensic Science International : Genetics*, *30*, 24–33.

Lowe, D. A., Degens, H., Chen, K. D., & Alway, S. E. (2000). Glyceraldehyde-3-phosphate dehydrogenase varies with age in glycolytic muscles of rats. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, *55*(3), 160-164.

Mistek, E., & Lednev, I. K. (2015). Identification of species' blood by attenuated total reflection (ATR) Fourier transform infrared (FT-IR) spectroscopy. *Analytical and Bioanalytical Chemistry*, *407*(24), 7435–7442.

Mistek, E., & Lednev, I. K. (2020). Discrimination between human and animal blood by attenuated total reflection Fourier transform-infrared spectroscopy. *Communications Chemistry*, *3*(1), 1–6.

Nascimento, G. M. d., (Ed.). (2018). Raman Spectroscopy. *IntechOpen*.

Nicholls, Li, H., & Liu, J.-P. (2012). GAPDH: A common enzyme with uncommon functions. *Clinical and Experimental Pharmacology & Physiology*, *39*(8), 674–679.

Old, J., Schweers, B. A., Boonlayangoor, P. W., Fischer, B., Miller, K. W. P., & Reich, K. (2012) Developmental Validation of RSIDTM-Semen: A Lateral Flow Immunochromatographic Strip Test for the Forensic Detection of Human Semen. *J. Forensic Sci.* 57, 489–499.

Oonk, S., Schuurmans, T. ., Pabst, M., de Smet, L. C. P. ., & de Puit, M. (2018). Proteomics as a new tool to study fingermark ageing in forensics. *Scientific Reports*, *8*(1), 16425–11.

Patel, E., Cicatiello, P., Deininger, L., Clench, M. R., Marino, G., Giardina, P., Langenburg, G., West, A., Marshall, P., Sears, V., & Francese, S. (2016). A proteomic approach for the rapid, multi-informative and reliable identification of blood. *Analyst* (London), 141(1), 191–198.

Quickenden, T. I., & Creamer, J. I. (2001). A study of common interferences with the forensic luminol test for blood. *Luminescence, 16*(4), 295–298.

Ramotowski. (2013). *Lee and Gaensslen's advances in fingerprint technology* (3rd ed.). CRC Press.

Rankin-Turner, S., Turner, M. A., Kelly, P. F., King, R. S. P., & Reynolds, J. C. (2019). Transforming presumptive forensic testing: in situ identification and age estimation of human bodily fluids. *Chemical Science (Cambridge)*, *10*(4), 1064–1069.

Reed. N. G., (2010). The History of Ultraviolet Germicidal Irradiation for Air Disinfection. *Public Health Reports (1974)*, *125*(1), 15–27.

Roeder, A. D., & Haas, C. (2013). mRNA profiling using a minimum of five mRNA markers per body fluid and a novel scoring method for body fluid identification. *International Journal of Legal Medicine*, *127*(4), 707–721.

Seraglia, R., Teatino, A., & Traldi, P. (2004). MALDI mass spectrometry in the solution of some forensic problems. *Forensic Science International*, *146*, S83–S85.

Souza, G. H. M. F., Guest, P. C., & Martins-de-Souza, D. (2016). LC-MSE, Multiplex MS/MS, Ion Mobility, and Label-Free Quantitation in Clinical Proteomics. *Methods in Molecular Biology (Clifton, N.J.)*, *1546*, 57–73.

Strohalm, M., Hassman, M., Kosata, B. & Kodicek, M. (2008) mMass dataminer: An open source alternative for mass spectrometric data analysis. *Rapid Commun. Mass Spectrom.* 22, 905–908.

Strohalm, M., Kavan, D., Novak, P., Volny, M. & Havlicek, V. (2010) mMass 3: A cross-platform software environment for precise analysis of mass spectrometric data. *Anal. Chem.* 82, 4648–4651.

Thacker, Yadav, S. P., Sharma, R. K., Kashou, A., Willard, B., Zhang, D., & Agarwal, A. (2011). Evaluation of Sperm Proteins in Infertile Men: A Proteomic Approach. *Fertility and Sterility*, *95*(8), 2745–2748.

Yang, H., Zhou, B., Deng, H., Prinz, M., & Siegel, D. (2013). Body fluid identification by mass spectrometry. *International Journal of Legal Medicine*, *127*(6), 1065–1077.

Zhang, L.-Y., Ying, W.-T., Mao, Y.-S., He, H.-Z., Liu, Y., Wang, H.-X., Liu, F., Wang, K., Zhang, D.-C., Wang, Y., Wu, M., Qian, X.-H., & Zhao, X.-H. (2003). Loss of clusterin both in serum and tissue correlates with the tumorigenesis of esophageal squamous cell carcinoma via proteomics approaches. *World Journal of Gastroenterology : WJG*, *9*(4), 650–654.

https://archive.seattletimes.com/archive/?date=19980309&slug=2738618 [last accessed: 06/06/22]

https://www.cbsnews.com/news/doggy-dna-hounds-murderers [last accessed: 06/06/22]

https://www.matrixscience.com/ [last accessed: 06/05/22]

https://www.waters.com/nextgen/us/en/shop/columns/186003539-acquity-uplc-hss-t3-column-100a-18--m-21-mm-x-100-mm-1-pk.html [last accessed: 04/03/22]

https://www.waters.com/waters/en_US/T3-Columns/nav.htm?cid=134914589&locale=en_US [last accessed: 04/03/22]

Chapter 4: Detection of human haemoglobin variants for forensic applications using MALDI MS

Introduction

Blood is a valuable type of forensic evidence that is often encountered at the scene of violent crimes. Investigators have sought to obtain information pertaining to a crime from the physical characteristics related to its presence (blood quantity, location, blood pattern analysis) for centuries. The discipline of blood pattern analysis (BPA) was formally initiated in the late 1800s by Piotrowski (**Piotrowski, 1895**). However, the information that can be revealed by analysing the molecular composition of blood is a relatively recent field of forensic research. Analytical chemistry, forensic toxicology or clinical chemistry approaches may be used for this purpose to detect compounds, drugs, alcohol or poisons. In addition, physiological discrepancies can be investigated to help build a 'biological profile' of an individual suspected of being involved with a crime, in an attempt to narrow down a pool of possible suspects.

Blood typing has been used since the beginning of the 1900s as evidence in the events of violent crimes. The different antigens on the surface of the red blood cells allow for the identification of one of 4 human blood types (A, B, AB and O). While this may be useful to exonerate individuals if blood found on items belonging to a suspect does not match that of a victim, it becomes problematic when trying to link individuals to crimes, as a large proportion of the population will share a given blood type; (A - 42%, B - 9%, AB - 3%, O - 46% in the British Caucasian population) (**Regan, 2017**). While these can be further categorised into positive and negative types, they are not useful for identification purposes due to the large groups of the population that they cover.

However, in the last 30 years, outside of the field of blood pattern analysis (BPA), blood is mostly used to derive DNA information. Developed

simultaneously but independently by Jeffery Glassberg in the U.S. and Alec Jeffreys in the U.K. in 1983/4, a sample of DNA material allows for the identification of an individual. Although 99.9% of human DNA sequences are identical, the variation in the remaining 0.1% is enough for individual identification. The technique was first successfully used in 1987, to catch double child rapist and murderer Colin Pitchfork in Leicestershire, and has been frequently used in forensics, and also paternity/maternity testing and immigration eligibility, ever since its inception (**Evans, 1998**). It has recently been reported that the ability to perform DNA profiling appears to be unaffected by several physical and chemical blood enhancement techniques. Stewart *et al.* assessed the application of powders, powder suspensions, cyanoacrylate fuming (CAF) and acid dyes prior to DNA extraction, with the majority (41 out of 45) yielding full DNA profiles, with the remaining samples showing allele drop out for 1 or 2 loci (**Stewart *et al.,* 2018**). Whilst DNA profiling is an invaluable resource, it has some drawbacks. The quantity of recovered DNA material might be insufficient for profiling to take place. It can be susceptible to degradation based on the length of time before recovery and environmental conditions such as temperature or depleted by precipitation. Finally, even if a DNA profile can be generated, an offender may not be present in a DNA database if this is their first offence.

Proteins, in contrast, are more resilient to degradation than DNA, which can still be detected in blood long after it has been shed. A first offender would not be present in a DNA database, but police could request access to a suspect's medical records to see if physio-pathological anomalies had been recorded in their medical history. In addition, it has been demonstrated that blood proteins can successfully be detected and visualised following the application of several

BETs. For example, Patel *et al.* report the detection of blood-specific peptides (haemoglobin, complement C3, apolipoprotein, A-1, α-1-antitrypsin, haemopexin, serotransferrin and α-2-macroglobulin) in a 9 year old blood-stained hand print that had been treated with Acid Black 1 (an amino acid targeting enhancement reagent) (**Patel *et al.,* 2016**), and Francese reported the detection of haemoglobin peptides in a 37 years old blood-stained fingermark that had been visualised with ninhydrin (**Francese, 2019**).

Haemoglobin is the most abundant protein in the blood, which makes up the vast majority of red blood cells (around 96% by dry weight, or 35% wet content). Haemoglobin is a heterotetramer protein usually made up of 2 pairs of α and β subunits, each housing an iron-containing heme group, an iron-containing porphyrin hetero cyclic ring, used for the transport of gases around the body, primarily oxygen. The majority of the global population has the 'normal' type of haemoglobin, characterised by 2 α and β subunits, otherwise known as HbA^0 (**Weatherall & Clegg, 2001, Weatherall *et al.,* 2006, Weatherall, 2013**). But at the time of writing, there are currently over 1,400 known haemoglobin variants, as listed on the HbVar website (**globin.bx.psu.edu/**), an online database of human haemoglobin variants and thalassemias. Haemoglobin variants are caused by mutations in the amino acid sequence of the globular proteins. Typically this is a single substitution mutation but can also be insertion or deletion mutations and some variants contain more than one instance. These mutations fall into two categories, those that result in a structural anomalies, changing the amino acid sequence resulting in unusual haemoglobins, or those that affect the rate of the production of globin chains, also known as thalassemias (**Modell & Darlison, 2008**).

In contrast to the limited number of blood types, the large number of possible haemoglobin variants makes them a more specific method of identification. Although Weatherall and Clegg reported that variants occur in around 7% of the global population, in some parts of the world, sickle cell haemoglobin (HbS) can be present in up to 40% of the population due to the protection it offers against malaria, and HbE in excess of 60% of the population, so could contribute useful information in terms of narrowing down a possible pool of suspects and/or ascertain bio-geographical provenance. In addition, it has been noted by staff working in Haematology that certain variants are far more prevalent in specific ethnic groups, or in a specific locality (*Dr Jason Eyre, personal communication, 2020*).

In a clinical setting, a number of analytical techniques are used to screen patient blood for haemoglobin variants. The NHS uses four analytical approaches to variant screening; high performance liquid chromatography (HPLC), capillary electrophoresis (CE), tandem mass spectrometry (MS/MS) and isoelectric focusing (IEF) (**Daniel & Henthorn, 2015**). HPLC incorporates an ion exchange resin stationary phase, which the variants interact with. Separation is performed as the variants are eluted at different retention times and detected via MS. CE involves the application of a voltage to a solution containing electrolytes. Due to electro-osmotic flow, the solution moves towards one of the electrodes, and variants within the solution are separated depending on their ionic charge and electrophoretic mobility. MS/MS requires coupling multiple mass analysers to achieve fragmentation of ions, which can provide more comprehensive structural information about biomolecules. HPLC is the most prevalent approach in the UK (**www.gov.uk/...newborn-screening, www.gov.uk/...acceptable-analytical-protocols,**

www.gov.uk/...interpretation-of-results). A combination of these methods may be utilised, depending on the laboratory, to detect as many variants as possible. The NHS specifies that "The analytical procedures employed must be capable of detecting haemoglobins [Hb]S, [Hb]C, [Hb] D-Punjab, [Hb]E, and [Hb] O-Arab, in addition to HbF and HbA" (**Daniel & Henthorn, 2017**). However, these techniques are time consuming, complex, and not appropriate for blood contaminated fingermarks, due to their destructive nature and their inability to map the physical position of the analytes within a sample.

In 2015, the National Health Service (NHS) also published the results of a pilot study, regarding analysis of newborn screening of haemoglobin variants (Hb S, Hb C, Hb D-Punjab, Hb O-Arab, and Hb E) and Thalassaemias. The authors used Electrospray Ionisation (ESI) tandem mass spectrometry (MSMS) at four different NHS sites. Two sites employed an AB Sciex API 4000 Triple Quadrupole (TQMS) coupled to a Shimadzu Prominence HPLC system; one site employed a Waters Xevo TQMS coupled to a Waters Acquity HPLC system; the final site employed a Waters Quadrupole-Time of Flight (Q-TOF) Quattro Premier XE instrument coupled to a Waters Acquity Ultra Performance Liquid Chromatography (UPLC) system. Three of the sites (those using the API 4000 and Xevo instruments) were able to successfully detect all of the target Hb variants, reporting no false negatives. However the site using the Premier XE instrument was unable to detect some Hb variants due to its lower levels of sensitivity. The study also reported high numbers of false positives, particularly with Hb D-Punjab carriers. The authors also reported drawbacks of using an ESI approach for the analysis, including it being a time consuming process, concerns regarding sensitivity and specificity, difficulties troubleshooting and

lack of commercially available standards for controls containing Hb's S, C, D, E and O-Arab (**Daniel & Henthorn, 2015**).

Following the 2015 pilot study, the NHS also employs electrospray method of ionisation for clinical analysis, interfaced to three quadrupoles, with Q1 and Q3 acting as mass analysers and Q2 between them acting as a collision cell (**Daniel & Henthorn, 2017**). IEF involves applying a voltage to agarose gel containing ampholytes, which migrate through the gel depending on their isoelectric points. Variants also migrate to different, characteristic positions due to the electric field and settle in thin bands where their isoelectric point is equal to the pH of the gel (**Daniel & Henthorn, 2017**).

It is national NHS hospital policy not to report detected haemoglobin variants other than HbS, HbC, Hb D-Punjab, HbE, Hb O-Arab, and Hb Bart's. However, labs must send samples with <1.5% HbA or spectra with signals eluting after HbA (HPLC) or to the right of HbA (CE) for secondary testing, to another NHS laboratory for confirmation (**Daniel & Henthorn, 2017**). Moreover, specialist haematology departments that believe they have detected novel or rare variants must send samples to another specialist haematology department in another region to have their finding corroborated (*Dr Jason Eyre, personal communication, 2020*).

Over the last seven years, Francese's research group has demonstrated the ability of MALDI MS to successfully detect and image blood-specific proteins in stain and fingermark samples (**Bradshaw *et al.,* 2014, Patel *et al.,* 2015, Deininger *et al.,* 2016, Francese, 2019, Kennedy *et al.,* 2020**)**.** Kamanna's group has simultaneously been undertaking similar research utilising MALDI in Australia (**Kamanna *et al.,* 2017a**), while Chaurand's group has been

investigating a laser desorption ionisation (LDI) approach to analyse bloody fingermarks (**Lauzon & Chaurand, 2018**). These three groups (Francese's, Kamanna's & Chaurand's) have demonstrated the effectiveness of these techniques following the prior enhancement of marks with blood visualisation techniques (BETs). The compatibility of these techniques in conjunction with currently used BETs, both at the scene by crime scene investigators (CSIs) and in forensic laboratories, means that laser-based mass spectrometric techniques have the potential to offer investigations additional information in an operational context. Furthermore, its effectiveness to provide results with aged blood samples, currently tested up to 37 years (**Francese, 2019**), means that it shows promising potential for cold case investigations. More recently, work has been undertaken to further establish the origin of blood (human or animal) down to the animal species, and discriminate blood from other biofluids and protein-containing matrices that may elicit a positive response to rapid, colorimetric or catalytic tests currently used for the presumptive identification of blood at crime scenes (**Kamanna *et al.,* 2016, Kamanna *et al.,* 2017a, Kamanna *et al.,* 2017b, Kennedy *et al.,* 2020, Witt *et al.,* 2021**).

Haemoglobin screening is well established in the healthcare sector. However, the above mentioned and currently used analytical techniques are destructive, depleting the available sample through swabbing, extraction and reconstitution. This is not suitable for forensic samples, which may need to be retained for posterity or further analysis. To the author's knowledge, no screening of haemoglobin variants in a forensic context has been reported in the literature to date. It is hypothesised that using a combined proteomics and Matrix Assisted Laser Desorption Ionisation Mass Spectrometry (MALDI MS) and Mass Spectrometry Imaging (MALDI MSI) approach, it is possible to retrieve

additional bio-chemical intelligence to assist in forensic investigations. Incidentally, this method could also be of interest in clinical haematology for triaging samples and help speeding up testing.

Haemoglobin provides the most intense signals when analysing blood samples using a bottom-up proteomic approach, due to its readily ionisable molecular structure (**Patel** *et al.,* **2016**). Bottom-up proteomics involves proteolytic digestion of proteins into peptides, followed by analysis via mass spectrometry and identification of the protein by the detection of said peptides. This process is called peptide mass fingerprinting (PMF). Matrix Assisted Laser Desorption Ionisation Mass Spectrometry (MALDI MS) has been selected for this study for the speed of analysis, the simplicity of simpler spectral interpretation given the detection of mainly singly charged ions, its ability to also provide imaging data and to its minimally destructive nature in this modality mode. When digesting haemoglobin proteins with trypsin, two [M+H]$^+$ haemoglobin peptides peaks at *m/z* 1274.725 and 1529.734 are generated which have a particularly high ionisation efficiency. These peptides originate from within the beta and alpha protein subunits respectively (**Table 19**).

Protein chain	*m/z*	Amino acid sequence	Position
α	1529.734	VGGHAAEYGAEALER	[18-32]
β	1274.725	LLVVYPWTQR	[32-41]

Table 19. The mass to charge ratio and amino acid sequence of the most intense peaks present in MALDI MS spectra of human HbA^0 haemoglobin.

These two abundant peptides are a good indication that blood is present within a sample. (Although false positive for both signals concurrently is unlikely, it is possible, so MSMS can be performed for absolute identification). Trypsin cleaves the protein at the carboxyl side of lysine and arginine (shorthand

notation K and R), therefore for 'healthy' α-haemoglobin subunits (αHb), it will cleave the protein into 21 potential peptide chains, and for 'healthy' β-haemoglobin (βHb) subunits into potential 26 peptide chains (**Table 20**). These potential peptides have been generated through *in silico* digestion using the PeptideMass function, allowing for up to 2 missed cleaveges, within UniProt (www.uniprot.org).

Possible normal haemoglobin α chain peptides

Mass [M+H]+	Position	Missed cleavages	Amino acid sequence
2996.489	63-91	0	VADALTNAVAHVDDMPNALS ALSDLHAHK
2967.612	101-128	0	LLSHCLLVTLAAHLPAEFTP AVHASLDK
2886.428	33-57	1	MFLSFPTTKTYFPHFDLSHG SAQVK
2582.271	18-41	1	VGAHAGEYGAEALERMFLSF PTTK
2485.258	09-32	2	TNVKAAWGKVGAHAGEYGAE ALER
2341.184	42-62	2	TYFPHFDLSHGSAQVKGHGK K
2213.089	42-61	1	TYFPHFDLSHGSAQVKGHGK
2043.004	13-32	1	AAWGKVGAHAGEYGAEALER
1833.892	42-57	0	TYFPHFDLSHGSAQVK
1684.938	02-17	2	VLSPADKTNVKAAWGK
1571.879	129-142	1	FLASVSTVLTSKYR
1529.734	18-32	0	VGAHAGEYGAEALER
1252.715	129-140	0	FLASVSTVLTSK
1171.668	02-12	1	VLSPADKTNVK
1087.626	92-100	1	LRVDPVNFK
1071.554	33-41	0	MFLSFPTTK
974.5418	09-17	1	TNVKAAWGK
818.4406	94-100	0	VDPVNFK
729.4141	02-08	0	VLSPADK
532.2878	13-17	0	AAWGK
526.3096	58-62	1	GHGKK

Possible normal haemoglobin β chain peptides

Mass [M+H]+	Position	Missed cleavages	Amino acid sequence
2827.519	97-121	1	LHVDPENFRLLGNVLVCVLA HHFGK
2809.478	122-147	2	EFTPPVQAAYQKVVAGVANA LAHKYH

2679.324	42-66	2	FFESFGDLSTPDAVMGNPKV KAHGK
2570.373	19-41	1	VNVDEVGGEALGRLLVVYPW TQR
2529.219	84-105	1	GTFATLSELHCDKLHVDPEN FR
2509.356	122-145	1	EFTPPVQAAYQKVVAGVANA LAHK
2286.111	42-62	1	FFESFGDLSTPDAVMGNPKV K
2228.167	10-31	1	SAVTALWGKVNVDEVGGEAL GR
2191.198	63-83	2	AHGKKVLGAFSDGLAHLDNL K
2058.948	42-60	0	FFESFGDLSTPDAVMGNPK
1866.012	02-18	1	VHLTPEEKSAVTALWGK
1797.986	67-83	1	KVLGAFSDGLAHLDNLK
1719.973	106-121	0	LLGNVLVCVLAHHFGK
1669.891	68-83	0	VLGAFSDGLAHLDNLK
1449.796	134-147	1	VVAGVANALAHKYH
1421.673	84-96	0	GTFATLSELHCDK
1378.7	122-133	0	EFTPPVQAAYQK
1314.665	19-31	0	VNVDEVGGEALGR
1274.726	32-41	0	LLVVYPWTQR
1149.674	134-145	0	VVAGVANALAHK
1126.564	97-105	0	LHVDPENFR
952.5098	02-09	0	VHLTPEEK
932.52	10-18	0	SAVTALWGK
767.4886	61-67	2	VKAHGKK
639.3936	61-66	1	VKAHGK
540.3252	63-67	1	AHGKK

Table 20. Possible healthy peptide chains when performing proteolytic digestion of haemoglobin with trypsin and allowing for up to 2 missed cleavages. (2a) Normal haemoglobin α chain peptides. (2b) Normal haemoglobin β chain peptides. (No missed cleavages shown in black, one missed cleavage site in blue and 2 missed cleavage sites in red).

