Universidade de Aveiro Departamento de Química
2013

Gonçalo Miguel Gomes Graça

Metabonomics of human amniotic fluid for prenatal diagnostics

A metabonómica do líquido amniótico humano no diagnóstico pré-natal

Universidade de Aveiro Departamento de Química
2013

Gonçalo Miguel Gomes Graça

Metabonomics of human amniotic fluid for prenatal diagnostics

A metabonómica do líquido amniótico humano no diagnóstico pré-natal

Tese apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Doutor em Química, realizada sob a orientação científica da Doutora Ana Maria Pissarra Coelho Gil, Professora Associada com Agregação do Departamento de Química da Universidade de Aveiro e do Doutor Brian James Goodfellow Professor Auxiliar do Departamento de Química da Universidade de Aveiro.

Apoio financeiro da Fundação para a Ciência e Tecnologia (FCT) - bolsa de Investigação SFRH/BD/41869/2007 financiada pelo Programa Operacional Potencial Humano (POPH) e projecto PTDC/QUI/66523/2006 financiado pelo Fundo Europeu de Desenvolvimento Regional (FEDER) e pelo Programa Operacional Factores de Competitividade (COMPETE); da Universidade de Aveiro - Centro de Investigação em Materiais Cerâmicos e Compósitos (CICECO); da empresa Bruker BioSpin GmbH; da Rede Nacional de RMN (PTNMR).

Dedico este trabalho à minha esposa Ana Margarida
pelo incondicional apoio em todos os momentos.

O júri
presidente

Prof. Doutor Vasile Staicu
Professor Catedrático do Departamento de Matemática da Universidade de Aveiro

Profª Doutora Maria Helena Dias dos Santos
Professora Catedrática Aposentada do Instituto de Tecnologia Química e Biológica (ITQB) da Universidade Nova de Lisboa

Doutor John Griffith Jones
Investigador do Centro de Neurociências e Biologia Celular (CNC) da Universidade de Coimbra

Doutora Isabel Marques Carreira
Professora Auxiliar com Agregação da Faculdade de Medicina da Universidade de Coimbra (FMUC), Directora do Laboratório de Citogenética e Genómica - FMUC, Coordenadora do Centro de Investigação Meio Ambiente, Genética e Oncobiologia (CIMAGO) da FMUC e membro do Centro de Ciências Forenses (CENCIFOR) do Instituto de Medicina Legal

Doutora Ana Maria Pissarra Coelho Gil
Professora Associada com Agregação do Departamento de Química da Universidade de Aveiro

Doutor Brian James Goodfellow
Professor Auxiliar do Departamento de Química da Universidade de Aveiro

Doutor António de Sousa Barros
Investigador Auxiliar da Unidade de Investigação de Química Orgânica, Produtos Naturais e Agro-alimentares (QOPNA) da Universidade de Aveiro

Acknowledgements

I would like to express my sincere gratitude to my supervisors Dr. Ana M. Gil and Dr. Brian J. Goodfellow for the opportunity to work with them in the exciting field of metabonomics, for their guidance, support and for the knowledge passed on which allowed me to grow as a researcher.

I express my gratitude also to Dr. Iola Duarte from whom I received my initial training in NMR acquisition and to Dr. António Barros, for the advice and fruitful discussions on multivariate analysis.

I also acknowledge all the people involved in sample and clinical data collection for this project including the doctors, nurses and secretaries, in particular Dr. Eulália Galhano, Dr. Ana Bela Couceiro and nurse Cristina Pita (from *Maternidade Bissaya Barreto*, Coimbra) and Prof. Isabel Carreira and Cláudia Pais (from *Laboratorio de Citogenética e Genómica, Faculdade de Medicina, Universidade de Coimbra*). *Centro Hospital de Coimbra* is also aknowledge for accommodating this research project.

A special thanks to Ana Moreira and Ana Correia for their collaboration in the MIR work, and to Sílvia Diaz for sharing the NMR spectra of maternal urine used in this work.

I acknowledge the collaboration with the Bruker BioSpin NMR Applications division, from which I thank specially Dr. Manfred Spraul, Dr. Li-Hong Tseng, Dr. Alexandre Schefer, Dr. Markus Godejohann and Dr. Ulrich Braumann.

The UPLC-MS work would not be possible without the collaboration with the Biomolecular Medicine group (Imperial College of London). I express my gratitude for the opportunity to work with them especially to Prof. John C. Lindon and Dr. Elizabeth Want, and also for the assistance provided by Dr. Konstantina Spagou, Dr. Kirill Veselkov, Dr. Perrine Masson and Dr. Olaf Beckonert.

I also would like to thank Dr. Eurico Cabrita from *Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa* (FCT-UNL) for performing the STD experiments and to Dr. Rosário Domingues from the *Departmento de Química* (*Universidade de Aveiro*) for the help provided in the interpretation of MS spectra.

Some people say doing a PhD is like crossing a desert. I certainly did not cross this “desert” alone, I was fortunate enough to have many colleagues and friends in *Departmento de Química* and CICECO who made this journey a much more pleasant one. I specially acknowledge my friends João Rodrigues, Sílvia Diaz, Joana Pinto, Cláudia Rocha, Inês Lamego, Joana Carrola and Joana Marques, for their companionship and support.

Finally I deeply thank my family, especially my parents and my wife, for their unconditional love, patience and support during these PhD years.

Palavras-chave

Líquido amniótico humano, urina materna, metabonómica, diagnóstico pré-natal, patologias pré-natais, espectroscopia de ressonância magnética nuclear (RMN), espectroscopia do infra-vermelho médio (MIR), cromatografia líquida de ultra eficiência acoplada a espectrometria de massa (UPLC-MS), análise multivariada (MVA).

Resumo

O trabalho apresentado nesta tese teve como principais objectivos contribuir para o conhecimento da composição do líquido amniótico humano (LA), colhido no 2º trimestre de gravidez, assim como investigar possíveis alterações na sua composição devido à ocorrência de patologias pré-natais, recorrendo à metabonómica e procurando, assim, definir novos biomarcadores de doenças da grávida e do feto.

Após uma introdução descrevendo o estado da arte relacionado com este trabalho (Capítulo 1) e os princípios das metodologias analíticas usadas (Capítulo 2), seguida de uma descrição dos aspectos experimentais associados a esta tese (Capítulo 3), apresentam-se os resultados da caracterização da composição química do LA (gravidez saudável) por espectroscopia de ressonância magnética nuclear (RMN), assim como da monitorização da sua estabilidade durante o armazenamento e após ciclos de congelamento-descongelamento (Capítulo 4). Amostras de LA armazenadas a -20°C registaram alterações significativas, tornando-se estas menos pronunciadas (mas ainda mensuráveis) a -70°C, temperatura recomendada para o armazenamento de LA. Foram também observadas alterações de composição após 1-2 ciclos de congelamento-descongelamento (a ter em conta aquando da reutilização de amostras), assim como à temperatura ambiente (indicando um período máximo de 4h para a manipulação e análise de LA). A aquisição de espectros de RMN de ^{1}H de alta resolução e RMN acoplado (LC-NMR/MS) permitiu a detecção de 75 compostos no LA do 2º trimestre, 6 dos quais detectados pela primeira vez no LA. Experiências de difusão (DOSY) permitiram ainda a caracterização das velocidades de difusão e massas moleculares médias das proteínas mais abundantes.

O Capítulo 5 descreve o estudo dos efeitos de malformações fetais (FM) e de cromossomopatias (CD) na composição do LA do 2º trimestre de gravidez. A extensão deste trabalho ao estudo dos efeitos de patologias no LA que ocorrem no 3º trimestre de gravidez é descrita no Capítulo 6, nomeadamente no que se refere ao parto pré-termo (PTD), pré-eclampsia (PE), restrição do crescimento intra-uterino (IUGR), ruptura prematura de membranas (PROM) e diabetes mellitus gestacional (GDM). Como complemento a estes estudos, realizou-se uma análise preliminar da urina materna do 2º trimestre para o estudo de FM e GDM, descrita no Capítulo 7. Para interpretação dos dados analíticos, obtidos por espectroscopia RMN de ^{1}H, cromatografia líquida de ultra eficiência acoplada a espectrometria de massa (UPLC-MS) e espectroscopia do infravermelho médio (MIR), recorreu-se à análise discriminante pelos métodos dos mínimos quadrados parciais e o método dos mínimos quadrados parciais ortogonal (PLS-DA e OPLS-DA) e à correlação espectral. Após análise por validação cruzada de Monte-Carlo (MCCV), os modelos PLS-DA de LA permitiram distinguir as FM dos controlos (sensibilidades 69-85%, especificidades 80-95%, taxas de classificação 80-90%), revelando variações metabólicas ao nível do metabolismo energético, dos metabolismos dos aminoácidos e glícidos assim como possíveis alterações ao nível do funcionamento renal. Observou-se também um grande impacto das FM no perfil metabólico da urina materna (medido por UPLC-MS), tendo no entanto sido registados modelos PLS-DA com menor sensibilidade (40-60%), provavelmente devido ao baixo número de amostras e maior variabilidade da composição da urina (relativamente ao LA). Foram sugeridos possíveis marcadores relacionados com a ocorrência de FM, incluindo lactato, glucose, leucina, valina, glutamina, glutamato, glicoproteínas e conjugados de ácido glucurónico e/ou sulfato e compostos endógenos e/ou exógenos (<1 μM) (os últimos visíveis apenas na urina). No LA foram também observadas variações metabólicas devido à ocorrência de vários tipos de cromossomopatias (CD), mas de menor magnitude. Os perfis metabólicos de LA associado a *pré*- PTD produziram modelos que, apesar do baixo poder de previsão, sugeriram alterações precoces no funcionamento da unidade fetoplacentária, hiperglicémia e *stress* oxidativo. Os modelos obtidos para os grupos *pré*- IUGR

pré- PE, *pré*- PROM e *pré*-diagnóstico GDM (LA e urina materna) registaram baixo poder de previsão, indicando o pouco impacto destas condições na composição do LA e/ou urina do 2º trimestre.

Os resultados obtidos demonstram as potencialidades da análise dos perfis metabólicos do LA (e, embora com base em menos estudos, da urina materna) do 2º trimestre para o desenvolvimento de novos e complementares métodos de diagnóstico, nomeadamente para FM e PTD.

Keywords

human amniotic fluid, maternal urine, metabonomics, prenatal diagnosis, prenatal disorders, nuclear magnetic resonance (NMR) spectroscopy, mid-infrared (MIR) spectroscopy, ultra-performance liquid chromatography-mass spectrometry (UPLC-MS), multivariate analysis (MVA).

Abstract

The work presented in this thesis aimed to contribute to knowledge of 2nd trimester human amniotic fluid (AF) composition and to investigate the possible metabolic effects of prenatal disorders on AF composition through metabonomics, in order to define new potential disorder biomarkers.

After an introduction describing the state-of-the-art (Chapter 1), the analytical methodologies used (Chapter 2) and the description of the experimental details of the work performed (Chapter 3), the results from the chemical characterization of AF (healthy pregnancy) by nuclear magnetic resonance (NMR) spectroscopy are presented, as well as the results of AF stability assessment during storage and after freeze-thaw cycles (Chapter 4). AF samples stored at -20°C registered significant compositional changes, less marked (but still measurable) at -70°C, the latter temperature being then recommended for AF storage. In addition, significant compositional changes were also observed after 1-2 freeze-thaw cycles (to be considered when sample re-usage is necessary), and at room temperature (indicating a maximum period of 4h for handling and analysis of AF). High resolution ^{1}H NMR and hyphenated NMR (LC-NMR/MS) analysis enabled the detection of 75 different compounds in 2nd trimester AF, 6 of which were detected for the first time in AF. Moreover, diffusion-edited spectroscopy (DOSY) experiments allowed the characterization of the main AF proteins in terms of diffusivity and, hence, average molecular weight.

Chapter 5 describes the study of the effects of fetal malformations (FM) and chromosomal disorders (CD) on the composition of 2nd trimester AF. In Chapter 6, this approach is extended, to the effects of 3rd trimester disorders, namely preterm delivery (PTD), preeclampsia (PE), intrauterine growth restriction (IUGR), premature rupture of the membranes (PROM) and gestational diabetes mellitus (GDM). These studies were complemented by a preliminary analysis of 2nd trimester maternal urine to study FM and GDM (Chapter 7). Interpretation of the analytical data obtained by ^{1}H NMR spectroscopy, ultra performance liquid chromatography-mass spectrometry (UPLC-MS) and mid-infrared spectroscopy (MIR) was performed through partial least squares and orthogonal partial least squares - discriminant analysis (PLS-DA and OPLS-DA) and statistical correlation spectroscopy. Monte-Carlo cross-validated (MCCV) PLS-DA models of AF revealed separation of FM cases from controls (sensitivities 69-85%, 80-95% specificities, classification rates: 80-90%), revealing disturbances in energy metabolism, amino acids and sugar metabolisms and possibly abnormal kidney function. A high impact of FM on maternal urine was also observed (by UPLC-MS), however, the models obtained were of lower sensitivity (40-60%), probably due to the low sample numbers and higher variability of urine composition (in relation to AF). Possible markers of FM were suggested including lactate, glucose, leucine, valine, glutamine, glutamate, glycoproteins and conjugation products of glucuronic acid and/or sulfate with endogenous and/or exogenous metabolites (<1 μM). In addition, metabolite variations were found in AF related to the occurrence of several types of chromosomal disorders (CD), although of smaller magnitude. Second trimester AF profiling associated with *pre*-PTD produced models, which, despite their low predictive power, enabled the detection of metabolite variations suggestive of early fetal-placental dysfunction, hyperglycaemia and oxidative stress. The models obtained for *pre*-IUGR, *pre*-PE, *pre*-PROM and *pre*-diagnostic GDM (both AF and urine) showed low predictive power, reflecting the small impact of these disorders in 2nd trimester AF and/or urine composition.

The results presented demonstrate the potential of metabolic profiling of 2nd trimester AF (and, although based on less studies, maternal urine) for the development of new complementary prenatal diagnosis methods, namely for FM and PTD.

List of publications including the work presented in this thesis

Papers in peer-reviewed journals:

Graça, G., Duarte, I. F., Goodfellow, B. J., Barros, A. S., Carreira, I. M., Couceiro, A. B., Spraul, M. and Gil, A. M.; Potential of NMR spectroscopy for the study of human amniotic fluid. *Analytical Chemistry*, **2007**, 79(21): 8367-8375.

Graça, G., Duarte, I. F., Goodfellow, B. J., Carreira, I. M., Couceiro, A. B., do Rosario Domingues, M., Spraul, M., Li-Hong, T. and Gil, A. M.; Metabolite profiling of human amniotic fluid by hyphenated nuclear magnetic resonance spectroscopy. *Analytical Chemistry*, **2008**, 80(15): 6085-6092.

Graça, G., Duarte, I. F., Barros, A. S., Goodfellow, B. J., Diaz, S., Carreira, I. M., Couceiro, A. B., Galhano, E. and Gil, A. M.; ^{1}H NMR based metabonomics of human amniotic fluid for the metabolic characterization of fetus malformations. *Journal of Proteome Research*, **2009**, 8(8): 4144-4150.

Graça, G., Duarte, I. F., Barros, A. S., Goodfellow, B. J., Diaz, S. O., Pinto, J., Carreira, I. M., Galhano, E. l., Pita, C. and Gil, A. M.; Impact of prenatal disorders on the metabolic profile of second trimester amniotic fluid: a nuclear magnetic resonance metabonomic study. *Journal of Proteome Research*, **2010**, 9(11): 6016-6024.

Graça, G., Diaz, S. O., Pinto, J., Barros, A. S., Duarte, I. F., Goodfellow, B. J., Galhano, E., Pita, C., Almeida, M. d. C., Carreira, I. M., and Gil, A. M. ; Can biofluids metabolic profiling help to improve healthcare during pregnancy?, *Spectroscopy: An International Journal*, **2012**, 27(5-6): 515-523.

Graça, G., Goodfellow, B. J., Barros, A. S., Diaz, S. O., Duarte, I. F., Spagou, K., Veselkov, K., Want, E. J., Lindon, J. C., Carreira, I. M., Galhano, E., Pita, C. and Gil, A. M.; UPLC-MS metabolic profiling of second trimester amniotic fluid and maternal urine and comparison with NMR spectral profiling for the Identification of pregnancy disorder biomarkers, *Molecular BioSystems*, **2012**, 8: 1243-1254.

Graça, G., Moreira, A. S., Correia, A.J.V., Goodfellow, B. J., Barros, A. S., Duarte, I. F., , Carreira, I. M., Galhano, E., Pita, C., and Gil, A. M.; Mid-infrared (MIR) metabolic fingerprinting of amniotic fluid: a possible avenue for early diagnosis of prenatal disorders?, *Analytica Chimica* Acta, **2013**, 764: 24–31.

Book chapter:

Gil, A. M. and Graça, G.; *Nuclear Magnetic Resonance Methods for Metabolomic Investigation of Amniotic Fluid. Methodologies for Metabolomics: Experimental Strategies and Techniques*. N. W. Lutz, J. V. Sweedler and R. A. Wevers, UK, Cambridge University Press, 2013.

Contents

Abbreviations and symbols

1D	One-dimensional
2D	Two-dimensional
ACN	Acetonitrile
ANN	Artificial neuronal networks
AF	Amniotic Fluid
ANOVA	Analysis Of Variance
ATR	Attenuated Total Reflectance
β-hCG	β-human chorionic gonadotropin
BPPLED	Bipolar pulse longitudinal eddy current delay
cAMP	Cyclic adenosine monophosphate nucleotide
CD	Chromosomal Disorders
CE	Capillary Electrophoresis
CPMG	Carr-Purcell-Meiboom-Gill
CV	Cross Validation
DAD	Diode Array Detector
dB	Decibel
DNA	Deoxyribonucleic Acid
DOSY	Diffusion Ordered Spectroscopy
DSS	Dimethylsilapentanesulfonic acid
F1	Frequency domain 1
F2	Frequency domain 2
FID	Free Induction Decay
FISH	Florescence *in situ* hybridization
FIR	Far-infrared
FM	Fetal Malformations
FT	Fourier Transform
GC	Gas Chromatography
GDM	Gestational Diabetes Mellitus
g.w.	Gestational weeks
HETCOR	Heteronuclear Correlation
HCA	Hierarchical cluster analysis
HILIC	Hydrophilic Interaction Chromatography
HPLC	High-Performance Liquid Chromatography
HSQC	Heteronuclear Single Quantum Correlation
HMQC	Heteronuclear Multiple Quantum Correlation
HSS	High-Strength Silica
Hz	Hertz
IEM	Inborn Errors of Metabolism
INEPT	Insensitive Nuclei Enhancement by Polarization Transfer
IR	Infrared
IUGR	Intrauterine Growth Restriction
IVH	Intraventricular hemorrhage
J-res	*J*-resolved spectroscopy
LB	Line Broadening
LC	Liquid Chromatography
LOESS	Locally weighted Scatter plot Smoothing
MCCV	Monte-Carlo Cross Validation
MHz	Megahertz

MIR	Mid-infrared
MLEV-17	Malcolm Levitt-17 sequence
MRI	Magnetic Resonance Imaging
MS	Mass Spectrometry
MS^n	n^{th} fragmentation MS experiment
MS^E	MS experiments at different collision energies
MVA	Multivariate Analysis
M_w	Molecular weight
NIR	Near-infrared
NLM	Non-linear mapping
NMR	Nuclear Magnetic Resonance
NMND	N-methyl-nicotinamide
NS	Number of scans (or transients)
OGTT	Oral glucose tolerance test
OPLS	Orthogonal Partial Least Squares
OPLS -DA	Orthogonal Partial Least Squares - Discriminant Analysis
PAPP-A	Pregnancy-associated plasma protein A
PCA	Principal Components Analysis
PE	Preeclampsia
PlGF	Placental growth factor
PLS	Partial Least Squares
PLS -DA	Partial Least Squares - Discriminant Analysis
PROM	Premature Rupture Of Membranes
PTD	Preterm Delivery
Q^2	Goodness of prediction or predictive power of a PLS-DA or OPLS-DA model
Q-TOF	Quadrupole-Time-Of-Flight
RD	Relaxation Delay
RF	Radiofrequency
RDS	Respiratory Distress Syndrome
ROC	Receiver Operative Characteristics
RP	Reverse Phase
RSPA	Recursive Segment-wise Peak Alignment
SGA	Small for Gestational Age
SHY	Statistical Heterospectroscopy
SI	Size of real spectrum
SIMCA	Soft Independent Modeling of Class Analogy
STD	Saturation Transfer Difference
SOM	Self organizing maps
STOCSY	Statistical Total Correlation Spectroscopy
SW	Spectral Width
TD	Time Domain
TLC	Thin-layer chromatography
TOCSY	Total Correlation Spectroscopy
TOF	Time-Of-Flight
TMS	Tetramethylsilane
TSP	3-trimethylsilyl propionate
UPLC	Ultra-Performance Liquid Chromatography
UV	Ultraviolet
UV scaling	Unit Variance scaling
VIP	Variable importance to the projection
WDW	Window function

1. Introduction

1.1. Pregnancy and fetal development

Human pregnancy normally has the duration of 38 to 40 gestational weeks (g.w.) and in order to better describe the events that occur throughout gestation it is usually divided into 3 trimesters. The 1st trimester of pregnancy comprises the period from fertilization until 12 g.w. It is during the 1st trimester that most important steps in the fetal organogenesis occurs, and is the most critical period of fetal development. The cardiac and central nervous system, for instance, are completely differentiated by the end of the 8th week and, by the 12th week the limbs appear completely differentiated (Moore and Persaud, 2008). Also at this stage, important fetal-maternal structures become completely differentiated ultimately forming the definitive annexes that support fetal growth and maintenance and also playing important regulatory and endocrine functions (Jauniaux and Gulbis, 2000; Graça and Barros, 2005). Some of the most important of these structures, shown in detail in Figure 1.1, are the chorionic villi and the capillary system, which form the placenta and the umbilical cord, respectively; the chorion, later forming the amniotic sac and the excoelomic cavity which is actively involved in the early amniotic fluid (AF) formation (Jauniaux and Gulbis, 2000).

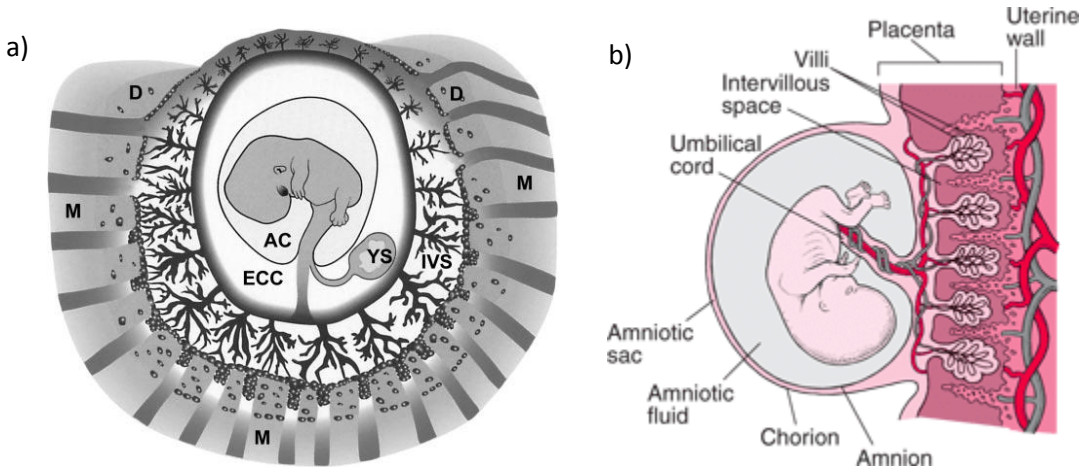


Figure 1.1 – The placenta and fetal annexes: a) diagram of the early gestational sac (8–9 g.w.) illustrating the fetal fluid compartments; amniotic fluid fills the amniotic sac (AC), and is surrounded by the exocoelomic cavity (ECC); the chorionic villous and intervillous fluid fills the intervillous space (IVS); D, Decidua; M, myometrium; YS, yolk sac (adapted from Jauniaux *et al.*, 2005); b) detailed representation of the placenta and fetal structures at the 10-12th g.w. (adapted from Brown, 2007).

The 2nd trimester of pregnancy (from 12 to 26 g.w.) is characterized by organ maturation and fetal growth. During the 2nd trimester the fetus starts to move more frequently, the fetal kidneys acquire the filtration capacity and start to produce fetal urine. The liver starts to secrete bile and the spleen initiates the production of erythrocytes. Towards the end of 2nd trimester fetal skin becomes keratinized and lung maturation is initiated with the production of the lung surfactant which prepares the fetal lung for its normal function after birth (Toot and Lu, 2004; Graça, 2005a).

Important physiological changes occur during the 3rd trimester (from 26 g.w. until labour), such as the initiation of fetal liver lipid biosynthetic function enabling the fetus to synthesize body fat in order to maintain temperature. It is also during the 3rd trimester that the bones of the skull become completely formed. A marked increase in fetal weight is registered also during this period. By the 37th week, fetal maturation is complete and pregnancy is at term, although labour can occur after this period, up to 41 or 42 weeks (Toot and Lu, 2004; Graça, 2005a).

In order to support the newly forming life, the maternal body has to adjust its physiology without compromising the maternal well-being. However, in some cases the disruptions that occur during gestation can cause severe maternal health problems and ultimately affect fetal development. For this reason it is important to ensure adequate pregnancy monitoring and eventually perform prenatal diagnosis procedures in case of enhanced risk of prenatal disorders, in order to allow proper intervention or, in most severe cases, pregnancy termination.

1.2. Pregnancy-related disorders and prenatal diagnosis: present status and challenges

Some of the most important prenatal disorders affecting the fetus, the mother or both are summarized in Table 1.1, together with their respective estimated prevalence, i.e. the portion of pregnancies registering such disorders.

Table 1.1 - Estimated prevalence of pregnancy-related disorders worldwide.

Disorder	Prevalence	reference
Intrauterine growth restriction (IUGR)	7-15% of all pregnancies	Cetin and Alvino, 2009
Gestational diabetes mellitus (GDM)	1-14% of all pregnancies	ADA, 2004
Premature rupture of the membranes (PROM)	8-10% of all pregnancies	Stuart *et al*, 2005
Preterm delivery (PTD)	9.6% of all live births	Gracie, *et al.* 2012
Preeclampsia (PE)	2-6% of all pregnancies	Walker, 2000
Fetal malformations (FM)	3-4% of all live births	Bacino, 2012
Chromosomal disorders (CD)	0.5% of live births	Lu and Hobel, 2004

Among the most prevalent prenatal disorders, worldwide, are the intrauterine growth restriction (IUGR), gestational diabetes mellitus (GDM), the premature rupture of the membranes and the preterm delivery (PTD) (Table 1.1). Nevertheless, most of the prenatal disorders mentioned induce significant pathophysiologic changes in the mother, fetus of both and, therefore, will be described in detail in the following subsections.

1.2.1. Fetal malformations (FM)

Fetal malformations (FM), often designated by congenital fetal malformations, congenital anomalies or birth defects, are anatomic alterations which occur during fetal development and may affect several organs, tissues and limbs (Table 1.2).

Table 1.2 - Some of the most common fetal malformations grouped according to the affected organ or system (Adapted from Chmait and Moore, 2004).

Affected organ/system	Malformations	
Central nervous system (CNS)	• Hydrocephaly • Anencephaly	• Spina Bifida • Ventriculomegaly
Face	• Cleft lip and/or palate	
Limbs	• Amelia (absence of limbs) • Club foot	• Polydactylia • Achondroplasia
Soft tissue	• Cystic hygroma	
Lungs	• Cystic adenomatoid malformation • Lung sequestration	• Diaphragmatic hernia
Cardiac	• Atrial septal defect • Ventricular septal defect • Tetralogy of Fallot	• Transposition of the great vessels • Arrhythmias
Abdominal	• Gastroschisis • Omphalocele	• Bowel atresia or obstruction
Urogenital	• Renal agenesis • Polycystic kidney disease	• Hydronephrosis • Posterior urethral valves
Bones	• Skeletal dysplasia	

Fetal malformations can be caused by genetic factors such as chromosomal alterations and gene mutations, environmental factors such as drugs and viruses, or multifactorial inheritance (genetic and environmental factors acting together in a complex manner) (Moore and Persaud, 2008). An example of the effects of an environmental agent is the infection with rubella virus during the critical period of development of the eyes, heart, and ears, which may cause severe anatomic anomalies, such as cataracts, cardiac defects, and deafness. Another well-known example relates to the effects caused by the drug thalidomide, taken by the mothers during early

pregnancy to reduce the effects of morning sickness, which caused severe fetal limb anomalies and other developmental disruptions. However, for most FM types the etiology is still unknown (Moore and Persaud, 2008).

The worldwide prevalence of FM is 3-4% of all live births (Table 1.1, page 2), being the brain defects the most common, occurring in every 10/1000 births, followed by heart defects, which occur in 8/1000 births (Miall *et al.*, 2003).

In terms of severity, FM types include those of minor severity, such as a club foot which can be corrected with surgery after birth, and those of major severity such as the cases of open spina bifida, transposition of great vessels or even incompatible with life such as anencephaly (Bacino, 2012). Some fetuses may also show characteristic patterns of abnormality which may suggest a "syndrome" diagnosis, such as those associated with chromosomal disorders, as for instance, Down syndrome (trisomy 21) (which will be discussed in detail in the next section) (Bacino, 2012).

The screening for FM types such as those presented in Table 1.2 is normally achieved by transabdominal ultrasound monitoring performed in late 1st trimester (11-14 g.w.) and during the 2nd trimester (around 16 g.w.). The transabdominal ultrasound is of paramount importance in prenatal diagnosis since it allows the correct determination of fetal gestational age, the evaluation of the biophysical profile of the fetus, which combines the heart rate, breathing activity, inspection of fetal movements and tone, as well as amniotic fluid volume evaluation (Chmait and Moore, 2004). The sensitivity for FM detection by ultrasound during the 2nd trimester ranges from 17% to 74%, which may be due to patient population variations or due to operator experience (Chmait and Moore, 2004). On the other hand, the specificity of ultrasound for the identification of a normal fetus is almost 100%, i.e. it has a very low false positive rate for the identification normal fetus (Chmait and Moore, 2004). Therefore, ultrasound can be very specific for ruling out FM, but it lacks sensitivity in detecting them.

Other diagnostic methods may be used when ultrasound imaging is equivocal or nondiagnostic. The main non-invasive alternative is the use of magnetic resonance imaging (MRI) for the diagnosis of complex central nervous system malformations such as ventriculomegaly, whereas laboratory determination of elevated α-fetoprotein levels in AF may be used to assist in the diagnosis of open neural tube and ventral wall defects, such as spina bifida and gastroschisis, respectively (Chmait and Moore, 2004).

During embryogenesis, the metabolic microenvironment profoundly influences the entire organogenesis processes (Prasad *et al.*, 2009). Indeed, specific metabolite variations were

described for different types of malformations such as neural tube defects namely increases in lactate, glutamate, acetate (Bock, 1994), *myo*-inositol and succinate (Groenen *et al.*, 2004). Decreases in folate and cobalamin levels were also found in association with neural tube defects, cleft lip and/or palate, heart defects or omphalocele (Brouns *et al.*, 2008). Moreover, inborn errors of metabolism (IEM) have been associated with the development of FM, particularly those of central nervous system (Prasad *et al.*, 2009). These IEM are caused by single gene defects resulting in enzymatic impairments within biochemical pathways, which may lead to accumulation of potentially toxic intermediate compounds. Several metabolite changes are then expected in such cases and may be used to further characterize FM cases and be used in the development of complementary diagnostic methods.

1.2.2. Chromosomal disorders (CD)

Chromosomal disorders (CD) reflect the occurrence of an atypical number of chromosomes or a structural alteration (chromosomal anomalies) in one or more of the 46 chromosomes (23 chromosome pairs) that constitute the human karyotype (Figure 1.2).

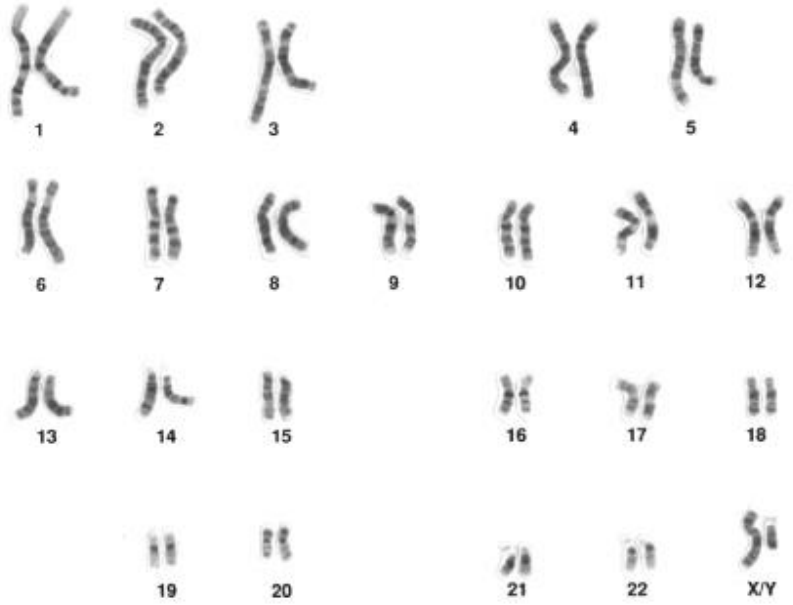


Figure 1.2 - The karyotype of a human composed of 23 pairs of chromosomes. The 23rd pair (XY) constitutes the sexual chromosomes of a male (XX in females). Reproduced from "Talking Glossary of Genetic Terms" (http://www.genome.gov/Glossary/).

Chromosomal anomalies occur in 0.5% of live births (Table 1.1, page 2) and can be caused by errors in cell division or by extrinsic factor such as radiation, chemical substances or viral infections (Lu and Hobel, 2004; Tavares, 2005). Women older than 34 years have increased risk of

giving birth to children carrying chromosomal anomalies and related FM types (Lu and Hobel, 2004), hence CD screening is recommended in pregnant women above this age.

Chromosome anomalies can be divided into three major types: numerical, structural and mosaicisms. Some of the most common are summarized in Table 1.3. Depending on the type, chromosome anomalies may range from small phenotypic alterations to more severe developmental and intellectual impairment such as malformations and mental retardation (Lu and Hobel, 2004).

Table 1.3 - Some of the most common types of chromosomal anomalies (Adapted from Lu and Hobel, 2004 and Tavares, 2005).

Anomaly	Details
Numerical	**Monosomies:** Turner syndrome (45,X)
	Trisomies: Down syndrome (T21), Edwards syndrome (T18), Patau syndrome (T13), Kleinfelter's syndrome (47,XXY), Triple X syndrome (47,XXX), 47,XYY syndrome
	Polysomies: Examples: 48,XXXX, 48,XXXY, 48,XXYY, 49,XXXXY
Structural	**Deletions** - chromosomal material is missing
	Duplications - duplication of chromosome portions
	Translocations - transfer of chromosomal material between chromosomes
	Inversions - inversion of chromosome portions
	Insertions - addition of chromosome material into other chromosomes
	Ring chromosome - simultaneous deletion and junction of telomeres
	Marker chromosome - extra chromosomal material
Mosaicisms	**Cell line with two or more karyotypes:** Example: mos 45,X / 46,XX (mosaic, Turner syndrome)

The most common numerical chromosomal anomalies are named aneuploidies and involve the repetition or absence of one or more chromosomes (Table 1.4). Numerical anomalies are responsible for the most common chromosomal disorders including the autosomal trisomies such as Down syndrome (trisomy 21), Edwards syndrome (trisomy 18) and Patau syndrome (trisomy 13), as well as the sex chromosomal aneuploidies (such as the Turner syndrome, Kleinfelter syndrome or the XYY syndrome) (Lu and Hobel, 2004; Tavares, 2005) (Table 1.4).

Table 1.4 - Characteristics of the most common chromosomal disorders (syndromes): prevalence, karyotype(s) and phenotype changes (Lu and Hobel, 2004; Tavares, 2005; Chen *et al.*, 2011).

Disorder	Karyotypes	Prevalence	Phenotypic changes
Down syndrome (trissomy 21)	47,XX+21 47,XY+21	1/700	• delayed cognitive ability and physical growth; microgenia (abnormally small chin); • oblique eye fissures with epicanthic skin folds on the inner corners; • cardiopathy, thyroid dysfunction, infertility, premature aging.
Edwards syndrome (trissomy 18)	47,XX+18 47,XY+18	1/6000	• hypertonia, microcephaly, prominent back portion of the head, micrognathia (abnormally small jaw), low-set years, clenched hands, rounded bottom feet, unusual shaped chest; • low birth weight and low survival rate (10%, <1 year); • associated with congenital malformations: cardiac defects, omphalocele and diaphragmatic hernia.
Patau syndrome (trissomy 13)	47,XX+13 47,XY+13	1/10000	• severe intra-uterine and post-natal development; • malformations: holoprosencephaly, polydactilia, cardiac defects, oral-facial clefts and ocular anomalies; • very low survival rate (few months); • found in association with spontaneous abortions.
Turner syndrome	45,X	1/5000	• female phenotype, low height (141-146 cm); • sexual immaturity, short and thick neck, lymphedema, large thorax and cardiac anomalies; • 99% of 45X are loss before birth.
Kleinfelter syndrome	47,XXY 48,XXYY 49,XXXXY	1/1000	• male phenotype with high stature, long limbs, sexual underdevelopment, infertility and gynecomasty.
XYY syndrome		1/1000	• male phenotype with high stature and apparently no behaviour alterations.

Structural chromosomal anomalies result from alterations within the portions of genetic material, either by rearrangement leading to no phenotypic consequences (balanced alterations), or by gain or loss of portions of genetic material with phenotypic changes (unbalanced alterations) whose severity depends on the chromosome portions and gene(s) affected (Tavares, 2005). Structural anomalies can be subdivided into deletions, (originating a monosomy for the affected chromosomal regions), duplications (originating trisomies for the affected chromosome regions), translocations, inversions, insertions, ring-shaped chromosomes (originating from deletions at the telomeres level and subsequent junction, resulting in chromosome inactivation

and functional monosomy) and marker chromosomes (small chromosomes of undetermined origin, which can be associated with malformations) (Tavares, 2005) (Table 1.3, page 6).

Chromosomal disorders may also arise from the co-existence of two or more cell lines with distinct karyotypes called mosaicisms, as for instance a mosaicism for Turner syndrome (Table 1.3, page 6). The clinical consequences of a mosaicism depend upon on the proportion of the abnormal cell line(s) in the several tissues and on their corresponding karyotype. Usually the presence of normal cell lines attenuates the effect of the abnormal ones (Tavares, 2005).

The screening for fetal chromosomal anomalies, especially trisomies, may be performed at 10-14 weeks with aid of a 1st trimester combined screening, which includes the ultrasound measurement of fetal nuchal translucency thickness in conjunction with the measurement maternal serum free β-human chorionic gonadotropin (β-hCG) and pregnancy-associated plasma protein A (PAPP-A) (Spencer, 2002). The 1st trimester screening has been associated with sensitivities of 89% for trisomy 21 and 90% for other chromosomal anomalies (trisomies 13 and 18, Turners syndrome, and triploidy) (Spencer, 2002) and the false positive rates of 3.2%–5.6% for 13, 18 and 21 trisomies (Chen *et al.*, 2011). Another screening procedure performed during the 2nd trimester, usually at 16-18 weeks, has also been used (Lu and Hobel, 2004). The 2nd trimester screening included the measurement of maternal serum α-fetoprotein, free or total β-hCG and maternal age, which registered lower sensitivities of 65-70% for trisomy 21 (Spencer, 2002). Nevertheless, the definitive diagnosis of CD involves chromosomal dosage through the use of cytogenetics procedures (Tavares, 2005; Wapner and Toy, 2011).

The most widely used cytogenetic technique in prenatal diagnosis is karyotype analysis, which consists in the observation of the chromosomes from cultivated fetal cell lines during the metaphase stage of cell division (Tavares, 2005). These cells are usually collected invasively, during the 2nd trimester through amniocentesis (to collect amniotic fluid, from which the fetal cells, the amniocytes, are obtained), or at the 1st trimester through chorionic villus sampling (CVS) to obtain fetal cells. Since it can be performed at an earlier stage of pregnancy, CVS (which has nearly the same risk of miscarriage as amniocentesis performed at the 2nd trimester, 0.5 to 1%) represents an advance in the early diagnosis of chromosomal disorders and has been increasingly used to replace amniocentesis for cell collection for cytogenetic analysis (Wapner and Toy, 2011). However, cell culture requires 3 to 7 days for chorionic villus cells and 7 to 10 days for amniocytes (Tavares, 2005), extending the period of analysis up to two weeks. Another important cytogenetic procedure is fluorescence *in situ* hybridization (FISH). This procedure consists of the identification of specific gene sequences through the use of fluorescence DNA probes, enabling the rapid

targeted assessment of chromosomal anomalies especially in non-cultured cells, at several stages of cell division, thus allowing results to be obtained in 24 hours. Nevertheless, the results must be complemented by conventional karyotype analysis (Tavares, 2005). FISH analysis also allows the detection of genetic syndromes (caused by gene deletions), such as Prader-Willi, Angelman-DiGeorge and Williams (Lu and Hobel, 2004).

Chromosomal disorders induce alterations in gene expression and, therefore metabolite variations can be expected. In fact, several metabolite variations were described in 2nd trimester AF of Down syndrome cases (Sims *et al.*, 1993a; Baggot *et al.*, 2008) including increases in 5-hydroxycaproate, methylsuccinate, α-ketoglutarate, adipate and phenylpyruvate thus suggesting impaired riboflavin and neurotransmitter metabolism (Baggot *et al.*, 2008). Another study considering 2nd trimester AF samples detected decreases in glutamine, gycine, taurine, valine, lysine, isoleucine, leucine and ornithine, as well as an increase in glutamate in Down syndrome cases (Amorini *et al.*, 2012). However, early metabolic variations resulting from other CD have not been explored, despite the occurrence of phenotypical alterations which may result in metabolic impairment, such as the impaired sexual development noted in Turner and Klinefelter syndromes (Tavares, 2005), which may have an impact in hormonal biosynthesis and regulation.

1.2.3. Preterm delivery (PTD) and premature rupture of the membranes (PROM)

Preterm delivery (PTD) is defined as the delivery occurring between 20 and 37 weeks of gestation (Hobel, 2004; Goldenberg *et al.*, 2005). Because prematurity is one of the leading causes of infant mortality and morbidity, ranging from 50-70% in the United States (Hobel, 2004) and increasing in the last 30 years worldwide (Gracie *et al.*, 2012), the major goals of obstetric care are to reduce the occurrence of such condition as well as to increase the gestational age of infants for which preterm delivery is unavoidable (Hobel, 2004). There are three etiological types of PTD: spontaneous (idiopathic); PTD due to the premature rupture of the fetal membranes prior to the occurrence of uterine contractions; and PTD due to induction of labour for medical indications (for instance in cases of intrauterine growth restriction or preeclampsia) (Hobel, 2004). The risk factors for PTD include socioeconomic factors such as poor antenatal care or poor nutritional status, previous PTD, repeated 1st trimester abortions, previous 2nd trimester abortions, urinary tract infections, multiple gestation, polyhydramnios (excess of amniotic fluid), uterine anomalies and incompetent cervix (Hobel, 2004).

The diagnosis of PTD for patients with ruptured and intact membranes is performed when the first symptoms appear and it is based on the documented uterine contractions (4 per 20 minutes or 8 per 60 minutes), on the documented cervical change (cervix effacement of 80% or dilatation of 2 cm or more) and on AF leakage in the cases of ruptured membranes (Hobel, 2004; Graça, 2005b). About 20% of patients without ruptured membranes usually cease contractions after rest and adequate hydration. However, when such measures fail to cease contractions, tocolytic (anti-contraction or labour repressant) therapy should be implemented to prevent PTD, provided there is no contraindication to it (Hobel, 2004; Graça, 2005b). On the other hand, most patients with ruptured membranes begin labour spontaneously within several days (Goldenberg *et al.*, 2008).

After diagnosis of PTD, it is usual to initiate antibiotic and glucocorticoids therapy. The antibiotic therapy is used as a prophylactic measure in order to prevent the progression of silent infections and the risk of fetal infections. Glucocorticoids therapy, on the other hand, accelerates pulmonary maturation of the preterm fetus, therefore reducing the risk for respiratory syndrome distress (RDS) and also reducing the risk of intraventricular hemorrhage (IVH) after birth (Hobel, 2004; Graça, 2005b).

Several markers have been associated with PTD including various cytokines in periodontal (crevicular) fluid, cytokines and microorganisms such as Chlamydia found in urine samples, salivary estriol measurements, and a wide variety of microorganisms, cytokines, and proteins (such as ferritin, fibronectin or metalloproteinase-9) in AF, serum, plasma and cervicovaginal secretions (Goldenberg *et al.*, 2005). A systematic review of 30 of potential early biomarkers of PTD has shown that only 3 of these biomarkers (the proteome profile and prolactin in cervicovaginal fluid, and matrix metalloproteinase-8 in AF) had some predictive value for the prediction of PTD (Conde-Agudelo *et al.*, 2011). However, these biomarkers were evaluated in small studies, and therefore were not considered clinically useful for the prediction of PTD (Conde-Agudelo *et al.*, 2011). The presence of the fetal protein fibronectin in the cervix or vagina at 22-24 g.w. has been suggested as a powerful predictor of PTD (Goldenberg *et al.*, 2005), provided that it is sampled in women with intact membranes (Roman *et al.*, 2005). Despite the availability of several potential markers, an early predictive test is still not available for PTD.

Besides the work presented in this thesis, which will be presented and discussed in later chapters, metabolite variations have been sought in 2nd trimester AF in order to identify woman at risk for PTD from a large group of pregnant women comprising: a) those who delivered at term with intact membranes, and those who delivered preterm b) with and c) without intra-amniotic

infection (Romero *et al.*, 2010). Several metabolites were suggested as markers of PTD, therefore showing that metabolic profiling can be used to assess the risk of PTD (Romero *et al.*, 2010).

Premature rupture of the membranes (PROM) is defined as the spontaneous rupture of fetal membranes after 37 weeks of gestation before the occurrence of uterine contractions (Stuart *et al.*, 2005). PROM is a quite common condition, occurring in approximately 8-10% of all pregnancies (Table 1.1, page 2) and increases the risk of maternal/neonatal sepsis, as well as the need for neonatal resuscitation at delivery (Stuart *et al.*, 2005). Similarly to PTD, an early diagnosis of PROM is not available, being diagnosed when the first symptoms occur. It is therefore important to initiate the antibiotic therapeutic prophylactics as soon as diagnosed, in order to reduce the risk of infections and sepsis (Hobel, 2004; Graça, 2005c). However, there seems to be an association between PROM and fetal membrane weakening mediated by biochemical processes involving degradation and dissociation of amnion membranes collagen with the involvement of the proteolytic enzymes, matrix metalloproteinases and apoptosis events (Stuart *et al.*, 2005; Moore *et al.*, 2006). A closer investigation of such processes should be useful to understand the pathophysiology of PROM and ultimately aid the development of predictive diagnostic methods.

1.2.4. Intrauterine growth restriction (IUGR)

Intrauterine growth restriction (IUGR) is an important cause of perinatal mortality and morbidity and occurs in 7-15% of all pregnancies worldwide (Table 1.1, page 2). IUGR can be defined as giving rise to birth weights below the 10th percentile for a given gestational age (Figure 1.3).

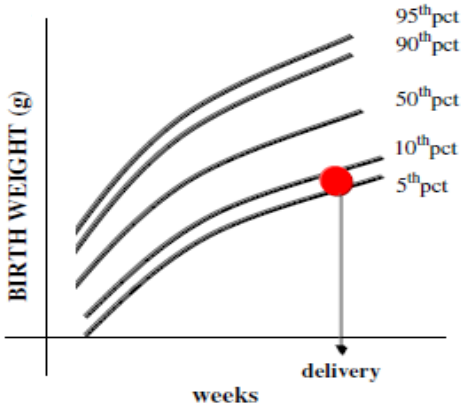


Figure 1.3 - Birth weight as a function of gestational weeks. Intrauterine growth restricted (IUGR) fetuses have birth weight below the 10th percentile (pct) (Adapted from Cetin and Alvino, 2009).

This definition is different from small for gestational age (SGA), which refers to small but anatomically normal fetus with an appropriate AF volume and growth rate, which means the outcome will usually be a normal, constitutionally small newborn (Hobel, 2004).

Two types of fetal IUGR have been described: symmetric, when growth of both fetal head and body is inadequate, which has been more commonly associated with FM and occurring as early as during the 2nd trimester; and asymmetric, when growth restriction occurs usually late in pregnancy, and head size is proportionally larger than the abdominal size (Hobel, 2004).

IUGR fetuses are prone to problems such as meconium (fetal feces) aspiration, asphyxia, polycythaemia, hypoglycaemia and mental retardation (Hobel, 2004). The risk factors, for the development of IUGR, result from the pre-existence of maternal diseases (autoimmune, endocrine, cardiovascular, hypertension, renal), pregnancy-induced hypertension (preeclampsia), alcohol or drug addition, smoking, acquired infections (cytomegalovirus, rubella, herpes simplex, malaria) and abnormal placental position. Moreover, FM can also lead to IUGR (Hobel, 2004; Cetin and Alvino, 2009). Also, IUGR is thought to be related to an abnormal placental phenotype and, as a consequence, to deficient placental metabolism and transport, thus modifying the normal supply of nutrients, which may lead to fetal undernutrition (Cetin and Alvino, 2009).

The diagnosis of IUGR relies solely on the identification of risk factors, the establishment of the correct fetal gestational age, the continuous assessment of fetal growth by fundal height (a measure of the size of the uterus, which may be used as first screening method) and by ultrasound (Hobel, 2004). In cases where IUGR is suspected, delivery can be performed at 34 weeks or later, provided that lung maturity is ensured. Also, regardless of the type of labour, the mother must be carefully monitored to detect the earliest sign of fetal distress. In order to avoid hypoglycaemia and respiratory distress in IUGR newborns, careful monitoring of blood glucose, and the respiratory function should be performed after birth (Hobel, 2004).

1.2.5. Preeclampsia (PE)

Preeclampsia (PE) is a multisystem disorder that is unique to pregnancy and is characterized by the new onset of hypertension and proteinuria in the latter half of pregnancy (Castro, 2004). It is associated with significant morbidity and mortality rates for the mother and the neonate, affecting about 2 to 6% of pregnancies (Table 1.1, page 2) (Walker, 2000; Farina *et al.*, 2011). Specifically, PE can cause several maternal disorders such as placental abruption, disseminated intravascular coagulation, renal failure, hepatic failure, permanent neurologic lesions and also

increased risk of essential hypertension after pregnancy. Fetal and neonatal complications include IUGR, prematurity and perinatal death (Castro, 2004).

PE is more prevalent in the first pregnancy (Raijmakers *et al.*, 2004), but it can also occur in women with previous pregnancies (multiparas), especially in those women with previous preeclampsia, twins, and in those with either pre-existing vascular disease or conditions associated with increased cardiovascular risk, including renal disease, hypertension, diabetes, thrombophilia, and obesity (Castro, 2004; Raijmakers *et al.*, 2004).

The diagnosis of PE follows two criteria: the development of hypertension (systolic blood pressures ≥140 mmHg or diastolic blood pressures ≥90 mmHg) after the 20th week of pregnancy in women with previously normal blood pressures; and the development of proteinuria, which may be defined as protein content >0.3 g in a 24 hour urine collection or an urinary protein/creatinine ratio 30 mg/mmol (Castro, 2004; Raijmakers *et al.*, 2004). When patients have liver dysfunction, thrombocytopenia, and haemolysis, they are classified as having HELLP syndrome (i.e., haemolysis, elevated liver enzymes, low platelets), which is a more severe form of PE (Raijmakers *et al.*, 2004).

Commonly, a distinction between early and late PE has been used since there is evidence that more severe disease is associated with early onset (prior to 34 g.w.) supporting the concept that the etiology of this disease may be different from that of later onset PE. Early onset disease appears to be mediated by the placenta, and is associated with abnormal uterine artery flow, IUGR, and adverse maternal and fetal outcomes. Later onset of the disease (after 34 g.w.) has been linked to maternal "constitutional" factors, such as body mass index (BMI), and may be associated with more favourable outcomes (Turner, 2010). Several metabolic factors have also been associated with PE, such as the metabolic syndrome (dyslipidaemia, hyperinsulinaemia, hyperglycaemia, hypertension and obesity), high maternal amino acid concentrations (Evans *et al.*, 2003; Turner *et al.*, 2008), altered maternal lipid and ketone bodies content (Turner *et al.*, 2007) and oxidative stress (Raijmakers *et al.*, 2004; Turner *et al.*, 2009).

Although PE appears suddenly in the 3rd trimester, the initial damage that triggers it, very likely occurs in the 1st trimester of pregnancy at the time when trophoblast cell differentiation/invasion initiates (Farina *et al.*, 2011). The identification of reliable screening markers would thus permit major improvements in obstetric care through better targeting and antepartum surveillance. Although a great number of possible predictive methods and markers have been investigated (Kenny *et al.*, 2010; Farina *et al.*, 2011; von Dadelszen *et al.*, 2011; Oh *et al.*, 2012), the results are diverse and some of them are conflicting and population dependent (Farina *et al.*, 2011). Delivery

is the only possible cure for PE, for which it is recommended preferably after 34 g.w. if there is no fetal compromise, or in more severe cases after a period of stabilization of the disease, regardless of the gestational age of the fetus (Castro, 2004).

1.2.6. Gestational Diabetes Mellitus (GDM)

Gestational diabetes mellitus (GDM) is defined as any degree of glucose intolerance with onset or first recognition during pregnancy (Metzger and Coustan, 1998), usually after 24-28 g.w. (ADA, 2004). The prevalence of GDM may range from 1 to 14% of all pregnancies (Table 1.1, page 2), depending on the population studied and the diagnostic tests employed (ADA, 2004).

GDM affected pregnancies have been associated to fetal macrossomia, delayed fetal organ maturity, FM, increased frequency of maternal hypertensive disorders, caesarean delivery (Gambone *et al.*, 2004), increased risk of post-natal maternal diabetes mellitus and, in more severe cases, an increased risk of intrauterine fetal death during the last 4–8 weeks of gestation (ADA, 2004). Several organ and vascular complications in pregnant women may then occur in GDM (Gambone *et al.*, 2004). Mothers with GDM also show evidence of altered ketogenesis and higher plasma concentrations of β-hydroxybutyrate (Tisi *et al.*, 2011) as well as altered placental amino acid exchange (Cetin *et al.*, 2005). Early perturbations of glucose and insulin has also been observed in women latter developing GDM (Star *et al.*, 1997; Tisi *et al.*, 2011).

GDM screening is generally performed between 24 and 28 weeks of gestation (ADA, 2004; Gambone *et al.*, 2004). However, women with clinical characteristics consistent with a high risk of GDM (marked obesity, personal history of GDM, glycosuria, or a strong family history of diabetes) should perform the screening as soon as feasible (ADA, 2004).

GDM screening can be performed by maternal serum or plasma glucose testing either fasting, or 1 hour after the oral intake of 50g of glucose (glucose challenge test or O'Sullivan's test) (Gambone *et al.*, 2004). If values indicative of glucose intolerance are obtained, that is glucose values >95 mg/dL or higher than ≥140 mg/dL for fasting and O'Sullivan's test, respectively, women are advised to undertake the 3 hours oral glucose tolerance test (OGTT). The OGTT consists in the measurement plasma glucose fasting and after 1, 2 and 3 hours after 100g oral glucose ingestion. The test has diagnostic value for GDM if at least two concentrations values exceed the threshold values shown in Table 1.5.

Strict metabolic control is needed after diagnosis of GDM to decrease the incidence of the adverse conditions described above. Therefore, an appropriate diet, insulin and exercise should

be offered to pregnant woman in order to achieve and maintain euglycaemia (Gambone *et al.*, 2004; Hod *et al.*, 2008).

Table 1.5 - Threshold concentration values for oral glucose tolerance test according to the American Diabetes Association (ADA, 2004).

Concentration	mg/dL	mmol/L
Fasting	95	5.3
1 h	180	10.0
2 h	155	8.6
3 h	140	7.8

1.3. Prenatal diagnosis research through biofluid composition

1.3.1. Chemical composition of human amniotic fluid

Amniotic fluid (AF) fills up the amniotic sac preventing contraction of the uterus, allowing space for the growth, movement and development of the fetus. It also has a protective role against potentially pathogenic bacteria (Toot and Lu, 2004). The formation of AF initiates in the 1st trimester from the free passage of molecules such as amino acids through fetal skin and from the coelomic fluid (CF), an ultrafiltrate of maternal plasma that fills the exocoelomic cavity (Figure 1.1, section 1.1, page 1) (Jauniaux and Gulbis, 2000). Gradually, AF volume increases mainly due to production of urine by the fetus, increasing from 30 mL at 10 g.w. to 350 mL at 20 g.w. and eventually reaching 1000 mL at 38 g.w. AF volume slowly declines until term mainly due to fetal swallowing (Lind, 1981; Graça and Barros, 2005).

The chemical composition of AF changes from 1st to 2nd trimester, particularly due to the keratinisation of the fetal skin which minimizes the free diffusion of small molecules and due to the development of fetal kidney filtration. Lung and tracheal fluid contribute to changes in AF composition the 3rd trimester due to an increase of lung surfactant contents such as lecithin (phosphatidylcholine) and surfactant proteins (Liu *et al.*, 1998b; Liu *et al.*, 2000). In result, several ions, metabolites and other small organic molecules can be found in AF, such as glucose, urea, creatine, creatinine, uric acid, amino acids, lactic acid, ketone bodies, lipids (triglycerides, phospholipids and cholesterol), fatty acids, long-chain alcohols, bile acids, bilirubin, vitamins, hormones and prostaglandins, whose compositions change with advancing gestational age (Lind, 1981; Sims *et al.*, 1996; Jauniaux *et al.*, 2005; Ottolenghi *et al.*, 2010). Some of the most important constituents of AF and their concentrations at the three trimesters of pregnancy are summarized in Table 1.6.

Table 1.6 - Concentrations of human AF constituents measured at the different gestational age ranges indicated inside brackets. [a] (Lind, 1981); [b] (Jauniaux *et al.*, 2005); [c](Groenen *et al.*, 2004), [d] (Ottolenghi *et al.*, 2010), [e] (Gao *et al.*, 2009). U, enzymatic units (μmol/min).

Compound	Concentrations (gestational weeks)		
	1st trimester	2nd trimester	3rd trimester
Electrolyte ions (mmol/L) [a]			
Sodium	134 (11-12)	134 (17-20)	131 (33-36)
Potassium	3.9 (11-12)	4.0 (17-20)	4.2 (33-36)
Chloride	103 (11-12)	111 (17-20)	108 (33-36)
Phosphate	0.99 (10-16)	-	0.44 (35-40)
Bicarbonate	16.7 (11-14)	16.5 (17-20)	15.7 (37-40)
Calcium	1.80 (11-12)	1.90 (17-20)	1.90 (33-36)
Magnesium	0.70 (11-12)	0.65 (17-20)	0.55 (33-36)
Amino Acids (μmol/L) [a]			
Alanine	401 (7-18)	312 (15-20)	143 (33-37)
α-aminobutyric acid	15 (7-18)	16 (15-20)	6 (33-37)
Arginine	15 (7-18)	43 (15-20)	20 (33-37)
Aspartic acid	8 (7-18)	17 (15-20)	6 (33-37)
Citruline	6 (7-18)	23 (15-20)	4 (33-37)
Cystine	89 (7-18)	63 (15-20)	57 (33-37)
Ethanolamine	-	20 (15-20)	13 (33-37)
Glutamic acid	149 (7-18)	227 (15-20)	76 (33-37)
Glutamine	50 (7-18)	186 (15-20)	120 (33-37)
Glycine	164 (7-18)	124 (15-20)	40 (33-37)
Histidine	102 (7-18)	76 (15-20)	40 (33-37)
Hydroxyproline	-	42 (15-20)	17 (33-37)
Isoleucine	47 (7-18)	45 (15-20)	13 (33-37)
Leucine	106 (7-18)	80 (15-20)	24 (33-37)
Lysine	319 (7-18)	239 (15-20)	150 (33-37)
Methionine	25 (7-18)	25 (15-20)	21 (33-37)
Methylhistidine	-	1 (15-20)	-
Ornithine	51 (7-18)	38 (15-20)	20 (33-37)
Phenylalanine	77 (7-18)	59 (15-20)	29 (33-37)
Proline	198 (7-18)	332 (15-20)	203 (33-37)
Serine	235 (7-18)	186 (15-20)	120 (33-37)
Threonine	-	186 (15-20)	120 (33-37)
Taurine	124 (7-18)	83 (15-20)	101 (33-37)
Tyrosine	64 (7-18)	51 (15-20)	21 (33-37)
Valine	200 (7-18)	194 (15-20)	69 (33-37)
Amino acids / nucleotide metabolites [a]			
Creatinine (μmol/L)	-	70 (15-22)	115 (33-34)
Uric acid (μmol/L)	-	250 (19)	547 (34)
Urea (mmol/L)	3.90 (10-16)	3.89 (19)	5.56 (35-40)

Table 1.6 - (*cont.*)

Compound	Concentrations (gestational weeks) 1st trimester	2nd trimester	3rd trimester
Sugars [a, b, c]			
Glucose (mmol/L)	3.0 (5-12)[b]	2.51 (19)[a]	2.25 (37)[a]
Myo-inositol (μmol/L)	100 (5-12)[b]	396 (15)[c]	137 (39)[c]
Mannose (μmol/L)	70 (5-12)[b]	-	-
Organic acids [a, d]			
Lactic acid (mmol/L)[a]	9.1 (10-16)	9.8 (17-20)	9.8 (33-36)
Citric acid (μmol/L)[a]	-	345 (19-24)	277 (33-36)
Succinic acid (μmol/L)[d]	8.4 (<15)	9.8 (15-18)	65.9 (29-36)
Pyruvic acid (μmol/L) [a]	-	-	19 (37-40)
Fumaric (μmol/L)[d]	6.3 (<15)	6.2 (15-18)	5.4 (29-36)
Malic acid (μmol/L)[d]	16.4 (<15)	16.9 (15-18)	13.3 (29-36)
β-hydroxybutyric acid (μmol/L)[d]	48.5 (<15)	40.9 (15-18)	45.8 (29-36)
α-oxoglutaric acid (μmol/L)[d]	2.5 (<15)	2.4 (15-18)	4.0 (29-36)
Pyroglutamic acid (μmol/L)[d]	33.9 (<15)	61.2 (15-18)	56.3 (29-36)
Methylmalonic acid (μmol/L)[d]	0.7 (<15)	0.7 (15-18)	0.6 (29-36)
Lipids [a]			
Total lipids (mg/L)	170 (8-24)		331 (35-39)
Cholesterol (mg/L)	196 (9-24)		57.4 (32-43)
Bile acids (mg/L)	-	-	0.5 (28-42)
Phosphatidylcholine (mg/L)	2.9 (12-14)	2.0 (16-23)	86.3 (36-39)
Sphingomyelin (mg/L)	4.2 (12)	5.0 (16-22)	21 (34)
Proteins [a, e]			
Total protein (g/L) [a]	3.3 (13-16)	5.4 (17-20)	2.5 (33-40)
Albumin (g/L)	-	4.41 (15)[e]	1.23 (36-40)[a]
Transferrin (mg/L)	-	330 (15)[e]	69 (31-42)[a]
IgG (mg/L) [a]	330 (11-15)	410 (16-20)	120 (36-40)
IgA (mg/L) [a]	28 (11-15)	32 (16-20)	29 (36-40)
α_1-antitrypsin (mg/L) [a]	-	290 (16-20)	132 (31-42)
β_2-microglobin (mg/L) [a]	-	200 (16-20)	6.3 (36)
α-fetoprotein (mg/L) [a]	15 (14)	16 (17)	1.3 (31-42)
Enzymes (U/L) [a]			
Lactate dehydrogenase	-	87 (19)	130 (33-42)
Alkaline phospatase	-	18.4 (19)	20.5 (37)
Bilirubin (μmol/L) [a]	-	2.2 (15)	3.9 (32-43)
Hormones [a]			
Estriol (μmol/L)	-	-	2.34
Cortisol (total) (nmol/L)	-	83 (20)	200 (35-40)
Chorionic gonadotropin (μg/L)	933-1139 (8-12)	589-783 (12-20)	42-57 (32-40)
Thyroxin (total) (nmol/L)	-	5.12 (15-19)	5.66 (35-42)

Regarding the ionic composition, initially, AF osmolarity is similar to that of maternal plasma, then decreasing to values lower than those of maternal plasma with increasing gestational age (Graça and Barros, 2005). Several electrolyte ions such as sodium, phosphate, bicarbonate and magnesium tend to decrease slightly with gestational age except, potassium, calcium which remain practically unchanged and chloride which increases slightly (Table 1.6). An increase in creatinine, uric acid and urea can also be observed during gestation (Table 1.6) reflecting kidney maturation and increased amino acid and nucleotide metabolisms (Lind, 1981; Sims *et al.*, 1996).

In terms of amino acids and sugars, during development, a general decrease in major amino acids (alanine, glutamate, glycine, histidine, isoleucine, leucine, lysine, phenylalanine, serine, threonine, tyrosine and valine) and glucose is observed, especially by the transition between the 2nd and 3rd trimester (Table 1.6), mostly due to increasing reabsorption by the kidney and increased protein synthesis (Lind, 1981). Another sugar, *myo*-inositol, tends to increase in concentration in AF as a result of enhanced biosynthesis from glucose and in order to participate in the biosynthesis of phosphatidylinositol (Groenen *et al.*, 2003).

With respect to organic acids, most of the major ones present in AF (lactic, citric, pyruvic, fumaric, malic, β-hydroxybutyric and α-oxoglutaric acids) do not vary significantly during normal pregnancy (Table 1.6), with the exception of succinic acid which shows an increase especially during the 3rd trimester and pyroglutamic acid which increases from 1st to 2nd trimester (Ottolenghi *et al.*, 2010).

Regarding lipids and proteins, the lipid content of AF generally increases during pregnancy, particularly during the 3rd trimester. Phospholipids are among the principle contributors to the total lipid content (Table 1.6). Their concentration increases sharply towards the end of gestation in parallel with lung maturation, since phospholipids are major constituents of human pulmonary surfactant (80%), from which phosphatidylcholine is the most abundant constituent (Liu *et al.*, 1998a). Protein content tends to increase from 1st to 2nd trimester primarily due to free diffusion through fetal skin and also due excretion through fetal urine, as occurs for example with α-fetoprotein and β_2-microglobulin (Lind, 1981). Protein content increases progressively until the 32nd week, decreasing afterwards until labour, mainly due to fetal swallowing (Lind, 1981). Following a similar trend is the heme degradation product bilirubin (Table 1.6). The enzymatic activities change during the course of pregnancy in a pattern roughly parallel to the protein concentration (Table 1.6). An exception is α-amylase, which may reach high levels during the 3rd trimester. Its most probable origin is from fetal urine and saliva (Lind, 1981).

Finally, several hormones, second messengers and prostaglandins can be found in AF (Table 1.6) most of them increasing in concentration during gestation (Lind, 1981). These include the non-steroid hormones thyroxin, chorionic gonadotropin, insulin, epinephrine and norepinephrine, the steroid hormones from which cortisols and estriols are the most abundant in AF, as well the second messenger cyclic adenosine monophosphate (cAMP) (Lind, 1981). Prostaglandins, from which the most abundant in AF are PGF_2 e $PGF_{2\alpha}$ rise sharply during labour due to their involvement in uterine contraction (Lind, 1981).

1.3.2. Human amniotic fluid as a diagnostic media

Amniotic fluid has long been considered the ideal media to assess fetal well-being and maturation. Indeed, collection of AF is essential for diagnosis of open tube defects and karyotype analysis, as described in previous sections 1.2.1 (page 4) and 1.2.2 (page 8), respectively. Moreover, AF analysis may be carried out in the diagnosis of infections (CMV, toxoplasmosis, rubella, etc.), to predict the severity of fetal haemolysis in red-cell alloimmunized pregnancies (determination of AF bilirubin levels) or in the assessment of fetal lung maturity through the measurement of phosphatidylcholine/sphigomyelin ratio at the 3rd trimester (Underwood *et al.*, 2005).

Extensive work has been conducted in the last three decades in AF composition, to find new markers of prenatal disorders, with great contributions from emerging analytical platforms, namely nuclear magnetic resonance (NMR) spectroscopy, infrared spectroscopy (IR), mass spectrometry (MS) and capillary electrophoresis (CE).

Analysis of AF by NMR spectroscopy was introduced in the early 80s, with first applications of low-field ^{1}H NMR *in-vivo* and *in-vitro* to diagnosis the presence of meconium (fetal faeces) in AF (Bene, 1980; Bene *et al.*, 1982; Borcard *et al.*, 1982). The potential of ^{1}H NMR spectroscopy to probe low molecular weight (M_w) molecules, especially clinically relevant metabolites, was later recognized (Nelson *et al.*, 1987). In this study 18 different metabolites were identified including several amino acids, lactate, glucose and creatinine using 60MHz and 360MHz spectrometers (Nelson *et al.*, 1987). Two subsequent studies applied ^{1}H NMR spectroscopy (at 300MHz) to quantify low M_w metabolites present in the mili- and micro-molar ranges, including creatinine, glucose, organic acids (acetate, citrate, and lactate) and several amino acids (alanine, histidine, leucine, phenylalanine, tyrosine and valine), using either standard addition procedures (McGowan *et al.*, 1993) or through the use of internal standards (Sims *et al.*, 1993b). Two-dimensional ^{1}H NMR spectroscopy was used for the first time to study AF (Sims *et al.*, 1993b). This thesis builds

on this type of knowledge, as it will be shown in chapter 4, in order to expand the knowledge of AF composition by NMR spectroscopy.

In terms of applications to the study of disorders, an early study using high field ^{1}H NMR spectroscopy (600 MHz) focused in PE, spina-bifida, trisomy 21 and GDM (Bock, 1994). Significant variations were noted in the levels of choline, succinate, acetate in PE and in the levels of lactate, glutamate, and acetate in spina-bifida cases (Bock, 1994). Other studies have found significant metabolite variations in AF due to the occurrence of disorders such as trisomy 21 (Sims *et al.*, 1993a), in cases of decreased digestive enzyme activities and cystic fibrosis (Le Moyec *et al.*, 1994), maternal insulin dependent diabetes mellitus (McGowan *et al.*, 1999) and spina-bifida (Groenen *et al.*, 2004). ^{1}H NMR spectroscopy was also used for a comparative study of AF composition between the 2nd and 3rd trimesters (Sims *et al.*, 1996; Cohn *et al.*, 2009) or to predict gestational age based on AF through quantitation of several metabolite signals (Cohn *et al.*, 2010).

The prediction of lung maturity has also been an important application of AF NMR spectroscopy *in-vitro*, particularly through the use of ^{31}P NMR spectroscopy to quantify phosphatidylcholine and sphingomyelin, which provided a reliable and straightforward alternative to the less reproducible method thin layer chromatography (TLC) commonly used to determine the phosphatidylcholine/sphingomyelin ratio (Pearce *et al.*, 1991; Pearce *et al.*, 1993). Spectral measurements of choline and related compounds by ^{1}H NMR were also attempted to study fetal lung maturation (Clifton *et al.*, 2006; Joe *et al.*, 2008), which resulted in studies of choline and related compounds *in vivo* using standard clinical magnetic resonance imaging (MRI) instruments (Fenton *et al.*, 1998; Clifton *et al.*, 2006; Kim *et al.*, 2008).

The use of infrared spectroscopy in AF analysis has focused in the use of near-infrared (NIR) and mid-infrared (MIR) methods as alternatives for the replacement of TLC for the measurement of phosphatidylcholine/sphingomyelin ratio (Kobayashi *et al.*, 1990; Liu *et al.*, 1997; Liu *et al.*, 1998b), and the fluorescence depolarization assays (Liu *et al.*, 2000) in the prediction of lung maturity. Other work has also explored the quantitation of glucose and lactate using MIR spectra to study preterm infection and fetal distress (Liu and Mantsch, 1999). More recently a NIR study of 2nd trimester AF has been reported, describing the development of preliminary predictive models for preterm birth, based on genetic algorithm analysis and showing the promise of NIR spectroscopy for early diagnosis (Power *et al.*, 2011).

The use MS coupled to a separation technique such as liquid (LC-MS) or gas chromatography (GC-MS) methods for 2nd trimester AF metabolic profiling has been introduced as an approach to identify patients at risk for preterm delivery (Romero *et al.*, 2010). More studies have been

carried out on 2nd trimester AF using both LC-MS and GC-MS methods in targeted approaches, for the prenatal diagnosis of several inborn errors of metabolism such as tyrosinaemia type I or propionic acidaemia (Kumps *et al.*, 2004), Smith-Lemli-Opitz syndrome (Griffiths *et al.*, 2008; Amaral *et al.*, 2010) and phenylketonuria (Amorini *et al.*, 2012).

Changes in the protein content of AF in connection with the occurrence of some disorders have also been the focus of several studies. Indeed, MS-based proteomics studies have revealed important changes in AF proteins in individuals affected by Turner syndrome (Mavrou *et al.*, 2008), Kleinfelter syndrome (Anagnostopoulos *et al.*, 2010), trisomies 18 and 21 (Wang *et al.*, 2009; Kolialexi *et al.*, 2011), congenital heart defects (Nath *et al.*, 2009), IUGR (Cecconi *et al.*, 2011), PTD (Buhimschi *et al.*, 2008; Bujold *et al.*, 2008; Romero *et al.*, 2008), PROM (Thadikkaran *et al.*, 2005; Wang *et al.*, 2011b) and PE (Joong Shin *et al.*, 2008; Oh *et al.*, 2012), which may provide important clues for biomarkers of such disorders (Horgan *et al.*, 2009).

With respect to capillary electrophoresis (CE), this technique has also been used in the characterization of 2nd trimester human AF, enabling the simultaneous detection of the major AF proteins (albumin, transferrin, and IgG) and also low M_w molecules such as uric acid (Stewart *et al.*, 2001; Gao *et al.*, 2008; Gao *et al.*, 2009). In one of such studies, authors have found an association with the increase in the 2nd trimester AF concentrations of transferrin and the later occurrence of PTD, and that the uric acid concentration is a good predictor of newborn birth weight (Gao *et al.*, 2008). CE has also been employed in the simultaneous analysis of AF amino acids and amino acid derivatives (Tuma *et al.*, 2006). More recent studies explored the use of CE electrophoretic profiles of 2nd trimester AF and chemometric methods for the early prediction of GDM and macrosomia (Boisvert *et al.*, 2012a; Boisvert *et al.*, 2012b).

1.3.3. Prenatal research using other biofluids

In spite of the great potential of AF for diagnosis and monitoring of pregnancy, amniocentesis still carries some risk of fetal loss and preterm birth. For this reason the search for novel less-invasive diagnosis methods has been an important field of activity, especially using biological materials obtained less invasively such as maternal blood and urine.

The analysis of circulating nucleic acids and cell free DNA from maternal plasma from the 2nd trimester have been tested for the prenatal diagnosis of genetic disorders (Ding *et al.*, 2004), the detection of trisomy 21 (Ghanta *et al.*, 2010; Bianchi *et al.*, 2012), trisomies 18 and 13 (Chen *et al.*, 2011; Bianchi *et al.*, 2012); and monosomy X (Turner syndrome) (Bianchi *et al.*, 2012). Cell free

fetal DNA was also found in measurable amounts in 3rd trimester maternal urine, however, its low concentration makes it inappropriate for non-invasive prenatal diagnosis (Majer *et al.*, 2007).

Specific metabolites and proteins from maternal serum and plasma have also been suggested as predictive markers of PE, IUGR and GDM. Recently, the maternal serum concentration of placental growth factor (PlGF) and the protein PAPP-A in the 1st trimester were evaluated as screening markers for early-onset PE and IUGR, thus revealing that only placental growth factor was a moderate predictor of PE (Vandenberghe *et al.*, 2011). Maternal blood samples obtained at 11–14 g.w. were analysed for 40 acylcarnitine species and 32 amino acids by LC-MS, of these only hydroxyhexanoylcarnitine, alanine, phenylalanine, and glutamate were able to predict PE (Odibo *et al.*, 2011). With respect to GDM, it has been found that maternal plasma insulin and adiponectin concentrations, when measured at 11 g.w. may be predictive of gestational diabetes (Georgiou *et al.*, 2008).

Maternal urine metabolite composition has also been explored. For instance, the amino acid composition of urine from preeclamptic women has been evaluated revealing disturbances in urinary amino acid excretion (Glew *et al.*, 2004). The increase in maternal urine inositol phosphoglycan P-type has recently been identified as a possible marker of PE (Scioscia *et al.*, 2007a; Paine *et al.*, 2010). The same metabolite was also found to be associated with GDM in another study, where an increased urinary excretion of inositol phosphoglycan P-type was detected in GDM cases and was positively correlated with blood glucose levels (Scioscia *et al.*, 2007b). Maternal urine measurements of steroid compounds were also found effective for detecting fetal steroid sulfatase deficiency in 2nd trimester pregnancies (Marcos *et al.*, 2009).

With respect to other fluids, cervicovaginal secretions and saliva analysis have found some use in the prediction of PTD. In fact, the high concentration of the fetal protein fibronectin detected at 22-24 g.w. may be predictive of PTD, whereas the salivary concentration of unconjugated-estriol levels seem to have predictive value for late PTD (near 36 g.w.) (Goldenberg *et al.*, 2005).

1.4. Metabonomics: principles and applications

1.4.1. Principles and general applications

Systems biology seeks to interpret of the complex interactions in an biological system using a holistic approach (Snoep and Westerhoff, 2005), and several disciplines have emerged from it namely, genomics (the study of genes), transcriptomics (the study of gene expression),

proteomics (the study of protein expression) and metabolomics (the nonbiased identification and quantification of all metabolites in a biologic system) (Fiehn, 2002; Ellis *et al.*, 2007).

The concept of metabonomics may be viewed as a successor of the term metabolomics, and can be defined as the quantitative measurement of the dynamic multi-parametric metabolic response of living systems to pathophysiological stimuli or genetic modification (Nicholson *et al.*, 1999). Both terms are often used interchangeably in the literature in association with terms such as "metabolic profiling" and "metabolic fingerprinting". The latter has also been defined as the global high-throughput analysis to provide sample classification from different biological status or origin, for instance to discriminate between disease/healthy subjects without the need for metabolite identification and quantification (Ellis *et al.*, 2007). In the context of systems biology, metabonomics has some advantages over genomics, transcriptomics or proteomics in that metabolites provide real biological endpoints, therefore providing markers for disease diagnosis or evaluation of beneficial or adverse drug effects (Lindon *et al.*, 2004).

Metabonomics involves the generation of metabolic databases based on tissue or biofluid samples (such as urine, blood, etc.), for control animals and humans, diseased patients, animals models used in drug safety testing, etc., allowing the simultaneous acquisition of multiple biochemical parameters on biological samples (Lindon *et al.*, 2004). This requires the use of analytical platforms that provide high-throughput, reproducibility and atom-specific molecular structural information. Such platforms are NMR spectroscopy and MS in tandem with liquid (LC-MS) or gas chromatography (GC-MS) (Lindon and Nicholson, 2008a). Nevertheless, vibrational spectroscopy (infrared and Raman spectroscopies) have been also used been used in an increasing number of metabonomics (metabolic fingerprinting) studies (Ellis and Goodacre, 2006). Vibrational spectroscopy methods presents the clear advantages of reproducibility, rapidity, low cost, and portability (Ellis and Goodacre, 2006). The interpretation of the complex metabolite information gathered in metabonomics experiments requires the use of multivariate statistical methods (Lindon and Nicholson, 2008a), which will be explained in detail in chapter 2.

Metabonomics has been used in several different contexts ranging from the study of drug toxicity (Azmi *et al.*, 2005; Coen *et al.*, 2007; Dong *et al.*, 2012), drug metabolism and development (Lindon *et al.*, 2006; Keun *et al.*, 2008), to evaluate the effects of nutritional intervention (Rezzi *et al.*, 2007) and in epidemiological studies such as the study of the human metabolic phenotype and its association with diet and blood pressure (Holmes *et al.*, 2008). Applications of metabonomics to the clinical research span several areas either using animal models or human subjects. Some of these applications include the study of diabetes (Griffin and

Vidal-Puig, 2008), diagnosis of coronary heart disease (Brindle *et al.*, 2002), ischemic heart failure (Kang *et al.*, 2011), several types of cancers including lung (An *et al.*, 2010; Carrola *et al.*, 2011), liver (Chen *et al.*, 2009) and kidney (Kind *et al.*, 2007); in the context of liver and kidney transplantation assessment (Duarte *et al.*, 2005; Chen *et al.*, 2012), autism (Yap *et al.*, 2010), glycogen storage disease (Duarte *et al.*, 2007), neurological diseases (Sinclair *et al.*, 2010), in the understanding of parasitic infections (Wang *et al.*, 2004; Garcia-Perez *et al.*, 2010) and also in neonatal screening (Atzori *et al.*, 2009; Dessì *et al.*, 2011).

1.4.2. Metabonomics in prenatal health research

Metabonomics studies involving prenatal diagnosis have been carried out using blood plasma, with strong emphasis on the study of PE (Kenny *et al.*, 2005; Turner *et al.*, 2007; Kenny *et al.*, 2008; Turner *et al.*, 2008; Kenny *et al.*, 2010; Heazell *et al.*, 2012). Indeed several metabolites were found by ^{1}H NMR to be altered in plasma leading to the observation of low lipid and ketone bodies and different aromatic amino acid profiles (Turner *et al.*, 2007; Turner *et al.*, 2008). Several metabolites detected by Gas Chromatography-Mass Spectrometry (GC-MS) (Kenny *et al.*, 2005) and Ultra-Performance Liquid Chromatography-Mass Spectrometry (UPLC-MS) were found to be associated with the onset of PE (Kenny *et al.*, 2008). In addition, an UPLC-MS study of maternal plasma (15 g.w.) allowed a multivariate predictive model to be built for PE based on 14 relevant metabolites (Kenny *et al.*, 2010). Moreover, a recent ^{1}H NMR metabonomics study has found metabolite differences in cord and maternal blood of very-low birth weight neonates (Tea *et al.*, 2012), while a study using UPLC-MS found variations in free fatty acids, glycerolipids, progesterone metabolites, sterol lipids, vitamin D metabolites, and sphingolipids associated with the occurrence of SGA, PTD and neonatal intensive care admission in 3rd trimester maternal plasma (Heazell *et al.*, 2012).

When compared to maternal blood plasma, metabonomics investigation on maternal urine is still underexplored. However, a recent report on 2nd trimester maternal urine and plasma ^{1}H NMR on FM, *pre*-diagnostic GDM, *pre*-PTD, *pre*-PROM and CD, revealed important metabolite variations, thus laying the ground for potential development of less-invasive diagnostic based on metabonomics approaches for such disorders (Diaz *et al.*, 2011).

Amniotic fluid is an obviously interesting media to study in a metabonomics perspective and this is the main theme of this thesis, where metabonomics studies of FM, CD, *pre*-PTD, *pre*-diagnostic GDM and *pre*-PROM (Graça *et al.*, 2010; Graça *et al.*, 2012; Graça *et al.*, 2013), as well as *pre*-PE and *pre*-IUGR, are described. Furthermore, once the changes in AF composition are

established, it was of interest to investigate if equivalent and complementary information could be obtained in a non-invasive way that is, using maternal urine (Graça *et al.*, 2012). This relationship between AF and urine was explored in chapter 7, using UPLC-MS.

A few metabonomics studies have also been performed using other biofluids and tissues. Indeed, metabolic profiling have been performed using cervicovaginal secretions, thus revealing several possible metabolite marker candidates of PTD which were still unassigned (Auray-Blais *et al.*, 2011). Placental tissues, namely on cultured placental explants, which have been used as models of hypoxia mimicking PE and SGA conditions have also been investigated (Dunn *et al.*, 2009; Horgan *et al.*, 2010; Heazell *et al.*, 2011).