However, haemoglobin variants are characterised by mutations in the amino acid sequence of the protein chains, compared to 'healthy' haemoglobin. The peptides that are generated through enzymatic proteolysis that contain these

modifications will have a different molecular weight to the 'healthy' haemoglobin peptides, and their detection provides the indication of a haemoglobin variant. The shorthand notation for 'normal' haemoglobin is HbA^0, with the 6 most common variants in the UK being HbC, HbE, HbS, Hb D-Iran, Hb D-Punjab and Hb J-Baltimore. They are named based on where their incidence was first discovered, but global migration has led to many variants being found worldwide, although, crucially, prevalence can vary based on their origin.

Hb Los-Angeles was among the first haemoglobin variants to be discovered, and was reported by Itano in 1951 to contain a substitution of glutamic acid for a glutamine on the 121st position in the β globin chain (**Itano, 1951**). Baglioni reported in 1962 that this was the same amino acid substitution as a variant that had been named Hb D-Punjab (**Baglioni, 1962**). Although their first instances had been recorded in different locations, they were identical variants, and are technically denominated as Hb D-Punjab/Los Angeles, although will now be referred to as Hb D-Punjab. Hb J-Baltimore was first reported in 1963 by Baglioni and Weatherall, and was discovered as the name suggests in Baltimore, in a family of African heritage (**Baglioni & Weatherall, 1963**). Dass *et al.* reported Hb D-Iran as a rare variant which was first discovered in a family in central Iran in 1973 (**Dass *et al.,* 2017**). Hb D-Iran is the second most common of the Hb D variants after Hb D Punjab (Los Angeles) (**Edwards, 2012**).

The globin protein structure of 'normal' HbA^0 contains 2 alpha and 2 beta subunits; α2β2. All of these variants have modifications which occur on the β chain. The variants are denominated by a superscript letter on the chain that has been modified shown in **Table 21**. **Table 21** also shows the amino acid

substitution and the position of the modification within the amino acid chain for each of these variants.

Variant	HbA^0	HbC	Hb D-Punjab	Hb D-Iran	HbE	Hb J-Baltimore	HbS
Chain	α2β2	α2β^{C}2	α2β^{D}2	α2β^{D}2	α2β^{E}2	α2β^{J}2	α2β^{S}2
Modification	NA	Glu>Lys	Glu>Gln	Glu>Gln	Glu>Lys	Gly>Asp	Glu>Val
Position	NA	6 β chain	121 β chain	22 β chain	26 β chain	16 β chain	6 β chain

Table 21. The chains, amino acid modification and modification position of each of the variants investigated in this study [*Adapted from Heaton et al., 2020*].

This chapter explores the use of MALDI MS profiling (MALDI MSP) to detect and identify six of the more common haemoglobin variants in blood stains, and the potential of MALDI MS imaging (MALDI MSI) to visualise specific HbVar peptides within the ridge pattern of bloody fingermarks. The results from this study show MALDI to be a sensitive, minimally destructive (ablation of a few micrometres of material from the surface) technique, leaving the ridge pattern intact. Whilst this is preliminary investigation into a small subset of the total number of known haemoglobin variants, compatibility has been demonstrated with at least one commonly used fingermark enhancement technique, illustrating its potential to be integrated into a forensic investigation workflow. Furthermore, with significant refinement, the method may offer a suitable alternative for the conventional clinical diagnostic methods currently utilised by the NHS for haemoglobin variant screening.

The study presented in this chapter has endeavoured to develop a novel method for the detection and visualisation of haemoglobin variants in blood stains and in blood fingermarks indicating the potential to support criminal investigations. This would prove particularly useful in scenarios where

identification through DNA profiling is not successful (such as those circumstances previously mentioned).

Experimental

Materials

Trifluoroacetic acid (TFA), α-cyano-4-hydroxycinnamic acid (α-CHCA), ammonium bicarbonate (AmBic), phosphorous red and TLC sheets were purchased from Sigma Aldrich (Poole, UK). Acetonitrile, (ACN) was purchased from Fisher Scientific (Loughborough, UK). Lyophilised sequencing grade modified Trypsin was purchased from Promega in 20 µg vials (Southampton, UK). RapiGest surfactant was purchased in 1 mg vials from Waters (Waters Corp, Manchester, UK). Double sided conductive tape was purchased from TAAB (Aldemastor, UK), Intellislides (glass slides) were provided by Bruker Daltonik (GmbH, Bremen, Germany). Indium tin oxide (ITO) slides were purchased from Sigma Aldrich (Poole, UK). Haemoglobin variant standards HbA^0 and HbS were purchased from Sigma Aldrich (Dorset, UK). A mixed haemoglobin AFSC standard was purchased from Sebia (Lisses, France) as they were not commercially available individually. Patient blood samples containing the 6 haemoglobin variants in this study (HbC, HbE, HbS, Hb D-Iran, Hb D-Punjab and Hb J-Baltimore) were procured from the Haemolysis Lab at Sheffield's Royal Hallamshire Hospital. 3D printed silicon replica fingers were produced in-house and used to deposit blood-contaminated fingermarks (to mitigate the biohazard risk). Patient blood was provided under Sheffield Hallam University Ethics Approval (Converis) and IRAS approval (approval codes ER6558932 and 160418, respectively).

Instruments and instrumental conditions

Three datasets were acquired on three separate mass spectrometers; the MALDI Q-TOF Synapt G2 HDMS (Waters Corp. Manchester, UK), MALDI TOF-TOF rapifleX and MALDI Q-TOF timsTOF flex (both Bruker Daltonik GmbH, Bremen, Germany).

Data acquisition: MALDI MSP

The Synapt G2 HDMS mass spectrometer is equipped with a neodymium-doped yttrium aluminium garnet (Nd:YAG) laser operating at 1 kHz. Acquisition was performed in the mass range m/z 600 - 2,500 in positive mode. A spot of 0.5 µL saturated phosphorus red solution dissolved in ACN was spotted on each slide and used as an internal calibrant, acquiring a calibration spectrum first within each of the sample acquisitions. MALDI MS/MS spectra were obtained using argon as the collision gas, with the trap gas set to 7.5 a.u. and the cooling at 10.0 a.u. The trap collision energy was ramped between 80 and 100 a.u.

The MALDI rapifleX mass spectrometer is equipped with a SmartBeam™ 355 nm Nd:YAG laser operating at 10 kHz. The instrument was calibrated in positive mode using Bruker's calibration standard mono peptide mix between m/z 1,046 - 2,455. Spectra were obtained in reflectron positive mode in the mass range m/z 600 - 3,200. Laser power was set to 32 a.u.

The MALDI timsTOF fleX mass spectrometer is equipped with a SmartBeam™ 3D laser operating at 10 kHz. The instrument was calibrated in positive mode using red phosphorus between m/z 100 - 3,000. In MS mode, 500 shots were

acquired per acquisition, and typically 10 acquisitions were summed to generate each sample spectrum. MS/MS spectra were obtained by ramping to collision energy between 80 - 130 eV. 5,000 laser shots were acquired per acquisition, and typically 10 acquisitions were summed to generate each sample spectrum.

Data acquisition: MALDI MSI

Imaging data were acquired on the same MALDI Synapt G2 HDMS mass spectrometer and MALDI timsTOF fleX mass spectrometer. On the Synapt instrument, two blood stains were analysed in imaging mode at a spatial resolution of 50 μm and acquired in the mass range *m/z* 600 - 2,000. The quadrupole resolution low mass (LM) was set at 4.4 and high mass (HM) at 11.0 a.u., with the laser power set at 250 a.u.

On the MALDI timsTOF flex mass spectrometer, separate blood fingermarks and overlapping blood marks were acquired at a spatial resolution of 50 μm^2 in positive mode in the mass range *m/z* 100 - 3,000. The laser power was set to 65% and accumulated 600 shots per pixel. The mass spectrometer was previously calibrated in ESI mode using Agilent low concentration calibration mix (G1969-85000).

Data pre-processing

For the 6 haemoglobin variants in this study, the amino acid sequence of the mutated β globin chains were identified using the HbVar database (https://globin.bx.psu.edu/globin/Hbvar). Theoretical *in silico* digestion was then performed by entering the variant amino acid sequence into UniProt (www.uniprot.org) (**Wilkins *et al.,* 1997, Gasteiger *et al.,* 2005**), using the PeptideMass function, which can be found at

(https://web.expasy.org/peptide_mass/), and selecting trypsin as the enzyme, allowing for up to 2 missed cleavages and displaying [M+H]$^+$ peptides up to *m/z* 3,000. From the resulting peptide list, peptides were selected which contained the modification region within 0, 1 or 2 cleavage sites. From this list, peptides were filtered out if they had an identical *m/z* to any 'healthy' or other HbVar peptides, as, if shared, these could not be used for variant identification. Peptides were also filtered out if they were within ±15 ppm of another peptide, to account for the mass accuracy of the instruments used. Peptides were further discounted from the final identification list if they were not detected in the Hb standards spectra **(Fig. 49)**.

Original Beta chain	MVHLTPEEKS AVTALWGKVN VDEVGGEALG RLLVVYPWTQ RFFESFGDLS TPDAVMGNPK VKAHGKKVLG AFSDGLAHLD NLKGTFATLS ELHCDKLHVD PENFRLLGNV LVCVLAHHFG KEFTPPVQAA YQKVVAGVAN ALAHKYH
Modified Beta chain	MVHLTPVEKS AVTALWGKVN VDEVGGKALG RLLVVYPWTQ RFFESFGDLS TPDAVMGNPK VKAHGKKVLG AFSDGLAHLD NLKGTFATLS ELHCDKLHVD PENFRLLGNV LVCVLAHHFG KEFTPPVQAA YQKVVAGVAN ALAHKYH

mass	position	#MC	modifications	peptide sequence
2827.5188	96-120	1		LHVDPENFRLLGNVLVCVLA HHFGK
2809.4783	121-146	2		EFTPPVQAAYQKVVAGVANA LAHKYH
2733.4933	1-26	2		VHLTPVEKSAVTALWGKVNV DEVGGK
2679.3235	41-65	2		FFESFGDLSTPDAVMGNPKV KAHGK
2569.4248	18-40	2		VNVDEVGGKALGRLLVVYPW TQR
2529.2190	83-104	1		GTFATLSELHCDKLHVDPEN FR
2509.3561	121-144	1		EFTPPVQAAYQKVVAGVANA LAHK
2286.1110	41-61	1		FFESFGDLSTPDAVMGNPKV K
2227.2193	9-30	2		SAVTALWGKVNVDEVGGKAL GR
2191.1981	62-82	2		AHGKKVLGAFSDGLAHLDNL K
2058.9477	41-59	0		FFESFGDLSTPDAVMGNPK
1836.0377	1-17	1		VHLTPVEKSAVTALWGK
1829.9755	9-26	1		SAVTALWGKVNVDEVGGK
1797.9857	66-82	1		KVLGAFSDGLAHLDNLK
1719.9726	105-120	0		LLGNVLVCVLAHHFGK
1671.9693	27-40	1		ALGRLLVVYPWTQR
1669.8907	67-82	0		VLGAFSDGLAHLDNLK
1449.7961	133-146	1		VVAGVANALAHKYH
1421.6729	83-95	0		GTFATLSELHCDK
1378.7001	121-132	0		EFTPPVQAAYQK
1313.7171	18-30	1		VNVDEVGGKALGR
1274.7255	31-40	0		LLVVYPWTQR
1149.6738	133-144	0		VVAGVANALAHK
1126.5639	96-104	0		LHVDPENFR
932.5200	9-17	0		SAVTALWGK
922.5356	1-8	0		VHLTPVEK
916.4734	18-26	0		VNVDEVGGK
767.4886	60-66	2		VKAHGKK
639.3936	60-65	1		VKAHGK
540.3252	62-66	1		AHGKK

Peptide sequences containing the modification region (6 β chain)			
Mass [M+H]$^+$	Position	Missed cleavages	Peptide sequence
2733.4933	2-27	2	VHLTPVEKSAVTALWGKVNVDEVGGK
1836.0377	2-18	1	VHLTPVEKSAVTALWGK
922.5356	2-9	0	VHLTPVEK

Figure 49. Peptide selection criteria using HbS as an example. The HbS variant contains two unmodified α chains and one or two modified β chains ($\alpha_2\beta^S_2$) (1a). (1b) shows the possible peptides when performing *in silico* proteolytic digestion on HbS using trypsin as the enzyme (values reported as singly charged ions [M+H]$^+$). (1c) shows those peptides with sequences spanning the modification range.

For the example of HbS, the amino acid substitution occurs on the 6th amino acid, where a Glutamic acid (E) mutates to a Valine (V). Peptides spanning the modification region were selected. The proteotypic peptide at nominal *m/z* 2733 was discounted from the final list as it was not observed in the HbS standard, likely due to its lower ionisation efficiency.

The final theoretical peptide list comprising of all of the 6 variants contained 19 peptides, with each variant having at least 2 proteotypic peptides. These are shown in **Table 22**, along with the position of the peptide segment containing the modification region within the β chain, the amino acid sequence, including the substitution mutation highlighted in red, and the *m/z* ([M+H]$^+$) of the peptide. There may be some miniscule discrepancies between the mass values obtained from UniProt and the reported values, as these were calculated accurately using IsoPro 3.1 Mass Spectral Isotopic Distribution Simulator (plus the mass of one proton), which uses slightly different amino acid mass values.

Hb variant	AA position	Proteotypic peptide sequence	*m/z*
HbC	(2-18)	VHLTP**K**EKSAVTALWGK	1865.0646
HbC	(2-9)	VHLTP**K**EK	951.5626
HbC	(2-7)	VHLTP**K**	694.4251
HbS	(2-18)	VHLTP**V**EKSAVTALWGK	1836.0381
HbS	(2-9)	VHLTP**V**EK	922.5361
HbE	(19-41)	VNVDEVGG**K**ALGRLLVVYPW TQR	2569.4253
HbE	(10-31)	SAVTALWGKVNVDEVGG**K**AL GR	2227.2197
HbE	(10-27)	SAVTALWGKVNVDEVGG**K**	1829.9760
HbE	(19-31)	VNVDEVGG**K**ALGR	1313.7175
HbE	(19-27)	VNVDEVGG**K**	916.4739

HbD-Punjab	(122-147)	**Q**FTPPVQAAYQKVVAGVANA LAHKYH	2826.5051
HbD-Punjab	(122-145)	**Q**FTPPVQAAYQKVVAGVANA LAHK	2508.3726
HbD-Punjab	(122-133)	**Q**FTPPVQAAYQK	1377.7164
HbD-Iran	(19-41)	VNVD**Q**VGGEALGRLLVVYPW TQR	2569.3889
HbD-Iran	(10-31)	SAVTALWGKVNVD**Q**VGGEAL GR	2227.1831
HbD-Iran	(19-31)	VNVD**Q**VGGEALGR	1313.6812
HbJ-Baltimore	(10-31)	SAVTALW**D**KVNVD**E**VGGEAL GR	2286.1726
HbJ-Baltimore	(1-18)	MVHLTPEEKSAVTALW**D**K	2055.0583
HbJ-Baltimore	(10-18)	SAVTALW**D**K	990.5259

Table 22. The theoretical peptide segments generated for each Hb variant, with the position within the chain, the amino acid sequence with the mutation highlighted and the *m/z.*

Data post-processing

MALDI MS and MS/MS spectra from each of the three mass spectrometers were manually exported to txt files and processed within mMass, an open source spectral processing program (**Strohalm *et al.,* 2008, Strohalm *et al.,* 2010**). The Synapt and rapifleX data were processed using smoothing and peak picking. Smoothing was performed using the Savitzky-Golay algorithm, an *m/z* window size of 0.2 and 1 smoothing cycle. Peak picking was performed using a signal-to-noise threshold of 4. The relative intensity threshold was set at 0.5 a.u. and picking height at 100 a.u. MALDI MS and MS/MS spectra acquired using the timsTOF fleX instrument were preliminarily processing using DataAnalysis software (Bruker Daltonik GmbH, Bremen, Germany) and exported as txt files for processing in mMass. For these spectra, no further smoothing was applied, and peak picking was performed using a signal-to-noise threshold of 20, relative intensity threshold of 0.5 a.u. and picking height of 98 a.u. The haemoglobin variant proteotypic peptides that were detected by the rapifleX were used to

build an internal mMass database which could automatically assign haemoglobin variants based on the proteotypic peptides detected in MS mode. For analysis on the Synapt mass spectrometer, the mass accuracy allowed for ± 13 ppm, for the rapifleX mass accuracy allowed for ± 11 ppm. Synapt MALDI MS images were processed using Waters proprietary HDI software (Waters Corp. Manchester, UK). Data was normalised using total ion current (TIC). timsTOF flex MALDI MS images were processed using SCiLS Lab (v. 2021a, Bruker Daltonik GmbH, Bremen, Germany). Mass accuracy was ± 15 ppm and data was normalised using root mean square (RMS). Brightness and contrast were adjusted to optimise the clarity of the ridge detail.

Trypsin digestion

In solution tryptic digests were carried out using the same protocol for all instruments. Blood was diluted 1:200 in deionised H_2O. 40 µL 40 mM AmBic was added to 10 µL of the diluted blood solution. 9 µL of 20 µg/mL trypsin was added (containing 0.1% (v/v) RapiGest surfactant). The solution was gently vortexed and incubated at 37 °C for 1 hr. Following incubation, 2 µL 5% TFA were added to stop the digestion.

In situ tryptic digests for imaging analysis were carried out as follows. For blood stains analysed on the Synapt instrument, 9 layers of trypsin gold at 100 µg/mL (containing 0.1% (v/v) RapiGest surfactant) were sprayed directly onto the blood stains using a SunCollect matrix sprayer (SunChrom, Germany) at a flow rate of 2 µL/min. The samples were then mounted on a polystyrene raft and incubated in a sealed Tupperware-style container, containing 50% MeOH at 37 °C for 3 hours. The digested samples were then coated with 5 layers of α-CHCA matrix at 5 mg/ml in 70:30 ACN:0.5% TFA at a flow rate of 4 µL/min.

For the blood-contaminated fingermark samples, 10 µL of patient blood containing an Hb variant was pipetted onto the tip of a custom in-house 3D printed silicon fingertip and smeared across the ridges. The contaminated silicon fingertip was then used to deposit bloody fingermarks on Intellislide glass slides. An overlapping mark sample was prepared as above, but after first depositing a HbC-contaminated mark, a HbJ-Baltimore contaminated mark was deposited, overlapping the first, with a clean silicon finger. These marks were left to dry for 24 hours before being enhanced with Acid Black 1 using the methanol formulation described in the Fingermark Visualisation Manual (**Bandey *et al.*, 2014**). These blood-contaminated fingermark samples were subsequently sprayed with sequencing grade trypsin (25 ng/µL in 20 mM ammonium bicarbonate) using the HTX M3 sprayer (HTX Technologies, LLC, U.S.). The spraying parameters were as follows. Spray temperature and pressure was 30 °C and 10 psi. The flow rate was 30 µL/min and the nozzle height was 40 mm. 8 layers of trypsin were sprayed with a nozzle velocity of 750 mm/min using a pre-defined criss-cross pattern with 2 mm track spacing. Following trypsin deposition, the Intellislides were mounted in a Tupperware box containing 40 mL saturated K_2SO_4 solution (75 g in 500 mL H_2O). 10 µL of H_2O was deposited on the edge of the slide. The samples were incubated at 50 °C for 2 hours. Samples were then removed from the incubation and analysed on a MALDI ttfleX mass spectrometer.

Matrix deposition

MS and MS/MS analysis

For samples analysed via MALDI MS and MS/MS mode on the Synapt G2 HDMS, α-CHCA was reconstituted in 70:30 ACN:0.5% TFA at 5 mg/mL. 0.5 µL

of sample and 0.5 μL of α-CHCA was then pipette mixed within wells on the target plate.

For samples analysed via MALDI MS mode on the rapifleX MS, α-CHCA was prepared as before. Matrix was then spotted onto 0.5 μL sample spots within MALDI target plate wells that had been allowed to fully dry in a vacuum desiccator.

For samples analysed via MALDI MS and MS/MS mode on the ttfleX MS, Hb standards and digested patient blood samples were individually mixed in a 1:1 ratio with α-CHCA in 70:30 ACN:0.1% TFA. 1 μL of this sample/matrix mixture was then spotted within wells on a MALDI Anchor 800 target plate. After being allowed to fully dry under ambient conditions, the well was washed with 3 μL H_2O containing 0.1% TFA.

MS Imaging analysis

For samples analysed via MALDI MSI mode on the Synapt G2 HDMS, 5 layers of α-CHCA were sprayed over the sample area at a concentration of 5 mg/mL in 70:30 ACN:0.5% TFA, using a SunCollect matrix sprayer at a flow rate of 2 μL/min.

For samples analysed via MALDI MSI mode on the ttfleX MS, 4 layers of α-CHCA were sprayed over the sample area at a concentration of 10 mg/mL in 70:30 ANC:0.2% TFA, using a HTX M3 sprayer at a flow rate of 120 μL/min. The nozzle height was set to 40 mm with 2 mm track spacing at 1200 mm/min. The nozzle temperature was set to 75 °C and the pressure was set to 10 psi.

Results and Discussion

This study was designed to develop a reliable method for the detection and identification of the six most common haemoglobin variants from whole blood samples. Although this research has an obvious application in a clinical setting, its focus was directed towards providing additional information in a forensic context. Clinical haematology labs already have established methods for identification and confirmation of haemoglobin variants, including electrophoresis-based techniques and HPLC, for health-based screening. However, due to the large number of possible known mutations, variants have the potential to be useful in including or ruling out suspects/victims from blood recovered from crime scenes. The six most common HbVar in the UK (excluding HbA^0 and HbA^2) are: HbC, Hb D-Iran, Hb D-Punjab, HbE, Hb J-Baltimore and HbS. A combined proteomics-MALDI MS profiling/imaging approach was employed for the rapid detection of proteotypic HbVar peptides, in both pre-mixed spotted samples and *in situ* blood-contaminated fingermarks.

Patient blood, containing each of the variants, as well as HbA^0 and HbA^2 as a control, was received in batches from the Haematology Department of Sheffield's Royal Hallamshire Hospital over the course of the study, over a period of 2 years. Initial batches were received with the variants labelled, for method development. Once the method had been refined, subsequent batches were received in a blind fashion, until the assignment to an HbVar or to HbA^0 had been made. The samples were then stored in a fridge at 4 °C and analysed as soon as possible (within 5 days) after receipt. A summary of the putative identification for each of the samples is reported in **Table 23**, below.