1.5. Aims and outline of this work

The main goal of the work presented in this thesis was to study 2nd trimester AF composition from women appointed to amniocentesis using metabonomics approaches, in order to contribute further to the knowledge of the biochemistry of prenatal disorders and eventually to the development of complementary diagnostic methods for FM and CD, as well as predictive methods of disorders occurring later in pregnancy, namely PTD, PE, IUGR, GDM and PROM. Second trimester maternal urine samples from some of the AF donors were also analysed in order to establish a relationship between AF compositional changes and those measured non-invasively through maternal urine. This should contribute to the evaluation of fetal/maternal synergism, in relation to prenatal disorders. Hence, the work presented comprised the following specific aims:

1 - Characterization of 2nd trimester AF by ^{1}H NMR spectroscopy: a) establishment of a protocol for the metabolic profiling of AF; b) assessment of compositional changes in AF during sample storage, handling and analysis; c) establishment of a database of metabolites of AF;

2 - Characterization and prediction of pregnancy disorders by metabonomics of 2nd trimester AF using ^{1}NMR, UPLC-MS and MIR: a) for the study of diagnosed FM and CD conditions; b) for the study of conditions occurring later in pregnancy, namely PTD, PROM, PE, IUGR and GDM;

3 - Characterization of the effects of FM and *pre*-diagnostic GDM on 2nd trimester maternal urine and their relation to those measured in AF: a preliminary investigation.

2. Principles of the analytical methods employed

2.1. NMR spectroscopy

2.1.1. Principles of NMR spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy uses the magnetic properties of atomic nuclei to provide information on the structure and dynamics of molecules. Atomic nuclei possess an intrinsic angular momentum with both magnitude and direction, the nuclear spin, *I*, which may be zero, a positive integer or a half integer (*I* = 0, 1/2, 1, 3/2, 2,...) (Harwood and Claridge, 1999). However, only the positive integer and half integer nuclei are NMR active, such as ^{1}H, ^{13}C, ^{19}F, ^{31}P (*I*=1/2); ^{2}H, ^{14}N (*I*=1); or ^{17}O (*I* =5/2). On the other hand, ^{12}C and ^{16}O have I=0 therefore are not NMR active.

NMR active nuclei possess a net positive charge and motion (spinning), which produces a weak magnetic field, and therefore a magnetic moment, **μ**. When placed in an external magnetic field (**B_0**), the nuclei experience a torque which forces them into precession about the axis of the external field (Figure 2.1a).

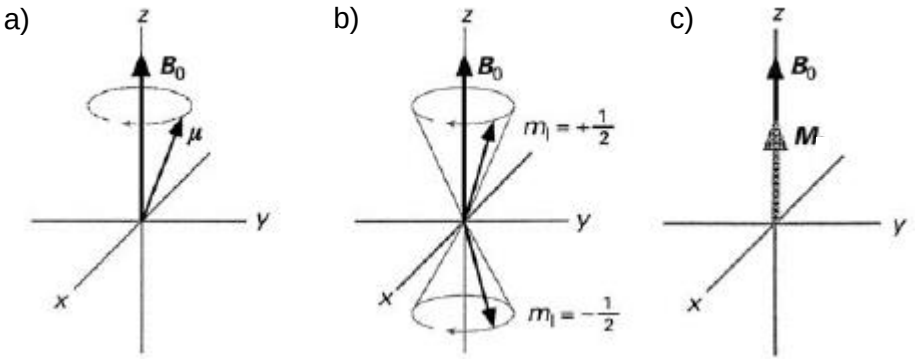


Figure 2.1 - a) The precession of magnetic moment **μ** in the presence of an external magnetic field **B_0**; b) the precession of magnetic moments in α (m_I =1/2) and β (m_I = - 1/2) spin states; c) the build-up of magnetization **M** due to the small excess of population of nuclei in the α spin state (Reproduced from Reynolds, 2000).

This motion is known as Larmor precession and it occurs at a frequency ω_0, the Larmor frequency, which is directly proportional to the strength of the applied magnetic field, **B_0**. The Larmor frequency (ω_0) is also dependent on the magnetogyric ratio of the nucleus, γ, and it can be given by the expression:

$$\omega_0 = \frac{\gamma \mathbf{B_0}}{2\pi} \qquad \text{[Eq. 2.1]}$$

In the presence of the external magnetic field, the nuclei can assume several possible orientations or energy levels, defined by the magnetic quantum number, m_I:

$$m_I = 2I + 1 \qquad \text{[Eq. 2.2]}$$

Hence, with *I*=1/2, two orientations are possible with two possible associated energy levels, parallel (lower energy, α) and anti-parallel (higher energy, β) to the field (Figure 2.1b). Therefore, energy transitions between levels involve a change in spin orientation.

The frequencies required to promote these transitions correspond to the Larmor frequency of the nucleus, which fall in the radiofrequency (*RF*) region of the spectrum and are expressed by Equation 2.1. However, at equilibrium the distribution of the spin states is defined by the Boltzman distribution as:

$$\frac{N_\alpha}{N_\beta} = e^{\Delta E/RT} \qquad \text{[Eq. 2.3]}$$

where *N* is number of nuclei in a spin state, ΔE, the energy of transition, *R*, the gas constant and *T*, the absolute temperature. According to this distribution, at equilibrium, there will be an excess of spins in the α state originating a bulk magnetization, **M**, parallel to the field $\mathbf{B_0}$ (Figure 2.1c).

In order to observe a signal, that is a resonance, the nuclei are irradiated with an *RF* pulse, usually of μs duration. In the case of *I*=1/2 nuclei, the *RF* pulse causes a change in the populations of the equilibrium spin states. Therefore, the magnetization vector, **M** (Figure 2.1c), shifts direction from the z axis to the rotating plane (*x*′-*y*′ plane), where the magnetization **M** starts precessing at the Larmor frequency (Figure 2.2a). Due to the change in the orientation of the magnetization, the *RF* pulse is named a 90° pulse (Keller, 1988).

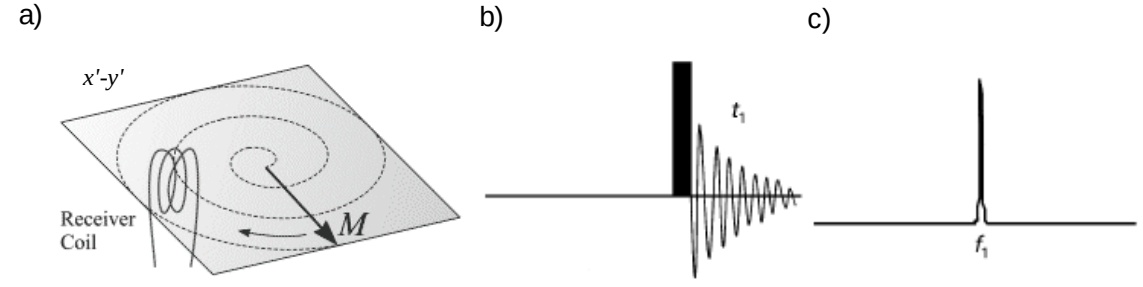


Figure 2.2 - The NMR experiment: a) After an *RF* pulse, **M** lies in the *x*′-*y*′ plane and precesses about the z-axis (Adapted from Keller, 1988); b) time domain free induction decay (FID) detected after the application of a *RF* pulse (solid bar) preceded by a preparation step and c) the same signal in the frequency domain after Fourier transformation of signal in (a) (Adapted from Rinaldi, 2000).

After the short *RF* pulse, the magnetization vector gradually returns back into its original orientation, a process known as spin relaxation and can be described by equations 2.4 and 2.5.

$$M_z(t) - M_0 \propto \mathrm{e}^{-t/T_1} \quad \text{[Eq. 2.4]}$$

$$M_{xy}(t) \propto \mathrm{e}^{-t/T_2} \quad \text{[Eq. 2.5]}$$

The spin relaxation is characterized by two time rate constants T_1 and T_2 describing the exponential decay of the longitudinal and transverse of magnetization components (Eq. 2.4 and 2.5), respectively (Claridge, 2009). The T_1 relaxation rate constant (Eq. 2.4) is related to the recovery of the z component of the magnetization (M_z) which involves the exchange of energy with the surroundings, being also known as spin-lattice relaxation (Claridge, 2009). T_2, also known as spin-spin relaxation is related to the loss of magnetization in the x'-y' plane components of the magnetization and differs from nuclei to nuclei due to the local inhomogeneity they experience inside a sample. T_2 is an important parameter in molecular dynamics since it is inversely related to molecular tumbling, and therefore to molecular size. Hence, smaller molecules are associated with higher T_2 relaxation constants (Claridge, 2009).

During relaxation, an oscillating current is created in the receiver coil of the NMR probe (Figure 2.2a) enabling the detection of a signal known as the free induction decay (FID) (Figure 2.2b). The signal intensity in the time domain, *f(t)*, is then transformed to the frequency domain, *f(ω)*, through Fourier transformation (FT):

$$f(\omega) = \int_{-\infty}^{+\infty} f(t)\, e^{-i\omega t}\mathrm{dt} \quad \text{[Eq. 2.6]}$$

resulting in the NMR spectrum (Figure 2.2c), a one-dimensional (1D), as explained better detail in the next section. However, according to the Heisenberg principle, there is an uncertainty in the spin energy levels, which means that in an NMR experiment the sample needs to be irradiated with a range of frequencies, rather than single frequency, during the 90° pulse (Harwood and Claridge, 1999).

Each nucleus is surrounded by electrons creating their own local magnetic field (B_{loc}), whose direction opposes that of the external magnetic field (B_0), creating a shielding effect in the nucleus. The effective field experience by the nucleus (B_{eff}) is then given:

$$\mathrm{B_{eff}} = \mathrm{B_0} + \mathrm{B_{loc}} \quad \text{[Eq. 2.7]}$$

In this way, each nuclei inside a molecule resonates at a different frequency (ν) depending on the chemical environment hence, each nuclei has a different position in the NMR spectrum.

Since the measured frequencies depend upon the strength of $\mathbf{B_0}$, the NMR spectrum scales are usually not reported as frequency (ν). Rather they are reported using a relative scale, the chemical shift, δ, a dimensionless property usually in the order of 10^{-6}, with units of parts per million (ppm).

The chemical shift is defined as:

$$\delta = \frac{\nu - \nu_{ref}}{\nu_{ref}} \times 10^6 \qquad \text{[Eq. 2.8]}$$

where ν_{ref} is the frequency of a nucleus in the $\mathbf{B_0}$, for instance 500MHz for ^{1}H or 125MHz for ^{13}C in a 11.7 Tesla magnet. The zero of the ppm scale is usually calibrated with the frequencies of the methyl groups bound to the silicon atom of the chemically inert compounds tetramethylsilane (TMS), dimethylsilapentanesulfonic acid (DSS) or trimethylsilyl propionate (TSP), which are used as ^{1}H and ^{13}C NMR chemical shift references. These are chosen specifically due to the shielding effect exerted by the silicon atom over the methyl groups, which results in lower observed frequencies. Figure 2.3 shows a series of ^{1}H and ^{13}C NMR chemical shifts ranges for various chemical environments (Atkins and de Paula, 2006).

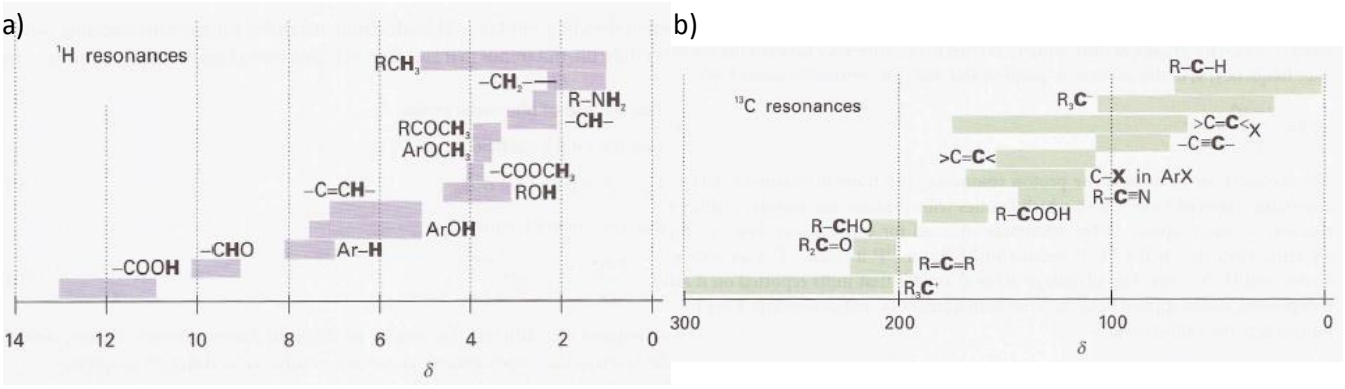


Figure 2.3 - The range of typical chemical shifts of a) ^{1}H resonances and b) ^{13}C resonances in several chemical groups (Adapted from Atkins and De Paula, 2006).

The more shielded ^{1}H or ^{13}C nuclei resonate at lower frequencies, therefore are situated at lower ppm values, whereas less shielded nuclei (higher frequencies) are situated at higher ppm values (Figure 2.3). It is this shielding effect that enables chemical group identification and ultimately molecular structure determination. In addition, NMR is inherently quantitative since the intensity of an NMR resonance is proportional to the number of respective nuclei producing this resonance, provided that quantitative conditions are employed during spectral acquisition. These conditions are achieved if a delay of duration of 5×T_1 of a isolated compound (or 5× the longest T_1 of a compound in a mixture), which is approximately the time needed for a I=1/2 nucleus to relax back to the equilibrium, is placed after each new *RF* pulse, thus enabling full relaxation of the nuclei under observation (Harwood and Claridge, 1999). A further effect relates to the interaction with neighbouring nuclei, known as spin-spin coupling, scalar coupling or *J*-coupling. The interaction between two coupling neighbouring spins originates new energy levels, hence transitions differing very slightly in energy (Harwood and Claridge, 1999). This energy

difference causes splitting of the resonances observed and it can be measured through the coupling constant, *J*, measured in Hz (Hore, 2000). This effect is illustrated in the ^{1}H NMR spectrum of ethanol shown in Figure 2.4.

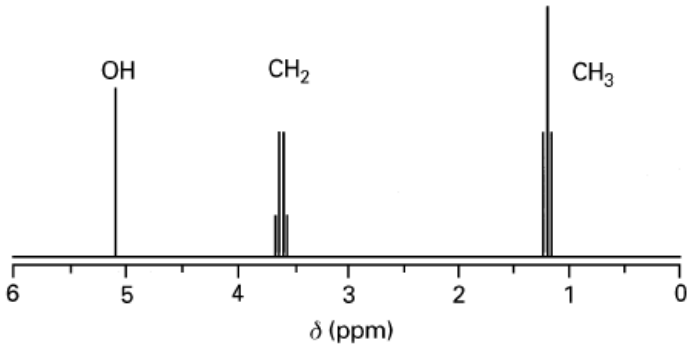


Figure 2.4 - ^{1}H NMR spectrum of liquid ethanol, C_2H_5OH. Splitting of the resonances at chemical shifts 1.2 and 3.6 corresponding to protons of CH_3 and CH_2 groups arises from spin-spin coupling of the two sets of protons. No splitting of OH group is observed due to the rapid intermolecular change of the hydroxyl proton (Reproduced from Hore, 2000).

2.1.2. One-dimensional (1D) and Two-dimensional (2D) NMR methods

One-dimensional (1D) NMR experiments, introduced in the previous section, are usually composed of a preparation step, where the population of nuclei are allowed to reach equilibrium and an acquisition step (t_1) immediately following the *RF* pulse or pulses. The resulting spectrum is obtained from the Fourier transformation (FT) of the free induction decay (FID) as shown previously in Figure 2.2b. There are 3 types of 1D ^{1}H NMR experiments which are particularly useful to study complex mixtures such as biofluids. One of such experiments, mentioned throughout this work as "standard" experiment, consist of an *RF* 90° pulse usually preceded by a presaturation pulse to saturate the intense water signal, therefore used to register the resonances from all compounds present in a mixture. The other two experiments, edited on the basis of molecular diffusion coefficients (diffusion-edited) and on T_2 relaxation times (such as the Carr-Purcell-Meiboom-Gill, CPMG spin-echo sequence, Meiboom and Gill, 1958) can be used to select only the contributions from macromolecules, or alternatively to select only the signals from the small molecules, respectively (Lindon and Nicholson, 2008a).

The resulting 1D NMR spectra can be difficult to interpret due to strong signal overlap, especially the spectra of complex mixtures such as biofluids which are composed of several hundreds of measurable resonances (Lindon *et al.*, 1999). Two-dimensional (2D) NMR spectroscopy methods are powerful tools to overcome spectral overlap, as the information is spread along two frequency dimensions, thus allowing a better identification of the individual

compounds resonances. In contrast with 1D, the 2D NMR experiments are composed of several steps: preparation, evolution (t_1), mixing and acquisition (t_2) (Rinaldi, 2000) (Figure 2.5).

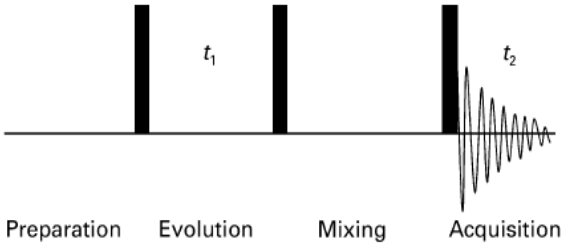


Figure 2.5 - An example of a 2D NMR experiment with its four basic steps represented. (Reproduced from Rinaldi, 2000).

Firstly, after the preparation step which allows the population of nuclei to reach equilibrium, one first 90° pulse brings the magnetization into the *x'*-*y'* plane as shown previously in Figure 2.2a. Then a series of pulses may be used to transfer the magnetization between nuclei of the same molecule during the evolution step, which is increased over time during the 2D experiment (Figure 2.5). After the evolution time (t_1) and depending on the experiment, additional pulses may be introduced in a step called mixing, in order to further manipulate magnetization before the actual acquisition of the NMR signal (t_2) (Figure 2.6a).

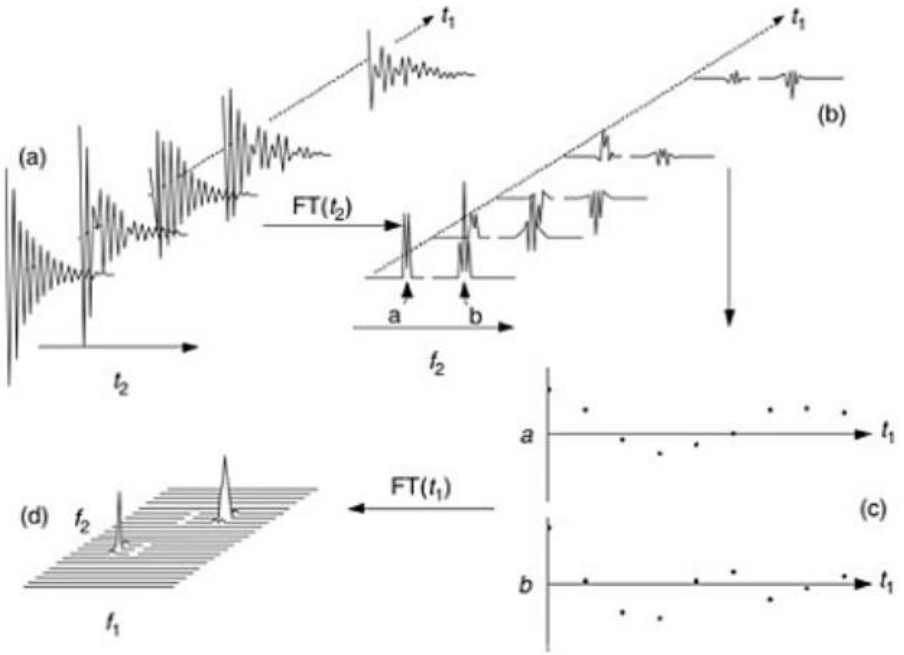


Figure 2.6 - Schematic illustration of the process used to produce a 2D NMR spectrum. (Reproduced from Rinaldi, 2000).

After acquisition, FT is performed in the t_2 acquired domain, originating a frequency domain (f_2) with different peaks intensities and phases according to t_1 (Figure 2.6b). A second FT is then applied with respect to the evolution time (t_1) (Figure 2.6c), originating the f_1 frequency domain (Figure 2.6d). The 2D spectrum is now a 3D plot with f_1 and f_2 frequency domains (1st and 2nd dimensions, respectively) and intensity represented as 3rd dimension (Figure 2.6d). However, 2D NMR spectra are more conveniently presented as 2D plots where intensity is represented through contour levels (Rinaldi, 2000).

The most commonly used 2D NMR experiments in complex mixture analysis in general and in metabonomics are the ^{1}H-^{1}H total correlation spectroscopy (TOCSY), ^{1}H-^{13}C heteronuclear single quantum correlation (HSQC) spectroscopy and ^{1}H *J*-resolved spectroscopy (*J*-resolved) (Beckonert *et al.*, 2007), which are described below.

TOCSY is the abbreviation for TOtal Correlation SpectroscopY, also known as HOHAHA (HOmonuclear HArtmann HAhn) (Hicks *et al.*, 1994). In the TOCSY pulse sequence, illustrated in Figure 2.7, after the evolution period t_1, the magnetization is spin-locked during the mixing time thus allowing its exchange through scalar coupling, a phenomena designated by isotropic mixing. This is achieved through the use of a composite pulse sequence, such as MLEV-17 sequence (Malcolm Levitt-17 sequence) which is a sequence of 16 composite 180° pulses, followed by a regular 180° pulse (Bax and Davis, 1985). During this spin-lock period, the magnetization behaves as a strongly coupled spin system enabling a coherence transfer between all coupled nuclei in a spin-system even if they are not directly coupled, thus enabling the assignment of the spin-system of a molecule.

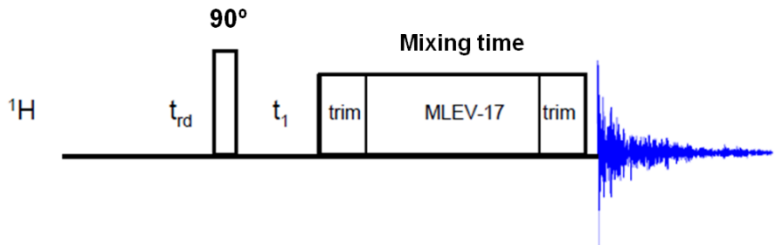


Figure 2.7 - ^{1}H-^{1}H TOCSY pulse sequence. Adapted from the Bruker Avance 1D/2D manual (Bruker, 2003).

The Heteronuclear Single Quantum Correlation (HSQC) experiment correlates protons with their directly attached heteronuclei (Bodenhausen and Ruben, 1980). The ^{1}H magnetization is detected (during t_2 - detection time) while the heteronuclei (such as ^{13}C) magnetization evolves during the evolution time - t_1 (Keeler, 2002), as shown in the pulse sequence in Figure 2.8.

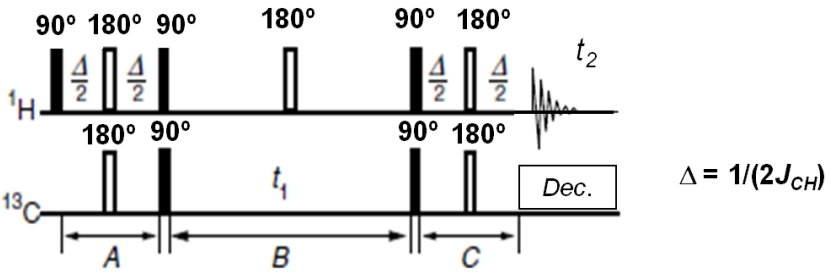


Figure 2.8 - ^{1}H-^{13}C HSQC pulse sequence depicting both ^{1}H and ^{13}C channels. A and C: INEPT sequences for ^{1}H-^{13}C magnetization exchange, B: evolution period. Acquisition takes place during t_2 with ^{13}C decoupling. J_{CH} is the ^{1}H-^{13}C single bond coupling constant which is about 150Hz. (Adapted from Keeler, 2002).

Magnetization transfer between the directly attached ^{1}H-^{13}C takes place through the use of an Insensitive Nuclei Enhancement by Polarization Transfer (INEPT) sequence, which is also used to recover the magnetization from ^{13}C to ^{1}H before acquisition (Figure 2.8). The 180° pulse in the mid evolution period removes the ^{1}H coupling and chemical shift effects (Figure 2.8). Because of the detection is performed on the high frequency nuclei (^{1}H), this sequence is much more sensitive than the sequences involving the direct detection of ^{13}C, such as the heteronuclear correlation (HETCOR) pulse sequence (Reynolds, 2000). The HSQC is usually applied to confirm the assignments obtained in 1D and 2D homonuclear correlation experiments (TOCSY). Another commonly used heteronuclear NMR experiment is the Heteronuclear Multiple Quantum Correlation (HMQC), which provides the same information as the HSQC. However, the latter offers better resolution in the heteronuclei dimension compared to HMQC, in which broadening of heteronuclei occurs due to homonuclear proton couplings (Claridge, 2009).

Finally, the ^{1}H *J*-resolved spectroscopy, also known as *J*-resolved (or just *J*-RES), involves the separation of the effects of chemical shift and *J*-coupling into separate dimensions (Ludwig and Viant, 2010). An example of a homonuclear 2D *J*-resolved pulse sequence is shown in Figure 2.9.

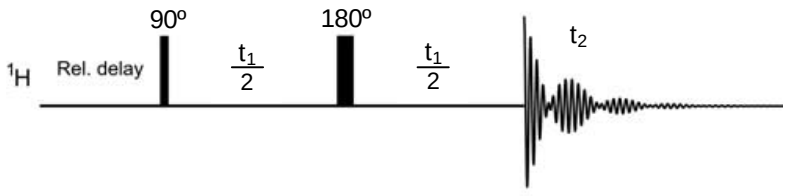


Figure 2.9 - Scheme for the 2D ^{1}H *J*-resolved basic pulse sequence. The increment time delay t_1 used to create the indirect time axis for the second dimension (t_2) is mediated by a 180° pulse. (Adapted from Ludwig and Viant, 2010).

After the 90° pulse to create magnetization, the chemical shift evolution created during the first $t_1/2$ delay is refocused by a 180° which then allows the observation of only the *J*-scalar couplings in the second dimension (Figure 2.9). This experiment allows the assignment of signals based in the measurement of *J*-scalar coupling. Additionally, the skyline projection of the 2D *J*-resolved produces a 1D-like spectrum without multiplicity information, i.e. all resonances are singlets. This representation which has been used in metabonomics and metabolite quantification studies, enables, in principle, the direct comparison of spectra acquired at different field strengths (such as 500 and 600 MHz), because of the absence of multiplicity (splitting) which is dependent of field strength (Ludwig and Viant, 2010).

2.1.3. Diffusion Ordered Spectroscopy (DOSY)

Molecular diffusion, the translational motion of the molecules in the liquid state or in solution, can be studied also by NMR, therefore providing additional information on the composition of complex mixtures such as biofluids (Smith *et al.*, 2007; Balayssac *et al.*, 2009). The measurement of diffusion by NMR spectroscopy is possible due to pulse field gradient techniques (Johnson, 1999). By the use of a gradient pulse applied in the same direction of the external magnetic field, $\mathbf{B_0}$, the molecules can be encoded according to their position in the sample (Kerssenbaum, 2006). Their motion after a certain diffusion time, Δ, can be decoded after a second gradient pulse. In this way, the measured NMR signal intensity is attenuated depending on the diffusion time (Δ), gradient strength (**G**) and duration (δ) (Kerssenbaum, 2006). The attenuated signal intensity is given by equation 2.9:

$$I = I_0 \, e^{-D \gamma^2 G^2 \delta^2 (\Delta - \frac{\delta}{3})} \qquad \text{[Eq. 2.9]}$$

where $\boldsymbol{I}$ is the observed (attenuated) intensity, $\boldsymbol{I_0}$ the unattenuated signal intensity, $\boldsymbol{D}$ the diffusion coefficient, γ the magnetogyric ratio of the observed nucleus, **G** the gradient strength, δ the gradient length and Δ the diffusion time (Kerssenbaum, 2006).

Pulsed field gradients are also used in the attenuation of signals from fast diffusing species in order to observe only signals from slower diffusion species, which is the basis for 1D NMR diffusion-edited experiments mentioned earlier in the previous section (Kerssenbaum, 2006). An example of a diffusion-type NMR experiment pulse sequence using two bipolar gradients (Wu *et al.*, 1995), i.e. two gradient pulses with opposite orientations, is illustrated in Figure 2.10.

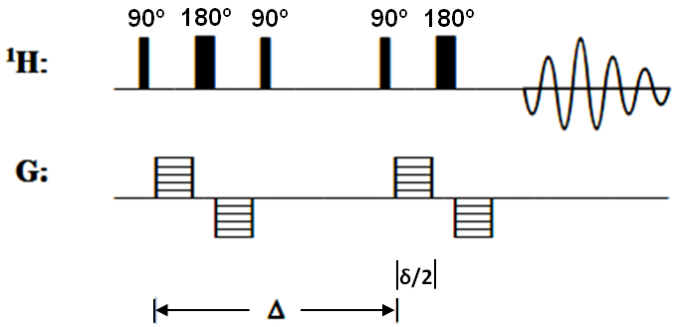


Figure 2.10 - An example of a diffusion-edited pulse sequence using bipolar gradients. Δ, diffusion time; δ, gradient length. Adapted from Kerssenbaum, 2006.

In diffusion ordered spectroscopy (DOSY), a series of 1D NMR spectra are obtained with increasing gradient strength (**G**) values. The resulting spectral intensities experience an exponential decay, as shown previously in equation 2.9 and exemplified in Figure 2.11a for a mixture of standard compounds. By fitting each measured signal intensity to an exponential function, it is possible to determine each respective diffusion coefficient and obtain a pseudo-2D representation of chemical shift *versus* diffusion coefficient as shown in Figure 2.11b.

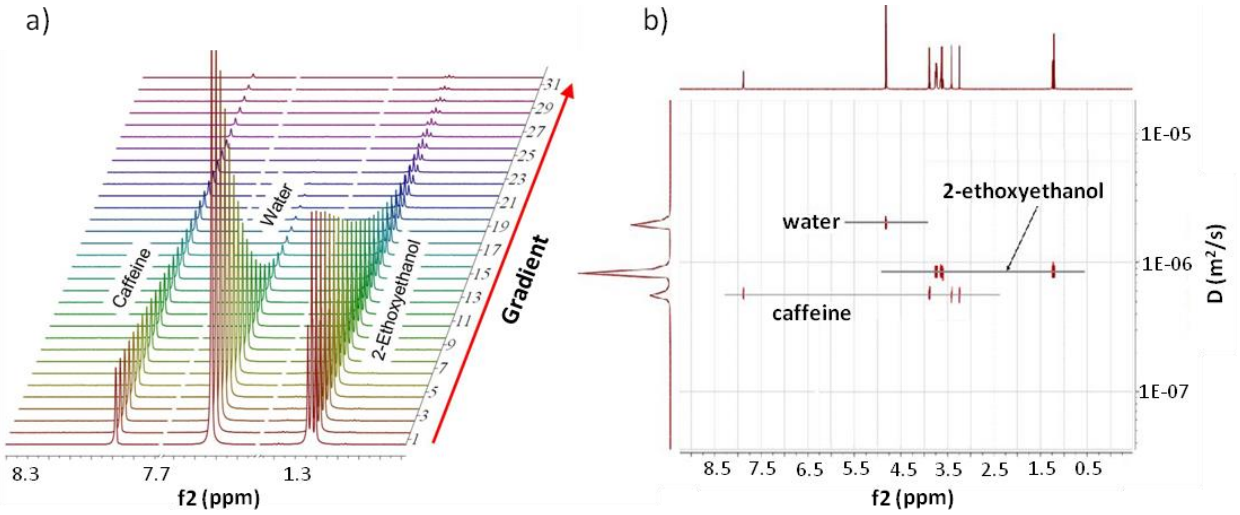


Figure 2.11 - DOSY of a standard mixture of caffeine, water and 2-ethoxyethanol. a) signals decay as function of gradient strength (**G**); b) pseudo-2D DOSY spectra. (Adapted from Mestrelab Research website: http://mestrelab.com/).

The information provided by the DOSY spectra is also very useful in the estimation of the molecular weight (M_w) of molecules. By assuming a globular shape for the molecules with fixed density, the M_w of a molecule is proportional to the inverse of the cube of the diffusion coefficient (Nilsson *et al.*, 2004). This relation enables the estimation of the M_w of an unknown compound from the resolved spin-systems in a DOSY spectra with the aid of a suitable M_w calibrant compound (Nilsson *et al.*, 2004).

2.1.4. Saturation Transfer Difference (STD) experiments

STD NMR experiments are generally used to characterize ligand-receptor complexes (Mayer and Meyer, 1999; Viegas *et al.*, 2011) and are useful to characterize specific interactions between high and low M_w molecules in complex mixtures. In an STD experiment, the resonances from a receptor macromolecule, for instance a protein bound to a small ligand, can be selectively saturated with an *RF* pulse. This saturation is spread over the entire protein by intramolecular saturation transfer (spin diffusion). The ligand molecules interacting with the receptor macromolecule are saturated by intermolecular saturation transfer. Through chemical exchange these saturated ligands are transferred into solution, provided there is a weak binding between ligand and receptor (dissociation constants, K_D, ranging from 10^{-8} to 10^{-3} mol/L) (Mayer and Meyer, 1999). By performing the difference between the 1D spectrum acquired with the selective saturation set to a frequency other than those of the ligand or the macromolecule (off-resonance) and the spectrum acquired with the selective saturation set only to one of the macromolecule resonances (on-resonance), positive signals from the ligand bound to the receptor will show up indicating that binding occurred (Figure 2.12).

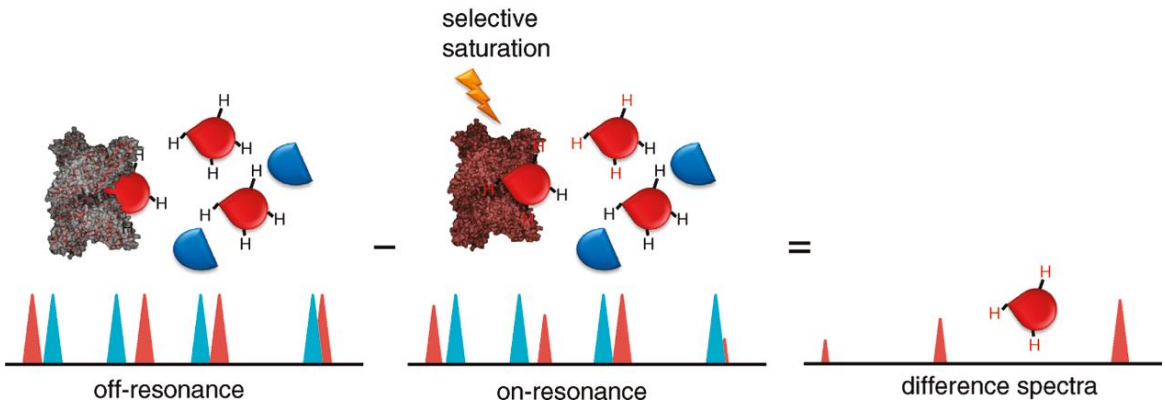


Figure 2.12 - Scheme of the STD-NMR experiment. The exchange between free and bound ligand allows intermolecular transfer of magnetization from the macromolecule to the bound small molecule (Reproduced from Viegas *et al*. 2011).

Because the chemical groups from the ligand in close contact with the macromolecule receive more saturation, the difference spectrum contains only the resonances from the ligand chemical groups which effectively bind to the macromolecule, thus aiding the mapping of the ligand binding sites (Mayer and Meyer, 1999).

2.1.5. Hyphenated NMR methods

The introduction of a chromatographic step before NMR analysis as in liquid chromatography-NMR (LC-NMR) enables further resolution of highly overlapped signals in NMR, such as those from higher abundance metabolites over the less abundant ones.

The LC-NMR setup consists of a computer operated high-performance liquid chromatography (HPLC) system connected in parallel with to an NMR system, as shown in Figure 2.13.

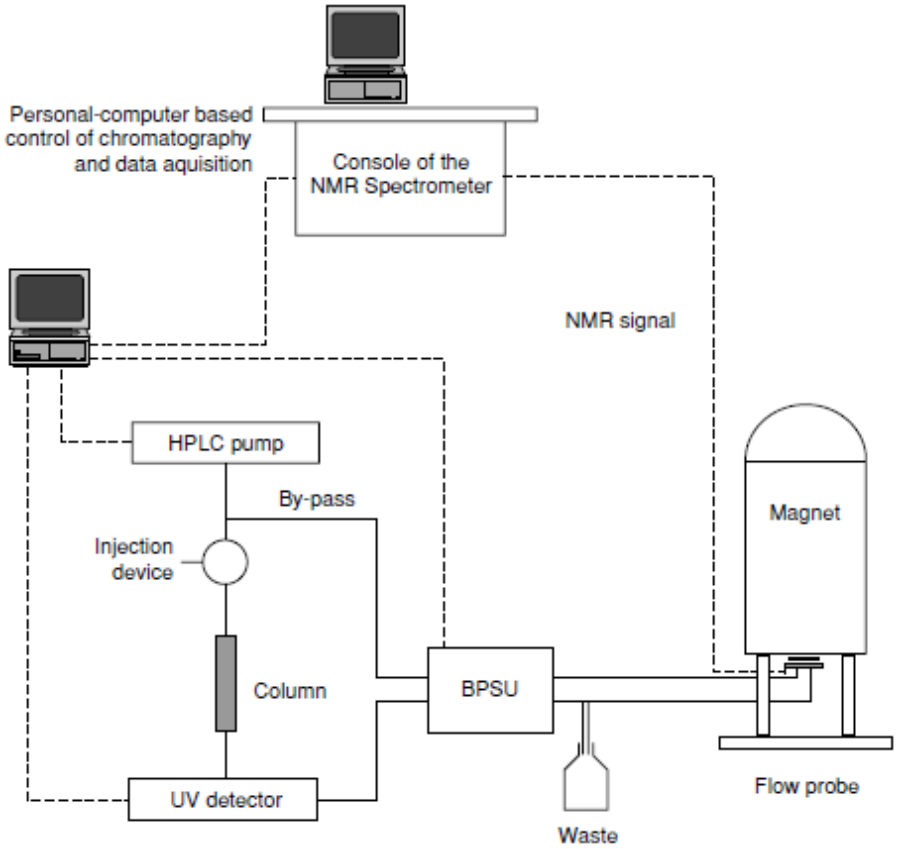


Figure 2.13 - Scheme of an experimental LC-NMR setup. BPSU, Bruker peak sampler unit; (—) capillary connections; (---) electronic connections (Reproduced from Albert, 2002).

The eluting fractions from the HPLC column pass through an UV detector, and are then directed to an NMR flow probe via an LC-NMR interface sample unit (Figure 2.13). The HPLC fractions can be either stored prior to NMR analysis (fraction storage or loop storage) or analysed on-line either in a continuous HPLC flow (on-flow experiment) or by interrupting the HPLC flow for NMR acquisition (stop-flow or time-slice experiment) (Albert, 2002). After NMR analysis, the HPLC fractions can be collected for further analysis or discarded (Figure 2.13).

Hyphenated NMR can be further extended by splitting the outlet of the HPLC column to a mass spectrometer (MS, discussed in the next section), in a LC-NMR/MS combination. The addition of a

parallel mass spectrometer enables a more complete structural characterization of a complex mixture (Lindon *et al.*, 2000b). In fact, hyphenated NMR techniques, namely LC-NMR and LC-NMR/MS, have been applied to the analysis of several types of biological matrices such as beverages and liquid food samples (Duarte *et al.*, 2003; Gil *et al.*, 2003), in phytochemistry (Wolfender *et al.*, 2005), natural products research (Jaroszewski, 2005) and in pharmaceutical research and drug discovery (Lindon *et al.*, 2000b; Corcoran and Spraul, 2003). The techniques have also been used in biofluid analysis, in the characterization of plasma lipoproteins (Daykin *et al.*, 2001), of polar metabolites in urine (Dear *et al.*, 2000; Godejohann, 2007) or in the identification of bile metabolites (Duarte *et al.*, 2009).

The introduction of cryo-probe technology and with the continuous increase in magnet field strengths, has led to significant sensitivity enhancements in hyphenated NMR, enabling the detection of metabolites at sub-micromolar concentration in complex matrices such as urine (Spraul *et al.*, 2003).

2.2. Liquid chromatography coupled to mass spectrometry (LC-MS)

In mass spectrometry (MS), the molecules from a sample are ionized and subsequently accelerated in an electric field under high vacuum, thus enabling the measurement of their mass-to-charge ratio (m/z) (Harwood and Claridge, 1999). One of the greatest strengths of MS is its high sensitivity compared to other spectroscopic methods used in complex mixtures analysis, such as NMR. MS is able to detect compounds down to picomolar and also has a greater dynamic range in terms of compounds concentrations to be detected (Lindon *et al.*, 2000b). The analytical capability of MS is usually extended by coupling it to a separation technique such LC, GC or CE. Biofluids contain hundreds or thousands of molecules in solution which makes the LC-MS the most suited hyphenated MS method for their comprehensive characterization. However, GC-MS has proven very useful in analysis of the more volatile compounds of complex biofluids such as urine (Rocha *et al.*, 2012).

The LC-MS setup is composed of an HPLC whose effluent is connected directly to the mass spectrometer source, usually an electrospray ionization (ESI) source, as shown in Figure 2.14a. The ESI ionization enables molecules to be taken directly from solution into the gas phase in an ionized state by passing the solution through the exit of a fine charged steel capillary, held at a fixed charged potential (Figure 2.14a). Positive or negative potential can be placed on the capillary, therefore enabling the production of singly or doubly charged positive and negative ions, respectively (Harwood and Claridge, 1999).

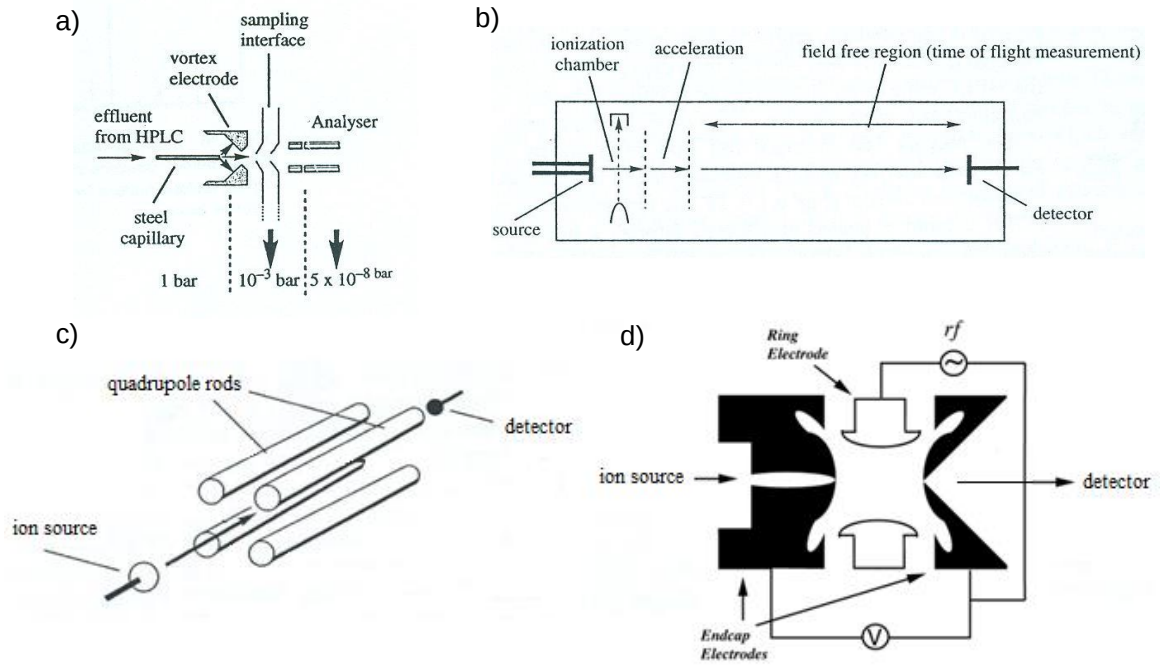


Figure 2.14 - Details of mass spectrometer components: a) an electrospray ionization (ESI) source; b) time-of-flight TOF mass analyser; c) quadrupole mass analyser (Adapted from Harwood and Claridge, 1999); d) ion trap mass analyser (Adapted from http://www.matrixscience.com/).

The ions produced, are conducted to a mass analyser such as the time-of-flight (TOF) mass analyser (Figure 2.14b). In a TOF mass analyser, the charged accelerated ions, subjected to the same potential difference are allowed to travel through a field free region between the ion accelerator and the detector (Figure 2.14b). The lighter ions will have a shorter time-of-flight compared to heavier ones, therefore registering lower m/z values (Harwood and Claridge, 1999). Commonly, TOF mass analysers also possess a reflectron mediating the accelerator and the detector, which functions as an ion mirror, thus increasing the resolution especially in the low m/z values. The TOF mass spectrometers are commonly used in metabolic profiling due to their high resolution and fast scanning capabilities (Want *et al.*, 2006). The quadrupole mass analysers (Figure 2.14c), another type of analyser, are commonly used in tandem with TOF mass analysers in the so called quadrupole-TOF (Q-TOF) instruments (Want *et al.*, 2006). In the quadrupole, the ions cross a path parallel to four voltage carrying rods (Figure 2.14c). Under the application of a specific *RF* and voltage current, only ions of a particular m/z will be collected and analysed (Figure 2.14c). In Q-TOF instruments, the ions travelling from the quadrupole analyser pass through a collision chamber. The introduction of a collision inert gas (He or N_2) in the collision chamber can be used to perform mass fragmentation (MS/MS) experiments which are very important for structure determination. The resulting mass fragments are then analysed in the TOF analyser.

Another commonly used type of mass analyser is the ion trap (Figure 2.14d), which can be used for both MS scanning and MS/MS studies. However, the ion traps have a poorer sensitivity, dynamic range, mass accuracy and resolution when compared with triple quadrupole and TOF mass analysers (Want *et al.*, 2006).

The analysis of complex mixtures by LC-MS comes at a cost of speed due to the introduction of a chromatographic separation step, which in HPLC can take from 20 to 60 min per sample, depending on the method and column used. An alternative approach to the traditional HPLC one is ultra-performance liquid chromatography (UPLC), which uses columns with much smaller particle size packing material (1.4-1.7 μm) than HPLC columns, thus allowing for improved separation and higher resolution (Wilson *et al.*, 2005a; Wilson *et al.*, 2005b) (Figure 2.15).

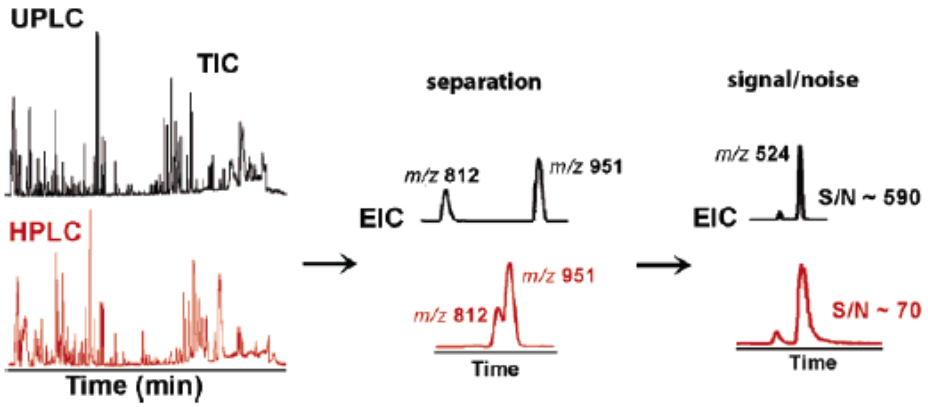


Figure 2.15 - Comparison between UPLC-MS and HPLC-MS chromatograms (Reproduced from Want *et al.*, 2006).

UPLC technology enables pumping and injection of liquids at pressures exceeding 10000 psi. Therefore sample analysis times can be reduced to as little as 5 min or less, resulting in much higher throughput (Want *et al.*, 2006). UPLC also allows narrower chromatographic peaks to be obtained (peak widths at half-height <1 s), resulting in increased peak capacity, lower ion suppression and improved signal-to-noise ratio, and thus increased sensitivity (Figure 2.15). In terms of sensitivity, UPLC can also detect more compounds than HPLC, with a 20% increase reported over the same chromatographic length and also enabling superior retention time reproducibility and signal-to-noise ratios over HPLC (Want *et al.*, 2006). Although both methods have been extensively used, UPLC-MS is increasingly becoming the method of choice for LC-MS-based metabolic profiling studies (Wilson *et al.*, 2005a; Want *et al.*, 2010).

2.3. Fourier Transform Infrared (FT-IR) Spectroscopy

The principle behind infrared (IR) spectroscopy lies in the ability to excite the molecular vibrations of chemical bonds with radiation in the infrared frequency region of the electromagnetic spectrum, provided that the electric dipole moment of the molecule changes during the vibration (Atkins and de Paula, 2006). The frequency of single bond (C-H, for instance) vibrations can be predicted by Hooke's law as:

$$\nu = \frac{1}{2\pi}\sqrt{\frac{k}{\mu}} \qquad \text{[Eq. 2.10]}$$

where, ν is the frequency of vibration, k is the bond resistance to vibration (or force constant) and μ is the reduced mass of a two atom system, that is for two atoms with masses m_1 and m_2, $\mu = (m_1 m_2)/(m_1+m_2)$ (Harwood and Claridge, 1999). However, transitions in IR are quantized, and non-linear molecules with n constituent atoms possess $(3n-6)$ fundamental vibrational modes, some of which may be degenerate (Harwood and Claridge, 1999).

The IR region of the electromagnetic spectrum is usually expressed as a function of the reciprocal of wavelength, the wavenumber (cm^{-1}), and can be divided into far-IR (FIR) 400-10 cm^{-1}, mid IR (MIR) between 200 and 4000 cm^{-1} and near-IR (NIR) between 14000-4000 cm^{-1}. The absorption of NIR radiation is commonly associated with the excitation of low-lying electronic excited states and overtones or combinations of molecular vibrations (Andrews, 2000). The FIR region relates mostly to low-frequency molecular vibrations and inversions, or rotations in small molecules (Andrews, 2000). In the MIR region, mostly fundamental molecular vibrations occur making this region the richest in chemical information (Spragg, 2000).

The two main molecular vibration modes active in the MIR range are the stretching (ν) and bending (δ) modes, corresponding to higher and lower energy transitions, respectively, and whose frequencies of vibration depend on the chemical functional groups and on the chemical bonds involved (Harwood and Claridge, 1999). The MIR spectra is most commonly acquired and represented as the percentage of transmittance (%T) as a function of wavenumbers, where the vibration bands are represented as dips (Harwood and Claridge, 1999). However, for quantitative and semi-quantitative applications, spectra are acquired in the absorbance mode, due to the linear relation between absorbance and analyte concentration predicted by Beer's law (Harwood and Claridge, 1999).

MIR spectroscopy covers a wide range of applications, from organic chemistry, through food science, forensics or quality control in industrial contexts. The MIR region has also been of great

interest in metabolic profiling (Ellis and Goodacre, 2006; Khaustova *et al.*, 2009), therefore it will be the focus of the remainder of this section.

Figure 2.16a shows a scheme of the optics system of a modern MIR spectrometer. Its mode of operation is based on the Michelson interferometer (Figure 2.16b).

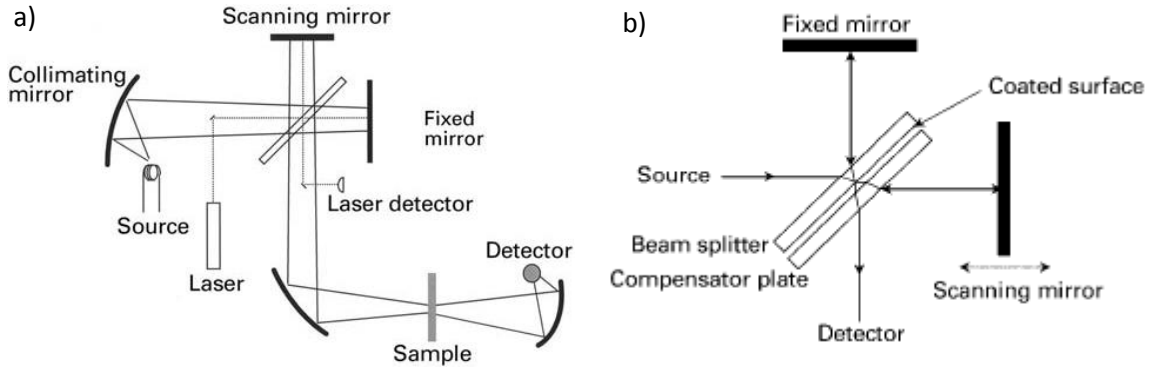


Figure 2.16 - Optics of an FT-MIR spectrometer. a) general optical scheme; b) the Michelson interferometer (Reproduced from Spragg, 2000a).

The single beam of IR radiation coming from the source falls on a beam splitter that divides it into two beams with approximately equal intensities (Figure 2.16a). These two beams strike mirrors that return them to the beam splitter where they recombine, creating interference, and then proceed to a detector, after passing through the sample (Figure 2.16a). Moving one of the mirrors creates a varying optical path difference between the two beams (Figure 2.16b). As the path difference changes, the two beams move in and out of phase with each other, so that there is alternating constructive and destructive interference (Spragg, 2000). Monochromatic radiation produces a sinusoidal signal, the frequency of which is directly proportional to the wavenumber. The signal from a broadband source consists of a combination of overlapping sinusoidal signals is called an interferogram. The interferogram is then converted into the spectrum of intensities at different wavenumbers by FT (Eq. 2.6, page 29) (Spragg, 2000). The samples to be analysed are placed in between a radiation path originating from the interferometer. Liquid samples can be placed in non-opaque cells, such as special sodium chloride cells, whereas solid samples are usually mixed with dispersing agents such as potassium bromide or the commercial mixture Nujol® (Harwood and Claridge, 1999). Alternatively, the MIR spectrometers can be equipped with an attenuated total reflectance (ATR) apparatus (Figure 2.17). The ATR is composed of a crystal (usually a ZnSe crystal), on top of which the liquid or solid sample is placed. The great advantage of the ATR is that the sample placed in the radiation path prior to spectra collection without the need of dispersing agents or special cells.

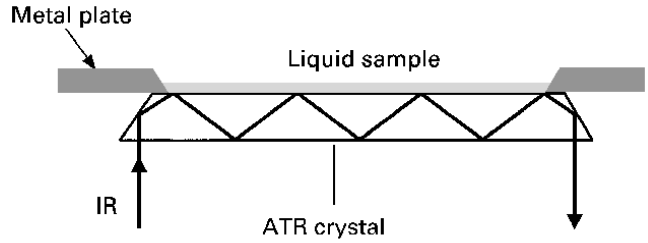


Figure 2.17 - A scheme of an Attenuated Total Reflectance (ATR) apparatus (Reproduced from Spragg, 2000b).

The radiation coming from the interferometer is thus reflected along the ATR crystal, passing through the sample several times, being then conducted to the detector (Figure 2.17), which allows an increase in sensitivity of detection.

In the MIR spectrum, the region of 4000-1500 cm^{-1} contains most of the interpretable bond stretching absorptions whereas the 1500-600 cm^{-1} region is usually referred to as the fingerprint region, because the rigorous assignments of the vibrations occurring in this region are not easily obtained (Harwood and Claridge, 1999). Table 2.1 summarizes some of the most intense MIR active vibrations for molecules commonly found in biofluids.

Table 2.1 - Mid-infrared stretching (ν) and bending (δ) vibrations of chemical functional groups from biofluid components. (Pretsch *et al.*, 1989; Harwood and Claridge, 1999; Liu and Mantsch, 1999).

Regions / bands (cm^{-1})	Vibrations	Molecules / assignments
4000-1500 cm^{-1} region		
3600	ν O-H	Free, non H-bonded (sharp band), water and proteins
3600-3000	ν O-H (broad)	H-bonded (broad band), water and proteins
3300	ν N-H (broad)	Amino acids and N-containing molecules
3000-2800	ν C-H	Sugars, proteins, glucose, amino acids and lipids
1750	ν C=O	Carboxylate groups of amino acids, organic acids, fatty acid esters (lipids)
1650	Amide I (ν C=O and δ C-NH)	Proteins
1600	ν_{as} COO^-, (asymmetric)	Amino acids and other organic acids
1560-1510	Amide II (δ N-H and ν C-N)	Proteins
1500-600 cm^{-1} (fingerprint) region		
1400	ν_{sy} COO^- (symmetric)	Amino acids and other organic acids
1230-1260	Amide III (very weak)	Proteins
1200-900	ν C-O and ν C-C	Sugars, proteins, glucose, amino acids and lipids

2.4. Data pre-processing and multivariate analysis (MVA)

Metabonomics data sets contain hundreds or thousands of variables thus requiring the appropriate statistical methods in order to extract meaningful biological and biochemical information. Conventional univariate statistical for instance Student's *t*-test, Mann-Whitney test or Analysis of Variance (ANOVA), which consist of separate analysis of each variable, can be unpractical for metabonomics analysis. Multivariate analysis (MVA), on the other hand, has the advantages of taking all the variables into account in one single analysis and, more importantly, allows the construction of predictive models. Therefore, MVA provides the appropriate tools for metabonomics data analysis.

2.4.1. Data pre-processing

The data sets gathered in metabonomics usually translate into two-dimension data tables of intensities, each line corresponding to a different observation or object (different sample, time-point, etc.) whereas each column represents a variable, for instance a chemical shift, wavenumber or MS feature (m/z, retention time pair) (Figure 2.18).

	variable $_1$	**variable $_2$**	**...**	**variable $_k$**
sample 1	I_{11}	I_{12}	...	I_{1k}
sample 2	I_{21}	I_{22}	...	I_{2k}
...	...	...	...	...
sample n	I_{n1}	I_{n2}	...	I_{nk}

Figure 2.18 - Example of a typical two-dimensional data table obtained in metabonomics, comprising n samples and k variables. I_{nk} is the intensity of variable k for sample n.

Data pre-processing is required to prepare the metabolic profiling data acquired in each of the previously described analytical platforms (NMR, LC-MS and MIR) for MVA. Generally, pre-processing can be considered into i) peak positional correction, ii) normalization and iii) scaling operations, which will be described in better detail for NMR, LC-MS and MIR as follows.

Regarding ^{1}H NMR, pH and ionic strength can originate drifts in peak position, therefore requiring the use of "binning" or peak alignment. Spectral "binning" or "bucketing" involves spectral integration over fixed spectral intervals whose width commonly ranges between 0.01 and 0.05 ppm, therefore accommodating the small peak deviations under the same intervals between samples. Since this procedure has an obvious drawback of reducing the spectral resolution, peak alignment can be used in order to correct the drifts in peak positions over the whole spectral

range (Savorani *et al.*, 2010). Several peak alignment algorithms are available such as fuzzy Hough transform (Alm *et al.*, 2009), recursive segment-wise peak alignment (RSPA)(Veselkov *et al.*, 2009), correlation shifting of spectral intervals combined with fast Fourier transform (iCOSHIFT) (Savorani *et al.*, 2010) or hierarchical Cluster-based Peak Alignment (CluPA) (Vu *et al.*, 2011).

Inter-sample dilution variation is a natural source of variation occurring in biofluids such as urine. In order to correct this source of variation, the ^{1}H NMR data sets are usually normalized to total spectral area or using other normalization methods such as probabilistic quotient normalization (also known as median fold-change normalization), quantile normalization, locally weighted scatter plot smoothing (LOESS) or group aggregating normalization (GAN)(Dieterle *et al.*, 2006; Dong *et al.*, 2011; Veselkov *et al.*, 2011). However, the choice of normalization method is dependent upon the sample type and problem under analysis (Dieterle *et al.*, 2006; Dong *et al.*, 2011; Veselkov *et al.*, 2011; Kohl *et al.*, 2012).

Complex samples such as biofluids are composed of metabolites with a large range of concentrations generating signals of varying intensities in ^{1}H NMR spectra. In order to avoid the dominance of the high intensity signals over lower ones in MVA it is necessary to apply scaling to the data sets (Veselkov *et al.*, 2011). In ^{1}H NMR it is common to apply the unit variance (UV) scaling (Ebbels and Cavill, 2009), which consists in the division of each element of the data table by the standard deviation of the respective column (variable). Alternatively, other scaling methods such as Pareto scaling (each element is divided by the square root of the standard deviation of the respective column) and logarithmic transformations (by application of the logarithm to each element of the data table), may be used (Veselkov *et al.*, 2011).

In relation to LC-MS data, small shifts in retention times may occur in data sets due to the variable interaction of the analytes with the stationary phase of the chromatographic column. Therefore, peak alignment strategies, such as those employed for ^{1}H NMR data are usually also performed after peak picking in LC-MS (Katajamaa *et al.*, 2006; Smith *et al.*, 2006). Prior to MVA, it is also common to apply normalization using similar methods to those described for ^{1}H NMR data, followed by scaling to UV, Pareto scaling and logarithmic transformation (Veselkov *et al.*, 2011).

In relation to MIR data, normalization is usually performed using standard normal variate (SNV) transformation, which consist in the division of each element of a line from the data table by the standard deviation of the respective line (Barnes *et al.*, 1989). Because the MIR bands are relatively broad, shifts due to small inter-sample pH and ionic strength variation are not usually observed, therefore, no peak alignment is necessary. Scaling is not usually performed for MIR data.

2.4.2. Multivariate Analysis (MVA) methods

The MVA methods used in metabonomics can be considered into supervised and unsupervised. In the unsupervised methods no a *priori* knowledge of the class information of the observations is considered, for instance which samples belong to the control or a disease group. These methods are employed to obtain an overview of the similarities and differences between the samples under analysis. The unsupervised methods which have been used in metabonomics include the principal components analysis (PCA), hierarchical cluster analysis (HCA), non-linear mapping procedures (NLM), self-organizing maps (SOM) and k-means clustering (Ebbels, 2007). On the other hand, when class membership of the samples is known, supervised methods may be used to explore the differences between classes and remove confounding factors resulting from inter-individual differences. Examples of supervised methods used in metabonomics include soft independent modeling of class analogy (SIMCA), partial least squares - discriminant analysis (PLS-DA), orthogonal PLS-DA (OPLS-DA), artificial neuronal networks (ANN) and genetic programming (Ebbels, 2007; Trygg and Lundstedt, 2007).

PCA, PLS-DA and OPLS-DA are the most used MVA methods in metabonomics, since they provide straightforward interpretation of the results, as it will be explained in the next sections.

2.4.2.1. Principal Components Analysis (PCA)

The multidimensional data table generated in metabonomics constitutes a matrix **X** with *K* variables and *N* objects (i.e. samples or observations) (Figure 2.18) or, in other words *N*×*K* dimensions, thus forming a scatter plot in the *K*-hyper-space (Figure 2.19).

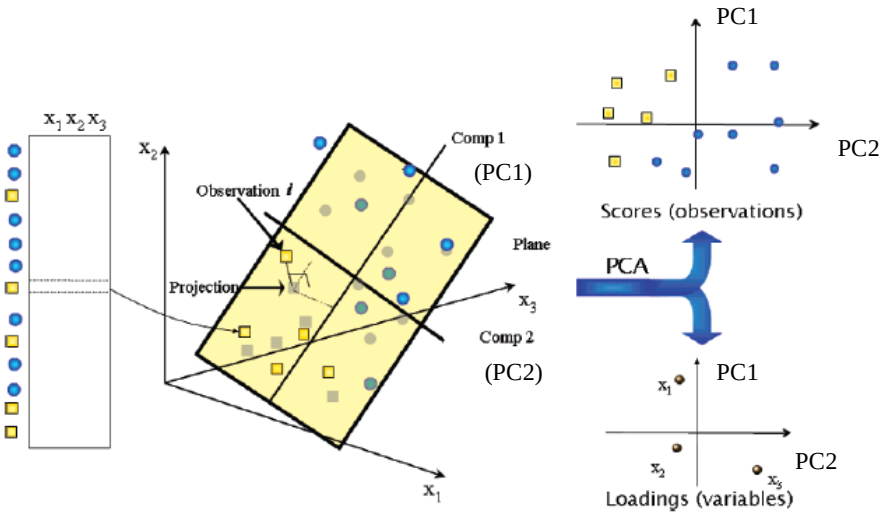


Figure 2.19 - Principal components analysis (PCA) geometric interpretation (Adapted from Trygg *et al.* 2007).

In PCA the *K*-space is reduced into a low-dimensional plane, representative of the objects in **X** (Figure 2.19). The new variables defining the plane are called the principal components (PCs) and are mutually linearly orthogonal to each other as shown in Figure 2.19 for a model with two PCs. The PCs result from the decomposition of the original data matrix **X** according to the equation:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathrm{T}} + \mathbf{E} \qquad \text{[Eq. 2.11]}$$

where **T** represents the scores matrix, i.e. the matrix containing the vectors which are used to define the low-dimensional planes that closely approximates the objects (samples) in **X** (Figure 2.19), meaning that position of each object in the planes (scores) is used to relate samples to each other. Hence, samples that are close to each other have a similar multivariate profile. The **P** matrix contains the weight vectors, or loadings, of the individual X-variables from the model. The loadings indicate which variables contribute for separation of different groups of objects in the scores, since directions in the score plot correspond to directions in the loading plot, and vice versa as illustrated in Figure 2.19 (right). The part of **X** that is not explained by the model, i.e. the residuals (**E,** equation 2.11) represents the distance between each point in *K*-space and its projection on the plane (Trygg *et al.*, 2007). Therefore, PCA provides a summary, or overview of all samples from the data table, making it possible to extract and display the systematic variation in the data.

2.4.2.2. Partial Least Squares and Orthogonal Partial Least Squares - Discriminant Analysis (PLS-DA and OPLS-DA)

In partial least squares (PLS), a quantitative relationship between two data tables is sought between a matrix **X**, usually comprising spectral or chromatographic data of a set of samples, and another matrix, **Y**, containing quantitative values, for example, concentrations of endogenous metabolites (Wold *et al.*, 2001; Trygg *et al.*, 2007). The **X** matrix, represented in the *K*-dimensional space (Figure 2.19), can be projected into hyper-planes defined by new variables, the latent variables (LVs), which are good predictors of the information contained in **Y** matrix. The PLS models can be described by the following equations:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathrm{T}} + \mathbf{E} \qquad \text{[Eq. 2.12]}$$

$$\mathbf{Y} = \mathbf{T}\mathbf{C}^{\mathrm{T}} + \mathbf{F} \qquad \text{[Eq. 2.13]}$$

where matrices **X** and **Y** are modeled by new variables, latent vectors (LVs) or components. The **P** and **C** matrices represent the **X** and **Y** loadings, respectively, whereas **E** and **F** matrices represent the residuals resulting from the modeling of **X** and **Y** matrices, respectively. The **T** matrices

represent the scores common to both **X** and **Y**. Moreover, the **T** scores are linear combinations of the original **X** variables and the weights **W** and, therefore, **T** can be expressed in the following equation:

$$\mathbf{T} = \mathbf{X}\,\mathbf{W}(\mathbf{P}^{\mathbf{T}}\mathbf{W})^{-1} \qquad \text{[Eq. 2.14]}$$

Combining equations 2.10 and 2.11, a PLS regression equation can be obtained:

$$\mathbf{Y} = \mathbf{XB} + \mathbf{F} \qquad \text{[Eq. 2.15]}$$

where **B** contains the PLS regression coefficients and $\mathbf{B} = \mathbf{W}(\mathbf{P}^{\mathbf{T}}\mathbf{W})^{-1}\mathbf{C}^{\mathbf{T}}$. The properties expressed in **Y** can then be predicted using this regression equation (Eq. 2.15).

While PLS is used to predict quantitative variables expressed in **Y**, it can also be used as a supervised discriminant analysis method, by using a **Y** matrix that encodes qualitative values (usually zeros and ones), called dummy variables, for example, to express class membership, gender, disease or treatment (Trygg *et al.*, 2007). This PLS variant is then designated by PLS-Discriminant Analysis (PLS-DA) (Wold *et al.*, 2001; Trygg *et al.*, 2007) (Figure 2.20, left).

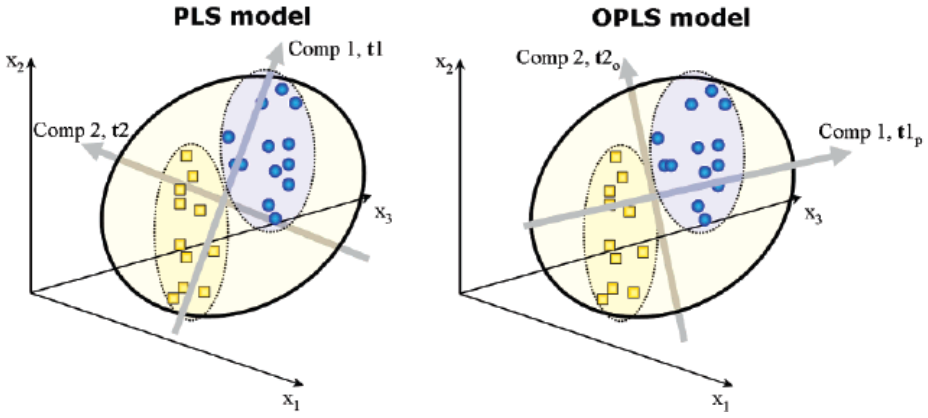


Figure 2.20 - A geometrical illustration of PLS and OPLS models. t_1, t_2 and t_{1p}, t_{2o} are the latent variables of PLS and OPLS models, respectively (Reproduced from Trygg *et al.*, 2007).

An important modification of the PLS method is the orthogonal-PLS (OPLS) method, in which the systematic variation in **X** is separated into two parts, one that is linearly related to **Y** and one that is unrelated (orthogonal) to **Y**. The resulting scores plots condensate the separation information into one single LV (Figure 2.20, right) therefore facilitating model interpretation when compared with PLS models (Trygg *et al.*, 2007). The OPLS model comprises two modeled components, the Y-predictive ($\boldsymbol{T_pP_p}^T$) and the Y-orthogonal ($\boldsymbol{T_oP_o}^T$) components, from which only the Y-predictive component is used for the modeling of Y ($\boldsymbol{T_pC_p}^T$) (Trygg and Wold, 2002), as described by the equations:

$$\mathbf{X} = T_p{P_p}^T + T_o{P_o}^T + \mathrm{E} \qquad \text{[Eq. 2.16]}$$

$$\mathbf{Y} = T_p{C_p}^T + \mathrm{F} \qquad \text{[Eq. 2.17]}$$

where **E** and **F** are the residual matrices of **X** and **Y**, respectively. OPLS can, analogously to PLS-DA, be used for discriminant analysis (OPLS-DA) (Bylesjö *et al.*, 2006).

The apparent separation present in PLS-DA and OPLS-DA scores alone is insufficient as a means to evaluate differences between the sample classes. In fact, for the purpose of classification of new samples, the models robustness needs to be assessed, which raises the need for appropriate model validation. The ideal validation of a model is such that it can precisely predict the **Y**-values of observations with new **X**-values, that is a validation data set of new samples (Westerhuis *et al.*, 2008). However, the samples gathered in metabonomics experiments are usually limited in size and almost all of them are used in the construction of the predictive model, i.e. to make a training set. In the absence of an external test set, validation can be performed by cross-validation (CV). The CV is performed by splitting the data sets in a number *G* of groups, usually five to nine, and then iteratively developing a number of parallel models (*G*) from reduced data set (training set), with one of the groups left out (used as validation set) (Wold *et al.*, 2001). The performance of the classification for each model in the CV procedure can be evaluated by means of a measure of the prediction error, given by goodness of prediction or predictive power (Q^2), which focuses on how efficiently the correct class label can be predicted from new data and can be defined as:

$$Q^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \qquad \text{[Eq. 2.18]}$$

where $\hat{y}_i$ refers to the predicted value of class membership for sample *i*, while $\bar{y}$ refers to the mean value of *y* for all samples (Westerhuis *et al.*, 2008). The value of Q^2 reported is usually the mean value of the *G*-fold CV iterations corresponding to each model produced. The optimal Q^2 value of 1 is difficult to obtain as it would require that the class prediction of each sample should be exactly equal to its class label, something that is hard to obtain due to the inherent variation between the individuals in the same class. The value of Q^2 depends, therefore, on the inter-class separation and also on the intra-class variability, which makes it is difficult to give a general Q^2 value that corresponds to a good classification (Westerhuis *et al.*, 2008). For this reason, it is usual to perform a permutations analysis, in which class memberships (**Y**-matrix) is randomly permuted, to provide a whole distribution of Q^2 values for models of incorrect class membership to relate the original Q^2 value (Westerhuis *et al.*, 2008).

Monte Carlo Cross Validation (MCCV) (Shao, 1993), is another validation method which can be performed in order to obtain the Q^2 distributions for both randomly permuted and original class memberships. In MCCV, 500-1000 random permutations are performed either by randomly

setting the **Y**-matrix (permuted classes) or by randomly sorting the original class memberships (true classes), i.e. randomly changing the order of samples in **X** and **Y** matrices simultaneously. The randomly created matrices (of permuted and true classes) are used to produce PLS-DA models which are evaluated through CV, the resultant Q^2 values being used to derive the Q^2 distributions.

Confusion matrices (Figure 2.21), which consist of the number of False Positive (FP), False Negative (FN), True Positive (TP) and True Negative (TN) classifications obtained by CV (Broadhurst and Kell, 2006), can also be used to derive other prediction measures to complement Q^2 distributions such as the classification rates (CR) and receiver operative characteristics (ROC) plots (Westerhuis *et al.*, 2008).

		Predicted classes	
		D	C
Actual classes	D	TP	FP
	C	FN	TN

Figure 2.21 - Example of a confusion matrix for two classes: D, disease and C, control. False Positive (FP), False Negative (FN), True Positive (TP) and True Negative (TN).

The classification rate is merely the ratio between the sum true positive (TP) and the true negative (TN) classifications and the total number of classifications:

$$\text{classification rate(CR)} = \frac{\text{TP+TN}}{\text{TP+FP+TN+FN}} \text{ x } 100 \qquad \text{[Eq. 2.19]}$$

The receiver operative characteristics (ROC) plots can be used to plot the true positives rate (TPR) and false positives rate (FPR) of a classification model (Fawcett, 2006). The TPR and FPR are equivalent measures to the sensitivity and 1-specificity of a model, respectively. The sensitivity (or true positive rate, TPR) is defined as the number of true positives found as a percentage of all positives (for instance, disease in a confusion matrix, Figure 2.21) as:

$$\text{sensitivity (TPR)} = \frac{\text{TP}}{\text{TP+FP}} \qquad \text{[Eq. 2.20]}$$

On the other hand, 1-specificity (the false positive rate, FPR) is the number of false positives (FP) as a percentage of all negatives (for instance, controls in the confusion matrix, Figure 2.21):

$$\text{1-specificity (FPR)} = \frac{\text{FP}}{\text{TN+FN}} \qquad \text{[Eq. 2.21]}$$

Each obtained PLS-DA model can be represented as a point with its associated sensitivity and 1-specificity value in a ROC space plot (Figure 2.22).

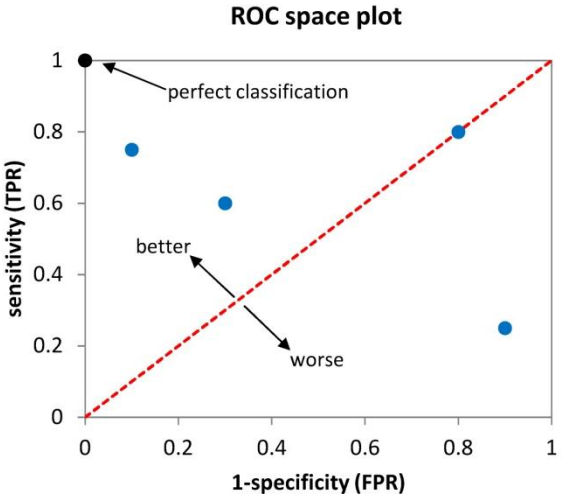


Figure 2.22 - ROC space plot of five example models (●). Perfect classification, i.e. sensitivity (TPR) of 1 and 1-specificity of 0 is illustrated. Dashed line, indicates the random classification.

The sensitivity values fall between 0 and 1 and should ideally be close to 1. Specificity values should ideally also be close to 1 and hence 1-specificity should be close to 0. A sensitivity value of 1 and 1-specificity of 0 indicates a perfect classification, whereas the values falling in the diagonal of the ROC space plot indicate a random classification performance (Figure 2.22).

2.4.2.3. Statistical total correlation spectroscopy (STOCSY) and statistical heterospectroscopy (SHY) methods

Statistical Total COrrelation SpectroscopY (STOCSY) enables the identification of linear relationships within spectroscopic data sets which may result from the same analytes (aiding spectral assignments) or from different analytes indicating dynamic processes, for instance metabolic reactions or physiologic processes (Lindon and Nicholson, 2008b).

Generalized two-dimensional correlation spectroscopy was introduced by Noda in an application to infrared spectroscopy to study time-course processes (Noda, 1989). The method was later extended to sample−sample correlation spectroscopy to obtain the information about the species perturbation-dependent dynamics in chemical processes (Šašić *et al.*, 2000). STOCSY was introduced more recently as a different approach to measure correlations in spectroscopic data sets. It is based on the principles in which the multicollinearity of the intensity variables in a set of spectra (for example 1D NMR) can be used to produce a pseudo-two dimensional spectrum showing correlation among the intensities of the various peaks across the whole sample (Cloarec *et al.*, 2005; Lindon and Nicholson, 2008b). This method allows the visualization of much more

information than that usually displayed, for instance, in a 2D NMR spectrum such as a TOCSY. An example is shown in Figure 2.23a for ^{1}H NMR spectra of human urine. In fact, in STOCSY two or more molecules involved in the same pathway can also present high intermolecular correlations or can even be anti-correlated (negatively correlated) due to common metabolic pathway control (Cloarec *et al.*, 2005).

STOCSY is based on the properties of the correlation matrix **C**, computed from a set of sample spectra according to the expression:

$$\mathbf{C} = \frac{1}{n-1}\,\mathbf{X_1^T X_2} \qquad \text{[Eq. 2.22]}$$

where $\mathbf{X_1}$ and $\mathbf{X_2}$ denote the autoscaled (UV scaled) experimental matrices of $n \times v_1$ and $n \times v_2$, respectively; n is the number of spectra (one for each sample) and v_1 and v_2 are the number of variables in the spectra for each matrix. **C** is therefore a correlation matrix of $v_1 \times v_2$, where each value is a correlation coefficient between two variables of the matrices $\mathbf{X_1}$ and $\mathbf{X_2}$. Correlation coefficient (***r***) can have values between -1 and 1, depending on whether two variables are varying on opposite directions or same direction, respectively. In real samples, $|r|$ will always be ≥0 and <1 because of spectral noise or peak overlaps from other molecules (Cloarec *et al.*, 2005). In practice, the correlation matrix from a set of spectra containing different amounts of the same molecule shows very high correlations between the variables corresponding to the signals of the same molecule. Plotting the correlation matrix provides a graphic representation of the multi-sample spectroscopic data set comparable to that of a 2D correlation experiment, such as an 2D NMR experiment (Cloarec *et al.*, 2005). The simplest application of STOCSY is the auto-correlation analysis where $\mathbf{X_1} = \mathbf{X_2}$, illustrated in Figure 2.23a.

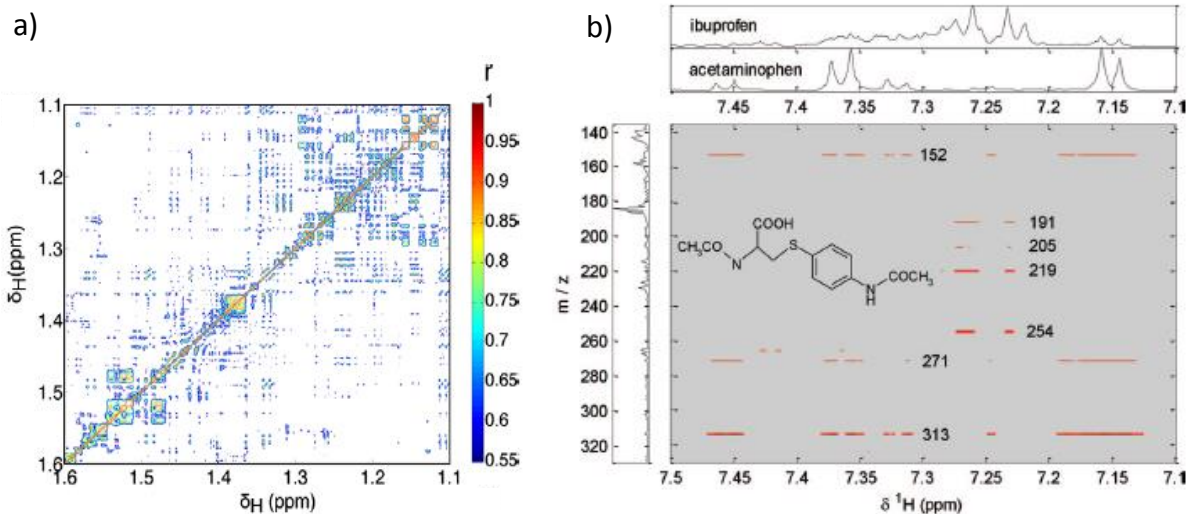


Figure 2.23 - STOCSY and SHY analysis of human urine samples. a) ^{1}H-^{1}H NMR STOCSY map (Adapted from Keun *et al.*, 2008); b) UPLC-MS / ^{1}H NMR SHY map shows positive correlations (red) between signals the of ibuprofen and acetaminophen in both spectroscopic domains (Adapted from Crockford *et al.*, 2008).

Furthermore, the use of two different matrices has been explored, by correlating ^{1}H NMR and ^{31}P NMR matrices (Coen *et al.*, 2007), and also notably ^{1}H NMR and UPLC-MS matrices to obtain an heterospectral correlation matrix, designated Statistical HeterospectroscopY (SHY) (Crockford *et al.*, 2006). In SHY, direct cross-correlation of chemical shifts from NMR and m/z data from MS provided improved structural information for a series of molecules in urine (Figure 2.23b).

Higher level biological information can also be obtained on metabolic pathway activity and connectivities by examination of different levels of the NMR to MS correlation and anti-correlations in rat and human urine (Crockford *et al.*, 2006; Crockford *et al.*, 2008). Besides NMR and MS, SHY analysis can be of general applicability to complex mixture analysis, if two or more independent spectroscopic data sets are available for any sample cohort (Crockford *et al.*, 2006). For instance, the SHY concept has been extended to study the metabolic correlations between ^{1}H NMR and MRS of different biofluids and tissues from HIV-1 individuals (Maher *et al.*, 2011). Another example is the application of SHY for the 1st time to correlate MIR and ^{1}H NMR of AF (Graça *et al.*, 2013), a work which will be presented in chapters 5 and 6 of this thesis.

It is sometimes convenient to express STOCSY results as a one-dimension spectral representation (1D STOCSY, Figure 2.24) instead of the direct representation of the correlation matrix (a two-dimensional result).

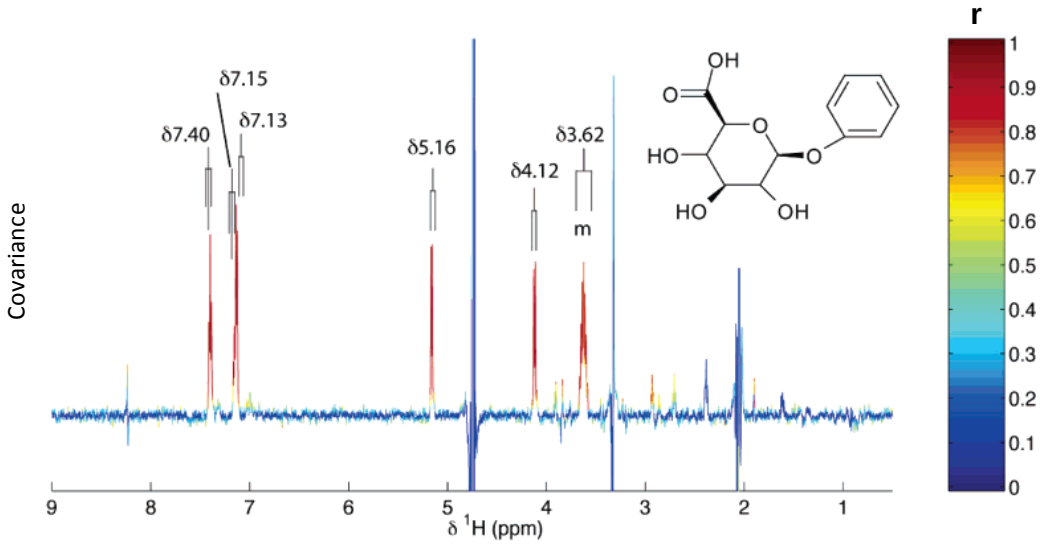


Figure 2.24 - ^{1}H NMR 1D STOCSY spectral representation. The representation was obtained from a series of ^{1}H NMR spectra obtained in a LC-NMR experiment by calculating the correlation (r) and covariance between the chemical shift 7.13 ppm and the full spectral range, thus highlighting the signals of phenyl glucuronide (Cloarec *et al.*, 2007).