HbA^0 samples			
Patient number	**Variant ID**	**Empirical identification**	**Results and replicate identification frequency**
IH140490	HbA^0	No variants detected	Samples analysed: **3**
IH087366	HbA^0	No variants detected	Samples where the general consensus identification was correct: **3**
IH087367	HbA^0	No variants detected	Correct variant identified in 3/3 replicates: **3** Correct variant identified in 2/3 replicates: **0** Correct variant identified in 1/3 replicates: **0** Inconclusive samples: **0**

HbC samples			
Patient number	**Variant ID**	**Empirical identification**	**Results and replicate identification frequency**
IH140065	HbC	Inconclusive	Samples analysed: **11**
IH140068	HbC	Inconclusive	
IH140073	HbC	HbC	Samples where general consensus identification
IH140076	HbC	HbC	was correct: **8**
IH140078	HbC	HbC	
IH140080	HbC	HbC	Correct variant identified in 3/3 replicates: **4**
IH140081	HbC	HbC	Correct variant identified in 2/3 replicates: **3**
IH140084	HbC	HbC	Correct variant identified in 1/3 replicates: **1**
IH087364	HbC	No variants detected	Inconclusive samples: **2***
IH140473	HbC	HbC	
IH140486	HbC	HbC	*in two instances, HbC was indicated along with
IH140488	HbC	HbC	another variant

Hb D-Iran samples			
Patient number	**Variant ID**	**Empirical identification**	**Results and replicate identification frequency**
IH140484	Hb D-Iran	Hb D-Iran	Samples analysed: **1** Samples where the general consensus identification was correct: **1** Correct variant identified in 3/3 replicates: **1** Correct variant identified in 2/3 replicates: **0** Correct variant identified in 1/3 replicates: **0** Inconclusive samples: **0**

Hb D-Punjab samples			
Patient number	**Variant ID**	**Empirical identification**	**Results and replicate identification frequency**
IH140071	Hb D-Punjab	No variants detected	Samples analysed: **7**
IH140072	Hb D-Punjab	Inconclusive	Samples where the general consensus identification was correct: **1**
IH140077	Hb D-Punjab	Hb D-Punjab	
IH140083	Hb D-Punjab	No variants detected	Correct variant identified in 3/3 replicates: **1**
IH140485	Hb D-Punjab	No variants detected	Correct variant identified in 2/3 replicates: **0** Correct variant identified in 1/3 replicates: **0**
IH140467	Hb D-Punjab	No variants detected	Inconclusive samples: **1*** *in one instance, each replicate indicated a different variant
IH087372	Hb D-Punjab	No variants detected	In five instances, no variant peptides detected

259

HbE samples			
Patient number	Variant ID	Empirical identification	Results and replicate identification frequency
IH140074	HbEE	HbE	Samples analysed: **10**
IH140075	HbAE	HbE	
IH140079	HbAE	HbE	Samples where the general consensus
IH140082	HbAE	HbE	identification was correct: **6**
IH087362	HbAE	No variants detected	Correct variant identified in 3/3 replicates: **5**
IH087365	HbAE	No variants detected	Correct variant identified in 2/3 replicates: **1**
IH140470	HbAE	HbE and Hb D-Iran	Correct variant identified in 1/3 replicates: **0**
IH140465	HbAE	HbE	Inconclusive samples: **1***
IH140489	HbAE	HbE	*in one instance, HbE and Hb D-Iran were indicated by one replicate each
IH140491	HbAE	HbE	In three instances, no variant peptides detected

Hb J-Baltimore samples			
Patient number	Variant ID	Empirical identification	Results and replicate identification frequency
IH140468	HbJ-Baltimore	No variants detected	Samples analysed: **1** Samples where the general consensus identification was correct: **0** Inconclusive samples: **1*** *no variant peptides detected in the sample analysed

HbS samples			
Patient number	Variant ID	Empirical identification	Results and replicate identification frequency
IH140064	HbSS	HbS	Samples analysed: **10**
IH140066	HbS	HbS	
IH140067	HbS (+HPFH)#	HbS	Samples where the general consensus
IH140069	HbS	HbS	identification was correct: **9**
IH140081	HbS	Inconclusive	
IH140086	HbS	HbS	Correct variant identified in 3/3 replicates: **5**
IH087370	HbS	HbS	Correct variant identified in 2/3 replicates: **3**
IH140464	HbS	HbS	Correct variant identified in 1/3 replicates: **1**
IH140471	HbS	HbS	Inconclusive samples: **1***
IH140466	HbS	HbS	*in one instance, HbS and HbC were both indicated by one replicate each

Table 23. MALDI MS putative identifications (from *m/z* of the peptides acquired in high mass accuracy) of the patient blood samples acquired on the Synapt instrument. A breakdown of the 3 technical replicates for each sample is included in the last column. # sample IH140067 also contained Hereditary Persistence of Foetal Haemoglobin (HPFH), a mutant gene that inhibits the synthesis of HbA^0 and HbA^2 (Sharma *et al.,* 2020).

All 3 of the HbA^0 samples were correctly identified as not containing any variant peptides. The HbC variant was detected in 8 out of the 11 total HbC samples. The Hb D-Iran variant was detected in 2 out of the 2 total Hb D-Iran samples. The Hb D-Punjab variant was detected in 1 out of the 7 total Hb D-Punjab samples. The HbE variant was detected in 6 out of the 10 total HbE samples.

The Hb J-Baltimore variant was detected in the only sample of this type analysed. The HbS variant was detected in 9 out of the 10 total HbS samples. It is hypothesised that the successful identification of this peptide is likely to relate in part to the ionisation potential of a low molecular weight (LMW) peptide of only 7 amino acids.

As quantitative analysis was not performed, it is not possible to distinguish between patients with homozygous and heterozygous genotypes. Patients can have one of three genotypes in terms of haemoglobin variants; homozygous with 2 copies of the 'normal' β-globin gene, homozygous with 2 copies of the 'abnormal' β-globin gene, or heterozygous with one 'normal' and one 'abnormal' copy of the gene. However, the ratios of each can vary from person to person and are rarely expressed 50:50 in heterozygous individuals or 100% in homozygous individuals. Indeed, even in patients with predominately HbA^0, there can still be low levels of HbA^2 and/or HbF (**Wild & Bain, 2017**). In an adult with 'normal' haemoglobin, the combinations are approximately HbA >95%, HbA_2 2%-3.4%, HbF <1% (www.gov.uk... understanding-haemoglobinopathies). In addition, the ratio of HbA_2 to other haemoglobin is clinically significant in the diagnosis of β-thalassemic traits (**Wahed & Risin, 2013**).

Therefore, data obtained in this study could not differentiate between normal and elevated levels of a given variant, or whether a patient was homozygous or heterozygous for a variant, only whether a variant was present in quantities above the level of detection of the method. For example, analysis could determine whether HbS was present, regardless of if the patient had the heterozygous trait (HbAS) or the homozygous pathology (HbSS).

Table 24 displays the experimental and theoretical *m/z* of the putative assignments and the peptides and reports the mass accuracy in parts per million (ppm) between them.

Patient no. and replicate	Variant (peptide position)	Exp. m/z (Th)	Mass Accuracy (ppm)	Experimental ID	True ID
IH140473_2	HbC (2-9)	951.571	9.0	HbC	HbAC
IH140486_1	HbC (2-9)	951.567	4.9	HbC	HbAC
IH140488_1	HbC (2-9)	951.562	0.2	HbC	HbAC
	Hb J-Baltimore (10-31)	2286.151	-9.3		
IH140073_2	HbC (2-9)	951.533	-9.4	HbC	HbAC
IH140076_3	HbC (2-9)	951.562	0.3	HbC	HbAC
IH140078_1	HbC (2-9)	951.564	1.7	HbC	HbAC
IH140080_1	HbC (2-9)	951.562	-0.2	HbC	HbAC
IH140084_3	HbC (2-9)	951.563	1.1	HbC	HbAC
IH087364_1	NA	NA	NA	no variants	HbAC
IH140065_3	HbC (2-9)	951.555	-8.0	Inconclusive	HbAC
	Hb D-Iran (19-31)	1313.674	-5.4		
IH140068_1	HbC (2-9)	951.562	0.1	Inconclusive	HABC

Patient no. and replicate	Variant (peptide position)	Exp. m/z (Th)	Mass Accuracy (ppm)	Experimental ID	True ID
IH140484_1	Hb D-Iran (10-31)	2227.183	0.2	Hb D-Iran	HbD-Iran

Patient no. and replicate	Variant (peptide position)	Exp. m/z (Th)	Mass Accuracy (ppm)	Experimental ID	True ID
IH087372_1	NA	NA	NA	no variants	HbAD-Punjab
IH140071_1	NA	NA	NA	no variants	HbAD-Punjab
IH140072_2	Hb D-Iran (19-31)	1313.671	-11.1	Inconclusive	HbAD-Punjab
	HbE (19-31)	1313.703	-7.6		
IH140077_1	Hb D-Punjab (121-133)	1377.705	-8.1	Hb D-Punjab	HbAD-Punjab
IH140083_1	NA	NA	NA	no variants	HbAD-Punjab
IH140485_1	NA	NA	NA	no variants	HbAD-Punjab
IH140467_1	NA	NA	NA	no variants	HbAD-Punjab

Patient no. and replicate	Variant (peptide position)	Exp. m/z (Th)	Mass Accuracy (ppm)	Experimental ID	True ID
IH140470_2	HbE (10-27)	1829.983	-4.0	HbE or Hb D-Iran	HbAE
	HbE (10-31)	2227.209	-4.6		
	HbD-Iran (10-31)	2227.209	11.8		
IH140465_1	NA	NA	NA	no variants	HbAE
IH087362_1	NA	NA	NA	no variants	HbAE
IH087365_1	NA	NA	NA	no variants	HbAE
IH140489_1	HbE (10-31)	2227.226	3.1	HbE	HbAE
IH140491_3	HbE (10-27)	1829.964	-1.7		
	HbE (10-31)	2227.201	-8.1		
	HbD-Iran 10-31	2227.201	8.2		
IH14074_2	HbE (19-31)	1313.713	-3.5	HbE	HbEE
	HbE (10-31)	2227.237	8.0		
IH14075_2	HbE (19-27)	916.475	1.3	HbE	HbEE
	HbE (19-31)	1313.714	-2.7		
	HbE (10-27)	1829.985	5.2		
	HbE (10-31)	2227.239	9.0		
IH14079_1	HbE (19-31)	1313.706	-8.6	HbE	HbAE
	HbE (10-31)	2227.211	-3.9		
IH14082_2	HbE (19-31)	1313.713	-3.1	HbE	HbAE

Patient no. and replicate	Variant (peptide position)	Exp. m/z (Th)	Mass Accuracy (ppm)	Experimental ID	True ID
IH140064_3	HbS (2-9)	922.538	2.8	HbS	HbSS
IH140066_1	HbS (2-9)	922.527	9.7	HbS	HbAS
IH140067_1	HbS (2-9)	922.54	4.8	HbS	HbSS
IH140069_2	HbS (2-9)	922.537	1.5	HbS	HbAS
IH140081_1	HbS (2-9)	922.547	12.1	Inconclusive	HbAS
	HbE (19-31)	1313.705	-9.3		
IH087370_1	HbS (2-9)	922.536	-5.9	HbS	HbAS
IH140464_3	HbS (2-9)	922.535	-0.8	HbS	HbAS
IH140471_2	HbS (2-9)	922.538	2.3	HbS	HbAS
IH140487_1	HbS (2-9)	922.542	7.1	HbS	HbAS
IH140086_2	HbS (2-9)	922.537	1.4	HbS	HbAS

Table 24. Data obtained on the Synapt reports the mass accuracy of the putative experimental assignments against the theoretical peptide mass-to-charge ratios, and the subsequent experimental identifications against the actual variant identification. Data for HbA^0 and Hb J-Baltimore is not reported here as no proteotypic peptides were detected in either case.

A batch of 9 further samples was acquired for a comparison study across 3 different mass spectrometers, to discern if MALDI MS instruments with enhanced sensitivity specifications could detect normal- and variant-derived peptides. These instruments were the Waters MALDI Synapt G2, the Bruker rapifleX and a Bruker timsTOF flex. This was to perform a direct empirical comparison of the sensitivity performance, as it is crucial to use the best-performing instrumentation. These specifications are all based on detecting [Glu1]-Fibrinopeptide B (GluFib), at *m/z* 1570.7, and are reported in **Table 25**.

Instrument	Synapt	rapifleX	timsTOF fleX
Limit of Detection	10 fmol	250 amol	150 amol
Signal-to-noise ratio	>90:1	>200:1	>100:1
Laser shots	*Not reported*	2000	2000

Table 25. The manufacturer reported limit of detection for a GluFib signal at *m/z* 1570.7 by 3 top-of-the-range mass spectrometric instruments in their specification documents.

Analyses were again carried out blind, and assigned using the custom built mMass database of proteotypic peptides that had been prepared (**Table 22**). IsoPro 3.1 mass spectral isotopic distribution simulator was used to calculate the molecular formula of each peptide using their amino acid sequence, based on the Yergey algorithm (**Yergey, 1983**).

Table 26 reports the mass accuracy in ppm as yielded by the three instruments. The ttfleX performed the best, correctly detecting and identifying variants in 8 out of the 9 samples, with mass accuracy ranging between -3.2 and 3.1 ppm. The Synapt was able to detect and identify variants in 7 out of the 9 samples, however it performed the least accurate mass measurements, ranging between -8.6 and 12.1 ppm. Finally the rapifleX detected and identified variants in 6 out of the 9 samples, but with better mass accuracy than the Synapt, ranging

between -3.7 and 10.5 ppm. The ttfleX was able to identify sample IH110848 as Hb D-Punjab, while the other two instruments failed to detect peptides relating to this variant. The ttfleX and Synapt were both able to identify sample IH110846 while the rapifleX erroneously identified this sample as containing Hb D-Iran. None of the instruments were able to detect peptides relating to Hb D-Punjab that were present in sample IH110849.

Samples IH110848 and IH110849, which were both Hb D-Punjab variants, were quantified by the Haematology Department as having 41.3% and 41.6% of the Hb D-Punjab variant respectively, with the remainder being 'healthy' HbA^0, and hence lack of detection could not be due to concentration. The concentration values do not help to explain why none of the MALDI mass spectrometer instruments were able to detect the variant in sample IH110849, as this sample had a marginally higher concentration that sample IH110848, in which variant peptides were detected. It is more likely to result from a combination of less that optimal digestion conditions and low ionisation efficiency specifically for these proteotypic peptides, which may generate analyte ions below the LOD. In fact, during preliminary method development, samples known to contain Hb D-Punjab frequently yielded no detectable proteotypic peptides (5 out of 7 samples yielded no distinguishable peptides, and only 1 sample was correctly identified as Hb D-Punjab).

Sample number	ttfleX			rapifleX			Syanpt		
	Peptide/s *m/z*	Mass accuracy (ppm)	Putative assignment	Peptide/s *m/z*	Mass accuracy (ppm)	Putative assignment	Peptide/s *m/z*	Mass accuracy (ppm)	Putative assignment
IH110843	2286.172	-0.1	Hb J-Baltimore	2286.171	-0.4	Hb J-Baltimore	2286.165	3.6	Hb J-Baltimore
IH110844	694.425 951.559	0 3.1	HbC	694.426	-1.4	HbC	951.570	8.2	HbC
IH110845	1865.057	-3.9	HbC	694.426 951.565	-1.4 -3.2	HbC	694.422 591.561	-4.2 -1.1	HbC
IH110846	1829.976 2227.217	0.3 -1.1	HbE	1313.680	-0.6	Hb D-Iran	916.468 1829.956 1313.706	-4.3 -5.6 -8.6	HbE
IH110847	1829.976	0.3	HbE	1313.708 2227.221	7.1 0.7	HbE	916.471 1829.971	-2.2 -2.4	HbE
IH110848	1377.716	-1.3	Hb D-Punjab	n/a	n/a	No variant	n/a	n/a	No variant
IH110849	n/a	n/a	No variant	n/a	n/a	No variant	n/a	n/a	No variant
IH110850	2286.172	-3.2	Hb J-Baltimore	2286.166	-3.5	Hb J-Baltimore	2286.145	12.1	Hb J-Baltimore
IH110851	1313.681 2227.183	-0.2 -0.1	Hb D-Iran	1313.676 2227.180	-3.7 -1.1	Hb D-Iran	1313.674 2227.190	-4.9 3.2	Hb D-Iran

Table 26. The haemoglobin peptides detected with each MALDI instrument and the mass accuracy of each measurement in parts-per-million. Green cells show when the correct variant was identified, red shows incorrect variant putatively identified. Samples were analysed in duplicate on the ttfleX and in triplicate on the rapifleX and Synapt.

MALDI MS/MS analysis was also performed to confirm the identity of the relevant HbVar peptides and hence to confirm the presence of a given HbVar. The tandem MS performances of the Bruker ttfleX and the Waters Synapt G2 instruments were compared. The Synapt was able to acquire confirmatory MS/MS parent ion and fragment signals for samples 843, 850 and 851, with a mass accuracy of between -18.5 and 7.8 ppm. The ttfleX however was able to acquire confirmatory MS/MS parent ion and fragment signals for samples IH110843, IH110844, IH110846, IH110847, IH110848, IH110850 and IH110851 (all samples except IH110845 and IH110849), with a mass accuracy of between -5.0 and 4.9 ppm. For samples IH110845 and IH110849, the parent ion was not detected and therefore not fragmented. The parent ion and fragments for each of the samples detected are displayed below in **Table 27**.

Sample 843	Theoretical m/z 2286.1726	ID: Hb J-Baltimore (10-31)			
Ion fragm.	Peptide sequence	m/z (ttfleX)	accuracy (ppm)	m/z (Synapt)	Accuracy (ppm)
[M+H]⁺	SAVTALWDKVNVDEVGGEALGR	2286.171	-0.8	2286.159	-5.8
b10	.SAVTALWDKV.n	1071.583	0.2	1071.579	-3.8
b11	.SAVTALWDKVN.v	1185.625	-1.0	--	--
b18	.SAVTALWDKVNVDEVGGE.a	1870.917	-0.8	1870.912	-14.0
b21	.SAVTALWDKVNVDEVGGEALG.r	--	--	2112.065	1.8
y1	g.R.	175.119	-1.2	--	--
y4	e.ALGR.	416.260	-4.8	416.261	-1.4
y6	g.GEALGR.	602.324	-2.0	602.324	-3.2
y7	v.GGEALGR.	659.345	-3.3	659.350	3.9
y8	e.VGGEALGR.	758.414	-1.9	758.408	-9.7
y9	d.EVGGEALGR.	887.457	-1.6	887.456	-2.6
y10	v.DEVGGEALGR.	1002.486	1.4	1002.474	-10.9
Y11	v.VDEVGGEALGR.	--	--	1101.535	-17.0
y12	v.NVDEVGGEALGR.	1215.601	3.6	1215.588	-6.8
y13	k.VNVDEVGGEALGR.	1314.666	0.6	1314.668	1.2
y14	d.KVNVDEVGGEALGR.	1442.757	-2.0	1442.752	-5.2
y15	w.DKVNVDEVGGEALGR.	1557.783	-2.1	1557.781	-4.4

Sample 844	Theoretical m/z 694.4251	ID: HbC (2-7)			
Ion fragm.	Peptide sequence	m/z (ttfleX)	accuracy (ppm)	m/z (Synapt)	accuracy (ppm)
b2	.VH.l	237.135	-2.8	--	--
b3	.VHL.t	350.219	-1.3	--	--
b4	.VHLT.p	451.266	-2.3	--	--
b5	.VHLTP.k	548.319	-3.3	--	--
y1	e.K			--	--
y3	l.TPK.	345.213	-4.9	--	--
y4	h.LTPK.	458.297	-2	--	--
y5	v.HLTPK.	595.356	-1	--	--
a2	.VH.l	209.14	-4	--	--
a3	.VHL.t	322.224	-1.8	--	--

Sample 843	Theoretical m/z 2286.1726	ID: Hb J-Baltimore (10-31)			
Ion fragm.	Peptide sequence	m/z (ttfleX)	Accuracy (ppm)	m/z (Synapt)	Accuracy (ppm)
[M+H]+	SAVTALWDKVNVDEVGGEALGR	2286.171	-0.8	2286.159	-5.8
b10	.SAVTALWDKV.n	1071.583	0.2	1071.579	-3.8
b11	.SAVTALWDKVN.v	1185.625	-1.0	--	--
b18	.SAVTALWDKVNVDEVGGE.a	1870.917	-0.8	1870.912	-14.0
b21	.SAVTALWDKVNVDEVGGEALG.r	--	--	2112.065	1.8
y1	g.R.	175.119	-1.2	--	--
y4	e.ALGR.	416.260	-4.8	416.261	-1.4
y6	g.GEALGR.	602.324	-2.0	602.324	-3.2
y7	v.GGEALGR.	659.345	-3.3	659.350	3.9
y8	e.VGGEALGR.	758.414	-1.9	758.408	-9.7
y9	d.EVGGEALGR.	887.457	-1.6	887.456	-2.6
y10	v.DEVGGEALGR.	1002.486	1.4	1002.474	-10.9
Y11	v.VDEVGGEALGR.	--	--	1101.535	-17.0
y12	v.NVDEVGGEALGR.	1215.601	3.6	1215.588	-6.8
y13	k.VNVDEVGGEALGR.	1314.666	0.6	1314.668	1.2
y14	d.KVNVDEVGGEALGR.	1442.757	-2.0	1442.752	-5.2
y15	w.DKVNVDEVGGEALGR.	1557.783	-2.1	1557.781	-4.4

Sample 844	Theoretical m/z 951.5626	ID: HbC (2-9)			
Ion fragm.	Peptide sequence	m/z (ttfleX)	Accuracy (ppm)	m/z (Synapt)	Accuracy (ppm)
MH+	VHLTPKEK	--	--	951.561	-1.5
b3	.VHL.t	--	--	350.218	-2.2
y1	e.K	--	--	147.114	-7.4

Sample 844	Theoretical m/z 951.5626	ID: HbC (2-9)			
Ion fragm.	Peptide sequence	m/z (ttfleX)	Accuracy (ppm)	m/z (Synapt)	Accuracy (ppm)
MH+	VHLTPKEK	--	--	951.561	-1.5
b3	.VHL.t	--	--	350.218	-2.2
y1	e.K	--	--	147.114	-7.4

Sample 846	Theoretical m/z 1313.717	ID: HbC (19-31)			
Ion fragm.	Peptide sequence	m/z (ttfleX)	Accuracy (ppm)	m/z (Synapt)	Accuracy (ppm)
MH+	VNVDEVGGKALGR	--	--	1313.709	-5.9
b3	.VNV.d	--	--	313.388	4.0
b5	.VNVDE.v	--	--	413.267	13.9
y4	k.ALGR.	--	--	557.264	13.9

Sample 846	Theoretical m/z 1829.976	ID: HbE (10-27)			
Ion fragm.	Peptide sequence	m/z (ttfleX)	Accuracy (ppm)	m/z (Synapt)	Accuracy (ppm)
MH+	SAVTALWGKVNVDEVGGK	1829.970	3.3	--	--
b4	.SAVT.a	359.193	0.1	--	--
b6	.SAVTAL.w	543.314	-4.2	--	--
b7	.SAVTALW.g	729.393	-4.7	--	--
b9	.SAVTALWGK.v	914.509	-0.3	--	--
b10	.SAVTALWGKV.n	1013.578	-3.8	--	--
b11	.SAVTALWGKVN.v	1127.621	-1.7	--	--
b14	.SAVTALWGKVNVDE.v	1470.759	-3.7	--	--
b15	.SAVTALWGKVNVDEV.g	1569.827	1.5	--	--
b17	.SAVTALWGKVNVDEVGG.k	1683.87	4.2	--	--
y10	g.KVNVDEVGGK.	1044.568	1.2	--	--
y13	a.LWGKVNVDEVGGK.	1400.753	-2.5	--	--
y15	v.TALWGKVNVDEVGGK.	1572.838	2.3	--	--

267

Sample 847	Theoretical m/z 1829.976			ID: HbE (10-27)	
Ion fragm.	Peptide sequence	m/z (ttfleX)	Accuracy (ppm)	m/z (Synapt)	Accuracy (ppm)
MH+	SAVTALWGKVNVDEVGGK	1829.976	0.1	--	--
a10	.SAVTALWGKV.n	985.586	3.5	--	--
b10	.SAVTALWGKV.n	1013.578	0.1	--	--
b11	.SAVTALWGKVN.v	1127.619	-1.8	--	--
b12	.SAVTALWGKVNV.d	1226.684	-4.4	--	--
b13	.SAVTALWGKVNVD.e	1341.716	-1.5	--	--
b14	.SAVTALWGKVNVDE.v	1470.759	-1.4	--	--
b16	.SAVTALWGKVNVDEVG.g	1626.856	4.6	--	--
b2	.SA.v	159.076	-0.9	--	--
b4	.SAVT.a	359.193	0.4	--	--
b6	.SAVTAL.w	543.312	-3.3	--	--
b7	.SAVTALW.g	729.393	-0.2	--	--
b8	.SAVTALWG.k	786.411	-4.1	--	--
b9	.SAVTALWGK.v	914.509	2.2	--	--
y11	w.GKVNVDEVGGK.	1101.589	-1.2	--	--
y12	l.WGKVNVDEVGGK.	1287.667	-1.4	--	--
y13	a.LWGKVNVDEVGGK.	1400.754	0.3	--	--
y15	v.TALWGKVNVDEVGGK.	1572.836	-1.2	--	--
y6	v.DEVGGK.	604.291	-4.1	--	--
y7	n.VDEVGGK.	703.366	4.9	--	--
y8	v.NVDEVGGK.	817.405	0.3	--	--
y9	k.VNVDEVGGK.	916.471	-2.1	--	--

Sample 848	Theoretical m/z 1377.7164			ID: Hb D-Punjab (121-133)	
Ion fragm.	Peptide sequence	m/z (ttfleX)	Accuracy (ppm)	m/z (Synapt)	Accuracy (ppm)
MH+	QFTPPVQAAYQK	1377.712	-2.8	N/A	N/A
b2	.QF.t	276.134	-1.4	N/A	N/A
b3	.QFT.p	377.181	-2.8	N/A	N/A
b4	.QFTP.p	474.234	-1.8	N/A	N/A
b5	.QFTPP.v	571.289	2.7	N/A	N/A
b6	.QFTPPV.q	670.355	-1.9	N/A	N/A
b7	.QFTPPVQ.a	798.413	-2.4	N/A	N/A
b8	.QFTPPVQA.a	869.448	-4.1	N/A	N/A
b9	.QFTPPVQAA.y	940.491	-3.4	N/A	N/A
b10	.QFTPPVQAAY.q	1103.550	-1.7	N/A	N/A
b11	.QFTPPVQAAYQ.k	1231.614	2.6	N/A	N/A
y2	y.QK.	275.171	-2.5	N/A	N/A
y3	a.YQK.	438.235	-0.5	N/A	N/A
y4	a.AYQK.	509.270	-2.7	N/A	N/A
y6	v.QAAYQK.	708.368	0.2	N/A	N/A
y8	p.PVQAAYQK.	904.491	2.4	N/A	N/A
y9	t.PPVQAAYQK.	1001.540	-1.1	N/A	N/A
y10	f.TPPVQAAYQK.	1102.588	-1.1	N/A	N/A
y11	q.FTPPVQAAYQK.	1249.653	-3.3	N/A	N/A
a7	.QFTPPVQ.a	770.420	0.1	N/A	N/A

Sample 850	Theoretical m/z 1313.6812			ID: Hb D-Iran (19-31)	
Ion fragm.	Peptide sequence	m/z (ttfleX)	Accuracy (ppm)	m/z (Synapt)	Accuracy (ppm)
MH+	.VNVDQVGGEALGR.	1313.677	-2.8	N/A	N/A
b2	.VN.v	214.118	-2.5	N/A	N/A
b6	.VNVDQV.g	655.339	-2.9	N/A	N/A
b8	.VNVDQVGG.e	769.383	-1.2	N/A	N/A
y3	a.LGR.	345.223	-3.6	N/A	N/A
y4	e.ALGR.	416.260	-4.8	N/A	N/A
y6	g.GEALGR.	602.324	-2.2	N/A	N/A
y7	v.GGEALGR.	659.345	-2.8	N/A	N/A
y8	q.VGGEALGR.	758.414	-1.9	N/A	N/A
y9	d.QVGGEALGR.	886.471	-3.4	N/A	N/A
y10	v.DQVGGEALGR.	1001.503	2.3	N/A	N/A
y11	n.VDQVGGEALGR.	1100.568	-3	N/A	N/A
a5	.VNVDQ.v	528.278	-0.9	N/A	N/A
a6	.VNVDQV.g	627.345	-1.4	N/A	N/A
a8	.VNVDQVGG.e	741.389	0.2	N/A	N/A

Sample 850	Theoretical m/z 2286.1726			ID: Hb J-Baltimore (10-31)	
Ion fragm.	Peptide sequence	m/z (ttfleX)	Accuracy (ppm)	m/z (Synapt)	Accuracy (ppm)
MH+	SAVTALWDKVNVDEVGGEALGR	2286.129	-1.7	2286.159	-5.8
b5	.SAVTA.l	430.227	-5	--	--
b6	.SAVTAL.w	543.311	-4.2	--	--
b10	.SAVTALWDKV.n	1071.583	-0.7	--	--
b11	.SAVTALWDKVN.v	1185.623	-2.8	--	--
b12	.SAVTALWDKVNV.d	1284.693	-1.6	--	--
b13	.SAVTALWDKVNVD.e	1399.722	-1.9	--	--
b15	.SAVTALWDKVNVDEV.g	1627.831	-0.8	--	--
b18	.SAVTALWDKVNVDEVGGE.a	1870.923	2.4	--	--
b21	.SAVTALWDKVNVDEVGGEALG.r	2112.056	-2.4	--	--
y4	e.ALGR.	416.262	-4.4	--	--
y6	g.GEALGR.	602.324	-2.2	602.324	-3.0
y7	v.GGEALGR.	659.345	-2.6	659.338	13.1
y8	e.VGGEALGR.	758.414	-1.9	758.420	6.3
y9	d.EVGGEALGR.	887.455	-3	887.456	-2.6
y10	v.DEVGGEALGR.	1002.484	-0.7	--	--
y11	n.VDEVGGEALGR.	1101.549	-4.3	--	--
y12	v.NVDEVGGEALGR.	1215.595	-1.3	--	--
y13	k.VNVDEVGGEALGR.	1314.661	-2.8	--	--
y14	d.KVNVDEVGGEALGR.	1442.756	-2.7	1442.760	6.3
y15	w.DKVNVDEVGGEALGR.	1557.789	1.5	--	--
y16	l.WDKVNVDEVGGEALGR.	1743.866	0.1	--	--
y18	t.ALWDKVNVDEVGGEALGR.	1927.986	-0.8	--	--

Sample 851	Theoretical m/z 2227.1831			ID: Hb D-Iran (10-31)	
Ion fragm.	Peptide sequence	m/z (ttfleX)	Accuracy (ppm)	m/z (Synapt)	Accuracy (ppm)
MH+	SAVTALWGKVNVDQVGGEALGR	2227.178	-2.2	2227.178	-2.2
b5	.SAVTA.l	430.230	-2.3	--	--
b6	.SAVTAL.w	543.314	-2.6	543.309	-8.1
b7	.SAVTALW.g	729.393	-2.2	--	--
b8	.SAVTALWG.k	786.414	-2.2	--	--
b9	.SAVTALWGK.v	914.509	-2.5	--	--
b10	.SAVTALWGKV.n	1013.578	-2.4	--	--
b11	.SAVTALWGKVN.v	1127.621	-2.3	1127.604	-14.8
b12	.SAVTALWGKVNV.d	1226.689	-3.5		
b13	.SAVTALWGKVNVD.q	1341.716	-2	1341.727	7.8
b15	.SAVTALWGKVNVDQV.g	1568.843	-1.9	--	--
b18	.SAVTALWGKVNVDQVGGE.a	1811.929	-1.9	--	--
b21	.SAVTALWGKVNVDQVGGEALG.r	2053.071	-2.1	--	--
y4	e.ALGR.	416.262	-4.3	416.261	-1.4
y5	g.EALGR.	545.304	-0.4	--	--
y6	g.GEALGR.	602.326	-2.2	--	--
y7	v.GGEALGR.	659.347	-2.8	--	--
y8	q.VGGEALGR.	758.416	-2.1	--	--
y10	v.DQVGGEALGR.	1001.501	-1.4	--	--
y11	n.VDQVGGEALGR.	1100.569	-1.8	--	--
y12	v.NVDQVGGEALGR.	1214.612	-2.7	--	--
y13	k.VNVDQVGGEALGR.	1313.681	-2.1	--	--
y14	g.KVNVDQVGGEALGR.	1441.776	-2.3	--	--
y15	w.GKVNVDQVGGEALGR.	1498.797	-1.9	--	--
y16	l.WGKVNVDQVGGEALGR.	1684.877	1.6	--	--
y17	a.LWGKVNVDQVGGEALGR.	1797.961	0.2	--	--
a7	.SAVTALW.g	701.398	-0.1	--	--
a8	.SAVTALWG.k	--	--	758.420	-1.0
a10	.SAVTALWGKV.n	985.583	-0.9	--	--
a12	.SAVTALWGKVNV.d	1198.694	-2.2	--	--
c21	.SAVTALWGKVNVDQVGGEALG.r	2070.098	-1.3	--	--

Table 27. Parent ion and fragment MS/MS assignments of haemoglobin peptides in patient blood samples. No data is shown for samples IH110845 and IH110849 due to no peptide parent ions being detected. Additionally no data is shown for sample IH110848 on the Synapt as no variants were detected.

Figure 50 shows MS and MS/MS spectra for sample IH110843 acquired on the Synapt and ttfleX, which were the Synapt spectra which yielded the most peptide fragments. However, the ttfleX demonstrated better sensitivity out of the two instruments in both MS and MS/MS mode, indicated by the higher intensity for the signals obtained by the ttfleX. **50a** and **50b** compare the MS spectra obtained by the ttfleX and Synapt, respectively. **50c** and **50d** compare the MS/MS spectra obtained by the ttfleX and Synapt, respectively.
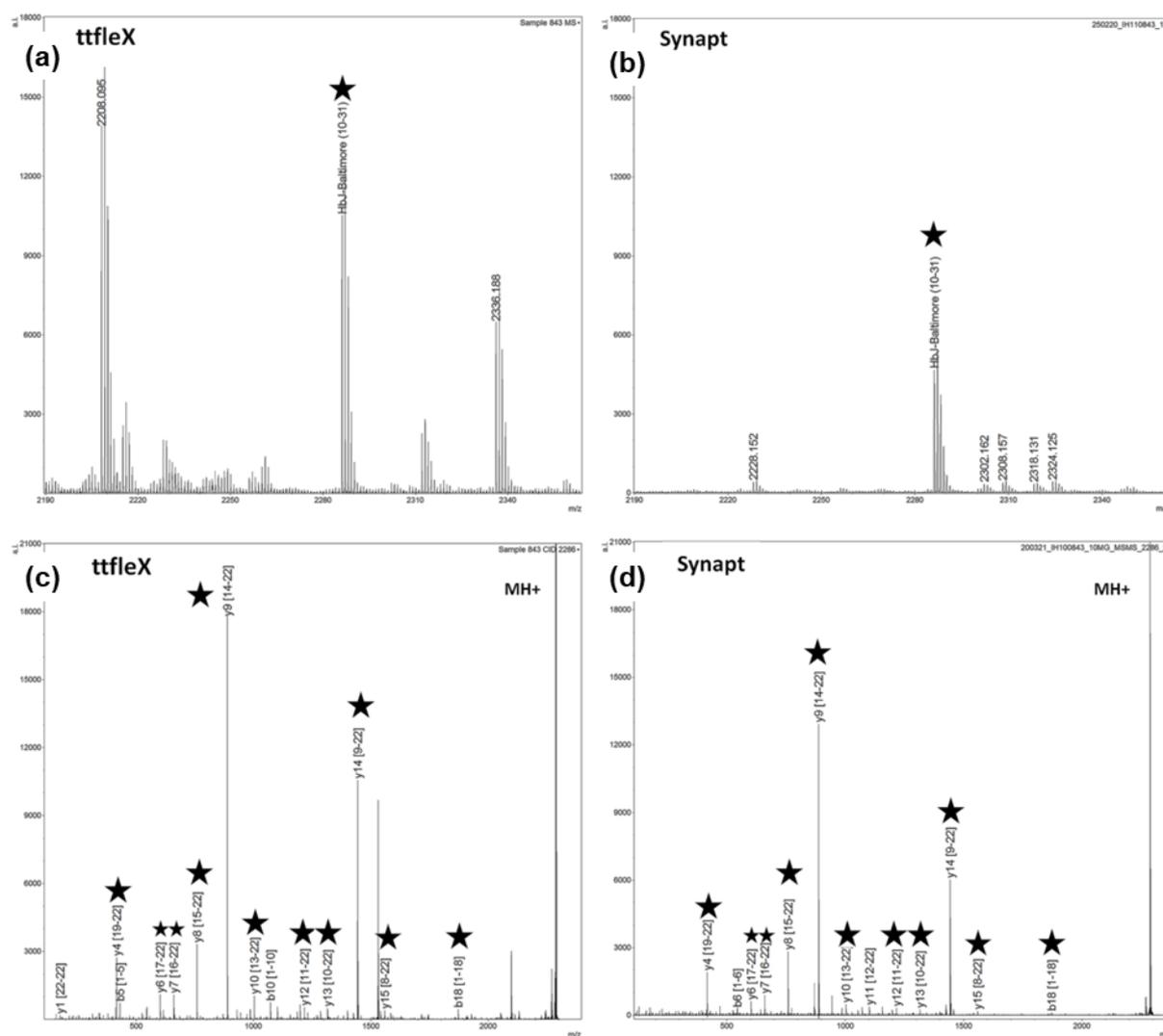

Figure 50. MS and MS/MS spectra from the Synapt and the ttfleX for the Hb J-Baltimore proteotypic peptide (sequence: SAVTALWDKVNVDEVGGEALGR, position: 10-31). (2a) shows the MS spectra obtained on the ttfleX. (2b) shows the MS spectra obtained on the Synapt. (2c) shows the MS/MS spectra obtained on the ttfleX. (2d) shows the MS/MS spectra obtained on the Synapt.

When acquiring in MS mode, the peak for the Hb J-Baltimore peptide present at nominal *m/z* 228 was twice as intense on the ttfleX compared to the Synapt. In addition, 14 ion fragments were detected with the ttfleX compared to 13 with the Synapt. Eleven of these were detected by both instruments. Although the disparity is not significant in terms of variant identification (as both were above the S/N threshold), the ttfleX was able to detect these ion signals with a greater intensity than the Synapt. Finally, it was noted that for samples IH110846 and IH110851, which contained variants HbE and Hb D-Iran respectively, the ttfleX was able to distinguish between the amino acid sequence in positions 10-31 within the proteotypic peptides of these variants. These proteotypic peptides only differed by a single amino acid substitution at different positions within the amino acid sequence. The HbE proteotypic peptide sequence SAVTALWGKVNVDEVGG**K**ALGR and the Hb D-Iran proteotypic peptide sequence SAVTALWGKVNVD**Q**VGGEALGR have theoretical *m/z* values of 2227.2197 and 2227.1831 respectively, resulting in a 0.037 Da (16.4 ppm) mass variation. The Synapt did not have sufficient resolution to discriminate between these two similar signals. The ion fragments detected and identified through MS/MS by the Synapt and ttfleX are given in **Table 24**.

To investigate specificity, the UniProtKB/Swiss-Prot protein database was searched against to determine if the proteotypic peptides identified for each variant could only have originated from the human β haemoglobin protein variants, and not from other human proteins. It must also be established whether they were specific to humans i.e. not present in other mammals, specifically those whose blood might reasonably be present in a domestic environment, such as porcine, bovine, equine, canine or feline. At the time of searching, in January 2021, UniProt contained 564,277 manually annotated and

reviewed proteins, comprising 14,014 species. The native peptide search tool was utilised to cross-reference the proteotypic haemoglobin variant peptides mass values. 8 out of the 19 total haemoglobin proteotypic peptides resulted in a match against the database; 7 of these were against human haemoglobin. The other was a match from the HbC peptide (amino acid sequence position 2-7, sequence VHLTP**K**, *m/z* 694.4251). This amino acid sequence was a match for peptides present in human thyroid adenoma-associated protein, as well as proteins within fungi, metazoa and bacteria. In humans, this thyroid adenoma-associated protein is expressed within the pancreas, adrenal medulla, thyroid, adrenal cortex, testis, thymus, small intestine and stomach, so its potential presence in the blood is likely to come from haemoglobin protein. There is a possibility that its potential presence in blood might originate from one of these organs if they were severed in the injury that caused the bloodshed, but this is outside the scope of the study. However, the modification region is potentially present in two other proteotypic peptides within the HbC protein, if allowing for up to 2 missed cleavages. These are at amino acid positions 2-9 and 2-18, giving *m/z* values of 951.5626 and 1865.0646 respectively (**Table 22**), which should be used to confirm the suspected presence of HbC rather than using the *m/z* 694 peptide alone. Additionally, the native Basic Local Alignment Search Tool (BLAST) feature in UniProt, which finds regions of similarity in sequences, was searched with the haemoglobin variant proteotypic peptide sequences. This tool can give an indication of evolutionary relationships between sequences and link gene families. The amino acid sequence in positions 2-7 within the HbC peptide was unique to the haemoglobin variant. This was also true for all of the other variant peptides, with the exception of HbE [10-27] and Hb J-Baltimore [10-31]; the HbE peptide was also found to be present in

primate β haemoglobin, and the Hb J-Baltimore peptide was also present specifically in the *Gorilla Gorilla* species.

While the detection of rare haemoglobin variants may help to inform an investigation, by helping to rule in or out that blood could have originated from a specific individual or not, more weight could be attributed to that evidence if the haemoglobin variant molecules could be mapped directly onto the ridge pattern of a bloody fingermark. This could link an individual, by identification through biometric fingermark information, to the time of (or very soon after) the bloodshed event. An experiment was designed to probe the feasibility of visualising haemoglobin variants within bloody fingermarks, utilising MALDI MS imaging. A preliminary experiment involved depositing two adjacent blood stains, one containing 'normal' HbA^0 blood and one containing 'sickle cell' HbS blood to see if they could be distinguished. These were digested, incubated and had matrix applied *in situ* as per in the experimental section (**Fig. 51**). These were then analysed via raster imaging on the same MALDI Synapt G2 as the profiling experiments.

Figure 51. Photograph of the adjacent blood stains with the sample on the left originating from a patient with 'normal' HbA^0 blood, and the sample on the right originating from a patient with HbS blood. Samples were deposited on a glass slide, and mounted on the MALDI target plate with double-sided carbon tape. The outline of the blood stains have been added to aid visualisation as they are faint in the image.