In the 1D STOCSY, the covariance between one selected variable of $\mathbf{X_1}$ matrix (for instance one resonance or MS intensity), and all the other variables of the $\mathbf{X_2}$ matrix (for instance a matrix of NMR spectra) is plotted with a colour code corresponding to the correlation between the selected variable from $\mathbf{X_1}$ and all the other variables of $\mathbf{X_2}$. In this way, the correlated peaks are highlighted by the projection of a colour coding to show degree of correlation (Cloarec *et al.*, 2005; Cloarec *et al.*, 2007). If, for instance X_2 is an NMR matrix (as in the example of Figure 2.24), the shape of the correlated peaks in 1D spectra is respected, therefore facilitating analysis.

3. Experimental procedures

The present chapter describes the sample and metadata collection procedures, as well as the experimental and data analysis strategies used to conduct the experimental work presented throughout the thesis.

3.1. Samples and clinical metadata

The AF and urine samples used in this work were collected between 2006 and 2011 at the *Centro de Diagnóstico Pré-Natal, Maternidade Bissaya Barreto, Centro Hospitalar de Coimbra*. Between January 2008 and June 2011, this work was supported by the Portuguese Foundation of Science and Technology (FCT) funding through the PTDC/QUI/66523/2006 project. Samples were collected under the approval of the ethical committee of *Centro Hospitalar de Coimbra* (Refs.18/04 and 29/09) and informed consents were obtained from each subject participating in the study (Annex I - Informed signed consent and questionnaire). The donors comprised Caucasian pregnant women, from the central part of Portugal, mostly in their second trimester of pregnancy recommended to undergo amniocentesis due to maternal age or for several medical reasons listed in Table 3.1.

Table 3.1 - Reasons for undergoing amniocentesis relating to the participants in the studies. AF and urine samples were collected between 2006 and 2011.* median (minimum-maximum) maternal age (years).

Reasons for amniocentesis	Description	Maternal age*	Numbers (%)
Advanced maternal age	Women over 35 for whom karyotype analysis is recommended	37 (35-46)	583 (64%)
Positive biochemical test	Positive trisomy 18 or 21 screening tests	31 (18-39)	67 (7%)
Suspected fetal malformation	Ultrasound detected malformation	30 (13-37)	73 (8%)
Ultrasound signs of alert	Ultrasound detected signs of alert such as nuchal translucence, growth restriction, etc.	32 (17-41)	30 (3%)
Previous chromosomal disorder	Previous child with chromosome disorder	33 (18-35)	17 (2%)
Parent with chromosomal disorder	One of the parents carried chromosomal disorder	30 (27-35)	8 (1%)
Maternal anxiety	Amniocentesis requested by mother with no apparent medical indications	36 (34-37)	6 (1%)
Others	AF collected for Rh immunization, mothers diagnosed with infections such as Toxoplasmosis, CMV, Rubella	31 (25-35)	24 (3%)
Undefined	Cases with incomplete clinical information	34 (18-40)	106 (12%)
Total			**914**

Every sample was given an arbitrary code number so that confidentiality was ensured after collection and during the analysis process. Regardless of the reason for undergoing amniocentesis, sample classification and inclusion in the study was performed after labour, when all relevant clinical information was available.

Due to medical restrictions in controlling/limiting the diet of pregnant women up to the time of sample collection, samples were obtained under non-fasting conditions during the morning period. However, the effects from the occurrence of disease are expected to have a stronger and long-term effect on metabolic profile than those resulting from the diet and life style.

3.1.1. Amniotic fluid and maternal urine collection

About 10-20 mL of AF was collected through ultrasound-guided trans-abdominal amniocentesis. The AF samples were transported to the cytogenetics laboratory of the *Faculdade de Medicina da Universidade de Coimbra* within a maximum period of 4 hours, during which samples were kept at room temperature. For this reason the subject AF stability at room temperature was addressed specifically in chapter 4 (section 4.2.3). Samples were then centrifuged at 177.41×*g* at 22°C for 10 minutes in order to collect the cells (pellet) for karyotype analysis. The supernatants were immediately frozen and stored at -70°C/-80°C. The AF samples containing blood were discarded.

Maternal urine samples (20-50 mL) were also collected from women who attended the prenatal diagnosis appointments to perform amniocentesis. Therefore, most of the urine samples were obtained from the same donors of AF. The urine samples were maintained and transported at room temperature to the cytogenetics laboratory of the *Faculdade de Medicina da Universidade de Coimbra*, within a maximum period of 4 hours. No significant metabolic changes are however expected in human urine in these conditions, based on a previous report (Barton *et al.*, 2008). Urine samples were then immediately stored at -80°C.

Prior to analysis all samples were transported to the *Departamento de Química*, *Universidade de Aveiro* in dry ice (approximately -70°C) and stored at -70°C/-80°C until analysis.

3.1.2. Sample classification

A total of 914 AF samples were obtained between 2006 and 2011 as shown in Table 3.1, which corresponded to the duration of the project presented in this thesis. A smaller number of 588 maternal urine samples were also obtained between 2006 and 2009, from which a smaller group

was used in the UPLC-MS study described in better detail in section 3.4. Information on pregnant women life-style (diet, physical exercise), medication, maternal and gestational age, known diseases and blood pressure were obtained from the questionnaire which accompanied the signed consent (Annex I - Informed signed consent and questionnaire).

Clinically relevant metadata regarding the time course of pregnancy such as pre-existing diseases, blood type, maternal height and weight before pregnancy (used to calculate the body mass index, BMI), smoking habits, obstetrical history, blood pressure records, clinical chemistry and urine analysis results, karyotype results, obstetrical data (including ultrasound monitoring and medical the diagnosis results in the cases of obstetrical disorders) were obtained through consultation of the medical team, after delivery. Information about the labour and pregnancy outcome, including fetal sex, weight and medical physical examination records up to the second day of life was obtained by consultation of the neonatology medical team. The gathered metadata allowed the classification of the samples into controls and several disorder groups. The samples comprising the control group were obtained from healthy non obese women, who were not under medication, and who did not register any obstetrical complication and delivered healthy term newborns (Table 3.2).

The fetal malformations (FM) group comprised pregnancies affected by some type of congenital malformation, diagnosed by morphological ultrasound examination, usually after 12 weeks of gestation and eventually confirmed by other imaging methods (such as magnetic resonance imaging). Third trimester (28-36 weeks) FM samples were also obtained, but their use was not considered given the extent of AF compositional differences occurring between trimesters (Lind, 1981; Sims *et al.*, 1996; Cohn *et al.*, 2009). Fetal malformations cases obtained are listed in Table 3.3, where they were divided by organic system affected, also with the indication of each FM type. One of the cardiac malformation cases also carried a 21 trisomy (Table 3.3) and another was from a twin pregnancy. Similarly to the FM group, an chromosomal disorders (CD) cases were considered as a group comprising cases from diverse types of chromosome disorders (numerical, structural and mosaicisms), usually diagnosed up to two weeks after amniocentesis (Table 3.3).

Second trimester samples were also obtained from women who later (during the 3rd trimester) developed GDM, PTD, PROM, PE and IUGR. These sample groups were designated *pre*-diagnostic GDM, *pre*-PTD, *pre*-PROM, *pre*-PE and *pre*-IUGR (Table 3.2) and were considered in order to evaluate the 2nd trimester AF composition predictive value for those prenatal disorders.

Table 3.2 - AF and maternal urine sample numbers subdivided by group and respective maternal ages (years), gestation ages (weeks) and pre-pregnancy body-mass index (BMI, kg/m^2) values. [a] samples with known classification at the time of analysis; [b] median values and corresponding range in brackets; [c] mean BMI value, in brackets: the number of cases with known BMI; [d] one FM case carrying a chromosomal disorder; [e] one case carrying chromosomal disorders; [f] one case also with diagnosed with PE; [g] one case of twin gestation, [f] either due to insufficient sample numbers, cases outside the scope of this work, existence of confounding effects or lack of clinical information.

	samples		***Age***		
Groups included in this study	**AF**	**urine[a]**	**gestational[b]**	**maternal[b]**	**BMI[c]**
Controls	90	68	17 (15-22)	36 (25-44)	22 (35)
Fetal Malformations (FM) [d,g]	36	13	22 (14-25)	29 (13-37)	23.8 (13)
Chromosomal Disorders (CD)	22	-	17 (15-23)	38 (22-40)	26.3 (14)
Prediagnostic Gestational Diabetes Mellitus (*pre*-diagnostic GDM) [e,g]	35	20	17 (16-21)	37 (30-44)	27.1 (21)
Pre-Preterm Delivery (*pre*-PTD) [g]	18	6	17 (16-17)	36 (30-39)	25.8 (10)
Pre-Premature rupture of membranes (*pre*-PROM)	45	-	17 (15-23)	36 (27-42)	23.4 (29)
Pre-Preeclampsia (*pre*-PE)	8	-	17 (16-17)	35 (24-41)	24.9 (7)
Intrauterine Growth Restriction (*pre*-IUGR) [g]	8	-	16 (15-17)	36 (31-39)	21.9 (2)
AF samples from groups not included in this study [f]					
Chronic Hypertension	5	Fetal death			2
Diabetes Mellitus	5	Other disorders/infections/medication			356
Macrosomia	15	No classification			267
Post-diagnostic Gestational Diabetes	2				

Table 3.3 - The different types of fetal malformations (FM) (*italic*), grouped by affected organic system (underlined); and chromosomal disorders (CD) and respective number. [a] one FM case carried a 21 trisomy; [b] one pre-diagnostic GDM case.

Malformation types and sample numbers					
Urogenital	5	Soft tissues	2	Poly-malformed	3
Polycystic kidney	*2*	*Cystic hygroma*	*2*	*Cleft-lip + tetramelia*	*1*
Hydronephrosis	*1*	Abdominal	2	*Jaw malformation+ club foot*	*1*
Multicystic dysplastic kidney	*1*	*Omphalocele*	*1*	*Cardiophaty + polydactylia*	*1*
Bladder agenesis	*1*	*Enlarged stomach*	*1*	Central Nervous System	13
Cardiac	5	Limbs	5	*Corpus callosum agenesis*	*3*
Complex cardiophaty	*2*	*Polydactylia*	*1*	*Complex cranioencephalic*	*2*
Great vessels malformation	*1*	*Club-feet*	*4*	*Holoprosencephaly*	*1*
Mitral valve agenesis	*1*	Lungs	1	*Hydrocephaly*	*1*
Septal malformation [a]	*1*	*Cystic adenomatoid*	*1*	*Spina bifida*	*2*
				Ventriculomegaly	*4*
Chromosomal disorder types and sample numbers					
Trisomy 21	7	XYY syndrome	2	Mosaicisms	3
Trisomy 18	1	XXX syndrome	1	Inversions	2
Kleinefelter syndrome [b]	3	Marker chromosome	1	Translocations	4

A *pre*-diagnostic GDM group, as opposed to *post*-diagnostic which refers to cases already diagnosed with GDM, was obtained considering only cases from women diagnosed with GDM only after amniocentesis (after 24–28 gestational weeks). This group contained one case also carrying a chromosomal disorder and one case of twin pregnancy (Table 3.3).

The *pre*-PTD group comprised cases from women who delivered before 37 gestational with or without membrane rupture (between 34 and 36 weeks). The *pre*-PROM group, on the other hand, was composed of cases from women who later had premature membrane rupture of the fetal membranes prior to labour after 37 gestational weeks (between 37 and 40 weeks). The PROM condition is clinically very similar to a healthy pregnancy, except for the fact that the membrane rupture occurs spontaneously at term before labour initiation. The number of *pre*-PROM cases registered was higher than those registered for other disorders considered here. Therefore only the *pre*-PROM cases with no overlapping conditions were considered to avoid misinterpretations of the results.

The *pre*-PE group comprised samples from women who later developed PE, diagnosed between 31 and 36 gestational weeks, whereas the *pre*-IUGR group was composed from cases where fetuses were diagnosed with IUGR at 34-38 gestational weeks. One case from this group was also diagnosed with PE later in pregnancy and another case was a twin pregnancy (Table 3.2).

It was not possible to obtain the pre-pregnancy BMI for all subjects, therefore, these were only considered for the samples indicated in Table 3.2. It is also important to mention that the cases registering obesity (BMI>30), multiple gestation and chronic disease were usually included in the respective disorder group, except for the *pre*-PROM group, as mentioned above. This was justified by the fact that obesity, multiple gestation and pre-existing chronic diseases are risk factors for many of the disorders studied and also due to the low sample numbers obtained.

Regarding the AF samples not included in this study (Table 3.2, bottom), these comprised AF samples obtained from donors which registered other conditions in the absence of the previously described prenatal disorders. These conditions included chronic hypertension (0.5%), diabetes mellitus (0.5%), macrosomia (1.6%), gestational diabetes (*post*-diagnostic) (0.2%) and fetal death cases (0.2%) (Table 3.2). Cases comprising other chronic diseases, maternal infections and cases of reported medication and the *pre*-PROM cases with overlapping conditions, accounting 38.9% of the total AF samples, were included in a separate group (other diseases/infections/medication) (Table 3.2). Samples for which no classification was possible (mainly due to incomplete clinical records, 29.2% of the total) were also not considered in this study (Table 3.2).

The maternal urine samples used in this study were collected between 2006 and 2009, and the numbers shown in Table 3.2 refer to the cases for which classification was known at the time of the analysis.

3.2. Chemicals

All solid chemicals, deuterated and non-deuterated solvents and solutions used in the experimental work performed were from analytical grade or equivalent and are listed in Table 3.4.

Table 3.4 - Chemicals used throughout the work. Manufacturer, purity or concentration (*) in case of solutions are indicated. Standard compounds for: [a]NMR and [b]MIR are also indicated.

Deuterated solvents/solutions (for NMR and LC-NMR)		
Name	***Purity***	***Manufacturer***
Acetonitrile (d_3)	97.0%	Euriso-top (Saclay, France)
Deuterium oxide (D_2O)	99.9%	Euriso-top (Saclay, France)
Deuterium oxide containing 0.75 % TSP (3-(trimethylsilyl)propionic-2,2,3,3-d_4 acid) sodium salt	99.9%	Aldrich (St Louis, USA)
Methanol (d_4)	99.8%	Euriso-top (Saclay, France)
Non-deuterated solvents/solutions		
Name	***Purity***	***Manufacturer***
Acetonitrile	99.9%	Fluka (Steinheim, Germany)
Formic acid	98.0%	Fluka (Steinheim, Germany)
Hydrochloric acid *	37.0%	Merk (Darmstadt, Germany)
L-lactic acid	98.0%	Sigma (Steinheim, Germany)
Methanol	99.9%	Sigma-Aldrich (St Louis, USA)
Sulphuric acid *	95.0%	Fluka (Steinheim, Germany)
Water (LC-MS grade)	100.0%	Fluka (Steinheim, Germany)
Solid chemicals		
Name	***Purity***	***Manufacturer***
Ammonium acetate	98.0%	Sigma (St. Louis, USA)
Alanine [a,b]	98.0%	Sigma (St. Louis, USA)
Bovine serum albumin (BSA) [a,b]	96.0%	Sigma (St. Louis, USA)
L-citrulline [a]	98.0%	Fluka (Steinheim, Germany)
Creatine [a]	98.0%	Aldrich (St Louis, USA)
Creatinine [a,b]	98.0%	Sigma-Aldrich (St. Louis, USA)
L-Cystine [a]	98.0%	Sigma (St. Louis, USA)
Ethanolamine [a]	98.0%	Sigma (St. Louis, USA)
D-Glucose [a,b]	99.0%	Riedel-de-Haën (Seelze, Germany)
Glutamic acid [a,b]	99.0%	Sigma (St. Louis, USA)
Glutamine [a,b]	99.0%	Sigma (St. Louis, USA)
Glycine [a,b]	99.0%	Sigma (St. Louis, USA)
α-ketoglutarate [a]	98.0%	Riedel-de-Haën (Seelze, Germany)
L-lysine [a,b]	98.0%	Sigma (St. Louis, USA)
L-malic acid [a]	99.0%	Aldrich (St Louis, USA)
Ornithine [a]	99.0%	Aldrich (St Louis, USA)
Potassium hydroxide	99.0%	Merk (Darmstadt, Germany)
Potassium di-hydrogenophosphate	99.0%	Merk (Darmstadt, Germany)
Shikimic acid [a]	99.0%	Sigma (St. Louis, USA)
Sodium hydroxide	99.0%	Merk (Darmstadt, Germany)
di-Sodium hydrogenophospate	99.0%	Fluka (Buchs, Switzerland)
Taurine [a]	99.0%	Sigma (St. Louis, USA)
Urea [b]	99.0%	Panreac (Barcelona, Spain)

3.3. High resolution NMR spectroscopy

3.3.1. Sample preparation for NMR

3.3.1.1. AF sample preparation

In this work, AF sample quality was evaluated by ^{1}H NMR to assess the effects of a) freeze-drying *versus* direct dilution, b) buffering, c) ultrafiltration and, hence, define the best protocol for AF sample preparation. Each AF sample was thawed at room temperature. The AF preparation methods are described in detail below.

Direct dilution: to 500µL of each AF sample 50µL of 0.24% TSP sodium salt solution in 100% D_2O were added, which resulted in pH values in the interval 8.75±0.25;

Freeze-drying: 1mL of each AF sample were frozen in liquid N_2 prior to freeze-drying during 8-12h (T=25°C, P=13Pa) and were then reconstituted into 720µL of D_2O and 80µL of 0.24% TSP in 100% D_2O solution, which resulted in an 1.3-fold concentration factor. The resulting pH values of the samples were within the interval 9.23±0.12;

Buffering: to each 500µL of sample 100µL of buffer A (0.5 M sodium phosphate buffer pH 7.20 ±0.10 in 100% D_2O) were added, which resulted in pH values in the interval 7.32±0.07;

Ultrafiltration: AF samples were also subjected to size-selective ultrafiltration using Amicon centricon 3kDa and 10kDa cut-off membrane centrifuge filters (Millipore Corporation, Billerica, MA, USA), as shown in Figure 3.1.

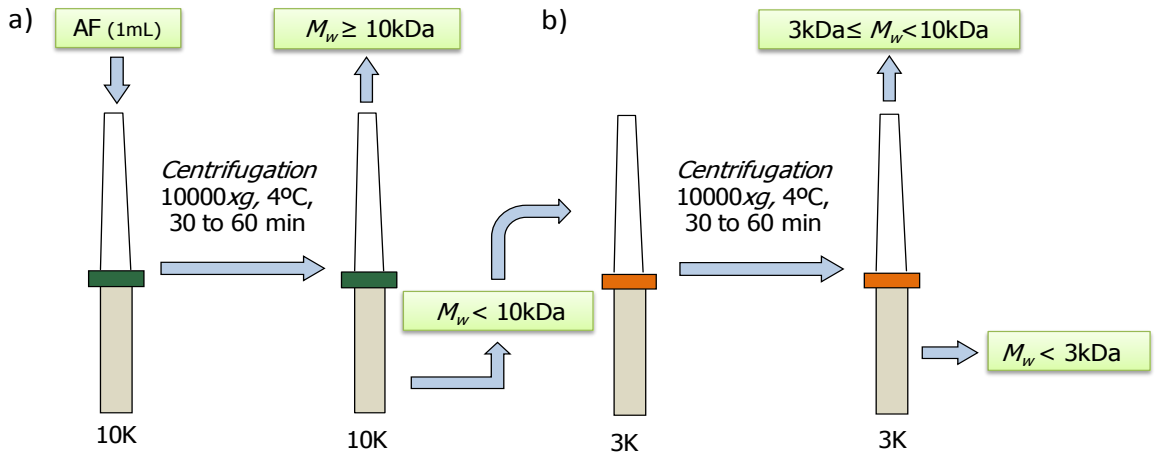


Figure 3.1 - Sequential amniotic fluid fractionation by ultrafiltration through centrifuge membrane filters: a) through 10kDa cut-off filter, fractions with M_w≥10kDa and M_w<10kDa were obtained; b) the M_w<10kDa fraction was further filtered through 3kDa cut-off filters resulting in one fraction with 3≤M_w <10 kDa and another with M_w<3 kDa.

Briefly, 1mL of each of six control samples was subjected to two sequential filtrations, the first employing 10kDa cut-off filters where fractions with M_w≥10kDa and M_w<10kDa were obtained (Figure 3.1a); secondly each fraction of M_w<10kDa was further filtered through 3kDa cut-off filters thus obtaining 3≤ M_w <10 kDa and M_w<3 kDa fractions (Figure 3.1b); the higher molecular weight fraction (M_w≥10kDa) was eventually washed several times with buffer A (0.5 M sodium phosphate buffer pH 7.20±0.10 in 100% D_2O) and was finally diluted back to the initial volume (1mL).

All AF samples and fractions were centrifuged (13400*xg*, 25°C, 5 min) and transferred into 5mm NMR tubes, prior to spectral acquisition. A scheme of the preparation procedures for AF samples is shown in Figure 3.2.

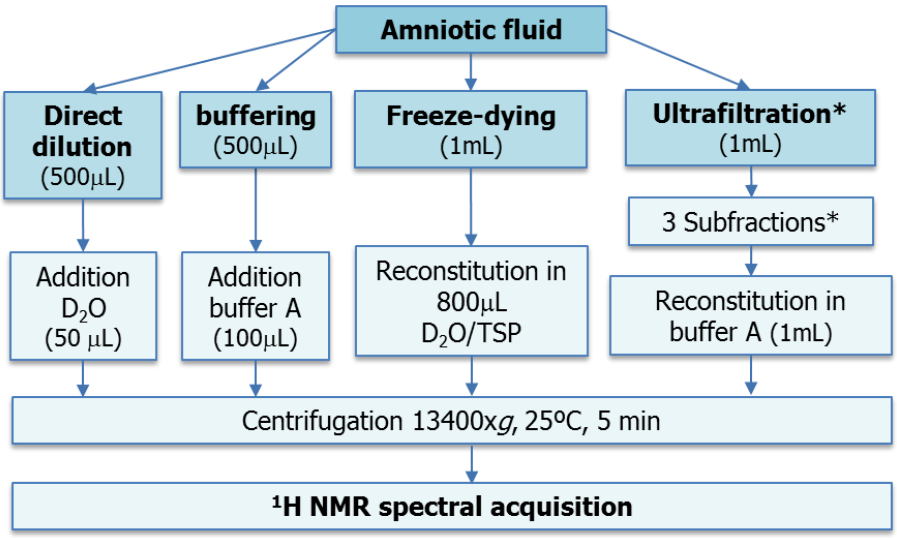


Figure 3.2 - Scheme of AF preparation procedures for ^{1}H NMR analysis. Buffer A (sodium phosphate buffer 0.5M at pH 7.2 in 100% D_2O); * detailed description in Figure 3.1.

3.3.1.2. Amniotic fluid stability evaluation for NMR

AF sample stability was assessed during storage, after freeze-thaw cycles and at room temperature by 1D ^{1}H NMR (using "standard" experiments, details in section 3.3.2) using healthy control AF samples. The stability tests are represented in the scheme of Figure 3.3 and explained in detail below.

Evaluation of stability with storage time: 3 AF samples were subdivided into aliquots, 6 aliquots per sample were stored at -20°C and other 6 were stored -70°C (Figure 3.3a). After each month one aliquot of each sample was thawed from -20°C and from -70°C, prepared with buffering as described in the previous subsection (page 63) and analysed by ^{1}H NMR (Figure 3.3a);

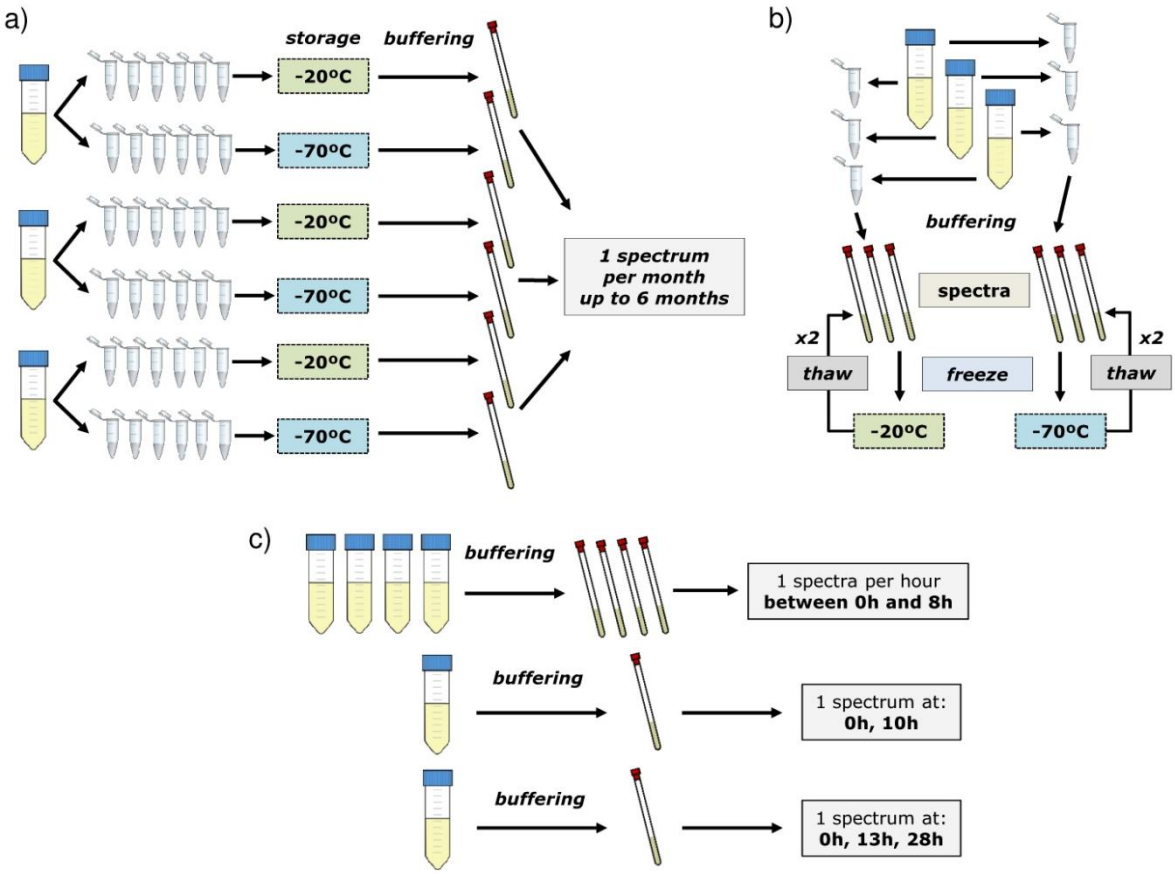


Figure 3.3 - Schematic representation of the tests to evaluate sample stability: a) during storage at -20°C and -70°C; b) with freeze-thaw cycles at -20°C and -70°C; c) at room temperature (22-25°C). Buffering: addition of 100μL of sodium phosphate buffer 0.5M at pH 7.2 in 100% D_2O to 500 μL of AF sample.

Evaluation of the effect of freeze-thaw cycles: 3 other AF samples were subdivided into 2 aliquots each and prepared with buffering and transferred to NMR tubes (Figure 3.3b). After analysis by ^{1}H NMR, 3 tubes were frozen at -20°C and other 3 tubes at -70°C. The tubes were thawed and analysed by ^{1}H NMR. After ^{1}H NMR analysis, the procedure was repeated once more (Figure 3.3b). The procedure was repeated once more for one of the samples.

Evaluation of stability at room temperature (22-25°C): 6 samples were prepared with buffering and then repeatedly analysed by ^{1}H NMR in regular intervals (when possible) of 1 hour up to 8 hours or at 0, 10, 13 or 28 hour periods (Figure 3.3c).

All AF samples and fractions were centrifuged and (13400*xg*, 25°C, 5 min) and transferred into 5mm NMR tubes, prior to spectral acquisition.

3.3.1.3. Urine sample preparation for NMR

Maternal urine was prepared by dilution with a phosphate buffer, according to a previously published preparation method (Bernini *et al.*, 2011), summarized in Figure 3.4.

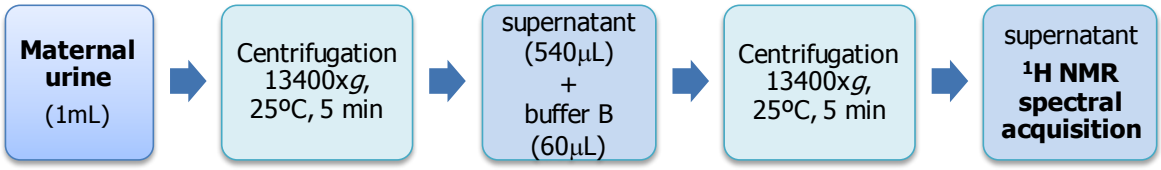


Figure 3.4 - Scheme of maternal urine sample preparation procedure for ^{1}H NMR analysis. Buffer B (potassium phosphate 1.5M, at pH 7.0 with 0,1% TSP and 2mM NaN_3 in 100% D_2O).

Each urine sample was thawed at room temperature and, after gentle mixing, 1mL of each sample was centrifuged (13400×*g*, 25°C, 5 min) to remove any particulate matter. To 540 µL of each urine supernatant, 60µL of potassium phosphate buffer 1.5M with 0,1% TSP with 2mM NaN_3 (used to prevent bacterial growth) in 100% D_2O at pH 7.00±0.10 (buffer B) were added to in order to minimize inter-sample pH differences (Figure 3.4). The pH of each buffered sample was measured once more and, if necessary, adjusted to pH 7.00±0.10 with volumes of 0.5 - 2µL of 4M HCl or 4M KOH solution prepared in 100% D_2O. The buffered samples were centrifuged (13400×*g*, 25 °C, 5 min) and 550µL of each prepared sample was transferred into 5 mm NMR tubes after for NMR spectral acquisition (Figure 3.4).

3.3.2. NMR experimental details

3.3.2.1. NMR equipment

The NMR spectrometers used and respective facilities where the experiments were performed are summarized in Table 3.5. The samples were kept at 22°C inside the sample changers prior to acquisition for periods no longer than 12 hours in order to prevent sample degradation. Access to 600MHz and 800MHz spectrometers was provided through courtesy of *Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa* (FCT-UNL) - *Laboratório associado* REQUIMTE, Portuguese NMR Network (Portugal); and NMR Application, Bruker BioSpin, NMR division (Reinstetten, Germany), respectively.

Table 3.5 - Equipment used to perform the NMR experiments and respective facility.

NMR facility	Spectrometer and equipment	Experiments
Departmento de Química, Universidade de Aveiro, Portugal	▪ Bruker Avance DRX 500MHz ▪ 5mm BBI probe ▪ operating at 300K ▪ gradient unit (53.5 G/cm, max. output) ▪ NMR-Case sample changer ▪ xwin-nmr 3.1 software	1D, 2D NMR and DOSY
FCT-UNL - Laboratório associado REQUIMTE, Portuguese NMR Network, Portugal	▪ Bruker Avance III 600 MHz ▪ TCI cryo-probe ▪ operating at 300K temperature ▪ TopSpin 2.1 software	Saturation transfer difference (STD)
NMR Applications, Bruker BioSpin NMR division, Germany	▪ Bruker Avance 800MHz ▪ 5 mm TXI probe ▪ operating at 298K ▪ gradient unit (53.5 G/cm, max. output) ▪ NMR-Case sample changer ▪ TopSpin 1.3 software	1D and 2D NMR

3.3.2.2. 1D and 2D ^{1}H NMR experiments

1D ^{1}H NMR experiments for AF analysis were acquired with water suppression using presaturation pulses with 58dB or 53.8dB power at 500MHz and 800MHz, respectively, set at the water frequency. Two "standard" 1D experiments which enabled the registration of signals from high and low M_w molecules were acquired, namely a *zgpr* (RD-90°-acquire) and *noesypr1d* (RD-90°-t_1-90°-t_m-90°-acquire) experiments from Bruker library. In the *zgpr* experiment water presaturation is performed during relaxation delay and was used for freeze-dried samples diluted in D_2O, whereas in *noesypr1d* water presaturation is performed during relaxation delay and mixing time thus enabling more efficient water suppression. The *noesypr1d* experiment was then performed for samples containing more water, namely buffered AF and urine. Moreover, T_2 relaxation edited ^{1}H spectra were acquired using the Carr-Purcell-Meiboom-Gill (CPMG) pulse sequence (RD-90°-(τ-180°-τ)$_n$-acquire) (Meiboom and Gill, 1958) (*cpmgpr1d*, from Bruker library) in order to filter signals from high M_w molecules. Diffusion-edited spectra were also recorded using the bipolar pulse longitudinal eddy current delay (BPPLED) pulse sequence with 2 sine gradient pulses (Wu *et al.*, 1995) (*ledbpgp2s1dpr*, from Bruker library) in order to filter low M_w molecule signals. The 1D pulse programs (from Bruker pulse program library, details shown in Annex II - ^{1}H NMR pulse programs) as well as the acquisition and processing parameters used are summarized in Table 3.6.

Table 3.6 - Acquisition and processing parameters used for 1D ^{1}H NMR 500MHz and 800MHz spectra of AF. NS, number of transients; RD, relaxation delay; t_1, fixed delay; t_m, mixing time; τ, echo time; *n*, number of (τ-180°-τ) loops; $2n\tau$, total spin-spin relaxation time; δ, gradient duration; Δ, diffusion time; *g*, strength gradient pulses as % of maximum gradient output (53.5 G/cm); SW, spectral width; TD, size of FID; SI, size of real spectrum; WDW, window function; LB, line broadening.

1D ^{1}H NMR Acquisition parameters - 800MHz (*Bruker library pulse program name*)			
Standard (*zgpr*)	**Standard** (*noesypr1d*)	**CPMG** (*cpmgpr1d*)	**Diffusion-edited** (*ledbpgp2s1dpr*)
NS: 256 **RD**: 5s **SW**: 9615.39 Hz **TD**: 32k points	**NS**: 512 **RD**: 5s **t_1**: 4μs **t_m**: 100 ms **SW**: 9615.39 Hz **TD**: 32k points	**NS**: 256 or 512 **RD**: 5s **τ**: 1ms ***n***: 80 **$2n\tau$**: 160ms **SW**: 9615.39 Hz **TD**: 32k points	**NS**: 512 **RD**: 5s **δ**: 2ms **Δ**: 200ms ***g***: 70% **SW**: 9615.39 Hz **TD**: 32k points
1D ^{1}H NMR Acquisition parameters - 500MHz (*Bruker library pulse program name*)			
Standard (*noesypr1d*)		**CPMG** (*cpmgpr1d*)	**Diffusion-edited** (*ledbpgp2s1dpr*)
NS: 256 **RD**: 3s **t_1**: 3μs **t_m**: 100 ms **SW**: 8012.82 Hz **TD**: 32k points		**NS**: 256 **RD**: 3s **τ**: 350 μs ***n***: 300 **$2n\tau$**: 210 ms **SW**:8012.82 Hz **TD**: 32k points	**NS**: 256 **RD**: 3s **δ**: 2ms **Δ**: 200ms ***g***: 95% **SW**: 8012.82 Hz **TD**: 32k points
***Processing Parameters* (500 and 800MHz)**			
Standard (*zgpr*)	**Standard** (*noesypr1d*)	**CPMG** (*cpmgpr1d*)	**Diffusion-edited** (*ledbpgp2s1dpr*)
SI: 64k points **WDW**:exponential **LB**: 0.3Hz	**SI**: 64k points **WDW**: exponential **LB**: 0.3Hz	**SI**: 64k points **WDW**: exponential **LB**: 0.3Hz	**SI**: 64k points **WDW**: exponential **LB**: 1Hz

The 800MHz experiments, aimed for AF characterization, were based on plasma and serum analysis acquisition parameters obtained from literature (Foxall *et al.*, 1993; Daykin *et al.*, 2002). The acquisition parameters of the experiments performed at 500MHz, used for AF routine analysis, were re-optimized from those of 800MHz, to allow a rapid acquisition (15 min per experiment) with an adequate signal-to-noise.

Spectra processing was performed using Bruker TopSpin 2.1 (Bruker BioSpin, Reinstetten, Germany). Prior to FT, the FIDs were zero-filled to 64k points (SI, Table 3.6) and multiplied by an exponential line-broadening window functions (WDW) according to the parameters shown in Table 3.6. The 1D spectra were manually phased, baseline corrected, and the chemical shifts

referenced internally to the α-glucose H1 resonance at 5.23 ppm. Although some samples contained TSP, the resonance of this compound was not used as reference for chemical shifts due to the line-broadening resulting from binding to AF proteins.

Maternal urine ^{1}H NMR spectra were acquired at 500MHz using a standard 1D experiment (*noesypr1d*) with acquisition conditions adapted from previous reports (Beckonert *et al.*, 2007; Bernini *et al.*, 2011) with water suppression using a water presaturation pulse with 60dB power. A list of acquisition and processing parameters of ^{1}H NMR spectra of maternal urine is shown in Table 3.7.

Table 3.7 - Acquisition and processing parameters used for 1D ^{1}H NMR 500MHz spectra of maternal urine. NS, number of transients; RD, relaxation delay; t_1, fixed delay; t_m, mixing time SW, spectral width; TD, size of FID; SI, size of real spectrum; WDW, window function; LB, line broadening.

Acquisition parameters *(noesypr1d)*		**Processing parameters**
NS: 128 **RD**: 4s **t_1**: 3μs	**t_m**: 100 ms **SW**: 10080.65 Hz **TD**: 64k points	**SI**: 64k points **WDW**: exponential **LB**: 0.3Hz **Chemical shift reference:** 0.00 ppm (TSP)

Maternal urine FIDs were zero-filled to 64k points (Table 3.7) and multiplied by an exponential line-broadening window functions (WDW) prior to FT. The 1D spectra were manually phased, baseline corrected, and the chemical shifts were referenced internally to TSP resonance at 0.00 ppm.

The 2D NMR experiments TOCSY, HSQC and J-resolved were performed for AF characterization both at 500MHz and 800MHz. The acquisition parameters used were based on previous studies (Foxall *et al.*, 1993; Daykin *et al.*, 2002) (Table 3.8). The TOCSY spectra were recorded in phase-sensitive mode using States-TPPI (time proportional phase incrementation) detection in t_1 with a MLEV-17 spin lock pulse sequence (Bax and Davis, 1985), with water presaturation using pulses of 58dB or 53.8dB power, for 500 and 800MHz respectively. The HSQC ^{1}H-^{13}C spectra were acquired with inverse detection and ^{13}C decoupling during acquisition.

The acquired FIDs were zero-filled according the data points defined in SI parameters shown in Table 3.8 and multiplied by a sine-bell-squared function (qsine) prior to FT. All spectra were manually phased, automatically baseline corrected and manually chemical shifts calibrated internally using α-glucose H1 resonance at 5.23 ppm using software TopSpin 2.0 (Bruker BioSpin, Reinstetten, Germany).

Table 3.8 - 2D NMR 500/800MHz spectra experimental acquisition and processing parameters: NS, number of transients; RD, relaxation delay; SW, spectral width; TD, time domain; F1, frequency domain 1; F2, frequency domain 2; pl10, power for TOCSY spin-lock; SI, size of real spectrum; WDW, window function; LB, line broadening.

2D ^{1}H NMR Experiment (*Bruker library pulse program names: 500/800MHz*)		
^{1}H-^{1}H TOCSY (*clmlevprtp* / *mlevphpr*)	**^{1}H-^{13}C HSQC** (*invietgpsi* / *hsqcetgpsp*)	***J*-Resolved** (*jresqfpr*)
Acquisition Parameters		
NS: 16 / 24 **RD**: 2s **t_m**: 600 ms **TD[F1]**: 300 / 256 points **TD[F2]**: 2k / 8k points **SW[F1]**: 8012.82 / 9615.39 Hz **SW[F2]**: 8012.82 / 9615.39 Hz **pl10**: 15.00 / 5.70 dB	**NS**: 64 **RD**: 2s **TD[F1]**: 300 / 256 points **TD[F2]**: 2k points **SW[F1]**: 25153.81 / 32258.06 Hz **SW[F2]**: 8012.82 / 9615.39 Hz	**NS**: 8 **RD**: 2s **SW**: 9615.39 Hz **TD[F1]**: 128 points **TD[F2]**: 2k points **SW[F1]**: 50 Hz **SW[F2]**: 9615.39 Hz
Processing Parameters		
SI[F1]: 1k points **SI[F2]**: 4k points **WDW**: qsine	**SI[F1]**: 1k points **SI[F2]**: 4k points **WDW**: qsine	**SI[F1]**: 1k points **SI[F2]**: 4k points **WDW**: qsine

3.3.2.3. Diffusion Ordered Spectroscopy (DOSY) experiments

DOSY experiments were performed on AF samples (prepared with buffer A as shown in Figure 3.2, page 64) and on AF fraction with M_w<3kDa and M_w≥10kDa (Figure 3.1, page 63). Spectra were acquired at 500 MHz, at 300K using the pulse sequence for diffusion measurement *ledbpgs2spr* from Bruker library, which uses a stimulated echo, BPPLED pulses for diffusion, and two spoil gradients (Wu *et al.*, 1995). The DOSY acquisition and processing parameters are summarized in Table 3.9.

Each FID was zero-filled and multiplied by exponential line broadening functions prior to FT according to the processing parameters shown in Table 3.9. Spectra were baseline corrected and chemical shifts referenced internally to the α-glucose signal H1 resonance at 5.23 ppm or the lactate methyl resonance at 1.32 ppm (for M_w≥10 kDa AF fraction), prior to exponential fitting for diffusion coefficient calculation. Shikimic acid, an organic acid which is not found in human AF, was used as a molecular weight reference for M_w estimation, and added to M_w≥10 kDa fraction prior to spectra collection. M_w values were estimated as the inverse of the cube of the diffusion coefficient of the resolved spin-systems in DOSY spectra, assuming a globular molecular shape (Nilsson *et al.*, 2004).

Table 3.9 - Acquisition and processing parameters form DOSY 500MHz spectra of AF and AF fractions of M_w<3kDa and M_w≥10kDa: RD, relaxation delay; NS, number of transients per gradient level; δ, gradient duration; Δ, diffusion time; TD[F1], number of gradient levels; TD[F2], size of each FID; SW, spectral width; SI, size of each real spectrum; WDW, window function; LB, line broadening.

DOSY ^{1}H NMR Experiments (500 MHz)		
Amniotic Fluid (AF)	**AF M_w<3kDa**	**AF M_w≥10kDa**
Acquisition Parameters		
NS: 128	**NS**: 64	**NS**: 160
RD: 0.5s	**RD**: 0.5s	**RD**: 3s
δ: 1.1 ms	δ: 1.1 ms	δ: 1.5 ms
Δ: 150 ms	Δ: 150 ms	Δ: 250 ms
TD[F1]: 16 levels	**TD[F1]**: 32 levels	**TD[F1]**: 32 levels
TD[F2]: 32k points	**TD[F2]**: 8k points	**TD[F2]**: 32k points
SW: 6009.62 Hz	**SW**: 6009.62 Hz	**SW**: 6009.62 Hz
Processing Parameters		
SI: 32k points	**SI**: 16k points	**SI**: 32k points
WDW: exponential	**WDW**: exponential	**WDW**: exponential
LB: 2 Hz	**LB**: 2 Hz	**LB**: 2 Hz

3.3.2.4. Saturation Transfer Difference (STD) experiments

STD experiments (Mayer and Meyer, 1999) were performed in order to determine the extent of lactate binding to AF proteins in both whole AF samples and the M_w≥10kDa AF fractions. The STD spectra were acquired using the Bruker library pseudo-2D saturation transfer difference pulse sequence, *stddiffesgp.3* (Annex II - ^{1}H NMR pulse sequences) with water suppression using excitation sculpting at 600MHz with cryo-probe (courtesy of *FCT-UNL - Laboratório Associado* REQUIMTE and Portuguese NMR Network, Portugal). Each FID consisted in 64 transients acquired with 3s relaxation delay, to 32k data points and using a spectral window of 12019.22 Hz. Four spectra were gathered per experiment: 2 spectra without signal saturation (off-resonance) and 2 spectra with saturation of protein signals (on-resonance). The on- and off-resonance frequencies were 4242 Hz (in the region of NH groups of the proteins) and 8000 Hz (baseline region), respectively. The spectra were processed using TopSpin 2.1 software (Bruker BioSpin, Rheinstetten, Germany) with zero-filing to 64k data-points, multiplied by an 0.3Hz line broadening function prior to FT, manually baseline corrected and chemical shifts referenced internally to the α-glucose signal H1 resonance at 5.23 ppm. The difference spectrum was then calculated as the difference between the two summed off-resonance spectra and the two summed on-resonance spectra.

3.3.2.5. Spectral assignments

Databases of NMR spectra of known standards, such as the Metabolic Profiling Database, Biorefcode 2.0.0 (courtesy of NMR Applications, NMR division, Bruker BioSpin, Rheinsteten, Germany), the Human Metabolome Database (HMDB) (Wishart *et al.*, 2009) and the BioMagResBank (BMRB) (Ulrich *et al.*, 2008), were used to aid spectral assignments. Spectra of solutions of selected pure compounds (Table 3.4, page 62) were also acquired at 500 MHz using the standard *noesypr1d* and the same acquisition and processing parameters used for AF spectra (Table 3.6) to confirm spectral assignments.

3.3.2.6. Pre-processing and analysis of NMR data

After acquisition and processing (described in section 3.3.2.2. page 67), 1D ^{1}H NMR of AF acquired at 500MHz were organized into rectangular matrices (X matrix, Figure 3.5).

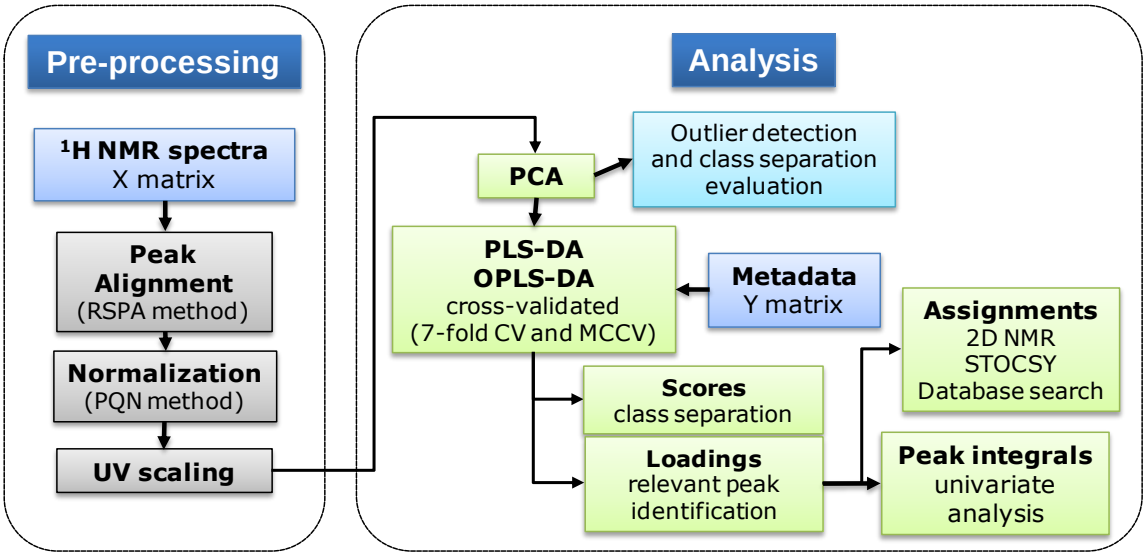


Figure 3.5 - Schematic representation of pre-processing and analysis of AF ^{1}H NMR data. RSPA: recursive segment-wise peak alignment; PQN: probability quotient normalization; UV, unit variance; CV, cross-validation; MCCV, Monte Carlo cross-validation.

The interfering water region (4.60–4.90 ppm) and the region containing the urea resonance (5.60-6.00 ppm, which is also affected by water suppression) were removed. The spectra were aligned using the segment-wise peak alignment (RSPA) algorithm (Veselkov *et al.*, 2009) and normalized using the probability quotient normalization (PQN) method in MATLAB version 7.9.0 (The Mathworks Inc., Natick, MA, USA). Prior to data analysis the X matrices were UV-scaled (each element of column was divided by the standard deviation of the column) (Figure 3.5).

Multivariate analysis was performed using SIMCA-P 11.5 software (Umetrics, Umeå, Sweden). PCA was performed followed by the PLS-DA and OPLS-DA where the sample classification and metadata were used to create sample groups. In the PLS-DA and OPLS-DA models, each disease was compared with a group of healthy control samples, and model predictive power was evaluated by Q^2 (goodness of prediction), a parameter obtained from 7-fold cross-validation (Figure 3.4). After model validation through MCCV using in-house software, the relevant metabolite signals were identified from the weight loadings plots of the PLS-DA or OPLS-DA models. The weight loadings plots consist of back-transformed weight loadings (i.e., each value from the loadings vector was multiplied by the standard deviation of each variable from the X matrix) and coloured by the variable importance in the projection (VIP) parameter. The VIP value is a measure of the magnitude of the contribution of each variable to the separation observed in the PLS-DA scores (Chong and Jun, 2005). VIP values were automatically calculated in SIMCA-P for the respective PLS-DA and OPLS-DA models. The signals identified from the weight loadings plots were assigned using 2D NMR, data base searching and eventually through STOCSY analysis (described in detail in section 3.6) (Figure 3.4). The metabolite signals were also analysed by univatiate Welch (*t*-test) or Mann-Whitney tests (Mann and Whitney, 1947) using the peak integrated areas of the ^{1}H NMR spectra. The distributions of peak integrals were also accessed with aid of box-and-whisker plots (box-plots) for group comparison (Figure 3.6).

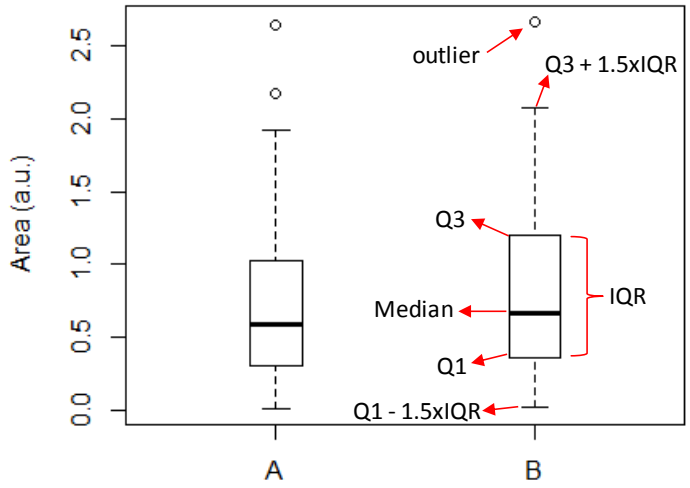


Figure 3.6 - An example of box-and-whisker representation of peak area distributions of two sample groups A and B (simulated data sets). Q1, quartile 25%; Q3, quartile 75%; IQR, inter-quartile range.

3.4. Liquid Chromatography coupled to NMR and MS (LC-NMR/MS)

3.4.1. Preparation of amniotic fluid samples for LC-NMR/MS

AF samples were directly ultrafiltered through 3kDa cut-off filters (see section 3.3.1, page 63), to remove protein content and prevent blockage of HPLC columns. About 1 mL of each AF M_w<3kDa fraction was freeze-dried and reconstituted in 100 μL of either D_2O or H_2O (without buffering) for LC-NMR/MS and LC-MS analysis respectively, in order to obtain 10-fold concentrated fractions.

3.4.2. Equipment and experimental setup

Hyphenated NMR experiments were performed for AF samples in two different NMR facilities and using three different LC-NMR/MS systems, whose setups are summarized in Table 3.10.

Table 3.10 - Description of the three LC-NMR/MS systems where the hyphenated NMR experiments were performed for AF samples.

NMR facility	System 1 *Departamento Química* *Universidade de Aveiro*	System 2 NMR Applications, NMR division, Bruker BioSpin, Rheinsteten, Germany	System 3 NMR Applications, NMR division, Bruker BioSpin, Rheinsteten, Germany
HPLC			
System	Agilent 1100	Agilent 1200	Agilent 1100
Column oven	-	yes	yes
Sample injection	manual	automatic	Automatic
UV-Vis detector	Bruker	Agilent	Agilent
NMR			
Spectrometer	Bruker Avance DRX500	Bruker Avance 600	Bruker Avance 500
Probes	LCSEI (120μL active volume)	Cryo-probe with 30 or 120μL cryo-fits	LCSEI (60μL active volume)
Interface	BPSU-36-2	BPSU-36-2	BPSU-36-2
MS			
Spectrometer	-	-	Bruker Esquire 3000
Interface	-	-	Bruker BNMI
SPE	-	Bruker Prospekt	Bruker Prospekt

The use of LC-NMR/MS systems 2 and 3 were a courtesy of NMR Applications, NMR division, Bruker BioSpin, (Rheinsteten, Germany). All LC-NMR/MS systems were composed of an HPLC systems equipped with quaternary pump and solvent degasser (Agilent Technologies, Inc. Waldbronn, Germany). LC-NMR/MS systems 2 and 3 were also equipped with an auto sampler and UV-Visible diode array detector (DAD) from Agilent, whereas system 1 was equipped with a Bruker UV-Visible DAD detector and manual sample injector (Rheodyne, CA, USA) (Table 3.10).

The HPLC systems were connected to NMR spectrometers of 500MHz (systems 1 and 3) and 600MHz (system 2) field strengths through BPSU-36-2 interface units (Bruker BioSpin, Rheinstetten, Germany). The BPSUP-36-2 units controlled the delivery of the HPLC flow to the NMR probe head and were also capable of storing up to 36 LC fractions into sample loops for post-chromatography NMR analysis (loop-storage). The NMR spectrometers of systems 1 and 2 were equipped with LC selective inverse (LCSEI) probe-heads with active volumes of 120 and 60 μL, respectively (Table 3.10). System 2 was equipped with a cryo-probe capable of accommodating cryofit flowcells of 30 and 120 μL active volume (Table 3.10). All probes were operated at 300K during spectral acquisition. Both systems 2 and 3 were connected to a solid phase extraction (SPE) interface, which enabled the trapping of LC separated compounds into hydrophobic cartridges. Finally, in system 3 the flow from the HPLC system was split into an NMR (BPSU-36-2) interface as described above and into an BNMI interface (Bruker BioSpin, Rheinstetten, Germany) which established the connection to an Esquire 3000 ion trap mass spectrometer (Bruker Daltonics, Bremen, Germany). The MS acquisition parameters used are summarized in Table 3.11. The MS spectrometer allowed also the acquisition of full MS and MS^n fragmentation spectra up to MS^4, which was set to automatic search for higher intensity peaks.

Table 3.11 - Ion trap MS acquisition parameters used in LC-NMR/MS experiments.

Condition	Values
Ionization mode	Alternating ESI^+/ESI^-
Mass range	50-1000m/z
Spray voltage	3.5kV
Desolvation temperature	350ºC
Sheath gas	N_2 at 60psi and 11L/min

3.4.3. Characterization of less polar metabolites

In order to characterize the less polar fraction of AF metabolites, chromatographic separation was performed in a reverse-phase (RP C18) column (Phenomenex, USA) at 35°C, using as a mobile phase (at pH 2.0) solution A (D_2O+ 0.1% formic acid deuterated to 90%) and solution B (acetonitrile + 0.1% formic acid deuterated to 90%), with a flow rate of 0.4mL/min. The following gradient program was used: 10-30% B from 0 to 40 min and then from 30 to 90% B from 40 to 45min. The UV detection was set at 210 nm and at ranges 200-210 and 254-280 nm. The injected sample volumes were 20μL and 100μL for on-line solid phase extraction (SPE) and fraction collection (loop-storage) respectively.

Parallel NMR and MS analysis were performed using system 3 (Table 3.10), for the selected chromatography separated fractions, using the loop-storage collection method. ^{1}H NMR spectra were recorded for each stored fraction using the *lc1pncwps* pulse sequence (a *noesypr1d* pulse sequence optimized for LC-NMR, shown in Annex II - ^{1}H NMR pulse programs) with double solvent presaturation employed to suppress both acetonitrile (ACN) and water peaks from the mobile phase. The number of transients acquired were 512, 1024, or 2048 (depending on the chromatographic peak intensity), with 16k data points, a spectral width of 10000Hz, and 0.8s acquisition time. Each FID was zero-filled to 32k, multiplied by an exponential function (1Hz line broadening), manually phased, and baseline corrected. For selected fractions, ^{1}H-^{1}H TOCSY spectra were also recorded. MS and MS^{n} spectra were acquired during the chromatographic run in positive mode and during fraction transfer in both positive and negative modes, using the mass range 50-1000 m/z.

Selected chromatographic fractions were trapped into SPE cartridges during chromatography, and were later extracted with either methanol-d_4 or acetonitrile-d_3 before transfer through a Prospekt interface into the cryo-probe head (using a cryofit flowcell with 30µL active volume). ^{1}H NMR spectra were recorded at 600.13 MHz using the same parameters used for loop-stored fractions in system 2 (Table 3.10).

3.4.4. Characterization of more polar metabolites

In order analyse the more polar compounds, the AF samples were injected onto a ION300 300×7.8mm cation-exchange column, with a particle size of 5.0µm (Interaction Chromatography Inc., San Francisco, CA, USA), a column specific for organic acids, alcohols and sugars, at a temperature of 22°C, under isocratic conditions, using a flow of 0.3mL/min with a total run time of 70 min. For LC-NMR, 100µL of each sample were injected and a mobile phase composed of 2.5mM H_2SO_4 in 100% D_2O (pH 2.4) was used. UV DAD detection was recorded at 210 nm.

The LC-NMR experiment was recorded at 600MHz (system 2, Table 3.10) using the time-slice stop-flow method, i.e., at every 30s of the chromatography run time, the flow was stopped and one 1D spectrum was acquired. The spectra were acquired with the *lc1pnf2* sequence from Bruker pulse program library (a *noesypr1d* pulse sequence optimized for LC-NMR, shown in Annex II - ^{1}H NMR pulse programs) with water presaturation at 2817.22Hz, with a total of 64 transients collected to 32k data points, using a spectral width of 12019Hz, and 1.36 s acquisition time. Each FID was multiplied by an exponential function (1Hz line broadening), manually phased, and baseline corrected. The processed 1D spectra were then reconstructed as pseudo-onflow 2D

spectrum. To confirm some of the results, repetition of 1D and 2D NMR on selected LC fractions were performed at 500MHz (system 1, Table 3.10) using the loop-storage method.

Due to experimental setup constraints, the MS detection for the ion exchange chromatography was collected separately as an LC-MS experiment using the same chromatographic conditions described above, but replacing the D_2O by ultrapure water in the mobile phase composition. The injected sample volume was also reduced to 20µL. Mass spectra were collected in both positive and negative modes, in the mass range of 90-1000 m/z, with automatic fragmentation of the most abundant ions (up to MS^4).

3.5. Ultra-Performance Liquid Chromatography - Mass Spectrometry (UPLC-MS) Analysis

3.5.1. Sample preparation for UPLC-MS

Amniotic fluid and maternal urine samples used in UPLC-MS profiling analysis were thawed at room temperature (25°C) and gently mixed. After thawing, the samples were prepared according to the schemes shown in Figure 3.7.

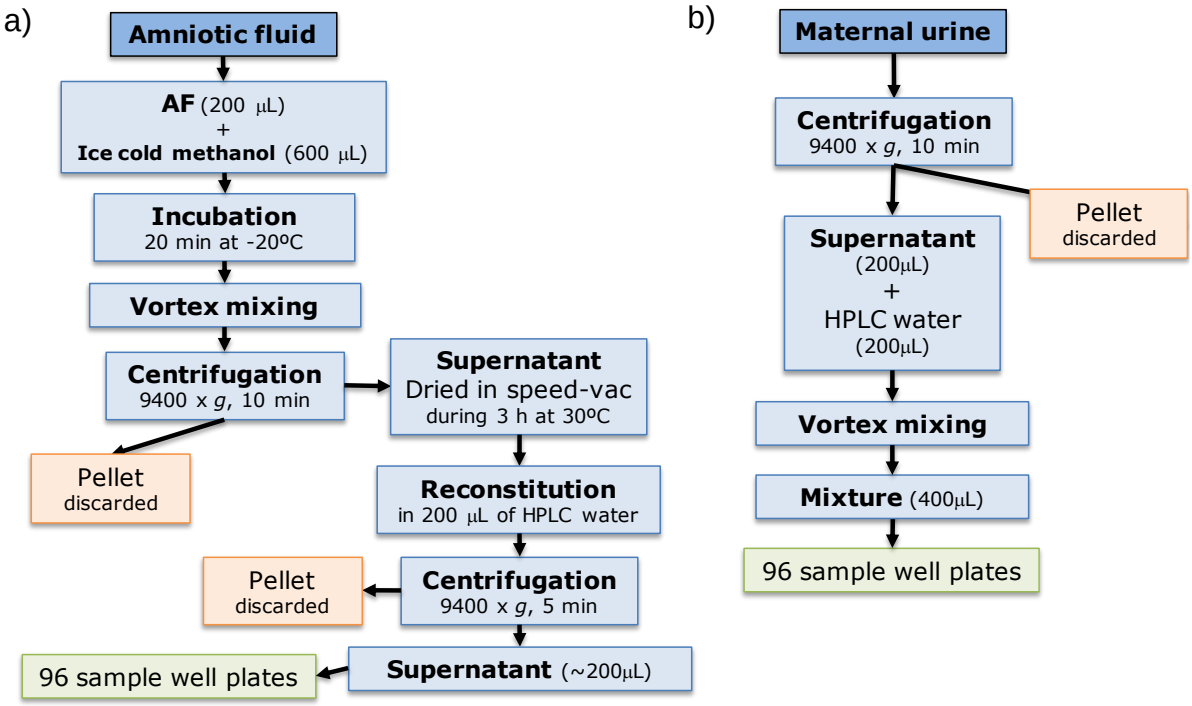


Figure 3.7 - Sample preparation procedures used prior to UPLC-MS analysis: a) Amniotic fluid and b) maternal urine samples.

In order to remove proteins which would otherwise damage the UPLC columns, AF samples were precipitated with ice-cold methanol (Figure 3.7a). After centrifugation, the supernatants of AF extracts were concentrated in a speed-vac sample concentrator (Eppendorf, Hamburg, Germany) at 30°C for 3 hours (Figure 3.7a). The dried samples were then reconstituted in HPLC grade water, sonicated in an ultrasound bath at room temperature (5 min), centrifuged (9400×*g*, 10 min) and transferred into 96 well plates.

Maternal urine samples were centrifuged (9400×*g*, 10 min), then diluted 1:1 with HPLC grade water and mixed after being transferred into 96 well plates (Figure 3.7b). Maternal urine and AF samples were kept inside the auto-sampler of the UPLC system during the analysis, which was maintained at a constant temperature of 4°C.

3.5.2. Equipment and experimental setup

UPLC-MS analysis was performed courtesy of Biomolecular Medicine Department, Division of Surgery and Cancer of the Faculty of Medicine, Imperial College of London. Profiling experiments were performed in Acquity UPLC systems (Waters Ltd. Elstree, UK) connected to an LCT Premier mass spectrometer (Waters MS Technologies, Ltd., Manchester, UK) or connected to a Q-TOF Premier mass spectrometer (Waters MS Technologies, Ltd., Manchester, UK). Fragmentation MS experiments (MS/MS on selected m/z and MS^E) were performed using an Acquity UPLC system (Waters Ltd. Elstree, UK) coupled to a Q-TOF Premier mass spectrometer (Waters MS Technologies, Ltd., Manchester, UK). Mass spectrometry was performed in both positive and negative ion electrospray (ESI^+ and ESI^-) modes separately, meaning that each sample batch was run in UPLC twice, one for each ESI mode. The MS parameters used are summarized in Table 3.12.

Table 3.12 - Acquisition parameters of the time-of-flight mass spectrometers (TOF-MS) used in UPLC-MS experiments.

Condition	Values
Capillary voltage	3.2kV (ESI^+) / 2.4kV (ESI^-)
Source temperature	120ºC
Desolvation temperature	350ºC
Desolvation gas flow	900L/h
Operation mode	V optics
Mass range	50-1000m/z
Acquisition mode	centroid
Acquisition rate	0.1s (0.01s interscan delay)

A solution of the peptide leucine enkephalin (monoisotopic mass = 555.2693 Da) with concentration 200pg/µL (in 50:50 ACN:H_2O) was infused into the mass spectrometers at 3µL/min via an auxiliary sprayer to provide a lock mass signal, which was registered every 15s and

averaged over 3 scans to perform mass correction. For fragmentation analysis (MS/MS and MS^E) collision energy ramps of 10-50 Volts were used.

Two different chromatographic columns were used due to their complementary performances in achieving a more complete characterization of a biofluid metabolic profile: High-Strength Silica (HSS) (Acquity BEH HSS 1.8 µm, 2.1 × 100 mm, Waters Corporation, Milford, U.S.A.), which shows improved retention for polar compounds over regular reversed phase (RP) columns (Masson *et al.*, 2011) and an Hydrophilic Interaction Chromatography (HILIC) (Acquity BEH HILIC, 1.7 µm, 2.1 × 100 mm, Waters Corporation, Milford, U.S.A.), which retains polar compounds, particularly those of basic nature (Gika *et al.*, 2008).

Gradient elution was used for both HSS and HILIC columns. For the HSS column a flow rate of 0.5mL/min was used with a gradient of the following mobile phases: water with 0.1% formic acid (A_{HSS}) and acetonitrile (ACN) with 0.1% formic acid (B_{HSS}) (Figure 3.8a). For the HILIC column a flow rate of 0.4mL/min was used and the mobile phases consisted in two solutions: 95% Ammonium acetate 10mM / 5% ACN with 0.1% formic acid (A_{HILIC}) and 50% Ammonium acetate 10mM / 50% ACN with 0.1% formic acid (B_{HILIC}), eluted according to the gradient shown in Figure 3.8b. All solvents and reagents were of HPLC grade or equivalent and both columns were constantly maintained at 40°C during the analysis.

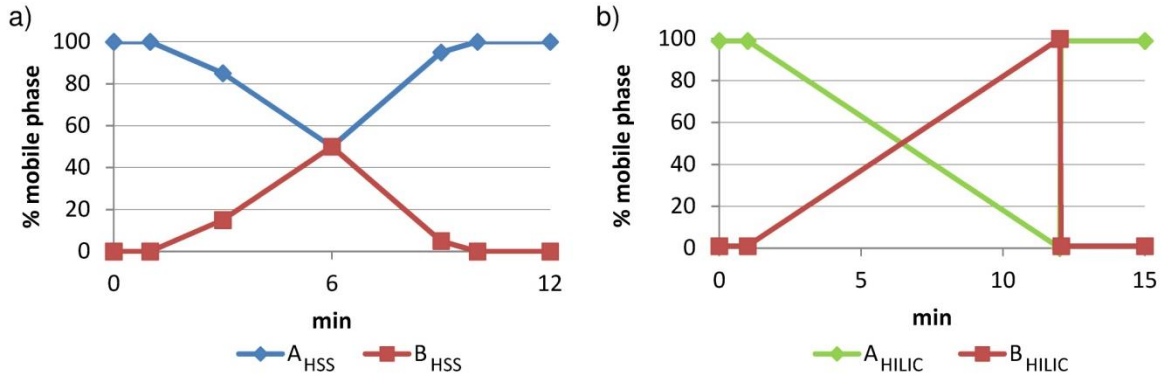


Figure 3.8 - Time profiles of the UPLC elution gradients: a) HSS column; b) HILIC column. A_{HSS}: H_2O+0.1% formic acid; B_{HSS}: ACN + 0.1% formic acid; A_{HILIC}: 95% Ammonium acetate 10mM/5% ACN+0.1% formic acid; B_{HILIC}: 50% Ammonium acetate 10mM / 50% ACN + 0.1% formic acid.

A sample volume of 5µL was injected twice per sample for each column (for positive and negative acquisition modes) in a randomised order. At the beginning of each run, a batch of 10 injections (5µL each) of a quality control (QC) sample was used to condition the column. The QC samples (of both AF and urine) consisted of mixtures of the control samples prepared in the same way as the study samples. The QC samples were repeatedly injected, after every 10 samples, in order to assess analytical reproducibility.

3.5.3. Pre-processing and analysis of UPLC-MS data

The pre-processing and analysis workflow used for the UPLC-MS data sets is summarized in Figure 3.9. The raw UPLC-MS chromatograms were converted to netCDF format using Waters MassLynx 4.1 Databridge software (Waters Corporation, Milford, USA), and then processed using the freely available software package XCMS (Smith *et al.*, 2006) in R 2.11.1 software (R Development Core Team, Vienna, Austria, 2011) using the *centWave* peak picking algorithm (Tautenhahn *et al.*, 2008) (Figure 3.9).

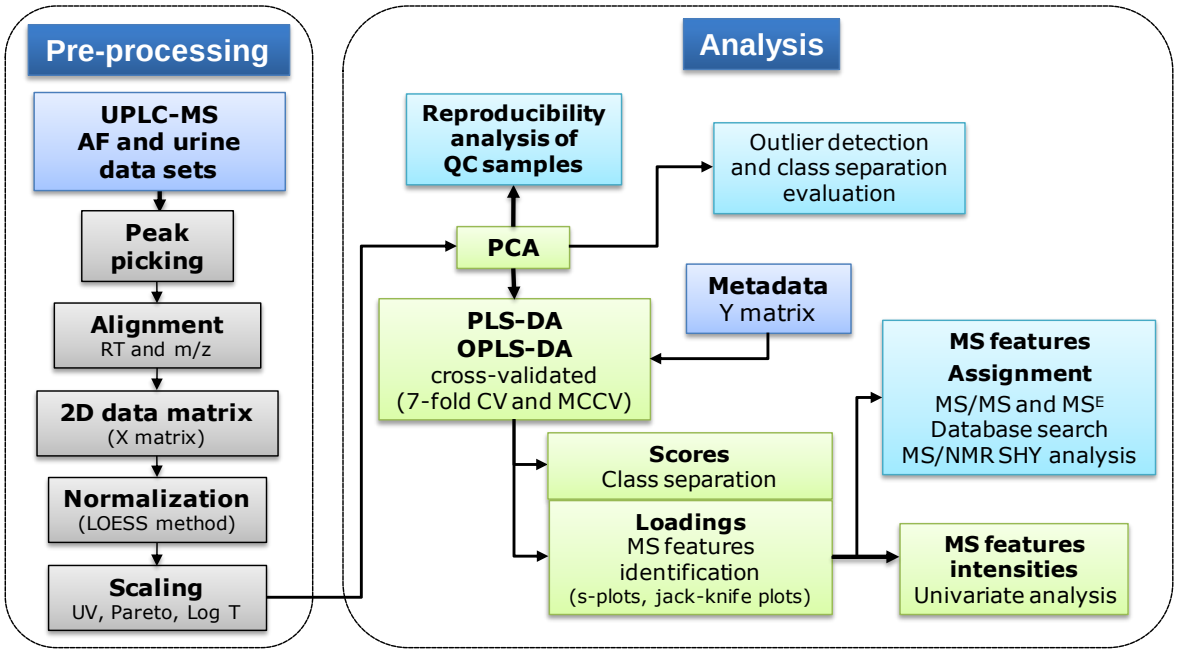


Figure 3.9 - Processing and analysis workflow of the AF and maternal urine UPLC-MS data sets. LOESS, locally weighted scatter plot smoothing; UV, unit variance; Log T, logarithmic transformation; CV, cross-validation; MCCV, Monte Carlo cross-validation; SHY, statistical heterospectroscopy.

Correction of the inter-sample peak retention time (RT) deviations was also performed using XCMS (Figure 3.9). Sample data sets and QC samples injected between runs were normalized using locally weighted scatter plot smoothing (LOESS) (Veselkov *et al.*, 2011) using R 2.11.1 software.

The normalized UPLC-MS data sets were arranged into a rectangular matrix (X matrix), where each element corresponded to the intensity of a retention time - m/z pair (column) for a given sample (row). The UPLC-MS data sets were subjected to scaling prior to mean centring using UV scaling, Pareto scaling, logarithmic transformation (Log transformed) and analysed by MVA (Figure 3.9) using SIMCA-P 11.5 software (Umetrics, Umeå, Sweden). Only the RT range of 1-8 min

was considered when the data sets were analysed, which corresponded to the range where sample metabolites eluted.

Contrary to NMR, which is a highly reproducible platform for large scale untargeted metabolite profiling studies (>100 samples), the reproducibility of MS platforms has to be evaluated and ensured throughout the analysis (Dunn *et al.*, 2012). The reproducibility of both AF and urine data sets was assessed through the analysis of QC sample distribution with the aid of PCA. The results of the reproducibility evaluation are presented in Annex III - UPLC-MS reproducibility evaluation.

For the PLS-DA and OPLS-DA models only controls *versus* disorders were considered. After validation by MCCV, the loadings from OPLS-DA were evaluated from S-plots of the latent variable most relevant for class separation. The S-plot combines covariance and correlation between the UPLC-MS variables (RT_m/z pairs) of the X data matrix and PLS-DA scores. A correlation threshold of $|r|>0.5$ were considered in the S-plot analysis. The relevant features were selected based on the ratio of absolute mean covariance loading (|p|) and its respective jack-knifed validated standard error (pcvSE) (Wiklund *et al.*, 2007).

3.6. Fourier-transform Mid Infrared (MIR) spectroscopy analysis of amniotic fluid

3.6.1. Equipment and experimental setup

The MIR spectrometer used was a PerkinElmer BX FT-IR spectrometer (PerkinElmer Inc., USA), equipped with a Golden-Gate single reflection ATR, with a ZnSe crystal (Specac Ltd, Slough, UK).

Spectra were acquired at 26±1°C and 38±2% relative humidity, in the range of 4000-600 cm^{-1} with 64 scans and 4 cm^{-1} resolution.

3.6.2. AF sample preparation and MIR spectra acquisition

The AF samples were thawed at room temperature and one drop of each AF sample (5μL) was applied on the ATR crystal. Spectra were consecutively collected during sample drying (at room temperature of 26±1°C) in which the AF sample forms a homogeneous film on top of the ATR crystal (Figure 3.10). The drying step is intended to remove excess of water which produces a strong absorption background and hampers the reproducible spectroscopy in the MIR region (Martin *et al.*, 2004; Petrich *et al.*, 2009).

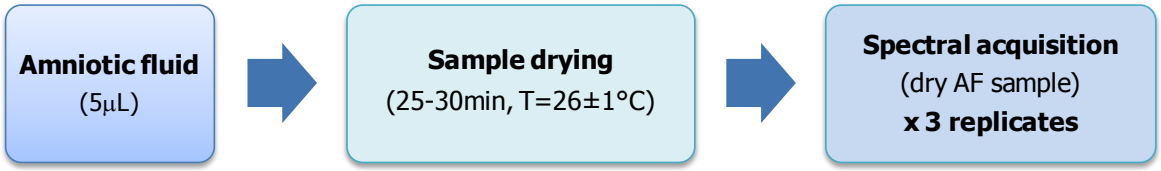


Figure 3.10 - Schematic procedure of acquisition of MIR spectra of AF samples.

Samples were considered dry after no visual differences between consecutive spectra were noted, usually after 25 to 30 min after sample application. A typical sequence of MIR spectra of AF recorded as a function of sample drying time is shown in in Annex IV - MIR reproducibility evaluation. After this, a set of 3 spectra were acquired (replicates) for each sample (Figure 3.10). In order to evaluate analytical reproducibility, a control sample was divided into 20 aliquots of 20µL and their MIR spectra were acquired between study samples (QC, quality control sample).

3.6.3. Pre-processing and analysis of MIR data

A rectangular matrix (X) was built using the 3 replicate spectra recorded for each dry sample, excluding the CO_2 absorbance region (2400-2200cm^{-1}) (Figure 3.11). The matrices used for correlation analysis contained only one spectrum replicate per sample.

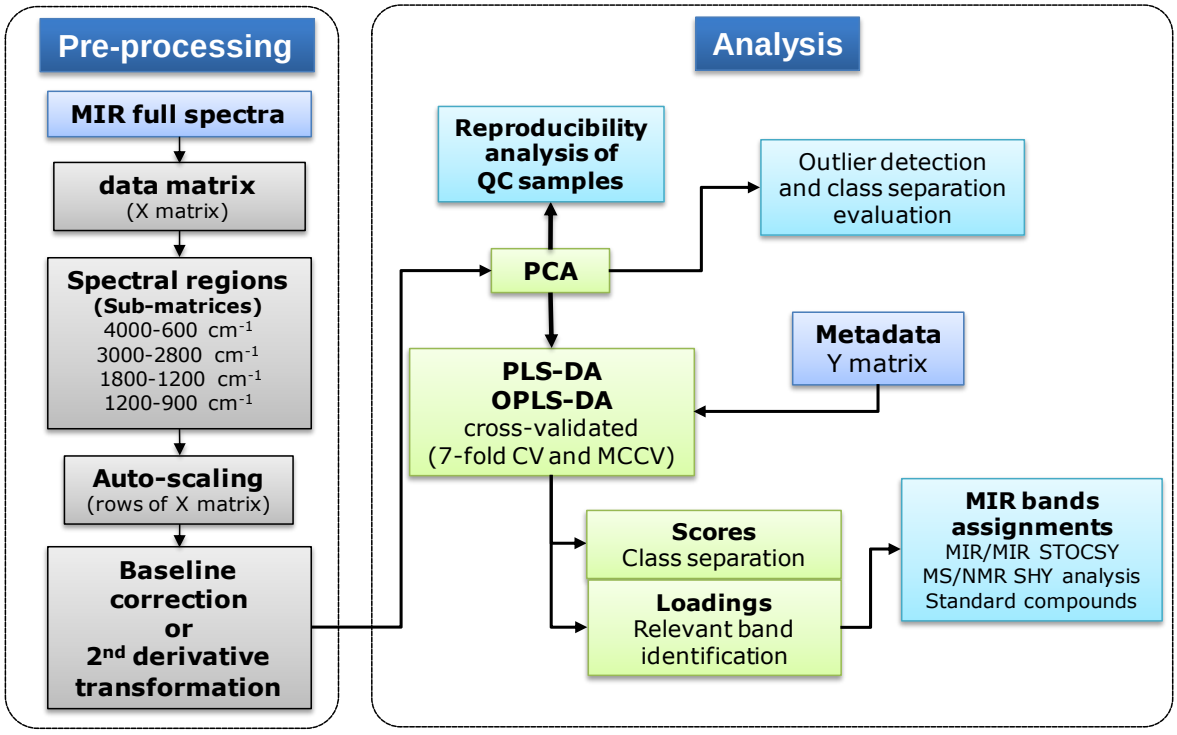


Figure 3.11 - Schematic procedure of acquisition and pre-processing and analysis of AF by MIR. CV, cross-validation; MCCV, Monte Carlo cross-validation; SHY, statistical heterospectroscopy.

The MIR matrices were split into smaller matrices (sub-matrices) corresponding to spectral regions: 3000-2800 cm^{-1}, 1800-1200 cm^{-1} and 1200-900 cm^{-1} (Figure 3.11) Each matrix was mean centred and normalization using standard normal variate (SNV) transformation (each intensity of a row, i.e. one spectra, was divided by the standard deviation of the intensities of the row) using R 2.11.1 software (R Development Core Team, Vienna, Austria, 2011) and baseline corrected using the software Cats (Barros, 1999). Second derivatives were obtained for the spectral regions considering 5 points windows using SIMCA-P 11.5 (Umetrics, Umeå, Sweden) (Figure 3.11).

All matrices were mean centred prior to MVA. The reproducibility of MIR data sets was assessed through the analysis of QC sample distribution with the aid of PCA, the results being presented in Annex IV - MIR reproducibility evaluation. PLS-DA models were obtained for controls *versus* disorder, being also subsequently evaluated through MCCV (Figure 3.11).

To aid spectral assignment (Figure 3.11), literature on MIR AF spectra was consulted (Liu *et al.*, 1998a; Liu *et al.*, 1998b; Liu and Mantsch, 1999; Liu *et al.*, 2000; Liu *et al.*, 2002), as well as tables on MIR spectra assignments of organic compounds (Pretsch *et al.*, 1989). Spectra of aqueous solutions of selected standards of 0.1M and pH 8.30±0.30 were also acquired, after drying, using the same acquisition conditions as AF samples. Some standards were acquired in their solid state using the Golden-Gate apparatus already described. The standards whose MIR spectra were acquired are indicated in Table 3.4 (page 62).

3.7. Statistical total correlation (STOCSY) and statistical heterospectroscopy (SHY) analysis

STOCSY analysis of ^{1}H NMR was performed using the 1D version of STOCSY, which consists in representing the covariance between the vector of intensities from a peak of interest (driver peak) and the full spectrum matrix, with the corresponding Pearson correlation coefficients (r) being projected as colour on top of the covariance (Cloarec *et al.*, 2005). The 1D STOCSY plots were obtained using all correlation coefficients calculated, i.e. no threshold was applied, to facilitate visualization of the correlated peaks.

STOCSY was also applied to obtain the MIR intra-correlation maps, using the full spectrum matrices. Pearson correlations coefficients, r, were calculated, and only those with $|r|>0.8$ with $p<0.001$ significance level were considered.

Inter-correlation maps of MIR/NMR were obtained using statistical heterospectroscopy (SHY) (Crockford *et al.*, 2006). The matrices of MIR spectral regions were correlated to the full ^{1}H NMR

standard spectra matrices. Only the Pearson correlation coefficients with $|r|>0.7$ and $p<0.001$ significance level were considered in the SHY maps. The same criterion was used for MS/NMR SHY analysis. However, the results were expressed as 1D STOCSY graphical representations of the covariance and correlation coefficients calculated between the matrices containing intensities of UPLC-MS features and the matrix containing full ^{1}H NMR CPMG spectra. The CPMG spectra (which filter high M_w compound signals) were used instead of the standard spectra (containing signals from both high and low M_w compounds) because UPLC-MS conditions were chosen to reflect mostly low M_w metabolites (section 3.5.1, pages 77-78).

Special mention has to be made for ^{1}H NMR maternal urine data, which was only considered for SHY analysis with the corresponding UPLC-MS data. The analysis of the ^{1}H NMR maternal urine data was performed as part of another PhD thesis and the data used here had been reported before (Diaz *et al.*, 2011). The maternal urine ^{1}H NMR spectra were aligned and normalized using the same procedure as that used for AF ^{1}H NMR spectra (section 3.3.2.6, page 72). The correlations were evaluated through scatter-plots, and those containing strong deviations from linearity or affected by the presence of outliers were not considered. All calculations and plots were performed using MATLAB version 7.9.0 (The MathWorks Inc., Natick, MA, USA).

4. Study of 2nd trimester human amniotic fluid by ^{1}H NMR spectroscopy: stability evaluation and compositional analysis

The first section of the present chapter is devoted to the qualitative evaluation of different protocols for AF preparation for ^{1}H NMR spectroscopy in order to establish a protocol for ^{1}H NMR metabolic profiling and AF characterization. In metabolic profiling, it is important to maintain sample integrity throughout the whole study by assuring that adequate sample storage and handling conditions are employed. While the effects of stability over storage and sample handling have been studied in some detail for biofluids such as serum, plasma and urine (Teahan *et al.*, 2006; Saude and Sykes, 2007; Barton *et al.*, 2008), similar systematic studies were, to the best of our knowledge, lacking for AF samples. Therefore, in section 4.2 the effects of storage under -20° and -70°C temperatures, freeze-thawing and storage at room temperature on human AF composition were evaluated by 1D ^{1}H NMR (Graça *et al.*, 2007). Once suitable sample preparation protocol was defined, comprehensive characterization of human AF was performed (sections 4.3 and 4.4) by high-resolution 1D ^{1}H NMR and 2D NMR spectroscopy (Graça *et al.*, 2007), and LC-NMR/MS methods (Graça *et al.*, 2008), thus extending previous knowledge of AF composition as viewed by ^{1}H NMR (Sims *et al.*, 1993b; Bock, 1994; McGowan *et al.*, 1999; Groenen *et al.*, 2004). The work described in this chapter establishes the ground for ^{1}H NMR based metabolic profiling studies of human AF studies described in the next chapters 5 and 6.

4.1. Comparison of different protocols for amniotic fluid analysis by ^{1}H NMR: establishment of a protocol for metabolic profiling

Effect of freeze-drying

Qualitative differences were noted between the spectra of an AF sample prepared by direct dilution (i.e., 500μL of AF diluted with 50μL D_2O/TSP) (Figure 4.1a) and the spectra of the same sample after freeze-drying and reconstitution in D_2O (1.3-fold concentrated) (Figure 4.1b). Indeed, freeze-drying the sample seemed to cause loss of acetone and/or acetoacetate, ethanol and urea signals (Figure 4.1a). Moreover, other changes seemed to have occurred due to freeze-drying such as the apparent reduction of pyruvate and formate signal intensities, the broadening of citrate peaks intensities and increases in intensities of broader signals from proteins, when compared with the diluted sample (Figure 4.1).