Figure 52 shows the result of the molecular images generated when selecting the nominal *m/z* 1275 βHb peptide (amino acid sequence: LLVVYPWTQR, position: 32-41) and *m/z* 1530 αHb peptide (amino acid sequence: VGGHAAEYGAEALER, position: 18-32) ion common to both blood samples. It also shows the molecular images generated when selecting the *m/z* 923 variant peptide (amino acid sequence: VHLTPVEK, position: 2-9) ion, indicating the presence of HbS in the bloodstain on the left, but not the right. This molecular image also revealed the presence of an as yet unidentified peptide at nominal *m/z* 761 in the bloodstain on the right, (a very faint trace of this ion can be seen in the location where sample 1 was deposited, but with a very low intensity).
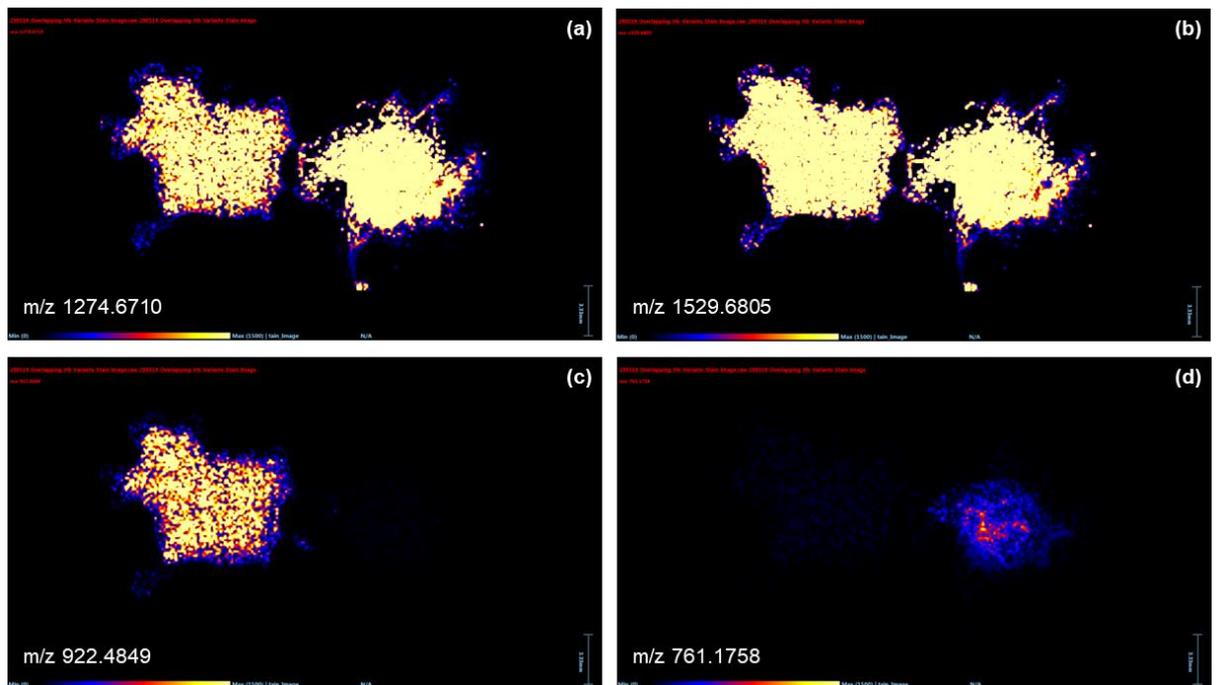
Figure 52. Shows four molecular ion images for two adjacent blood stains, with the sample on the left originating from a patient with HbS-containing blood, and the sample on the right originating from a patient with HbA^0-containing blood. (a) shows the presence of an ion at *m/z* 1274.671, resulting from a β haemoglobin peptide, present in both samples (patient 1 and 2). (b) shows the presence of an ion at *m/z* 1529.681, resulting from an haemoglobin peptide, also present in both samples (patient 1 and 2). (c) shows the presence of an ion at *m/z* 922.485, resulting from a proteotypic HbS peptide, which is only present in the sample on the left (patient 1). (d) shows the presence of an unidentified ion at *m/z* 761.176, which is only clear in the sample on the right (patient 2). This could not be assigned to a variant peptide, but demonstrates the ability of MALDI MSI to detect molecules unique to one individuals blood within the same analysis.

As can be seen from **Fig. 52**, the two bloodstains could be easily identified as blood from the detection of the α and β H peptides at *m/z* 1529.681 and 1274.671, respectively (in addition to the β haemoglobin peptide present in the spectra at *m/z* 922.485), but crucially, the two blood stains from different individuals could easily be discriminated between by the presence/absence of the proteotypic peptide in the sample on the left from the patient with sickle cell anaemia, compared to the sample on the right from a patient with 'normal' haemoglobin.

Following the successful visualisation of haemoglobin variants within adjacent blood stains, the experiment was repeated, this time using overlapping fingermarks contaminated with blood containing the same two variants, HbA^0 and HbS (**Fig. 53**), which were then prepared in an identical way to the blood stains, and analysed using the Synapt G2 MALDI instrument in raster imaging mode.

Figure 53. Overlapping blood-contaminated fingermarks deposited on at TLC sheet, mounted on a Synapt target plate. The mark on the left was contaminated with blood from a patient with 'normal' HbA^0 haemoglobin and the mark on the right was contaminated with blood from a patient with sickle cell anaemia, HbS.

Unfortunately, this experiment was unsuccessful in visualising haemoglobin variant peptides in a molecular ion image of the fingermark ridge pattern. Subsequently, individual marks were deposited on the same TLC substrate containing the six variants previously used in this study and imaged using the more sensitive ttfleX MALDI instrument.

Generally, the haemoglobin variant peptides that had been detected via MS (along with confirmatory MS/MS analysis) on the ttfleX were also detected via MALDI MSI in imaging mode. It was possible to map these variant peptides onto the fingermark minutiae when overlapping the molecular images obtained over optical images of the fingermarks (taken using a Nikon Cool Scan slide imager). The results of this experiment are shown in **Fig. 54**, showing proteotypic peptide molecular ion images for HbC [2-7] at *m/z* 694.420 (**i**), HbC [2-9] at *m/z* 951.563 (**ii**), HbE [10-27] at *m/z* 1829.976 (**iii**), HbE [10-31] at *m/z* 2227.220 (**iv**), HbD-Punjab [121-133] at *m/z* 1377.719 (**v**), HbD-Iran [19-31] at *m/z* 1313.687 (**vi**), HbD-Iran [10-31] at *m/z* 2227.183 (**vii**), HbJ-Baltimore [10-31] at *m/z* 2286.173 (**viii**) and HbS [2-9] at *m/z* 922.536 (**ix**), within a mass accuracy of +/-10 ppm.
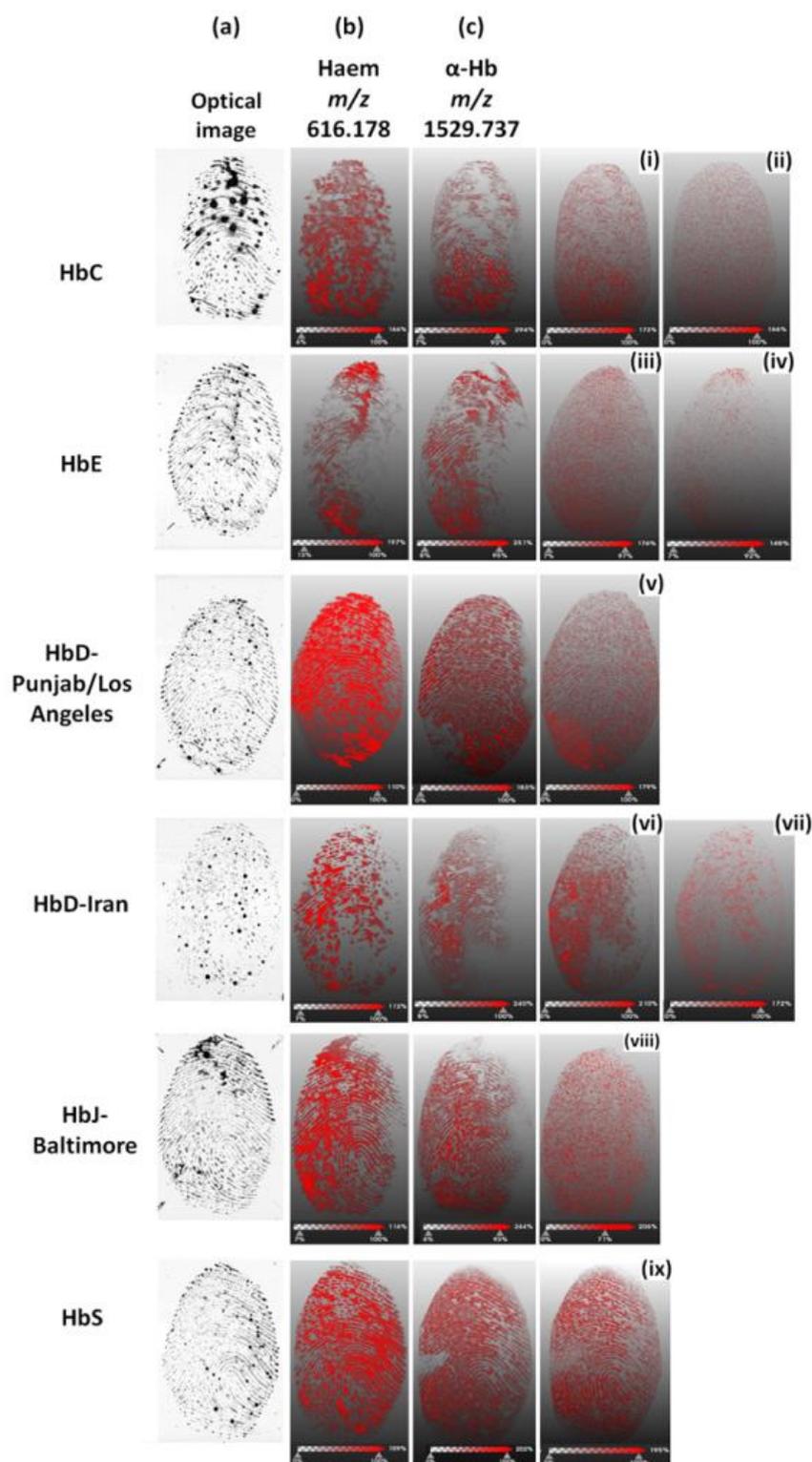
Figure 54. MALDI MSI molecular ion images for marks containing each of the haemoglobin variants investigated in the study. Column a shows an optical image for each prepared mark. Column b shows the haem signal for each mark. Column c shows the αHb chain at *m/z* 1530 for each mark. Proteotypic peptide signals are also shown; HbC [2–7] at *m/z* 694.420 (i), HbC [2–9] at *m/z* 951.563 (ii), HbE [10–27] at *m/z* 1829.976 (iii), HbE [10–31] at *m/z* 2227.220 (iv), Hb D-Punjab [121–133] at *m/z* 1377.719 (v), Hb D-Iran [19–31] at *m/z* 1313.687 (vi), Hb D-Iran [10–31] at *m/z* 2227.183 (vii), Hb J-Baltimore [10–31] at *m/z* 2286.173 (viii) and HbS [2–9] at *m/z* 922.536 (ix).

The *m/z* 992.5 HbS [2-9] proteotypic peptide molecular image gave the most comprehensive ridge detail coverage, which can be attributed to the enhanced ionisation efficiency of this peptide's size and sequence, being the second-lowest molecular weight, but with more readily ionisable amino acids that the lowest molecular weight peptide, *m/z* 694.4 HbC [2-7]. The other proteotypic variant peptide which gave very comprehensive coverage of the ridge detail in the molecular ion image was the *m/z* 1377.7 Hb D-Punjab [121-133]. Although a higher molecular weight peptide, it may also have enhanced ionisation efficiency. Ionisation efficiency is a complicated function of many characteristics of the peptide and matrix.

Whilst it was attempted to standardise the deposition conditions of each replicate mark (same analyst depositing marks with the same replica silicon finger, same volume of blood, approximately the same deposition pressure, same substrate surface, same digestion and incubation conditions, same matrix application, same analysis parameters), there was slight variation in the resulting blood-contaminated fingermarks. Some marks were fainter than others and several had some areas of blood pooling within the marks. The two replicates which generated the most complete molecular ion images were two of the faintest marks with minimal blood pooling and high clarity of the optical images. It could be concluded that these fainter marks, which did not exhibit blood pooling, provided a more optimum ratio between the enzyme/analyte which allowed effective digestion of the blood peptides by the trypsin. Conversely, the darker marks may have resulted in an excess of analyte, which meant that the enzyme was outweighed by the quantity of blood present in the mark, that a) added noise to the spectral profiles and b) reduced the relative

quantity of peptides generated, making them harder to detect against the background with MALDI MSI.

However, an exception to this observation would be the results for the Hb J-Baltimore peptide molecular image. The original mark appears visually similar to the HbS and Hb D-Punjab in the optical images, but the ion mapping distribution appears incomplete and fragmented. It is hypothesised that the Hb J-Baltimore peptide at m/z 2286.2 has a lower ionisation efficiency than HbS and Hb D-Punjab, due to its amino acid sequence and analyte-matrix interactions. It was observed as a general trend that in both MS and MSI analysis, peptides exceeding m/z 2000 were detected with lower relative intensity than the low molecular wight peptides. CHCA ionises peptides better than intact proteins, and a different matrix such as sinapinic acid is needed for the analysis of proteins.

In addition to the proteotypic peptides specific to each variant, **Fig. 54** also displays the molecular ion map of haem at m/z 616.178 and the HbA^0 α-chain peptide at m/z 1529.737. The haem ion maps were generally of better quality (more intense signal and more ridge detail coverage) than the variant-specific peptide ion signals. This could be because the haem molecule is more readily ionisable being a much smaller compound, or because there is a greater quantity of haem in the sample compared to the variant peptides. This would support the observation that the use of haem can help to generate the biometric information whilst the visualisation of HbVar provides additional physiological intelligence. Another interesting observation is that in the Hb D-Punjab molecular ion image, the region where the haem molecule is detected with high intensity, the variant peptide is detected with low intensity and vice versa. This suggests that the presence of haem might be suppressing the detection of the

variant peptide, either in terms of the haem signal overwhelming the variant peptide signal, or because of preferential ionisation of the smaller molecule in that region.

While the haem signal provides a better quality molecular image of the *minutiae* than the variant peptide signals, the clarity is slightly below what can be achieved with conventional enhancement techniques and digital photography, as modern digital cameras are able to achieve smaller image pixel sizes than current state-of-the-art MALDI MSI. The reduced image quality is partially due to sample preparation; the inflexibility of the silicon model finger, lack of natural eccrine and sebaceous material that would be present on a finger and would act as an adhesive, and the uneven distribution of blood in this pseudo-realistic setting, and partially due to inherent instrumental limitations; laser spot size, raster width and pixel size. Silicon model fingers were used in this study to mitigate the biohazard risk of applying patient blood directly to the researcher's fingers, despite their limitations, and as a result, the marks deposited were not of optimum quality. However this is somewhat representative of crime-scene marks, which are also unlikely to be ideal fingermarks, which may be partial, smudged or overlapping.

Overlapping marks are frequently encountered at crime scenes. The overlayed ridge patterns pose a real challenge to successful biometric identification, so marks with overlapping *minutiae* are often discarded or simply not recovered. This scenario was mimicked in a subsequent experiment. Overlapping marks were deposited using replica silicon fingers; one of which had been contaminated with blood from a patient with the HbC variant and the other by using blood from a patient with the Hb J-Baltimore variant (**Fig. 55**). These

overlapping marks were enhancing with Acid Black 1 to also assess compatibility of the method with a frequently used BET.
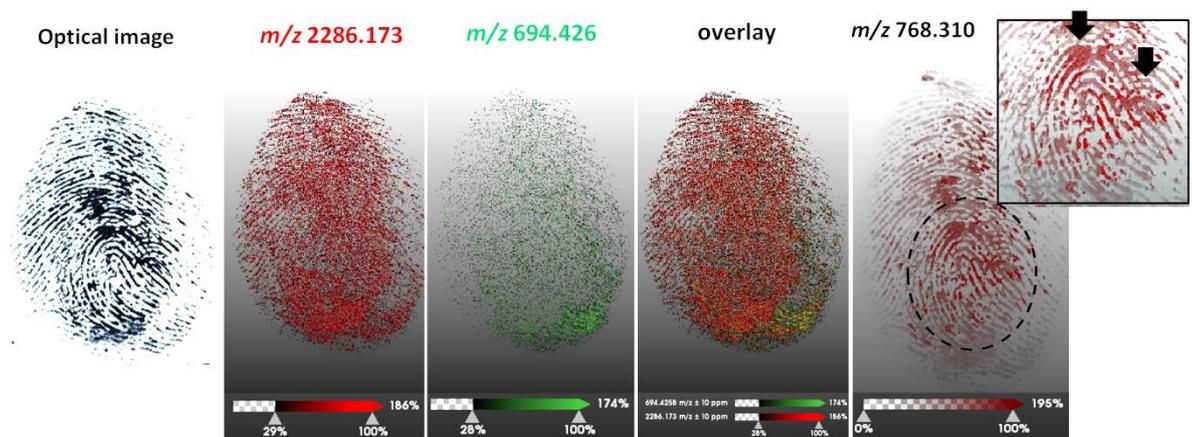

Figure 55. Optical image of the overlapping marks, with an *m/z* 2286 molecular ion map attributed to HbC, an *m/z* 694 molecular ion map attributed to Hb J-Baltimore, and an overlayed ion image of both variant peptides, demonstrating the ability to resolve overlapping marks using variant haemoglobin peptides specific to each mark. In addition, a prominent blood peptide at *m/z* 768 has been included, which clearly shows the presence of 2 ulnar loops, (enlarged, inset) to confirm the presence of two overlapping marks.

The marks were deposited almost immediately after one another, without giving the first deposition time to dry. This was to more accurately replicate what might be encountered operationally, where multiple blood shed events might occur simultaneously as the result of a violent interaction, rather than further bloodshed occurring after an initial bloody mark had fully dried. It is likely that depositing a second mark on top of a wet mark led to some blood mixing at the interface where the two marks made contact, making the resolving of the ridge patterns more difficult. Conventional BETs would not be able to resolve overlapping marks, as they target molecules (haem, amino acids, proteins) that would be common to any source of blood, even if that blood came from different individuals. However, unlike these conventional BETs, which have a limited range of specific target molecules, MALDI offers the ability to probe the composition of a mark, providing information on molecules which fall outside the

three categories of BETs. In addition, it has been demonstrated that MALDI MS can be utilised to differentiate blood from different donors through detection and identification of HbVar, a capability that presumptive colorimetric tests does not offer.

In this experiment, these marks could be partially resolved visually by MALDI MSI, although the ridge detail was not of high quality. The partial resolution was probably due to blood mixing and hence the visualisation of HbVar peptides in some areas of both marks. However, the experiment also did confirm the possibility to detect two HbVar in two different overlapping fingermarks, the partial separation of which could permit biometric identification in association with physiological intelligence.

While these experimental conditions would also apply to crime scene marks, the use of a silicon finger presents a further limitation that would not be encountered operationally. The lack of naturally occurring oils (eccrine pores from sweat pores in the palms and sebaceous oils from the finger's incidental contact with the hair and face) which would ordinarily act as a kind of adhesive were not present to 'stick' the blood to the substrate. This, in combination with the reduced flexibility of the replicate fingers, resulted in marks with less clarity than those deposited with human fingers. Furthermore, to accurately reproduce the diversity of crime-scene marks would require contaminating the fingertips of many donors (to account for inter-donor variability) with patient blood, which could not be ethically justified in a research environment due to health and safety concerns arising from the biohazards associated with handling biofluids. However, further studies could examine artificially impregnating the silicon fingers with sebaceous material to achieve both more realistic and higher clarity marks.

In this proof-of-concept study, MALDI MSI was able to resolve two overlapping fingermarks in blood, a feat not possible with conventional visualisation techniques. While the clarity of the minutiae obtained in the molecular images was not sufficient to perform a biometric identification, it highlights the ability of mass spectrometric imaging to detect differing components within fingermarks, in a less destructive method than swabbing and performing LC/MS, as the ridge detail conformation is not disturbed. It has also demonstrated compatibility with a common fingermark enhancement technique (AB1) establishing its potential to be used as a tool within the fingermark visualisation workflow. The method presented in this study has emphasised the effective digestion, ionisation, detection and identification of proteotypic haemoglobin peptides, to aid with the inclusion/exclusion of suspects, and mapped these onto fingermarks, a potential source of biometric identification. In the final experiment, the detection of peptides resulting from both HbC and Hb J-Baltimore reveal that the blood of at least two individuals is present.

The NHS scientist co-authoring the publication of this study (**Heaton *et al.,* 2021**) at Sheffield Royal Hallamshire Hospital had made the observation over several years of analysing the haemoglobin variants of patients from the surrounding area. They noted that blood samples containing the Hb D-Punjab variant almost exclusively came from patients located in the Rotherham area, specifically from a region with a large South Asian population (predominantly of Pakistani heritage when the Hb D-Punjab variant is present in approximately 0.8% of the population). Similarly, they noticed that blood samples containing the Hb J-Baltimore variant came almost exclusively from the Greater Leeds area (*Dr Jason Eyre, personal communication, 2020*). Some variants are more prevalent in certain areas of the world, for example, "β-thalassemias are the

most common autosomal recessive disorder in populations of Mediterranean, Middle and Far Eastern, Asian/Indian and African descent" (**Patrinos *et al.,* 2004**). It is hypothesised that using global haemoglobin variant detection, individual's biogeographical provenance could be inferred. For example, sickle cell anaemia is known to give a level of protection against malaria (**Weatherall *et al.,* 2006**), so, as a result of evolution, is prevalent in tropical areas where mosquitos are found; Africa and Asia, Central and South America and parts of the Middle East. Sickle cell haemoglobin is found in communities "throughout Sub-Saharan Africa, the Indian subcontinent, and the Middle East, where carrier frequencies range from 5 to 40 percent or more" (**Weatherall *et al.,* 2006**), highlighted in **Fig. 55b**. In addition, HbE is prevalent "in the eastern half of the Indian subcontinent and throughout Southeast Asia, where carrier rates may exceed 60 percent" (**Weatherall *et al.,* 2006**). Weatherall also notes that "There is now abundant evidence that protection against the severe complications of *Plasmodium falciparum* malaria is afforded to heterozygotes for the sickle cell or HbC genes and for heterozygotes and homozygotes for the mild form of α thalassemia, that is, α+ thalassemia. There is also evidence, though not so strong, that the same applies to β thalassemia and HbE" (**Weatherall, 2013**). Their global distribution is highlighted in **Fig. 55a**. Whilst it would not be uncommon to find patients with variant haemoglobin in these regions, outside of these areas, where the variant is less common, its presence in a patient would give an indication of their heritage. Some of the rarer haemoglobin variants, like HbC, originated in very specific locations, and would indicate that an individual with this variant had heritage originating in West Africa. This might be useful information in a forensic context to add to a suspect suspect/victim profile if blood containing rare variants was detected at a crime scene.
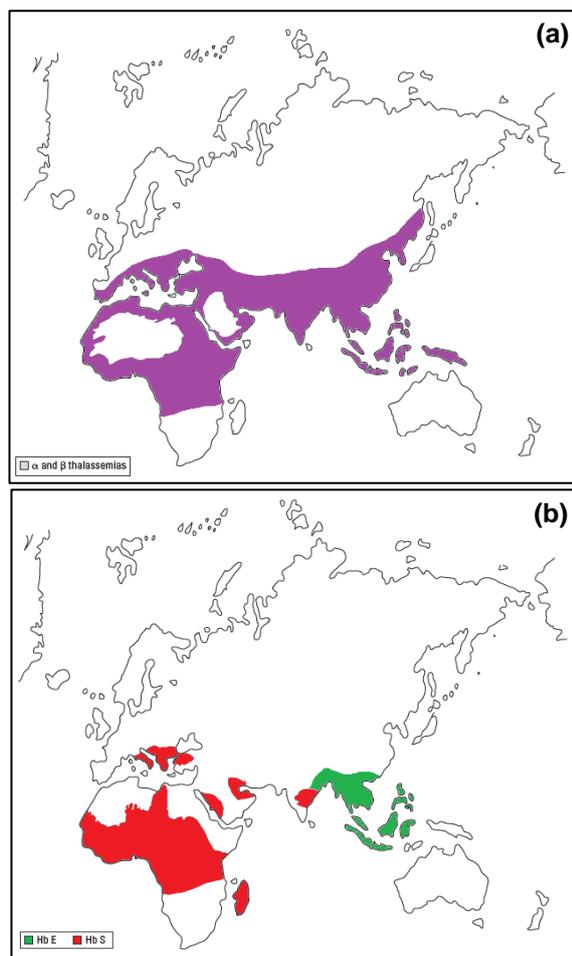
Figure 56. World map showing the distributions of Haemoglobin variants. (a) the general global distribution of α and β thalassemias, shown in purple. (b) the global distribution of HbE, shown in green, and HbS, shown in red [*adapted from Weatherall et al., 2006*].