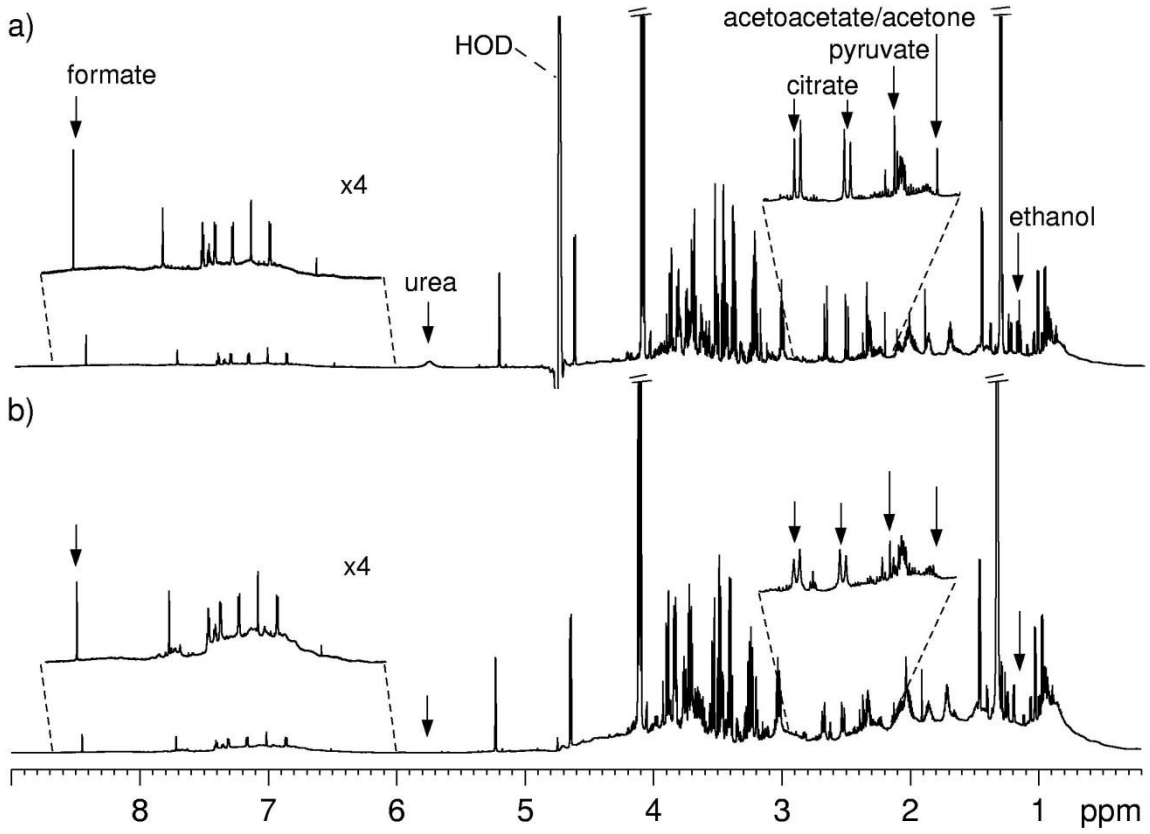


Figure 4.1 - Standard 1D ^{1}H NMR 800 MHz spectra of a human AF: a) diluted with D_2O and b) freeze-dried reconstituted in D_2O and 1.3-fold concentrated. The signals affected by freeze-drying the sample are indicated in with arrows; HOD, residual water peak.

The total or partial loss of the more volatile compounds acetone, acetoacetate, formate, pyruvate and ethanol was expected in upon freeze-drying the sample. In fact similar compound losses have been observed after freeze-drying in human urine samples (Lauridsen *et al.*, 2007). The disappearance of the urea signal in the freeze-dried sample resulted from the slow exchange the NH protons with the deuterium from the solvent since the sample was reconstituted in 100% D_2O solution. Upon reconstitution, the freeze-dried samples were 1.3 fold concentrated, which justifies the higher intensities noted for protein signals (Figure 4.1b). The broadening and intensity reduction of the citrate resonances (Figure 4.1a) may have resulted from binding to proteins, as observed previously in plasma samples (Daykin *et al.*, 2002). Because of the observed compound loss without a significant increase in concentration, freeze-drying process is not recommended for routine analysis of AF. Instead direct sample analysis of AF after dilution with D_2O to provide NMR field lock reference seems the more adequate choice.

The effect of sample pH

Another important experimental issue to take into consideration is the pH at which the AF sample spectra are collected. In fact the chemical shifts are dependent on pH of the sample (Fan, 1996). Since the human AF samples analysed in this work registered natural pH variability, with values ranging from 6.30 to 8.75 with an average value of 8.15±0.34 (mean±standard deviation), the effect of pH on AF spectra was investigated. Figure 4.2 shows 3 spectral expansions of spectra acquired for one AF sample after adjusting its pH values to 7.00, 8.00 and 9.00.

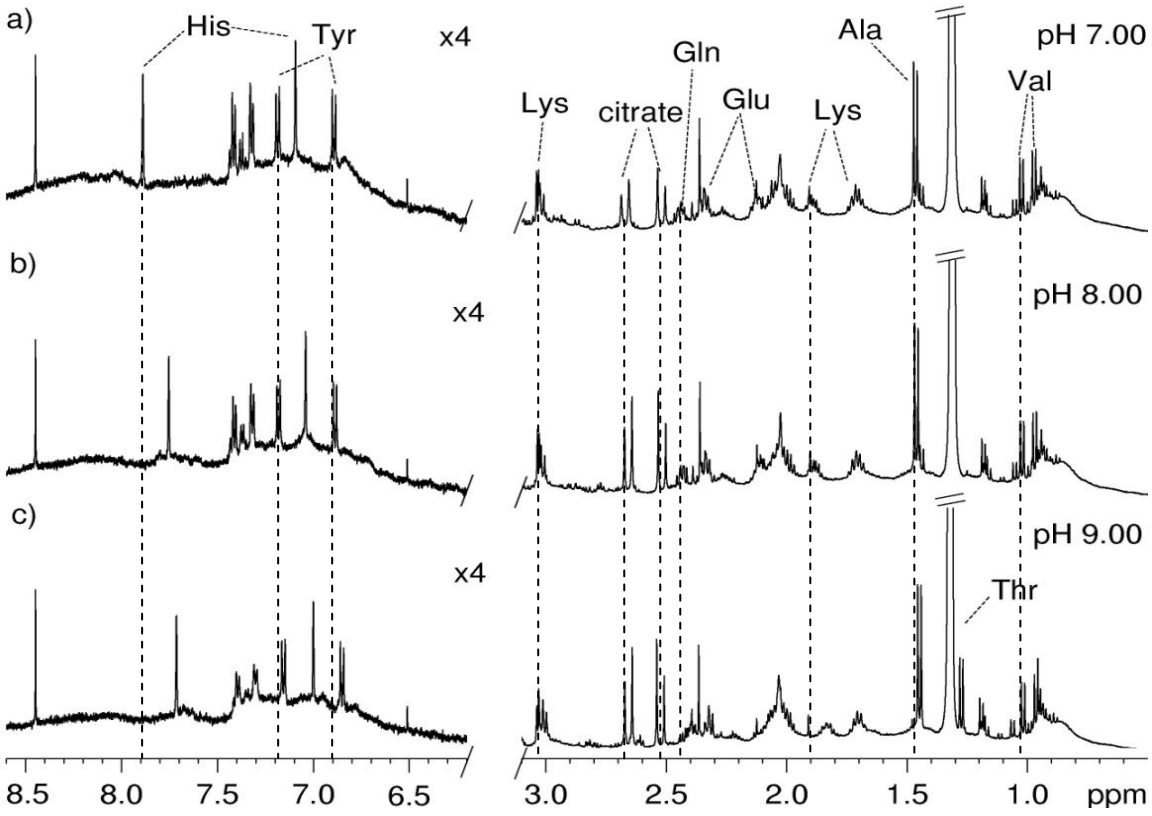


Figure 4.2 - Spectral expansions of standard 1D ^{1}H NMR 500 MHz spectra obtained for the same AF sample with pH adjusted to: a) pH 7.00, b) pH 8.00 and c) pH 9.00; vertical dashed lines indicate the changes noted in chemical shifts of the indicated resonances; amino acids indicated with 3 letter code.

The chemical shifts of resonances from histidine, tyrosine, lysine, glutamine, glutamate, alanine, threonine, valine and citrate varied with pH (Figure 4.2), which is expected since these compounds contain protons near acidic groups and/or basic groups prone to differential protonation (Fan, 1996). In addition, the signals of citrate also experienced broadening from pH 9.00 to pH 7.00, which may be due to an increased binding of citrate to AF proteins (such as albumin) and metal ions (such as Fe^{3+}) as a result from pH lowering, as previously suggested for human plasma where similar effects occurred (Daykin *et al.*, 2002).

It is common to perform buffering at around pH 7.0 in biofluids such as urine in order to minimize chemical shift variations, since at this pH values relatively few metabolites are expected to have major chemical shift variations (Lindon *et al.*, 2000a). Therefore, a buffering procedure was adapted from a previous study on urine (Solanky *et al.*, 2003) and tested for AF samples. As described in chapter 3 (section 3.3.1, page 63), the procedure consisted in the addition of 100μL of sodium phosphate 0.5M buffer solution at pH 7.20 to each sample. After buffering, the samples tested registered pH values in the interval 7.32±0.07. Hence, buffering at pH 7.2 was selected for routine sample preparation in ^{1}H NMR profiling analysis described in this work.

Effect of ultrafiltration

Subfractionation of AF by ultrafiltration was also used to obtain fractions differing in M_w (Figure 3.1, chapter 3, page 63). The resulting 1D ^{1}H NMR spectra of whole AF sample and the respective M_w<3 kDa and high M_w≥3 kDa fractions are presented in Figure 4.3a, b and c, respectively.

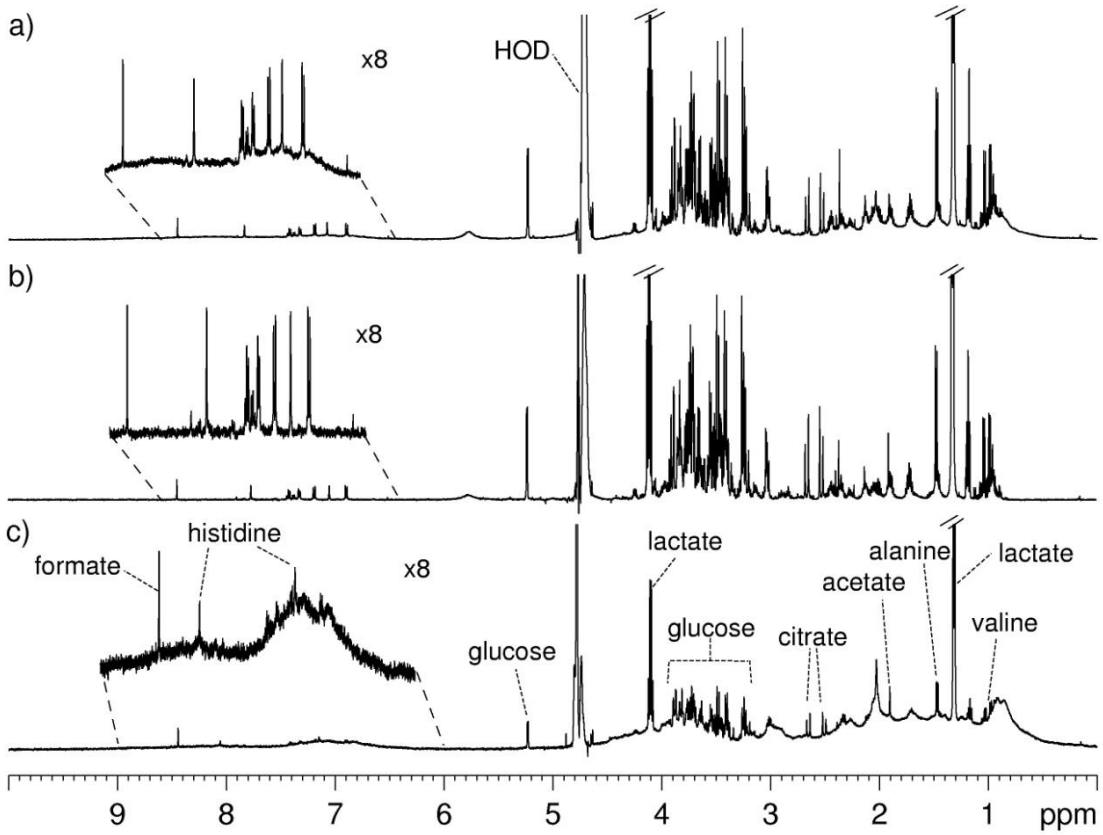


Figure 4.3 - Standard 1D ^{1}H NMR 500 MHz spectra of: a) AF diluted sample and the corresponding fractions with b) M_w<3kDa and c) M_w≥3kDa, all registering pH 7.40.

The broad signals originating from proteins, which were overlapped with sharp signals from low M_w compounds seen in the intact sample spectrum (Figure 4.3a), were lost after filtration as expected, which resulted in spectra composed only of sharp peaks from small molecules for M_w<3 kDa fraction (Figure 4.3b). The spectra of the latter, as it will be shown in later in this chapter (section 4.3.1), is very similar to the T_2-edited spectrum (CPMG) of the intact AF sample. However, some low M_w compounds signals, namely acetate, lactate, citrate, alanine, valine, glucose, histidine and formate, were still present in the M_w≥3 kDa fraction, indicating incomplete filtration and/or possibly strong binding to proteins. As a consequence, quantitative measurements using solely the low M_w fraction, as those reported by Groenen *et al* (2004), may not be representative of the exact concentration of a portion of metabolites inspected. The procedure of ultrafiltration is therefore very useful for qualitative analysis, but may be unsuited for quantitative purposes.

4.2. Study of the human amniotic fluid sample stability

4.2.1. Stability during storage

In order to identify any possible changes occurring during storage, the composition of 3 control AF samples (A, B and C) stored at -20°C and -70°C was monitored by ^{1}H NMR monthly, during a 6 months period (as illustrated previously in Figure 3.3a, page 65). The sample pH was also followed during the storage period, and the increases of ΔpH = 1.15 ± 0.34 and ΔpH = 0.58 ± 0.28 registered for -20°C and -70°C storage temperatures, respectively. PCA scores scatter plot of the spectra from the stored samples at -20°C and -70°C are presented in Figure 4.4.

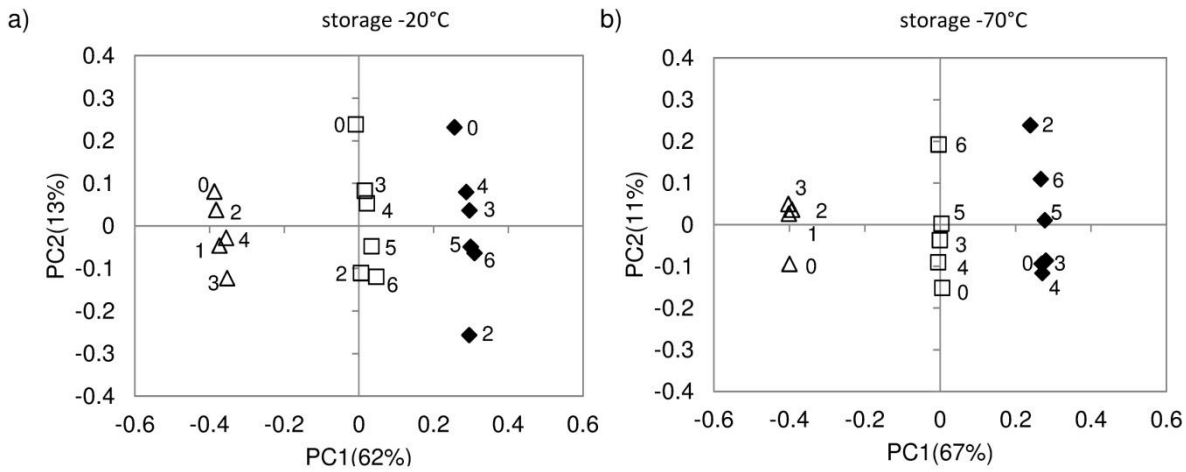


Figure 4.4 - PCA scores scatter plots obtained for control AF samples stored during 6 months (represented by numbers 0-6) at: a) -20°C and at b) -70°C; symbols represent the 3 different from the 3 different sample donors: Δ (A), □ (B) and ♦ (C).

As shown in both PCA scores scatter plots, the 3 different sample groups were clearly distinguished, reflecting the different inter-individual sample composition (Figure 4.4). There was however separation in within each sample group, characterized by dispersion of the data points in the PC2 axis in both storage temperatures (Figure 4.4). It is worth mentioning that samples from both -20 and -70°C groups corresponding to 1 month (sample B and C) as well as 5 and 6 months of storage (sample A) were removed from the analysis, as they were strong outliers in the PCA scores (not shown), most probably due to instrumental bias (field inhomogeneity) during spectral acquisition. Moreover, the -70°C sample corresponding to 4 months of storage could not be acquired. Nevertheless, a close inspection of the scores plots reveals that the dispersion along PC2 was however slightly more evident for -20°C samples (Figure 4.4a). The dispersion within each group indicated underlying compositional changes occurred over the storage period, which became apparent after visual inspection, as shown in Figure 4.5 for storage at -20°C.

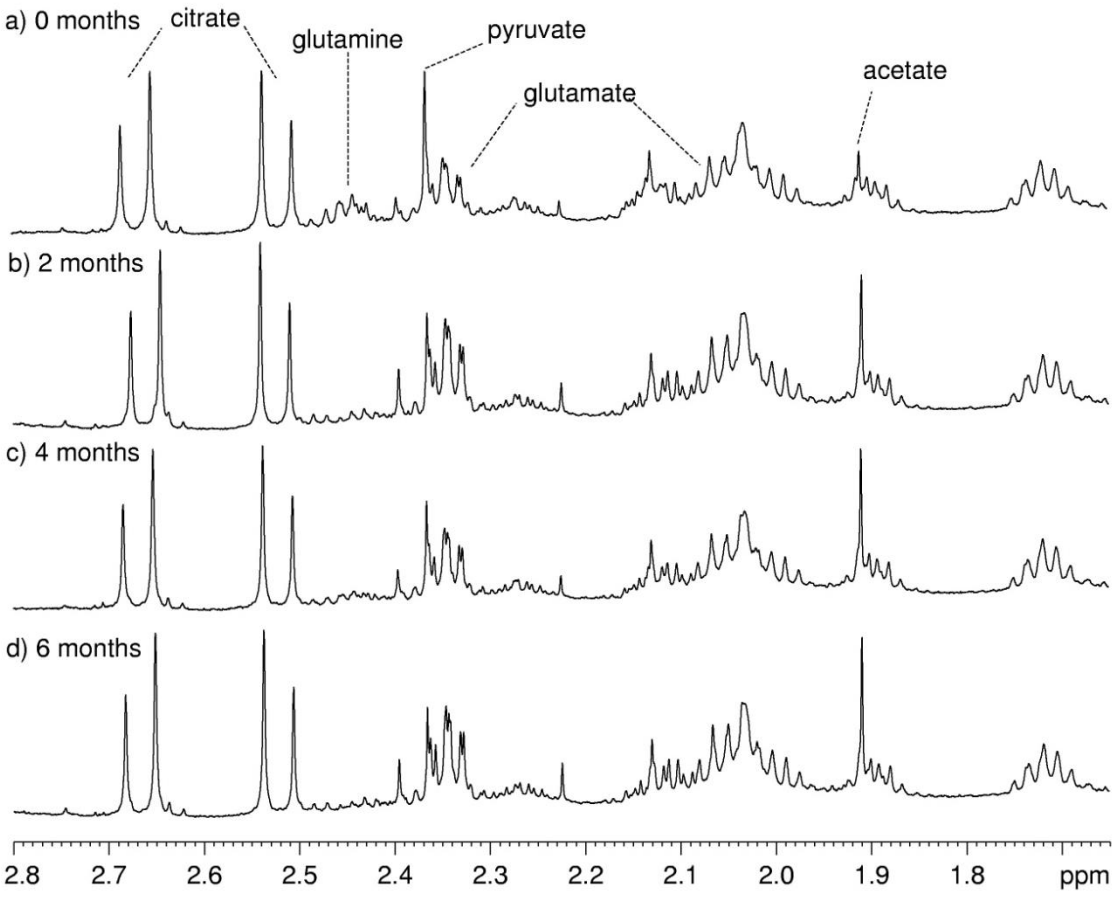


Figure 4.5 - Spectral expansions (1.5 - 2.8 ppm) of standard 1D ^{1}H NMR 500 MHz spectra obtained for AF sample C stored at -20°C acquired at: a) 0 months; b) 2 months; c) 4 months; d) 6 months.

These variations included the increases in signals of acetate and glutamate and decreases in the signals of glutamine, pyruvate as well as subtle decreases in the signals of citrate (Figure 4.5). The ratios between the integrals of the peaks of acetate, pyruvate, glutamine, glutamate and citrate at each time point (month) and the integrals at 0 month (before storage) were obtained and plotted as function of storage time (Figure 4.6). It may be seen that most of the variations observed at -20°C seem to occur during the first two months of storage, then stabilizing throughout the remaining analysed period of time. The only exception was noted for the variation of acetate which seems to increase slowly up to 6 storage months (Figure 4.6a). Comparatively, only small changes occurred at -70°C for the same metabolites, with only apparent small increase in acetate and decrease of pyruvate integral ratios being noted in the after one month of storage, then returning to initial values (Figure 4.6b).

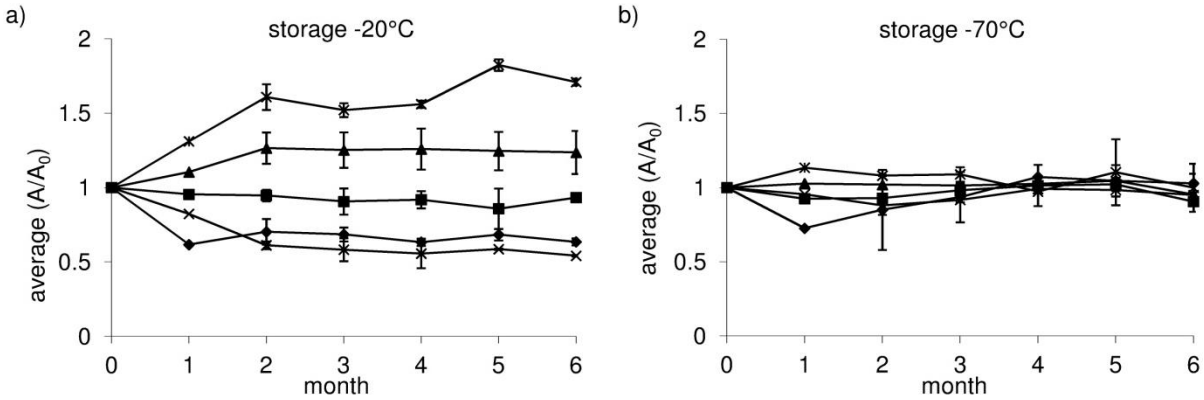


Figure 4.6 - Variations in the signals of acetate (*, 1.91 ppm), citrate (■, 2.52 ppm), glutamate (▲, 2.34 ppm), glutamine (×, 2.46 ppm) and pyruvate (♦, 2.37 ppm), occurring during sample storage at the temperatures a) -20°C and b) -70°C. Normalized plots of the ratio A/A_0 (A_0 is the integral of the peak at t = 0 months).

Increases in acetate have been noted for other biofluids namely in urine during long-term storage (up to 6.5 months) at 4°C, but not seen at -25°C (Lauridsen *et al.*, 2007). Acetate increase has been interpreted as signs of microbial activity (Lauridsen *et al.*, 2007). However, the bacterial activity in carbohydrate rich media (such as AF), would lead to an increase in acetate and concomitant decrease in carbohydrates (such as glucose) (Saude and Sykes, 2007). This was not observed in AF, meaning that changes in increases in acetate contents had a different origin. Similarly, in rat plasma slight increases in acetate and decreases in pyruvate were observed after 1 week at 4°C, which were not prevented with the addition of an anti-bacterial agent sodium azide being therefore not attributed to bacterial activity (Deprez *et al.*, 2002). Glutamate and glutamine changed in opposite directions at -20°C, which may indicate to the possible degradation of the later by loss of an NH_2 group to form glutamate and NH_3. The conversion of

NH_3 to NH_4^+ in solution may have led to the increase in pH noted during storage. Similar decreases in amino acids arginine and glutamine were observed in rat and human plasma stored at -20°C during 6 months, but were not observed after plasma deproteinization prior to storage, which indicated the presence of enzymatic activity at -20°C (Van Eijk *et al.*, 1994). These variations were not observed when samples were stored -70°C which led to the conclusions that freezing at temperatures lower than -20°C was necessary in order to inhibit enzymatic activities (Van Eijk *et al.*, 1994). Similarly, the metabolite variations registered in AF samples possibly have resulted from enzymatic activity that may have occurred at -20°C but were significantly reduced at -70°C.

Therefore, the use of lower storage temperatures such as -70°C or lower is strongly recommended in order to minimize sample compositional alterations. This is not always possible, especially at the place of sample collection, where -70°C freezers may not be readily available, and -20°C are much more common. As most of the changes in human AF occurred in the first two months of storage at -20°C, in cases where lower temperature storage is not possible, it is strongly recommend that analysis should be performed within a few days after collection.

4.2.2. Stability upon freeze-thaw cycles

Sample re-freezing after analysis may be required for sample re-analysis, especially when limited volumes of sample are available. Therefore the effects of freeze-thawing in AF composition needed to be investigated. For this purpose, spectra were acquired for 3 different AF control samples (C, D and E) before and after freeze-thaw cycles (Figure 3.3b, page 65). PCA scores plots, resulting from the analysis of the spectra, are shown in Figure 4.7.

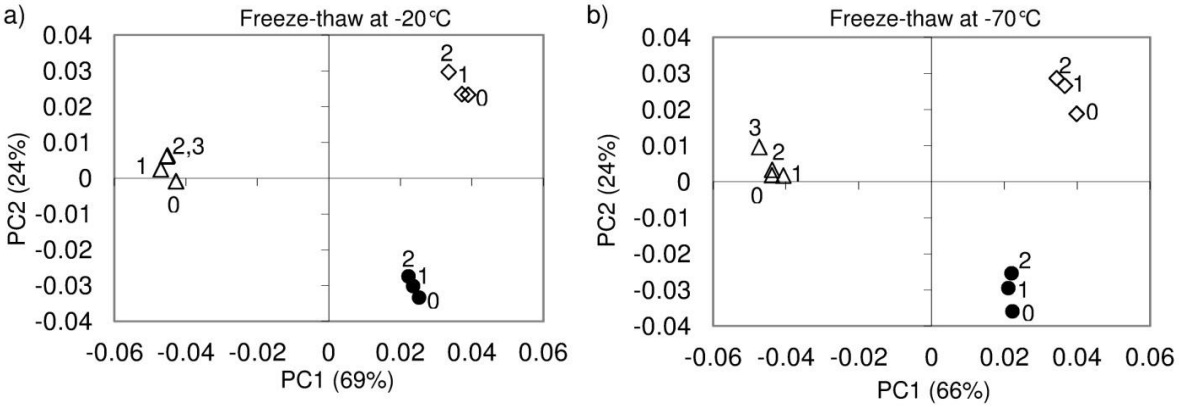


Figure 4.7 - PCA scores scatter plots for control AF samples analysed after consecutive freeze-thaw cycles at: a) -20°C and b) -70°C. Symbols represent sample from 3 different donors: Δ (C), ● (D) and ◊ (E). Numbers 1-3 indicate the freeze-thaw cycle number and 0 indicates the spectrum acquired the prior to freezing the sample.

Both PCA scores scatter plots show 3 distinct groups of data points, each corresponding to one of the 3 samples used in the study. Trends from the first spectra prior to freezing and up to the spectra of 2 or 3 freeze-thaw cycles were observed in the PC2 dimension for each sample. This dispersion resulted from small differences observed between spectra after each consecutive freeze-thaw cycle both at -20°C and -70°C. After spectra visual inspection, differences were found between consecutive freeze-thaw cycles, namely subtle diminution of the intensity of the broad signals across the whole spectral range, noted both at -20°C and -70°C. In fact, protein precipitation was confirmed visually after the first freeze-thaw cycle by the appearance of pellet in the NMR tube. Also at -20°C a decrease in carnitine signal at 3.42 ppm was noted. Most of the variations between consecutive freeze-thaw cycles identified were, however, highly dependent upon the sample as shown in Table 4.1.

Table 4.1 - Sample-specific metabolite variations noted in ^{1}H NMR spectra of control AF samples (C, D and E from Figure 4.7) with freeze-thawing cycles. Arrows indicate increase (↑) or decrease (↓) in signal intensity.

Temperature	Sample C	Sample D	Sample E
-20ºC	acetate (1.91 ppm, ↑) creatinine (3.04 ppm, ↑) betaine (3.26 ppm, ↓)	acetate (1.91 ppm, ↑) pyruvate (2.37 ppm,↓) betaine (3.26 ppm, ↓) glutamine (3.77, ppm ↓) formate (8.45 ppm, ↑)	creatinine (3.04 ppm, ↓) betaine (3.26 ppm, ↓) glutamine (3.77, ppm ↓)
-70ºC	acetate (1.91 ppm, ↑) glutamate (3.75, ppm ↓)	creatinine (3.04 ppm, ↓) betaine (3.26 ppm, ↓) carnitine (3.42 ppm, ↓) glutamine (3.77, ppm ↓)	dimethylglycine (2.92 ppm, ↑) creatinine (3.04 ppm, ↓) betaine (3.26 ppm, ↓) carnitine (3.42 ppm, ↓) glutamine (3.77, ppm ↓)

These results contrast with those reported in a previous study, in which the AF spectra recorded at 300MHz was unaffected by prior freezing-thawing of AF samples (Sims *et al.*, 1993b). Therefore, the work presented resulted in a significant increase in sensitivity and resolution in detection of AF chemical changes occurring after freezing-thawing the samples. Indeed, compositional differences were also observed in human plasma, for which mostly the high M_w species (lipoproteins) were affected by freeze-thawing (Teahan *et al.*, 2006). Similar changes were observed for rat plasma samples, also accompanied by increases in sharp resonances, possibly due to protein precipitation originated by freeze-thawing cycles (Deprez *et al.*, 2002). It is possible, therefore that the same type of phenomena is responsible for the intensity reduction observed in AF after the freeze-thaw cycles. Moreover, several metabolite changes were observed in human urine samples freeze-thawed during a period of one month, some of which

were also observed in some of the AF samples, namely an increase in acetate and decrease in creatinine (Saude and Sykes, 2007).

In conclusion, small variations were registered in ^{1}H NMR spectra of AF after consecutive freeze-thaw cycles at either temperature tested, most of which were sample-dependent. Therefore, freeze-thawing of human AF samples should be minimized to one or two cycles, and only when necessary.

4.2.3. Stability at room temperature

Amniotic fluid samples may stay for several hours at room temperature in several stages of sample analysis and preparation. These may include the period mediating the sample collection and their transport to the cytogenetics laboratory in order to collect the amniocytes (as mentioned earlier in chapter 3 (section 3.1.1, page 58); during sample handling and preparation for analysis; and also during high-throughput analysis, where samples may stay for several hours at room temperature in the automatic sample changer while the acquisition of other samples take place. Since sample cooling to lower temperatures (4°C for example) is not always available in those situations, it is necessary to evaluate any possible composition variations occurring at room temperature.

In order to evaluate human AF stability at room temperature (22°C) over periods of from 1 up to 13 hours (typical duration of an overnight run), spectra were acquired for 4 samples in regular intervals 1 hour up to 8 hours (samples F, G, H and I) or acquired before and after periods of 10 hours (for 2 samples: J and K), 13 and 28 hours (for sample K only) (as illustrated in Figure 3.3c, page 65). The spectra obtained were analysed by PCA (Figure 4.8).

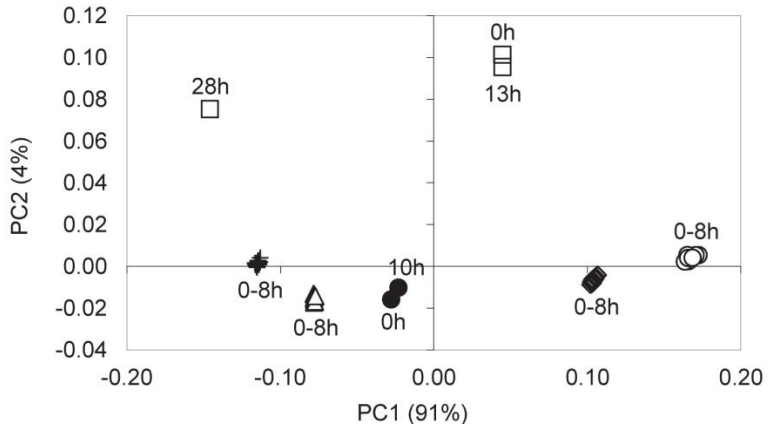


Figure 4.8 - PCA scores scatter plots of spectra from control AF samples analysed at several times (hours) at room temperature (22°C). Symbols represent sample from 6 different donors: ○(F), Δ(G), ♦(H) +(I), ●(J) and □(K).

The PCA scores scatter plot (Figure 4.8) shows that time points for the spectra collected between 0-8h are well clustered together (samples F,G, H and I) as well as those corresponding to 0h and 10h (sample J) and 0h and 13h of sample K (Figure 4.8). This means that the differences between spectra of the different samples were far greater than the differences between spectra acquired for the same sample. The only exception is noted for time-point at 28h for sample K, which clearly deviates from the time-points at 0h and 13h, therefore indicating that strong compositional variation occurred after 28h at room temperature. Visual inspection of the spectra of sample K for 0h and 28h time-points revealed intensity decreases of the broad profile as well as sharp intensities resonances with an increase of ethanol intensity. The ethanol present in the sample originated from the disinfectant used in sample collection and manipulation. Similarly to other metabolites such as lactate and citrate, ethanol may be bound to some extent to protein (Avdulov *et al.*, 1996). In fact, increases in peak intensities of tyrosine and phenylalanine have also been reported in ^{1}H NMR spectra of plasma after 15 and 24 h at room temperature (Deprez et al., 2002), which were also attributed to the release of these aromatic amino acids from precipitating macromolecules. Therefore, the increase in ethanol signal in AF may be linked to protein precipitation (visible in the NMR tube).

Visual inspection of the spectra of samples F, G, H and I revealed consistent variations between 0-8h. These variations comprised decreases in glutamine and increases in acetate and glutamate, which were integrated and plotted as the ratio between the integral at each time-point and the integral at the initial spectrum (t=0 h), and illustrated in Figure 4.9.

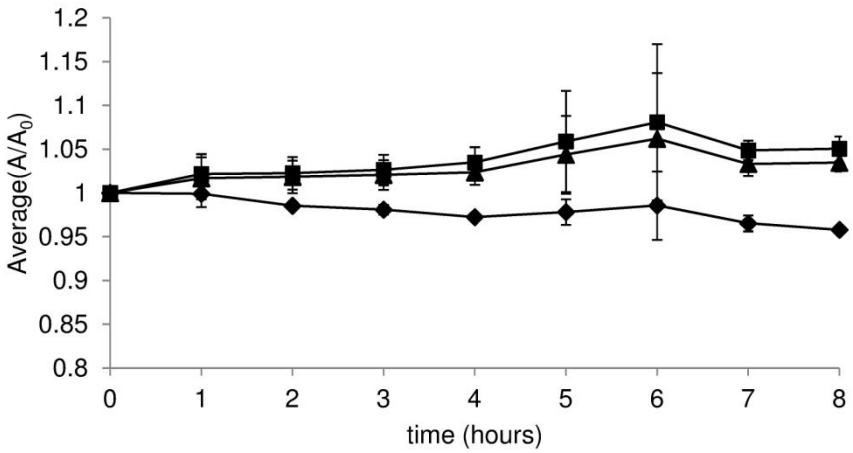


Figure 4.9 - Variations of acetate (▲, 1.91 ppm), glutamate (■, 2.34 ppm) and glutamine (♦, 2.46 ppm) occurring at room temperature (22°C). Normalized plots of the ratio A/A_0 (A_0 is the integral of the peak at t = 0 hours).

Sample-specific variations were also found, namely decreases in citrate (sample H) and decreases in creatinine (sample I), decreases in pyruvate (F, H, I) and increases in citrate (sample F). Decreases in glutamine and increases in acetate and glutamate were also seen between 0h and 10h (sample J), as well as between 0h and 13h (sample K), the latter also registering decreases in creatinine, histidine and citrate. Interestingly, some of the variations noted at room temperature in AF, namely those of acetate, glutamine and glutamate were observed during storage at -20°C, although with smaller magnitude at room temperature (Figure 4.9). Once more, the enzyme activity of the samples could be responsible for the changes observed, rather than any bacterial activity since, for instance, no concomitant decreases in glucose were observed with the increases in acetate. However, since no sample preserving agent was used nor any enzyme activity assays were performed, none of the hypothesis can be ruled out.

Comparatively to AF, changes in metabolites have also been reported for blood plasma after 15 and 24h at room temperature as mentioned above (Deprez *et al.*, 2002), as well as in urine after 1 week at 22°C (Saude and Sykes, 2007) with changes occurring in the levels of lactate, benzoate, hippurate, glycine and creatinine which were attributed to bacterial activity. Hence, AF seems to be more susceptible to metabolite changes at room temperature than plasma and urine.

In conclusion, no large AF profile differences seem to occur after 13h at room temperature, as seen by PCA (Figure 4.8). However, close inspection of the spectra reveals that small but measurable variations are expected after 1h, which remain nearly constant after 4h (Figure 4.9). It is then recommended that acquisition of AF spectra should be performed within 4h. If larger time windows up to 8h are to be considered, then variations in the levels of acetate, glutamine and glutamate should be taken into account. Although it was not possible to evaluate the AF composition during the period between collection and storage (a period not greater than 4 hours as mentioned in chapter 3, section 3.1.1, page 58), the observations described identify what metabolite changes may have occurred during this period.

4.3. Compositional analysis of 2nd trimester human amniotic fluid by high resolution NMR and hyphenated NMR

4.3.1. Analysis by high resolution 1D and 2D ^{1}H NMR spectroscopy

In order to characterize human amniotic fluid composition, AF samples were analysed at the field strengths of 500MHz and 800MHz and prepared as described in chapter 3 (section 3.3.1, page 63) for direct dilution, with buffer addition or freeze-dried and reconstituted in D_2O to

obtain more concentrated samples (1.3-fold concentrated). Figure 4.10 shows an example of a ^{1}H NMR spectra acquired at 500MHz for an AF sample prepared with buffering.

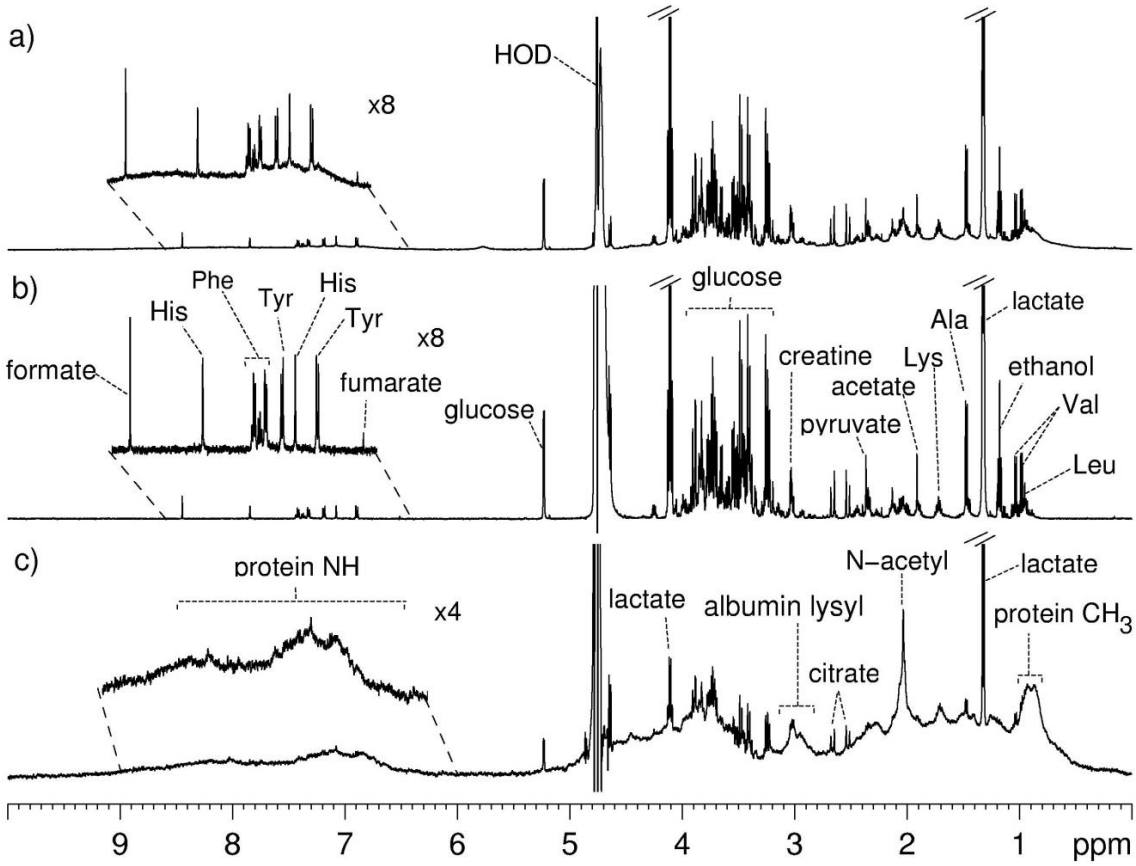


Figure 4.10 - 1D ^{1}H NMR 500 MHz spectra of a control AF sample prepared with buffering (pH 7.22): a) standard, b) T_2-edited (CPMG) and c) diffusion-edited; N-acetyl, glycoprotein N-acetylated groups; amino acids presented in tree-letter code.

The standard 1D spectrum (Figure 4.10a) shows sharp well resolved signals resulting from low M_w molecules, overlapping with a broad envelop of signals arising from macromolecules, mostly proteins, which illustrates the complexity of AF. Since the small molecules and macromolecules differ in both T_2 relaxation times (longer in low M_w molecules) and diffusion rate in solution (higher for low M_w molecules), it was possible to selectively filter their respective resonances using 1D spectral edited experiments namely T_2-edited (CPMG) and diffusion-edited experiments. In the CPMG experiment, a spectrum composed only of sharp resonances and a flat baseline was obtained (Figure 4.10b). The CPMG spectrum is similar to the spectrum of the M_w<3kDa fraction of AF (Figure 4.3, page 88). The diffusion-edited spectrum obtained (Figure 4.10c) was predominantly composed of broad signals originating from the AF proteins such as the methyl proton (CH_3) resonances of from amino acid residues at 0.5-1.0 ppm and resonances from NH protons (from amide and side-chains of amino acid residues) at 6.5-8.5ppm (Figure 4.10c). A

group of signals observed around 3.0 ppm could be assigned to lysine side-chain protons of albumin, which is the most abundant protein in AF (with concentrations around 4mg/mL during 2nd trimester, according to Gao *et al.*, 2009 and Lind, 1981) and with basis on reported 1D ^{1}H NMR spectra of blood plasma (Foxall *et al.*, 1993). A sharp resonance at 2.03 ppm was assigned to the N-acetylated groups of amino sugar moieties in highly mobile regions of glycoproteins, from which transferrin, an iron transport protein, is one the most abundant in 2nd trimester AF found at concentrations of 0.3mg/mL (Gao *et al.*, 2009).

The standard and CPMG spectra acquired for one control AF sample (prepared by direct dilution) at 500MHz showed comparable results to those acquired for another control sample at 800MHz, as it is shown in Figure 4.11a and b, respectively. However, as expected, the spectra collected at the higher field strength registered better resolution, as illustrated in the expansions of Figure 4.11b and also from half height width values (for alanine doublet at 1.47 ppm: $\Delta\nu/2$ = 1.47Hz at 800MHz and $\Delta\nu/2$ = 1.52Hz at 500MHz). Higher sensitivity was also obtained at 800MHz with signal-to-noise ratio (s/n) of 221.7 at 800MHz, compared to 63.4 registered at 500MHz.

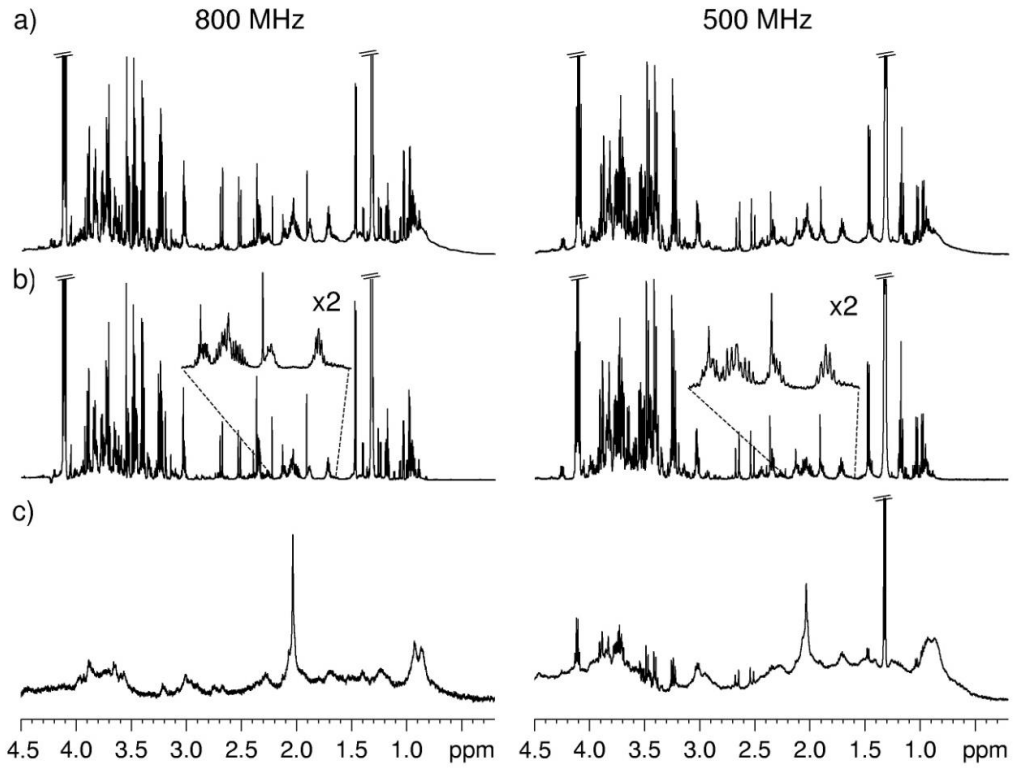


Figure 4.11 - Sub-region of 0 - 4.5 ppm of 1D ^{1}H NMR spectra of a control AF sample acquired at 800MHz (left) and another control sample acquired at 500MHz (right): a) standard, b) CPMG (with 1.6-2.2 ppm expansion) and c) diffusion-edited. All spectra consisted in 256 transients except the diffusion-edited obtained at 800MHz acquired to 512 transients.

Figure 4.11c (left) shows a diffusion-edited spectrum registered at 800MHz. The acquisition conditions required the use of a gradient pulse of 2ms (using 70% of maximum gradient strength) and diffusion time of 200ms, which enabled the complete elimination of sharp resonances (from low M_w molecules) but required the double number of transients (NS=512) in order to clearly visualise the resonances (which doubled the acquisition time of this experiment when compared to the standard or CPMG). The acquisition parameters used for 800MHz diffusion experiment were also tested at 500MHz. Because a poorer spectrum was obtained, a re-adjustment of the gradient pulse, gradient strength and diffusion time was needed in order to obtain a spectrum with better s/n at 500MHz. The best compromise between acquisition time and s/n was obtained by changing the values of gradient pulse to 1 ms using 95% of maximum gradient strength and diffusion time to 100ms. The resulting spectrum registered the higher signal-to-noise ratio of 94.4 when compared with the signal-to-noise ratio of 71.7 obtained for the 800MHz spectrum. However, a complete attenuation of some signals of the most abundant low M_w species was not achieved (Figure 4.11c, right). Nevertheless, the spectrum obtained at 500MHz was considered satisfactory as the best compromise between acquisition speed, low M_w signal attenuation and s/n of the resulting spectrum and these conditions were used for the subsequent metabolic profiling studies.

Figure 4.12 shows the expansions of the 800MHz TOCSY and HSQC spectra of a human AF freeze-dried reconstituted sample, where several peaks are indicated, corresponding to assigned and unassigned spin-systems. The assigned spin-systems are listed in Table 4.2.

The analysis of 2D spectra enabled the assignments of previously identified metabolites (Bock, 1994; Le Moyec *et al.*, 1994; McGowan *et al.*, 1999; Groenen *et al.*, 2004), thus confirming their identification, as well as the identification of about 20 new metabolite assignments marked with [a] in Table 4.2. The newly assigned metabolites comprised several families of metabolites including amino acids: α-aminobutyrate (ABA), arginine, aspartate, citruline, cystine, hydroxyproline, ornithine and taurine; organic acids: acetoacetate, α-hydroxyisobutyrate, α-hydroxyisovalerate, fumarate, α-ketoglutarate and oxaloacetate; and others such as acetone, ethanolamine, urea, trimethylamine-N-oxide (TMAO), N-acetyl group (from glycoproteins) and ethanol. The latter, as mentioned in the previous section, may originate from the disinfectant used in the amniocentesis procedure rather that from endogenous metabolism.

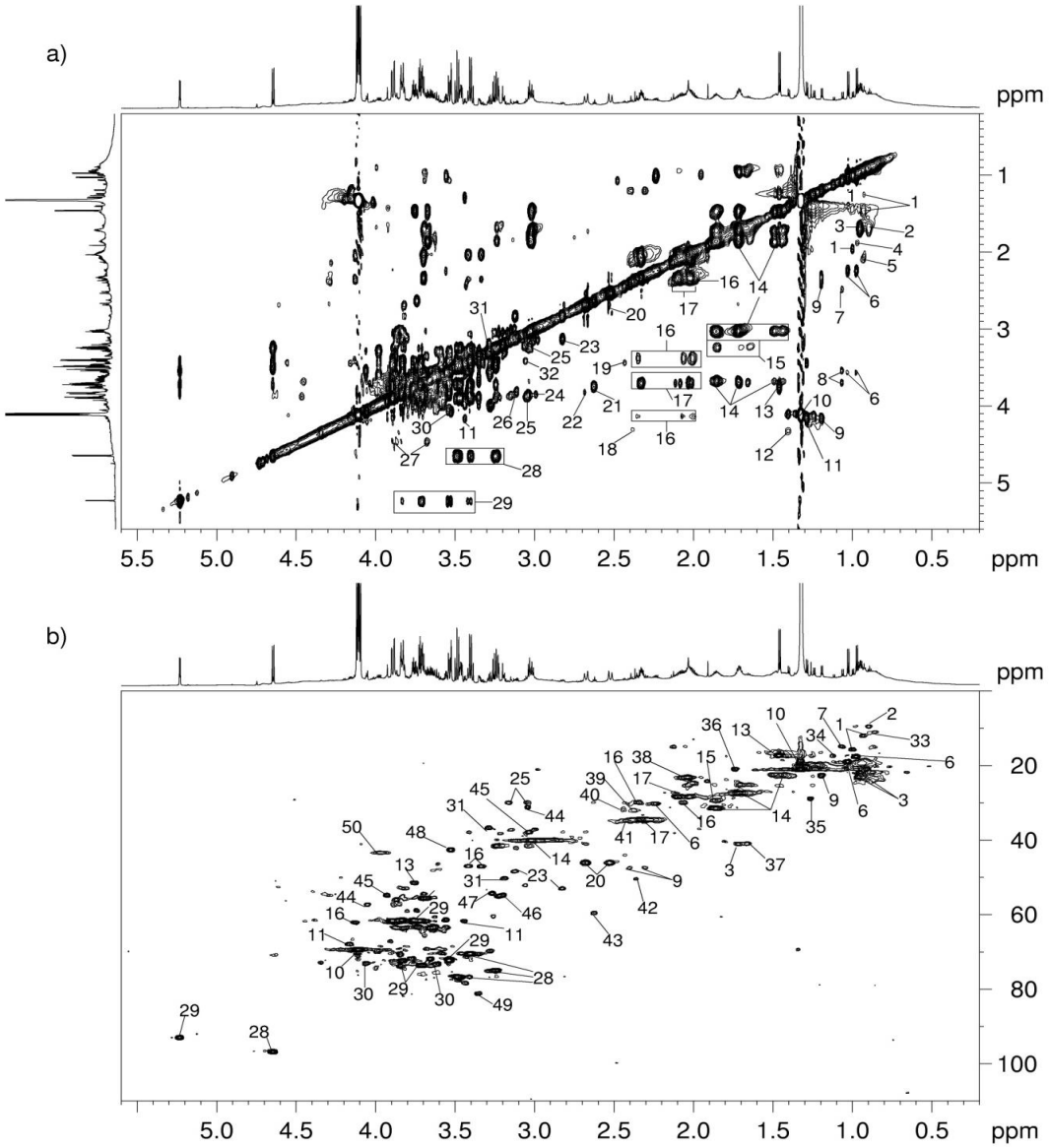


Figure 4.12 - Expansions of 800MHz: a) TOCSY spectrum and a b) ^{1}H/^{13}C HSQC correlation spectrum of a control human AF sample (pH 9.25). Assignments: 1, isoleucine, 2, α-hydroxybutyrate; 3, leucine; 4, α-aminobutyrate; 5, U6; 6, valine; 7, U8; 8, U9; 9, β-hydroxybutyrate; 10, lactate; 11, threonine; 12, U12; 13, alanine; 14, lysine; 15, arginine; 16, proline; 17, glutamate+glutamine; 18, hydroxyproline; 19, carnitine; 20, citrate; 21, methionine; 22, U13; 23, U22; 24, tyrosine; 25, histidine; 26, ethanolamine; 27, U27; 28, β-glucose; 29, α-glucose; 30, myoinositol; 31, taurine; 32, U24; 33, U4; 34, α-oxoisovalerate; 35, α-hydroxyisobutyrate; 36, U17; 37, U16; 38, N-acetyl groups (glycoproteins); 39, pyruvate; 40, α-ketoglutarate; 41, succinate; 42, oxaloacetate; 43, U19; 44, creatinine; 45, creatine; 46, choline; 47, betaine; 48, glycine; 49, U28; 50, U15.

Table 4.2 - ^{1}H NMR assignments for control human AF samples. δ^{1}H, δ^{13}C and *J*-coupling constants observed at 800 MHz and pH≈9.0. In square brackets: δ^{1}H observed at pH≈7.3 (500MHz). [a], identified for the 1st time by NMR in this work; [b], tentative assignment; s, singlet; d, doublet; dd, doublet of doublets; t, triplet; q, quadruplet; m, multiplet.

compound	δ^{1}H in ppm (multiplicity, *J* in Hz, assignment) / δ^{13}C in ppm
acetate	1.91(s,CH_3)/24.02
acetone/acetoacetate [a]	2.22(s, CH_3)
N-acetyl groups (glycoproteins) [a]	2.03(s, CH_3)/23.06
alanine	1.46(d, 7.4, βCH_3)/17.2, *[1.47]*; 3.76(q, αCH)/51.4, *[3.78]*
α-aminobutyrate (ABA) [a,b]	0.97(t, γCH_3)/9.30; 1.87(m, βCH_2)/25.25
arginine [a]	1.67(m, γCH_2)/25.06, *[1.66]*; 1.85(m, βCH_2)/29.32, *[1.90]*; 3.23(t, δCH_2)/41.47
aspartate [a]	2.68(m, βCH_2)/37.74, *[2.86]*; 2.82(m, $\beta' CH_2$)/37.74, *[2.95]*; 3.81(m, αCH)/52.89
betaine	3.26(s, $N(CH_3)_3$)/54.24; 3.90(s, CH_2)/67.20
carnitine	2.44(dd, αCH_2)/44.09; 3.23(s, $N(CH_3)_3$)/41.46; 3.43(m, γCH_2)/71.51
citrate	2.52(d, 15.7, α,β,CH)/45.99; 2.67(d, 15.3, α',β',CH)/45.99, *[2.66]*
citrulline [a,b]	1.54(m, γCH_2)/26.77; 1.84(m, βCH); 3.12(m, δCH_2)
choline	3.20(s, $N(CH_3)_3$)/54.94, *[3.19]*
creatine	3.03(s, NCH_3)/37.74; 3.93(s, NCH_2)/54.69
creatinine	3.04(s, NCH_3)/31.07; 4.05(s, NCH_2)/57.24
cystine [a,b]	3.04(dd, βCH_2); 3.22(dd, $\beta' CH_2$); 3.82(m, αCH), *[3.85]*
N,N-dimethylglycine	2.92(s, $(CH_3)_2$); 3.79(s, CH)
ethanol [a]	1.19(t, CH_3)/19.83, *[1.18]* ; 3.89(q, CH_2)/63.43, *[3.66]*
ethanolamine [a]	3.11(m, NCH_2)/41.91, *[3.15]*; 3.80(m, OCH_2)/58.93, *[3.82]*
formate	8.45(s, CH)
fumarate [a]	6.50(s, CH)
α-glucose	3.41(t, C4H)/70.42; 3.53(dd, C2H)/70.20; 3.71(t, C3H)/73.52; 3.76(dd, 5.8, C6H)/61.42; 3.83(m, C5H)/72.73; 3.83(, C6H')/61.42); 5.23(d, 3.9, C1H)
β-glucose	3.23(dd, C2H)/74.97; 3.40(t, C4H)/70.40; 3.47(dd, C5H)/61.51; 3.49(t, C3H)/76.66; 3.71(dd, C6H')/61.54; 3.89(m, C6H)/61.54; 4.64(d, 8.1, C1H)/96.69
glutamate	2.01(m, βCH_2)/28.22, *[2.05]*; 2.03(m, $\beta' CH_2$)/28.22, *[2.13]*; 2.33(m, γ'CH2)/34.32, [2.35]; 2.49(m, γCH_2), *[2.35]*; 3.69(t, αCH)/55.49, *[3.75]*
glutamine	2.09(m, βCH_2)/28.25, *[2.06]*; 2.11(m, $\beta' CH_2$)/28.25, *[2.13]*; 2.33(m, γCH2)/31.87, [2.45]; 3.69(t, αCH)/55.42, *[3.77]*
glycine	3.52(s, αCH)/42.50, *[3.56]*
histidine	3.03(dd, βCH_2)/29.77; 3.15(dd, $\beta' CH_2$)/29.83; 3.85(dd, αCH)/57.22 7.01(s, C4H, ring)/117.70, *[7.08]*; 7.72(s, C2H, ring)/137.20, *[7.85]*
α-hydroxybutyrate	0.89(t, 7.5, CH_3)/9.41; 1.65(m, CH_2); 1.73(m, CH_2); 4.00(m, CH)/74.16
β–hydroxybutyrate	1.19(d, 6.5, CH3)/22.54; 2.29(m, CH2)/47.32, [2.31]; 2.39(m, CH2)/47.41, [2.40]; 4.13(m, CH)/66.57, [4.12]
α-hydroxyisobutyrate[a]	1.26(s, CH_3)/28.76

Table 4.2 *(cont.)*

compound	$\delta^{1}H$ in ppm (multiplicity, *J* in Hz, assignment) / $\delta^{13}C$ in ppm,
α-hydroxyisovalerate [a,b]	0.83(d, CH_3)/23.45; 0.85(d, CH_3)/21.50; 1.99(m, βCH_2)
hydroxyproline [a,b]	2.13(m, βCH_2)/38.00; 2.38(m, β'CH2); 4.29(m, γCH_2)
isoleucine	0.93(t,7.2, γCH_3)/11.90; 1.00(d, 7.0, βCH_3)/15.60; 1.25(m, γCH_2)/25.23, *[1.27]*; 1.45(m, γ'CH_2)/25.39, *[1.46]*; 1.94(m, βCH)/36.96, *[1.98]*; 3.62(m, αCH)/60.52, *[3.68]*
indoxyl sulfate [b]	7.01(m); 7.17(m); 7.30(s, NCH); 7.53(d), *[7.54]*; 7.72(d), *[7.73]*
α-ketoglutarate [a]	2.44(t, βCH_2)/31.60, *[2.43]*; 2.99(t, γCH_2)
lactate	1.32(d, 7.0, CH_3)/20.83; 4.11(q, 6.9, CH)/69.30, *[4.14]*
leucine	0.95(d, 6.1, δCH)/21.72; 0.96(d, 6.1, δCH)/22.73; 1.69(m, γCH)/25.00; 1.71(m, βCH2)/40.90; 3.69(t, αCH)/54.30, *[3.74]*
lysine	1.45(m, γCH_2)/22.38, *[1.47]*; 1.71(m, δCH_2)/27.31, *[1.72]*; 1.86(m, bCH2)/31.37, [1.90]; 3.01(t, εCH_2)/39.97, *[3.03]*; 3.67(t, αCH)/55.50, *[3.75]*
methionine	2.13(m, βCH_2)/27.18; 2.14(s, SCH_3)/14.85, *[2.19]*; 2.62(t, 6.2, γCH2)/29.78, *[2.63]*; 3.73(t, αCH)/55.34, *[3.86]*
myo-inositol	3.23(t, C5H)/75.21; 3.58(dd, C1H, C3H)/70.27; 3.62(dd, C4H, C6H)/73.42 4.05(t, C2H)/73.03
ornithine [a,b]	3.02(t, δCH_2); 3.67(t, αCH)
oxaloacetate [a]	2.36(s, CH_2)/50.30
α-oxoisovalerate	1.12(d, 7.2, CH_3)/17.31; 3.02(m, βCH)
phenylalanine	3.03(dd, β'CH_2)/38.12; 3.19(dd, βCH_2)/38.17; 3.83(dd, αCH)/55.00; 7.32(m, C2H, C6H, ring)/130.10; 7.40(m, C4H, ring)/128.60, *[7.35]*; 7.41(m, C3H, C5H, ring)/129.80, *[7.42]*
proline	2.01(m, γCH)/24.59; 2.06(m, βCH)/29.80; 2.35(m, β'CH)/29.80; 3.33(m, δCH)/46.91, *[3.34]*; 3.41(m, δ'CH)/46.91; 4.12(m, αCH)/62.03, *[4.13]*
pyruvate	2.37(s, CH_2)/31.87
succinate	2.39(s, CH_2)/34.67, *[2.40]*
taurine [a]	3.19(t, SCH_2)/50.11, *[3.12]*; 3.27(t, NCH2)/36.60, *[3.24]*
threonine	1.28(d, γCH_3)/20.22, *[1.32]*; 3.44(d, αCH)/61.64, *[3.58]*; 4.14(m, βCH2)/67.91, *[4.25]*
trimethylamine-N-oxide (TMAO) [a]	3.26(s, $N(CH_3)_3$)/60.43
tyrosine	2.99(dd, β'CH)/37.00; 3.14(dd, βCH)/37.15; 3.84(d, αCH)/57.12 6.86(d, C3H, ring)/116.85, *[6.89]*; 7.17(d, C2H, ring)/131.58, *[7.19]*
urea [a]	5.77 (broad s, NH_2)
valine	0.97(d, 7.0, γCH_3)/17.42, *[0.98]*; 1.03(d, 7.0, γ'CH_3)/18.83; 2.24(m, βCH)/30.10, *[2.27]*; 3.57(d, αCH)/61.33, *[3.61]*

The 2D *J*-resolved spectrum (Figure 4.13) enabled the resolution of highly overlapped multiplets, for instance the multiplets of glutamine and glutamate at 2.35 ppm (region 6, Figure 4.13), as well as the measurement of *J*-coupling constants, as for instance of the triplet of isoleucine which was 7.2 Hz (region 1 in Figure 4.13), thus enabling further confirmation of their assignments.

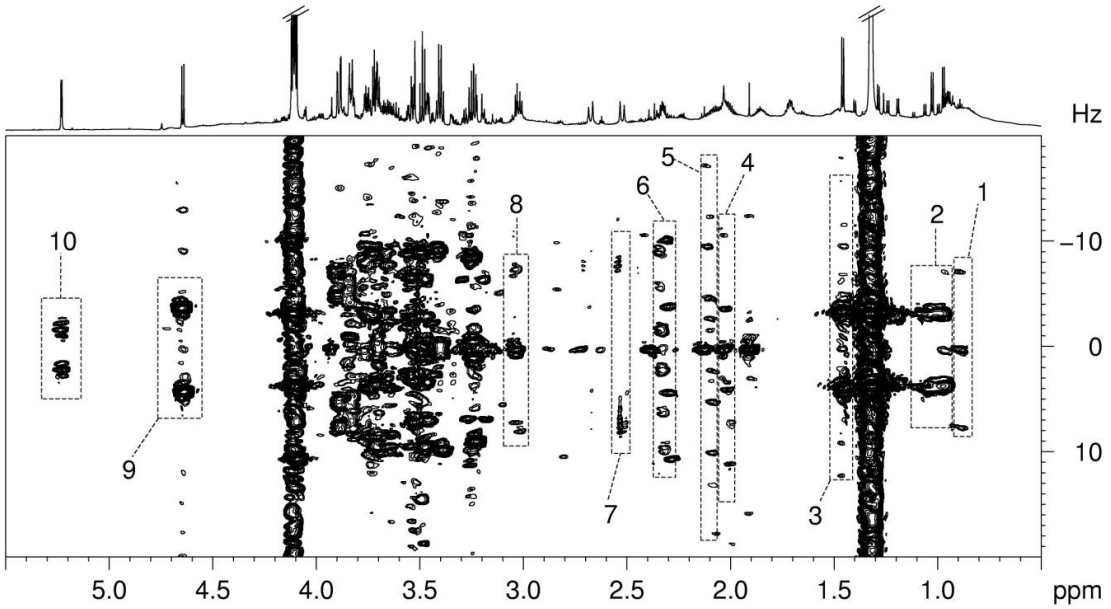


Figure 4.13 - Expansion of 800MHz *J*-Resolved spectrum of a human AF sample (pH 9.25). Numbered regions contain signals corresponding to: 1, isoleucine; 2, leucine, valine, U8 and α-oxoisovalerate; 3, alanine, lysine; 4, proline, glutamine, glutamate, N-acetyl (glycoproteins); 5, proline, glutamine, glutamate; 6, glutamine, glutamate; 7, citrate; 8, lysine, histidine; 9, β-glucose; 10, α-glucose.

A list of 52 spin-systems, for which no assignment was possible by NMR alone, is presented in Table 4.3. This shows on one hand, the potential of 2D analysis to detect many more metabolites thus extending the knowledge gathered by 1D NMR experiments in previous reports (Bock, 1994; Le Moyec *et al.*, 1994; McGowan *et al.*, 1999; Groenen *et al.*, 2004) and, on the other hand, the need for complementary methods such as hyphenated NMR to improve spectral assignment. The differences in sensitivity between 800MHz and 500MHz is also noted here by highest number of unassigned signals detected in 800MHz, most of them were not detected at 500MHz (Table 4.3).

The overall detection limit achieved in the 1D and 2D ^{1}H NMR was of the order of a few micro-molar (μM) since the weakest signals observed in the spectra correspond to compounds expected to occur in this order of concentration in human AF at the 2nd trimester, for instance α-aminobutyrate (ABA) which registered an average concentration of 7μM in human AF at 17-19 weeks of gestation (Tuma *et al.*, 2006).

Table 4.3 - List of still unassigned spin systems. $\delta^{1}H$, $\delta^{13}C$ observed at 800 MHz and pH≈9. In square brackets: $\delta^{1}H$ observed at pH≈7.3 (500MHz).

	δ ^{1}H in ppm (observed multiplicity) / δ ^{13}C in ppm		
U1	0.66/21.70	U27	3.35/81.08; 3.67/69.10;
U2	0.77/20.23; 0.85/20.23		3.87/55.86; 4.46
U3	0.86/14.90	U28	3.41/76.62
U4	0.86/10.88	U29	3.52/68.42; 3.80/68.42; 3.97/68.42
U5	0.91/20.84	U30	3.56/63.19; 3.65/63.43
U6	0.93(broad signal)/19.17, *[0.92]*; 2.07, *[2.08]*;	U31	3.59/70.21
	2.60; 3.44; 3.97, *[3.86]*; 4.16	U32	3.74/58.78
U7	1.04/20.84	U33	3.74/63.19
U8	1.06(d, CH_3)/14.79; 2.48/35.12	U34	3.83/63.32; 3.88/63.43
U9	1.06; 2.48/45.75; 3.54; 3.71, *[3.70]*	U35	3.93; 5.18(d)
U10	1.34/69.29	U36	3.99/69.45
U11	1.36; 4.09	U37	4.35/72.80, *[4.34]*
U12	1.40(m)/17.27; 4.31	U38	5.40(m)
U13	1.54; 1.72; 2.67/40.95; 3.83/63.52	U39	5.66(s), *[5.68]*
U14	1.62; 3.94	U40	6.91(s)
U15	1.65; 3.28/69.61; 3.44/78.25; 3.97/43.27	U41	7.57(s)
U16	1.65/40.89; 1.83/27.74; 2.98/20.08	U42	7.59(m)
U17	1.74/20.79	U43	7.63
U18	1.80/40.24	U44	7.68(s)
U19	2.51; 2.63/59.51; 3.61/46.31; 3.85; 5.60(m)	U45	7.70
U20	2.53; 2.73; 4.90(d)	U46	7.75(m)
U21	2.75; 5.28(m)	U47	7.80(m)
U22	2.83(m)/52.96, *[2.94]*; 3.12(m)/48.26, *[3.15]*;	U48	7.85(m)
	3.82/69.39	U49	8.57
U23	3.04/41.36	U50	8.89(s)
U24	3.06/51.98; 3.41/37.81 or 76.62	U51	9.28(m)
U25	3.09; 5.36	U52	1.17; 4.02
U26	3.15; 3.85/68.97		

4.3.2. Analysis by DOSY and STD spectroscopy

Figure 4.14 shows the aliphatic and aromatic regions from the DOSY spectrum of a healthy human AF sample recorded at 500MHz (pH 7.34). The resonances observed in the DOSY spectrum correspond mainly to the most abundant low M_w species characterized by high diffusivity values in the order of 10^{-9} m^2/s (Figure 4.14). However, in the aliphatic region the signal at 2.03 ppm showed a relatively lower diffusivity value of 1.95×10^{-9} m^2/s (peak 4, Figure 4.14a). The stretched shape of the N-acetyl signal along the diffusion dimension resulted from the overlap with other resonances occurring in the same region, namely the multiplet of proline (Figure 4.14a).

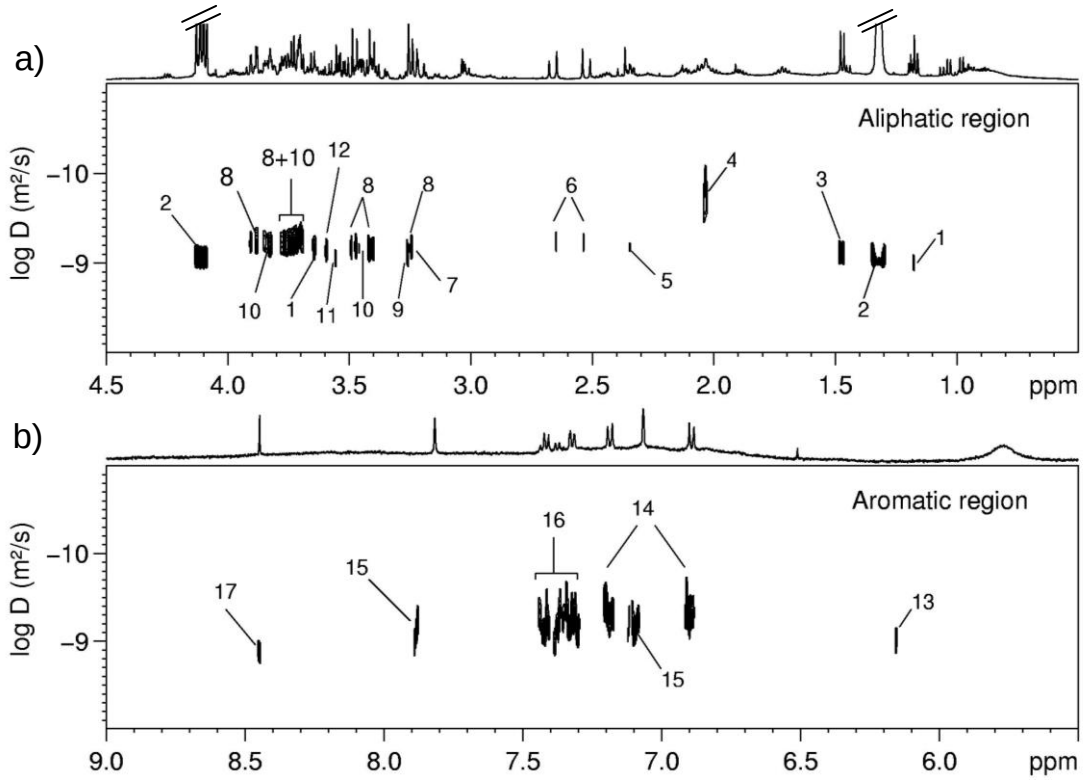


Figure 4.14 - Expansions of the aliphatic and aromatic regions of the ^{1}H NMR 500 MHz DOSY spectrum of a healthy human AF sample at pH 7.34. External projection: standard 1D ^{1}H NMR spectrum of the same sample; assignments: 1, ethanol; 2, lactate; 3, alanine; 4, N-acetyl groups (glycoproteins); 5, pyruvate; 6, citrate; 7, choline; 8, β-glucose; 9, betaine; 10, α-glucose; 11, glycine; 12, threonine; 13, processing artifact; 14, tyrosine; 15, histidine; 16, phenylalanine; 17, formate.

The estimates of M_w for N-acetyl groups of glycoproteins using the pyruvate signal (at 2.37 ppm, assignment 5 in Figure 4.14a) calibration reference gave values of about 3kDa (diffusion was measured at the centre of the peak). The signal at 2.03 ppm is related to N-acetylated sugar moieties of glycoproteins, therefore estimated M_w values in the range of 52-150kDa should be expected, corresponding to some of the most abundant glycoproteins in human AF α1-Anti-trypsin and Immunoglobulin G (Lind, 1981), respectively. Other macromolecules signals, such as the broad resonances observed previously in diffusion-edited spectra (shown in Figure 4.10c, page 97) were not observed in DOSY spectrum of AF, possibly due to their lower relative intensities and broader line widths.

In order to decrease the overlapping effects and allow better resolution and more accurate estimates of the macromolecules M_w, subfractionation by ultrafiltration of AF was performed prior to DOSY analysis (Figure 3.1, page 63). Three AF fractions were obtained differing in average M_w of their constituents and their respective 1D ^{1}H NMR spectra are shown in Figure 4.15.

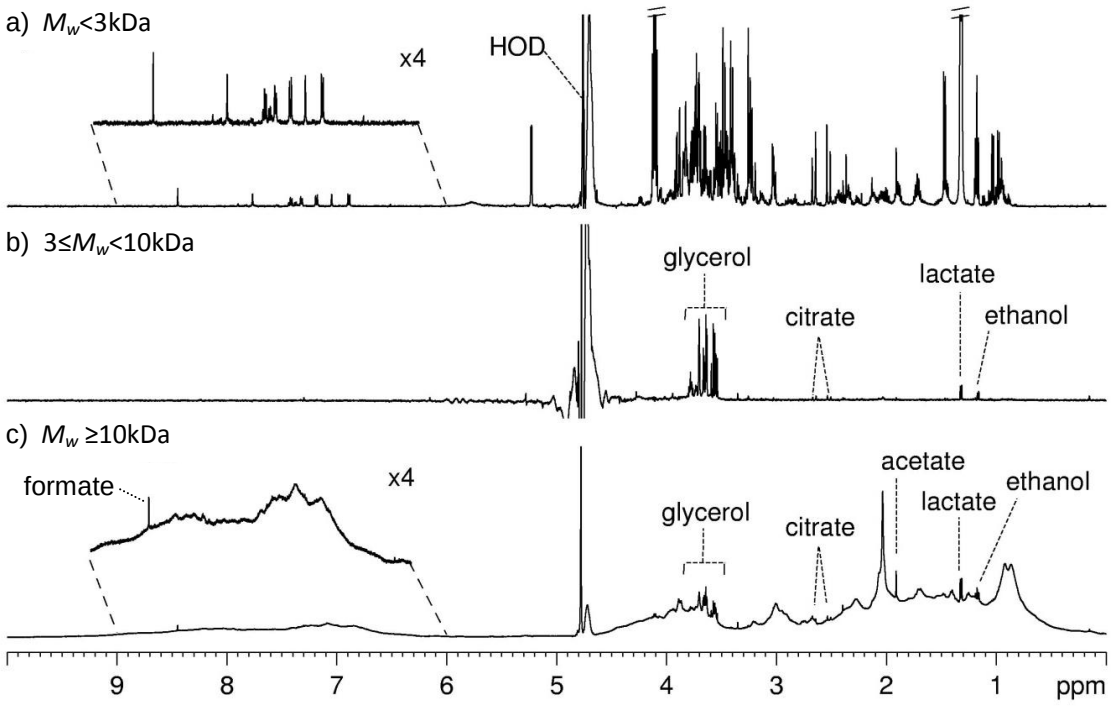


Figure 4.15 - Standard 1D ^{1}H NMR 500 MHz spectra of subfractions of human AF of different M_w (pH≈7.3): a) M_w<3kDa, b) 3≤M_w<10kDa and c) M_w ≥10kDa. HOD, residual water peak.

As seen in Figure 4.15, most of the low M_w molecules were registered in the spectrum of M_w <3kDa fraction (Figure 4.15a), whereas the spectrum of 3≤M_w<10kDa fraction (Figure 4.15b) registered only signals from glycerol from the filter membrane (multiplets at 3.52 - 3.81 ppm), lactate (1.32 ppm), ethanol (1.16 ppm) and small amounts of citrate (2.52 and 2.66 ppm). The spectrum of M_w≥10kDa fraction showed mostly broad signals from proteins as well as signals from organic acids, namely lactate (1.32 ppm), acetate (1.91 ppm), citrate (2.52 and 2.66 ppm), formate (8.45 ppm) and also from glycerol (from the membrane filter) and ethanol used for membrane filter washing and storage (Figure 4.15c). Considering that the fraction M_w≥10kDa was washed with phosphate buffer and filtered several times prior to spectral acquisition, it is suggested that these organic acids were tightly bound to proteins. A similar binding effect was observed between citrate and human serum albumin in solution (Daykin *et al.*, 2002).

In order to further investigate the binding of lactate and other organic acids to AF proteins, a saturation transfer difference (STD) ^{1}H NMR experiment was performed using an intact AF sample (Figure 4.16). In this experiment, two spectra were collected, with saturation of one frequency in the noise region (off-resonance, Figure 4.16a) and with saturation of one frequency of NH groups of the protein (on-resonance, Figure 4.16b).

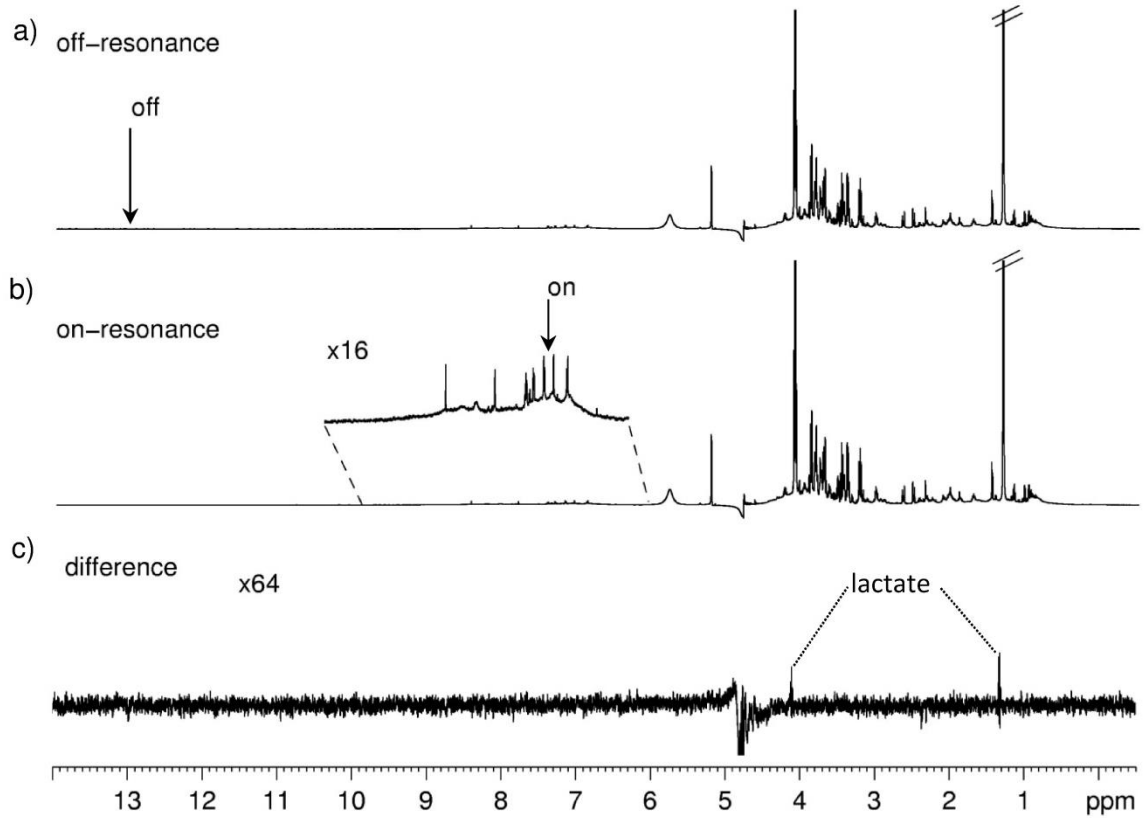


Figure 4.16 - Saturation transfer difference (STD) experiment performed in a control human AF sample: saturated off- (a) and on-resonance (b) ^{1}H NMR 600 MHz spectra and respective difference spectrum between off and on resonance spectra (c).

The difference between the off-resonance and the on-resonance spectrum (Figure 4.16c) showed a small increase in lactate signals thus indicating the saturation transfer from protein signals to bound lactate. However, no signals from citrate or acetate were observed, probably because these molecules are bound to AF proteins to a lesser extent.

DOSY spectrum was then acquired for M_w<3kDa fraction using the same gradient parameters used for the intact AF sample (Table 3.9, chapter 3, page 71). The same signals from low M_w compound observed previously in the DOSY of the intact AF sample (Figure 4.14) are present in the DOSY of the M_w<3kDa fraction as shown in Figure 4.17. In addition the signals from the amino acids leucine, valine, alanine and lysine, which registered some overlap with broad profile resonances in the intact sample 1D spectra, were also detected in the DOSY spectrum of M_w<3kDa fraction (Figure 4.17). Most of the signals registered diffusivity values around 10^{-7} m^2/s whereas in the DOSY spectrum of intact AF sample their diffusivity values were of the order of 10^{-9} m^2/s, which could be justified by the higher viscosity of the latter, which has direct effect on the measured diffusion.

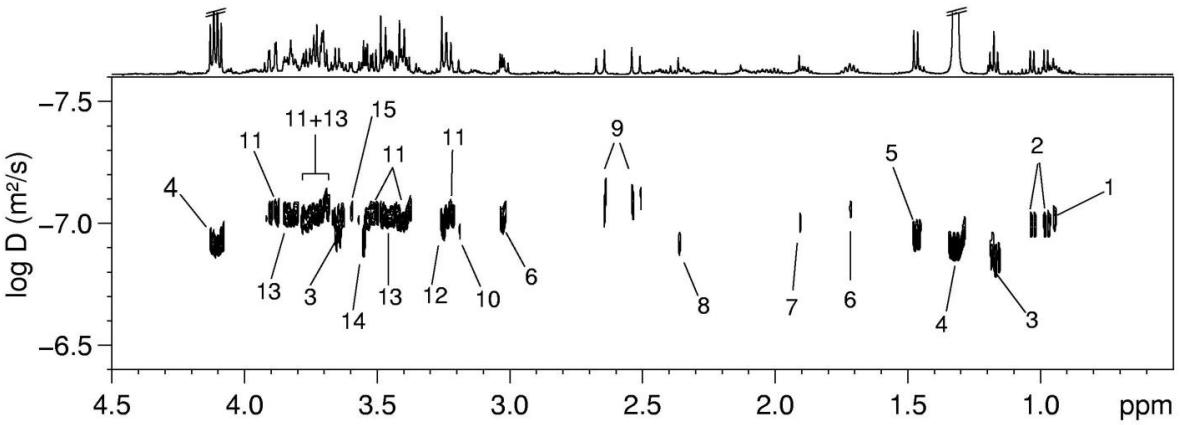


Figure 4.17 - Expansion of 500 MHz ^{1}H NMR DOSY spectrum of an M_w<3kDa fraction of a control human AF sample (pH 7.40). Assignments: 1, leucine; 2, valine; 3, ethanol; 4, lactate; 5, alanine; 6, lysine; 7, acetate; 8, pyruvate; 9, citrate; 10, choline; 11, β-glucose; 12, betaine; 13, α-glucose; 14, glycine; 15, threonine.

The DOSY spectrum of M_w<3kDa fraction improved detection of low M_w compounds. However, no new assignment information was obtained.

DOSY spectra were also acquired for the fraction of M_w≥10kDa, but modifying the gradient conditions to gradient pulse of 1500μs and diffusion time of 250 ms in order to accommodate the entire signal decay of the slow diffusing species (those of higher M_w). A DOSY spectrum obtained for an AF M_w≥10kDa is shown in Figure 4.18.

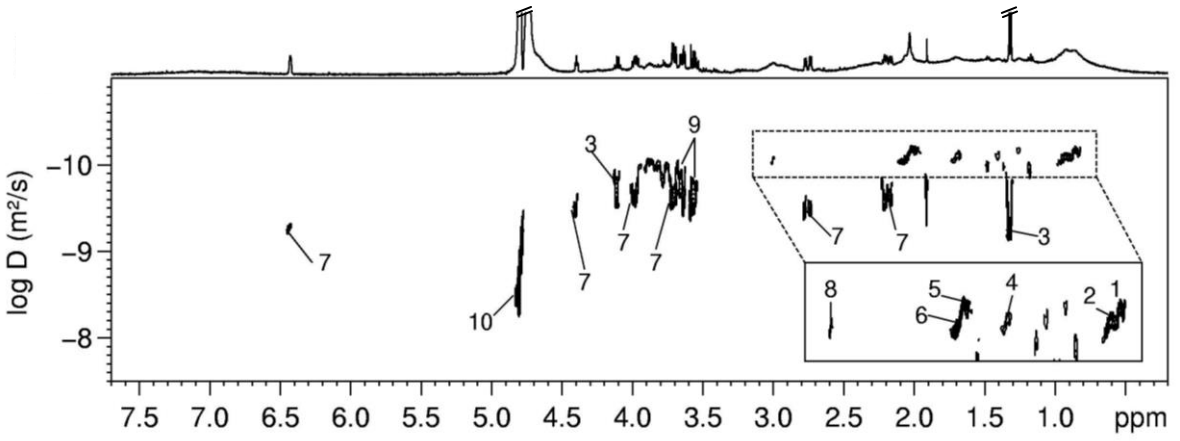


Figure 4.18 - Expansion of ^{1}H NMR DOSY spectrum of an M_w≥10kDa fraction of AF, spiked with shikimic acid, at 500 MHz and pH 7.40 and respective standard 1D ^{1}H NMR spectrum (external projection). Assignments: 1 – protein, 0.86 ppm, 2 – protein, 0.92 ppm, 3 – lactate, 1.32 and 4.11 ppm, 4 – protein, 1.70 ppm, 5 - glycoprotein (N-acetyl group), 2.03 ppm, 6 - glycoprotein (N-acetyl group), 2.06 ppm, 7 – shikimic acid, 2.19(dq, C7H), 2.75(dd, C7H), 3.71(dd, C6H), 3.98(m, C6H), 4.40(t, C4H) and 6.43(c, C3H) ppm, 8 - protein, 3.00 ppm, 9 - glycerol, 3.56 and 3.65 ppm.

The aliphatic region of the DOSY spectrum showed many resolved low diffusivity signals (insert of Figure 4.18) and, based on diffusivity values three families of macromolecules (most probably proteins) could be identified: A, resonances 1 (0.86 ppm) and 5 (2.03 ppm) with diffusivity of 7.16×10^{-11} m^2/s; B, resonances 2 (0.92 ppm) and 4 (1.7 ppm) with diffusivity of 8.20×10^{-11} m^2/s; and C, resonances 6 (2.06 ppm) and 8 (3.00 ppm) with diffusivity 8.99×10^{-11} m^2/s (Figure 4.18).

In order to estimate the average M_w of these families of proteins based on their diffusivity values an appropriate calibrant had to be chosen. Usually in DOSY, signals from compounds of known M_w can be chosen as internal references for M_w estimation. Examples include HOD signal in model D_2O aqueous solutions (Crutchfield and Harris, 2007), the signal of formic acid in wine (Nilsson *et al.*, 2004), or compounds added for chemical shift referencing such as TMS in virgin olive oil (Crutchfield and Harris, 2007). In the case of the M_w≥10kDa fraction of AF, the only compounds of known M_w such as lactate, acetate, ethanol or formate registered signal overlap with the broad signals of higher M_w species. There was also the possibility that most of these compounds were bound to proteins, therefore, both factors interfere with the apparent diffusivity values measured. The HOD signal was also not a good choice since the signal presented a stretched shape, spanning several diffusivity values. The alternative was to use an external calibrant, shikimic acid ($C_6H_6(OH)_3COOH$, M_w = 174.15 g/mol), which had at least one non-overlapped signal when added to AF M_w≥10kDa fraction (signal at 6.43 ppm) as it can be seen in the external projection of Figure 4.18. The M_w estimated values for the 3 protein spin-systems using shikimic acid signal at 6.43 ppm (diffusivity of 5.46×10^{-10} m^2/s) as calibration reference are presented in Table 4.4.

Table 4.4 - Estimated M_w values obtained for low diffusivity protein spin-sytems identified in the DOSY spectrum of M_w≥10kDa AF fraction.

Spin-system	Signals (ppm)	Diffusivity ($\times 10^{-11}$ m^2/s)	Estimated M_w (kDa)
A	0.86, 2.03	7.16	77
B	0.92, 1.70	8.20	50
C	2.06, 3.00	8.99	40

It is difficult to dwell on the identity the protein families identified solely based in these NMR measurements. Nevertheless, at least for protein system A, which includes the N-acetyl group resonance, the estimated M_w roughly matches the M_w of transferrin (80 kDa), one of the most abundant glycoproteins in AF (Gao *et al.*, 2008). The M_w of protein system B (50kDa) is close to the M_w of albumin, which is about 67kDa, although the difference in average M_w is considerable.

However, the resonance at 3.00 ppm, usually attributed to albumin, was found associated with protein system C, which must be related to another protein since it registered a lower M_w (40kDa) compared to the M_w of albumin (67kDa). The results show that a deeper characterization analysis is needed and, possibly correlation with gel electrophoresis analysis, in order to fully assign the spin-systems. However, DOSY of M_w≥10kDa fractions of human AF may help to detect qualitative changes in the protein profile of the biofluid, which may be an important indicator of the onset and course of disorders.

4.3.3. Analysis by LC-NMR/MS

Human AF metabolite composition was further explored by LC-NMR/MS. Since human AF is mainly composed of polar metabolites, two main chromatographic separation approaches were followed prior to NMR and/or MS analysis: a reverse phase-type chromatography, which focused on less polar metabolites, and an ion exchange type chromatography, which focused mainly on sugars and organic acids.

4.3.3.1 Characterization of less polar metabolites

Reverse-phase chromatography (RP-C18) using gradient elution was performed for the analysis of 10 times concentrated M_w<3kDa AF fractions. Figure 4.19 shows the UV and MS (ESI^+) chromatograms obtained.

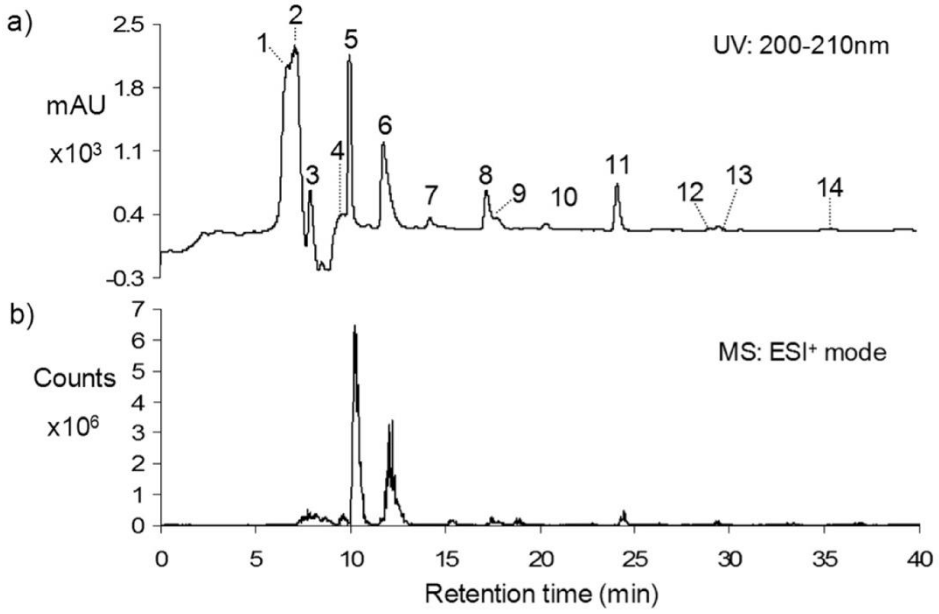


Figure 4.19 - RP-C18 chromatograms of a human AF 10x concentrated M_w<3kDa fraction: a) UV trace, b) MS ESI^+ trace; numbers: the peaks corresponding to the fractions collected (loops).

The eluted fractions corresponding to the resolved peaks of the chromatogram were collected into sample loops (fractions numbered in the UV trace, Figure 4.19a) being subsequently analysed in parallel by NMR and MS.

The most intense peak in the UV chromatographic trace occurred around 6.5 min (Figure 4.19a). The corresponding ^{1}H NMR spectrum, loop 1 collected at 6.51 min. (Figure 4.20b), contains a good portion of the most intense signals observed in the spectrum of the whole M_w<3kDa fraction (Figure 4.20a).

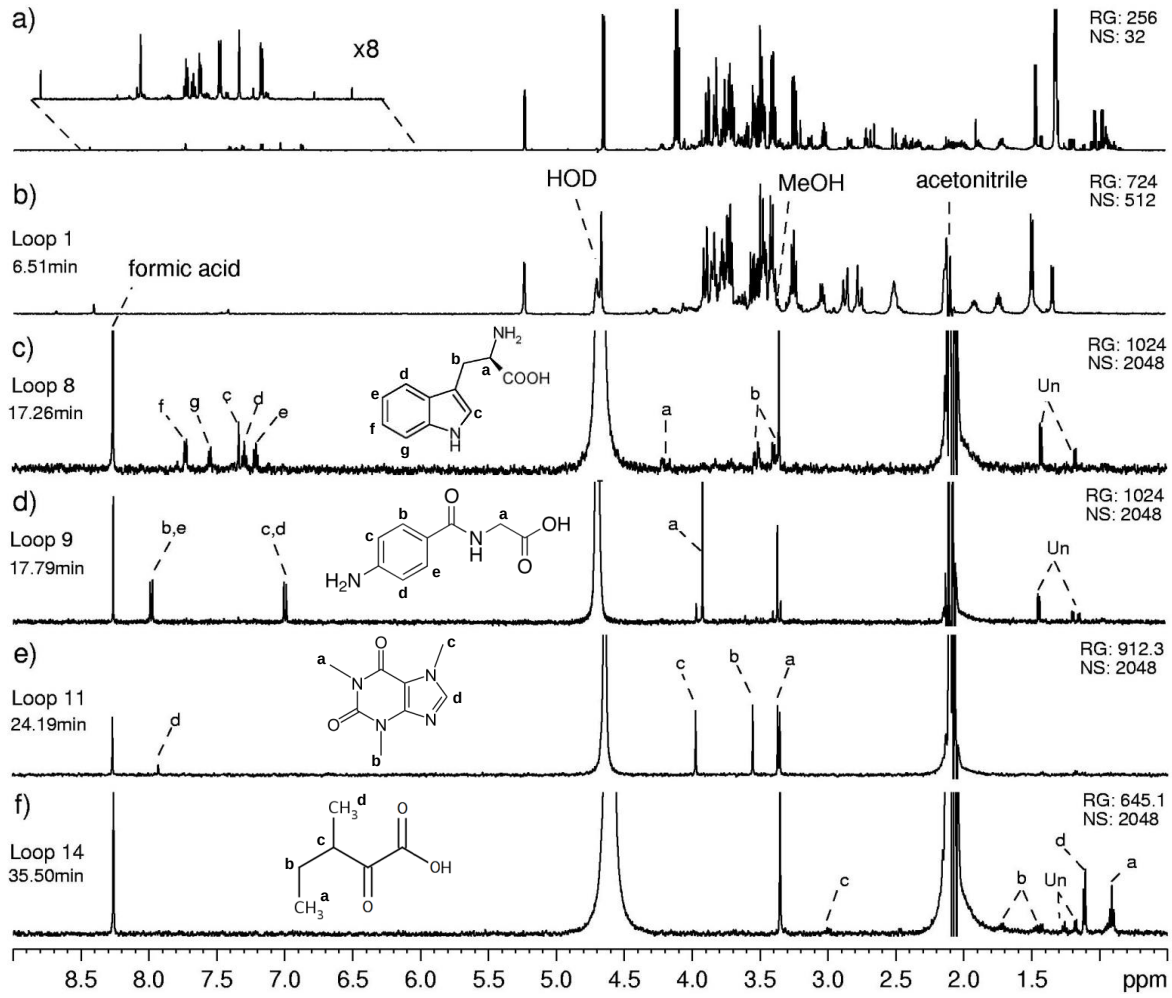


Figure 4.20 - 1D 500MHz ^{1}H NMR spectra of a) 10× concentrated AF M_w<3kDa fraction, and several fractions resulting from its analysis by RP-C18 LC-NMR/MS: b) loop 1, several co-eluting at 6.5 min; c) loop 8, containing tryptophan; d) loop 9, containing 4-aminohippurate; e) loop 11, containing caffeine; loop 14, containing 3-methyl-2-oxovalerate. Loop numbers correspond to the fractions indicated in figure 4.19. Un., unassigned peak; RG, receiver gain; NS, number of scans; HOD, residual water peak; MeOH, residual methanol peak (from the LC system).

Highly polar metabolites, eluting first in RP-C18 columns such as glucose, *myo*-inositol, citrate and the amino acids, alanine glutamine, glutamate, lysine and threonine where identified in the spectrum of loop 1 (Table 4.5). In the spectrum of loop 2 collected at 7.12 min, several compounds also co-eluted including choline, betaine, creatine, lactate, glycine, histidine, proline and valine, together with mannose which was newly identified by NMR (Table 4.5). Co-elution of polar metabolites was also confirmed in the spectrum of subsequent loops 4 (9.50 min) and 5 (10.01 min), where creatinine and methionine; leucine, isoleucine and tyrosine, were assigned respectively (Table 4.5). The first well resolved metabolite identified was the amino acid phenylalanine which eluted at 12.11 min (loop 6, Table 4.5). Up to this point mostly the more polar metabolites were eluted and their assignments confirmed both by ^{1}H NMR and MS and MS/MS measurements in ESI^{+} mode. Apart from mannose, all these metabolites had already been observed by NMR alone (Table 4.5).

The ^{1}H NMR and MS spectra collected for the loops 7 (14.27 min), 8 (17.26 min), 11 (24.19 min) and 13 (24.19 min) allowed the assignment of further metabolites which had not been previously observed in human AF by NMR: paraxanthine, tryptophan (Figure 4.20c), caffeine (Figure 4.20e) and hippurate, respectively. These metabolites had been detected in human AF previously by targeted analytical methods such as GC-MS and capillary electrophoresis (Sommer *et al.*, 1975; Ng *et al.*, 1982; Stewart *et al.*, 2001; Tuma *et al.*, 2006). Therefore, LC-NMR/MS presents a clear advantage over these targeted methods by enabling the detection of several different metabolites in one single experiment. Tryptophan is a known component of AF, although its detection by ^{1}H NMR had not been reported before, probably due to its lower concentration when compared to other amino acids such as phenylalanine, which has signals in the same spectral regions as tryptophan, being however 5 times more concentrated (Table 4.5). Caffeine possibly resulted from free passage from maternal circulation to the fetus, after consumption of caffeine containing beverages, then being excreted by the fetus into AF (Sommer *et al.*, 1975). Consultation of clinical records of the sample donors revealed that most women reported the consumption of coffee, which confirms the finding. Paraxanthine is of the three most abundant metabolites of caffeine found in blood circulation after coffee consumption. It structurally resembles the structure of caffeine but lacks the methyl group (the b methyl group in structure of caffeine shown in Figure 4.20e). Paraxanthine possibly enters the AF by the same mechanism as caffeine (Sommer *et al.*, 1975). Hippurate is a metabolite of phenolic compounds of mainly dietary origin (Holmes *et al.*, 2008), which may also cross the placenta being excreted by the fetal kidney then passing to into AF (Ng *et al.*, 1982).

Table 4.5 – Assignments of LC-NMR/MS on human AF using a RP-C18 column at pH≈2.0. Loop numbers according to Figure 4.19. [a] new assignment by NMR; [b] tentative assignment, no MS information available. MS and MS/MS m/z values correspond to $[M+D]^+$ adducts except: [c]$[M+Na]^+$, [d]$[M+H]^+$. Expected concentration values (2nd trimester of pregnancy) are shown in the last column, according to references: [e], (Tuma *et al.*, 2006); [f], (Lind, 1981); [g], (Groenen *et al.*, 2004).

Loop/ RT (min)	MS (m/z)		δ^1H, ppm (multiplicity, *J* in Hz, assignment)	Assignment	expected concentration in μM (*ref*)
	MS	MS/MS			
1/6.51	-	-	1.49(d, 7.3, CH_3)	alanine	323 [e]
	193.8[c]	-	2.77(d, 15.3, CH_2), 2.86(d, 15.0, CH_2)	citrate	345 [f]
	202.8[c]	-	3.41(t, H4), 3.52(dd, H2), 3.70(t, H3), 3.76(dd, H6), 3.83(m, H5), 5.23(d, H1)	α-D-glucose	2510 [g]
	202.8[c]	-	3.24(dd, H2), 3.40(t, H4), 3.49(t, H3), 3.47(dd, H5), 3.73(dd, H'6), 3.90(dd, H6), 4.64(t, H1)	β-D-glucose	
	153.1	-	1.49(m, γCH_2), 1.73(m, δCH_2), 1.91(m, βCH_2), 3.03(t, 7.5, εCH_2)	lysine	202 [e]
	153.0	-	2.13(m, βCH_2), 2.50(m, γCH_2)	glutamine, glutamate	186,227 [f]
	202.8[c]	-	3.28(t, H5), 3.62(t, H4/H6), 4.06(t, 2.9, H2)	*myo*-inositol	396 [g]
	-	-	1.33(d, 6.6, γCH_3), 4.27(m, βCH_2)	threonine	156 [e]
2/7.12	-	-	3.27(s, CH_3)	betaine	66 [g]
	-	-	3.05(s, CH_3)	creatine	99.8 [g]
	-	-	3.20(s, CH_3), 4.06(m, CH_2)	choline	11 [e]
	90.0[c]	-	1.38(d, CH_3), 4.26(q, CH)	lactate	9800 [f]
	160.9	-	7.41(s, ring H4), 8.48(s, ring H2)	histidine	97 [e]
	-	-	2.00(m, γCH), 2.35(m, γCH'), 4.13(m, αCH)	proline	181 [e]
	-	-	3.56(s, CH_3)	glycine	167 [e]
	-	-	4.89(d, H'6), 5.18(d, H6)	D-mannose[a]	
	-	-	1.00(d, 7.0, CH_3), 1.05(d, 7.0, CH_3), 2.29(m, βCH)	valine	153 [e]
4/9.50	-	-	3.14 (s, CH3)	creatinine	44 [e]
	153.9	133.9	2.14 (s,SCH_3), 2.24 (m, βCH_2), 2.67 (t, γCH_2), 4.01(dd, αCH_2)	methionine	14 [e]
5/10.01	135.9	88.3	0.98(t, 6.2, δCH_3), 1.74(m, γCH), 1.81(m, βCH_2), 3.91(m, αCH)	leucine	68 [e]
	135.9	88.3	0.96(t, 7.4, δCH_3), 1.04(d, 7.0, βCH_3), 1.32(m, γCH), 1.51(m, γCH), 3.86(d, 3.7, αCH)	isoleucine	29 [e]
6/12.11	169.8 166.8[d]	122.0	3.16(m, βCH), 3.32(m, βCH'), 4.12(dd, αCH), 7.34(m, ring H6), 7.39(m, ring H4), 7.44(m, ring H3/H5)	phenylalanine	53 [e]
7/14.27	183.0	-	3.50(s, CH_3), 3.95(s, CH_3), 7.93(s, CH)	paraxanthine[a]	-
8/17.26	209.9 202.8[c]	190.8	3.38(m, βCH), 3.52(m, βCH'), 4.21(m, αCH), 7.21(t, ring H5/H6), 7.29(t, ring H5/H6), 7.33(s, ring H2), 7.55(d, ring H7), 7.72(d, ring H4)	tryptophan[a]	10 [f]
9/17.79	-	-	3.91(s, CH_2), 6.98(d, ring H3/H5), 7.98(d, ring H2/H6)	4-amino-hippurate[a,b]	-
11/24.19	196.0 194.9[d]	138.8	3.37(s, CH_3), 3.56(s, CH_3), 3.98(s, CH_3), 7.92(s, ring H2)	caffeine[a]	-
13/29.54	-	-	4.19(s, CH_2), 7.58(t, H3/H5), 7.86(d, H2/H6)	hippurate[a]	-
14/35.50	-	-	0.91(t, 7.4, CH_3), 1.11(d, 7.0, CH_3), 1.45(m, CH), 1.72(m, CH'), 3.00(m, CH)	3-methyl-2-oxovalerate[a]	-

The ^{1}H NMR spectrum from loop 9 (17.79 min) was tentatively assigned to 4-amino-hippurate (Figure 4.20d). This compound had been mainly used as ultrasound imaging contrast medium used to measure amniotic fluid volume (Kirshon *et al.*, 1990). However this procedure was apparently not used in any of the donors of AF samples used, therefore the compound may be of endogenous origin. The signals from ^{1}H NMR spectrum of loop 14 collected at 35.5min (Figure 4.20f) were assigned to metabolite 3-methyl-2-oxovalerate which was identified for the first time in AF. Six spin-systems were found for which no assignment was possible at this stage and also the spin-systems appearing at several retention times (system contaminants) (Table 4.6).

Table 4.6 - Unassigned spin systems and signals from contaminants detected in the LC-NMR/MS experiments of human AF using a RP-C18 column. Loop numbers according to Figure 4.19.

Loop/ RT(min)	Unassigned spin systems δ^{1}H, ppm (multiplicity)	Signals from contaminants δ^{1}H, ppm (multiplicity)	
2/7.12	3.30(s); 3.23(s); 3.60(s); 4.48(s); 7.37(s), 7.43(s)	3.56(dd), 3.65(dd), 3.77(m)	glycerol (membrane filter)
3/7.93	1.15(d); 1.25(d); 1.10(m)		
4/9.50	1.14(d), 3.45(m), 3.53(dd); 3.88(d)	3.35(s)	methanol (system)
5/10.01	1.14(d); 1.18(t); 3.14(s); 3.65(s)	0.93(t); 0.95(s); 1.14(s);	unidentified (system)
9/17.79	3.33(s); 3.91(s); 3.97(s)	1.18(d); 1.43(d); 1.26(t)	unidentified (system)
12/29.09	3.19(s); 3.66(s); 4.19(s); 7.37(t), 7.58(m)		

The pre-concentration of the peaks eluted between 15 and 40 min (less polar metabolites) into hydrophobic SPE cartridges for LC-SPE-NMR analysis was attempted. However, retention of compounds in these cartridges was unsuccessful since no ^{1}H NMR signals were obtained after elution of the cartridges.

4.3.3.2 Characterization of more polar metabolites

The more polar fraction of AF metabolites co-eluted in the very early peaks of reverse-phase chromatography, as shown in the previous section, thus hampering their resolution and detection of those compounds present in lower concentrations. Therefore, in order to characterize the more polar compounds of human AF, cation exchange chromatography was performed prior to NMR and MS analysis. Several peaks can be observed across the time range of the chromatogram UV trace obtained (Figure 4.21a) showing that good retention was achieved for the separated metabolites. The MS trace (Figure 4.21b) shows that the MS signal was quite low when compared with the obtained reverse-phase MS signal (Figure 4.19b, page 110) with signals occurring only at 10-20 min. This is justifiable by the lack of organic solvents in the composition of the mobile phase, which facilitate the electrospray ionization process (Kebarle and Verkerk, 2009). However, their use was not possible due to the incompatibility with this particular column.

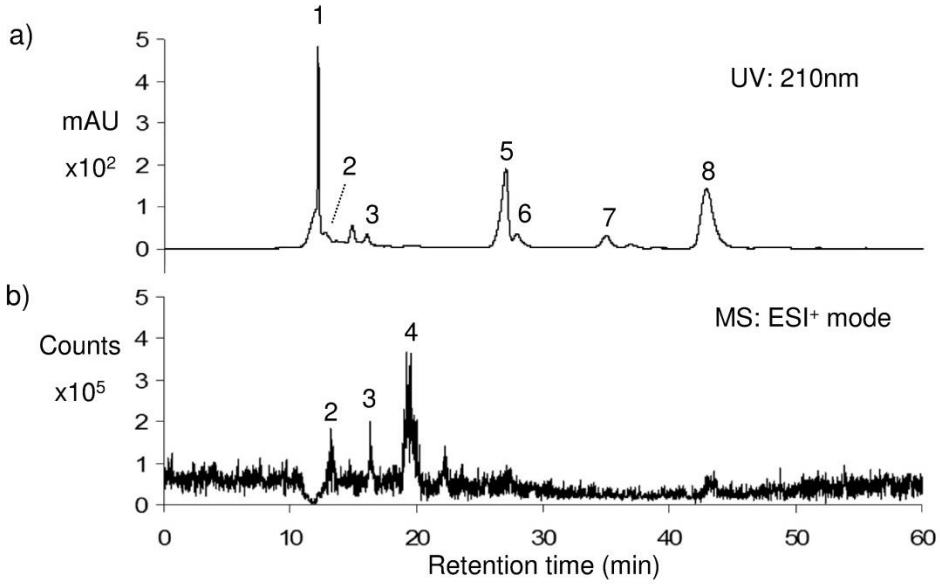


Figure 4.21 - Cation-exchange chromatograms of a 10× concentrated M_w<3kDa human AF fraction: a) UV trace, b) MS ESI^+ mode trace. 1, column void elution volume; 2, oligosaccharides; 3, pentose; 4, monosaccharides (glucose, *myo*-inositol, mannose and fructose); 5, lactate; 6, succinate and formate; 7, acetate; 8, pyroglutamate.

Figure 4.22a shows the pseudo-2D LC-NMR time-slice chromatogram obtained for a 10× concentrated M_w<3kDa AF fraction using cation-exchange chromatography. Each line corresponded to an 1D ^{1}H NMR spectrum obtained at every 30s of the chromatography as exemplified in Figure 4.22b for 4 spectra obtained at 4 different retention times of pseudo-2D LC-NMR chromatogram.

Table 4.7 lists the 32 compounds assigned in this experiment, 16 of which having been already assigned with ^{1}H NMR alone. The spin-systems of metabolites such as *myo*-inositol, α-hydroxybutyrate, α-hydroxyisobutyrate, β-hydroxybutyrate, α-hydroxyisobutyrate, α-ketoglutarate and taurine were partially overlapped with spin-systems of other compounds in 1D and 2D ^{1}H NMR spectra being completely resolved in this LC-NMR experiment, thus allowing their confirmation. Two other metabolites, which were observed in the reverse-phase experiment, namely mannose and 3-methyl-2-oxovalerate (Figure 4.22b), were also identified in the cation exchange experiment, thus confirming their assignments.

Newly identified metabolites included malate (Figure 4.22b), hydroxypropionate (Figure 4.22b), D-fructose, sorbitol, acetylcholine and pyroglutamate (a cyclic derivative of glutamate) which were previously found in human AF in other studies (Kumps *et al.*, 2004; Jauniaux *et al.*, 2005; Tuma *et al.*, 2006). These compounds were completely resolved by cation exchange LC which enabled their identification for the first time using ^{1}H NMR.

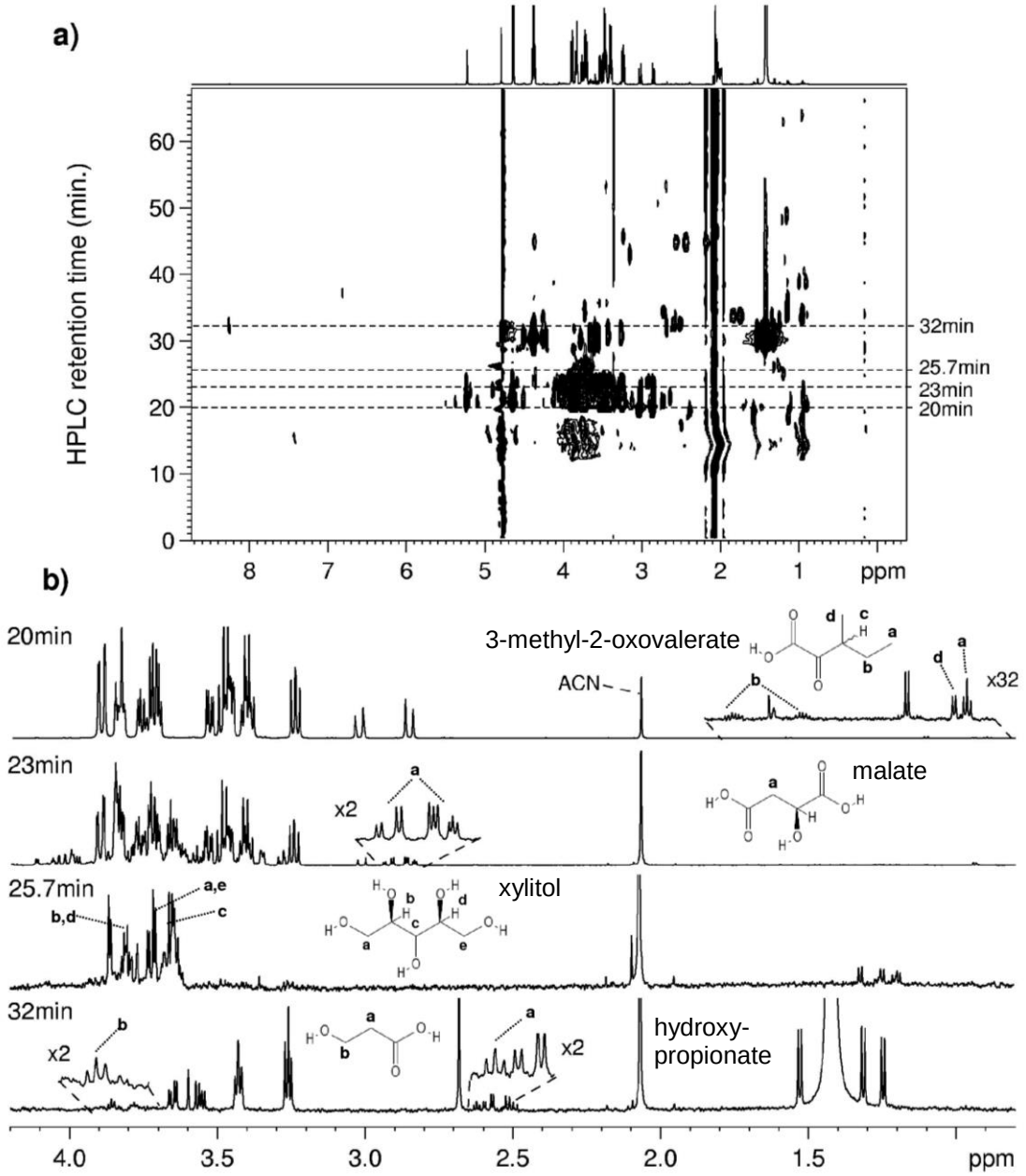


Figure 4.22 - a) Pseudo-2D LC-NMR chromatogram of a 10-fold concentrated AF fraction with M_w<3kDa recorded with an cation-exchange LC-NMR experiment at 600MHz with cryo-probe (the apparent distortion at earlier RT's is due to empty volume effects) and b) selected lines taken from a), for retention times of 20 min (containing 3-methyl-2-oxovalerate), 23 min (containing malate), 25.7 min (containing xylitol) and 32 min (containing hydroxypropionate). RG: 724 and NS: 64 for all lines. Un: unassigned peak.

Table 4.7 – Results of LC-NMR on human AF using an cation-exchange column. [a], newly found by NMR; [b], tentative assignment. MS and MS/MS m/z values correspond to $[M+Na]^+$ adducts except [c], $[M+H]^+$. Expected concentration values (2nd trimester pregnancy) are shown in the last column and respective references: [d], (Lind, 1981); [e], (Groenen *et al.*, 2004); [f], (Kumps *et al.*, 2004).

RT (min)	MS (m/z)		δ^1H, ppm (multiplicity, *J* in Hz, assignment)	Compound	expected concentration in μM (*ref*)
	MS	MS/MS			
14.0			3.2 (s), 3.3 (s), 3.6-4.0 (broad profile)	methoxylated oligosaccharide (U2)[a]	
16.0	407.0	214.5	3.6-3.9, 4.35, 4.99 (d), 7.40 (s)	pentose[a,b]	
19.0			0.94(d, 7.0, γCH_3)	isovalerate[a,b]	
19.5			2.85(d, CH_2, 15.8), 3.03(d, 15.8, CH_2)	citrate	345 [d]
			2.38(s, CH_3)	pyruvate	
			1.58(s, CH_3)	pyruvate (hydrate form)	
20.0			0.89(t, 7.5, δCH_3), 1.09(d, 6.8, $\beta'CH_3$), 1.46(m, γCH), 1.68(m, γ'CH), 2.92(m, βCH)	3-methyl-2-oxovalerate[a]	
20.5	202.8		3.41(t, H4), 3.52(dd, H2), 3.70(t, H3), 3.76(dd, H6), 3.83(m, H5), 5.23(d, H1)	α-D-glucose	2510 [d]
	202.8		3.24(dd, H2), 3.40(t, H4), 3.49(t, H3), 3.47(dd, H5), 3.73(dd, H'6), 3.90(dd, H6), 4.64(t, H1)	β-D-glucose	
	324.9	162.8[c]	3.12(dd), 3.29(m), 3.36(m), 3.59(m), 3.66(m), 3.96(dd), 4.01(dd), 4.12(m), 4.25(d), 4.32(t), 4.50(d), 5.50(d)	hexose disaccharide[a]	
21.0			0.94(d, 6.8, γCH_3), 2.09(m, βCH), 2.64(d, αCH)	2-oxoleucine[a]	
21.5	202.8		3.26(t, H5), 3.61(dd, H4/H6), 4.05(t, H2)	*myo*-inositol	396 [e]
22.0	202.8		3.56(t, H3), 3.64(m, H3/H4), 3.93(m, H2), 4.90(d, H'6), 5.18(d, H6)	mannose[a]	
22.5	202.8		3.57(m, H1), 3.80(d, H6), 3.99(m, H4), 4.03(dd, H5), 4.11(d, H3)	fructose[a]	
23.5			2.85(m, βCH), 2.92(m, βCH'), 4.57(dd, αCH)	malate[a]	
24.5	182.8[c]	164.8[c] 65.9[c]	3.63(m, H1/H4/H6), 3.73(d, H6), 3.76(m, H3), 3.82(d, H1), 3.84(m, H2/H5)	sorbitol[a]	
25.7			3.65(m, H3), 3.72(dd, H1/H5), 3.80(m, H2/H4)	xylitol[a]	
30.5			1.42(d, 7.1, CH_3), 4.38(q, 7.0, CH)	lactate	9800 [d]
30.8			3.26(t, 6.6, CH_2), 3.43(t, 6.6, CH_2)	taurine	83 [d]
31.5			2.62(t, 6.0, CH_2), 3.86(t, 5.9, CH_2)	hydroxypropionate[a]	9.5-27 [f]
32.0			2.68(s, CH_2)	succinate	
			8.45(s, CH)	formate	
32.5			1.25(d, 6.4, γCH_3), 2.51(m, βCH), 2.59(m, βCH'), 4.24(m, αCH)	β-hydroxybutyrate	120 [d]
33.5			1.34(s, CH_2), 2.57(s, CH_2)	α-hydroxyisobutyrate	
33.7			0.96(t, 7.5, γCH_3), 1.74(m, βCH), 1.84(m, βCH'), 4.25(m, αCH)	α-hydroxybutyrate	50 [d]
35.0			2.32(s, CH_3)	acetoacetic	
35.5			2.09(s, CH_3)	acetate	
37.2			6.81(s, CH)	fumarate	
38.5			0.90(d, 6.9, γCH_3), 1.00(d, 6.9, $\gamma'CH_3$), 2.06(m, βCH), 4.12(d, 4.3, αCH)	α-hydroxyisovalerate	
44.0			2.06(s, NCH_3), 3.96(s, αCH_2)	N-acetylglycine[a]	
45.0			2.17(m, γCH), 2.43(m, βCH_2), 2.56(m, γCH'), 4.37(dd, αCH)	pyroglutamate[a,b]	[19-50] [f]
46.0			3.23(s, CH_3)	acetylcholine[a,b]	
46.5			1.41(d, 7.3, CH_3), 2.03(s, NCH_3)	N-acetylalanine[a]	
53.5			2.69(t, 7.0, CH_2), 3.45(t, 7.1, CH_2)	α-ketoglutarate[b]	

Further compounds were identified for the first time in human AF namely D-xylitol (Figure 4.22b), 2-oxoleucine, the amino acids derivatives N-acetylalanine, N-acetylglycine, as well as the tentatively assigned isovalerate (Table 4.7).

Again, 31 spin-systems were also found for which no assignment could be advanced (Table 4.8). The corresponding MS data did not contribute much to the assignment of these spin-systems due to the poor ionization mentioned previously. Apparently, there is no link between the systems listed in Table 4.8 and the unassigned spin-systems found by ^{1}H NMR (Table 4.3), which seems to suggest that the unassigned spin-systems detected in the LC-NMR cation-exchange originate from less concentrated compounds, therefore undetected by ^{1}H NMR alone.

Table 4.8 – Unassigned spin-systems obtained from the LC-NMR experiment of human AF using a cation-exchange column.

RT (min)	δ^{1}H, ppm (multiplicity)	RT (min)	δ^{1}H, ppm (multiplicity)	RT (min)	δ^{1}H, ppm (multiplicity)
12.5	1.0 (m), 1.6 (d), 3.9 (s) (U1)	25.5	3.62(m), 3.68(m), 3.78(dd)	35.8	3.44(dd), 3.54(dd)
15.0	3.91(s)		3.56(dd), 4.35(d)	40.5	0.93(d), 3.74(t)
16.0	3.20(s), 3.33(s)	25.8	1.26(d), 1.32(d)	40.8	2.46(s)
	3.77(m), 3.97(d), 4.99(m),	26.5	1.27(d), 1.43(s)	42.5	1.18(t)
	7.46(s) (U3)	28.8	3.86(t), 4.54(s), 4.58(s)	43.0	3.15(s)
16.5	2.50(d), 2.75(m)	29.5	3.60(s)	43.5	1.10(t)
18.0	0.95(d)	30.0	4.21(s)	48.0	1.22(s)
	1.13(d)	35.5	1.14(d, 7.1), 2.72(m),	49.0	1.15(d)
20.3	2.73(dd), 2.88(dd)		3.72(d, 6.2)	50.5	2.80(s)
23.5	3.65(m)		3.67(s)	63.0	1.20(d)
24.5	1.19(d)	36.0	2.20(s)	64.0	0.96(d)

Regarding sugar and sugar-like compounds, the newly found D-xylitol and sorbitol, like caffeine and paraxanthine, may also possibly originate from dietary sources, such as from the consumption of sweeteners or some fruits (plums, berries, etc.). There is also the possibility of an endogenous origin of these metabolites, due to biosynthesis by the embryo during the first trimester of pregnancy, as it has been suggested in a previous study (Jauniaux *et al.*, 2005). An hexose disaccharide that co-eluted with glucose, was also identified at 20.5 min retention time and confirmed by MS (Table 4.7) but for which no further assignment was possible at this stage. In addition, an interesting group of weaker signals resonating in the sugar region may be seen at retention times between 12 and 18min (Figure 4.22a). These signals reflect the overlap of at least 3 different spin-systems (U1, U2 and U3) shown in detail in Figure 4.23, and confirmed by ^{1}H-^{1}H TOSCY in the corresponding fraction collected by loop-storage at about 16 min (not shown). The further characterization of these spin-systems was attempted by DOSY, but this was hindered by the relatively low signal-to-noise characterizing the 12-18 min retention time region.

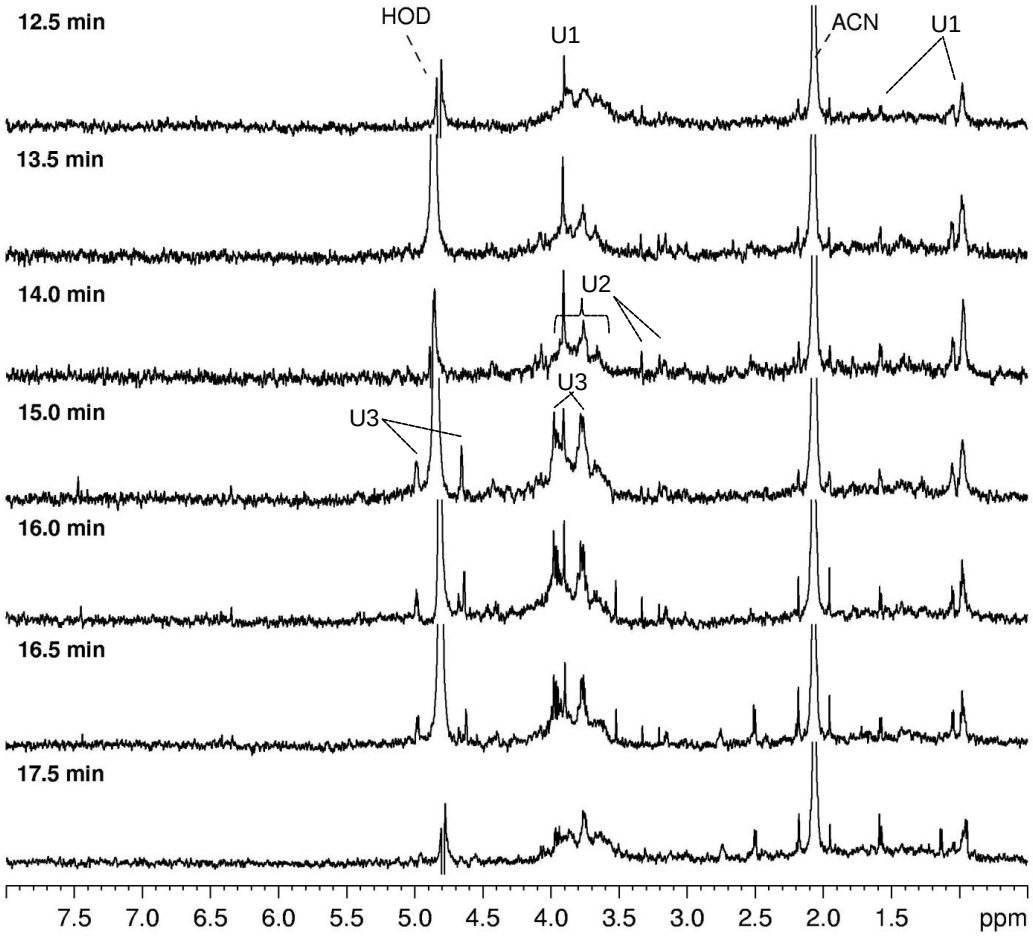


Figure 4.23 - ^{1}H NMR spectra taken from the LC-NMR chromatogram shown in figure 4.22 for retention times between 12 and 18 min. The spectra were re-processed with 2Hz line broadening and recalibrated to the ACN peak.

Although the identification of U1 was not possible at this stage, the profiles of U2 and U3 resemble sugar-like profiles. The relatively broader profile of U2 and the singlets at 3.2 and 3.3 ppm (Figure 4.23) may indicate a possibly methoxylated oligosaccharide (Table 4.7). The U3 system shows clearly resonances at the anomeric region, together with an unassigned resonance at 7.4 ppm, while the MS spectra at this retention time have revealed that this compound shows ions compatible with a pentose sugar (Table 4.7). Therefore U3 was tentatively assigned to a pentose sugar (Table 4.7).

It is interesting to note that none of the amino acids observed previously in the reverse-phase experiment (Table 4.5) was observed in the LC-NMR cation-exchange experiment. This may have resulted from the protonated states of most amino acids at the pH of the mobile phase (pH≈2.4), especially the protonation of the NH_2 groups, which then confers high positive charge densities. This may have led to a higher affinity of these compounds to the stationary phase of the column and, as a result the amino acids are not eluted. In spite of this, taurine (Figure 4.24a) and the

newly detected amino acid derivatives N-acetylglycine, N-acetylalanine and pyroglutamate were observed in the chromatograms (Figure 4.24b-c).

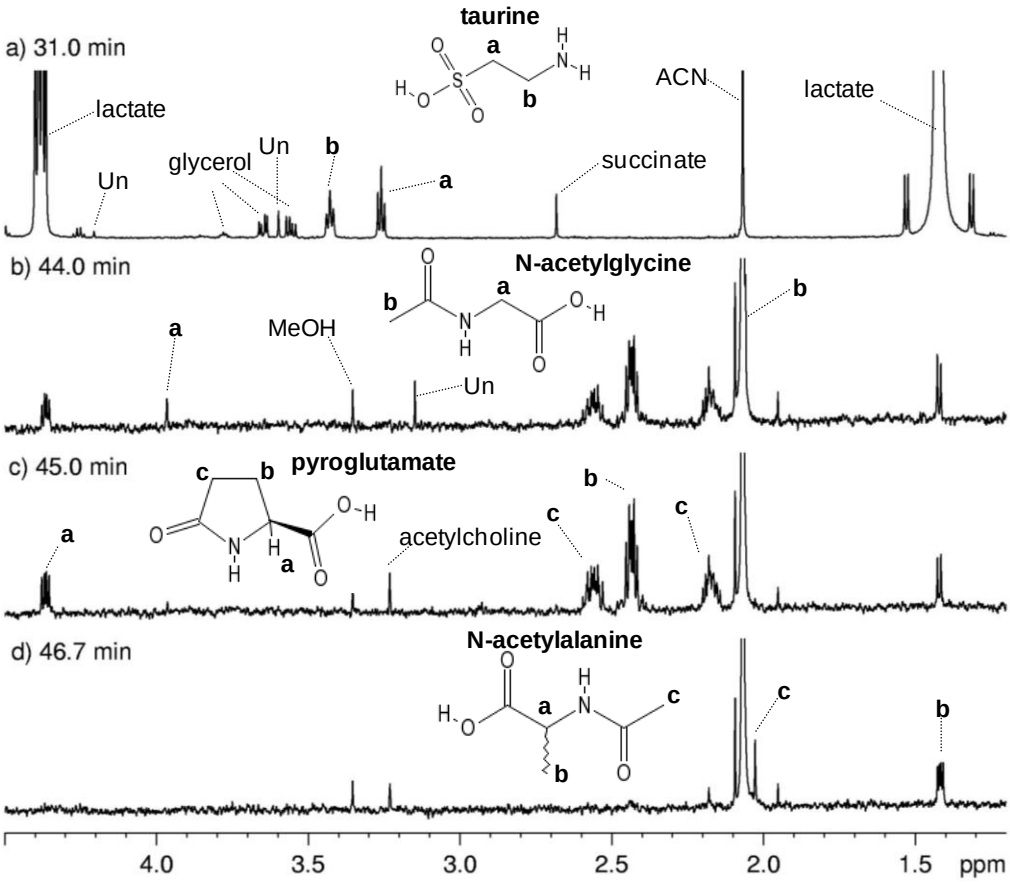


Figure 4.24 - ^{1}H NMR spectra taken from the LC-NMR chromatogram shown in Figure 4.22, showing the elution of some amino acids derivatives: a) taurine, b) N-acetylglycine, c) pyroglutamate and d) N-acetylalanine.

In these compounds, the NH_2 group is replaced by a NHR group, where R can be an acetyl or a methylene group (Figure 4.24b-c). Furthermore, in the case of taurine, which has free NH_2 group (Figure 4.24a), the charge density is compensated by the negatively charged sulfonic group. These effects have resulted in a decrease in the overall positive charge of these compounds at the pH≈2.4, which then could have resulted in the high retention times observed for these compounds.

Overall, the results from reverse-phase and cation-exchange LC-NMR/MS experiments enabled the detection and assignment of 21 new metabolites, which adding to those identified by 1D and 2D NMR enabled the establishment of a database of 75 compounds detected in healthy human AF. The complete list of compounds can be found in Annex V – Full list of compounds identified in AF by high resolution NMR and LC-NMR/MS.

5. Study of the effects of fetal malformations (FM) and chromosomal disorders (CD) on 2nd trimester human amniotic fluid

The work presented in the first part of this chapter aims to further explore the metabolic effects of fetal malformations on 2nd trimester human amniotic fluid using a metabonomic approach comprising three complementary analytical techniques: ^{1}H NMR, UPLC-MS and MIR. The second part of this chapter explores the human AF metabolite variations resulting from chromosomal disorders using ^{1}H NMR and MIR metabonomics. The ^{1}H NMR, UPLC-MS and MIR analytical techniques were chosen since they have different dynamic ranges in terms of metabolites detected, therefore, providing complementary information on the complex AF metabolite composition.

5.1. Study of the effects of FM on 2nd trimester human amniotic fluid

5.1.1. ^{1}H NMR profiling to evaluate the effects of maternal age, gestational age and malformation type

The fetal malformations group of samples considered in this study comprised different types of malformations grouped according to the affected organic system, as summarized in Table 5.1. Due the time-scale of sample collection, the number of samples (control and FM) varied between the three analytical studies as indicated in the same table.

Table 5.1 - Sample numbers and corresponding median age values of gestational and maternal ages of control and fetal malformation (FM) groups. Numbers in brackets: age range.

	Age		*Sample numbers per study*		
	Gestational	**Maternal**	**^{1}H NMR**	**UPLC-MS**	**MIR**
Control:	17 (15-22)	36 (25-44)	90	26	41
Fetal Malformation(FM) type:					
Urogenital	20 (14-24)	30 (25-33)	5	5	5
Cardiac	22 (15-23)	22 (13-32)	5	3	4
Soft tissues	14, 21	21, 22	2	1	2
Abdominal	16, 23	27, 37	2	2	2
Limbs	22 (16-23)	29 (19-31)	5	1	5
Lungs	17	35	1	1	1
Poly-malformed	19 (15-25)	27 (17-36)	3	2	3
Central Nervous System	22 (16-24)	32 (24-35)	13	7	13
Total FM sample numbers:			36	22	35

Considering the median age values, differences of 3 to 5 weeks in gestational age and of 4 to 15 years in maternal age were noted between control and FM samples (Table 5.1). This was justified by fact that amniocentesis performed for controls was usually carried out on women with ≥ 36 years of age, at around 16 weeks of gestation. On the other hand, for FM cases (which can be detected at any maternal age) amniocentesis was usually performed after diagnosis of FM, and can thus span a wider range of gestational ages of 2nd and 3rd trimester (14-36 weeks). In order to evaluate the possible influence of gestational and maternal ages in the AF composition, ^{1}H NMR standard spectra of control samples (15-22 g.w.) were analysed by PCA, without including any disorder spectra in this analysis. The PCA scores scatter plot obtained illustrates that the gestation age range of the analysed samples spread across the scatter plot (Figure 5.1a), thus, showing that no measurable trends were observed with respect to gestational age.

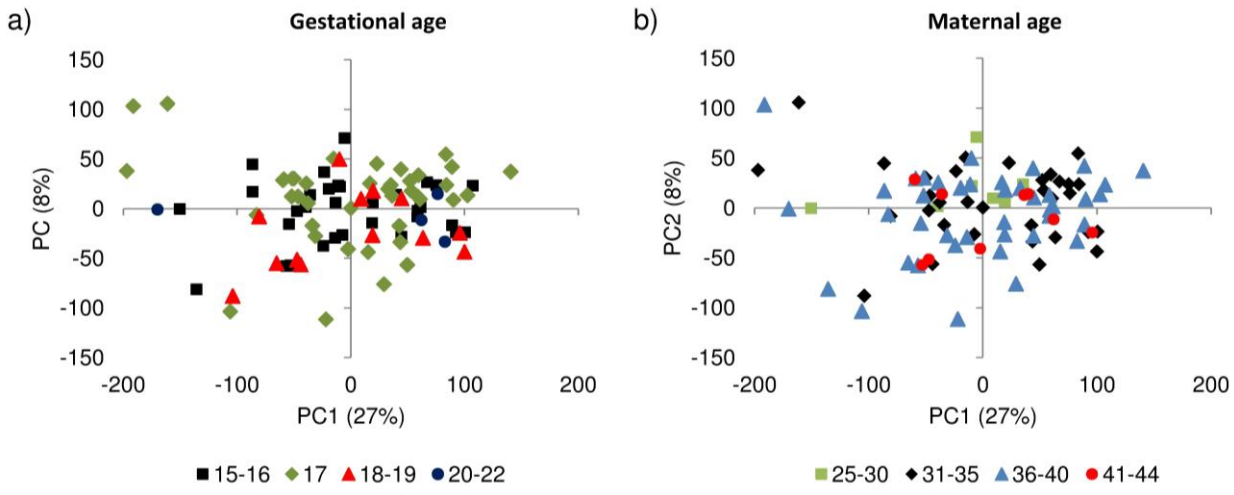


Figure 5.1 - PCA scores scatter plots of standard ^{1}H NMR spectra of control samples (n=90) with: a) gestational and b) maternal ages sub-ranges represented.

Similar results were obtained when the PCA scores scatter plot data points were coloured according to the maternal age ranges (Figure 5.1b). The results shown indicate that any AF compositional variations due to gestational or maternal ages, in the considered ranges, are not significant and should be much smaller that the inter-individual changes, thus showing the non-confounding influence of these variables in the data sets considered. The results are not surprising since the major compositional changes in AF occur mainly between trimesters (>10 g.w.), especially between the 2nd and the 3rd trimester as a response to fetal organ maturation and involving changes in the contents of amino acids, proteins, choline and related compounds and creatinine (Lind, 1981; Sims *et al.*, 1996; Cohn *et al.*, 2009).

The FM group was composed of malformations affecting several organic systems (Table 5.1). In order to inspect the differences between the FM types, their spectra were analysed by PCA using the different ^{1}H NMR spectra collected: standard, CPMG and diffusion-edited (Figure 5.2).

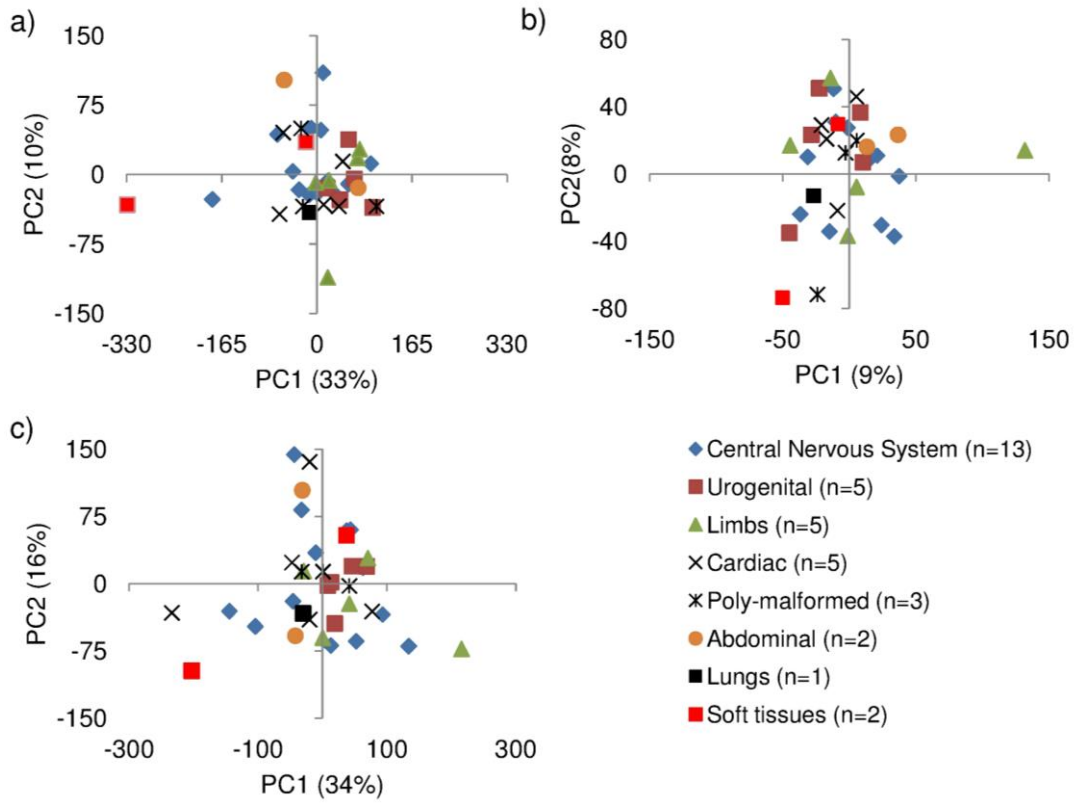


Figure 5.2 - PCA scores scatter plots of ^{1}H NMR spectra of AF samples from different fetal malformations grouped according to the organic system affected: a) standard, b) CPMG, c) diffusion-edited spectra.

From the inspection of the PCA scatter plots (Figure 5.2) it is clearly seen that different types of FM overlap, showing that the corresponding ^{1}H NMR spectra were indistinguishable. This may result partially from the low sample numbers corresponding to different FM organic types. Moreover, each organic type of FM considered was composed of different types of FM (as shown in chapter 3, Table 3.3, page 60), which may also justify the spread observed in the PCA scores for the samples of the same organic type (Figure 5.2). PLS-DA models were also obtained to evaluate the differences between central nervous system samples (largest group, n=13) and other FM (n=23) types (not shown), since the former group could not be differentiated from the other FM in PCA. However, the models could not be validated by MCCV, hence reflecting the absence of significant differences between both groups.

Due to the low sample number available for each malformation type, the FM group was considered as a whole when compared with the control group in the analysis described below, therefore attempting to identify general metabolic effects related to FM.

5.1.2. ^{1}H NMR profiling for comparison of control and FM groups

When analysed by PCA, the spectra of control and FM samples showed separation tendencies in the first two components (PC1 and PC2) for standard and CPMG spectra (Figure 5.3a and b, left) indicating that compositional differences between control and FM were registered in this two experiments.

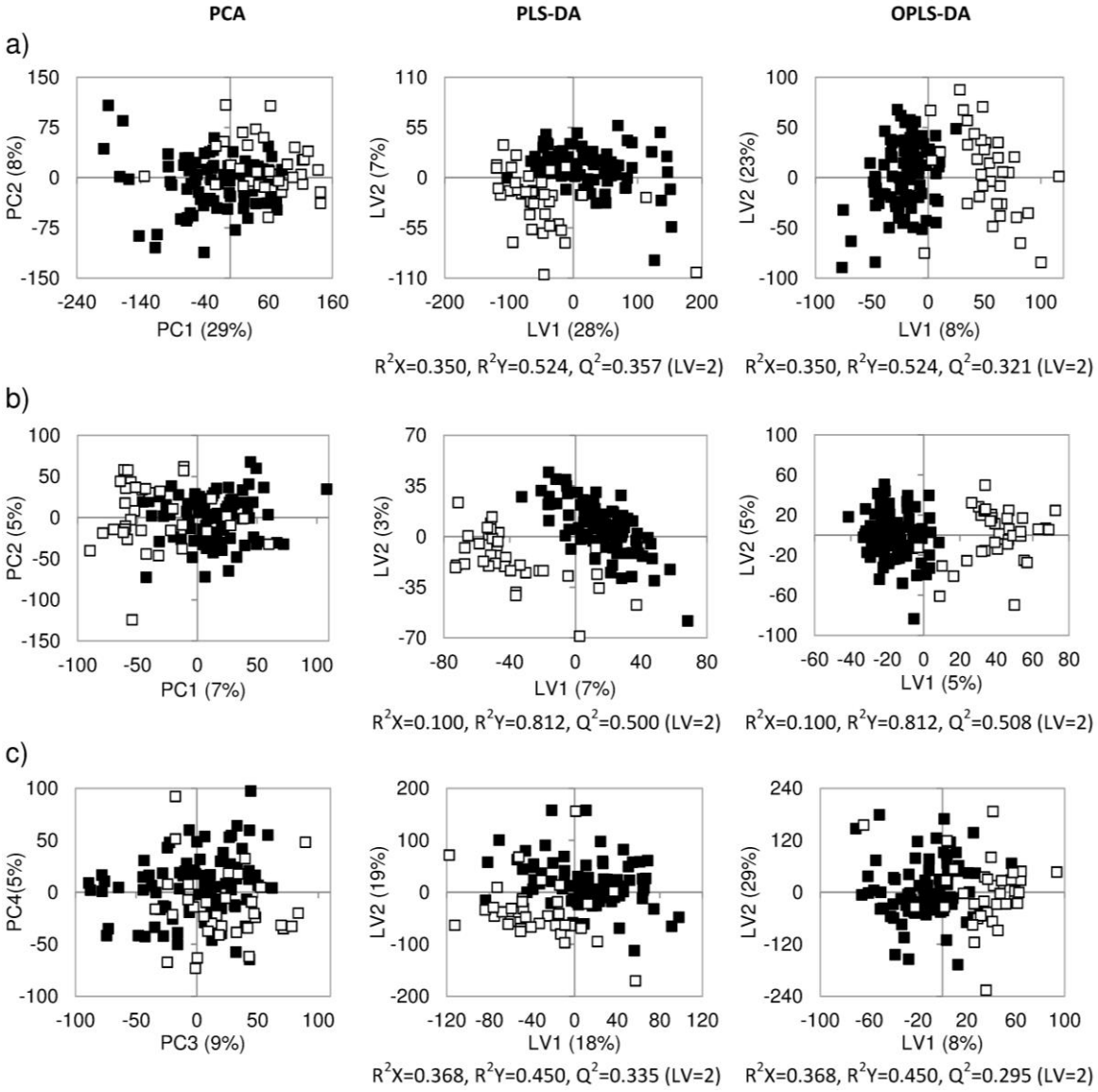


Figure 5.3 - Scores scatter plots from PCA (left), PLS-DA (centre) and OPLS-DA (right) models of control (■, n=90) *vs* FM (□, n=36) samples: a) standard; b) CPMG; c) diffusion-edited ^{1}H NMR spectra.

A small separation tendency was also obtained in the PCA scores scatter plot of diffusion-edited spectra, but only when PC3 and PC4 components were considered (Figure 5.3c, left) which indicate less compositional differences between control and FM regarding larger compounds (proteins mainly). PLS-DA enabled a better separation of control and FM samples (Figure 5.3, centre) although with some overlap of FM with control samples seen for standard and diffusion-edited spectra (Figure 5.3a and c, centre) and an almost complete separation obtained for CPMG spectra in the LV1 axis (Figure 5.3b, centre). Similar separation trends were obtained in the OPLS-DA scores scatter plots in the first component (LV1) (Figure 5.3, right). The PLS-DA models obtained registered moderate predictive power (Q^2) values between 0.335 and 0.500, and the full validation of these models was pursued by Monte-Carlo cross-validation (MCCV) (Figure 5.4).

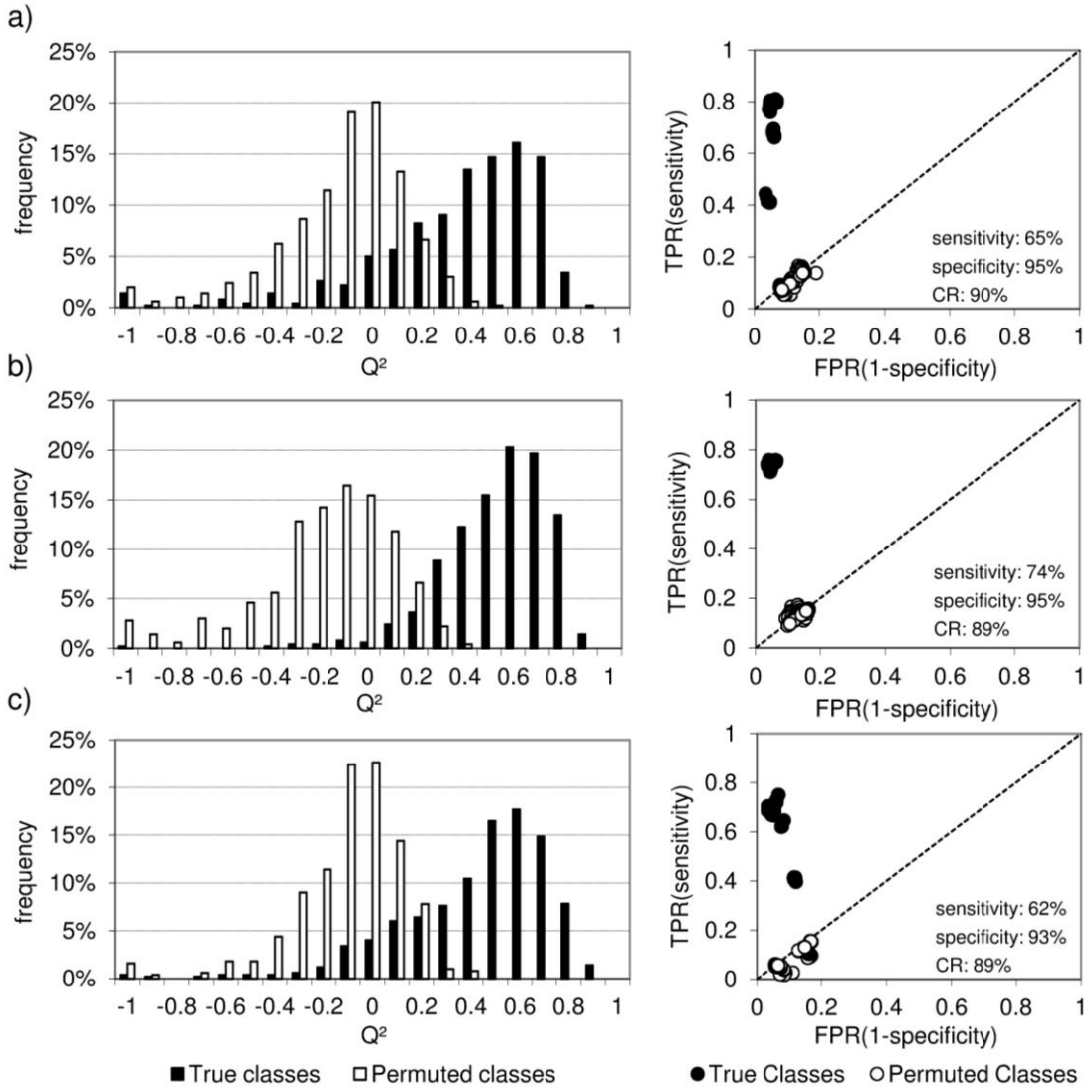


Figure 5.4 - MCCV results for PLS-DA models of control *vs* FM: a) standard, b) CPMG, c) diffusion-edited ^{1}H NMR spectra. Q^2 distributions (left) and ROC plots (right) of true classes and permuted classes models are shown. The true classes average sensitivity, specificity and classification rate (CR) are indicated in the ROC plots.

Separation of Q^2 distributions of models of true and permuted classes were obtained for all models (Figure 5.4, left), with less overlap observed for the CPMG data set (Figure 5.4b, left). The ROC plots show also complete separation of true and permuted classes models (Figure 5.4, right) and high specificity values >80% for all data sets. In terms of sensitivity, the CPMG models registered average values of 74% for true class models (Figure 5.4b, right), the standard spectra registered sensitivity values between 40%-80% (Figure 5.4a, right) and diffusion-edited registered sensitivity values between 40%-75% (Figure 5.4c, right). The dispersion in sensitivity observed for these models indicates a poorer predictive ability when compared with CPMG models. The classification rates obtained were however similar for the 3 data sets: 90% for standard spectra models and 89% for CPMG and diffusion-edited spectra. The overall MCCV results demonstrate the robustness of human AF ^{1}H NMR PLS-DA models in the discrimination of control and FM samples, with the CPMG spectra providing the most robust models, which means that low M_w metabolites seem to contribute more importantly for that discrimination. Interestingly, the sensitivity values obtained for the ^{1}H NMR models for FM detection (62-74%) were within the range of sensitivities for FM detection associated with ultrasound methods (14-74%, Chmait and Moore, 2004) with a slight improvement towards intermediate values. However, the specificity of ultrasound methods is higher for FM detection is higher (100%, Chmait and Moore, 2004) compared to the specificity of ^{1}H NMR models (93-95%).

The ^{1}H NMR signals responsible for the separation observed in the PLS-DA scores scatter plots shown in Figure 5.3 (centre) were identified from the corresponding loadings weights plots. In the loadings weights plots (Figure 5.5) the signals pointing in the positive direction of the separating latent variables (LV1 for standard and for CPMG; and LV2 for diffusion-edited data sets) are associated with higher content of the corresponding metabolites in the control group, which were characterized by positive values of the latent variable in the corresponding scores scatter plots (Figure 5.3, centre); hence these signals are associated with lower contents of the corresponding metabolites in the samples of the FM group. The negative signals, on the other hand, are associated with higher contents of the corresponding metabolites in the samples of the FM group. Moreover the signals in orange-red colour, registering higher variable importance to the projection (VIP) values, indicate the signals which had a higher contribution to the class separation observed (Figure 5.5). In this way, several metabolite signals decreased in FM group were identified from the loadings of standard spectra (Figure 5.5a) namely the amino acids isoleucine, leucine, valine, alanine, lysine, proline, glutamine, histidine, phenylalanine and tyrosine, together with decreases in glucose, mannose, *myo*-inositol, lactate, choline, creatine,

citrate, pyruvate, formate, fumarate, α-oxoisovalerate and two unassigned spin-systems, Un1 (2.84 ppm, multiplet) and Un2 (2.93, 3.15 ppm, multiplets). Broad signals were found increased in the spectra from FM group samples, namely at 0.60 and 7.00 ppm (Figure 5.5a).

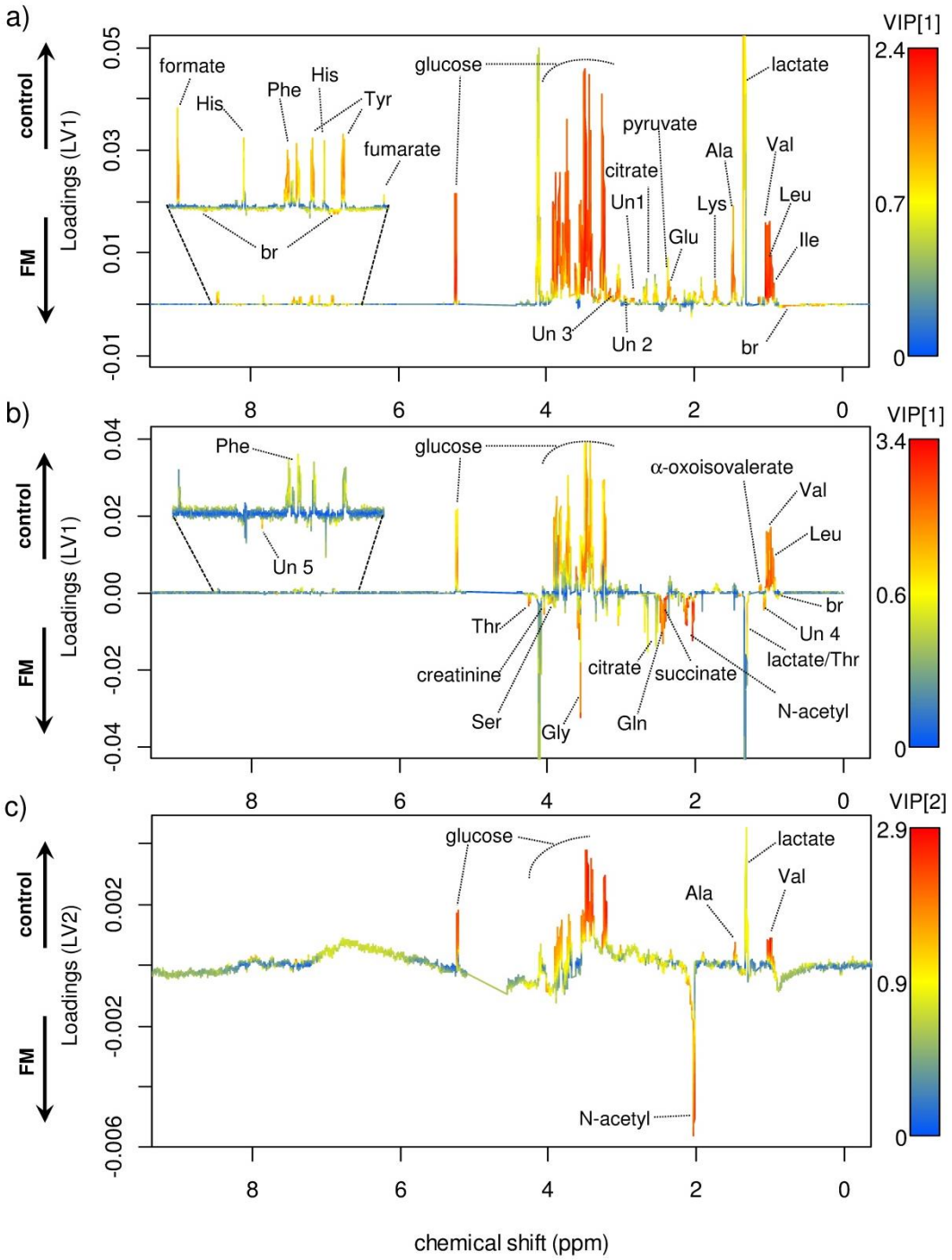


Figure 5.5 - Loadings from PLS-DA models of control *vs* FM: a) standard, b) CPMG and c) diffusion-edited ^{1}H NMR spectra colour-coded with VIP values, corresponding to the scores in Figures 5.3a,b,c (centre), respectively. Ui, unassigned signals; br, broad profile; amino acids in three-letter code.

The CPMG spectra loadings (Figure 5.5b) further confirmed the decreases of isoleucine, leucine, valine, α-oxoisovalerate, glucose, phenylalanine and tyrosine in the samples of the FM group, as previously noted in the loadings of standard spectra. Furthermore, increases in the signals of N-acetyl groups (from glycoproteins), glutamine, succinate, creatinine, glycine, serine, ascorbate and threonine as well as in 2 unassigned signals, Un3 (1.06 ppm) and Un4 (7.67 ppm), were noted in the FM group. The diffusion-edited spectra loadings (Figure 5.5c) confirmed the decrease in lactate as well as the decreases in glucose, alanine and valine as well the increase in N-acetyl groups of glycoproteins in the FM group (Figure 5.5c).

Interestingly, lactate and citrate signals showed an inverse variation in CPMG loadings when compared to standard spectra and diffusion-edited spectra loadings (Figure 5.5) indicating an increase of these metabolites in FM. As discussed in chapter 4 (pages 106-107), a small portion of citrate and lactate may be found bound to the protein. It seems then possible that the two forms of lactate and citrate may co-exist in AF samples: one free in solution and other tightly bound to proteins. The variations seen in the CPMG loadings should be mainly due to the "free" forms of citrate and lactate, whereas those of lactate in diffusion-edited spectra loadings should be mainly due to the "bound" form (both forms contributing to the standard spectra).

All the metabolite variations were further evaluated through univariate analysis after integration of the respective ^{1}H NMR signals. The list of significant metabolite variations obtained is shown in Table 5.2, with indication of the respective direction of variation and magnitude of change (mean fold change, i.e. the ratio between the average of integrals (A): $A_{control}/A_{FM}$ or $A_{FM}/A_{control}$, if $A_{FM}>A_{control}$). The univatiate tests confirmed the variations previously identified in the PLS-DA loadings. The integrals from selected metabolites identified as changing more significantly in the AF of FM samples (including glucose, valine, glutamine, glutamate, serine, succinate, N-acetyl group from glycoproteins) were expressed in box-plots (Figure 5.6). The box-plots corresponding to the variation of both forms of lactate ("free" and "bound") are also shown to better illustrate their change in controls and FM group (Figure 5.6).

Most of the metabolite variations identified had also been found in the early stages of the work presented in this thesis where similar analysis was performed using several FM types but with reduced numbers of both FM (n=27) and control (n=82) groups (Graça *et al.*, 2009; Graça *et al.*, 2010), therefore confirming the obtained findings. However, the variations of "free" and "bound" citrate, choline, creatine, fumarate, mannose, proline, unassigned signals U1, U2 and U3 signals, were found here for the first time in AF of FM samples, which may be solely due to the use of larger FM (n=36) and control (n=90) groups in the present work.

Table 5.2 - List of significant metabolite variations (p<0.05) noted in 2nd trimester human AF of FM affected pregnancies. [a], signal within the metabolite spin-system chosen for integration; [b], signal integral measured in CPMG spectra. s, singlet; d, doublet; t, triplet; m, multiplet; br, broad signal; Ui, unassigned signals. ( ), very small changes corresponding to magnitude (mean fold change) values >1.0 and <1.05.

metabolite	δ^1H [a]	direction and magnitude of change		*p*-value
alanine	1.47(d)	↓	1.3	7.43×10^{-9}
α-oxoisovalerate	1.11(d)	↓	1.1	6.53×10^{-9}
ascorbate [b]	4.01(m)	↑	1.3	2.95×10^{-3}
broad signal (proteins)	0.60(br)	(↑)	1.0	5.71×10^{-7}
choline	3.19(s)	↓	1.2	1.36×10^{-5}
citrate	2.67(d)	↓	1.1	9.77×10^{-3}
citrate ("free" form) [b]	2.67(d)	↑	1.2	6.75×10^{-4}
creatine	3.92(s)	↓	1.1	4.56×10^{-4}
creatinine [b]	4.05(s)	↑	1.1	1.04×10^{-4}
formate	8.45(s)	↓	1.3	3.11×10^{-7}
fumarate	6.51(s)	↓	1.1	4.52×10^{-4}
glucose	5.23(d)	↓	1.7	1.70×10^{-10}
glutamate	2.35(m)	↓	1.2	3.48×10^{-7}
glutamine [b]	2.45(m)	↑	1.6	2.04×10^{-6}
glycine [b]	3.55(s)	↑	1.2	4.60×10^{-6}
histidine	7.07(s)	↓	1.1	1.86×10^{-8}
isoleucine	0.94(t)	↓	1.1	1.32×10^{-8}
lactate	1.32(d)	↓	1.3	1.76×10^{-4}
lactate ("free" form) [b]	1.32(d)	↑	1.1	2.63×10^{-2}
leucine	0.95(t)	↓	1.3	3.77×10^{-11}
lysine	1.72(m)	↓	1.1	1.24×10^{-8}
mannose	5.19(d)	↓	1.2	3.54×10^{-4}
myo-inositol	4.06(t)	↓	1.2	1.97×10^{-4}
N-acetyl (glycoproteins) [b]	2.03(s)	↑	1.3	7.66×10^{-12}
phenylalanine	7.38(m)	↓	1.2	1.28×10^{-10}
proline	2.00(m)	↓	1.1	1.58×10^{-5}
pyruvate	2.37(s)	↓	1.2	4.37×10^{-6}
serine [b]	3.95(m)	↑	1.1	6.93×10^{-6}
succinate [b]	2.40(s)	↑	1.4	4.72×10^{-9}
threonine [b]	4.26(m)	↑	1.2	4.72×10^{-6}
tyrosine	6.89(d)	↓	1.2	4.30×10^{-10}
valine	1.03(d)	↓	1.4	1.62×10^{-11}
Unassigned resonances	**δ^1H [a]**	**direction and magnitude of change**		***p*-value**
Un1	2.84(m)	↓	1.1	5.79×10^{-7}
Un2	2.93(m)	↓	1.1	6.93×10^{-6}
Un3	3.10(m)	↓	1.2	3.77×10^{-11}
Un4 [b]	1.06(d)	↑	1.1	2.34×10^{-3}
Un5 [b]	7.67(s)	↑	1.7	4.09×10^{-4}

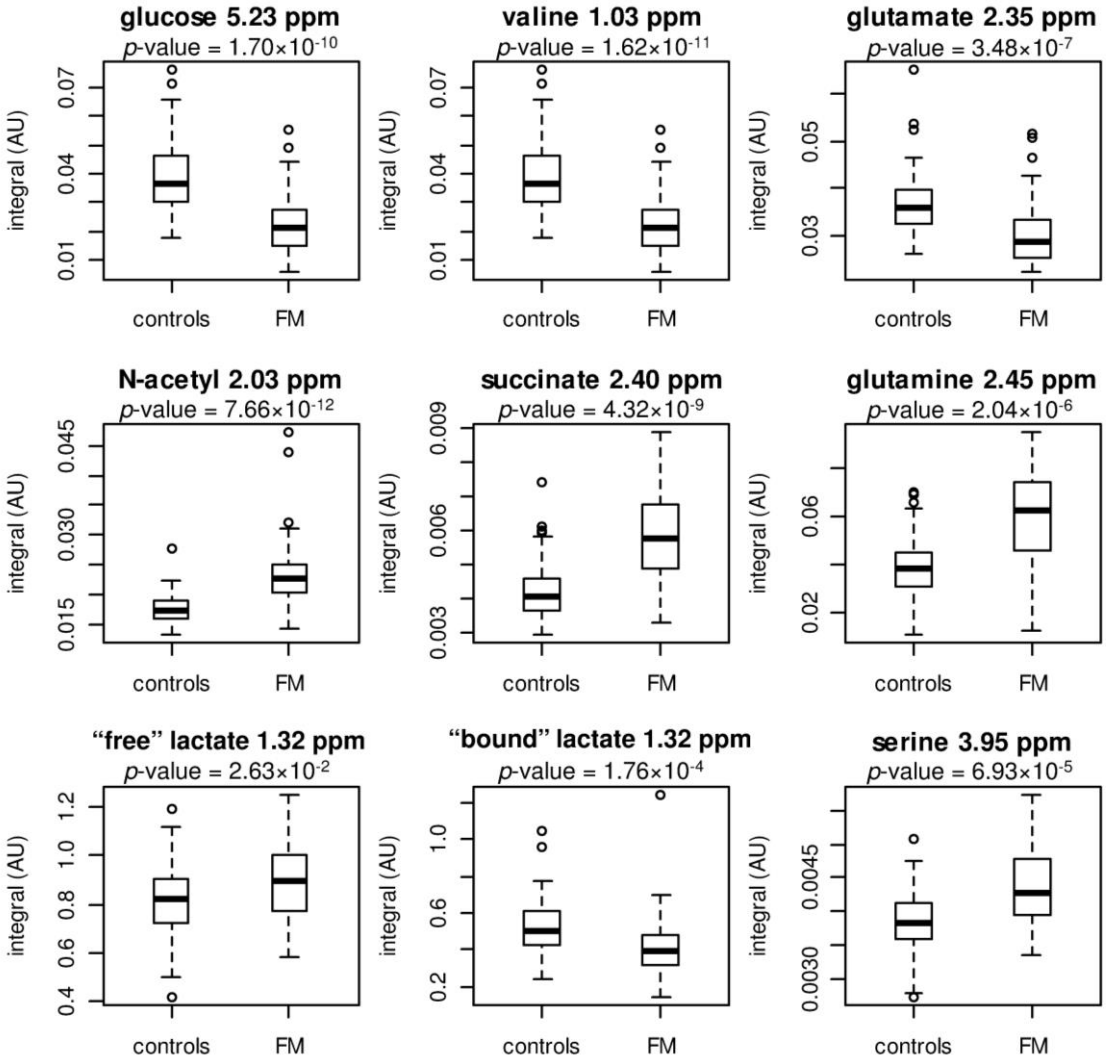


Figure 5.6 - Box-plots for selected metabolites identified as changing significantly in the loadings of AF FM ^{1}H NMR PLS-DA models. The respective *p*-values are also shown.

5.1.3. UPLC-MS profiling for comparison of control and FM groups

Further characterization of FM cases was performed through human AF profiling using UPLC-MS, a technique which offered more sensitivity and greater dynamic range when compared with ^{1}H NMR. Figure 5.7, illustrates the UPLC-MS chromatograms obtained for a control human AF using two complementary chromatographic columns, HSS and HILIC, in both ESI^{+} and ESI^{-} modes. Some assignments based on standards retention time are indicated.