However, a systematic study of the geographical provenance of haemoglobin variants is required in order for the hypothesis to be tested. Furthermore, globalisation has accelerated since the 1700s, due to the greater accessibility of travel resulting from infrastructure and technological advances, and this might have had some effect on the incidence and distribution of given variants.

Conclusion

This work has presented a MALDI-based proteomics method for the confirmatory analysis of the six most common haemoglobin variants. Although

CE-HPLC is currently the gold standard for haemoglobin discrimination in a clinical setting, the destructive nature of obtaining the sample is less than ideal in a forensic context. It is also laborious in both sample preparation and analysis, and is a time consuming process, that requires complex interpretation. MALDI MSI analysis is only a partially destructive process, ablating a few micrometres from the sample surface, and leaving the ridge pattern confirmation intact when analysing fingermarks, which would have to be swabbed (destroying the ridge pattern) to perform an HPLC analysis. Furthermore, MALDI requires relatively minimal sample preparation, requiring the spray application of enzyme and matrix, with an incubation step. Finally, interpretation of digested haemoglobin spectra would be relatively simple, if compared with a pre-determined database of potential proteotypic haemoglobin variant peptides. This would require a much larger-scale study to first calculate all of the potential peptides for all known variants, ensure there were no identical molecular masses across all known variants, and to empirically investigate whether these proteotypic peptides were ionisable and detectable.

This study has presented a preliminary investigation into the profiling and imaging of a subset of haemoglobin variants, and compared the results obtained by three different MALDI mass spectrometer instruments, with varying specifications, particularly in regards to sensitivity. The detection of proteotypic peptides specific to a single variant through MALDI MS profiling analysis is a rapid and robust confirmatory test that can reveal the presence of rare haemoglobin variants in blood stains, and could help to include/exclude suspected individuals as the source of the bloodshed.

Furthermore, this study has demonstrated how molecular 'chemical' imaging can be combined with physical 'optical' imaging. The MALDI MS molecular

images can be superimposed over optical images in a 1:1 ratio, to visualise the distribution of analytes within an exhibit such as a bloody fingermark, which may be absent from the background. This would provide a link between an individual potentially identified through biometric fingerprint matching, and, crucially, their contact with the analytes detected. The imaging experiments revealed that haem and non-variant haemoglobin peptides (such as those at nominal *m/z* values 1275 and 1530) generated fingermark images characterised by higher ridge pattern coverage, likely due to their greater quantities within the blood and/or greater ionisation efficiency. The detection of variant peptides are less useful for the reconstruction of ridge detail, but could be used to include/exclude individuals as the source of the bloodshed by detecting the HbVar which could be cross checked with medical records, to include/exclude a suspect from an investigation.

The molecular ion images generated by MALDI MSI analysis provided a slight reduction in clarity as a result of the limitations surrounding the artificial deposition of pseudo-operational marks (using replica fingers, lack of sebaceous material). Acquisition time and file sizes are also constraints (better spatial resolution could be achieved by setting smaller laser spot sizes, smaller pixel dimensions and oversampling raster patterns, but this has negative implications for speed of acquisition, processing time and the size of data file generated).

One of the challenges in getting this technique deployed operationally would be a fundamental shift in the 'recovery' of fingermark evidence. CSIs typically enhance blood marks at the scene, and photograph these *in situ*, particularly if the marks are visualised on large items such as doors and windows. The blood marks are very rarely lifted. In order to use MALDI MSI, these marks would

need to be physically recovered, either by cutting and removing the substrate, or through the use of gels lifters (which pose an issue when trying to establish a vacuum in the sample chamber). Smaller items, such as knives and guns, drugs packaging and gloves are routinely recovered back to a forensics lab, where they might be subjected to further enhancement techniques in the forensic laboratory, for example cyanoacrylate fuming. However, currently limitations with the mass spectrometric instrumentation dictate that samples reside on a flat, thin surface, enabling it to fit inside a MALDI sample chamber.

Despite some drawbacks, this study has presented initial data towards the development of a molecular 'suspect profile' that could be exploited by law enforcement and that also has potential applications in healthcare. Further research is necessary to investigate some of the further 1300 haemoglobin variants, and evaluate their ionisation and detection via mass spectrometric techniques. Compatibility with the different categories of BETs should also be examined, as well as the effect of time and varying environmental conditions and substrates that might complicate analysis. Finally, further investigation of the link between rare variants and ethnicity as a function of geographical location could be undertaken to provide more robust conclusions about variants provenance and origin.

This work was the subject of an article published in Analyst, entitled 'Detection and mapping of haemoglobin variants in blood fingermarks by MALDI MS for suspect "profiling"' (**Heaton *et al.,* 2021**).

Contributions

Data acquisition on the Bruker MALDI qTOF timsTOF flex and sample preparation using the HTX M3 sprayer was performed by Matthias Witt (Bruker Daltonik GmbH, Bremen, Germany).

References

Baglioni C. (1962) Abnormal human haemoglobins. VII. Chemicalstudies on haemoglobin D. Biochim Biophys Acta. 59(2):437–49

Baglioni, C. & Weatherall, D.J., (1963) Abnormal human hemoglobins IX. Chemistry of hemoglobin J Baltimore, Biochim Biophys Acta. 78, 637-643.

Bandey, H. L., Bleay, S. M., Bowman, V. J., Downham, R. P., Sears, V. G., (2014) Fingermark Visualisation Manual, 1st ed., Centre for Applied Science and Technology (CAST) St. Albans

Bradshaw, R., Bleay, S., Clench, M. R., & Francese, S. (2014). Direct detection of blood in fingermarks by MALDI MS profiling and Imaging. *Science & Justice*, *54* (2), 110–117.

Dass, J., Gupta, A., Mittal, S., Saraf, A., Langer, S., Bhargava, M., (2017) Comparison of the characteristics of two hemoglobin variants, Hb D-Iran and Hb E, eluting in the Hb A2 window, Blood Res. 52, 130-134.

Daniel, Y. and Henthorn, J. (2015), Sickle Cell and Thalassaemia Tandem Mass Spectrometry for Sickle Cell and Thalassaemia Newborn Screening Pilot Study, Screening Programmes, NHS available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/att achment_data/file/1029323/Sickle_cell_and_thalassaemia_screening_tandem_ mass_spectrometry_newborn_screening_pilot_study.pdf [last accessed: 10/07/22]

Daniel, Y. and Henthorn, J. (2017) NHS Sickle Cell and Thalassaemia Screening Program. Handbook for newborn laboratories, NHS & Public Health England *available at*

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/att achment_data/file/585126/NHS_SCT_Handbook_for_Newborn_Laboratories.pd f [last accessed: 28/07/21]

Deininger, L., Patel, E., Clench, M. R., Sears, V., Sammon, C., & Francese, S. (2016). Proteomics goes forensic: Detection and mapping of blood signatures in fingermarks. *Proteomics (Weinheim)*, *16*(11-12), 1707–1717.

Edwards, R. L., Griffiths, P., Bunch, J., & Cooper, H. J. (2012). Top-Down Proteomics and Direct Surface Sampling of Neonatal Dried Blood Spots: Diagnosis of Unknown Hemoglobin Variants. *Journal of the American Society for Mass Spectrometry*, *23*(11), 1921–1930.

Evans, C., (2007) [1998]. The Casebook of Forensic Detection: How Science Solved 100 of the World's Most Baffling Crimes (2nd ed.). New York: Berkeley Books. p. 86–89.

Eyre, J., Personal Communication, 2020

Francese, S., (2019) Criminal profiling through MALDI MS based technologies - breaking barriers towards border free forensic science, Australian Journal of Forensic Sciences, 51, 623-635.

Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A. (2005) Protein Identification and Analysis Tools on the ExPASy Server; (In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press, pp. 571-607

Haynes, C.A., Guerra, S.L., Fontana, J.C., DeJesús, V.R., (2013) HPLC–ESI-MS/MS analysis of hemoglobin peptides in tryptic digests of dried-blood spot

extracts detects HbS, HbC, HbD, HbE, HbO-Arab, and HbG-Philadelphia mutations

Itano. H.A., (1951) A third abnormal hemoglobin associated withhereditary hemolytic anemia. Proc Natl Acad Sci. U.S.A. 37(12):775–84

Kamanna, S., Henry, J., Voelcker, N. H., Linacre, A., & Kirkbride, K. P. (2016). Direct identification of forensic body fluids using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *International Journal of Mass Spectrometry*, *397-398*, 18–26.

Kamanna, S., Henry, J., Voelcker, N. H., Linacre, A., & Paul Kirkbride, K. (2017a). A mass spectrometry-based forensic toolbox for imaging and detecting biological fluid evidence in finger marks and fingernail scrapings. *International Journal of Legal Medicine*, *131*(5), 1413–1422.

Kamanna, S., Henry, J., Voelcker, N., Linacre, A., & Kirkbride, K. P. (2017). "Bottom-up" in situ proteomic differentiation of human and non-human haemoglobins for forensic purposes by matrix-assisted laser desorption/ionization time-of-flight tandem mass spectrometry. *Rapid Communications in Mass Spectrometry*, *31*(22), 1927–1937.

Kennedy, K., Heaton, C., Langenburg, G., Cole, L., Clark, T., Clench, M. R., Sears, V., Sealey, M., McColm, R. & Francese, S., (2020). Pre-validation of a MALDI MS proteomics-based method for the reliable detection of blood and blood provenance. *Scientific Reports*, *10*(1), 17087–17104.

Lauzon, N., & Chaurand, P., (2018). Detection of exogenous substances in latent fingermarks by silver-assisted LDI imaging MS: perspectives in forensic sciences. *Analyst (London)*, *143*(15), 3586–3594.

Modell, B., & Darlison, M., (2008). Global epidemiology of haemoglobin disorders and derived service indicators, *Bulletin of the World Health Organization*, 1-10, *available at* http://www.who.int/bulletin/volumes/86/6/06036673/en/ [last accessed: 28/07/21]

Patel, E., Cicatiello, P., Deininger, L., Clench, M. R., Marino, G., Giardina, P., Langenburg, G., West, A., Marshall, P., Sears, V. & Francese, S., (2016). A proteomic approach for the rapid, multi-informative and reliable identification of blood. *Analyst (London)*, *141*(1), 191–198.

Patrinos, Giardine, B., Riemer, C., Miller, W., Chui, D. H. K., Anagnou, N. P., Wajcman, H., & Hardison, R. C. (2004). Improvements in the HbVar database of human hemoglobin variants and thalassemia mutations for population and sequence variation studies. *Nucleic Acids Research*, *32*(suppl-1), D537–D541.

Piotrowski, E., (1895) Ueber Entstehung, Form, Richtung und Ausbreitung der Blutspuren nach Hiebwunden des Kopfes [*On the formation, form, direction, and spreading of blood stains after blunt trauma to the head*] Vienna, Austria.

Rees, D., (2017) An overview of thalassemias and complications.

Regan, F.A.M., (2017) Blood Cell Antigens and Antibodies: Erythrocytes, Platelets and Neutrophils (Ch 21) in Bain, B.J., Bates, I., Laffan, M.A. Dacie and Lewis Practical Haematology (12th Ed.)

Sharma, D.C., Singhal, S., Woike, P., Rai, S., Yadav, M. and Gaur, R. (2020) Hereditary persistence of fetal haemoglobin, Asian J Transfus Sci., **14** (2) 185–186

*Stewart, V., Deacon, P., Zahra, N., Uchimoto, M.L. and Farrugia, K.J. (2018) The effect of mark enhancement techniques on the presumptive and confirmatory tests for blood, Science and Justice, **58**, 386-396.*

*Strohalm, M., Hassman, M., Košata, B., & Kodíček, M. (2008) mMass Data Miner: An Open Source Alternative for Mass Spectrometric Data Analysis, Rapid Commun Mass Spectrom. **22** 905-908.*

Strohalm, M., Kavan, D., Novakand, P., & Havlicek, V., (2010) mMass 3: A Cross-Platform Software Environment for Precise Analysis of Mass Spectrometric Data, Anal Chem. 82, 4648-4651.

Thein, S.L., (1998) 3 β-Thalassemia, Baillèire 's Clinical Haematology, **11**, (1) 91-126.

Wahed, A., & Risin, S., (2013) Issues with Immunology and Serology Testing (Ch 18), in Accurate Results in the Clinical Laboratory: a guide to error detection and correction. *Dasgupta, A. & Sepulveda, J. L., Elsevier, 295-304.*

Weatherall, D., Akinyanju, O., Fucharoen, S., Olivieri, N., and Musgrove, P., Inherited Disorders of Hemoglobin (Ch 34) *in* Jamison, D.T., Breman, J.G., Measham, A.R., Alleyne, G., Claeson, M., Evans, D. B., Jha, P., Mills, A., Musgrove, P., (2006) Disease Control Priorities in Developing Countries. 2nd ed., Oxford University Press, 663-680.

Weatherall, D. J., & Clegg, J. B. (2001). Inherited haemoglobin disorders : an increasing global health problem. *Bulletin of the World Health Organization*, *79*(8), 704–712.

Weatherall, D. J. (2013). The role of the inherited disorders of hemoglobin, the first "molecular diseases," in the future of human genetics. *Annual Review of Genomics and Human Genetics*, *14*(1), 1–24.

Wild & Bain, Investigation of variant haemoglobins and thalassaemias (Ch 14) 282-311. in Dacie & Lewis, (2017) Practical Haematology, 12th ed., Elsevier, 282-311.

Wilkins, M.R., Lindskog, I., Gasteiger, E., Bairoch, A., Sanchez, J.-C., Hochstrasser, D.F., & Appel, R.D. (1997) *Detailed peptide characterisation using PEPTIDEMASS - a World-Wide Web accessible tool;* Electrophoresis 18(3-4), 403-408.

Witt, M., Kennedy, K., Heaton, C., Langenburg, G., Francese, S., (2021) Forensic visualisation of blood and blood provenance in old fingermarks by MALDI MS Imaging, *Bruker Application Note*

Yergey, J., (1983) A general approach to calculating isotopic distributions for mass spectrometry, Int. J. Mass Spectrom. Ion Phys. 52, 337-349.

https://globin.bx.psu.edu/hbvar/menu.html *[last accessed: 10/07/22]*

https://www.gov.uk/government/publications/handbook-for-sickle-cell-and-thalassaemia-screening/understanding-haemoglobinopathies *[last accessed: 24/06/21]*

https://www.gov.uk/government/publications/sct-screening-handbook-for-newborn-laboratories/newborn-screening *[last accessed: 10/07/22]*

https://www.gov.uk/government/publications/sct-screening-handbook-for-newborn-laboratories/acceptable-analytical-protocols *[last accessed: 10/07/22]*

https://www.nhlbi.nih.gov/health-topics/thalassemias *[last accessed: 17/06/21]*

Chapter 5: Conclusions

This body of work has presented three novel applications of MALDI MS and MALDI MSI for the investigation of biological molecules. Whilst the application of MALDI MS and MALDI MSI spans over 4 and 2 decades respectively, it is only in the last decade or so that the technique became interesting in a forensic science application context. However its use, reported in 2009 by Wolstenhome *et al.* was restricted to an academic environment until 2017 when Bradshaw *et al.* first demonstrated its application to operational casework. The slow uptake is not surprising as the transition in an operational environment requires, amongst other things, funding and constant end user input and comprehensive validation studies. The techniques have certainly been propelled by drastic improvements in acquisition time, mass accuracy and mass resolution. Instrument sensitivity has also increased, which has allowed laser spot sizes to reduce (as a sufficient ion yield can be obtained from less surface material), enabling spatial resolution to improve for MSI analysis. Additionally, recent configuration improvements, such as the development of the pioneering MALDI-2 source, have enabled further increases in ion yields, resulting in sensitivity 2-3 orders of magnitude greater than conventional MALDI configurations (**Henkel *et al.*, 2020**). These technological improvements have unlocked the potential for further applications than were possible previously and an acceleration of the operational transition is expected. In imaging mode, spatial resolution of the most state-of-the-art instruments is now at a sub-cellular level, and hence instrumentation is more than capable of reconstructing fingermark ridge detail, allowing for the visualisation of level 2 details (*minutiae*) and with particularly high resolution analysis, level 3 details (pores) can also be observed, as reported by Ferguson *et al.* (**Fig. 57**) (**Ferguson *et al.*, 2013**).
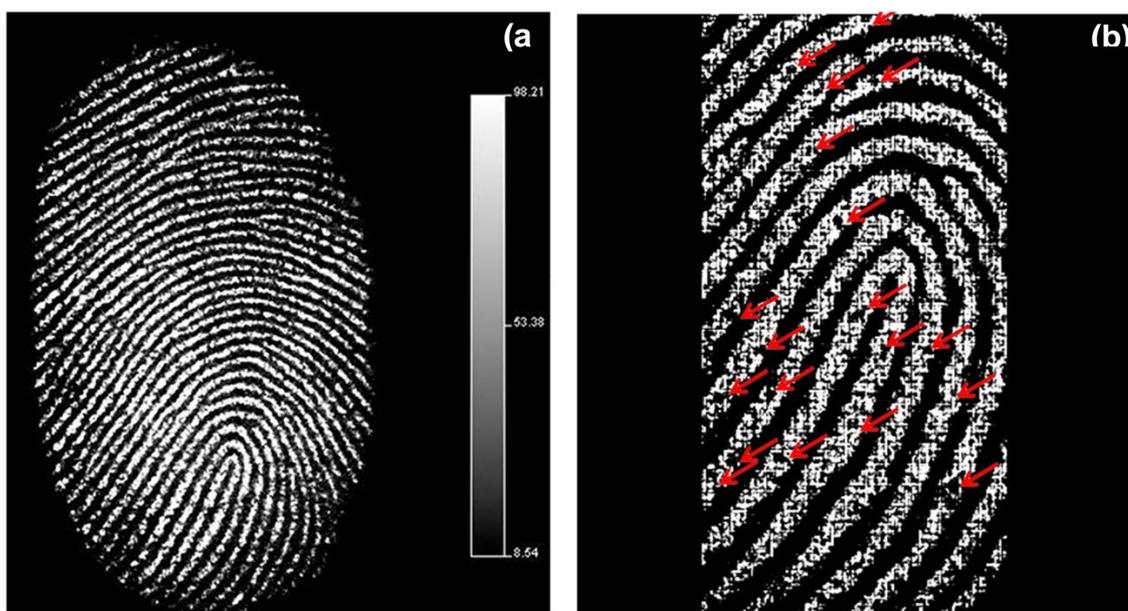
Figure 57. Molecular MALDI image of a signal at *m/z* 283.3 attributed to oleic acid in a fingermark. The first panel exhibits good clarity of level 2 fingermark detail of the *minutiae* (1a). The second panel depicts a zoomed in image of level 3 fingermark detail, with pores marked by arrows, present within the ridges of the mark (1b). *[Adapted from Ferguson et al., 2013].*

The hyphenation of chromatography with mass spectrometry is utilised routinely for operational forensic casework, and liquid chromatography (LC) mass spectrometry (or LC-MS/MS) is still the preferred analytical approach in most forensic laboratories, from toxicological screening to flammable accelerants. However, usually the sample preparation, separation step and data acquisition are time consuming processes. Whilst the generation of multi-charged ions coupled with prior separation permits the detection of a greater number of ions and better MS/MS capabilities, MALDI based methods are very rapid and the spectra are straightforward to interpret. Both features sit well in a forensic analysis context. In addition, chromatography-based techniques lack the spatial dimension that MALDI MS imaging can provide, which can be particularly useful for linking biometric information such as the identifying fingermark ridge pattern with the molecular information that has generated the image.

Conventional fingermark enhancement techniques (FET) are used by CSI and crime labs to visualise fingermarks on different surfaces. They generally target one class of biomolecules that are normally present in fingermarks such as lipids, amino acids, proteins or electrolytes. These methods would not permit fingermark visualisation by targeting specific contaminants that may be present.