The UPLC-MS data sets registering good reproducibility (as shown in Annex III - UPLC-MS reproducibility evaluation), i.e. Log transformed HSS and HILIC ESI^{+} and ESI^{-} data sets, as well as well as the HSS/ESI^{+} (Pareto scaled) and HILIC/ESI^{+} (UV scaled), were analysed by MVA.

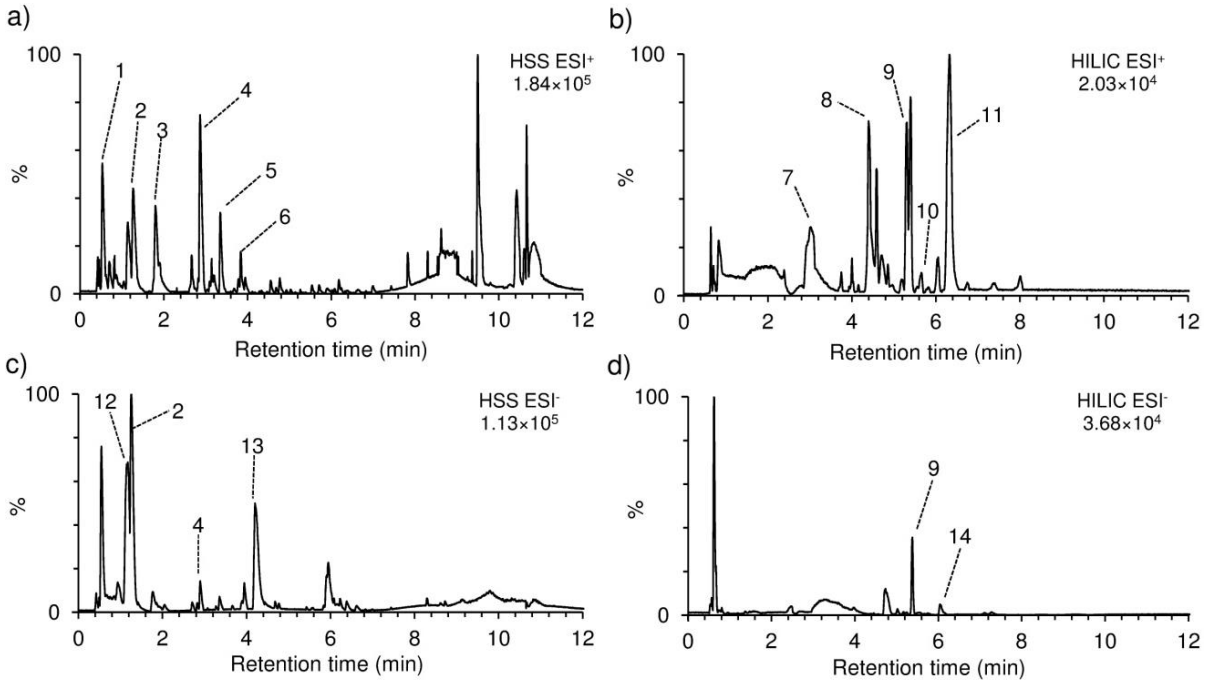


Figure 5.8 - Base peak ion (BPI) UPLC-MS chromatograms of control human AF sample obtained using two different columns and acquired in both ionization modes: a) HSS/ESI$^+$; b) HILIC/ESI$^+$; c) HSS/ESI$^-$; d) HILIC/ESI$^-$. Maximum peak intensities are indicated in each chromatogram. Assignments: 1, glucose; 2, uric acid; 3, tyrosine; 4, phenylalanine; 5, tryptophan; 6, caffeine; 7, creatinine; 8, choline; 9, betaine; 10, creatine; 11, carnitine; 12, citrate; 13, indoxyl sulfate; 14, glutamine/glutamate.

Similarly to the results obtained with ^{1}H NMR data (page 123), no distinction was observed for different malformation types, with basis on PCA the UPLC-MS profiles of AF (not shown). PCA models were then obtained for control *vs* FM (all FM types as one single group).

Although very weak tendencies were visible in PCA scores (Figure 5.8, left), separation was only obtained through PLS-DA, with most of the separations being observed in the LV1 component (Figure 5.8, centre). OPLS-DA was also performed in the data sets, resulting in increased separation in the LV1 component (Figure 5.8, right). The HSS/ESI$^+$ Pareto scaled and the HILIC/ESI$^+$ UV scaled data sets originated similar PLS-DA and OPLS-DA results to those presented in Figure 5.8a and c, respectively (not shown). The PLS-DA and OPLS-DA model parameters obtained for all data set/scaling combinations used are summarized in Table 5.3. The prediction power, which was expressed as Q^2 values, was between 0.4-0.6, therefore indicated robustness of the models obtained (Table 5.3). The PLS-DA models were subjected to MCCV, which resulted in some overlap of Q^2 distributions of true and permuted classes with, however a clear separation tendency (Figure 5.9a-d).

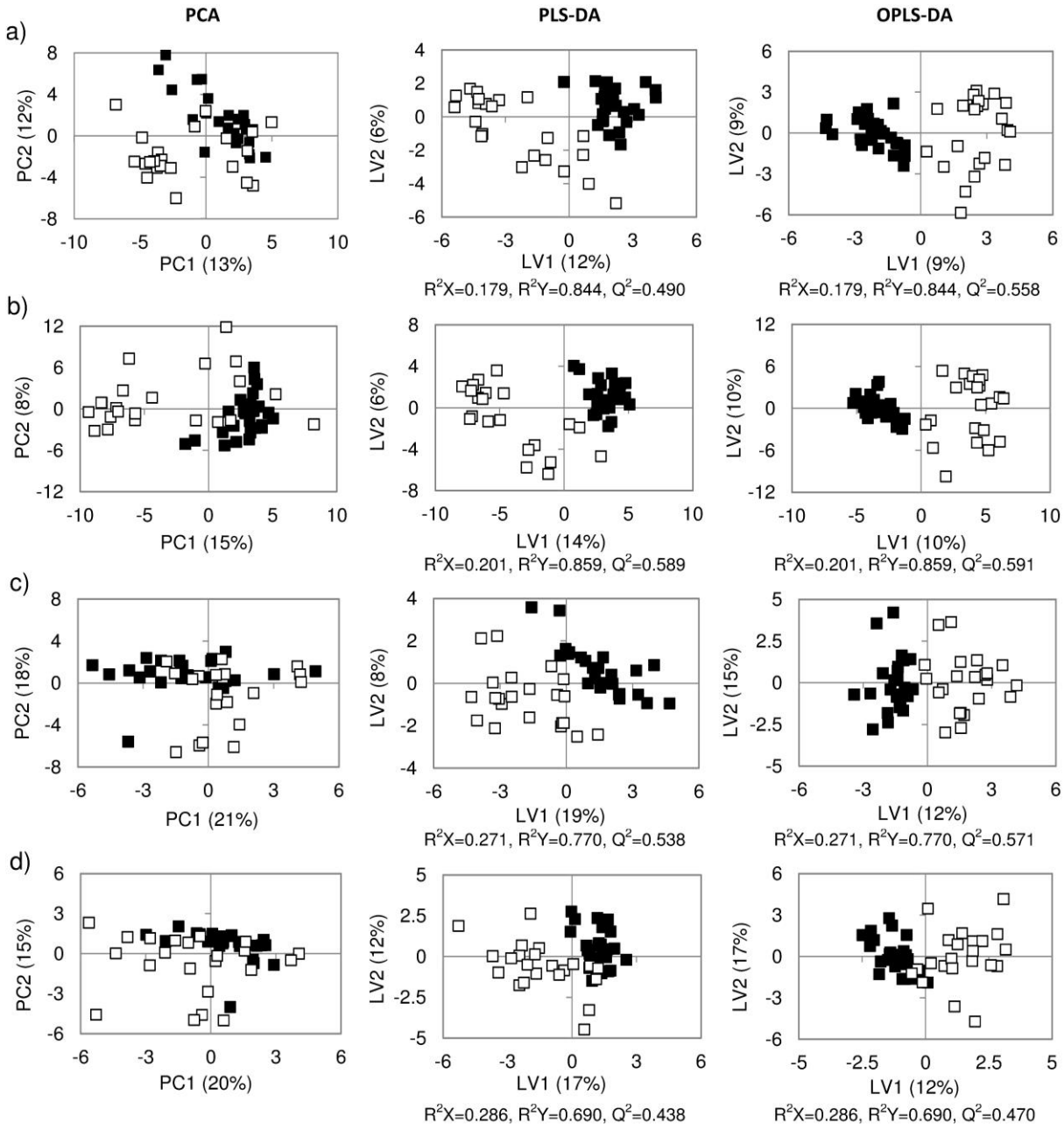


Figure 5.7 - Scores scatter plots from PCA (left), PLS-DA (center) and OPLS-DA (right) models of control (■, n=26) *vs* FM (□, n=22) of Log transformed UPLC-MS data sets: a) HSS/ESI$^+$, b) HSS/ESI$^-$, c) HILIC/ESI$^+$, d) HILIC/ESI$^-$.

Table 5.3 - Parameters of UPLC-MS PLS-DA and OPLS-DA models of control *vs* FM.

UPLC column	detection mode	Scaling	*PLS-DA model parameters*				*OPLS-DA model parameters*			
			R²X	R²Y	Q²	LV	R²X	R²Y	Q²	LV
HSS	ESI+	Pareto	0.187	0.815	0.471	2	0.187	0.815	0.558	1+1
HSS	ESI+	Log	0.179	0.844	0.490	2	0.179	0.844	0.558	1+1
HSS	ESI-	Log	0.201	0.859	0.589	2	0.201	0.859	0.591	1+1
HILIC	ESI+	UV	0.243	0.819	0.623	2	0.243	0.819	0.535	1+1
HILIC	ESI+	Log	0.271	0.770	0.538	2	0.271	0.770	0.571	1+1
HILIC	ESI-	Log	0.286	0.690	0.438	2	0.286	0.690	0.470	1+1

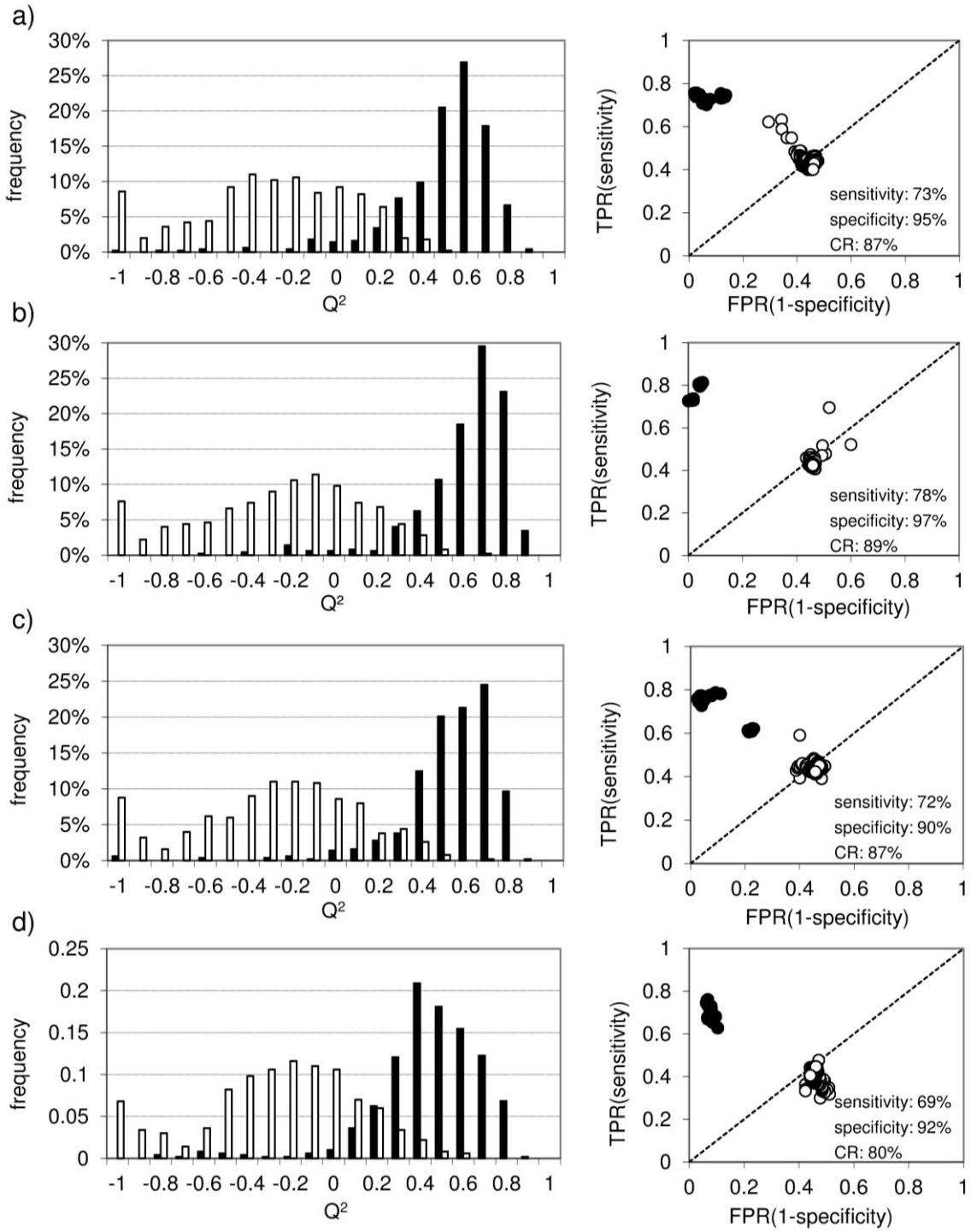


Figure 5.9 - MCCV results for control *vs* FM PLS-DA models of log transformed UPLC-MS data sets: a) HSS/ESI$^+$, b) HSS/ESI$^-$, c) HILIC/ESI$^+$, d) HILIC/ESI$^-$. Q^2 distributions (left) of ■true classes, □ permuted classes and ROC plots (right) of true (●) and permuted (○) classes models are shown. Average sensitivity, specificity and classification rate (CR) values for true classes are shown in the ROC plots.

True class models of HSS data sets registered high specificity and sensitivity of about 80%-98% and 70%-75% respectively in ROC plots (Figure 5.9a,b). The HILIC/ESI$^-$ data set models registered specificity and sensitivity of about 90% and 60%-80%, respectively, for true classes in ROC plot

(Figure 5.9d), whereas the HILIC/ESI$^+$ registered more dispersion in the true class models, with some models registering specificity of 80% and sensitivity of 60% with the remaining registering sensitivity close to 80% and specificity values above 80% (Figure 5.9c). Similar MCCV results were obtained for HSS/ESI$^+$ Pareto and HILIC/ESI$^+$ UV data sets (not shown). This means these models have good predictive performances. It is clear, however, that the most robust models were obtained with HSS data sets.

Comparing the MCCV of UPLC-MS with those obtained previously by ^{1}H NMR (Figure 5.4, page 125), it is interesting to note an improvement in sensitivity of the models based on the HSS UPLC-MS, possibly due to the detection of higher number of more stable signals. On the other hand, the balanced number of controls (n=26) and FM (n=22) samples used in UPLC-MS PLS-DA models possibly resulted in more stable MCCV parameters obtained through the validation process, therefore contributing to the higher overall stability in terms of sensitivity.

The OPLS-DA models were considered for the purpose of identification of MS signals or features (consisting of retention time - m/z pairs) with impact in separation between control and FM samples. This was justified because OPLS-DA models obtained registered Q^2 values equivalent to those obtained for the corresponding PLS-DA models (Table 5.3), and have the advantage of providing a more straightforward loadings interpretation since separation is condensed in LV1 (Figure 5.8, right). The loadings from the OPLS-DA models were interpreted with the aid of S-plots (Figure 5.10a), as explained previously in chapter 3 (section 3.5.3, page 81).

In the S-plot, the features with correlation values |p(corr)[1]|>0.5 and covariance values p[1] falling outside the diagonal (|p[1]|>0.05) were selected and further evaluated through covariance ratio plots jack-knife plot (Figure 5.10b). The latter enables the selection of only the most stable features, i.e. those registering lower standard error of covariance across the cross-validation rounds (p[1]cvSE). The threshold for feature selection (dashed line in Figure 5.10b) was placed above the ratio value 4, which is the ratio where p[1]cvSE is 25% the value of the covariance value.

The selected features were then further evaluated by univariate analysis. Considering all the data sets analysed, a total of 83 features were found as varying significantly (p<0.05) between control and FM groups, 21 of which were assigned and listed in Table 5.4.

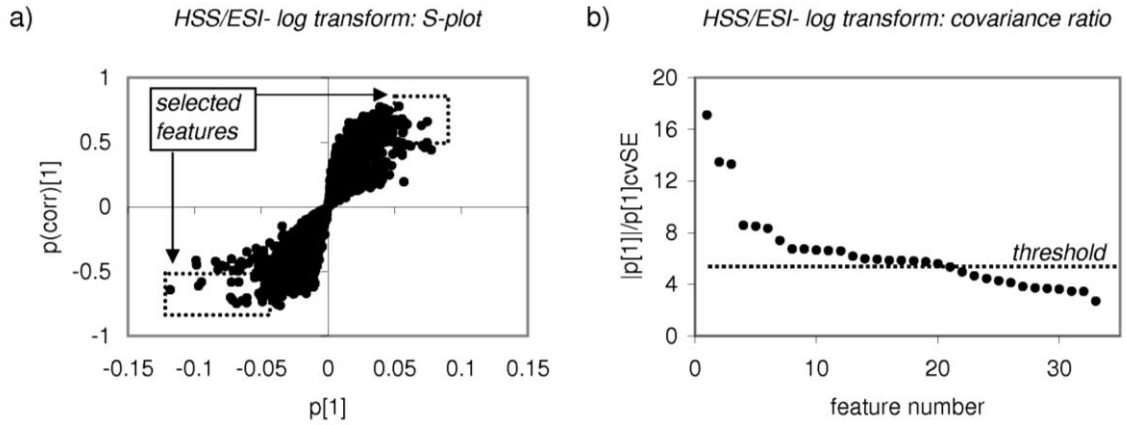


Figure 5.10 - Selection of the MS features with higher impact in the OPLS-DA model obtained for control *vs* FM AF HSS/ESI$^-$ log transformation. a) S-plot and b) covariance ratio plot of the separating latent variable. Dashed boxes in a) represent the groups of selected features (for $|p(\text{corr})[1]|>0.5$ and $|p[1]|>0.05$) and the dashed line in b) indicates the threshold (>4) of the covariance ratio above which the features where selected.

Table 5.4 - List of assigned significant UPLC-MS features obtained for 2nd trimester AF of the FM compared to the control group. Feature intensities are indicated as: very high (>10^5), high (>10^4, <10^5), low (>10^3, <10^4) and very low (<10^3). [a] metabolite fragment; [b] metabolite adduct; [c] isotope of the molecular peak or adduct. *, newly identified metabolite variations; †, tentatively assigned MS features.

metabolite	UPLC column	detection mode	m/z	RT (min.)	feature intensity	PLS-DA Model	direction/ magnitude of variation		*p*-value
leucine/isoleucine	HSS	ESI^+	132.09	1.8	high	Pareto	↓	1.4	1.80×10^{-05}
			86.08[a]	1.8	low	Pareto			
carnitine* †	HSS	ESI^+	162.00	0.4	low	Log	↑	2.5	1.80×10^{-05}
polyol* †	HSS	ESI^+	205.07[b]	0.5	high	Pareto	↓	1.4	8.78×10^{-04}
pyroglutamate*	HSS	ESI^+	130.04	1.4	high	Pareto	↑	1.6	2.24×10^{-06}
valine	HSS	ESI^+	118.07	0.9	high	Pareto	↓	1.4	7.49×10^{-07}
estriol-3-glucuronide/ estriol-16-glucuronide*†	HSS	ESI^-	463.20	4.9	high	Log	↑	3.8	3.90×10^{-05}
estriol-3-sulfate-16-glucuronide* †	HSS	ESI^-	543.16	4.3	high	Log	↑	3.1	1.47×10^{-04}
leucine/isoleucine	HSS	ESI^-	491.19[b]	1.9	low	Log	↓	2.2	2.24×10^{-06}
	HSS	ESI^-	462.24[b]	1.9	low	Log			
	HSS	ESI^-	576.27[b]	1.9	low	Log			
	HSS	ESI^-	547.32[b]	1.9	low	Log			
glutamine	HILIC	ESI^+	130.05[a]	5.7	low	UV	↑	2.7	2.81×10^{-04}
	HILIC	ESI^+	84.04	5.7	low	Log			
	HILIC	ESI^+	527.14[b]	5.7	very low	Log			
isoleucine	HILIC	ESI^+	312.04[b]	4.9	low	UV	↓	1.7	4.06×10^{-06}
lysine	HILIC	ESI^+	84.08[a]	7.2	low	UV	↓	1.4	1.74×10^{-03}
valine	HILIC	ESI^+	72.08[a]	5.2	low	UV	↓	1.5	3.58×10^{-07}
	HILIC	ESI^+	298.03[b]	5.2	very low	UV			
	HILIC	ESI^+	369.10[b]	5.2	very low	UV			
glutamate	HILIC	ESI^-	146.04	6.1	low	Log	↓	1.5	7.01×10^{-03}

The features corresponding to glutamate, leucine/isoleucine, lysine and valine were found decreasing in FM group together with an increase in glutamine, which is in agreement with the variations previously observed by ^{1}H NMR. New metabolite variations were also observed by UPLC-MS (Table 5.4). Indeed, the feature 205.07 m/z found in HSS/ESI^{+} data sets was found decreasing in FM group. This feature co-eluted with glucose at about 0.5min and was tentatively assigned to a sodium adduct of a polyol, possibly sorbitol (Figure 5.11a).

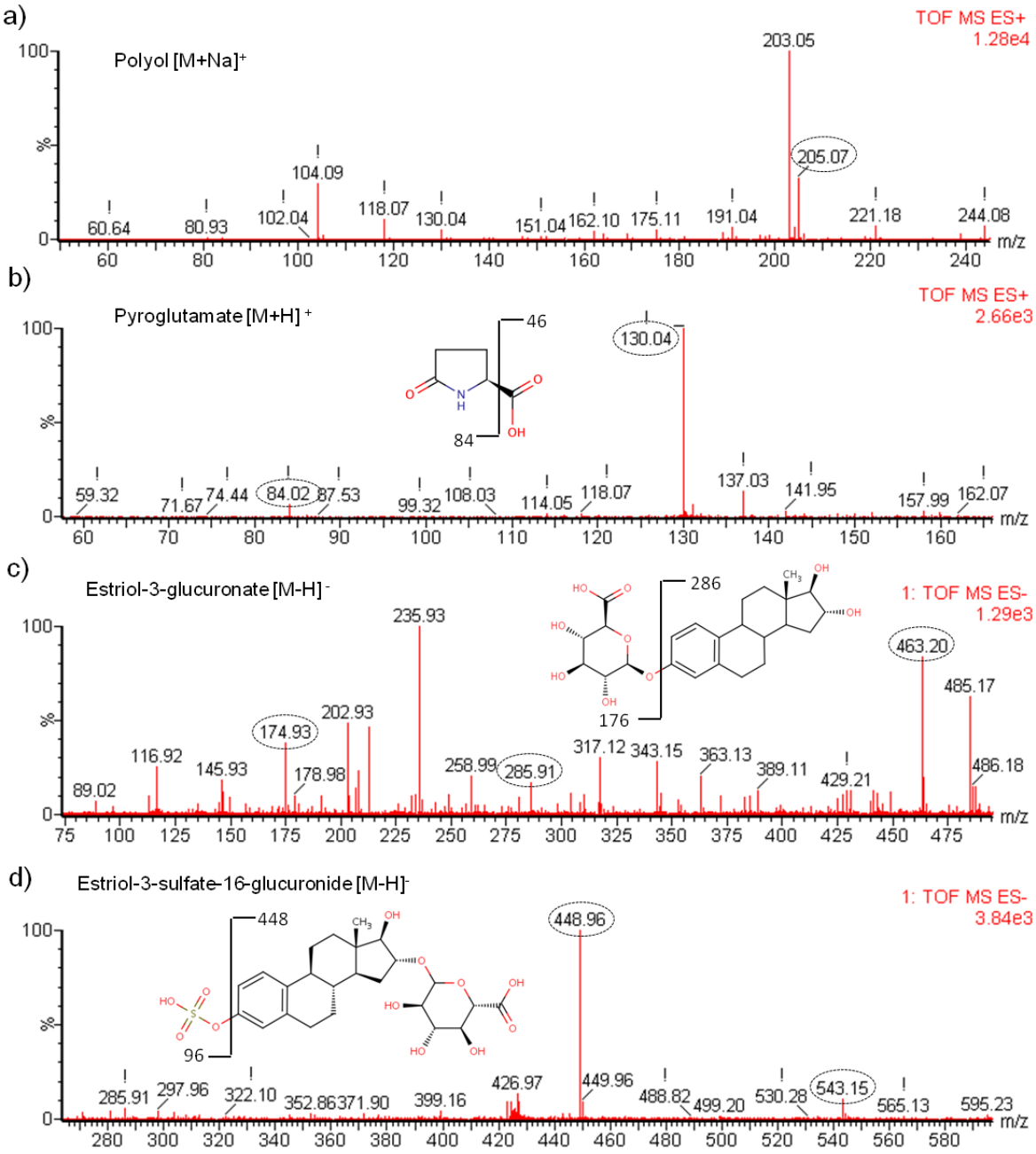


Figure 5.11 - Mass spectra of newly and tentatively assigned features which had significant variations in FM samples: a) polyol, possibly sorbitol, b) pyroglutamate, c) estriol-3-glucuronide, d) estriol-3-sulfate-16-glucuronide. Peaks corresponding to each feature and to respective fragments are indicated, together with the compound structure and possible fragments.

Additionally, pyroglutamate (Figure 5.11b), carnitine and the tentatively assigned to glucuronidated estriol (estriol-3-glucuronate, Figure 5.11c or estriol-16-glucuronate) and estriol-3-sulfate-16-glucuronate (Figure 5.11d) were found increased in FM group. Although pyroglutamate, carnitine and polyols had been previously detected in human AF by high field NMR and LC-NMR (chapter 4, Table 4.2 - page 101 and Table 4.5 - page 113, respectively), their variation had not been detected in ^{1}H NMR. This was possibly due to strong overlap of their resonances with those of the more abundant one glutamine and glutamate (2.0-2.6 ppm) and glucose (3.2-3.5 ppm). Estriol conjugates, on the other hand are expected only in nM levels in human AF (Lind, 1981), therefore, it is not surprising not to detect estriol signal variations by ^{1}H NMR. A list of 62 features found varying significantly (p<0.05) between control and FM group and for which no assignment was obtained at this stage is also presented in annex (see Annex VI - Lists of UPLC-MS unassigned features detected in AF). However, it is important to mention that further assignment work could reveal other important metabolite variations and should be considered in future analysis.

5.1.4. UPLC-MS/^{1}H NMR SHY analysis for FM cases

In order to confirm the obtained MS assignments and find further metabolic relationships, the intensities of each significant MS feature of the FM group were correlated with their corresponding ^{1}H NMR CPMG spectra using statistical heterospectroscopy (SHY) analysis (chapter 3, section 3.7, page 83). The results were analysed in the form of 1D STOCSY (Figure 5.12).

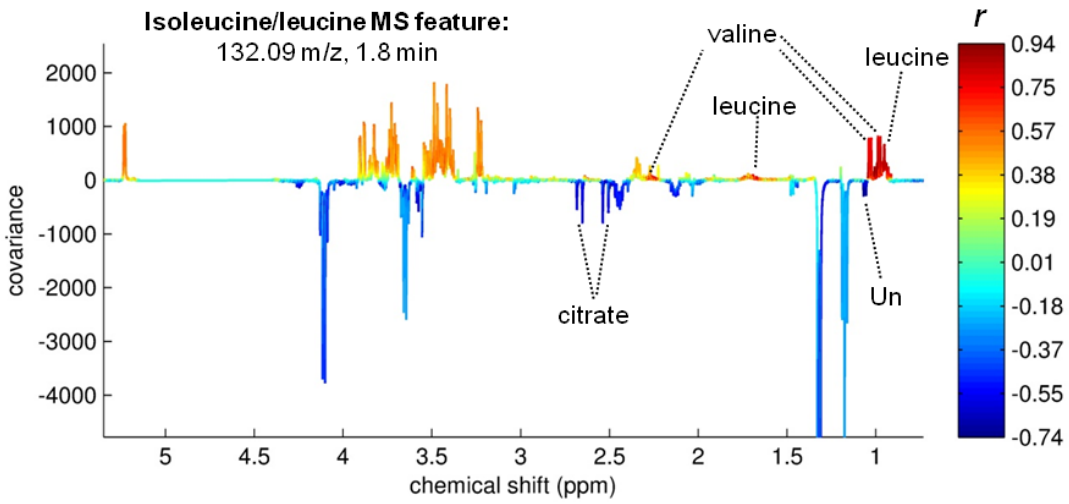


Figure 5.12 - Example of UPLC-MS and ^{1}H NMR SHY analysis for leucine/isoleucine HSS/ESI^{+} MS feature (132.09 m/z, 1.8min.). The assigned resonances correspond only to those registering correlation values $|r|\geq 0.70$ and (p<0.001).

In the 1D STOCSY plot shown in Figure 5.12, the ^{1}H NMR signals with positive covariance and correlation show variation in the same direction as the MS feature whereas those with negative covariance and correlation show opposite variation with that of the MS feature. Only the positive and negative correlations values with $|r| \geq 0.70$ and $p < 0.001$ significance were considered based on previous reports (Crockford *et al.*, 2006; Maher *et al.*, 2011). The results of the SHY analysis of ^{1}H NMR data and the relevant MS features are presented in Table 5.5.

Table 5.5 - List of UPLC-MS/^{1}H NMR correlations obtained from SHY analysis of significant AF MS features (shown in Table 5.4) and AF ^{1}H NMR CPMG spectra of FM. Only maximum or minimum correlations values $|r| \geq 0.70$ and ($p < 0.001$) were considered. Assigned spin systems are indicated in bold. [a] metabolite fragment; [b] metabolite adduct; [c] isotope of the molecular peak or adduct; s, singlet; d, doublet; t, triplet; m, multiplet; dd, doublet of doublets; * variation relative to control group; Un, still unassigned resonances.

metabolite	m/z	direction of variation*	RT (min)	^{1}H NMR resonances correlated positively to MS feature / ppm	^{1}H NMR resonances correlated negatively to MS feature / ppm
leucine/ isoleucine	132.09	↓	1.8	**leucine:** 0.95(t); **valine:** 0.98(d), 1.03(d), 2.25(m) **isoleucine:** 0.93(t), 1.00(d), 1.70(m)	**citrate:** 2.51; Un: 1.06(d)
leucine/ isoleucine [b]	491.19	↓	1.9	**isoleucine:** 0.93(t), 1.01(d); **leucine:** 0.95(t), 1.70(m) **valine:** 0.98(d), 1.03(d), 2.25(m), 3.60(d) **glucose:** 3.23(dd), 3.40(m), 3.53(dd), 3.72(m), 3.82(m), 3.89(dd)	Un: 1.06(d)
isoleucine [b]	312.04	↓	4.9	**isoleucine:** 0.93(t), 1.01(d); **leucine:** 0.95(t), 1.70(m) **valine:** 0.98(d), 1.03(d), 2.25(m), 3.60(d) **tyrosine:** 6.89(d), 7.19(d)	-
glutamine [a]	130.05	↓	5.7	**glutamine:** 2.13(m), 2.45(m)	**glutamate:** 2.35(m)
glutamate	146.04	↑	6.1	**glutamate:** 2.06(m), 2.35(m)	-
valine	118.07	↓	0.9	**leucine:** 0.95(t); **tyrosine:** 6.89(d), 7.19(d) **valine:** 0.98(d), 1.03(d), 2.25(m), 3.60(d) **isoleucine:** 0.93(t), 1.00(d), 1.70(m) **glucose:** 3.23(dd), 3.40(m), 3.45(m), 3.48(m), 3.53(dd), 3.72(m), 3.82(m), 3.88(dd), 5.23(d)	Un: 6.92(s), 7.67(s), 4.01, 1.06(d)
valine [a]	72.08	↓	5.2	**isoleucine:** 0.93(t), 1.01(d); **leucine:** 0.95(t), 1.70(m) **valine:** 0.98(d), 1.03(d), 2.25(m), 3.60(d)	-
pyroglutamate	130.04	↑	1.4	**threonine:** 3.58(d); Un: 0.88(t) **tyrosine:** 6.89(d), 7.19(d)	**phenylalanine:** 7.32(m), 7.40(m); **tyrosine:** 6.89(d), 7.19(d);
lysine [a]	84.08	↓	7.2	**phenylalanine:** 7.32(m)	-

Positive autocorrelations were found for leucine/isoleucine, valine, glutamine and glutamate, which simply confirm the MS assignments of these metabolites. In addition, positive correlations were found, in several instances, between leucine/isoleucine, valine, tyrosine and glucose, indicating the concomitant decreases of these metabolites (Table 5.5). The leucine/isoleucine MS feature (decreased in FM) also correlated negatively with citrate, which indicates that citrate accumulates at a similar rate to that of glucose and the above amino acid consumption. Other UPLC-MS/^{1}H NMR correlations found for the FM group comprise the correlation of the pyroglutamate MS feature (increased in FM) with resonances arising from threonine (positive correlation), phenylalanine and tyrosine (negative correlations). Furthermore, the lysine MS feature (decreased in FM) correlated positively with resonances of phenylalanine, and the glutamine MS feature (increased in FM) correlated negatively with glutamate resonances. It is also noted that several unassigned NMR resonances were seen to correlate to several of the MS features (Table 5.5), which should be considered in future assignment work.

5.1.5. MIR profiling for comparison of control and FM groups

Figure 5.13 shows a typical ATR-MIR spectrum of a dry human AF sample. Most of the signals occur in the fingerprint region (1800-600 cm^{-1}) for which proteins, lipids, organic acids (mainly lactate), amino acids, and glucose contribute importantly. Also some signals related to stretching vibrations (ν) of simple C-H bonds appear in the region of 3000-2800 cm^{-1}, together with O-H and N-H stretching vibrations (ν), more visible around 3600-3100 cm^{-1} (Figure 5.13).

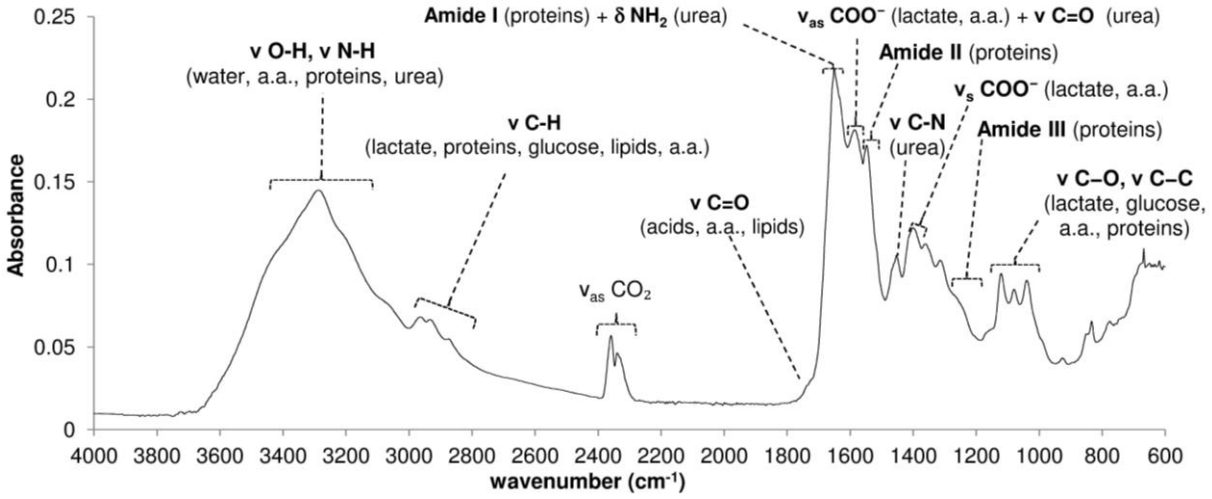


Figure 5.13 - ATR - MIR spectra of a dry control AF sample. Main assignments are shown. a.a., amino acids. Spectral assignments based on literature data (Liu *et al.*, 1998a; Liu *et al.*, 1998b; Liu and Mantsch, 1999; Liu *et al.*, 2000; Liu *et al.*, 2002).

The MIR spectra obtained for control and FM AF samples were used to produce PLS-DA models. Three spectral replicates were considered for each sample. PLS-DA models were obtained for spectral regions for which spectral reproducibility was obtained (Annex IV - MIR reproducibility evaluation): 3000-2800 cm^{-1}, 1800-1200 cm^{-1}, 1200-900 cm^{-1}, and for 2nd derivative transformed data of the 1200-900 cm^{-1} region. The corresponding scores scatter plots are shown in Figure 5.14.

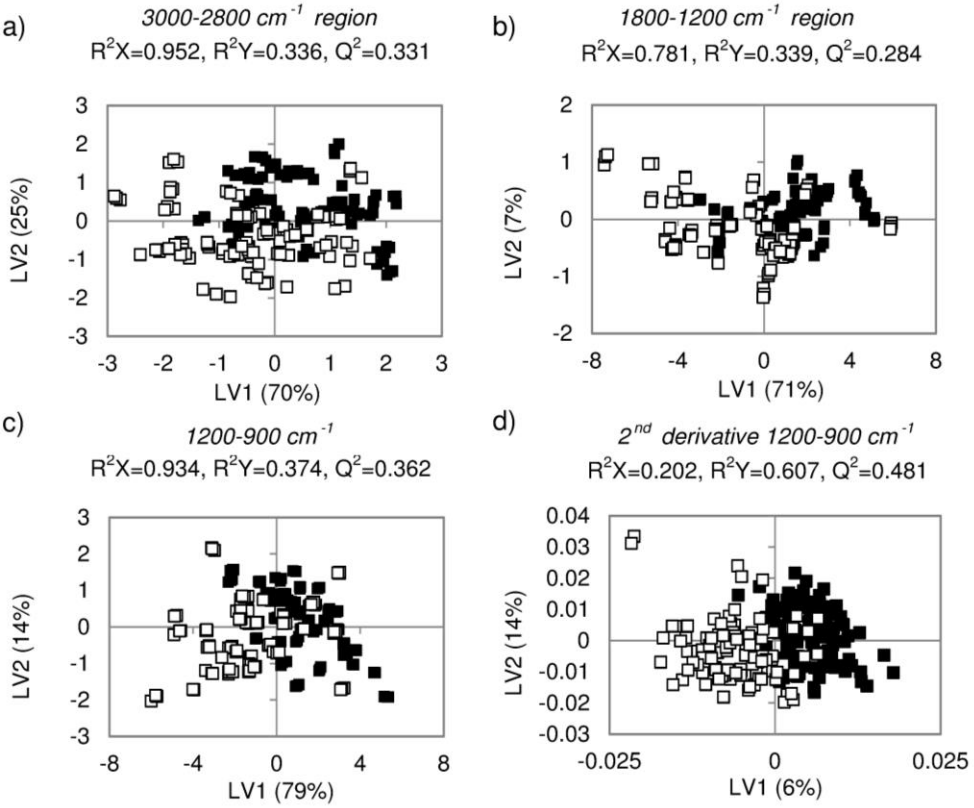


Figure 5.14 - PLS-DA scores scatter plots of obtained for MIR spectra of AF of control (■, n=40) *vs.* FM cases (□, n=34) with 3 spectral replicates were used per sample for the spectral regions: a) 3000-2800 cm^{-1}, b) 1800-1200 cm^{-1}, c) 1200-900 cm^{-1}, d) 2nd derivative transformed data for the 1200-900 cm^{-1} region.

The results show a significant overlap of control and FM spectra in the models obtained for the region 3000-2800 cm^{-1} and 1800-1200 cm^{-1} (Figure 5.14a and b, respectively). However, a tendency for separation of control and FM spectra was noted when the spectral region of 1200-900 cm^{-1} was used (Figure 5.14c). This tendency is improved when the 2nd derivative of the same data is considered (Figure 5.14d), probably resulting from resolution increase which unveils small spectral differences between control and FM spectra. One outlier was noted in the scores scatter plot from the latter model which corresponds to one of three poly-malformed fetuses cases (the other 2 are overlapped with the remaining FM samples, Figure 5.14d). No other clinical info was found to justify the separation of this sample from the remaining samples of the FM group.

The validation of the PLS-DA models of 1200-900 cm^{-1} region and 2nd derivative transformed data for the 1200-900 cm^{-1} region was assessed with MCCV (Figure 5.15). The PLS-DA model obtained using 1200-900 cm^{-1} region, revealed mostly separated true and permuted Q^2 distributions, 80% classification rate, 73% of sensitivity and 85% sensitivity (Figure 5.15a), whereas the PLS-DA model corresponding to the 2nd derivative transformed data of the 1200-900 cm^{-1} region resulted in mostly separated of true and permuted Q^2 distributions, a classification rate of 80%, 69% sensitivity and 87% specificity (Figure 5.15b). The apparent outlier noted in data points of permuted classes of Figure 5.15b (right) results from the random permutation of classes, a process that naturally has some probability of originating models similar to the true class models, and hence similar values of sensitivity and specificity. However, such an artefact does not affect the interpretation of the MCCV results that provide evidence of the validity of the PLS-DA models of the 1200-900 cm^{-1} region and 2nd derivative transformed data of the 1200-900 cm^{-1} region.

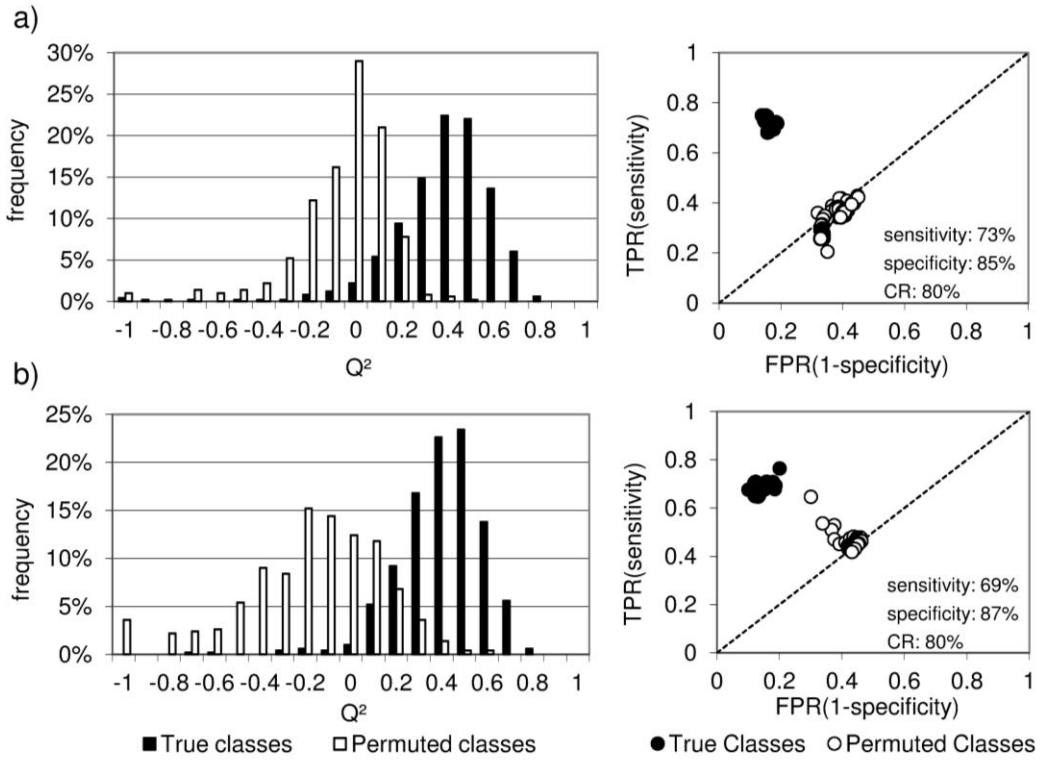


Figure 5.15 - MCCV results for the PLS-DA models obtained for: a) 1200-900 cm^{-1} region and b) 2nd derivative of the 1200-900 cm^{-1} region of the MIR spectra of control *vs* FM groups. Left: Plots of the Q^2 values distribution for true and permuted classes; Right: ROC curves for true (●) and permuted (○) classes. Average sensitivity, specificity and classification rate (CR) values for true classes are shown in the ROC plots.

The separation observed between control and FM samples in the PLS-DA scores of the 1200-900 cm^{-1} region and 2nd derivative of the 1200-900 cm^{-1} region (Figures 5.14c,d) occurred mainly in LV1. The respective LV1 loadings for both models are shown in Figure 5.16).

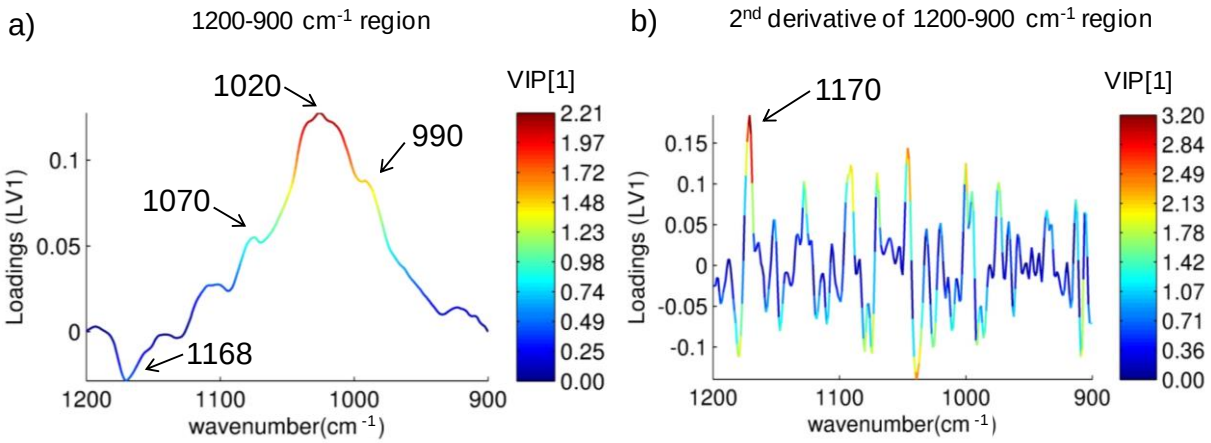


Figure 5.16 - PLS-DA LV1 loadings from MIR models of control *vs* FM: a) 1200-900 cm^{-1} spectral region; b) 2nd derivative of 1200-900 cm^{-1} region. VIP values for LV1 are represented as colour.

Inspection of LV1 loadings from the PLS-DA 1200-900 cm^{-1} region model (Figure 5.16a) revealed decreases in bands at 1070, 1020 and 990 cm^{-1}, and a decrease in the band 1168 cm^{-1} in of the FM group. However, the VIP values indicate that the band 1020 cm^{-1} has the highest impact in the separation between control and FM groups.

The LV1 loadings of the 2nd derivative 1200-900 cm^{-1} region PLS-DA model suggested the involvement of the band 1170 cm^{-1} (Figure 5.16b). To aid interpretation, a comparison of the LV1 loadings (5.17a) with a typical record of 2nd derivative of the 1200-900 cm^{-1} spectral region (Figure 5.17b) and also with the respective untransformed spectral region (Figure 5.17c) was performed. It is important to notice that the 2nd derivative transformation originates signal inversion when the transformed (Figure 5.17b) and untransformed (Figure 5.17c) spectral regions are compared. Therefore, the variations observed in the loadings should be interpreted as pointing in the opposite direction to that observed in the scores scatter plots of (Figure 5.14d). In this way, the band 1170 cm^{-1} is found increased in FM group. Moreover, comparison of 2nd derivative of 1200-900 cm^{-1} spectral region (Figure 5.17b) and the untransformed 1200-900 cm^{-1} region (Figure 5.17c) clearly shows a very weak 1170 cm^{-1} band in the 1200-900 cm^{-1} region (Figure 5.17c), which becomes more visible after 2nd derivative transformation (Figure 5.17b), thus demonstrating the usefulness of such procedure.

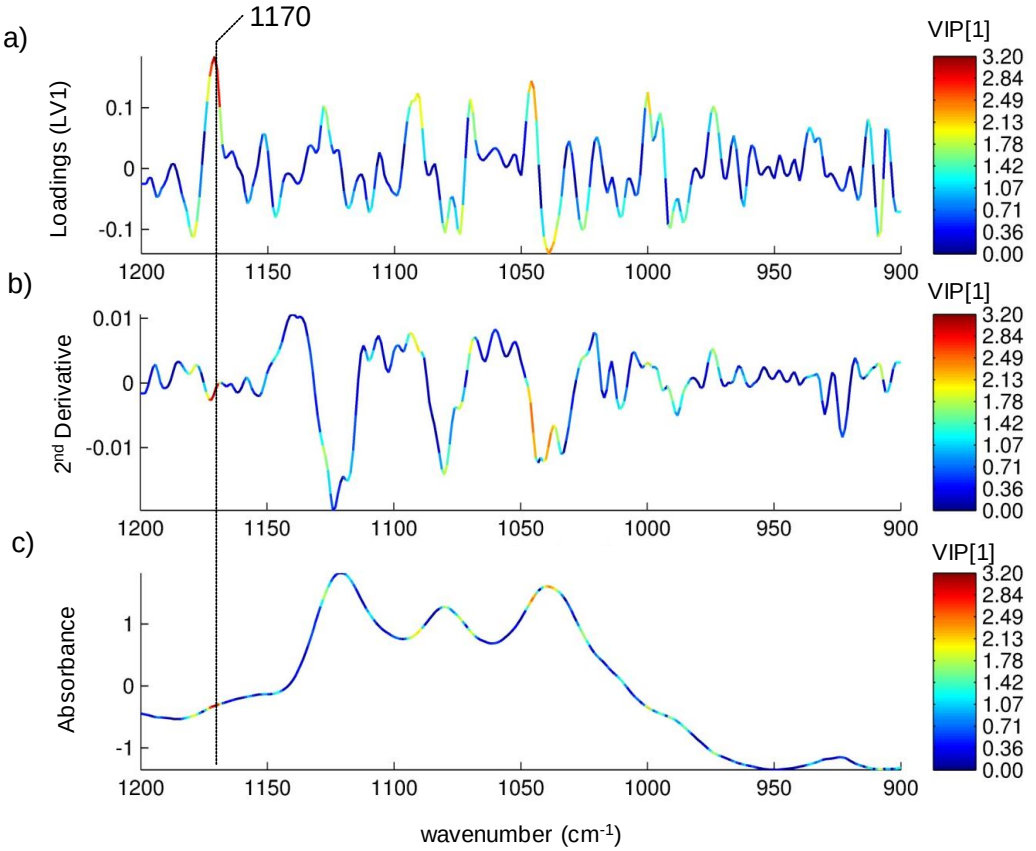


Figure 5.17 - PLS-DA loadings analysis of the 2nd derivative of 1200-900 cm^{-1} region: a) LV1 loadings; b) 2nd derivative of the 1200-900 cm^{-1} spectral region from a representative control sample; c) 1200-900 cm^{-1} spectral region from the same control sample. Dashed lines indicate the position of the 1170 cm^{-1}.

5.1.6. MIR STOCSY analysis of FM cases

The identification of intra-correlations between MIR bands may help in the assignment of the relevant bands identified in PLS-DA loadings and also reveal biochemical relationships. An MIR intra-correlation map was obtained for the 1750-900 cm^{-1} region of the spectra from FM samples, using STOCSY (Figure 5.18).

The high degree of correlation in MIR data leads to intra-correlation maps with large correlation areas spanning several correlation values. Therefore, a threshold of $|r|>0.8$ and a significance level of $p<0.001$ were chosen with basis on previous reports using STOCSY (Crockford *et al.*, 2006; Maher *et al.*, 2011) in order to eliminate spurious correlations. The bands from the loadings of the FM PLS-DA models were searched along the vertical dimension of the STOCSY map and the off-diagonal correlation maxima were identified (Figure 5.18). At the correlation and significance thresholds used, all the correlations obtained were positive as shown in Figure 5.18 and Table 5.6.

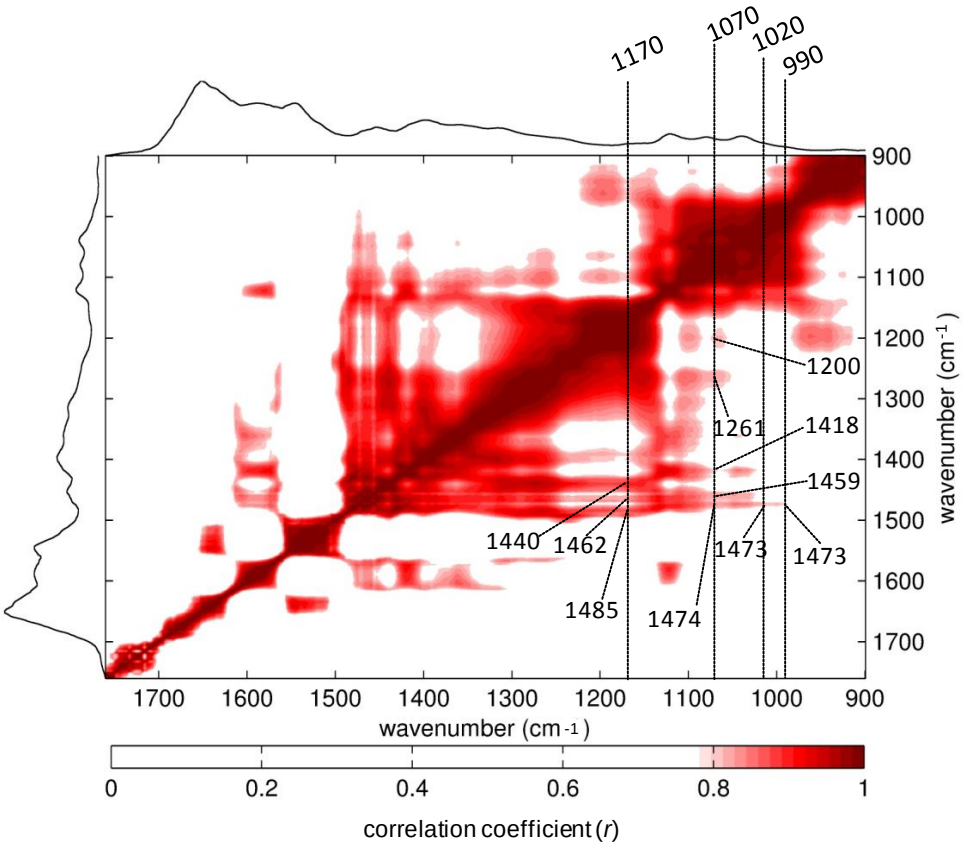


Figure 5.18 - MIR (1750-900 cm^{-1}) intra-correlation map for the FM group (n=34), with correlation threshold $|r|>0.8$ ($p<0.001$). Dashed vertical lines indicate the bands of interest identified through loadings analysis and their correlations across the 1750-900 cm^{-1} spectral region.

Table 5.6 - MIR bands from the PLS-DA loadings and respective correlations from the intra-correlation map shown in Figure 5.16. *observed in 2nd derivative models; ν_s, symmetric stretching; ν_{as}, asymmetric stretching; δ_s, symmetric bending; in brackets: putative assignments (Pretsch *et al.*, 1989).

MIR bands (cm^{-1})	change	MIR/MIR correlations (cm^{-1})
1170 (νC-C/C-O) *	↑	1485, 1462, 1440 (δ_sNH_2)
1070 (νC-C/C-O)	↓	1474, 1459 (δ_sNH_2); 1418 (ν_sCOO^-); 1261 (δ/νOC-OH); 1200 (νC-C/C-O)
1020 (νC-C/C-O)	↓	1473 (δ_sNH_2)
990 (νC-C/C-O)	↓	1473 (δ_sNH_2)

From the bands decreasing in FM, the one at 1070 cm^{-1} was found to correlate positively with vibrations of functional groups from carboxylate and amine groups (ν_sCOO^-, δ/νOC-OH, δ_sNH_2) and with νC-C and/or νC-O; 1020 and 990 cm^{-1} were found correlated with amine group vibrations (δ_sNH_2) (Table 5.6). These vibrations may be related with amino acid content (decreasing in FM), which is consistent with previously observed changes found by ^{1}H NMR and UPLC-MS (Table 5.2, page 129 and Table 5.4, page 135, respectively).

The 1170 cm^{-1} band (νC-C and/or νC-O) was correlated to bands at 1485, 1462 and 1440 cm^{-1} all possibly corresponding to $\delta_s NH_2$ vibrations (Table 5.6). This suggests a possible relationship to amine group containing molecules, such as amino acids, urea or amines. Moreover, the 1170 cm^{-1} band may also be related to protein content. Indeed, a small band is seen in the albumin spectrum at 1170 cm^{-1} (Figure 5.19a), also registering a relatively weak 2nd derivative signal compared to the region of highest intensity (Figure 5.19b). Other correlations expected with basis on this suggestion, for instance in relation to Amide I and Amide II bands may remain undetectable due to spectral overlap. However, protein signals in the ^{1}H NMR spectra were found to be increased in FM cases, as shown in section 5.1.2 (Table 5.2, page 129) and also in previously reported work (Graça *et al.*, 2009; Graça *et al.*, 2010).

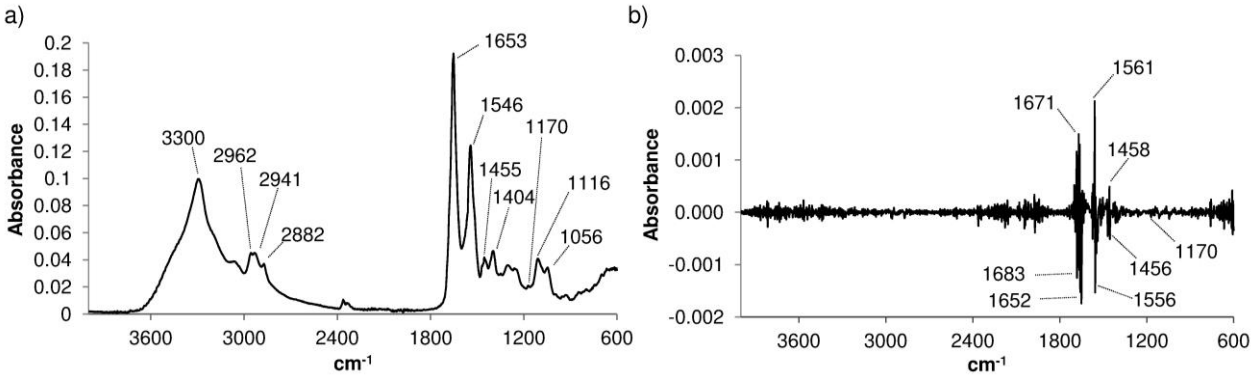


Figure 5.19 - ATR-MIR spectra of dried standard aqueous solution of bovine serum albumin (7 mg/mL, pH 8.60), serving as a model for human AF albumin: a) MIR spectra and b) respective 2nd derivative function.

5.1.7. MIR/^{1}H NMR SHY analysis for FM cases

SHY analysis of MIR and ^{1}H NMR of AF samples from FM group was used to search for further assignments and metabolite relationship information. SHY was performed by correlating the 1200-900 cm^{-1} spectral region of the FM spectra with the corresponding full standard ^{1}H NMR spectra. The SHY map obtained is shown in Figure 5.20. The bands from the LV1 loadings of MIR PLS-DA models were searched along the horizontal dimension of the SHY map, enabling the identification of several positive and negative significant correlations (Table 5.7). Some of the correlations corresponded to metabolites already identified to vary with the occurrence of FM (Table 5.7).

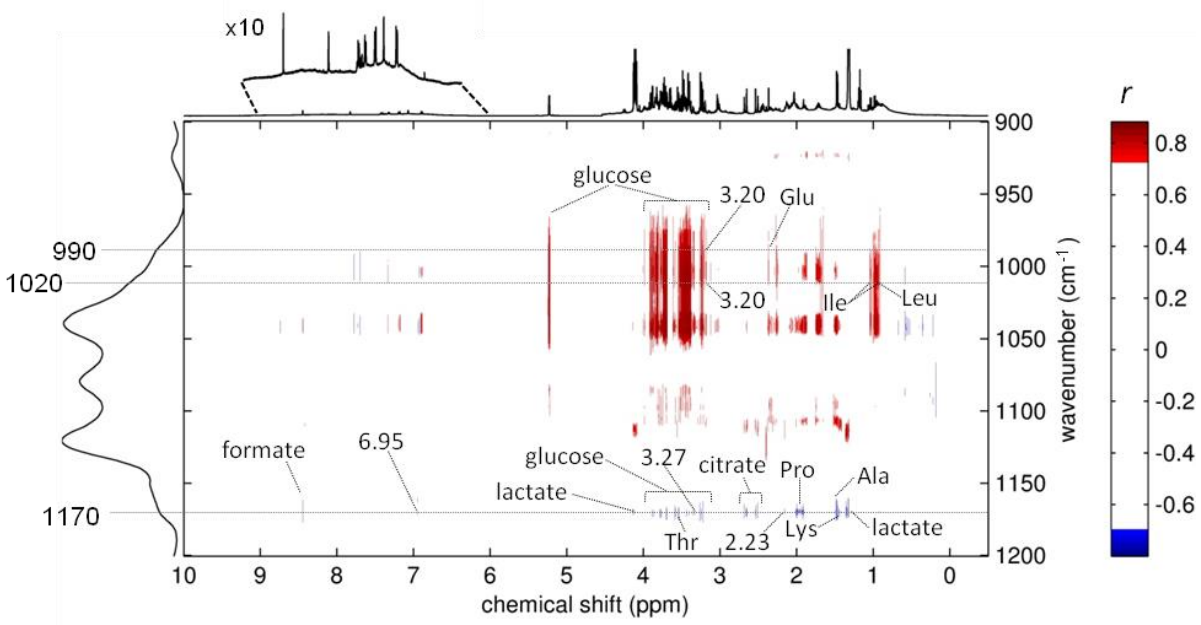


Figure 5.20 - Statistical heterospectroscopy (SHY) maps of MIR (1200-900 cm^{-1}) / ^{1}H-NMR (0-10 ppm) with correlation threshold of $|r|>0.7$ ($p<0.001$) for the FM group (n=34). Dashed horizontal lines indicate the bands of interest identified through loadings analysis and their correlations across the NMR 0-10 ppm region. Unassigned signals are indicated with their chemical shift values.

Table 5.7 - List of SHY heterocorrelations found between MIR bands and standard spectra ^{1}H-NMR signals. Assignments are given in brackets; amino acids are represented in 3 letter code; Un: unassigned signals from ^{1}H-NMR. In bold: metabolites whose variation in FM group was previously reported (Table 5.2, page 129 and Table 5.4, page 135 and (Graça *et al.*, 2009; Graça *et al.*, 2010)).

MIR bands (cm^{-1})	change	Positively correlated ^{1}H-NMR signals	Negatively correlated ^{1}H-NMR signals
1170 (νC-C/C-O)*	↑	NH groups (6.95 ppm)	**citrate, glucose,** formate, **lactate, Ala, Lys,** Pro, Thr, 2 Un (2.23, 3.27 ppm)
1020 (νC-C/C-O)	↓	**glucose, Ile, Leu,** 1 Un (3.20 ppm)	-
990 (νC-C/C-O)	↓	**glucose, Glu,** 1 Un (3.20 ppm)	-

The bands at 1020 and 990 cm^{-1} (decreasing in FM) were positively correlated to specific amino acids and glucose (Table 5.7). The presence of an intense band at around 1025 cm^{-1} in the MIR spectrum of a dried glucose solution (Figure 5.21) seems to indicate the relationship of the band at 1020 cm^{-1} with glucose. The concomitant decrease of glucose and several amino acids (including leucine and isoleucine, correlating positively with glucose) has been observed previously in FM group by ^{1}H NMR and UPLC-MS (sections 5.1.2 and 5.1.4, respectively) seems to confirm the consumption of these metabolites in FM.

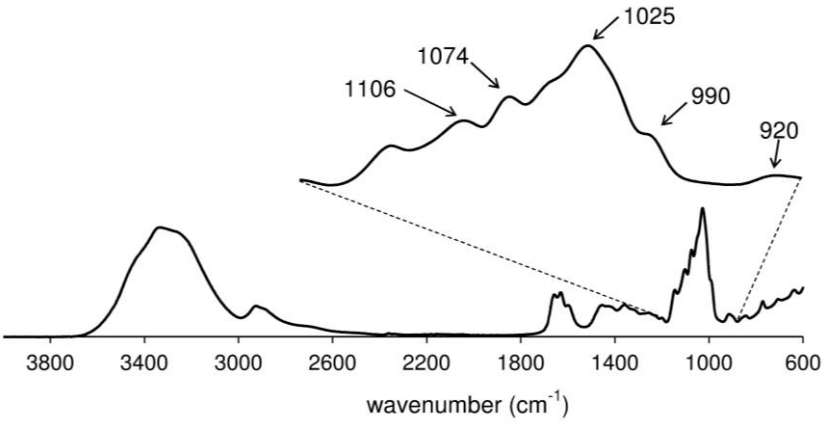


Figure 5.21 - ATR-MIR spectrum of dried standard aqueous solution of glucose (0.1M, pH 8.36).

The 1170 cm^{-1} band noted to increase in the FM group correlated positively and weakly to a signal at 6.95 ppm (NH/aromatic region) and negatively to glucose and some organic and amino acids. The correlation to NH/aromatic region may also be related to protein content, as suggested in the previous section by the MIR intra-correlations. Further NH resonances in NMR may, however, remain undetectable due to spectral and signal broadening in the NMR spectra. The negative correlation of the 1170 cm^{-1} band with alanine, lysine, and proline may possibly indicate a concomitant decrease of the amino acids as response to an increasing demand for protein synthesis in FM cases. The negative correlation with glucose, lactate and citrate, although with no apparently direct metabolic relation, is in agreement with the previously observed increases in proteins and decreases of glucose, citrate and lactate in FM cases (Table 5.2, page 129 and Table 5.4, pages 135).

5.1.8. Proposed interpretation of metabolic alterations occurring in FM cases

The metabolite variations found indicate the possibility of several biochemical alterations common to the majority of FM cases. Indeed, as observed by ^{1}H NMR, higher levels of "free" lactate were observed in FM group when compared with the control group, accompanied by a concomitant decrease in the levels of glucose, mannose, pyruvate, and also a decrease in amino acids (supported by UPLC-MS, UPLC-MS/^{1}H NMR and MIR/^{1}H NMR SHY analysis) including alanine, lysine, isoleucine, leucine, glutamate, histidine, phenylalanine, tyrosine and valine. These variations may suggest derangement in the energy metabolism, which may involve a higher utilization of the glycolytic pathway under anaerobic (hypoxic) conditions (Figure 5.22), where glucose and possibly mannose are being used for as substrates for energy production. In addition,

there seems to be a higher utilization gluconeogenesis and ketogenesis pathways, where the amino acids may be used as substrates for glucose and ketone bodies synthesis, respectively (Figure 5.22) (Murray *et al.*, 2003).

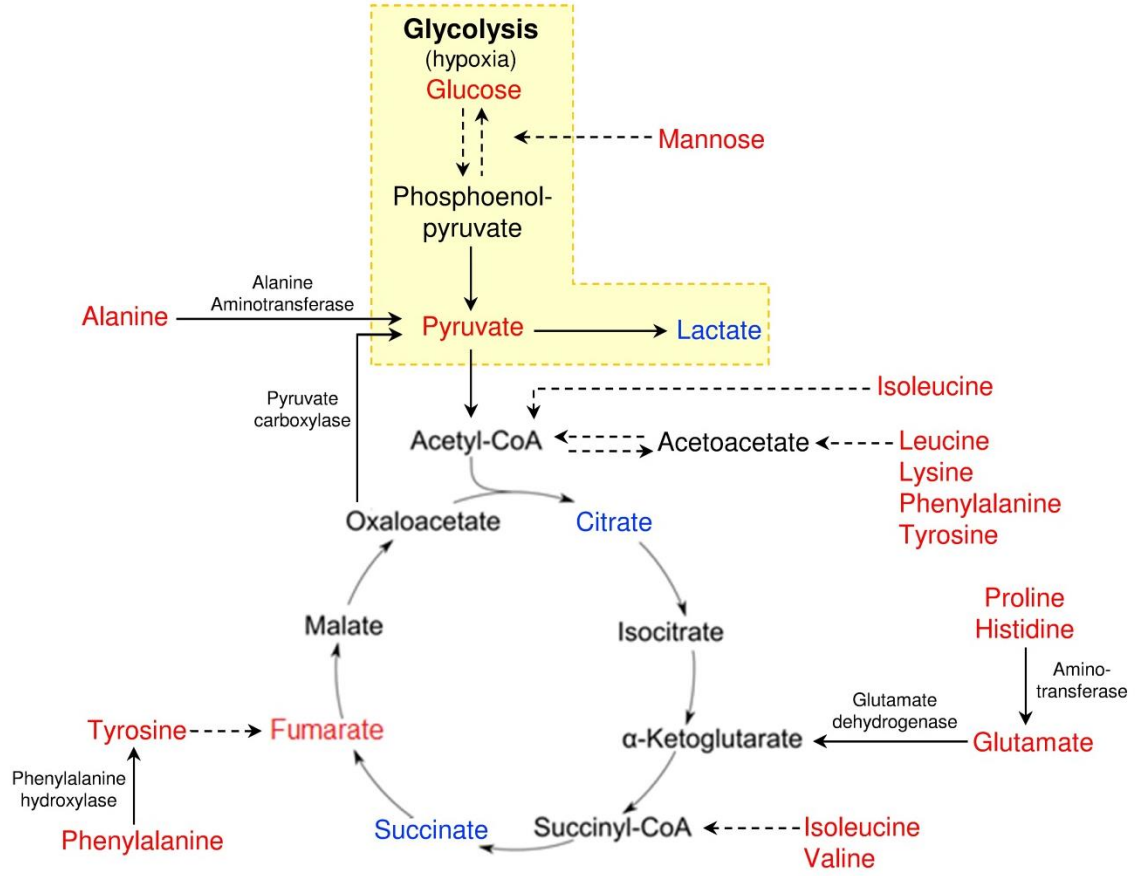


Figure 5.22 - Schematic representation of the glycolytic and gluconeogenic/ketogenic pathways possibly affected in FM cases. The shaded region indicates the glycolytic pathway under hypoxic conditions. Metabolites in blue were increased and those in red decreased in the FM group. Dashed arrows indicate transformations occurring through several reactions.

Under aerobic conditions, metabolic energy production occurs inside the mitochondria through respiratory chain and oxidative phosphorylation, and although the necessary enzymatic machinery is completely formed in the fetal tissues at 9-17 weeks gestation, it is not completely active until birth, when oxygen partial pressure and energetic demands rise (Minai *et al.*, 2008). However, several FM including, hydronephrosis, vertebral abnormalities, anal atresia, cardiac abnormalities, tracheoesophageal fistula/atresia, renal agenesis and dysplasia, limb defects and gastrointestinal malformations, have been associated with deficiencies in the function and regulation of the mitochondrial respiratory chain and oxidative phosphorylation during fetal development, and are also thought to be involved in the mechanisms triggering malformations (von Kleist-Retzow *et al.*, 2003). Succinate and fumarate, which are the substrate and product

respectively of the enzyme complex II of the respiratory chain (Succinate dehydrogenase complex) (Murray *et al.*, 2003), were here found altered (succinate increased and fumarate decreased) in FM cases by ^{1}H NMR spectroscopy. This observation can be interpreted as a possible indicator of the deficient use of the respiratory chain for energy production, therefore shifting the energy production to glycolysis.

Still in connection with the energy metabolism, the decreases in polyol content in FM fetuses (detected by UPLC-MS) suggest perturbations in polyol metabolism. Polyol metabolic pathways are very active in the fetal-placental unit during first trimester in order to maintain ATP concentrations and cellular redox potential at physiological levels, where fetal environment conditions are very poor in oxygen (Jauniaux *et al.*, 2005). In this pathway, glucose has two possible fates apart from glycolysis: one is to generate sorbitol and $NADP^{+}$; and a second one is to generate NADPH and pentose-phosphates (nucleotide precursors), as well as other polyols, such as ribitol or erythritol, and NAD^{+} (Jauniaux *et al.*, 2005). It seems possible that the decreases in glucose contents may have occurred early in pregnancy leading to the decreases in polyols observed. On the other hand, a polyol such as sorbitol could possibly be indirectly channelled into energy production by conversion to fructose which is a substrate in the glycolytic pathway (Jauniaux *et al.*, 2005).

To replenish glucose levels, the amino acid catabolism could be increased in FM in order to feed gluconeogenic pathway, which is already active in the developing liver around 16 weeks (Jauniaux *et al.*, 1999). Indeed, the amino acids alanine, glutamate, valine, proline, histidine, phenylalanine, tyrosine and isoleucine, were found decreased in FM in connection with the decrease in glucose (as observed in MS/NMR and MIR/NMR SHY analysis). Leucine and lysine, both ketogenic amino acids, were also found in decreased in FM cases which indicate that the ketogenesis pathway may also be active in order to generate acetyl-CoA to feed the tricarboxylic acid (TCA) cycle (Figure 5.22). Phenylalanine and tyrosine were also found decreased in FM cases (by ^{1}H NMR and MS/NMR SHY analysis) and, being both gluco- and ketogenic amino acids, may also be used in the ketogenesis pathway. Another argument supporting the increasing need of amino acid by FM fetuses refers to the increase in pyroglutamate in FM fetuses detected by UPLC-MS. Pyroglutamate may be functioning as an extracellular signal to activate the uptake of amino acids at the placenta, as noted before in a mammalian model (Vina *et al.*, 1989). In fact, the pyroglutamate correlated negatively (in MS/NMR SHY analysis) with phenylalanine and tyrosine and positively with threonine suggesting a possible relationship to the uptake of these amino acids.

Interestingly, other glucogenic and ketogenic amino acids glycine, glutamine, serine, and threonine were found increased in FM cases, which suggests their possible involvement in other pathways. For instance, a higher glutamine level (and concomitant lower levels glutamate) is in agreement with the hypothesis advanced in a previous study for spina-bifida cases which proposed that glutamine to glutamate conversion is hindered in fetal kidneys reflecting underdevelopment of these organs (Groenen *et al.*, 2004). Consistently this could also be the reason for the accumulation of proteins and glycoproteins (increases in N-acetyl signals in ^{1}H NMR and in protein signal in MIR) and creatinine.

Glycine and serine increases detected by ^{1}H NMR suggest a disturbance in connection to folic acid pool regulation, with possible reflections in amino acid biosynthesis and choline metabolism. This suggestion is further supported by the observed decreases in choline and formate in FM cases, both of which are also connected the folic acid pool. Other indicators of disturbance of amino acid biosynthesis relate to the marked decreases in leucine and α-oxoisovalerate, metabolites which share a common pathway arising from valine, also decreased in the FM cases. This may indicate a higher demand for protein synthesis in malformed fetuses. The increase in ascorbate may also reflect amino acid and protein synthesis alterations since this compound is involved in collagen, carnitine, and tyrosine synthesis (Murray *et al.*, 2003).

Other metabolic variations did suggest the disturbance of other pathways in FM cases, namely lipid metabolism. For instance, the reduction in the contents of choline and *myo*-inositol in FM cases (noted by ^{1}H NMR), may affect biomembrane and lung surfactants phosphatidylcholine and phosphatidylinositol biosynthesis, which have been previously associated with the neural tube defects (Groenen *et al.*, 2003). The accumulation of citrate in FM cases may indicate, for instance, a possible decrease in lipid biosynthesis pathway, which is a metabolic branching point from TCA cycle through citrate (Murray *et al.*, 2003).

An increase in carnitine was detected in FM cases by UPLC-MS. This may be linked to an increase in fatty acid transport to the mitochondria for β-oxidation, another important source of subtracts for ketone bodies (Murray *et al.*, 2003). Therefore, carnitine increase may be interpreted as a response to higher energy demand in FM fetuses.

Finally, two metabolites detected by UPLC-MS and tentatively assigned to estriol-3-sulfate-16-glucuronide and estriol-3-glucuronide/estriol-16-glucuronide were found increased in FM cases. Estriol conjugates are formed in the fetal-placental unit from fetal precursor dehydroxyepiandrosterone which are then conjugated with glucuronic acid and/or sulfate in maternal liver to be finally excreted in urine (Yang *et al.*, 2003). Reduced unconjugated maternal

serum estriol has been described in Down's syndrome and anencephaly cases (Haddow *et al.*, 1992; Yaron *et al.*, 1998). However, the increase of conjugated estriol in human AF of FM affected pregnancies has not been reported before and no further hypothesis can be put forward with the available data.

5.2. Study of the effects of CD on 2nd trimester human amniotic fluid

The effects of chromosomal disorders (CD) on 2nd trimester AF were investigated using the ^{1}H NMR and MIR profiling, since only a small number of CD samples (about 5) was available at the time of analysis by UPLC-MS.

5.2.1. ^{1}H NMR profiling to evaluate the effects of CD type

The CD samples considered for ^{1}H NMR and MIR analysis are shown in Table 5.8, together with the corresponding number of AF samples. Several types of CD were considered (Table 5.8), therefore, PCA analysis was performed initially to evaluate any differences between the AF profiles of different CD types (Figure 5.23). No separation or tendency of any CD type was observed in the scores scatter plots (Figure 5.23), meaning that no differences between the ^{1}H NMR spectra of CD types seem to occur. The sample number of each type of CD was quite low, which may justify in part the result obtained.

Table 5.8 - Sample numbers corresponding to the chromosomal disorders considered in ^{1}H NMR and MIR profiling studies.

Chromosomal disorders	Sample numbers	
	^{1}H NMR	MIR
Trisomy 21	7	7
Trisomy 18	1	1
Marker chromosome	1	1
Kleinefelter syndrome	3	2
XYY syndrome	2	2
XXX syndrome	1	1
Mosaicisms	3	3
Inversions	2	2
Translocations	4	4
Total	24	23

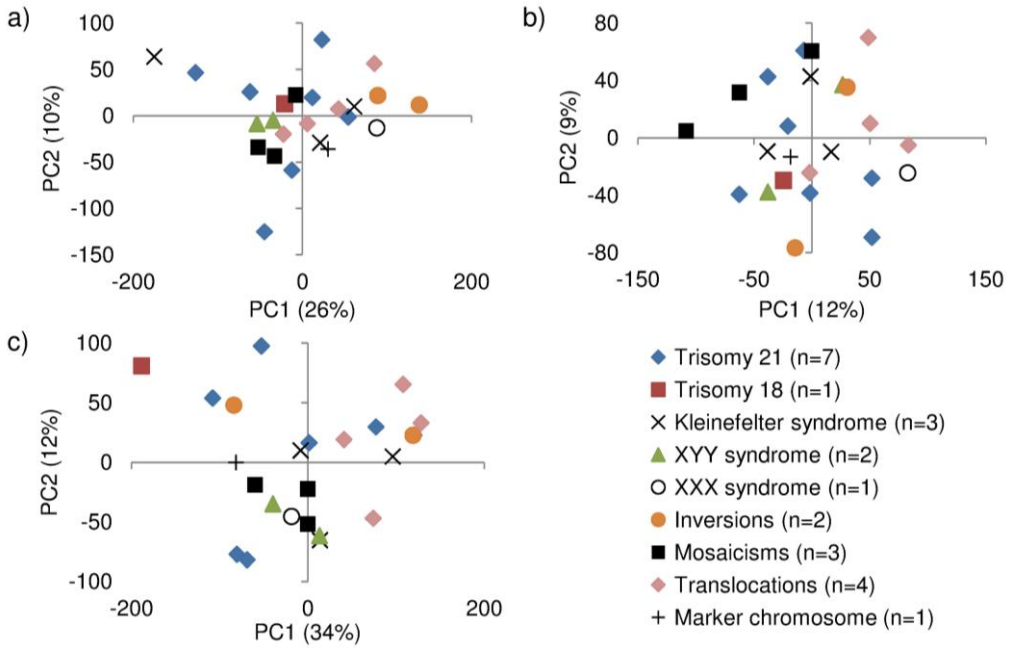


Figure 5.23 - PCA scores scatter plots of ^{1}H NMR spectra from AF samples affected by different types of CD: a) standard, b) CPMG and c) diffusion-edited spectra.

The larger group of CD, Trisomy 21 (n=7), was compared to the remaining CD types (n=17) using PLS-DA (not shown) and the models obtained could not be validated through MCCV, which means that no significant differences were found between Trisomy 21 and the other CD types. In this way, all samples CD were considered as one group to be compared with control samples. Therefore, the profiling studies presented consisted of exploratory studies to search for metabolic variations common to the majority chromosomal disorders.

5.2.2. ^{1}H NMR profiling for comparison of control and CD groups

Figure 5.24 shows the scores plots from the PLS-DA analysis of standard, CPMG and diffusion-edited ^{1}H NMR spectra of control *vs* CD samples. The PLS-DA scores plots revealed separation tendencies for all three spectra (Figure 5.24a,b,c), which were however not visible in PCA (not shown). Also no spectral differences were apparent by visual inspection of the spectra. The separation tendencies were maintained in OPLS-DA scores plots (Figure 5.24d,e,f).

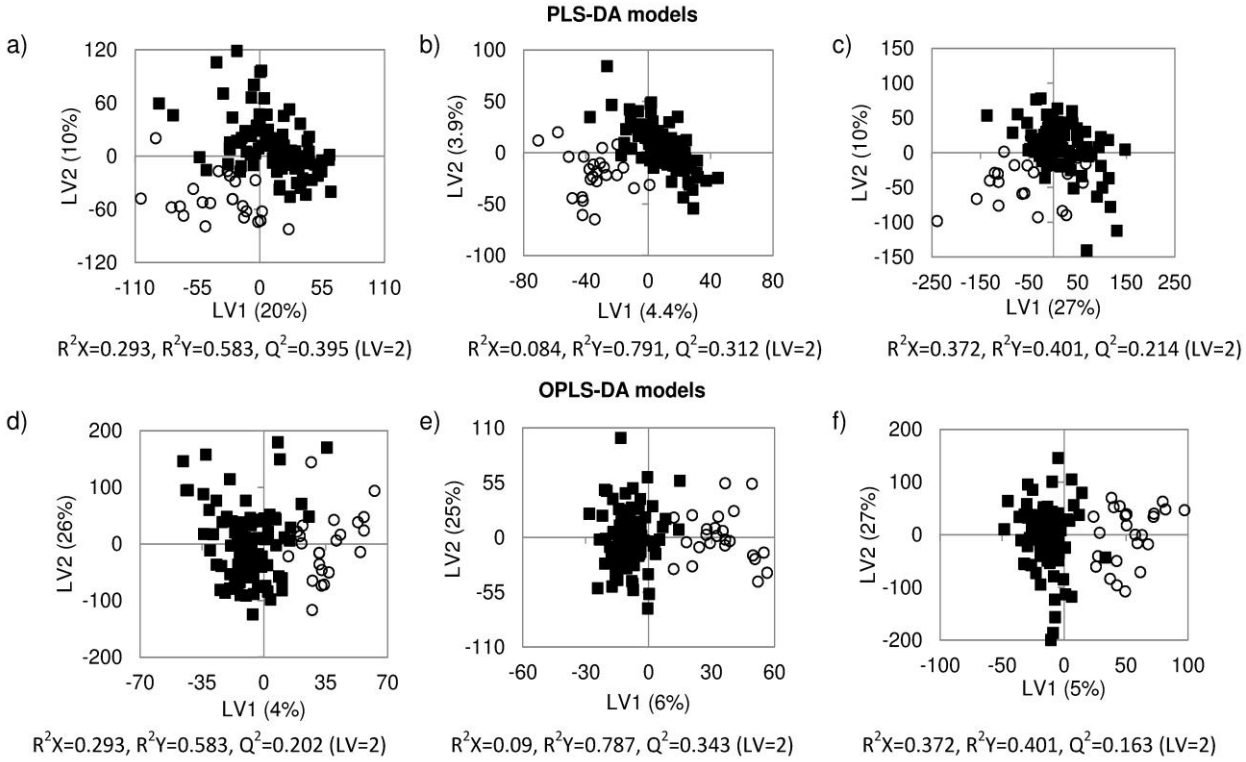


Figure 5.24 - Scores scatter plots from PLS-DA and OPLS-DA models of control (■, n=90) *vs* CD (○, n= 24). PLS-DA models: a) standard; b) CPMG; c) diffusion-edited ^{1}H NMR spectra; OPLS-DA models: d) standard; e) CPMG; f) diffusion-edited.