Conversely, the versatility of MALDI MS and MSI in the detection of a vast range of molecules in a wide mass range permits for wider opportunities to visualise (multiple) images of a fingermark, thus maximising ridge detail (**Bradshaw *et al.,* 2021**). The multiplex nature of MALDI MSI allows for the stitching of partial mark images from the signals originating from multiple detected compounds to improve ridge continuity, a clear advantage over a single image provided by most FET (**Francese *et al.,* 2013**), or be used to resolve overlapping marks that may be comprised of different substances (**Bradshaw *et al.,* 2012**). It is not essential to identify these compounds if the goal is the enhanced visualisation of the ridge detail, though MALDI MS/MS (and/or IMS) can be applied if it is anticipated that identification of the compounds might be useful for the provision of forensic intelligence around the suspect or the circumstances of the crime (**Francese *et al.,* 2017**). Altogether, these capabilities increase the likelihood to find a match with a fingerprint record in the National Database and contribute to informing and steering investigations. Furthermore extensive compatibility of MALDI based methods with current FETs have been demonstrated over the past 15 years making this method certainly operational within the current investigative workflow.

The work reported in these chapters has demonstrated the capability of MALDI MS instrumentation combined with a proof-of-concept proteomic approach for three forensic applications namely the determination of sex from fingermarks,

species identification from human and animal blood and a novel method for the detection of haemoglobin variants in human blood. Whilst these concepts are in relatively early stages, it is hoped that the published work generates some level of engagement with end users.

Discussion on the state-of-the-art operational significance of the methods developed as well as additional MALDI MS profiling and imaging research to further drive integration into operational deployment, and enhance the molecular 'suspect profile' available to investigators, is presented in the context of each chapter.

Chapter 2 reported an investigation into the determination of sex through fingermark analysis. Following from the proof of concept study by Ferguson *et al.,* in 2012, the work presented in this chapter confirmed the ability to determine the sex of an individual through the detection of peptide and protein profiles via MALDI MS in combination with comprehensive statistical modelling The increase in sample size (a cohort of 172 versus 80 in *Ferguson et al., 2012*) increases the validity of the results whilst the use of natural marks (versus ungroomed in *Ferguson et al., 2012*) with and without the prior application of the white enhancement powder employed increases the significance.

The accuracy of the predicted sex of an individual was 86.1%, using the XGBOOST classifier. This appears to represent only a mild improvement to the 85% reported by Ferguson *et al.* (**Ferguson *et al.,* 2012**). However, in this previous work, the final prediction was based on the sex predicted for the majority of 9 technical replicates. Some of the replicates for each mark did not match the 'general consensus' prediction. Therefore if each replicates prediction

was taken at face value, the accuracy would fall to 68.9% for female and 74.4% for male spectra in the Ferguson *et al., 2012* study. The 86.1% prediction accuracy reported in Chapter 2 was obtained using 'full consensus scoring' i.e. the prediction is only considered correct if 3 out of 3 replicates are in agreement. This was achieved using a recent machine learning algorithm, the XGBOOST classification model, offering more comprehensive decision making in the classification process. Whilst the application of this scoring system inevitably leads to a reduction in the number of classified marks (no classification for nearly 50% of the samples), it does enable correct classification for the remaining samples with an 86.1% of accuracy of prediction. In a forensic context, whilst this predictive power is insufficient to eliminate a suspect from the investigation, the Police collaborators co-authoring the study deem the method suitable for triaging crime scenes presenting several marks.

The observations reported in this study regarding sample storage, time since deposition and ion yield should be used to inform future studies. Samples in the initial preliminary study were analysed over a four month period, whereas all of the samples for the present study reported here, were acquired and analysed within one month of deposition. This, along with advances in instrumentation, contributed to the detection of greater numbers of peptides with signals of higher intensity. In future work, fingermark deposits could be collected in multiple small batches, so that it was possible to analyse the whole batch in an even shorter time since deposition. This would mitigate prolonged sample storage, potentially minimising sample degradation, thus also potentially resulting in a higher signal population with signals of even higher intensity. However, the instrumental conditions (general performance and environmental conditions) could still be a negative variable.

Acquiring data from the 'freshest' marks would be useful for the validation of the method to try to improve the prediction accuracy and identify the most prominent discrepancies between the fingermarks from opposite sexes, under laboratory conditions. While every attempt was made to minimise time between fingermark deposition and analysis, access to the laboratory and mass spectrometer used in this study (in Maastricht) meant this was the shortest time feasible. Future work could involve analysis of marks in an even shorter time frame, if the researchers have instrumentation in-house.

The effect of sample storage and time period (and therefore potential degradation) between the studies could not be accurately investigated within the scope of this research, as a) different donors were used, and b) donors were anonymous as part of the study design. To directly compare the effect of degradation, the same identified donors would be required for consistency. This could also be the focus of a future study.

Instrumentation advances would clearly help in increasing the peptide/protein yield (sensitivity), and mass resolution and mass accuracy as it is already known that the specifications of the RapifleX (the MALDI-TOF/TOF used in this work) have been surpassed by more state-of-the-art instrumentation.

As mentioned above, white powder was employed on a subset of collected marks to preliminarily enhance them prior to the matrix application. Powders are the most common FET used in the UK. Three common fingermark powders were assessed for their compatibility with this method, and the powder that caused the least analyte suppression with MALDI was selected. Although the other two powers exhibited some level of analyte suppression, peptide and protein signals were still present, which demonstrates promise for the

compatibility of the prior application of these powders at a crime scene. Future work could include an assessment of the compatibility with further enhancement powders, and with further FETs such as dyes as well as the analysis of tape lifts after enhancement.

All marks were deposited on Bruker conductive IntelliSlides for compatibility with the MALDI instrumentation. Future work could involve investigating additional deposition surfaces, as Bruker RapifleX instruments require conductive slides, but other MALDI MS instrumentation do not. This could involve an assessment of additional surface types, encompassing both smooth and textured and porous and non-porous surface types. Laser focus might present an issue on undulating surfaces, and components such as amino acids may be absorbed into porous surfaces such as paper over time. Already in 2011, Ferguson *et al.* demonstrated that latent marks could be enhanced with MALDI matrix and lifted before MSI analysis (**Ferguson *et al.,* 2011**), and Bradshaw *et al.* have already evaluated the feasibility of MALDI MS and MSI to be used on lifted marks post-enhancement (**Bradshaw *et al.,* 2017**). Whilst this method of sample acquisition would be desirable, a validation study would have to be carried out to assess feasibility for this application, where peptides and proteins require successful lifting and detection.

The donors were from a fairly homogenous sample population, comprising of University staff and students and Police volunteers. While no exclusion criteria were applied, the participants lacked diversity. Some 88 % of the participants identified as 'White British', with 13 other self-reported ethnicities making up the remaining 12% (**Fig. 58**). Furthermore, nearly a quarter of the donors were in the 22-29 years age range, with over half of the participants within the age range of 20-39 years (**Fig. 59**). To make the model more robust for operational

deployment, training data should be acquired from an equally diverse sample population, balance for sex, across a wider age range and a more ethnically diverse sample population.
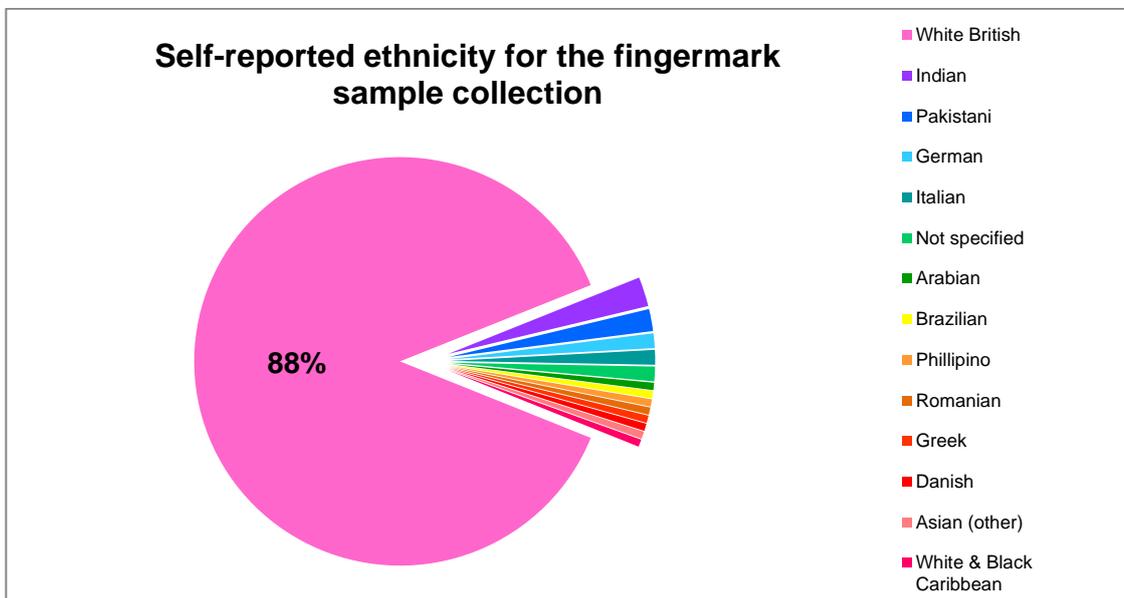

**Figure 58.** Ethnic distribution of fingermark donors.
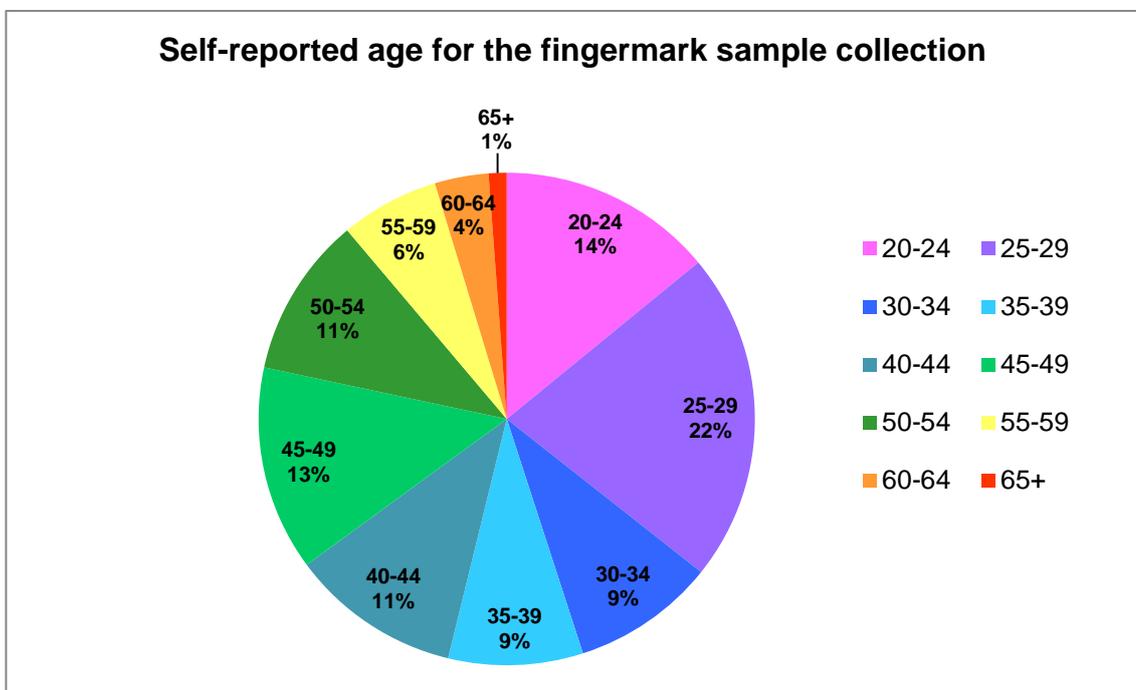
Figure 59. Age range distribution of fingermark donors (years).

Finally, additional donor information was recorded during sampling to account for any intra-sample variability that may have been observed, namely: smoker

status, which toiletry products which may have come into contact with the hands previously on the day of the collection, any medication the donor may be using and any known illnesses. While no discernable differences were observed between these groups, the populations of each were not of sufficient size to enable statistically significant conclusions to be drawn. However, future work could further investigate the impact on sex differentiation for marks of smokers versus non-smokers, those that used toiletries versus those that did not (or those that had used different categories of toiletry products, by looking at different polymers for example) and those that exhibited illnesses that might affect an individual's protein/peptide expression in sweat.

Chapter 3 reported a validation study assessing the capabilities of MALDI MS and MSI combined with bottom-up proteomics for the detection, identification and provenance of blood. Samples comprising of human blood, animal blood (bovine, porcine (pig and wild boar) and chicken), other human biofluids (saliva, semen, sweat) and other non-biofluid interferents (e.g. egg, ketchup and paint) were analysed in a blind fashion. Samples were provided as both stains and contaminated fingermarks. A subset of these substances were enhanced using Acid Black 1, Acid Yellow 7 or Leucocrystal Violet as blood enhancement techniques (BET). The method was developed on 56 samples and validated on 13 samples, demonstrating feasibility regardless of the enhancement technique employed and in some cases refuting the false positive indication given by these presumptive tests. By the time these samples were analysed, they were three-and-a-half years old, demonstrating the robustness of the model on aged samples. A key and unexpected finding was the ability to distinguish between the blind blood samples initially commissioned for the study and blood obtained

locally from butchers via intravenous sampling through the detection of the skeletal muscle Glyceraldehyde 3-phosphate dehydrogenase (GAPDH).

Incidentally, during method development, identification and differentiation of semen was possible through the detection and identification of semen-specific markers which are also used by current forensic testing. Three semenogelin peptide markers were detected in the study and subsequently proposed as biomarkers for the presence of this biofluid. This occurrence highlights the ability of MALDI to detect fluid-specific biomarkers, and further research could be applied to additional biofluids outside of the scope of the study, for example sweat, urine or vaginal fluid, which could all provide evidential value in investigations. This demonstrates the capability of MALDI MS to operate as a multiplexed approach for the initial screening of suspected biofluid samples in a forensic context.

Specifically, in the final validation batch of 13 samples, human blood was correctly identified as blood and of human origin in both instances (2/2). Animal blood was correctly identified in all instances (6/6), with species identification correct in all cases except in one instance where a conclusive statement on the species could not be made. Non-blood samples were correctly assigned as non-blood in all instances (5/5), which were further narrowed down, with semen successfully identified in two out of the three semen samples (2/3), with the final semen sample, whilst correctly identified as non-blood, was inconclusive for exact biofluid identification. The final proportion correctly identified was 92%.

These preliminary results set the foundation for further investigation into the application of MALDI MS to discriminate biofluid samples, as discussed in the following section.

The animal species investigated within this study, in addition to humans, were limited to the most commonly consumed animals in the UK and US, and therefore included avian, bovine and porcine (including wild boar) blood. However, future work could expand the number of species investigated and assess more widely the proteotypicity of the blood biomarkers discovered. Indeed, during the course of the study, Kirkbride's group in Australia were investigating the differences in the amino acid sequence between human haemoglobin and that of some Australian native mammals, also using MALDI MS and a bottom-up proteomics approach for forensic purposes (**Kamanna *et al.,* 2017**). While these mammalian species would be particularly uncommon in the UK, including other species regularly consumed in the UK such as sheep, duck, turkey etc. would be relevant. Also relevant might be the inclusion of domesticated species that can commonly be found within the home such as cats, dogs rabbits etc.

A rapid method for distinguishing human from animal blood could have applications in violent crime scene investigation, as well as for wildlife crimes. Proteomics-based methods are a viable alternative to expensive and time consuming DNA profiling in specialist DNA laboratories, and do not use up their finite resources.

In this study, the enhancement techniques employed to enhance a subset of the blind samples to probe compatibility of the proposed MALDI MS-based method were Acid Black 1, Acid Yellow 7 and Leucocrystal Violet. Acid Black 1 and Acid Yellow 7 are two of the most common protein dyes, and Leucocrystal Violet is a common haem-reactive reagent. All three are commonly used both in the UK and the US for presumptive blood visualisation (**Bleay *et al.,* 2017**). BETs broadly fall into three categories; protein stains (**Farrugia *et al.,* 2011a**),

peroxidase reagents (**Farrugia *et al.,* 2011b**) and amino acid stains (**Farrugia *et al.,* 2013**) and they are applied at the scene of violent crimes without particular concern for potential subsequent forensic tests. To consider MALDI-based analyses, end users would require validation studies that demonstrate compatibility of MALDI with all currently used blood enhancement techniques. Whilst a milestone, the present study should be followed by a larger validation study to evaluate the prior enhancement of samples with all currently used enhancement reagents from the three categories, to give end users the necessary reassurance to use MALDI MS both as a confirmatory test as the tool for further intelligence such as blood provenance.

In addition, all of the fingermark and stain samples were deposited on a single substrate, aluminium slides. Using the same surface eliminated one variable from the assessment of the effectiveness of the technique. However, a larger validation study should also include an assessment of various surfaces of deposition (both porous and non-porous substrates) to assess any surface-dependant ion suppression, particularly when blood is analysed *in situ*. Furthermore, all samples were prepared in a laboratory environment and not exposed to varying environmental conditions. The effect of different temperatures, climates, light/dark cycles and time since deposition should be investigated to assess the rate of degradation of samples across these variables, to account for variables that would be encountered in the field.

The development of this method was a manual and iterative process, with experimental findings feeding back into the optimisation of the method. To be a more viable option for operational deployment, this process would require automation. For an optimised workflow, statistical processing would be available natively within the instrument software. A statistical approach using PCA could

be employed to categories different sample types against a known dataset, in a tiered approach similar to that described in this study. The first tier could determine if a sample was blood or not based on whether it fell within pre-determine confidence ellipses on the plot, with subsequent tiers determining whether the blood was of human or animal origin, and attempting species identification. Other tiers could indicate if a sample was not blood, whether it was another human (non-blood) biofluid or a non-biofluid, using the same approach. Reference databases could be updated with blood-derived peptides from local species depending on the location. Ideally, the reference peptides should be confirmed experimentally, as not all theoretical peptides are detected due to variation in ionisation efficiency. Theoretical peptide lists would need to be evaluated to identify proteotypic peptides, and those shared between additional species not investigated within the scope of this study. Furthermore, the differences between different human biofluids should be even more pronounced than between the blood of different species, which in some cases can share identical peptides from the same protein, for example the prominent m/z 1275 haemoglobin peptide, present in human and bovine blood. In 2016 The Kirkbride group reported the direct MALDI MS identification of blood, saliva, semen and urine using between 7 and 22 peptides per protein (**Kamanna *et al.,* 2016**). Proteotypic peptides such as those reported could be added to the model for rapid discrimination.

Future work could be carried out to determine specific markers for the remaining human biofluids that might be relevant in a forensic context; sweat, saliva, mucus, urine, vaginal fluid and menstrual blood. Markers are reported in the literature for these biofluids (**Illiano *et al.,* 2018**), but mainly by utilising an LC-based method, and it would be interesting to see if these could be detected

through MALDI MS, particularly by MSI to precisely locate particular biofluids within a sample area, such as within a contaminated mark, as this could support/refute an eyewitness account.

Finally, all analyses described in this chapter were carried out in a laboratory on large and cumbersome MS instrumentation that currently requires mains power and a fixed gas supply. The sample chamber also requires low vacuum necessitating large roughing pumps. Consequently, this analysis can only currently be carried out in a laboratory environment and not *in situ* at a crime scene. To be of maximum use to end users, portable instrumentation is desirable to triage suspected biofluid deposits at the crime scene, something not possible with current technology. Over the next decade, further work should consider the miniaturisation and portability of devices to perform rapid, robust identification at the point of evidence recovery to deliver the largest benefit to investigators.

This chapter describes an investigation into the detection and identification of human haemoglobin variants using a MALDI MS and MSI approach, to narrow down the pool of suspects. This evidence gains more incriminatory value the lower the incidence of the variant in a given population. The World Health Organisation reports that haemoglobin gene variants are present in between 1 and 1.5 out of every 1000 people in "all sizeable populations". However, in regions where malaria is prevalent, those variants that offer some protection against sickle cell anaemia (HbS, HbC, HbE, Hb D-Punjab) may account for up to 40% of the population (**Modell & Darlison, 2008**). Conversely, some rare variants such as Hb Belliard have only a handful of documented case reports

(**Murthy & Benavides, 2017**). The local incidences would need to be taken into account before determining the evidential value of a detected variant.

The six most prevalent variants (in the UK) were identified through detection of their proteotypic peptides via MALDI MS and MSI following *in solution* and *in situ* digestion. An *in silico* method was employed to assess specificity of the variant peptides (proteotypic so were specific to the variant haemoglobin protein and species-specific, so only found in humans). This was performed by screening the variant peptides against the UniProt protein database, against over 500,000 manually annotated and review proteins from over 14,000 species (www.uniprot.org). Using the Peptide Search function 8/19 variant peptides yielded a match against human haemoglobin, with the remaining 11 yielding no match. (One of the 8 human haemoglobin matches, HbC (VHLTPE, [2-7]) however also yielded a match to proteins found in Fungi, metazoa, bacteria and human).

To the author's knowledge, this is the first time haemoglobin variants have been investigated in a forensic context for identification purposes. This knowledge could be used by crime scene investigators to rule in/out potential individuals of interest when blood is encountered if; a) they have pathological haemoglobin variant and b) their medical information was known to investigators. This would be particularly significant when identification through DNA profiling may not be possible, for the reasons discussed in Chapter 4.

The amino acid sequences of these proteotypic peptides were verified through MALDI MS/MS and *De Novo* sequencing. Furthermore, MALDI MSI was employed to map the molecular signals from the proteotypic peptides within bloody fingermarks. At least one proteotypic peptide was successfully detected

and a molecular image generated for each variant. It was demonstrated that two bloodstains, containing differing variants, could be resolved using molecular imaging, an application not possible with current conventional blood enhancement techniques. In addition, separation of overlapping bloody fingermarks, contaminated with different variants, was attempted following enhancement with Acid Black 1. Although not successful in completely resolving the ridge detail of each mark (likely due to fresh blood mixing), the technique was able to establish the presence and location of the two variants following enhancement, demonstrating the ability of MALDI MSI to reveal information not available when using BETs alone. Furthermore, this data would not be available when undertaking the 'gold standard' of forensic identification techniques, GC and LC-MS based analysis, which does not offer 2D spatial information. The successfully acquisition of this data post-enhancement demonstrates that this technique could integrate into current forensic analysis workflows.