The Q^2 values obtained for the PLS-DA models (between 0.20-0.39) motivated their further validation through MCCV (Figure 5.25). MCCV revealed only modest results when compared with the results obtained for control *vs* FM models (Figure 5.4, page 125), with only some separation tendency noted between true and permuted distributions of Q^2 values and the overlapping of distributions nearly exceeding 50% (Figure 5.25a,b,c, left). Although high specificity values were obtained for true class models in the ROC plots (>90%), low sensitivities were obtained: ranging from 10-50% for standard spectra models (Figure 5.25a, right), 20-60% for CPMG models (Figure 5.25b, right) and 20-50% for diffusion-edited spectra (Figure 5.25c, right). Therefore, CPMG model registered the best predictive ability, also expressed by a classification rate of 86% (against classification rates of 82%, and 79% for standard and diffusion-edited models, respectively).

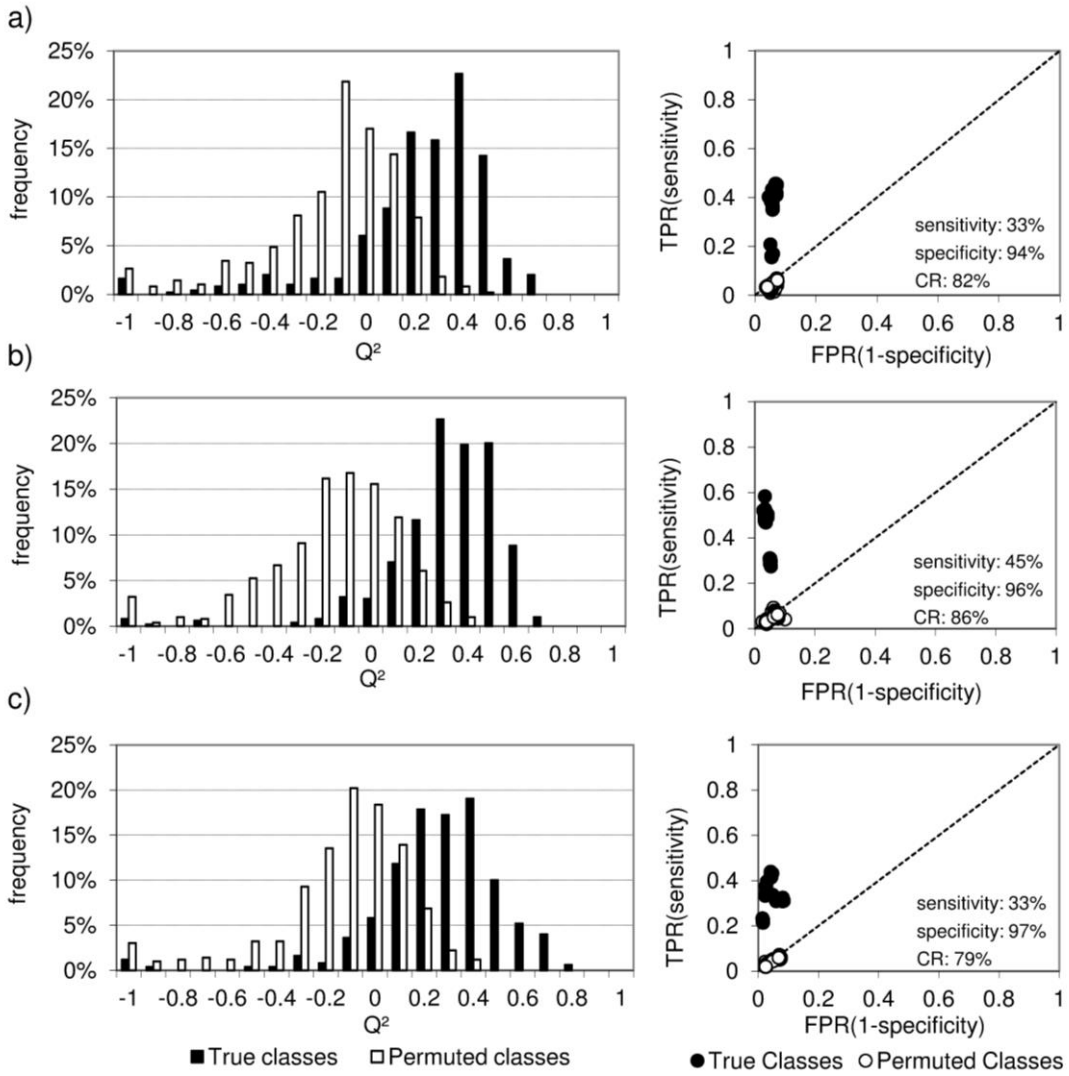


Figure 5.25 - MCCV results for PLS-DA models of control (n=90) *vs* chromosomal disorders (n=24) shown in Figure 5.21: a) standard, b) CPMG, c) diffusion-edited ^{1}H NMR spectra. Left: Q^2 distributions; right: ROC plots. Average sensitivity, specificity and classification rate (CR) values for true classes are shown in the ROC plots.

In spite of the low predictive ability of the models, the loadings were analysed in order to identify possible small metabolite variations accompanying the occurrence of CD. The loadings analysis was pursued using the LV1 loadings from OPLS-DA models from Figure 5.24d,e,f, to ease model interpretation. The LV1 loadings from standard spectra model (Figure 5.26a) revealed increases in α-oxoisovalerate, alanine, citrate, pyruvate, methionine, *myo*-inositol, threonine and glucose in CD, together with decreases in α-hydroxybutyrate, 3 unassigned spin-systems and some broad signals in the broad region of protein NH groups (6.94, 7.15 and 8.00 ppm).

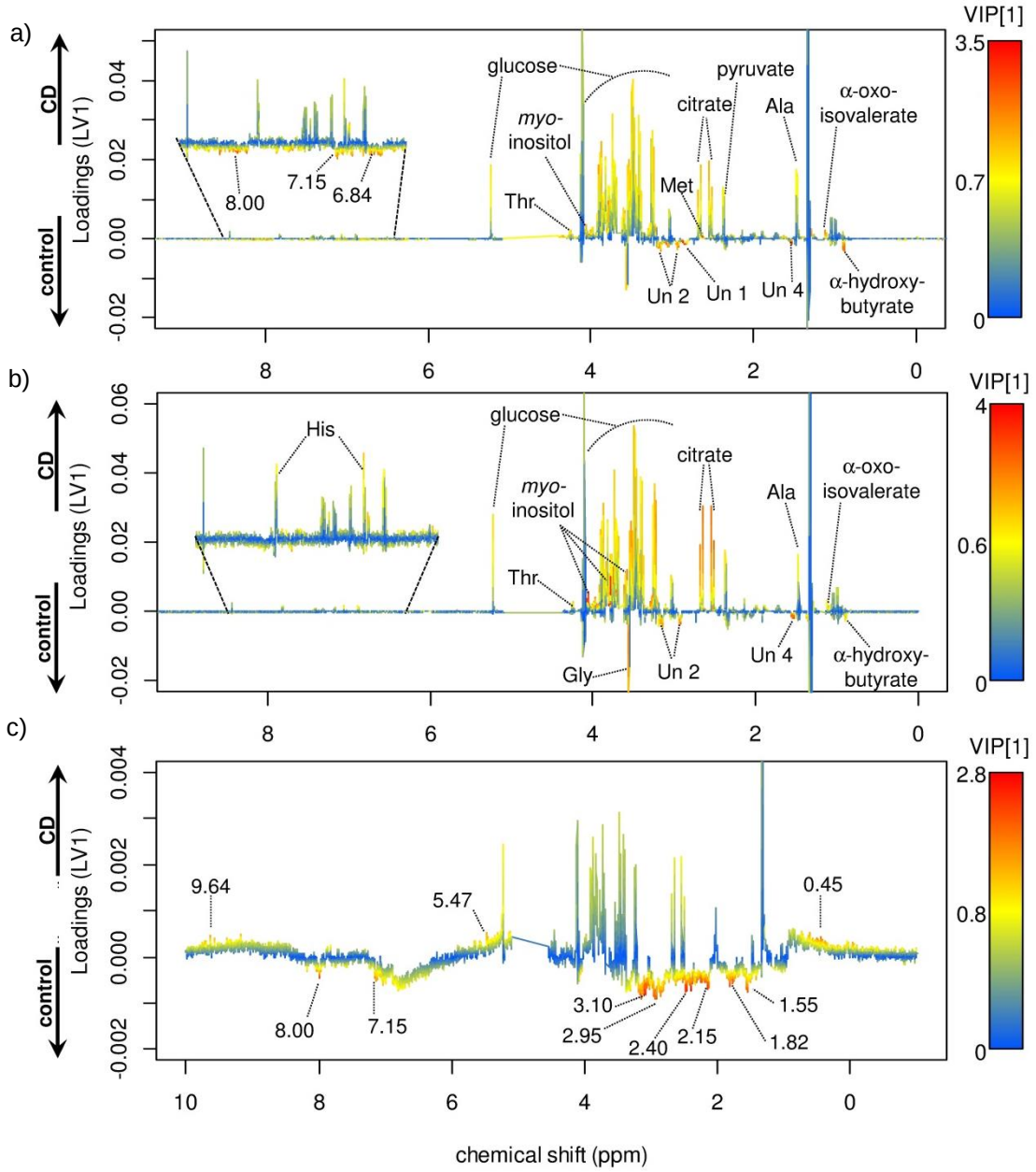


Figure 5.26 - Loadings from OPLS-DA models of control *vs* CD: a) standard, b) CPMG and c) diffusion-edited ^{1}H NMR spectra colour-coded with VIP values. Ui, unassigned signals; amino acids in three-letter code.

The increases in α-oxoisovalerate, alanine, citrate, *myo*-inositol, threonine and glucose as well as the decreases in α-hydroxybutyrate, and 2 unassigned spin systems were confirmed in the LV1 loadings of CPMG model (Figure 5.26b). Two further variations were observed in the CPMG loadings, namely the decrease in glycine and increase in histidine (Figure 5.26b). The diffusion-edited spectra loadings showed that several signals from the protein profile were decreased in CD cases including two broad signals already observed in standard spectra (centered at 7.15 and 8.00 ppm), and increases in three other signals in the broad profile (Figure 5.26c).

The significance of the identified variations was tested after signal integration using univariate analysis, with the significant variations being presented in Table 5.9. The *p*-values obtained for the metabolite variations noted in CD were in the range of 10^{-2}-10^{-5}, which in comparison to those detected previously for FM (in the range of 10^{-3}-10^{-11}) seem to indicate that smaller changes occurred in CD. Significant small increases confirmed in CD cases included α-oxoisovalerate, citrate, methionine, threonine, *myo*-inositol, glucose and two broad profile signals 0.45 and 9.64 ppm (Table 5.9). On the other hand, glycine, the 3 unidentified spin-systems (U1, U2 and U4) and the broad signals (1.55, 1.82, 2.15, 2.40, 2.95, 3.10, 6.84, 7.15 and 8.00 ppm, which may be related to proteins) were all found significantly decreased in CD cases. The results obtained present an advance compared to those obtained in a previous stage of this work with less control (n=82 *vs* n=90 here) and CD cases (n=10 *vs* n=24 here) (Graça *et al.*, 2010). In that study no significant metabolite variations were reported for the same significance level (p<0.05). The integrals obtained were also used to produce box-plots for better visualization of the variation of metabolites in controls and CD groups (Figure 5.27).

Table 5.9 - List of significant metabolite variations (p<0.05) noted in 2nd trimester human AF of CD affected fetuses. [a], signal within the metabolite spin-system chosen for integration; [b], signal integral measured in CPMG spectra; [c], signal integral measured in diffusion-edited spectra. s, singlet; d, doublet; t, triplet; m, multiplet; br, broad signal. ( ), very small changes corresponding to magnitude (mean fold change) values >1.0 and <1.05.

metabolite	δ^1H [a]	direction and magnitude of change		*p*-value
α-oxoisovalerate [b]	1.11(d)	↑	1.2	2.47×10^{-3}
citrate [b]	2.67(d)	↑	1.1	9.98×10^{-4}
glucose	5.23(d)	↑	1.2	4.00×10^{-2}
glycine [b]	3.55(s)	↓	1.1	3.26×10^{-2}
methionine [b]	2.64(t)	↑	1.1	2.59×10^{-2}
myo-inositol [b]	4.06(t)	↑	1.1	3.14×10^{-4}
threonine	4.26(m)	↑	1.1	2.71×10^{-2}
Un1	2.84(m)	↓	1.2	1.10×10^{-4}
Un2 [b]	2.93(m)	↓	1.8	2.15×10^{-4}
Un4 [b]	1.54(m)	↓	1.2	1.20×10^{-5}

Broad profile signals

signal	direction/magnitude of change		*p*-value	*signal*	direction/magnitude of change		*p*-value
0.45(br) [c]	↓	1.1	2.22×10^{-3}	3.10(br) [c]	↓	1.1	6.72×10^{-5}
1.55(br) [c]	(↓)	1.0	3.70×10^{-3}	5.47(br) [c]	↑	1.3	2.90×10^{-2}
1.82(br) [c]	(↓)	1.0	1.19×10^{-3}	6.84(br)	(↓)	1.0	3.24×10^{-3}
2.15(br) [c]	(↓)	1.0	8.18×10^{-5}	7.15(br) [c]	(↓)	1.1	1.44×10^{-3}
2.40(br) [c]	(↓)	1.0	4.51×10^{-4}	8.00(br) [c]	(↓)	1.0	6.41×10^{-3}
2.95(br) [c]	↓	1.1	2.38×10^{-4}	9.64(br) [c]	↑	1.2	2.92×10^{-2}

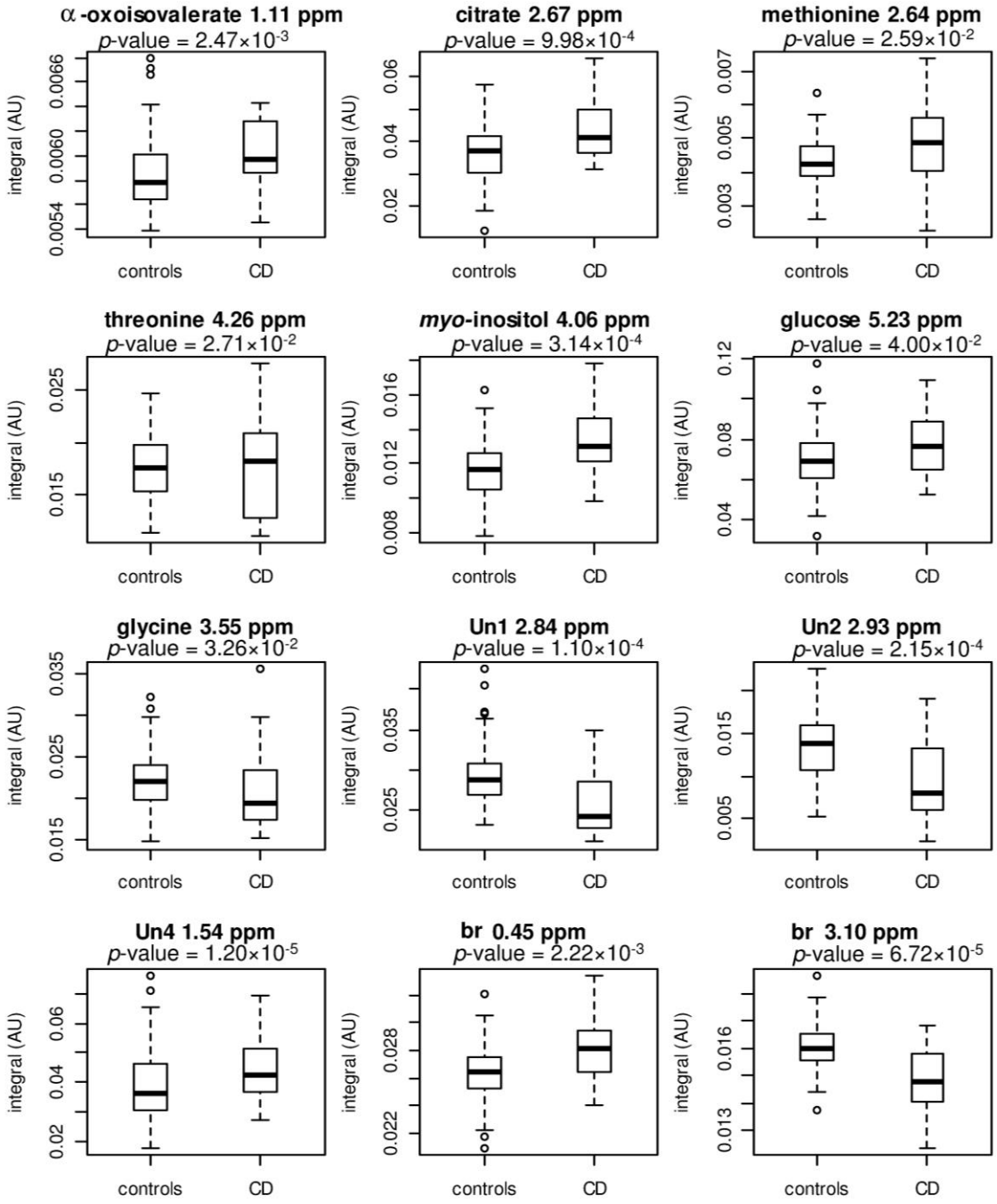


Figure 5.27 - Box-plots for selected metabolite signals identified from the loadings of AF CD OPLS-DA models. The respective *p*-values are also shown. br, broad signals (possibly from proteins).

From the close inspection of box-plots of α-oxoisovalerate, threonine, methionine, glycine and glucose, it is clear that despite the differences in their median values, all box-plots show overlap of the CD and control distributions (Figure 5.27). Moreover, the dispersion of integral values of the above mentioned metabolites is greater in CD than in controls, which may reflect the variability due to different types of CD cases under study. The integrals of citrate, *myo*-inositol, glucose, one unassigned metabolite (spin-system Un4, 1.54 ppm) and broad signals at 0.45 and

9.64 ppm showed increases in CD group, together with decreases in two unassigned metabolites (spin-systems Un1, 2.84 ppm and Un2, 2.93 ppm) and broad signals in the region of 1.55-8.00 ppm (possibly protein signals) (Table 5.9). No assignment was possible for the unassigned spin-systems with basis on previous knowledge from LC-NMR or 2D NMR experiments. The described variations form a metabolic response common to all or to the majority of the analysed CD cases.

Regarding the trisomy 21 cases, which comprise 7 out of 24 cases in the CD group used in the present study (Table 5.8), new PLS-DA models were performed in order to compare the trisomy 21 cases to control (n=90). The scores plots showed initially some separation, particularly for diffusion-edited spectra (not shown). However, MCCV analysis of the models (from standard, CPMG and diffusion-edited) registered very low sensitivity, with average values below 19% (not shown), therefore, the models were not considered for further analysis. The low sensitivity may due to the reduced number of trisomy 21 cases considered, but on the other hand it may also be a result of very small metabolite changes resulting of trisomy 21.

5.2.3. MIR profiling for comparison of control and CD groups

The MIR analysis of a group of CD (n=23) in comparison with a control group (n=40) was attempted. The results scores scatter plots obtained in PLS-DA analysis are shown in Figure 5.28).

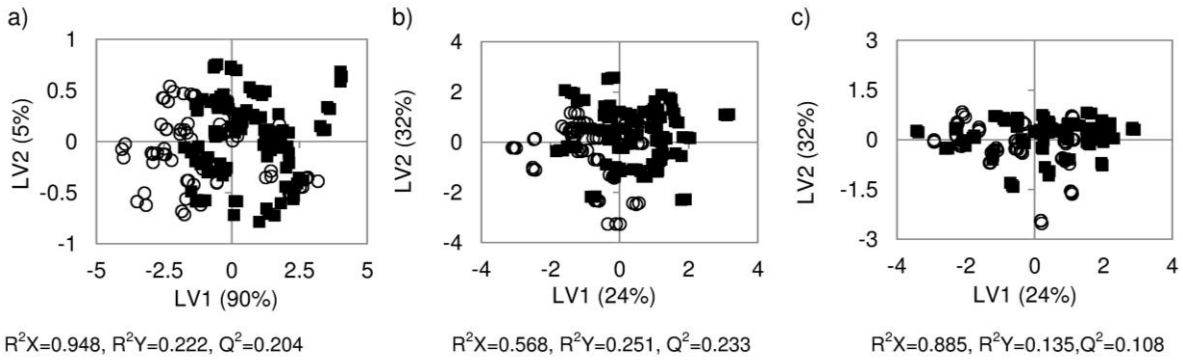


Figure 5.28 - PLS-DA scores scatter plots of MIR spectra models of controls (■, n=40) *vs* CD (○, n=23) for the spectral regions: a) 3000-2800 cm^{-1}, b) 1800-1200 cm^{-1}, c) 1200-900 cm^{-1}.

There was no separation noted in the PLS-DA scores scatter plots of the spectral regions 3000-2800 cm^{-1}, 1800-1200 cm^{-1} and 1200-900 cm^{-1}, with a complete overlap of between controls and CD samples being observed (Figure 5.28). The results obtained for 2nd derivative transformed spectral regions were quite similar (not shown).

PLS-DA analysis considering trisomy 21 samples (n=7) and control samples (n=40) was also performed considering the same MIR spectral regions as well as their respective 2nd derivative

transformations. The resulting PLS-DA models showed mainly no separation in the scores plots (not shown), being also accompanied by low Q^2<0.2 values. The lack of separation in PLS-DA MIR is not surprising, given the small number of metabolite variations detected by ^{1}H NMR. On the other hand, the strong overlap of the MIR bands of more abundant metabolites (such as glucose, lactate, albumin), may hinder detection of less abundant metabolites.

5.2.4. Proposed interpretation of the metabolite variations occurring in CD cases

Based on the variations observed, slight alterations in pathways of sugar metabolism (with the involvement of *myo*-inositol and glucose) and lipid and biomembranes biosynthesis (due to the increase of *myo*-inositol and citrate, which are involved in the biosynthesis of phospholipids and fatty acids, respectively) (Murray *et al.*, 2003), can be suggested as possible metabolite derangements in CD cases. Amino acids metabolism may also be affected, as it is suggested by the average changes observed, particularly the increases of α-oxoisovalerate, threonine and methionine, together with a decrease in glycine (Murray *et al.*, 2003). However, the high inter-individual variability in the levels of these amino acids was also observed inside the CD group, as shown in the respective box-plots (Figure 5.27, page 157) hinders the correct interpretation of the amino acid changes. Variations in amino acid concentrations in 2nd trimester AF, namely decreases in isoleucine, leucine, glycine, lysine, taurine, valine, ornithine and glutamate and increases in glutamine have been found of trisomy 21 cases (n=24) (Amorini *et al.*, 2012). This shows the need for an improved characterization work using larger numbers of samples of this type of CD.

Alterations in the ^{1}H NMR protein profile of CD cases (decrease of broad signals) suggest decreased fetal protein expression/excretion into AF (Table 5.9). In fact, the differential protein expression in human AF have been described for several chromosomal disorders including 21 trisomy (Cho *et al.*, 2010), Kleinefelter syndrome (Anagnostopoulos *et al.*, 2010) and Turner syndrome (Mavrou *et al.*, 2008). However, the protein variations observed by ^{1}H NMR may be limited to the most abundant AF proteins as seen in chapter 4 (section 4.3.2, pages 108-110). It is particularly difficult to assign the broad signals varying in CD to specific proteins, however, the involvement of the major AF glycoproteins can be ruled out due to the absence of variation of the N-acetyl group characteristic resonances in the loadings plots. One possibility is the decrease in albumin in CD which should have resonances in the 1.55-8.00 and a characteristic group of resonances around 3.00 ppm. However, no evidence was found in literature to support this hypothesis.

The variation of the unassigned spin-systems, Un1, Un2 and Un4 was also observed in CD cases. The occurrence of variations of these signals reinforces the need for further assignment work. Interestingly, two of these spin-systems, Un1 and Un2 were both found decreased in FM (p-values$\approx 10^{-5}$) and CD (p-values$\approx 10^{-4}$), thus indicating the possibility of a common metabolite alteration, possibly a common stress effect, which is however less pronounced in CD group due to the lower significance of the variation.

The results obtained seem to suggest that small metabolite changes occur in CD cases, therefore encouraging the further characterization of larger sample populations of each type of chromosomal disorder by ^{1}H NMR and by MIR.

6. Study of the effects of disorders occurring later in pregnancy (PTD, PROM, PE, IUGR and GDM) on 2nd trimester human amniotic fluid

The work presented in this chapter aims at exploring the metabolite composition of 2nd trimester human AF as a means to devise predictive multivariate methods for disorders affecting both mother and fetus later in pregnancy (during the 3rd trimester). The disorders include preterm delivery (PTD), preeclampsia (PE), intrauterine growth restriction (IUGR), gestational diabetes mellitus (GDM) and premature rupture of the membranes (PROM). The diagnosis of such disorders may be performed 7 - 20 weeks after amniocentesis (average values), depending on the disorder. Since AF was collected prior to the occurrence of the disorders, the groups were designated throughout the chapter *pre*-PTD, *pre*-PE, *pre*-IUGR, *pre*-diagnostic GDM and *pre*-PROM.

It is important to mention that due to the sampling time-scales, the numbers of samples (both controls and disorders) used in ^{1}H NMR, UPLC-MS and MIR profiling studies was not the same.

6.1. Analysis of 2nd trimester amniotic fluid composition for the prediction of PTD (*pre*-PTD group)

6.1.1. ^{1}H NMR profiling for comparison of control and *pre*- PTD groups

The ^{1}H NMR spectra obtained for 2nd trimester AF samples from women which delivered preterm between 34 and 36 weeks, *pre*-PTD group (n=18) were compared with a control (n=90) group. There is an obvious imbalance between the numbers of controls and *pre*-PTD, however, the downsizing of control group was avoided to prevent the biased selection of controls.

PCA of the ^{1}H NMR spectral data of controls *vs pre*-PTD did not reveal any separation trend between the both classes (not shown). Separation tendencies were only observed between control and *pre*-PTD in PLS-DA and OPLS-DA models of standard and CPMG ^{1}H NMR spectra (Figure 6.1a and b, respectively). A tendency for separation was also observed in diffusion-edited spectra models (not shown) which registered low Q^2 values (Q^2<0.2) indicating lower predictive ability. This result indicates that the contribution of high M_w compounds to the differentiation between classes is not significant. All models were subjected to further validation testing through MCCV.

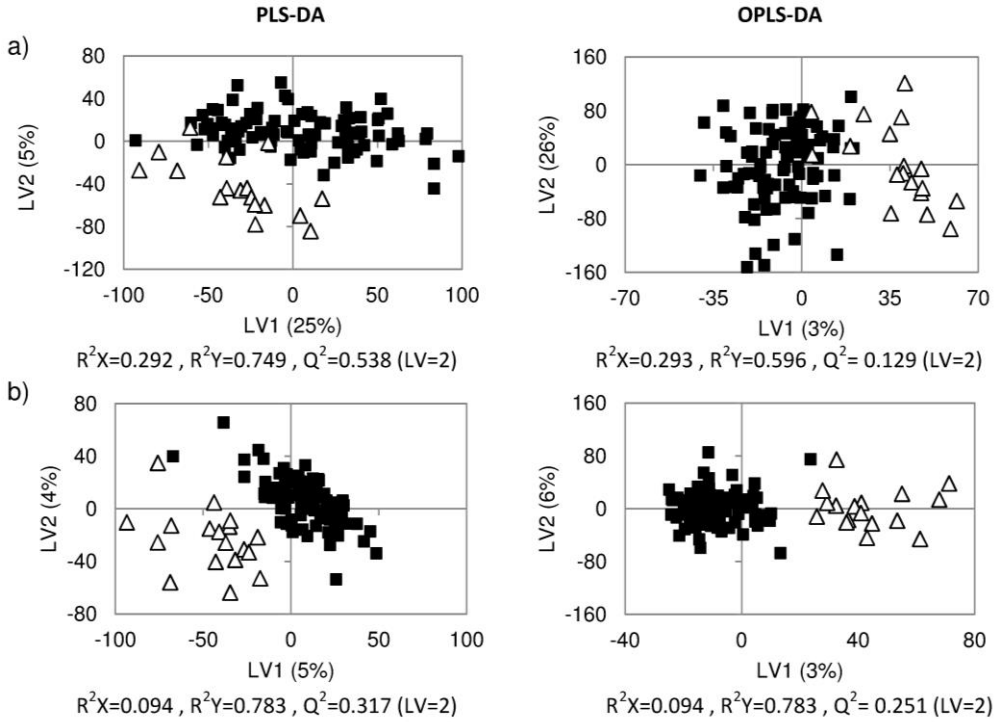


Figure 6.1 - Scores scatter plots from PLS-DA and OPLS-DA models of ^{1}H NMR spectra 2nd trimester AF of control (■, n=90) and *pre*-PTD (Δ, n=18) groups and respective parameters: a) standard spectra, b) CPMG spectra.

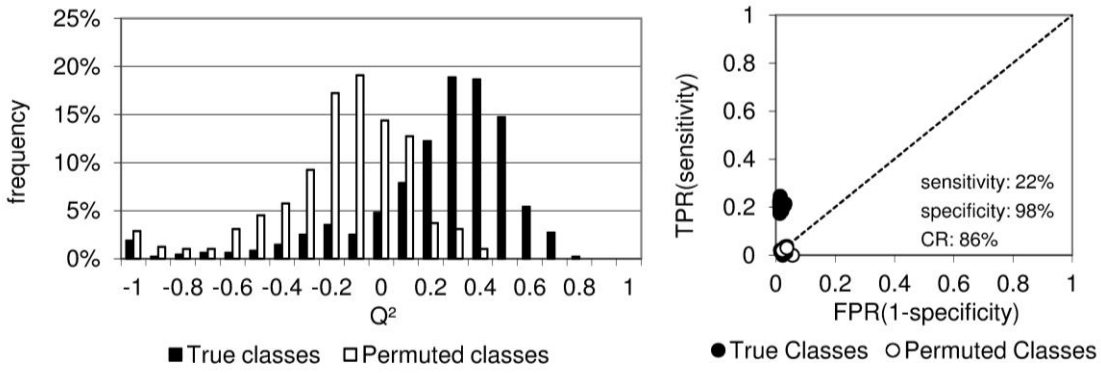


Figure 6.2 - MCCV results for PLS-DA model of CPMG ^{1}H NMR human AF spectra of control (n=90) and *pre*-PTD (n=18). Q^2 distributions (left) and ROC plots (right) of true classes and permuted classes models are shown. The true classes average sensitivity, specificity and classification rate (CR) are indicated in the ROC plot.

Despite the apparently lower Q^2 value obtained in 7-fold CV for the CPMG model when compared with the standard spectra model (Q^2=0.317 *vs* Q^2=0.538, respectively), the former registered better results in MCCV (Figure 6.2). As shown in Figure 6.2(left), the Q^2 distributions of true and permuted class models of CPMG spectra data show a tendency for separation, despite the large overlap noted. The sensitivity of the models noted in ROC plot (average 22%) indicated a low true positive rate for true classes (Figure 6.2, right). Despite the high specificity noted

(average 98%) and classification rate of 86%, the MCCV results denote a low predictive ability of PLS-DA model of CPMG spectra. The PLS-DA models of standard and diffusion-edited spectra registered a poorer predictive ability, characterized by an almost complete overlap of Q^2 distributions of true and permuted class models (not shown), sensitivities ranging from 0-20% in ROC plots (not shown) and classification rates of 84%. A low predictive ability was also obtained in an early stage of this work (not shown), where slightly smaller sample groups were considered (*pre*-PTD n=12, controls n=82) (Graça *et al.*, 2010). In another study, Romero *et al.* (2010) reported accuracies of about 88% for the correct prediction of occurrence of PTD from women delivering without membrane rupture (n=40), with membrane rupture (n=33), and with membrane rupture with intra-amniotic infection (n=40), based on analysis of 2nd and 3rd trimester AF using GC-MS and LC-MS and multivariate analysis. However, the study was performed closer to the occurrence of PTD, at the time women experienced the first symptoms of PTD (Romero *et al.*, 2010), which may have some impact on the results obtained. Nevertheless, the study of Romero *et al.* (2010) also shows that the use of high number of samples may help improve the predictive ability of the method.

Despite the low predictive ability of the standard and CPMG spectra models, loadings analysis was performed to further explore possible small metabolites variations occurring in *pre*-PTD. The LV1 loadings from OPLS-DA model of standard spectra (which showed improved separation in LV1 component compared to the corresponding PLS-DA model) and also the LV1 loadings from PLS-DA model of CPMG (Figure 6.3a and b, respectively) were inspected. The loadings from the standard spectra model, indicated decreases in the contents of several amino acids in *pre*-PTD group, including alanine, isoleucine, leucine, lysine, glutamine, histidine, methionine, phenylalanine, tyrosine and valine; together with decreases in the contents of *myo*-inositol, citrate and formate (Figure 6.3a). On the other hand, several signals were found increasing in *pre*-PTD group, such that of allantoin and the unassigned signals at 1.10, 1.22, 2.98 and 7.06 ppm (Figure 6.13a). The loadings from CPMG PLS-DA model indicated decreases in *pre*-PTD group of methionine and the unassigned signals 1.80 and 3.11, as well as increases in the signals of acetoacetate, β-hydroxybutyrate, pyruvate, choline, serine, glucose, allantoin, creatinine, threonine and the unassigned signal at 7.87 ppm in *pre*-PTD (Figure 6.3b). Increases in *pre*-PTD signals 0.85, 0.92 (both broad signals) and 2.03 (N-acetyl groups) were also noted, raising the possibility of relation to glycoprotein content.

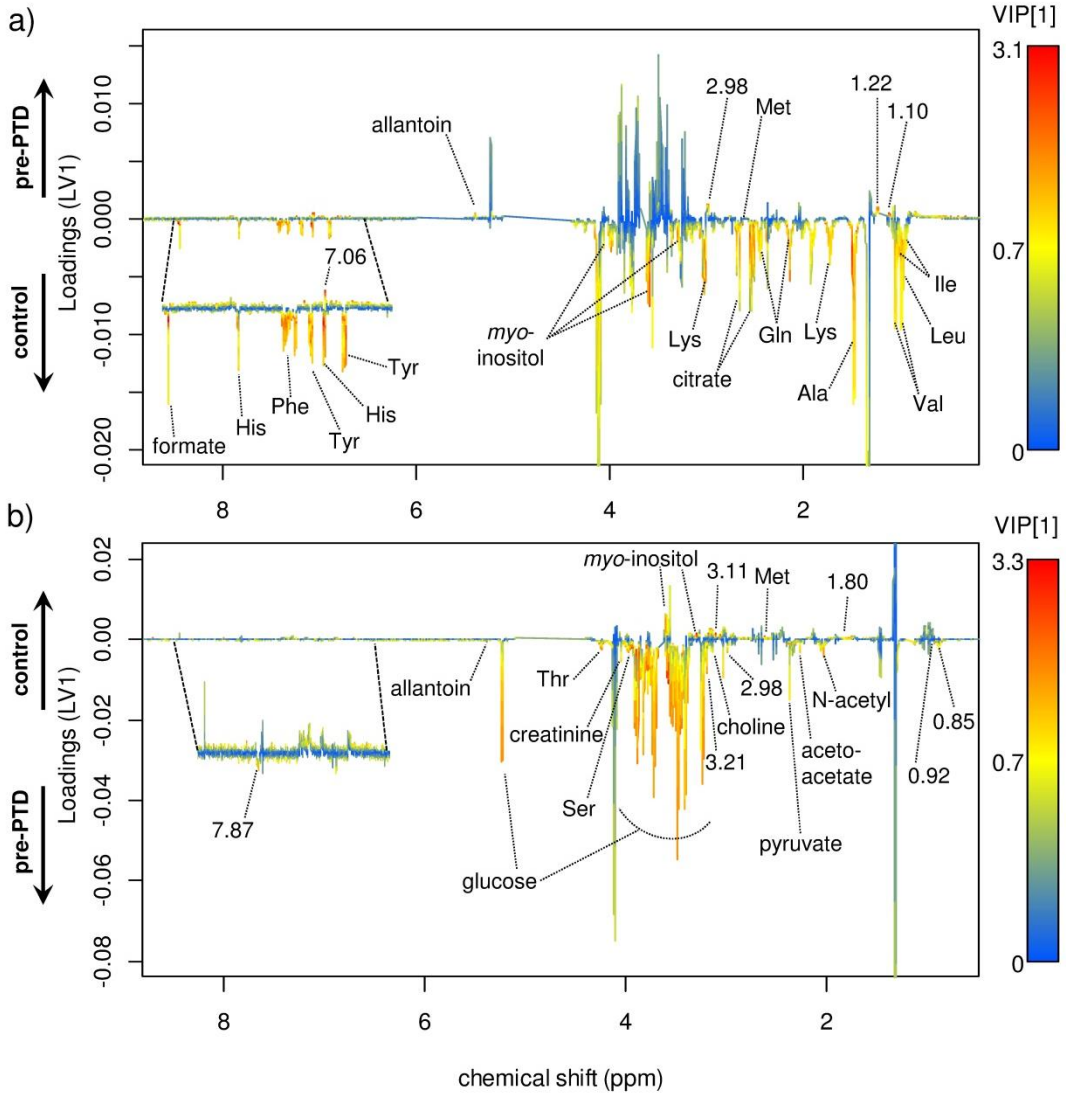


Figure 6.3 - LV1 Loadings from control *vs pre*-PTD ^{1}H NMR models colour-coded with VIP values: a) OPLS-DA LV1 loadings of standard spectra and b) PLS-DA LV1 loadings of CPMG spectra. Amino acids in three-letter code; unassigned signals are indicated with their respective values.

Interestingly, only allantoin (increasing in *pre*-PTD), methionine and *myo*-inositol (decreasing in *pre*-PTD) appear to vary consistently in the standard (Figure 6.3a) and CPMG (Figure 6.3b) experiments. This may be attributed to overlap of broad and sharp resonances occurring in the standard spectra, which interferes in the detection of small differences in sharp signals.

The integrals of all ^{1}H NMR signals were obtained and evaluated using univariate analysis and the statistically significant variations (p<0.05) were summarized in Table 6.1. The variations validated through univariate analysis included the increases in *pre*-PTD for allantoin, acetoacetate, glucose, choline, serine, threonine, N-acetyl group (possibly associated glycoprotein resonances), β-hydroxybutyrate and a broad signal at 0.92 ppm (the latter two registering the

most significant increases), as well as 3 unassigned signals (Table 6.1). On the other hand, significantly decreases in *pre*-PTD were registered for the amino acids isoleucine, leucine, alanine, lysine, glutamine, histidine, methionine, phenylalanine, tyrosine, formate and 2 unassigned signals (registering the most significant decreases) (Table 6.1).

Table 6.1 - Significant metabolite variations (p<0.05) noted in 2nd trimester human AF samples of *pre*-PTD relative to healthy controls. [a], signal within the metabolite spin-system chosen for integration; all variations measured in CPMG spectra except *, measured in standard spectra. s, singlet; d, doublet; t, triplet; m, multiplet; br, broad signal. ( ), very small changes corresponding to magnitude (mean fold change) values >1.0 and <1.05.

metabolite	δ^1H [a]	direction and magnitude of change		*p*-value
acetoacetate	2.27 (s)	↑	1.1	2.58×10^{-2}
alanine *	1.47 (d)	↓	1.1	2.08×10^{-2}
allantoin *	5.39 (s)	↑	1.4	2.19×10^{-3}
β-hydroxybutyrate	2.38 (m)	↑	1.1	8.29×10^{-4}
choline	3.19 (s)	↑	1.1	6.77×10^{-3}
formate *	8.45 (s)	↓	1.1	2.17×10^{-2}
glucose	5.23 (d)	↑	1.2	1.44×10^{-3}
glutamine *	1.45 (m)	↓	1.1	4.93×10^{-2}
histidine *	7.83 (s)	↓	1.1	2.17×10^{-2}
Isoleucine *	1.00 (d)	(↓)	1.0	1.56×10^{-2}
leucine *	0.95 (t)	↓	1.1	3.48×10^{-2}
lysine *	1.72 (m)	(↓)	1.0	4.43×10^{-2}
methionine	2.64 (t)	↓	1.1	5.61×10^{-3}
phenylalanine *	7.42 (m)	(↓)	1.0	2.52×10^{-3}
serine	3.95 (m)	↑	1.1	2.42×10^{-2}
threonine	4.26 (m)	↑	1.1	2.86×10^{-2}
tyrosine *	6.91 (d)	↓	1.1	8.24×10^{-3}
valine *	1.03 (d)	(↓)	1.0	1.29×10^{-2}
Possibly glycoprotein related:				
N-acetyl	2.03 (s)	↑	1.1	6.44×10^{-3}
0.85 ppm	0.85 (br)	↑	1.2	5.26×10^{-3}
0.92 ppm	0.92 (br)	↑	1.1	3.63×10^{-4}
Unassigned:				
1.10 ppm *	1.10 (d)	(↑)	1.0	3.08×10^{-3}
1.22 ppm *	1.22 (m)	↑	1.1	1.39×10^{-2}
1.80 ppm	1.80 (br)	↓	1.1	3.20×10^{-4}
3.11 ppm	3.11	↓	1.2	1.92×10^{-5}
7.87 ppm	7.87	↑	1.9	2.58×10^{-2}

Comparing the results obtained in the present work with those from an early stage involving less sample numbers (not shown) (Graça *et al.*, 2010), the only common metabolites variations detected were the increase in allantoin and decrease in alanine. Moreover, the significant decreases in citrate and *myo*-inositol observed previously were not found in the present study, possibly due to the use of an increased number of samples. Some of the significant metabolite variations (p<0.05) obtained in the present study were also expressed in the box-plots shown in Figure 6.4.

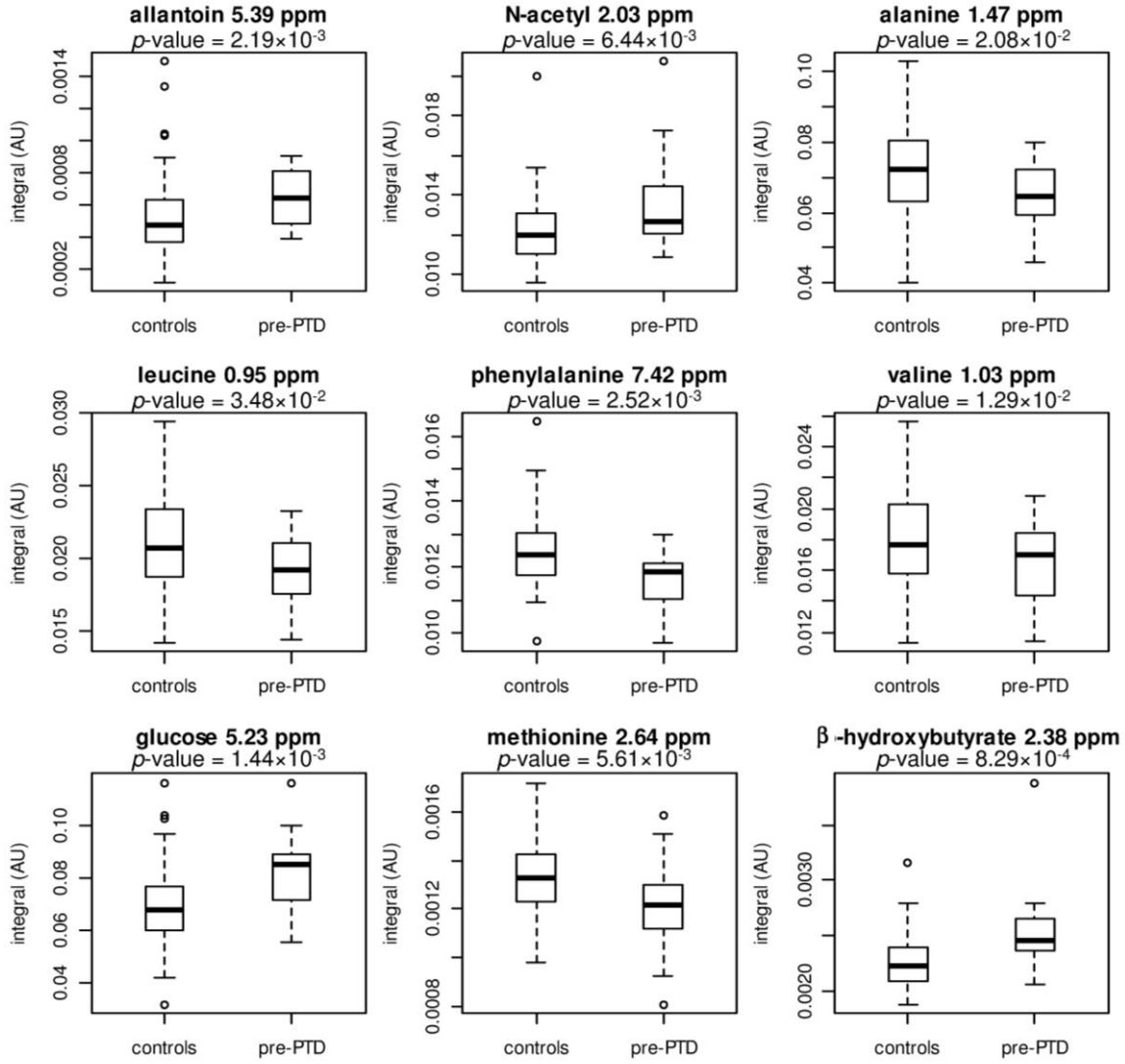


Figure 6.4 - Box-plots of selected ^{1}H NMR signals from metabolites registering significant variations in *pre*-PTD (n=18) relative to control group (n=90).

The distribution of integrals for controls and *pre*-PTD of alanine, leucine and valine show some overlap (Figure 6.4), with a similar trend observed for histidine, isoleucine and lysine indicating only slight differences in the levels of these metabolites between groups. Clear decreases in

methionine and phenylalanine in the *pre*-PTD cases were observed when compared with controls (Figure 6.4). The *pre*-PTD cases showed an average increase in the content of allantoin when compared with control group. However, 4 out of 90 control samples also registered high allantoin levels being noted as outliers in the corresponding box-plot (Figure 6.4). Allantoin is a product of oxidation of uric acid and can be regarded as a marker of oxidative stress (Tolun *et al.*, 2010). Therefore, extrinsic effects linked to diet or life style may have resulted in higher values of allantoin in the control samples.

6.1.2. UPLC-MS profiling for comparison of control and *pre*-PTD groups

The AF samples from control and *pre*-PTD were analysed by UPLC-MS using the conditions previously described in chapter 3 (section 3.5, pages 77-79).

No separation was observed in PCA scores scatter plots (not shown) which motivated the application of PLS-DA to the data sets. Class separation was then obtained for the data sets HSS/ESI$^+$ (Pareto scaled) (Figure 6.5a), HILIC/ESI$^+$ (Log transformed) (Figure 6.5a) and HSS/ESI$^-$ (Log transformed) (not shown). A list of parameters from the PLS-DA and OPLS-DA UPLC-MS models, where separation between control and *pre*-PTD was observed is shown in Table 6.2.

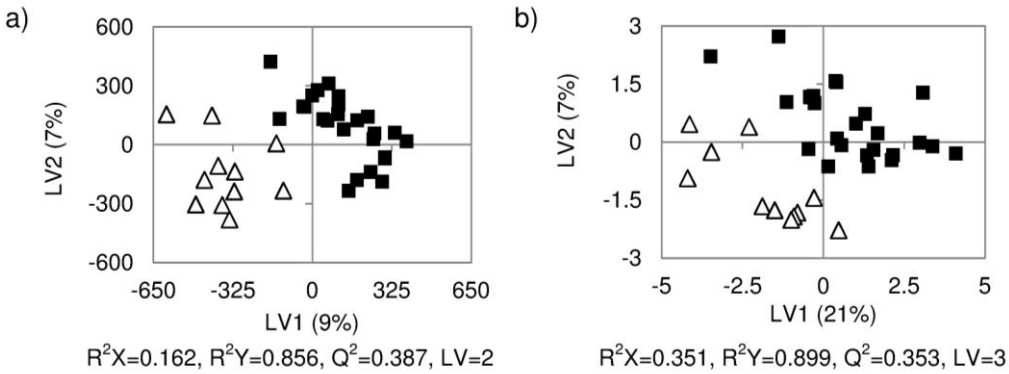


Figure 6.5 - Scores scatter plots for PLS-DA and OPLS-DA UPLC-MS models of control (■, n=26) vs *pre*-PTD (Δ, n=11): a) HSS/ESI$^+$ (Pareto scaled), b) HILIC/ESI$^+$ (Log transformed).

Table 6.2 - Parameters from control *vs pre*-PTD PLS-DA and OPLS-DA models from UPLC-MS data.

UPLC	detection	scaling or	*PLS-DA model parameters*				*OPLS-DA model parameters*			
column	mode	transformation	R²X	R²Y	Q²	LV	R²X	R²Y	Q²	LV
HSS	ESI$^+$	Pareto	0.162	0.856	0.387	2	0.162	0.856	0.460	2
HSS	ESI$^-$	Log	0.122	0.929	0.296	2	0.122	0.929	0.282	2
HILIC	ESI$^+$	Log	0.351	0.899	0.353	3	0.275	0.795	0.031	2

MCCV was performed for all separating models as illustrated in Figure 6.6 for the PLS-DA model of HSS/ESI$^+$ (Pareto scaled) data set. Mostly overlapped Q^2 distributions of true and permuted classes models were obtained for the PLS-DA model of HSS/ESI$^+$ (Pareto scaled) data set, with a small tendency of separation being observed (Figure 6.6). The ROC plot revealed separation between true and permuted class models and average specificity values of 84%, with an associated low average sensitivity of only 38% (Figure 6.16). The models from HSS/ESI$^-$ and HILIC/ESI$^+$ data sets registered similar MCCV results (not shown). The sensitivity values were superior to those registered for ^{1}H NMR models (around 20%), but still indicating a very low predictive power of the UPLC-MS models.

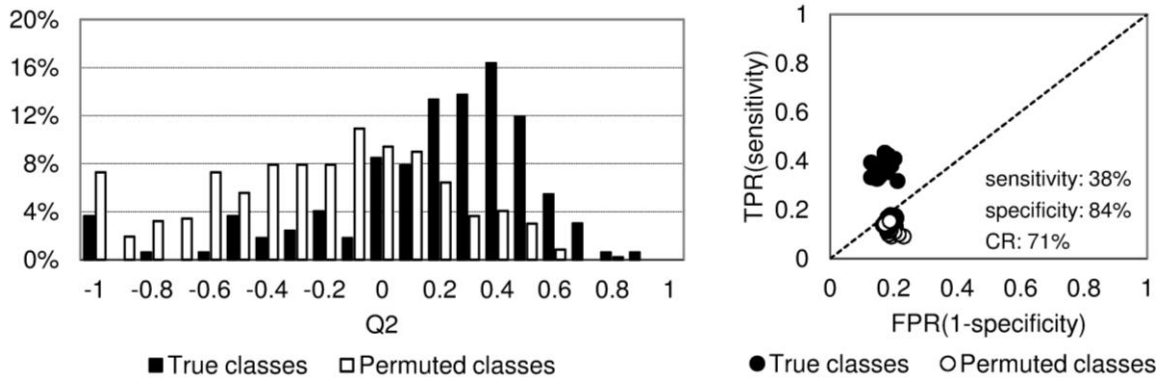


Figure 6.6 - MCCV results for UPLC-MS PLS-DA model of control *vs pre*-PTD of HSS/ESI$^+$ (Pareto scaled). The average sensitivity, specificity and classification rate (CR) for the true class models are shown in the ROC plot.

The classification rates registered in UPLC-MS models were much lower (71–75%) then those registered by ^{1}H NMR models (84-86%), raising the need for tentative further improvement, for instance using larger sample numbers. Bearing in mind that any eventual metabolite changes found through loadings analysis are merely indicative and do not bear any predictive power in the current study, loadings interpretation was performed using the procedure described previously in chapter 5 (section 5.1.3, pages 134-135), and several features were identified from which 38 registered significant variations (p<0.05) in *pre*-PTD cases. From these features, 10 features could be assigned (Table 6.3). Small decreases were noted in histidine, isoleucine/leucine, methionine, phenylalanine and valine, together with an increase in the sodium adduct of an hexose MS feature 203.05 m/z (probably arising from glucose, since this is the most abundant monosaccharide in human AF), (Table 6.3). These variations were also observed by ^{1}H NMR, as shown in the previous section (Table 6.1, page 165).

Table 6.3 - List of assigned significant UPLC-MS features obtained for 2nd trimester AF of *pre*-PTD group compared to the control group. Feature intensities are indicated as: very high (>10^5), high (>10^4, <10^5), low (>10^3, <10^4) and very low (<10^3). [a] fragment; [b] adduct; [c] isotope of the molecular peak or adduct.

Metabolite	UPLC column	detection mode	m/z	RT (min.)	feature intensity	PLS-DA model	Direction and fold change		*p*-value
hexose	HSS	ESI^+	203.05[b]	0.5	very high	Pareto	↑	1.4	2.04×10^{-03}
	HSS	ESI^+	383.12	0.5	Low	Pareto			
isoleucine/leucine	HSS	ESI^+	133.09[c]	1.9	Low	Pareto	↓	1.3	5.45×10^{-03}
methionine	HSS	ESI^+	150.05	1.0	high	Pareto	↓	1.3	4.26×10^{-04}
	HSS	ESI^+	133.02[a]	1.0	high	Pareto			
	HSS	ESI^+	104.04[a]	1.0	Low	Pareto			
phenylalanine	HSS	ESI^+	167.08[c]	2.9	high	Pareto	↓	1.1	1.37×10^{-03}
valine	HSS	ESI^+	118.07	0.9	high	Pareto	↓	1.2	4.85×10^{-03}
histidine	HSS	ESI^-	309.98[b]	0.8	Low	Log	↓	1.7	6.72×10^{-04}
phenylalanine	HILIC	ESI^+	210.05[b]	4.8	Low	Log	↓	1.6	1.18×10^{-02}

Other ^{1}H NMR detected variations such as those of alanine, allantoin, choline, β-hydroxybutyrate, acetoacetate, formate and serine were not detected by UPLC-MS probably due to their poor chromatographic retention under the conditions used. Tyrosine and threonine were also not detected by UPLC-MS as varying significantly, possibly due to the difference in the sample numbers used, rather that due to chromatographic retention issues, since these amino acids are well retained under the HSS chromatography conditions used. A list of 28 features with significant variation in *pre*-PTD subjects (p<0.05) for which no assignments were possible at this stage was also obtained (see Annex VI - Lists of UPLC-MS unassigned features detected in AF). The occurrence of such variations may indicate the presence of more significantly affected metabolites, which should be investigated in future studies. In fact a large family of compounds has been shown to be associated with the occurrence of PTD, including several amino acids, sugars and organic acids, detected by a combination of LC-MS and GC-MS approaches, which were not detected in the present study (Romero *et al.*, 2010). However, the low predictive power of UPLC-MS models should be improved as well, for instance by increasing the number of cases, which should lead to the identification more significant metabolite variations.

6.1.3. MIR profiling for comparison of control and *pre*-PTD groups

MIR spectra were acquired for *pre*-PTD AF samples (n=14) and control samples (n=40). PLS-DA models were then obtained considering the spectral regions previously introduced in chapter 5 (3000-2800 cm^{-1}, 1800-1200 cm^{-1} and 1200-900 cm^{-1} before and after 2nd derivative transformation). Separation was only obtained when the spectral region of 1200-900 cm^{-1} was considered after 2nd derivative transformation (Figure 6.7a).

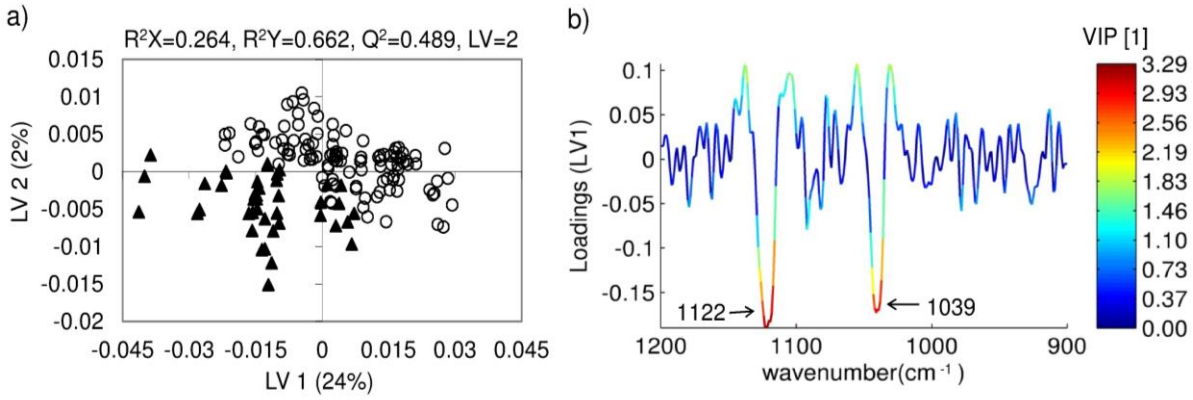


Figure 6.7 - PLS-DA scores scatter plot of *pre*-PTD (▲, n=14) *vs* control (○, n=40) of 2nd derivative of 1200-900 cm^{-1} spectral region using 3 technical sample replicates: a) scores scatter plot; b) LV1 loadings plot coloured by VIP.

The MCCV of the 2nd derivative 1200-900 cm^{-1} PLS-DA model revealed some tendency in separation of Q^2 distributions of true and permuted class models (Figure 6.8). In fact, 70% of true class models registered Q^2>0.2, while permuted classes models Q^2 values spanned from -1 to 0.6 with the most frequent value being Q^2=0, i.e. permuted classes were largely non predictive, as expected (Figure 6.8, left). The ROC plot shows that true class models registered sensitivity values in the range 30% to 70% (average 41%) and specificity values >80% (average 87%) (Figure 6.8, right). Despite the spread in sensitivity values, the PLS-DA model of the MIR of the 2nd derivative of 1200-900 cm^{-1} region has, apparently, an improved predictive ability when compared with UPLC-MS and ^{1}H NMR PLS-DA models of controls *vs pre*-PTD. However, the true class models of MIR data were characterized by a classification rate of 75%. Once more, a low value that may be increased perhaps using larger sample numbers.

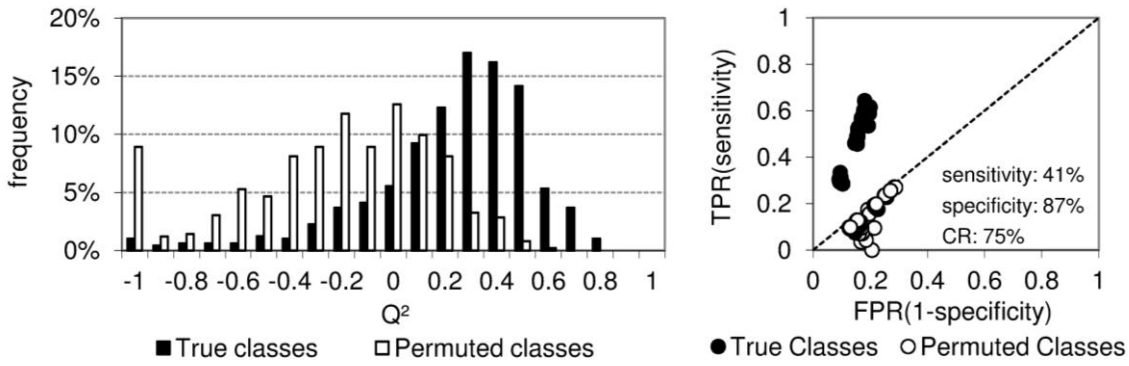


Figure 6.8 - MCCV results for the MIR PLS-DA model of control (n=40) *vs pre*-PTD (n=14) 2nd derivative 1200-900 cm^{-1} region: Q^2 distributions (left) and ROC plot (right) of true and permuted class models. The average sensitivity, specificity and classification rate (CR) for the true class models are shown in the ROC plot.

The loadings of the LV1 of 2nd derivative 1200-900 cm^{-1} PLS-DA model presented in Figure 6.7b indicate the variation the bands 1039 and 1122 cm^{-1}. Considering the 2nd transformation of the data set, the negative LV1 direction indicates a decrease of these two bands in *pre*-PTD relative control. Both 1122 cm^{-1} and 1039 cm^{-1} bands, which were not found in previous MIR studies involving biofluids, occur in the region of C-C and C-O stretching vibrations (Pretsch *et al.*, 1989). This region of 1200-900 cm^{-1} has been used for the quantification of lactate and glucose in AF in a previous MIR study of preterm infection, fetal distress, and fetal lung maturity (Liu and Mantsch, 1999). However, several compounds such as organic and amino acids, sugars and proteins exhibit vibrations absorbing in the same region.

6.1.4. STOCSY and SHY analysis of *pre*-PTD cases

6.1.4.1. UPLC-MS/^{1}H NMR SHY analysis

SHY correlation analysis was performed by correlating the MS relevant features obtained from the UPLC-MS PLS-DA control *vs pre*-PTD models and the full CPMG ^{1}H NMR of the *pre*-PTD AF samples as described in chapter 3 (section 3.7, page 83). The significant correlations obtained are summarized in Table 6.4.

Table 6.4 - List of UPLC-MS/^{1}H NMR SHY correlations between the AF MS features (shown in Table 6.3) and AF ^{1}H NMR spectra of *pre*-PTD. Only maximum or minimum correlations values $|r| \geq 0.70$ and ($p<0.001$) were considered. Assigned spin-systems are indicated in bold. *variation relative to control group.

metabolite	m/z	direction of variation*	RT (min)	^{1}H NMR resonances correlated positively to MS feature / ppm	^{1}H NMR resonances correlated negatively to MS feature / ppm
histidine	309.98	↓	0.8	unknown: 6.67	unknown: 2.19
methionine	150.05	↓	1.0	-	**glucose:** 3.23(dd), 5.23(d) **histidine:** 7.82(s)
phenylalanine	167.08	↓	2.9	-	unknown: 3.18, 5.67

No UPLC-MS/^{1}H NMR positive correlations were found between the MS detected metabolites and their respective ^{1}H NMR resonances (auto-correlations) as it should be expected. This may possibly justified, at least partially, by the low number of samples used (n=11) which is a known issue in the quality of the correlation calculation (Camacho *et al.*, 2005). However, negative correlations were registered between methionine MS feature (decreased in *pre*-PTD) and the glucose spin-system, indicating their concomitant opposite variations, which seems to be in

agreement with their variation identified in ^{1}H NMR. An inverse variation was also observed between methionine MS feature and a signal from histidine. It is interesting to note that despite the decrease in histidine and methionine in *pre*-PTD relative to control group, apparently these two metabolites vary in opposite directions within the *pre*-PTD group (Table 6.4), which could be interpreted as specific metabolic/physiologic relationship in connection with PTD. Some unassigned NMR resonances were also found to correlate with MS features (Table 6.4).

6.1.4.2. MIR STOCSY analysis

MIR intra-correlation analysis was performed using the *pre*-PTD spectra, in order to further explore the origin of the relevant LV1 loadings bands from the 2nd derivative 1200-900 cm^{-1} region PLS-DA model (1039 and 1122 cm^{-1}). The 1750-900 cm^{-1} expansion of the intra-correlation STOCSY map obtained for *pre*-PTD MIR spectra is shown in Figure 6.9, where the vertical lines mark the positions of the bands at 1039 and 1122 cm^{-1}.

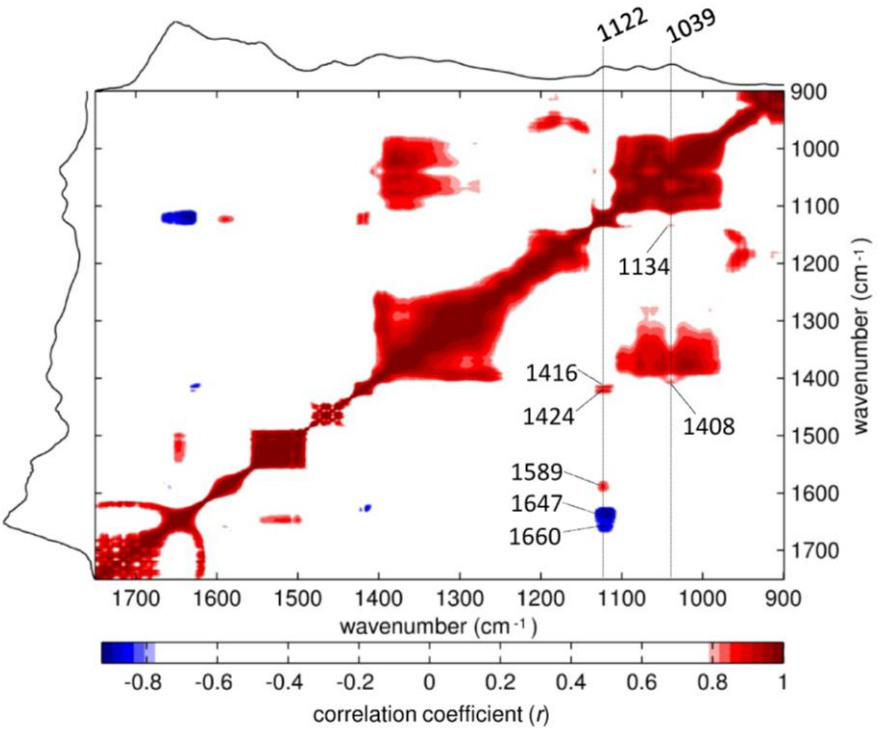


Figure 6.9 - MIR spectra (1750-900 cm^{-1}) intra-correlation STOCSY map for the *pre*-PTD group (n=14), with correlation threshold $|r|>0.8$ and $p<0.001$. Dashed vertical lines indicate the bands of interest identified through loadings analysis and their correlations across the 1750-900 cm^{-1} spectral region.

The correlation map shows firstly that both bands 1122 cm^{-1} and 1039 cm^{-1} are not correlated, meaning they should correspond to distinct metabolites. The significant correlations found along the vertical lines (Figure 6.9) are listed in Table 6.5, together with their possible assignment.

Table 6.5 - List of MIR intra-correlations and from STOCSY analysis of *pre*-PTD spectra.

MIR bands (cm^{-1})	change *vs* controls	MIR/MIR correlations (cm^{-1})	
		positive correlations	negative correlations
1122 (νC-C/C-O)	↓	1589 ($\nu_{as}COO^-$); 1424, 1416 (ν_sCOO^-);	1647, 1660 (Amide I)
1039 (νC-C/C-O)	↓	1408 (ν_sCOO^-); 1134 (νC-C/C-O)	-

In this way the band at 1122 cm^{-1} was found positively correlated to the asymmetric (1589 cm^{-1}) and symmetric (1424, 1416 cm^{-1}) stretching bands of carboxylate groups suggesting a possible connection to carboxylate containing metabolites such as organic and amino acids (Pretsch *et al.*, 1989; Liu and Mantsch, 1999) (decreasing in *pre*-PTD) (Table 6.4). Interestingly the same 1122 cm^{-1} band was found anti-correlated to the bands at 1647 and 1660 cm^{-1} both belonging to the protein Amide I region (Pretsch *et al.*, 1989; Liu and Mantsch, 1999) (Figure 6.9 and Table 6.5), thus suggesting an inverse relation with the protein content, which seems to be in agreement with the ^{1}H NMR results which indicated an increase in glycoprotein signals (Table 6.1, page 165). In addition, the band at 1039 cm^{-1} correlated only positively with the band at 1408 cm^{-1}, (symmetric stretching of carboxylate group) and also positively with the 1134 cm^{-1} (C-C/C-O stretching vibration) (Pretsch *et al.*, 1989), thus indicating possible relations to organic acids found decreasing in *pre*-PTD.

6.1.4.3. MIR/^{1}H NMR SHY analysis

Further analysis was pursued using MIR/^{1}H NMR SHY analysis. A MIR/^{1}H NMR SHY map was calculated by correlating the MIR 1200-900 cm^{-1} spectral region to the full range of standard ^{1}H NMR spectra (0-10 ppm) for *pre*-PTD samples (Figure 6.10). All the significant correlations identified in SHY analysis are listed in Table 6.6.

The 1039 cm^{-1} band correlated positively with amino acids resonances namely of alanine, isoleucine, leucine, histidine, lysine, phenylalanine, proline, threonine, tyrosine, valine, and also with α-oxoisovalerate, citrate and glucose (Figure 6.10, Table 6.6). In addition, negative correlations with resonances in the regions of the ^{1}H NMR spectrum corresponding to CH_3 groups (Figure 6.10) were observed. On the other hand no significant correlation was registered for band 1122 cm^{-1} (Figure 6.10), meaning that this band may be related to compounds present in concentrations below the NMR detection limit but which can be detected by MIR, or to compounds with no ^{1}H NMR visible protons or absent of protons, such as an inorganic ion.

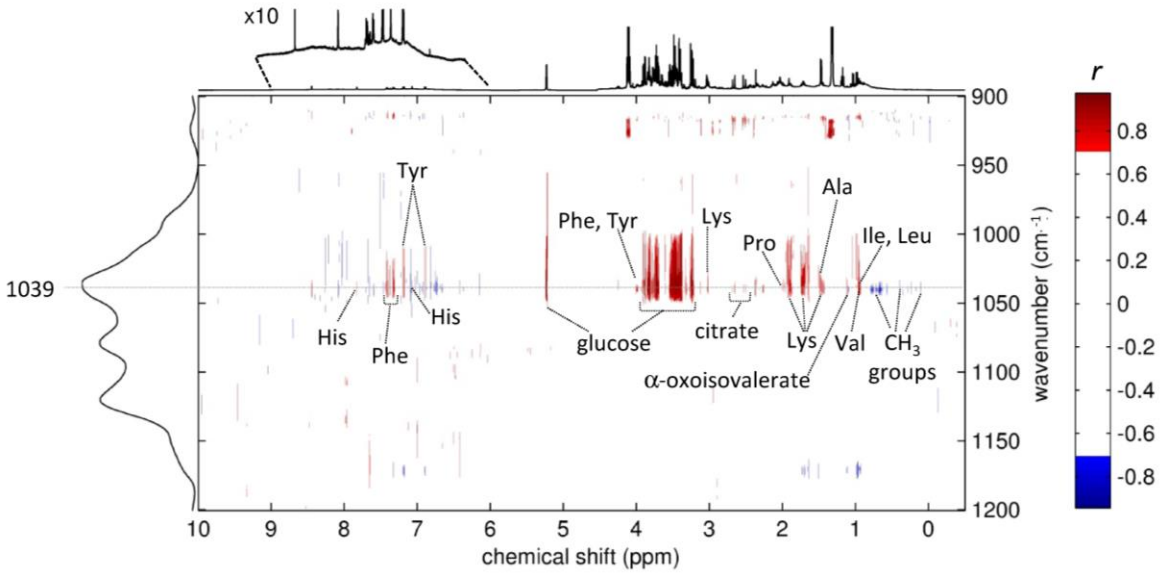


Figure 6.10 - Statistical heterospectroscopy (SHY) maps of MIR (1200-900 cm^{-1}) / ^{1}H-NMR (0-10 ppm) with correlation threshold of $|r|>0.7$ and $p<0.001$ for the *pre*-PTD group (n=14). Dashed horizontal lines indicate the bands of interest identified through loadings analysis and their correlations across the NMR 0-10 ppm region.

Table 6.6 - List of MIR/^{1}H NMR SHY correlations obtained between MIR 1200-900 cm^{-1} spectral region and the standard ^{1}H-NMR (0-10 ppm) spectra of *pre*-PTD.

MIR bands (cm^{-1})	change *vs* controls	Positively correlated ^{1}H-NMR signals	Negatively correlated ^{1}H-NMR signals
1039 (νC-C/C-O)	↓	α-oxoisovalerate, citrate, glucose, Ala, Ile, Leu, His, Lys, Phe, Pro, Thr, Tyr, Val	CH_3 groups (in 0.08-0.79 ppm range)

The positive correlation between the amino acids, citrate and α-oxoisovalerate resonances and 1039 cm^{-1} band in *pre*-PTD (found decreasing when compared to controls) seems to be in agreement with the decreases of those metabolites in pre-PTD group when compared to controls, as observed previously by ^{1}H NMR and UPLC-MS. Glucose, found positively correlated with 1039 cm^{-1} band, seems to vary in a different direction in *pre*-PTD (a decrease) when compared with the its increase relative to controls detected by ^{1}H NMR and UPLC-MS. This apparent contradiction may be justified by the fact that the most intense glucose bands occur in the region of 1065-1000 cm^{-1}. Therefore, the positive correlation observed at 1039 cm^{-1} may simply result the positive correlation of glucose signals of both MIR and ^{1}H NMR spectral domains. However, there is also the possibility that the 1039 cm^{-1} band results from a different compound, whose vibrations overlap with those of glucose, or that this compound contains no ^{1}H NMR visible protons or is absent of protons, such as an inorganic ion.

6.1.5. Proposed interpretation of the metabolite variations in *pre*- PTD cases

The multivariate analysis of ^{1}H NMR, UPLC-MS and MIR data sets from control and *pre*-PTD 2nd trimester AF samples have led to PLS-DA models generally characterized by low predictive ability, which could be in principle improved by the use of a larger number of *pre*-PTD samples. Nevertheless, the careful analysis of such models provided a metabolite signature of possible early markers of PTD. This includes a small decrease in the amino acids alanine, isoleucine, leucine, glutamine, histidine, lysine, methionine, phenylalanine, proline, tyrosine and valine a well as a decrease in formate, α-oxoisovalerate and citrate in *pre*-PTD cases. These variations were also accompanied by small increases in glucose, acetoacetate, β-hydroxybutyrate, allantoin, choline, serine, threonine and proteins (including glycoproteins). Regarding the possible origin of the decreases of these amino acids, decreases in isoleucine have been reported in 2nd trimester AF of *pre*-PTD subjects (Ramón Y Cajal *et al.*, 2007). In addition, significant fluxes of methionine and phenylalanine were observed after nutrient administration in PTD newborns when compared with term newborns (Thomas *et al.*, 2008). Also glutamine, a non-essential amino acid, can become conditionally essential in preterm low-birth-weight newborns after administration of amino acids up to 90 hours after birth (Kadrofske *et al.*, 2006). The decreases in the amino acids alanine, isoleucine, leucine, glutamine, histidine, lysine, methionine, phenylalanine, proline, tyrosine and valine could then indicate an early perturbation of the fetal-placental unit which plays a key role in amino acid transfer (Cetin, 2001). On the other hand, the increase noted in the essential amino acid threonine is not easily justifiable, especially in a context of decreases in most amino acids as observed previously. One possible explanation could be an increased need of threonine for protein synthesis as suggested by the increased incorporation of threonine in enteric tissues observed in preterm newborns after feeding (van der Schoor *et al.*, 2007).

Allantoin can be regarded as a non-enzymatic oxidation product of uric acid in humans (Shestopalov *et al.*, 2006; Tolun *et al.*, 2010). Therefore, the increase in allantoin noted in *pre*-PTD group suggests the possibility of occurrence of oxidative stress mediated by reactive oxidation species (ROS) (Wall *et al.*, 2002). In fact, increased allantoin levels have been found in plasma and bronchoalveolar lavage fluid of preterm newborns who developed chronic lung disease when compared with healthy preterm newborns (Ogihara *et al.*, 1998).

The increase in glucose for the *pre*-PTD group may indicate hyperglycaemia, a condition that has been found in very low birth weight infants, which in turn is a condition associated with prematurity and other risk factors (Beardsall *et al.*, 2010). Also the increases in ketone bodies β-hydroxybutyrate and acetoacetate noted in *pre*-PTD (by ^{1}H NMR) can indicate deficient glucose

utilization and the need for alternative energetic substrates, also compatible with hyperglycaemic states such as does occurring in diabetes (Murray *et al.*, 2003). However, no clinical indications of hyperglycaemia or diabetes have been found in the newborns corresponding to the *pre*-PTD cases studied, which complicates the interpretation of this finding.

Choline, which was found increased in *pre*-PTD AF samples, was also found increased in 2nd trimester maternal urine samples (Diaz *et al.*, 2011), thus suggesting accumulation of choline in both intra and extra-placental environments. Choline is metabolically connected to methionine (decreased in *pre*-PTD) and folate metabolism (Zeisel, 2006). Formate, decreased in *pre*-PTD, is in turn related to folate metabolism through formyltetrahydrofolate (Murray *et al.*, 2003). Therefore, these variations in *pre*-PTD seem to suggest early perturbations of these metabolic pathways. There is also a metabolic connection between serine (increased in *pre*-PTD) and folate metabolism (Murray *et al.*, 2003), thus establishing a possible link between the variation of serine and that of choline and formate.

Finally, the increase in N-acetyl signals (seen by ^{1}H NMR and by MIR intra-correlations) may possibly indicate high glycoprotein content in the *pre*-PTD group. In fact, the most abundant AF glycoprotein, transferrin, has been related to PTD in a previous report in relation to oligohydramnios and the occurrence of maternal anaemia, both conditions also associated with the occurrence of PTD (Gao *et al.*, 2008). Another study regarding proteomic analysis of 3rd trimester AF found high levels of other proteins including another AF glycoprotein, α-1-microglobulin, in relation to PTD in the absence of intra-amniotic infection and also transferrin in relation to PTD with intra-amniotic infection (Bujold *et al.*, 2008). However, identification of the glycoproteins underlying the variations observed in this study would require the use of proteomics methodologies, which was outside the scope of the present work.

6.2. Analysis of 2nd trimester amniotic fluid composition for the prediction PE, IUGR and GDM (*pre*-PE, *pre*-IUGR groups, *pre*-diagnostic GDM)

The following sections describe the results from the analysis of 2nd trimester AF metabolite ^{1}H NMR and MIR profiles of preeclampsia (*pre*-PE), intrauterine growth restriction (*pre*-IUGR) and *pre*-diagnostic GDM cases when compared to a control group. UPLC-MS results were only obtained for *pre*-diagnostic GDM.

In order to reduce the imbalance between *pre*-PE (n=9), *pre*-IUGR (n=8) and control groups in MVA of ^{1}H NMR profiling data, a smaller group of control samples (n=40) was used (the same

group of control samples used for MIR analysis). Due to the reduced numbers of *pre*-PE and *pre*-IUGR cases the respective profiling studies were still largely exploratory.

6.2.1. ^{1}H NMR and MIR profiling for comparison of control and *pre*- PE groups

The ^{1}H NMR spectra collected for comparison of control (n=40) and *pre*-PE (n=9) cases were analysed using PLS-DA, since no separations or trends were observed in PCA analysis. Separation tendencies were observed in the PLS-DA scores scatter plots for standard spectra (Figure 6.11a) and a complete separation was obtained for CPMG spectral data (Figure 6.11b). The diffusion-edited spectra originated a PLS-DA model with Q^2<0.2 (obtained from 7-fold CV), too low to be considered for further analysis (not shown).

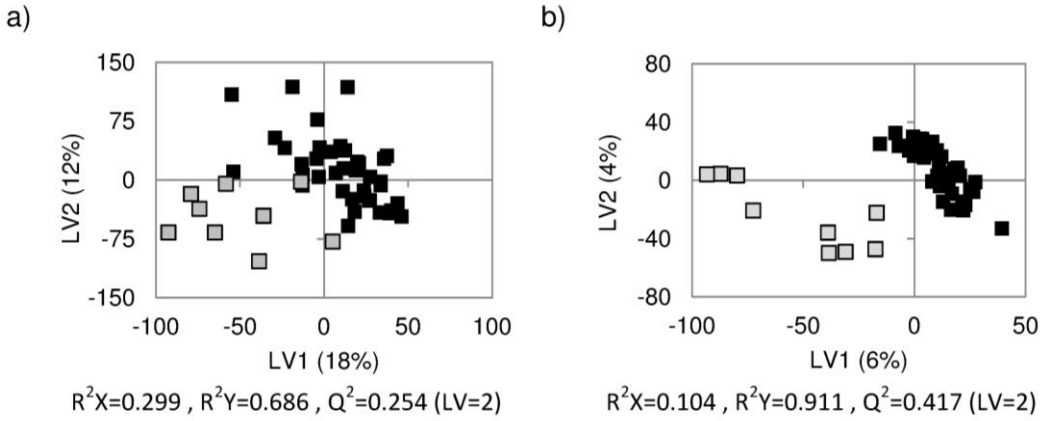


Figure 6.11 - Scores scatter plots from PLS-DA models of ^{1}H NMR spectra 2nd trimester AF of control (■, n=40) and *pre*-PE (□, n=9) groups: a) standard spectra, b) CPMG spectra.

The MCCV results obtained for standard and CPMG models are shown in Figure 6.12. Both standard and CPMG data sets showed largely overlapped Q^2 distributions of true and permuted class models, although with a small tendency for the separation of both distributions (Figure 6.12a,b, left). However, the remaining MCCV parameters obtained for standard (85% classification rate, 0-40% sensitivity, average 95% specificity) (Figure 6.12a, right) and CPMG (89% classification rate, average 40% sensitivity and 100% specificity) (Figure 6.12b, right), demonstrate that the latter shows better performance, despite the still low predictive ability registered. In order to explore any possible variations occurring in *pre*-PE group, the PLS-DA loadings of the CPMG model were considered (Figure 6.13).

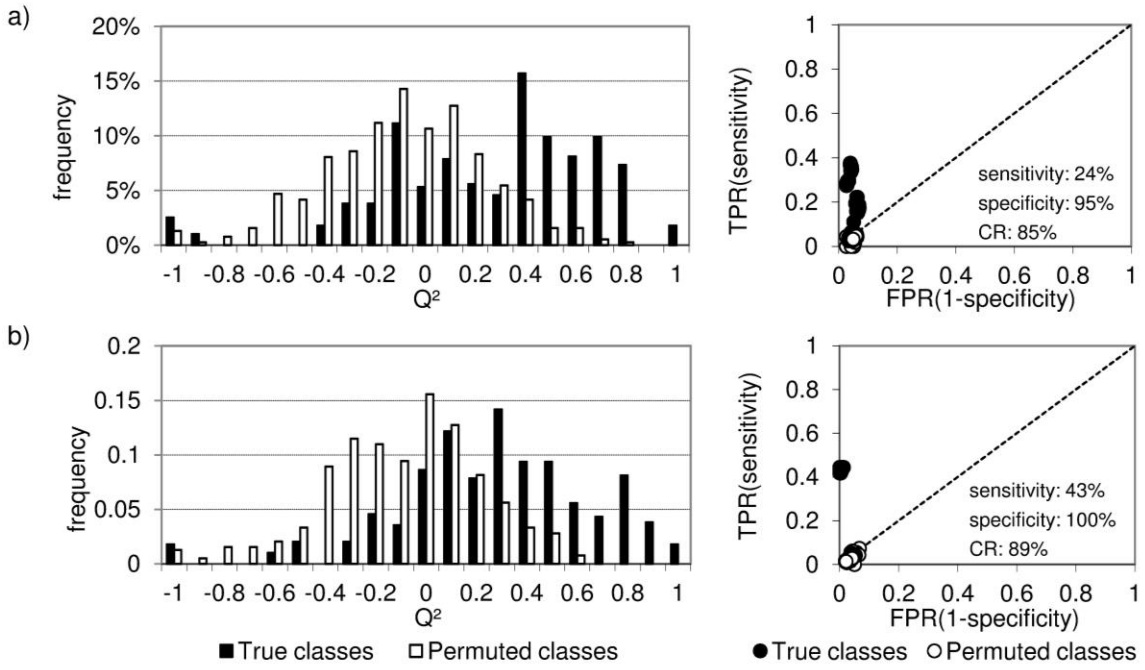


Figure 6.12 - MCCV results for PLS-DA models for a) standard and b) CPMG ^{1}H NMR AF spectra of control (n=40) and *pre*-PE (n=9) groups. Q^2 distributions (left) and ROC plots (right) of true classes and permuted classes models are shown. The average sensitivity, specificity and classification rate (CR) for the true class models are shown in the ROC plots.