Finally, although healthcare providers have established screening programs and methods for haemoglobin identification, this MALDI-based protocol proposed in this chapter has potential clinical diagnostic applications, as it could offer high-throughput screening to relieve the analytical burden of the time consuming methods currently employed in hospitals in the UK.

Whilst the research reported in Chapter 4 focussed on the 6 most common haemoglobin variants in the UK, but on the HbVar (Haemoglobin Variant) database (**https://globin.bx.psu.edu/cgi-bin/hbvar/counter**), there are currently 1856 total entries, 1417 of which are haemoglobin variants and 536 of which are thalassemias, (69 of which fall in both categories) listed. Weatherall & Clegg reported in 2001 that only three structural variants occur in high frequency (**Weatherall & Clegg, 2001**). Future work could involve a larger

study to extend this preliminary research to incorporate additional variants. Consultation with a local haematology department could reveal those variants which had been detected in patients and these could be targeted in a more comprehensive study. This would require calculating *in silico* the theoretical peptides expected from these variants, and determining if these variants had proteotypic peptides unique to each one to facilitate identification.

To streamline this process, identification could be automated, in an approach similar to that proposed for the automated discrimination of biofluids. A large database could be created with the theoretical peptides from variants present in a given region, and statistical processing such as PCA applied to determine unknown samples against a known dataset. In an ideal scenario, this would identify the haemoglobin variant within a blood sample, if proteotypic peptides were detected, or give a shortlist of potential variants if peptides common to multiple variants were detected.

In an operational context, this could help to narrow down potential persons of interest involved in a violent crime that had resulted in bloodshed. Although haemoglobin variants are currently detected within a clinical context for medical diagnosis, this is usually carried out through IEF, HPLC or CE and would not afford investigators the molecular information in combination with the spatial information that can be gained through MALDI MSI, which could prove critical to an investigation.

This study demonstrated the compatibility of a MALDI MS – 'bottom-up' proteomics based approach for the detection of peptides in samples that had undergone prior enhancement with Acid Black 1, a commonly used BET. Samples were also all deposited on the same substrate to remove that variable

within the scope of this preliminary work. However, further investigation on the compatibility with additional BETs and surfaces of deposition would be required to ensure the method yielded good results with all of the enhancement categories and surface types.

Collaborators from the Haematology Department at the Royal Hallamshire Hospital in Sheffield (UK), co-authoring the research, observed that patients with certain haemoglobin variants came from specific geographic areas. They reported that the overwhelming majority of their patients exhibiting the Hb J-Baltimore variant came from South Asian communities in certain areas of Leeds, and virtually all of their patients displaying the Hb D-Iran variant came from South Asian communities in certain areas of Rotherham. It was hypothesised that this was the result of immigration of people from South Asian nations (predominantly of Pakistani heritage) preferentially settling in close-knit areas to be near to family and/or people from their native community. Variants are usually named after the location they were first discovered in. For example, the first abnormal haemoglobin variant reported in an English family was Hb Norfolk in 1958, so called because the family lived in Norfolk (**Ager *et al.,* 1958**). Hb J-Baltimore was first reported by Baglioni in 1963 (**Baglioni & Weatherall, 1963**), and the group of Hb D haemaglobinopathies was first reported by Itano in 1951 (**Itano, 1951**), with Rahbar (**Rahbar, 1973**) independently designating the β22 Glutamic Acid (E) → Gutamine (Q) substitution as D-Iran in 1973 (**Chandel *et al.,* 2015**). However, due to global migration and the fact that they are most often inherited characteristics these Hb variants may change in incidence. Whilst they are no longer a definitive marker of where an individual may live or of their ethnicity, their detection may

provide a certain level of intelligence that is worth pursue during an investigation.

These mutations can persist within the population if they are advantageous to the populace, such as with the example of sickle cell anaemia providing protection against malaria. However, despite recent globalisation, it is hypothesised that the detection of haemoglobin variants may be able to rule in/out a suspect or victim as the source of a bloodstain. Detection of a rare haemoglobin variant within a bloodstain at a crime scene may be able to indicate the likelihood of an individual's ethnic heritage, based on the prevalence of the variant globally. For example, incidence of the sickle cell gene is concentrated in sub-Saharan Africa (and also present in Saudia Arabia and India), in areas most prone to malaria, where carrier frequencies for HbS can reach >40% of the population (**Weatherall & Clegg, 2001**). Given the observation by colleagues in the Haematology Department, this approach could be used at a more local level, for example, narrowing a suspect pool to a more localised geography, to a certain level of confidence. However, this observation is based on anecdotal evidence, and a more rigorous investigation would be required to draw any meaningful statistical conclusions based on sample size and variant incidence.

Finally, it is important to remember that some of the variants have clinical implications, for example HbS, or sickle cell anaemia, which can, in extreme cases, prove fatal. As such, discussion is needed around the ethical implications of using forensically derived information to potentially identify individuals but that also carry medical information and whether they should be disclosed to the individual. This challenge extends to the analysis of any medically relevant endogenous substances detected from biological material

such as biofluids or fingermarks. Recent research carried out by Francese's group investigated the potential non-invasive detection of breast cancer markers through fingertip smears (**Russo *et al.,* 2022**). If investigators were to incidentally reveal markers for a specific disease when investigating the molecular composition of evidential samples, is it ethical to divulge this information to the individuals? Would judicial cases require Police to share how they potentially identified an individual? While an answer is not immediately obvious to these ethical discussions, implications of biochemical analysis should be carefully considered before operational deployment. The mass spectrometric techniques discussed are already used in medical settings as they are clearly effective in many clinical applications. The line between analyses of biological samples for clinical or forensic applications is blurred and should be cautiously considered when the potential for revealing intimate personal information increasingly becomes a reality.

References

Baglioni, C., & Weatherall, D.J. (1963). ABNORMAL HUMAN HEMOGLOBINS. IX. CHEMISTRY OF HEMOGLOBIN J-BALTIMORE. Biochimica et biophysica acta, 78, 637-43.

Bradshaw, R., Wolstenholme, R., Blackledge, R. D., Clench, M. R., Ferguson, L. S., & Francese, S. (2011). A novel matrix-assisted laser desorption/ionisation mass spectrometry imaging based methodology for the identification of sexual assault suspects. *Rapid Communications in Mass Spectrometry*, *25*(3), 415–422.

Bradshaw, R., Rao, W., Wolstenholme, R., Clench, M. ., Bleay, S., & Francese, S. (2012). Separation of overlapping fingermarks by Matrix Assisted Laser Desorption Ionisation Mass Spectrometry Imaging. *Forensic Science International*, *222*(1), 318–326.

Bradshaw, R., Denison, N., & Francese, S. (2017). Implementation of MALDI MS profiling and imaging methods for the analysis of real crime scene fingermarks. *Analyst (London)*, *142*(9), 1581–159.

Bradshaw, R., Wilson, G., Denison, N., & Francese, S. (2021). Application of MALDI MS imaging after sequential processing of latent fingermarks. *Forensic Science International*, *319*, 110643–110643.

Bleay, S., Sears, V., Downham, R., Bandey, H., Gibson, A., Bowman, V., Fitzgerald, L., Ciuksza, T., Ramadani, J., Selway, C., (2017). Fingerprint Source Book v2.0 (second edition), *CAST Publication 081/17*. 1-666.

Caprioli, R. M., Farmer, T. B., & Gile, J. (1997). Molecular Imaging of Biological Samples: Localization of Peptides and Proteins Using MALDI-TOF MS. *Analytical Chemistry (Washington)*, *69*(23), 4751–4760.

Chandel, R. S., Roy, A. & Abichandani, L. G., (2015). *Serendipity: A Rare Discovery of Haemoglobin D-Iran in An Indian Female During Routine Antenatal Screening for β-Thalassemia*, *J Clin of Diagn Res. 9*(7), BD01-BD02.

Farrugia, K. J., Savage, K. A., Bandey, H., & Nic Daéid, N. (2011a). Chemical enhancement of footwear impressions in blood on fabric – Part 1: Protein stains. *Science & Justice*, *51*(3), 99–109.

Farrugia, K. J., Savage, K. A., Bandey, H., Ciuksza, T. & Nic Daéid, N. (2011b). Chemical enhancement of footwear impressions in blood on fabric — Part 2: Peroxidase reagents. *Science & Justice*, *51*(3), 110–121.

Farrugia, Bandey, H., Savage, K., & NicDaéid, N. (2012). Chemical enhancement of footwear impressions in blood on fabric — Part 3: Amino acid staining. *Science & Justice*, *53*(1), 8–13.

Ferguson, L., Bradshaw, R., Wolstenholme, R., Clench, M., & Francese, S. (2011). Two-Step Matrix Application for the Enhancement and Imaging of Latent Fingermarks. *Analytical Chemistry (Washington)*, *83*(14), 5585–5591.

Ferguson, L. S., Wulfert, F., Wolstenholme, R., Fonville, J. M., Clench, M. R., Carolan, V. A., & Francese, S. (2012). Direct detection of peptides and small proteins in fingermarks and determination of sex by MALDI mass spectrometry profiling. *Analyst (London)*, *137*(2), 4686–4692.

Ferguson, L. S., Creasey, S., Wolstenholme, R., Clench, M. R., & Francese, S. (2013). Efficiency of the dry-wet method for the MALDI-MSI analysis of latent fingermarks. *Journal of Mass Spectrometry.*, *48*(6), 677–684.

Francese, S., Bradshaw, R., Ferguson, L. S., Wolstenholme, R., Clench, M. R., & Bleay, S. (2013). Beyond the ridge pattern: multi-informative analysis of latent fingermarks by MALDI mass spectrometry. *Analyst (London)*, *138*(15), 4215–4228.

Francese, S., Bradshaw, R., & Denison, N. (2017). An update on MALDI mass spectrometry based technology for the analysis of fingermarks - stepping into operational deployment. *Analyst (London)*, *142*(14), 2518–2546.

Henkel, C., Koch, A., Becker, M., Orth, T., Kobarg, J. H., Cornett, S., Silvescu, C., Barsch, A. & Easterling, M. (2020) The combination of MALDI-2 and timsTOF flex brings targeted drug imaging to the next level, *Bruker Daltonik GmbH*, Application Note.

Holmna, C. A., Smith, A. J., Whimster, W. F., Beale, D., & Lehmann, H. (1964). Haemoglobin J Baltimore in a Kent Family. *BMJ*, *2*(5414), 921–922.

Itano H. A., (1951) A third abnormal haemoglobin associated with hereditary haemolytic anemia. Proc Natl Acad Sci U S A. 37(12):775–84.

Kamanna, S., Henry, J., Voelcker, N. H., Linacre, A., & Kirkbride, K. P. (2016). Direct identification of forensic body fluids using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *International Journal of Mass Spectrometry*, *397-398*, 18–26.

Modell, B., & Darlison, M. (2008). Global epidemiology of haemoglobin disorders and derived service indicators. *Bulletin of the World Health Organization*, *86*(6), 480–487.

Murthy, S., & Benavides, R. (2017). A Rare Hemoglobin Variant, Hb Belliard. *Proceedings - Baylor University. Medical Center*, *30*(2), 184–185.

Rahbar, S., (1973) Haemoglobin D Iran: 2 22 glutamic acid leads to glutamine (B4).[2] Br J Haematol. 24(1):31–35.

Russo, C., Wyld, L., Da Costa Abreu, M., Bury, C., Heaton, C., Cole, L. M. & Francese, S. (2022) Non-invasive screening of breast cancer from fingertip smears - a proof of concept study, *Scientific Reports*, *[pre-print]*

https://globin.bx.psu.edu/cgi-bin/hbvar/counter [last accessed: 05/06/22]

https://web.expasy.org/peptide_cutter [last accessed: 05/06/22]

https://www.uniprot.org/ [last accessed: 05/06/22]

Publications List

<u>Heaton, C.,</u> Bury, C. S., Patel, E., Bradshaw, R., Wulfert, F., Heeren, R. M. A., Cole, L. M., Marchant, L., Denison, N., McColm, R., Francese, S., (2020). Investigating sex determination through MALDI MS analysis of peptides and proteins in natural fingermarks through comprehensive statistical modelling, *Forensic Chemistry* 20, 100271, https://doi.org/10.1016/j.forc.2020.100271

Kennedy, K., <u>Heaton, C.,</u> Langenburg, G., Cole, L. M., Clark, T., Clench, M. R., Sears, V., Sealey, M., McColm, R., Francese, S., (2020). Pre-validation of a MALDI MS proteomics-based method for the reliable detection of blood and blood provenance, *Scientific Reports,* 10, 17087. https://doi.org/10.1038/s41598-020-74253-z

Russo, C., <u>Heaton, C.,</u> Flint, L., Voloaca, O., Haywood-Small, S., Clench, M. R., Francese, S., Cole, L. M., (2020). Emerging applications in mass spectrometry imaging; enablers and roadblocks, *Journal of Spectral Imaging* 9, a13. https://doi.org/10.1255/jsi.2020.a13

Kennedy, K., Bengiat, R., <u>Heaton, C.,</u> Herman, Y., Oz, C., Levin-Elad, M., Cole, L. M., Francese, S., (2021). "MALDI-CSI": A proposed method for the tandem detection of human blood and DNA typing from enhanced fingermarks by MALDI MSP and MSI, *Forensic Science International*, *323*, 110774–110774. https://doi.org/10.1016/j.forsciint.2021.110774

Witt, M., Kennedy, K., <u>Heaton, C.,</u>, Langenburg, G., Francese, S., (2021). Forensic visualisation of blood and blood provenance in old fingermarks by MALDI MS Imaging, Bruker Application Note,

file:///C:/Users/Owner/Downloads/asms-2021-fp-216-maldi-imaging-of%20blood-and-blood-provenance.pdf

Heaton, C., Witt, M., Cole, L. M., Eyre, J., Tazzyman, S., McColm, R., Francese, S., (2021). Detection and mapping of haemoglobin variants in blood fingermarks by MALDI MS for suspect "profiling." *Analyst*, 146(13), 429–432. https://doi.org/10.1039/d1an00578b

Bury, C., Heaton, C., Cole, L. M., McColm, R., Francese, S., (2022). Exploring the problem of determining human age from fingermarks using MALDI MS-Machine learning combined approaches. *Analytical Methods*, *14*(8), 789–797. https://doi.org/10.1039/d1ay02002a

Fischer, T., …Heaton, C., ... Francese, S., (2022). Profiling and imaging of forensic evidence - a pan-European forensic Round Robin Study Part 1: Document forgery, *Forensic Science International*, 62(4), 433-447. https://doi.org/10.1016/j.scijus.2022.06.001

Russo, C., Wyld, L., Da Costa Abreu, M., Bury, C., Heaton, C., Cole, L. M., Francese, S., (2022). Non-invasive screening of breast cancer from fingertip smears – a proof of concept study, *Scientific Reports*, [pre-print]

Appendices

Appendix 1

1(i) Participant Information Sheet template

Sheffield Hallam University | Biomolecular Sciences Research Centre

[dstl] Defence, Science & Technology Laboratory

Participant Information Sheet

Title of project: *Human identification though advanced forensic mass spectrometry of fingermarks*

The participant should read this sheet in its entirety

You are being invited to take part in this research project. Before you decide to do so, it is important you understand why the research is being done and what it will involve. Please read the following information carefully and ask for any clarification or if you would like more information. Take time to decide whether or not you wish to take part.

It is your choice whether you take part or not. Non-participation will not affect you in any way. You are also able to withdraw from this project once it has commenced without affecting you in any way. However once we leave site we will not be able to identify your data so it cannot be withdrawn after this point.

This study aims to build a chemical and biological profile of an individual based on analysis of fingerprint residue. This will include determination of sex and age of the donor.

It is envisaged that this research will benefit the wider community in terms of aiding suspect identification in serious crime investigations

If you agree to take part your fingerprint will be taken and some basic information (including age, sex and ethnicity) collected in the form of a participant questionnaire. This should take 1 minute. This information will be assigned an anonymous code so you will not be able to be identified once the fingerprint has been taken.

The results of this study may be published, which might include data and images. You will not be identifiable in any report or publication. If you wish to view a copy of any reports that result from participation in this study, please provide an email address to a member of the research team.

This project is a collaboration between Sheffield Hallam University Biomedical Science Research Centre and the Defence, Science and Technology Laboratories, Porton Down.

This research project has been approved by the research ethics committee at Sheffield Hallam University.

If you have any complaints about the project, in the first instance please contact a member of the research team.

Thank you for taking part in this research.

Contact details:

Cameron Heaton, BMRC, Sheffield Hallam University
Email: ▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮

1. **Legal basis for research for studies.** The University undertakes research as part of its function for the community under its legal status. Data protection allows us to use personal data for research with appropriate safeguards in place under the legal basis of public tasks that are in the public interest.

2. **A full statement of your rights can be found at** https://www.shu.ac.uk/about-this-website/privacy-policy/privacy-notices/privacy-notice-for-research. However, all University research is reviewed to ensure that participants are treated appropriately and their rights respected. This study was approved by UREC with Converis number ER6558932. Further information at https://www.shu.ac.uk/research/ethics-integrity-and-practice

3. Information in box below (this can all be in smaller print but needs to be legible.)

You should contact the Data Protection Officer if:	You should contact the Head of Research Ethics (Professor Ann Macaskill) if:
- you have a query about how your data is used by the University - you would like to report a data security breach (eg if you think your personal data has been lost or disclosed inappropriately) - you would like to complain about how the University has used your personal data DPO@shu.ac.uk	you have concerns with how the research was undertaken or how you were treated a.macaskill@shu.ac.uk

Postal address: Sheffield Hallam University, Howard Street, Sheffield S1 1WB

Telephone: ▇▇▇▇▇▇

1(ii) Participant Consent Form template

Sheffield Hallam University | Biomolecular Sciences Research Centre

[dstl] Defence, Science & Technology Laboratory

Participant Consent Form

Title of project: *Human identification though advanced forensic mass spectrometry of fingermarks*
The participant should complete the whole of this sheet themselves

Have you read and understood the Participant Information Sheet? Yes☐   No☐

Have you had an opportunity to ask questions and discuss this study? Yes☐   No☐

Have you received satisfactory answers to all of your questions? Yes☐   No☐

Have you received enough information about this study? Yes☐   No☐

Do you understand that you are free to withdraw from the study:

- At any time during sampling

- Without giving a reason

- Without affecting your future in any way Yes☐   No☐

Have you had sufficient time to consider the nature of this project? Yes☐   No☐

Do you agree to take part in this study? Yes☐   No☐

Signed...   Date.....................

Name (BLOCK CAPITALS)...

1(iii) Participant Questionnaire template

Sheffield Hallam University | Biomolecular Sciences Research Centre

[dstl] Defence, Science & Technology Laboratory

Participant Questionnaire

Title of project: *Human identification though advanced forensic mass spectrometry of fingermarks*

The participant should complete this questionnaire sheet to the best of their knowledge

Sample number:

Sex:

☐ Male

☐ Female

Age: ....

Smoker?:

☐ No

☐ Yes   if yes, how many per day:................

Did you use any toiletry products (including hair products) or make up today?

☐ No

☐ Yes   if yes, please specify:...

....

Are you taking regular medication?

☐ No

☐ Yes   if yes, please specify:...

Do you have any major health issues/illnesses?

☐ No

☐ Yes   if yes, please specify:...

Ethnicity:

Please choose one of the sections below and tick the appropriate box:	
White	**Black or black British**
☐British	☐Caribbean
☐Irish	☐African
☐Any other white background (please specify)	☐Any other black background (please specify) ....
Mixed	**Asian or Asian British**
☐White and black Caribbean	☐Indian
☐White and black African	☐Pakistani
☐White and Asian	☐Bangledeshi
☐Any other mixed background (please specify)	☐Any other Asian background (please specify) ....
Chinese or other ethnic group	**Other**
☐Chinese	☐Would prefer not to say
☐Any other (please specify)	☐Unknown

Thank you for completing this participant questionnaire

Appendix 2

2(i) R scripts for sex classification model

The .r file is an R script used for the initial processing of raw supplied spectra, and performs the peak-picking and TIC normalisation using the MALDIquant R package.

[Available upon request]

2(ii) Python scripts for sex classification model

The other scripts are .py python files. These python scripts take the pre-processed spectra dataset files from the R script referenced above, and then perform all the model training/validation tasks.

- The "master" script is called "main.py" and it is the one that controls the others (you would run "python main.py" on the command line to kick off model training).

- The "constants.py" script contains a series of hardcoded variables that specify what model types to train, where to write the output statistics files, the number of k-folds to perform for cross-validation etc. You'll see that these variables specified in constants.py are referenced within many of the other python files.

- "train_model.py" contains the main functions used to train models and perform cross validation - you'll see that this file contains some pretty big functions, but most of the logic in these functions is actually to do with (a) passing/retrieving information to/from other functions in other .py files

- "models.py" contains some quick functions to retrieve different ML model types (e.g. Random forest, XGBOOST etc) - for reference here, I always

think it is quite interesting to see how little code actually getting the models takes, compared to the massive amount of code required for processing data, model training and validation(!).

- "read_data.py" contains a series of functions for reading in the processed M/Z spectra files (output from the R script referenced above). In particular, we make heavy use of pandas dataframes here to easily store the data. You'll see functions contained here for doing things like adding extra columns to the dataset for the sex, age and flagged contamination per sample.

- "preprocessing.py" contains a series of utility functions for applying transformations for the dataset, once it is read in. In particular, there are functions here to apply common scaling strategies to the dataset prior to modelling (e.g. standardisation) as well as feature selection strategies to optionally discard some m/z peak positions prior to training based on Chi-squared statistics or the VIP scoring protocol.

- "vip_scoring.py" is a separate set of functions that contains a python-implemented VIP feature-selection protocol - these functions are directly used by "preprocessing.py" in the above bullet point.

- "contam_filter.py" contains a quick utility function which can optionally be applied to the dataset to filter flagged contaminated samples out from the training dataset prior to modelling.

- "model_scorer.py" contains a function which calculates summary performance metrics given lists of true and predicted values output from the model. Depending on the specified model type (i.e. is it binary classification or regression), different metrics are applied (i.e. accuracy for classification and R2 for regression)

- "results.py" contains a class to handle the model performance statistics generated during cross-validation. It essentially stores all the cross validation metrics as the big training loop progresses, and then writes these metrics to file at the end of the training session.

[Available upon request]