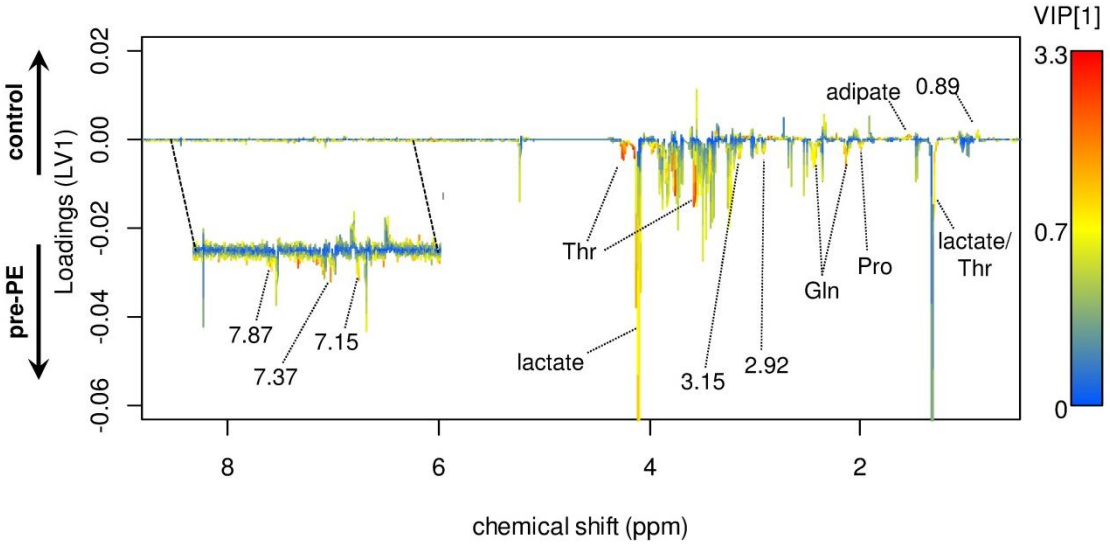


Figure 6.13 - Loadings from the PLS-DA models of CPMG spectra from control *vs pre*-PE colour-coded with VIP values. Amino acids in three-letter code; unassigned signals are indicated with their respective chemical shift values.

The resonances of lactate, glutamine, threonine, proline, unassigned spin-system 2.92 and 3.15 ppm, and unassigned resonances 7.15 and 7.87 ppm were found increased in *pre*-PE cases (Figure 6.13). Conversely, the resonances of adipate and unassigned resonance 0.89 ppm were

found decreasing in *pre*-PE cases compared with controls (Figure 6.13). Peak integration and univariate analysis confirmed the statistical significance of such variations (p<0.05) (Table 6.8). The significant variations detected comprised those of adipate (decreasing in *pre*-PE), glutamine, proline and threonine, all increased in *pre*-PE (Table 6.8).

Table 6.8 - List of significant metabolite variations (p<0.05) noted in 2nd trimester human AF samples of *pre*-PE relative to control group. [a], signal within the metabolite spin-system chosen for integration; m, multiplet.

metabolite	δ^1H [a]	direction and magnitude of change		*p*-value
adipate	1.54 (m)	↓	1.3	1.24×10^{-4}
glutamine	2.45 (m)	↑	1.2	3.61×10^{-2}
proline	2.01 (m)	↑	1.1	4.13×10^{-3}
threonine	4.26 (m)	↑	1.2	2.23×10^{-2}
Unassigned				
2.99 ppm	2.99 (m)	↑	1.6	4.70×10^{-2}
3.78 ppm	3.78	↑	1.2	5.97×10^{-3}
3.77 ppm	3.77	↑	1.1	7.13×10^{-3}
3.97 ppm	3.97	↑	1.2	2.04×10^{-2}
3.98 ppm	3.98	↑	1.2	1.63×10^{-2}
7.37 ppm	7.37	↑	1.2	1.00×10^{-2}
7.87 ppm	7.87	↑	2.5	7.78×10^{-3}

The variations detected in 2nd trimester AF in the present study were not consistent with those obtained in a previous ^{1}H NMR study, which compared 3rd trimester AF samples from PE affected cases (n=7) and controls (n=36) using univariate analysis of a few integrated resonances (Bock, 1994). In the latter, significant decreases in the contents of acetate, choline, succinate and possibly citrate were found in the PE affected group (Bock, 1994). Bearing in mind that only a small number of PE cases were used (n=7), as in the present study (n=9), both studies must be considered as largely exploratory, thus requiring further improvement with the use of increased sample numbers.

Metabolite variations were observed in connection to *pre*-PE found in 2nd trimester AF in other studies such as the decreases in palmitic acid (Witkin *et al.*, 2012) and increases in oxidative stress marker F_2-isoprostane (Wang *et al.*, 2011a). These metabolites are usually present in sub-μM concentrations and, therefore, the absence of variation of such molecules ^{1}H NMR spectra is not surprising. Protein changes have also been observed in 2nd trimester AF in connection with the later occurrence of PE, such as the elevated levels of matrix metalloproteinase-2 and respective inhibitors (Lavee *et al.*, 2009), elevated levels of oxidized transthyretin (Vascotto *et al.*, 2007) and

the over-expression of proteins albumin and apolipoprotein A-1 (Oh *et al.*, 2012). However, protein identification using ^{1}H NMR is very limited, as previously discussed in chapter 4 (section 4.3.2, pages 108-110).

No associations were found between the decrease noted in adipate and PE, which leads to no conclusions regarding its variation. Regarding the amino acids variations, increases in maternal serum and cord blood amino acid concentrations have been reported for PE cases, from which glutamine, proline and threonine were significantly increased in maternal blood, and also proline in cord blood when compared with controls (Evans *et al.*, 2003). Other amino acids variations were found in maternal blood in PE cases, such as a reduced glutamine (Hsu *et al.*, 2005), increased only in phenylalanine (Glew *et al.*, 2004), or only increases in histidine (Turner *et al.*, 2008). The different results obtained in those studies seem to reflect the different sampling procedures used, which makes it difficult to correlate with variations observed here in AF. Despite the many alterations observed in AF, maternal and cord blood, a more complete understanding of the biochemical interrelationship between maternal and fetal side in PE would benefit from improved AF analysis. In order achieve this goal, improved ^{1}H NMR based models with larger sample sizes should be considered as a further approach in the study of PE as a complement to other approaches using maternal plasma/serum and urine.

Regarding the MIR data, no apparent differences were found by visual comparison between MIR spectra representative of *pre*-PE (n=9) and controls (n=40). Also a complete overlap between both groups was obtained in PCA scores plots of the 3000-2800 cm^{-1}, 1800-1200 cm^{-1}, 1200-900 cm^{-1} spectral regions and 2nd derivative of 1200-900 cm^{-1} region (not shown). However, apparent separations were observed in the PLS-DA scores plot of 2nd derivative transformed 1200-900 cm^{-1} region, although with a very low Q^2 value (<0.2) (Figure 6.14).

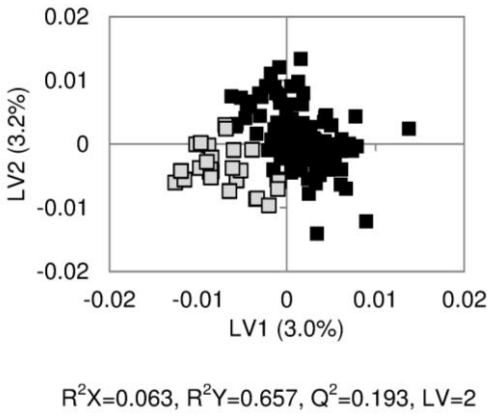


Figure 6.14 - Scores scatter plot of MIR PLS-DA model of the 2nd derivative of 1200-900 cm^{-1} region for control (■, n=40) *vs pre*-PE (□, n=9).

Despite the apparent tendency noted, MCCV results showed model with complete overlap of true and permuted class models in Q^2 distributions and ROC plots (not shown), meaning no distinctions were found between true and permuted models. Therefore, the models were non-predictive.

The absence of predictive ability of MIR models is not surprising, given the poor predictive ability obtained in the respective ^{1}H NMR models. Because the MIR spectrum is dominated by the most abundant AF components, namely glucose, lactate, proteins and some amino acids, these metabolites may mask the variation of less abundant metabolites, as noted in the previous chapter. Another important issue is the reduced number of samples composing the *pre*-PE group (n=9). Therefore, increasing sample size should, in principle, increase the models predictive ability of MIR analysis.

6.2.2. ^{1}H NMR and MIR profiling of *pre*- IUGR cases

The PLS-DA scores of ^{1}H NMR spectra of controls and *pre*-IUGR cases, representing the best results in terms of class separation are shown in Figure 6.15.

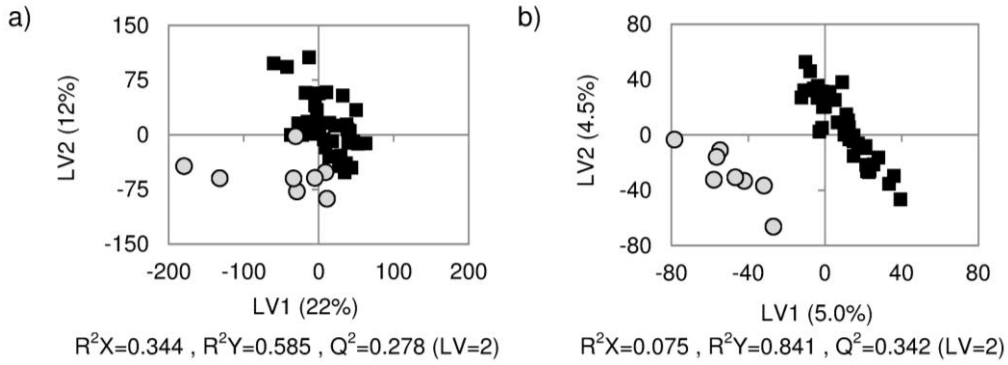


Figure 6.15 - Scores scatter plots from PLS-DA of ^{1}H NMR spectra 2nd trimester AF of *pre*-IUGR (○, n=8) and healthy controls (■, n=40): a) standard spectra, b) CPMG spectra.

Indeed, a complete separation of *pre*-IUGR and control groups was obtained for PLS-DA model of CPMG spectra (Figure 6.15b), whereas only a separation tendency was obtained for the PLS-DA model of standard spectra (Figure 6.15a). The PLS-DA model obtained from diffusion-edited spectra was associated with a low Q^2value (<0.2) and, therefore, not considered for further analysis. The MCCV analysis performed for the standard and CPMG PLS-DA models revealed also models associated with very low predictive ability. In fact, both standard and CPMG PLS-DA models showed completely overlapped Q^2 distributions, as well as sensitivities below 20%, resulting in very low true positive rates. Figure 6.16 illustrates the MCCV results obtained for the

CPMG model. Due to the lack of predictive ability, the models loadings were not considered at this stage.

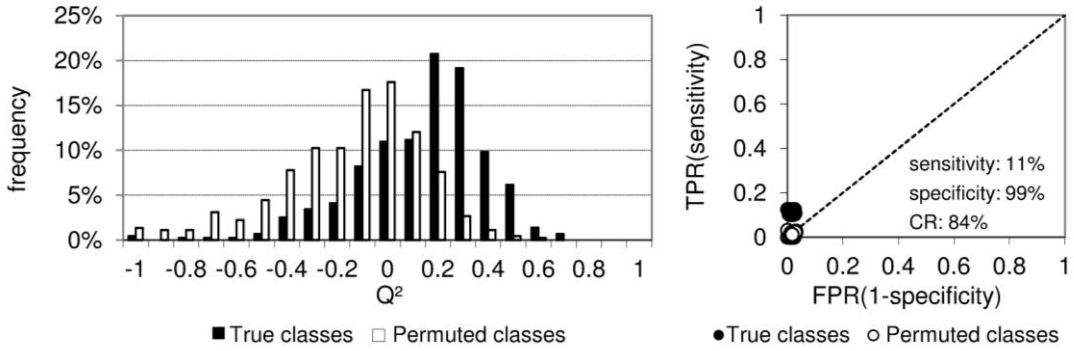


Figure 6.16 - MCCV results for PLS-DA models for CPMG ^{1}H NMR AF spectra of controls (n=40) and *pre*-IUGR (n=8). Q^2 distributions (left) and ROC plots (right) of true classes and permuted classes models are shown. The average sensitivity, specificity and classification rate (CR) for the true class models are shown in the ROC plot.

The PLS-DA analysis of MIR spectra from control (n=40) and *pre*-IUGR (n=8) groups showed an apparent separation, only observed in the 2nd derivative transformed 1200-900 cm^{-1} data set (Figure 6.17). However, the MCCV analysis of the PLS-DA model revealed a complete overlap of true and permuted class models in Q^2 distributions and ROC plots (not shown) hence confirming the absence of predictive ability.

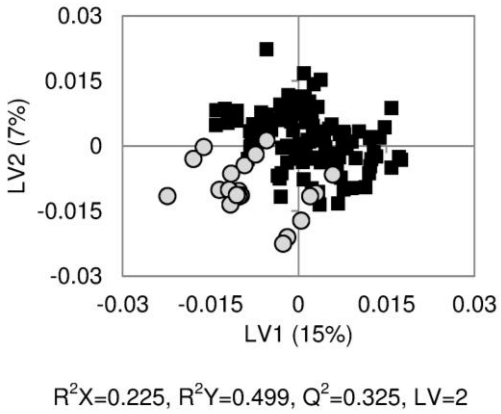


Figure 6.17 - Scores scatter plot of MIR PLS-DA models of the 2nd derivative of 1200-900 cm^{-1} region of control (■, n=40) *vs pre*-IUGR (○, n=8).

Once more it is important to recall the reduced number of *pre*-IUGR cases, which may be regarded as the main cause of low performance of the ^{1}H NMR and MIR PLS-DA models obtained. Nevertheless, several metabolite variations noted in maternal and fetal blood plasma were found in connection with IUGR (Cetin *et al.*, 1996; Marconi and Paolini, 2008; Favretto *et al.*, 2012). Also, AF metabolite variations were found in *pre*-IUGR, namely increases in the levels of isoprostanes

(present in nanomolar concentrations) (Longini *et al.*, 2005), as well as changes in catecholamines and respective metabolites (Divers *et al.*, 1981); and also in 3rd trimester AF such increased levels of phosphatdylglycerol (Gross *et al.*, 1982); and increases in nitric oxide (Di Iorio *et al.*, 1997). All these metabolites are present in AF in low concentration (not being detected by ^{1}H NMR and MIR), therefore further research using enlarged sample sizes and possibly more sensitive analytical methods, namely UPLC-MS, should be employed to improve the predictive ability of AF models in relation to IUGR.

6.2.3. Analysis of 2nd trimester amniotic fluid composition for the prediction of GDM (*pre*-diagnostic GDM group)

Profiling studies were performed on 2nd trimester AF of *pre*-diagnostic GDM, compared with control group, by ^{1}H NMR (*pre*-diagnostic GDM, n=35, control, n=90), UPLC-MS (*pre*-diagnostic GDM, n=23, control, n=26) and MIR (*pre*-diagnostic GDM, n=33, control, n=40). The PLS-DA analysis of the ^{1}H NMR, UPLC-MS and MIR data sets resulted in models with separation tendencies between the *pre*-diagnostic GDM and control groups although characterized by low Q^2<0.3 (not shown). However, the PLS-DA models were not validated by MCCV, showing in general large overlaps of true class and permuted class models in Q^2 and ROC plots and hence were not further explored.

These results are in general agreement with those obtained in an early stage of this work by ^{1}H NMR profiling of 2nd trimester human AF (Graça *et al.*, 2010), which used reduced sample sizes of both control (n=82) and *pre*-diagnostic GDM (n=27) groups with the corresponding PLS-DA model obtained (based on ^{1}H NMR standard spectra) registering Q^2=0.269 (based on 7-fold CV), which however was not evaluated by MCCV. Several significant metabolite variations were then reported including decreases in acetate, creatinine, formate, glycine, proline, serine, taurine, glycerophosphocholine and glutamate, together with a marked increase in glucose content in *pre*-diagnostic GDM cases (Graça *et al.*, 2010). In order to verify the occurrence of such variations in the present data sets, univariate analysis was performed on the assigned integrals of the corresponding ^{1}H NMR signals and MS features (detected by UPLC-MS). The only variation thus confirmed was the one regarding glucose, which was expressed both as an increase in glucose ^{1}H NMR signals, and also as an increase in the intensity of the hexose-related feature detected in UPLC-MS (HSS/ESI$^+$ data set) as 203.05 m/z (sodium+hexose MS adduct) (Figure 6.18).

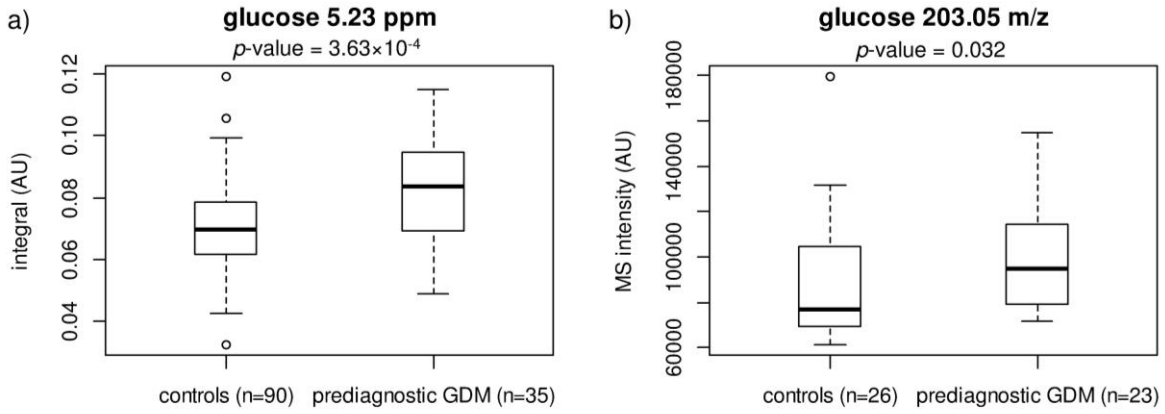


Figure 6.18 - Box-plots illustrating the statistical significant variation of glucose related signals in *pre*-diagnostic GDM relative to control: a) integrals of glucose ^{1}H NMR signal of *pre*-diagnostic GDM (n=35) and controls (n=90); b) hexose-related MS feature intensity of *pre*-diagnostic GDM group (n=23) and controls (n=26).

The decrease of glutamate was also confirmed in UPLC-MS data (p=0.03) (MS feature 146.07 m/z), but not in ^{1}H NMR. The different results obtained for glutamate are certainly due to the differences in sample size used in both analytical platforms (^{1}HNMR: *pre*-diagnostic GDM, n=35 and control, n=90 *vs* UPLC-MS: *pre*-diagnostic GDM, n=23 and control, n=26).

It seems clear that only glucose is increased consistently in 2nd trimester AF of *pre*-diagnostic GDM subjects, as shown by ^{1}H NMR and UPLC-MS analysis. Glucose increase has been noted previously in 2nd trimester AF samples (at around 15 weeks gestation) using larger samples sets of both *pre*-diagnostic GDM (n=52) and control (n=356) (Tisi *et al.*, 2011). Such an increase in glucose may result from increased glucose placental transport in GDM cases, which is then thought to trigger the fetal insulin production in insulin sensitive tissues (Star *et al.*, 1997). This increase in insulin has also been measured in 2nd trimester AF GDM-diagnosed women (Star *et al.*, 1997; Tisi *et al.*, 2011).

Given the pathophysiologic determinants of GDM which are primarily expressed in the maternal side, a more fruitful approach would be to focus on the search for early markers in maternal blood plasma or urine for instance. In fact, recent metabonomics work performed using 2nd trimester blood plasma and urine has led to the detections of a possible set of early markers of the occurrence of GDM (Diaz *et al.*, 2011), therefore opening a new avenue to explore new markers of GDM.

6.3. Analysis of 2nd trimester amniotic fluid composition for the prediction of PROM (*pre*-PROM group)

The following sections describe the results from the analysis of 2nd trimester AF metabolite ^{1}H NMR and MIR profiles of PROM (i.e., premature rupture of the membranes occurring after 37 g.w. prior to the initiation of uterine contractions) (*pre*-PROM group). Due to constraints in instrument availability at the time of analysis, UPLC-MS could not be acquired for *pre*-PROM samples.

6.3.1. ^{1}H NMR profiling for comparison of control and *pre*-PROM groups

Standard, CPMG and diffusion-edited ^{1}H NMR spectra obtained for 2nd trimester AF *pre*-PROM samples (n=45) were compared with spectra of matching control samples (n=90) using MVA. No separation tendencies were observed in the corresponding PCA scores scatter plots, hence the data sets were analysed by PLS-DA. A separation tendency was only obtained in PLS-DA of standard (Figure 6.19) whereas in the scores of PLS-DA of CPMG spectra the groups showed almost complete separation (not shown), however, the CPMG model registered Q^2<0.1, too low to be considered. Diffusion-edited spectra originated a model with no separation between controls and *pre*-PROM in the corresponding scores plots, therefore it was also not considered for further analysis.

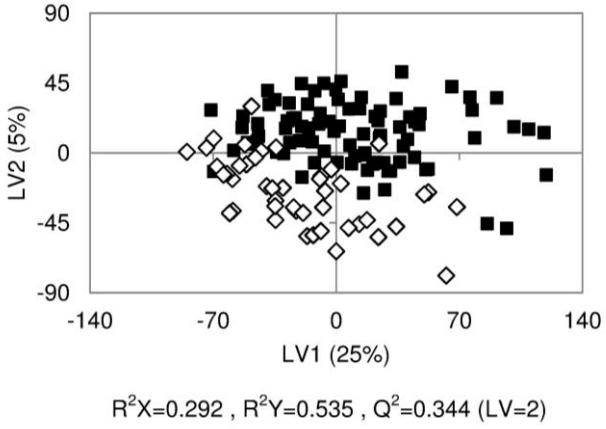


Figure 6.19 - Scores scatter PLS-DA models of ^{1}H NMR standard spectra of 2nd trimester AF of control (■, n=90) and *pre*-PROM (◊, n=45) groups.

The MCCV analysis of standard spectra model from Figure 6.19, showed overlapped Q^2 distributions of true and permuted class models (Figure 6.20, left), and in the ROC plot, the sensitivity and specificity for true class models were in the ranges of 20-55% (average 36%) and 80-90% (average 84%), respectively (Figure 6.20, right), as well as with a classification rate of 67%.

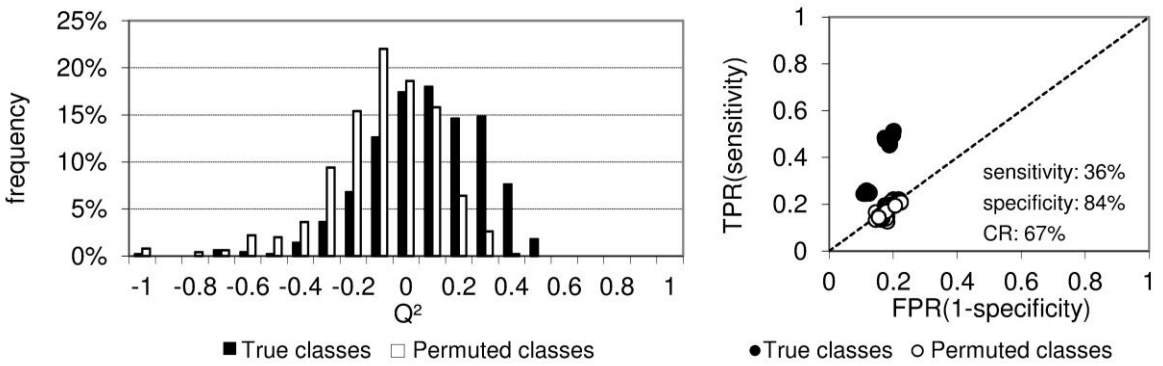


Figure 6.20 - MCCV results for PLS-DA models from standard ^{1}H NMR AF spectra of control (n=90) and *pre*-PROM (n=45) groups. Q^2 distributions (left) and ROC plots (right) of true classes and permuted classes models are shown. The average sensitivity, specificity and classification rate (CR) for the true class models are shown in the ROC plot.

The MCCV indicate that the obtained model has no predictive value, and therefore no loadings analysis was performed. These results are in agreement with those obtained in an early stage of this work where similar groups were considered (*pre*-PROM (n=34) and control (n=82)) and a similar separation tendency was obtained in the PLS-DA scores plots of standard spectra (Q^2=0.276, not validated by MCCV) (Graça *et al.*, 2010). Moreover, significant metabolite variations were validated by univariate analysis in the previous study namely increases in methionine, glutamine and decreases in threonine and in 2 unassigned resonances (not shown).
In order to verify the occurrence of such variations in the present data sets, univariate analysis was performed on the assigned integrals of the corresponding ^{1}H NMR signals. The univariate analysis only confirmed the variation of one of the previously unassigned resonances centred at 3.29 ppm (*p*=0.01), which means that the results obtained are very sensitive to the samples used. Therefore, it is not surprising to obtain ^{1}H NMR models of *pre*-PROM with no predictive ability.

6.3.2. MIR profiling for comparison of control and *pre*-PROM groups

Amniotic fluid samples from *pre*-PROM (n=37) and controls (n=40) were also analysed using MIR spectroscopy. Initially, no apparent spectral differences were observed by visual inspection or by PCA. From the PLS-DA models obtained, only those from sub-spectral region 3000-2800 cm^{-1} and from the 2nd derivative transformed 1200-900 cm^{-1} region, showed some separation tendency (Figure 6.21). Both 3000-2800 cm^{-1} and 2nd derivative transformed 1200-900 cm^{-1} spectral region PLS-DA models were subjected to MCCV, the results being shown in Figure 6.22.

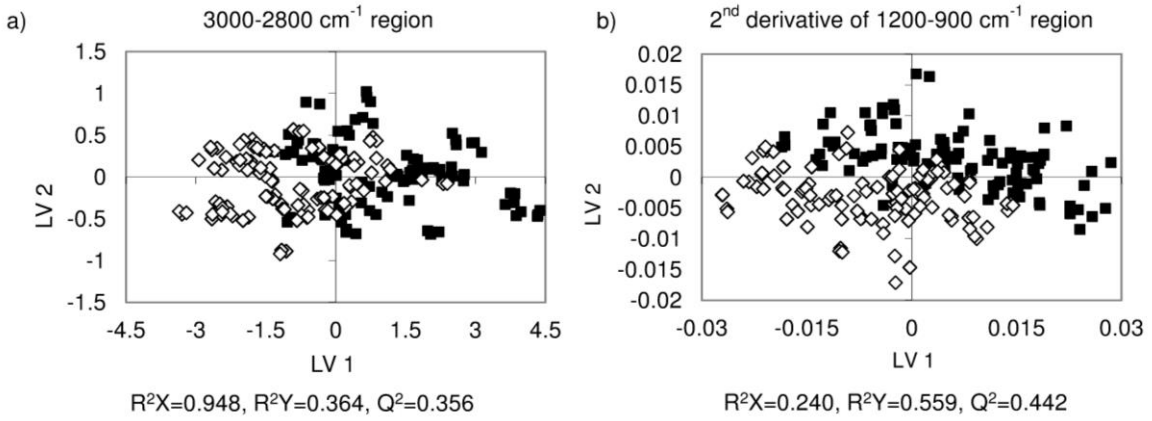


Figure 6.21 - Scores scatter plot of control (■, n=40) *vs pre*-PROM (◇, n=37) MIR PLS-DA models for the spectral regions with 3 technical replicates per sample: a) 3000-2800 cm^{-1}, b) 2nd derivative of 1200-900 cm^{-1} regions.

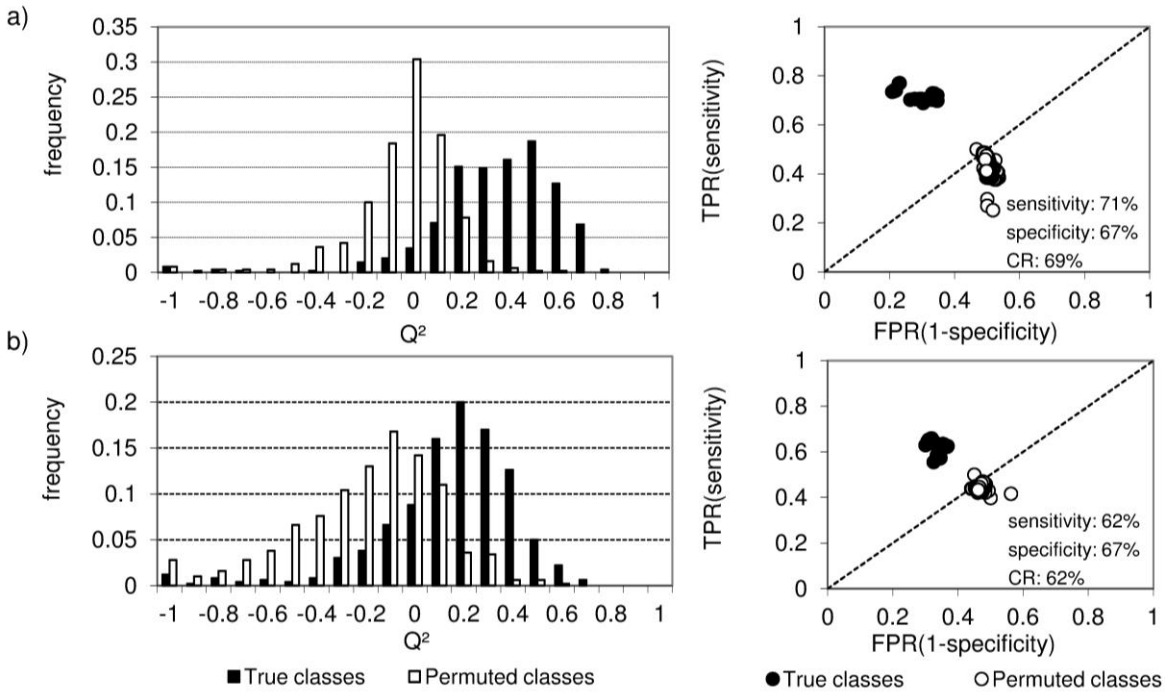


Figure 6.22 - MCCV results of MIR PLS-DA models of control (n=40) *vs pre*-PROM (n=37): a) 3000-2800 cm^{-1} region; b) 2nd derivative of 1200-900 cm^{-1} region. Q^2 distributions (left) and ROC plots (right) of true classes and permuted classes models are shown. The average sensitivity, specificity and classification rate (CR) for the true class models are shown in the ROC plots.

The low predictive ability of the models is expressed by the MCCV results (Figure 6.22), which apart from tendencies of separation noted in Q^2 distributions of true and permuted class models (Figure 6.22, left), register very modest MCCV parameters for both 3000-2800 cm^{-1} region model (69% classification rate, 70% sensitivity, 75% specificity) and for 2nd derivative of 1200-900 cm^{-1} region model (62% classification rate, 60% sensitivity, 70% specificity) (Figure 6.22, right).

The MIR MCCV results show an improvement in predictive ability, expressed by tendencies of Q^2 distributions and in sensitivity (true positive rates) when compared with the ^{1}H NMR model from the previous section. Therefore, the loadings from PLS-DA models were considered for further analysis (Figure 6.23).

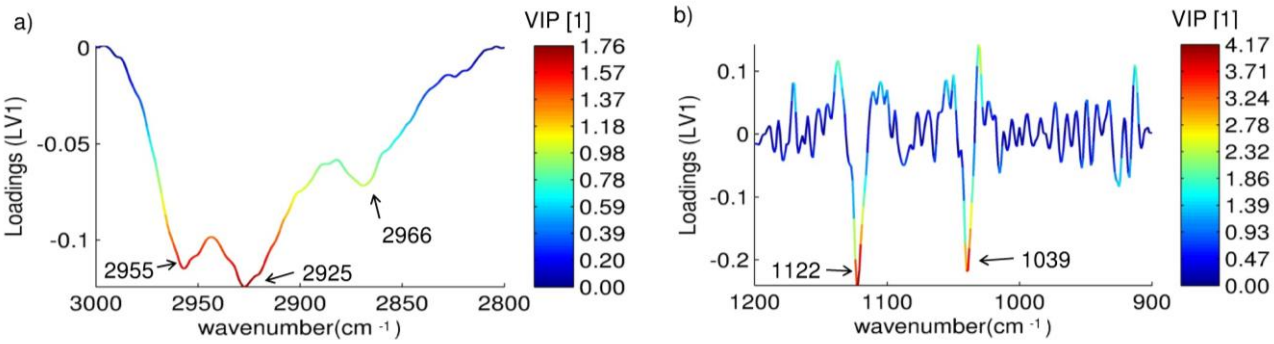


Figure 6.23 - LV1 loadings from PLS-DA model of control (n=40) *vs pre*-PROM (n=37) for the 2nd derivative of 1200-900 cm^{-1} region.

The whole 3000-2800 cm^{-1} region, corresponding to C-H stretching vibrations (Liu and Mantsch, 1999), containing the 3 maxima indicated in Figure 6.23a, seemed to be increased in *pre*-PROM cases, whereas for the 2nd derivative transformed 1200-900 cm^{-1} region model, two bands at 1122 cm^{-1} and at 1039 cm^{-1}, were identified in the LV1 loadings as decreased in *pre*-PROM cases relative to healthy controls (Figure 6.23b). These vibrations can be attributed to a wide range of compounds, namely lactate, proteins, lipids and amino acids, as shown previously in the spectral assignments in Figure 5.13 (chapter 5, section 5.1.5, page 139). Further assignment of such vibrations was attempted with aid of STOCSY analysis, as discussed in the following sections.

6.3.3. Statistical spectroscopy correlation analysis of *pre*-PROM cases

6.3.3.1. MIR STOCSY analysis

STOCSY analysis was performed using all the *pre*-PROM samples (n=37). The resultant STOCSY intra-correlation map obtained is shown in Figure 6.24. The region 3000-2800 cm^{-1} did not add much more information since only a broad correlation was found along the 3000-2300 cm^{-1} range, and no significant correlations along the remaining spectrum were found (Figure 6.24a). On the other hand, the bands at 1122 and 1039 cm^{-1} from the LV1 loadings of 2nd derivative PLS-DA model showed correlations only in the region of 1750-900 cm^{-1} (Figure 6.24a, dashed box) shown

in detail in Figure 6.24b. The significant correlations were searched along vertical dimension of the intra-correlation map 1750-900 cm^{-1} expansion (Figure 6.24b, Table 6.9).

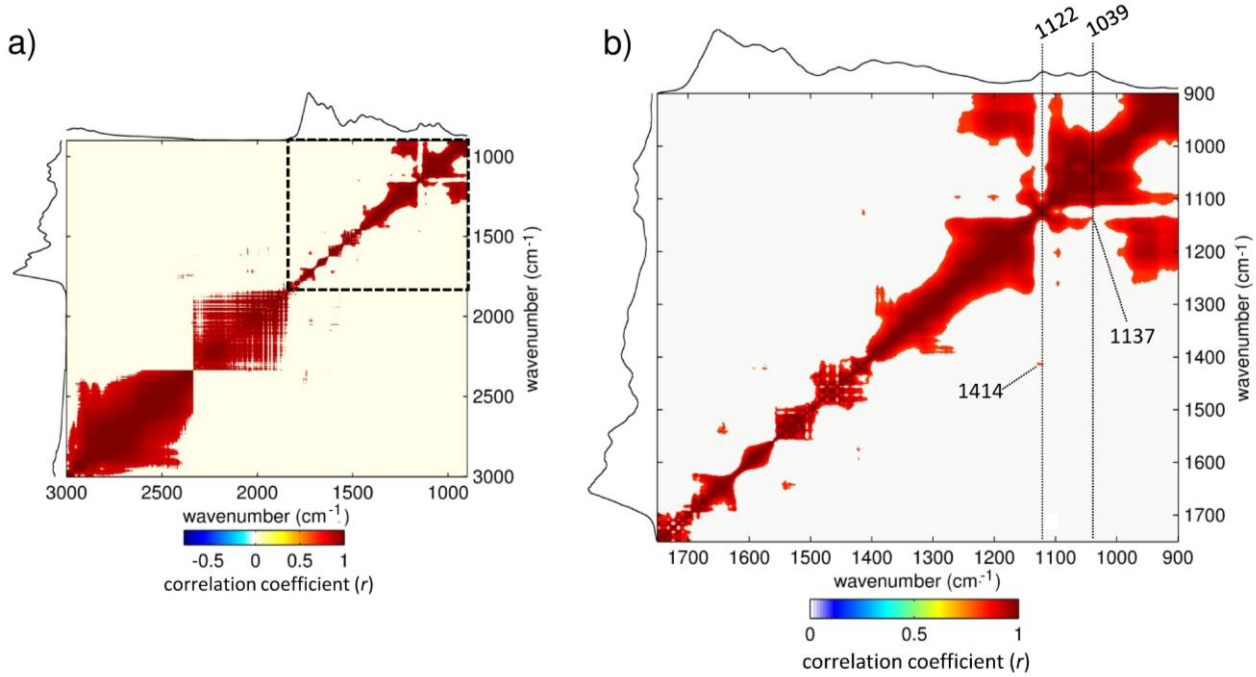


Figure 6.24 - MIR STOCSY map for the *pre*-PROM group (n=37): a) region 3000-900 cm^{-1}; b) 1750-900 cm^{-1} spectral expansion. Correlation threshold of $|r|>0.8$ and $p<0.001$ was used. Dashed vertical lines indicate the bands of interest identified through loadings analysis and their correlations across the 1750-900 cm^{-1} spectral region.

Table 6.9 - Correlations found in the MIR intra-correlations correlation maps obtained using *pre*-PROM spectra (Figure 6.24). The main MIR bands affected in *pre*-PROM by comparison with control group, with respective direction of change are also indicated. Assignments are given in brackets.

MIR bands (cm^{-1})	change relative to control	MIR / MIR positive correlations (cm^{-1})
1122 (νC-C/C-O)	↓	1414 ($\nu_s COO^-$)
1039 (νC-C/C-O)	↓	1137 (νC-C/C-O)

No direct correlation between the two bands at 1122 cm^{-1} and at 1039 cm^{-1} was found, indicating the bands have originated from different molecules. Positive correlations only were found between 1122 cm^{-1} and 1414 cm^{-1}; and for 1039 cm^{-1} and 1137 cm^{-1} (Figure 6.24b, Table 6.9), which are also different from the correlations obtained for *pre*-PTD (considering that the bands 1122 cm^{-1} and 1039 cm^{-1} were found decreasing in both conditions). Therefore, the correlations indicate the involvement of different metabolites in *pre*-PROM.

The band at 1122 cm^{-1} which could be attributed to C-C/C-O stretching vibrations was found correlated with the 1414 cm^{-1} symmetric stretching carboxyl vibration, thus indicating the possible involvement of an organic or amino acid (Pretsch *et al.*, 1989). On the other hand, the band at 1039 cm^{-1}, attributed to C-C/C-O stretching vibrations was found correlated with a band at 1137 cm^{-1} also attributed to C-C/C-O stretching vibrations (Pretsch *et al.*, 1989; Liu and Mantsch, 1999). This correlation could indicate the involvement of a carbohydrate molecule, sugar alcohol or alcohol in the variation of the band at 1039 cm^{-1} (Pretsch *et al.*, 1989; Liu and Mantsch, 1999).

6.3.3.2. MIR and ^{1}H NMR SHY analysis

In order to further investigate the origin of the MIR bands of interest in *pre*-PROM, correlation analysis was performed using MIR/^{1}H NMR SHY using both MIR and ^{1}H NMR spectra of *pre*-PROM, as shown in the corresponding SHY map obtained (Figure 6.25).

From the 3000-2800 cm^{-1} region, only the band at 2955 cm^{-1} registered significant correlations (Figure 6.25a). Indeed the band at 2955 cm^{-1} was found positively correlated with several resonances in the region of CH_3 groups and NH/aromatics regions (possibly related to proteins), whereas correlating negatively with the resonances of lysine, phenylalanine and *myo*-inositol (Figure 6.25a, Table 6.10). Interestingly, from SHY analysis the 1039 cm^{-1} band was found positively correlated with resonances of glucose, whereas the band at 1122 cm^{-1} was positively correlated with resonances of lactate (Figure 6.25b, Table 6.10). The SHY results seem to be in agreement with the intra-correlation results, which pointed to the relation of 1122 cm^{-1} with an organic acid molecule (lactate) and 1039 cm^{-1} with sugar or sugar alcohol type molecule (glucose).

Table 6.10 - MIR/NMR SHY correlation obtained using *pre*-PROM spectra (Figure). The main MIR bands affected in *pre*-PROM by comparison to control group, with respective direction of change are also indicated. Assignments are given in brackets; amino acids are represented in 3 letter code.

		MIR/NMR SHY correlations	
MIR bands (cm^{-1})	**change relative to control**	**Positively correlated ^{1}H-NMR signals**	**Negatively correlated ^{1}H-NMR signals**
2955 (νC-H)	↑	(CH_3 groups 0.24-0.72 ppm), (NH/aromatics 7.60-8.78 ppm)	Lys, Phe, *myo*-inositol
1122 (νC-C/C-O)	↓	Lactate	-
1039 (νC-C/C-O)	↓	Glucose	-

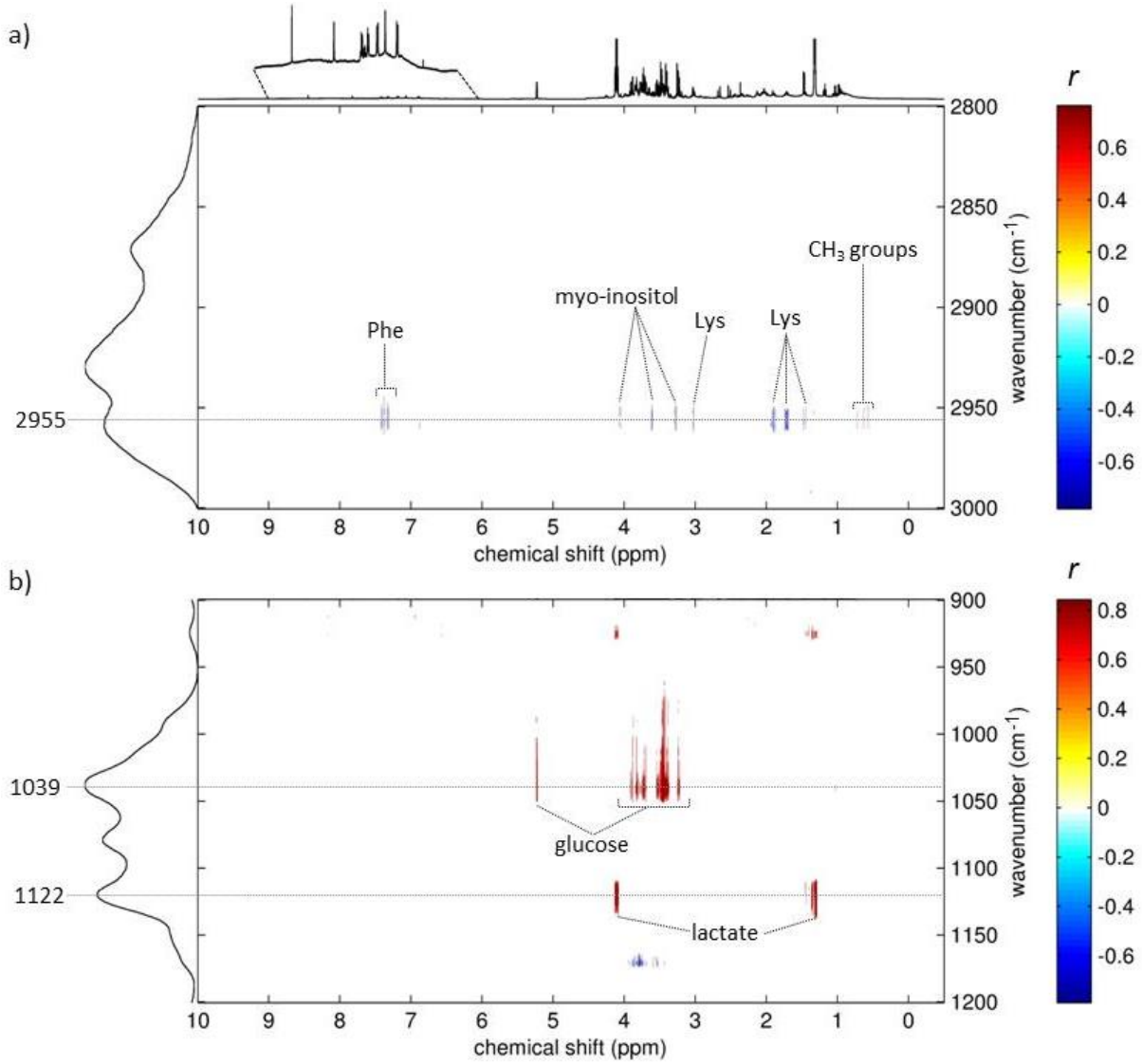


Figure 6.25 - MIR/^{1}H-NMR statistical heterospectroscopy (SHY) maps for *pre*-PROM group (n=37) of ^{1}H-NMR spectra (0-10 ppm region) and the MIR spectral regions: a) 3000-2800 cm^{-1}, b) a) 1200-900 cm^{-1}. A correlation threshold of $|r|>0.7$ was considered. Dashed horizontal lines indicate the bands of interest identified through loadings analysis and their correlations across the NMR spectrum.

6.3.4. Proposed interpretation of the metabolite variations detected in *pre*-PROM cases

Generally, the ^{1}H NMR and MIR of 2nd trimester AF PLS-DA models obtained for the early prediction PROM showed poor predictive power, certainly resulting from the small number of variations detected by both methodologies. Although slight improvements were achieved in the predictive ability of MIR PLS-DA models, the interpretation of the results from the MIR PLS-DA loadings and STOCSY and SHY is very limited and must be interpreted with care. The MIR analysis did suggest the possible increase of proteins (νC-H vibration and MIR/^{1}H NMR correlations with

CH_3 and NH groups) and concomitant decreases in lysine, phenylalanine and *myo*-inositol, together with decreases in glucose and lactate (suggested with basis on MIR/^{1}H NMR). However, no association has been found in literature linking these variations with PROM. On the other hand, evidence gathered in other studies has suggested that PROM is related to biochemical processes involving degradation and dissociation of amnion membranes collagen with the involvement of proteolytic enzymes, matrix metalloproteinases (Stuart *et al.*, 2005; Moore *et al.*, 2006) and due to the action of reactive oxygen species (ROS) (Caughey *et al.*, 2008). Ascorbic acid, which plays an essential role in the synthesis of collagen and in the maintenance of the membrane integrity contributing to the antioxidant activity against free-radical damage, was found reduced in PROM membranes at 37 weeks of gestation (Stuart *et al.*, 2005). In fact, one signal from ascorbate (multiplet at 4.01 ppm) was observed in ^{1}H NMR CPMG spectra of 2nd trimester AF as mentioned in chapter 5 (section 5.1.2, Table 5.2, page 129). The comparison between control and *pre*-PROM based on the ascorbate signal revealed an average decrease of this signal in *pre*-PROM group. Although the difference was not statistically significant (p=0.059), it may indicate possible early deficiency in the AF ascorbate contents of PROM cases.

It seems clear, however, that the use of more sensitive techniques, such as MS based methods could shed some more light in the small compositional changes occurring in 2nd trimester AF of *pre*-PROM cases (which escape the detection of both ^{1}H NMR and MIR) in order to provide clues to the early membrane degradation process and possibly lead to the development of predictive methods of PROM.

7. Preliminary study of 2nd trimester maternal urine by ultra-performance liquid chromatography mass spectrometry (UPLC-MS) for the study and prediction of pregnancy disorders (FM and GDM)

The work described in the present chapter constitutes an UPLC-MS profiling study of 2nd trimester maternal urine, whose principal objective was to explore any metabolic variations occurring in the maternal side as a result of prenatal disorders, and identify any possible relationships with those variations occurring in the fetal side (identified through AF analysis). The maternal urine samples were analysed separately from AF due of the reduced urine sample numbers available at the time of analysis compared to those of AF (as shown in Table 3.2, page 60). Hence, the results presented here are only preliminary. For UPLC-MS profiling only two disorder groups were considered: FM (n=13) and *pre*-diagnostic GDM (n=20). To further complement UPLC-MS analysis, UPLC-MS / ^{1}H NMR SHY analysis was performed using some of the previously published ^{1}H NMR maternal urine spectra (Diaz *et al.*, 2011), corresponding to the same samples analysed by UPLC-MS in the present study.

The analysis of other disorders by UPLC-MS was not possible at this stage, due to the small number of samples gather up to the time of analysis, namely of *pre*-PTD (n=5), CD (n<5), *pre*-PE and *pre*-IUGR (n<4). However, the analysis of maternal urine of such groups is also worth pursuing by UPLC-MS, as it has been initiated by ^{1}H NMR metabonomics for CD and *pre*-PTD (the subject of another PhD thesis) in an effort to develop less invasive diagnosis tools. Moreover, several studies using targeted analysis maternal urine, have suggested possible markers in connection with PE (Glew *et al.*, 2004; Scioscia *et al.*, 2007a) and early markers of oxidative stress in connection PE and PTD (Stein *et al.*, 2008).

7.1. Analysis of 2nd trimester maternal urine for the study of FM cases

7.1.1. UPLC-MS profiling for comparison of control and FM groups

The UPLC-MS data sets analysed originated clustered QCs sample distributions when compared to all maternal urine samples in their respective PCA scores scatter plots after Pareto and UV scaling (Annex III - Results from the UPLC-MS reproducibility evaluation). These observations indicated a good reproducibility of the data sets which were then considered suited for further MVA.

The PLS-DA scores scatter plots from controls *vs* FM models of the negative mode (ESI^-) of both HSS and HILIC columns (Figure 7.1), showed that good separation between control and FM groups was achieved, with Q^2 values of 0.528-0.736 (Table 7.1) indicating high predictive ability. On the other hand, the PLS-DA models corresponding to positive mode (ESI^+) data sets (not shown) registered Q^2 values of 0.242-0.363 (Table 7.1), which indicated lower predictive abilities. All models were the evaluated by MCCV.

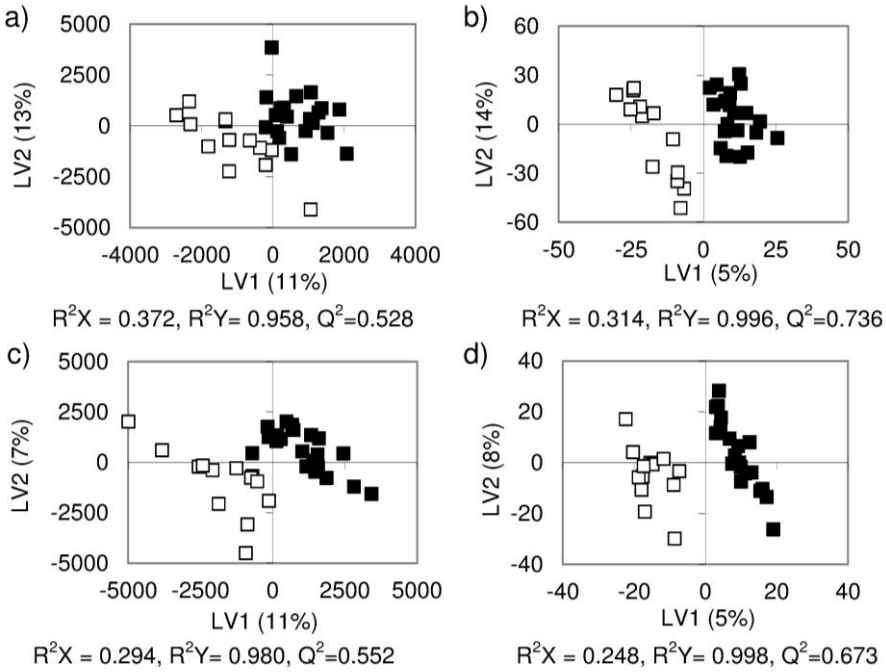


Figure 7.1 - PLS-DA scores scatter plots of UPLC-MS models of control (n=21) vs FM (n=13): a) HSS/ESI^-, Pareto scaled; b) HSS/ESI^-, UV scaled; c) HILIC/ESI^-, Pareto scaled; d) HILIC/ESI^-, UV scaled.

Table 7.1 - Maternal urine UPLC-MS based PLS-DA parameters of control *vs* FM models.

UPLC column	detection mode	scaling or transformation	PLS-DA model parameters			
			R^2X	R^2Y	Q^2	LV
HSS	ESI^+	Pareto	0.772	0.242	0.242	2
HSS	ESI^+	UV	0.337	0.834	0.363	2
HSS	ESI^-	Pareto	0.372	0.958	0.558	4
HSS	ESI^-	UV	0.340	0.996	0.736	4
HILIC	ESI^+	Pareto	0.248	0.795	0.331	2
HILIC	ESI^+	UV	0.251	0.822	0.313	2
HILIC	ESI^-	Pareto	0.294	0.980	0.552	4
HILIC	ESI^-	UV	0.248	0.998	0.673	4

Overlapping was observed between the true and permuted class models Q^2 distributions, but with clear tendency for separation of true class models, with the majority of Q^2 values located around Q^2=0.4 (Figure 7.2, left). ROC plots indicated average specificities of 95% for the true class models of both data sets, sensitivity between 40-60% (average 47%) for HSS/ESI$^-$ (UV scaling) data set (Figure 7.2a, right) and a sensitivity between 50-60% (average 59%) for HILIC/ESI$^-$ (UV scaling) data set (Figure 7.2b, right). The good predictive ability of the presented models was also expressed in the classification rates obtained: 97% for HSS/ESI$^-$ (UV scaling) and 83% for HILIC/ESI$^-$ (UV scaling). Similar predictive abilities were obtained in MCCV of Pareto scaled HSS/ESI$^-$ and HILIC/ESI$^-$ data sets (not shown).

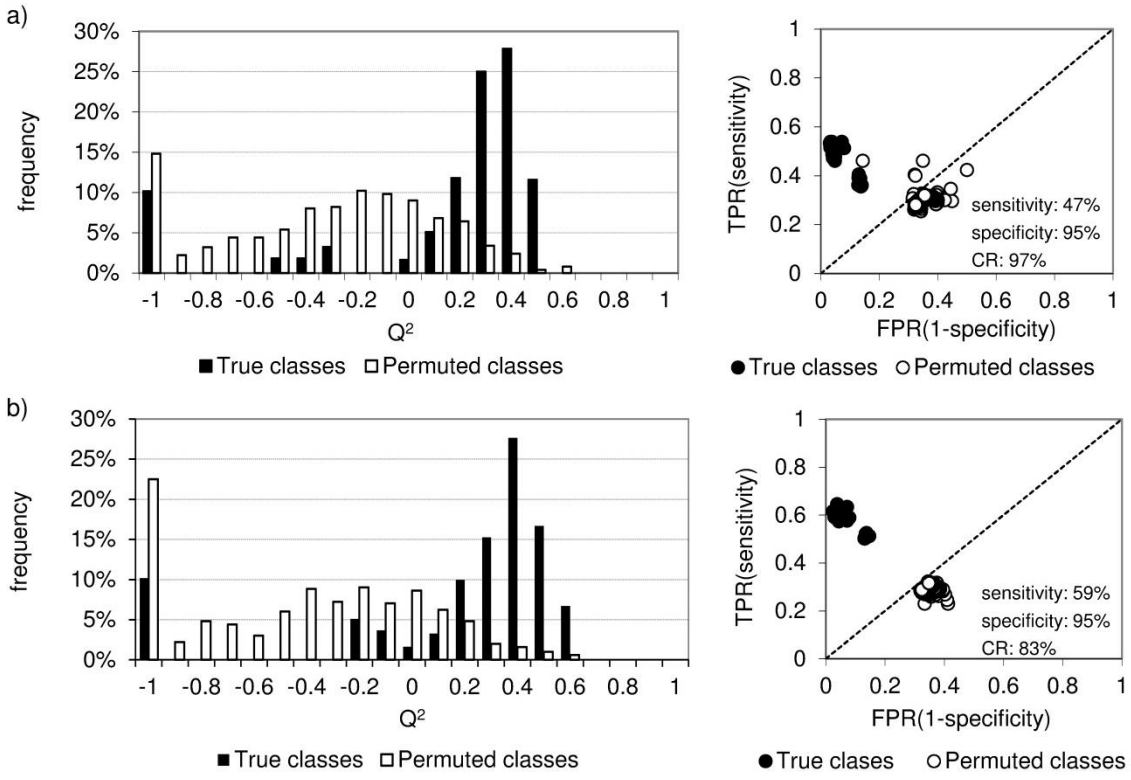


Figure 7.2 - MCCV results of PLS-DA UPLC-MS models of control (n=21) *vs* FM (n=13): a) HSS/ESI$^-$ (UV scaling); b) HILIC/ESI$^-$ (UV scaling).

Despite the moderate Q^2 obtained for positive mode data sets, their respective PLS-DA models were not validated through MCCV, which were characterized by completely overlapped true and permuted classes in both Q^2 distributions and ROC plots (not shown).The higher predictive ability of the PLS-DA models seem to indicate to a higher impact of metabolites ionizing preferentially in negative mode (ESI$^-$) in the urinary profile of FM cases.

Therefore, PLS-DA loadings analysis was performed for ESI$^-$ models in order to identify the MS features contributing to the separation observed in the controls *vs* FM PLS-DA models using the

same approach used for AF UPLC-MS data sets described in chapters 5 and 6. The resulting S-plot and covariance ratio plots obtained for LV1 of HILIC/ESI$^-$ UV model (Figure 7.3) illustrates the loadings analysis. The MS features were firstly selected from each S-plot of the separating component from PLS-DA scores scatter plots (Figure 7.3a), considering a correlation threshold of |p(corr)[1]|>0.5 (threshold lines, Figure 7.3a). The S-plot shown in Figure 7.3a, was derived from UV scaled data, therefore it different from the S-plots obtained from the Pareto scaled or log transformed data sets, such as the one shown in Figure 5.10a (chapter 5, page 135). The linear plot obtained in S-plots of UV scaled data results from the fact that the covariance (p[1]) which is scaled by standard deviation of each original variable (i.e. before scaling). In such an S-plot, only the correlation is needed to define the thresholds for feature selection.

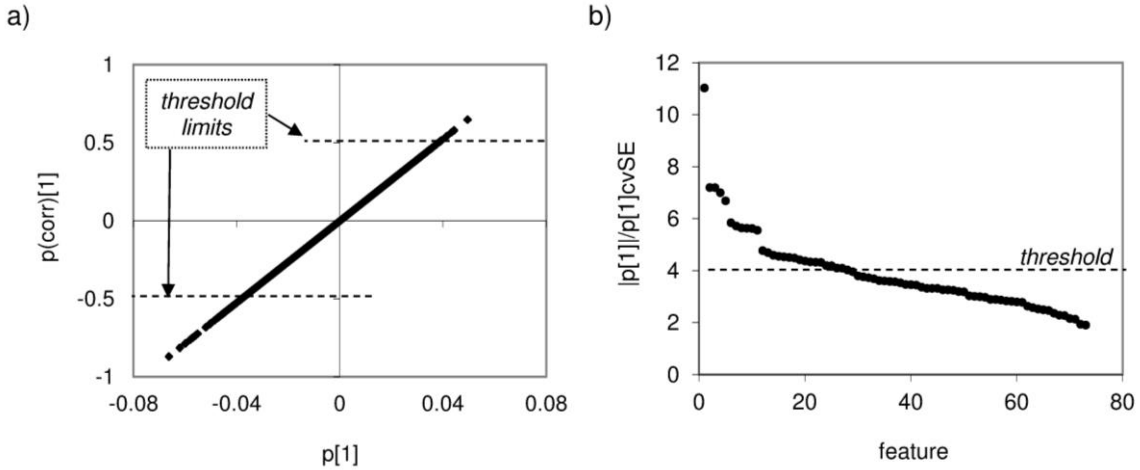


Figure 7.3 - PLS-DA control *vs* FM HILIC/ESI$^-$ UV model loadings selection: a) S-plot and b) covariance ratio plots of LV1 with correlation threshold limits indicated by dashed lines; dashed lines in the covariance ratio plot indicates the threshold for features selection.

The features obtained from the S-plots were further evaluated according to their intrinsic variability over cross-validation by the use of covariance plots (Figure 7.3b). The features for which the standard error of the covariance was not greater than 25% than the covariance value (threshold of 4 indicated in the dashed line of Figure 7.3b) were the features with the most stable variation in the PLS-DA models, therefore being considered the MS features with highest impact in class separation.

The significance of each of the MS features was also tested by univariate analysis. A list of significant assigned and partially assigned MS features together with the respective model of origin, variation in FM group and *p*-value is given in Table 7.2.

Table 7.2 - List of assigned significant UPLC-MS features obtained for 2nd trimester urine of the FM affected pregnancies compared to healthy controls. Feature intensities are indicated as: very high (>10^6), high (>10^5, <10^6), low (>10^4, <10^5) and very low (<10^4). [a] dimer and [b] adduct of hippurate. The names in brackets correspond to partially assigned features.

metabolite	UPLC Column	detection mode	m/z	RT(min.)	feature intensity	model scaling	direction and magnitude of variation	*p*-value
hippurate	HSS	ESI^-	178.06	3.9	very high	Pareto	↓ 1.5	6.57×10^{-03}
	HSS	ESI^-	357.13	3.9	very high	Pareto		
	HSS	ESI^-	442.19	3.9	high	UV		
(glucuronidated)	HSS	ESI^-	384.07	3.4	low	UV	↓ 5.5	1.16×10^{-02}
(sulfated)	HSS	ESI^-	242.13	3.9	very low	UV	↑ 1.5	1.82×10^{-03}
(glucuronidated)	HILIC	ESI^-	375.13	5.1	very low	UV	↓ 3.1	4.02×10^{-03}
(sulfated)	HILIC	ESI^-	190.00	0.6	low	UV	↓ 2.3	1.30×10^{-02}

From the MS features presented in Table 7.2 only hippurate (decreasing in FM) and its respective MS adducts could be fully assigned (Table 7.2). Significant variations were also found for features corresponding to two glucuronidated metabolites (decreasing in FM), one sulfated metabolite (increased in FM) and another sulfated metabolite (decreased in FM). The assignment of these features were only partially achieved from their respective MS/MS fragmentation spectra shown in Figure 7.4, which lead to the identification of mass losses of negatively charged sulfate residue or the loss of a neutral glucuronic acid residue (neutral loss) (Figure 7.4, page 198).

No further assignments were achieved for the remaining peaks detected in the MS/MS spectra. A list of 17 features found varying significantly (p<0.05) between control and FM group and for which no assignment was obtained at this stage is also presented in Annex VII - Lists of UPLC-MS unassigned features detected in maternal urine. It is worth mentioning that one of the unassigned relevant features which is found to be increased in urine of FM subjects (unassigned 385.17 m/z) was also found to be increased in AF of FM subjects (Annex VII - Lists of UPLC-MS unassigned features detected in maternal urine) thus suggesting a possible connection between maternal and fetal environment. However, further assignment work is needed to identify the origin of such feature.

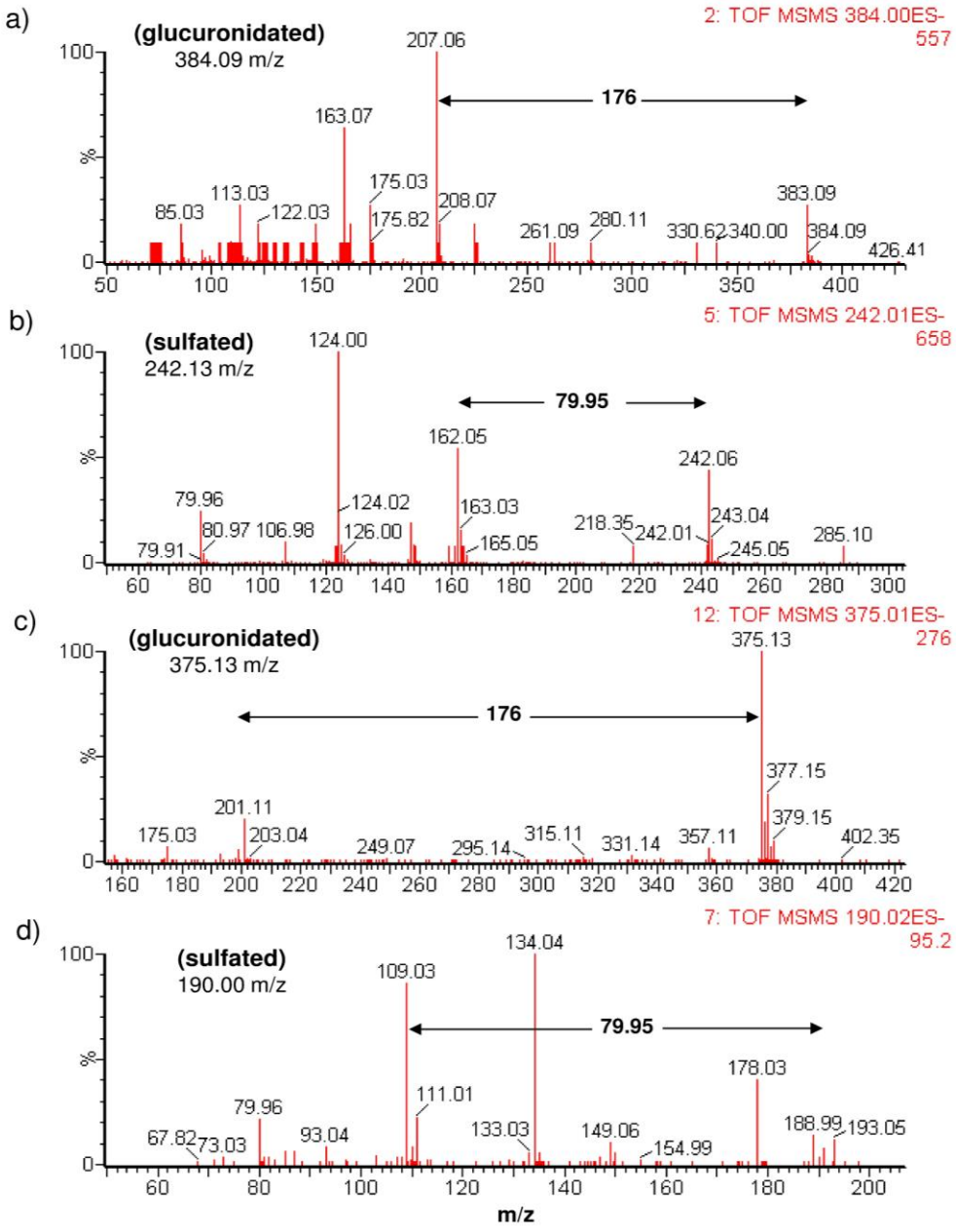


Figure 7.4 - Fragmentation mass spectra (MS/MS) of the partially assigned features: a) glucuronidated, 384.09 m/z, 3.4 min; b) sulfated, 242.13 m/z, 3.9 min; c) glucuronidated, 375.13 m/z, 5.1 min; d) sulfated, 190.00 m/z, 0.6 min. The mass differences of 176 and 79.95 correspond to the glucuronic acid neutral loss and the sulfate residues respectively.

7.1.2. UPLC-MS/^{1}H NMR SHY analysis of FM cases

Similarly to what has been done in the previous chapters 5 and 6, UPLC-MS/^{1}H NMR SHY analysis was performed using 1D STOCSY in order to find the correlations between the significant urine MS features obtained from the FM models from Table 7.2 and the full ^{1}H NMR spectra of the maternal urine samples from FM affected cases. An 1D MS/NMR STOCSY plot obtained for

hippurate MS feature (178.06 m/z, 3.09 min) shown in Figure 7.5. The MS/NMR SHY representation highlights the signals from the ^{1}H NMR spectrum that correlate with the MS feature (through correlation coefficient, *r*) as well as the direction of change (through covariance). In the example from Figure 7.5, the MS feature of hippurate correlated positively with the ^{1}H NMR hippurate resonances, and negatively with several signals, which however registered non-significant correlations (i.e., $p>0.001$). The complete list of lists the correlations found between maternal urine MS features and the full ^{1}H NMR spectrum is shown in Table 7.3.

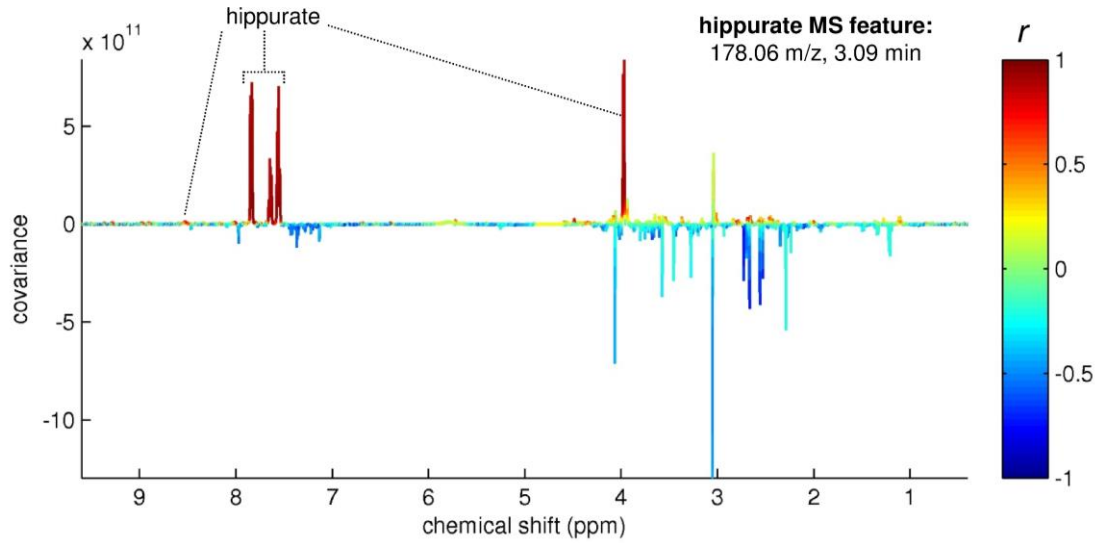


Figure 7.5 - MS/NMR SHY analysis of hippurate MS feature (178.06 m/z, 3.09 min) and full ^{1}H NMR spectra of urine. The indicated peaks registered significant correlations values: $|r|\geq 0.70$ and ($p<0.001$).

Table 7.3 – List of MS/NMR correlations obtained from SHY analysis of significant urine MS features and urine ^{1}H NMR spectra of FM. Maximum or minimum correlations values $|r|\geq 0.70$ and ($p<0.001$) were considered. Assigned spin systems are indicated in bold. s, singlet, d, doublet, t, triplet, m, multiplet, bs, broad singlet; * variation relative to control group.

metabolite	m/z	RT(min)	direction of variation*	^{1}H NMR resonances correlated: positively to the MS feature / ppm	negatively to the MS feature / ppm
hippurate	178.06	3.9	↓	**hippurate:** 3.97(d), 7.56(t), 7.64(t), 7.84(d), 8.52(bs)	-
glucuronidated	384.07	3.4	↓	unknown: 1.83(m)	-
sulfated	242.13	3.9	↑	unknown: 2.23, 2.25, 3.19(s), 5.58	unknown: 8.42(d)
sulfated	239.07	3.3	↑	unknown: 5.34(s)	-

Besides the expected direct correlation between the hippurate MS feature and its own ^{1}H NMR resonances (Figure 7.5), three of the remaining significant MS features (one glucuronidated compound and two sulfated metabolites) showed correlations with ^{1}H NMR resonances that are unassigned at this stage (Table 7.3). Therefore, no further assignment or metabolic information could be retrieved through SHY analysis.

7.1.3. Proposed interpretation of the urinary metabolite variations occurring in FM cases

As viewed by UPLC-MS, FM cases have a considerably less impact on 2nd trimester maternal urine than in 2nd trimester AF, as shown previously in chapter 5. The main difference noted in maternal urine relates to decreased hippurate in FM-affected pregnancies, a change which was however not detected in previous ^{1}H NMR work (Diaz *et al.*, 2011), even though hippurate signals were clearly visible in ^{1}H NMR spectra. This may reflect the different number of samples of both control (n=25 *vs* 21 here) and FM groups (n = 26 *vs* 13 here) in each study, and thus hippurate variations should be interpreted with care. Decreases in hippurate have been observed in the urine of lung cancer patients (Carrola *et al.*, 2011), autistic children (Yap *et al.*, 2010), and in cases of high blood pressure (Holmes *et al.*, 2008), and have been associated with abnormalities in gut microflora. However, no link has been found, between hippurate and prenatal disorders. The other metabolite changes found in maternal urine from FM affected cases by UPLC-MS comprised two glucuronidated metabolites (375.13 and 384.07 m/z, decreased), one sulfated metabolite (190.00 m/z, decreased) and a second sulfated metabolite (242.13 m/z, increased), which have low/very low intensity in MS (Table 7.2) and may be present in small amounts not being detected by ^{1}H NMR. Although only the partial assignment was only possible for the metabolites, several metabolites candidates can undergo conjugation reactions with glucuronic acid and sulfate prior to excretion into urine, including endogenous compounds (such as steroid hormones) and xenobiotics (Murray *et al.*, 2003). None of the pregnant woman reported the use of any medication, nor did the m/z values fit the major drug conjugates (such as paracetamol, aspirin or ibuprofen). No match was also found between these metabolites and the major steroid hormones expected in urine during pregnancy (such as estriol, pregnenolone or progesterone). It seems possible, however, that conjugation reactions involving other endogenous or exogenous metabolites may be specifically affected in the FM cases. Also regarding steroid hormones, the feature at 463.20 m/z found at higher levels in 2nd trimester AF of FM affected subjects and tentatively assigned to estriol-3-glucuronide/estriol-16-glucuronide (as shown in chapter 5, Table

5.4, page 135) was also found higher in maternal urine of FM cases although with a lower significance (p = 0.033) when re-analysed by univariate methods, thus not meeting the loadings selection criteria used. This interesting observation provides a link between variations observed on both fetal and maternal sides. Once more, evidence was found on the connection between elevated estriol-3-glucuronide/estriol-16-glucuronide and the occurrence of FM.

It is interesting to note that some variations observed previously by ^{1}H NMR on maternal urine of FM affected pregnancies (e.g. increases in isoleucine, valine, choline, N-methyl-nicotinamide (NMND), threonine, cis-aconitate and hypoxanthine) (Diaz *et al.*, 2011) were not observed in the present study. The specific UPLC-MS conditions chosen may play an important role in this apparent discrepancy since some of these compounds (namely threonine, cis-aconitate and hypoxanthine), known to be present in maternal urine in concentrations higher than a few mM could not be identified in the acquired urine datasets. On the other hand, isoleucine, valine, choline, N-methyl-nicotinamide (NMND), which could be identified in the maternal urine UPLC-MS data sets, were not significantly altered in FM cases (p>0.05), as confirmed by univariate analysis. The above mentioned different composition of the control and FM groups may also have an influence in the different observed results. This demonstrates the need for an increase in sample size, re-analysis by UPLC-MS and possibly the search for these specific variations using targeted analysis.

7.2. Analysis of 2nd trimester maternal urine for the study of *pre*-diagnostic GDM cases

PLS-DA models were obtained for *pre*-diagnostic GDM and control groups. The PLS-DA models corresponding to HSS/ESI^{+} and HSS/ESI^{-} UV scaled data sets models (Figure 7.6), where the most robust (Q^2=0.478 and Q^2=0.413, respectively) as shown in Table 7.4. Most of the separation between control and *pre*-diagnostic GDM cases was noted in LV2 (Figure 7.6), with similar separations being obtained for the remaining data sets as summarized in Table 7.4 (not shown).

Despite their apparent good separation, none of the models could be validated through MCCV (results not shown), therefore the models showed no predictive power.

These results contrast with those obtained previously for maternal urine by ^{1}H NMR which resulted in predictive models for *pre*-diagnostic GDM (Diaz *et al.*, 2011). In the ^{1}H NMR study, several possible early markers of GDM were suggested such as 3-hydroxyisovalerate, 2-hydroxyisobutyrate, choline, N-methyl-2-pyridone-5-carboxamide (2PY) and N-methyl-nicotinamide (NMND) (Diaz *et al.*, 2011).

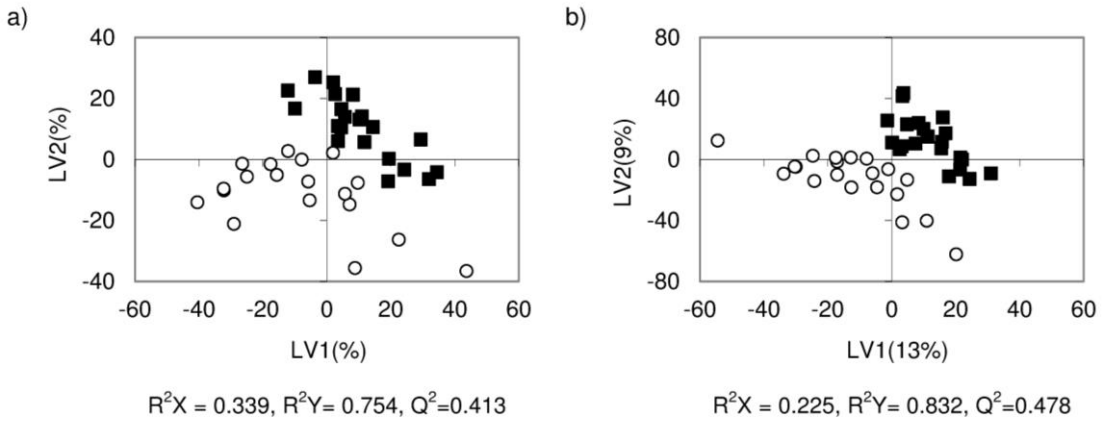


Figure 7.6 - PLS-DA scores scatter plots of UPLC-MS models of control (n=21) *vs pre*-diagnostic GDM (n=20): a) HSS/ESI$^+$, UV scaled; b) HSS/ESI$^-$, UV scaled.

Table 7.4 - Maternal urine UPLC-MS based PLS-DA parameters of control *vs pre*-diagnostic GDM models.

UPLC column	detection mode	scaling or transformation	PLS-DA model parameters			
			R²X	R²Y	Q²	LV
HSS	ESI+	Pareto	0.262	0.666	0.176	2
HSS	ESI+	UV	0.339	0.754	0.413	2
HSS	ESI-	Pareto	0.255	0.676	0.217	2
HSS	ESI-	UV	0.255	0.832	0.478	2
HILIC	ESI+	Pareto	0.262	0.670	0.164	2
HILIC	ESI+	UV	0.239	0.776	0.389	2
HILIC	ESI-	Pareto	0.123	0.831	-0.038	2
HILIC	ESI-	UV	0.150	0.883	0.322	2

Both choline (194.10 m/z in ESI$^-$) and NMND (137.05 m/z, in ESI$^+$), were however identified in UPLC-MS data sets, whereas the remaining metabolites were not detected, possibly due to early elution under the chromatographic conditions used. After re-analysis by univariate analysis, only choline was significantly increased (p = 0.020) in the *pre*-diagnostic GDM group, thus confirming the variation observed by ^{1}H NMR which indicated a possible disturbance of choline metabolism (Diaz *et al.*, 2011). Once more, the different results may simply reflect the different dimensions of both control (n=25 *vs* 21 here) and *pre*-diagnostic GDM (n=26 *vs* 20 here) groups used in both studies. Also the specific UPLC-MS conditions used may have hindered the detection of significant variation between controls and *pre*-diagnostic GDM cases, namely the variations of 3-hydroxyisovalerate, 2-hydroxyisobutyrate and 2PY. The verification of such variations requires the re-analysis of control and *pre*-diagnostic GDM samples under UPLC-MS, using enlarged sample groups.

8. Conclusions and future perspectives

The work presented in this thesis represents a contribution to the study of 2nd trimester AF composition through metabonomics, in order to deepen the understanding of the biochemistry and pathophysiological alterations occurring in prenatal disorders namely, FM and CD, occurring during the 2nd trimester; PTD, PE, IUGR, GDM and PROM occurring during the 3rd trimester of pregnancy. This knowledge was used in order to explore new markers of FM and CD, and to devise predictive markers for the occurrence of PTD, PE, IUGR, GDM and PROM.

The identification of the adequate AF storage and handling conditions necessary to ensure that minimal compositional changes occur during storage and analysis was of paramount importance for the proposed metabolic profiling studies. Therefore, a systematic study of AF compositional stability with time and temperature of storage using ^{1}H NMR was performed for the first time to our knowledge. The study consisted of monitoring of the effects freeze-thaw cycles, storage at the temperatures of -20°C, -70°C and at room temperature (22°C) in AF composition. Compositional changes were noted in the contents of acetate, citrate, glutamate, glutamine and pyruvate during the first two months of storage at -20°C, suggesting storage at lower temperatures (-70°C or lower) for longer storage times. Variations were noted in the levels of acetate, glutamate and glutamine during the first 4 hours at room temperature, which remained nearly stable up to 8 hours. Freeze-thaw cycles, introduced only minimal and sample dependent compositional changes, which should be minimized by avoiding freeze-thawing cycles.

Further compositional characterization studies using ^{1}H NMR spectroscopy were performed with the aim of increasing the knowledge of the composition of 2nd trimester AF from healthy pregnancies. Indeed, the work performed using high field ^{1}H NMR and especially LC-NMR/MS methods, allowed the detection of metabolites present in concentrations of less than 10 μM, and most notably enabled the identification of 6 metabolites from exogenous and endogenous origin not detected previously in AF by any other analytical technique. Moreover, the broad AF protein signals observed in ^{1}H NMR were characterized for the first time using DOSY, uncovering three proteins spin-systems, and enabling also the estimation of their M_w. Although their assignment was not possible at this stage, at least one spin-system could be roughly attributed to the most abundant glycoprotein in AF transferrin. With these results it was possible to build a database of 75 different compounds present in 2nd trimester AF of healthy pregnancies.

Metabolic profiling studies performed on 2nd trimester AF using a combination of ^{1}H NMR, UPLC-MS and MIR in connection with multivariate analysis, were used to explore the

compositional changes due to the occurrence of FM and CD. The most robust multivariate models were obtained for FM. In fact, irrespectively of the analytical platform considered, PLS-DA models of controls *vs* FM were associated with sensitivities of about 80% as well as specificities and classification rates above 80%. This resulted from the highest impact on 2nd trimester AF observed due to the occurrence of FM. The metabolic variations observed in all or the majority of analysed FM cases seemed to suggest impaired energy, amino acids and sugar metabolisms and also possibly abnormal kidney function. It was not possible, however to dwell on the effects of each particular type of FM on AF due the small sample size of each sub-group of FM obtained. In relation to diagnosed CD, PLS-DA models showing some discrimination between CD and control samples were only obtained with ^{1}H NMR data, and were less robust when compared to FM models, being associated with sensitivities lower than 60%. The poorer predictive ability obtained could be attributed to the small number of metabolite changes expected to occur in such a heterogeneous group of CD (composed of several CD types). Once more, the small sample size of each sub-group of CD cases hindered the exploration of the effects of each CD type on the 2nd trimester AF composition. Nevertheless, the further analysis of AF composition using increased sample numbers of each CD type may be of great importance, for instance in relation to trisomy 21, in order to devise diagnostic methods complementary to established diagnostic cytogenetics procedures.

Regarding the groups of disorders occurring in 3rd trimester pregnancy, namely *pre*-PTD, *pre*-PE, *pre*-IUGR, *pre*-diagnostic GDM and *pre*-PROM, 2nd trimester AF profiles were used primarily to devise multivariate models of early prediction of such disorders for which no established predictive methods are available. In general, the PLS-DA models obtained for these groups registered low predictive power (sensitivities lower than 70%). This is justifiable due to the smaller number of metabolite variations detected, or their absence, which may result from the temporal distance between AF collection and the occurrence of each disorder, but also, in the cases of *pre*-PTD, *pre*-PE and *pre*-IUGR, from the small number of samples analysed. Nevertheless, in spite of the poor models obtained, for which higher sample numbers should be considered in order to achieve predictive models, metabolites signals selected from the PLS-DA loadings were analysed by univariate analysis to explore possible biochemical variations, which may indicate the potential of markers for those disorders. Indeed a panel of significant metabolite changes were then detected in *pre*-PTD cases, thus indicating the potential of AF analysis for the development of prediction models of PTD. The variations were observed in the contents of amino acids, glucose, ketone bodies and allantoin suggesting earlier occurrence fetal-placental dysfunction,

hyperglycaemia and oxidative stress in *pre*-PTD subjects. Given the small number of metabolites variations detected for *pre*-PE (n=8) and their absence in *pre*-IUGR (n=9) cases and given the small sample sizes considered the analysis of these groups were largely exploratory at this stage. The increase in the sample numbers of the *pre*-PTD, *pre*-PE and *pre*-IUGR should then be used improve the sensitivity of the PLS-DA models. The results obtained for *pre*-diagnostic GDM, for which larger sample sizes were considered (n=23, 33 and 35, for UPLC-MS, MIR and ^{1}H NMR respectively) seem to reflect the small number of metabolite changes found, mainly resulting from glucose. Similarly, the results obtained in the analysis of pre-PROM cases (n=45), seem suggest that few differences exited between 2nd trimester AF. Therefore, the analysis of maternal biofluids in the case of *pre*-diagnostic GDM, and the use of more sensitive analytical approaches such as UPLC-MS for AF analysis of *pre*-PROM, should be explored in order to improve the sensitivity of the respective models.

UPLC-MS analysis of 2nd trimester maternal urine was also performed to explore the possible metabolic variations occurring in the maternal side as a result of prenatal disorders, and identify any possible relationships with those variations occurring in the fetal side (identified through AF analysis. The results were consistent with the higher impact of FM on metabolism, as seen through AF analysis. In comparison to AF, the maternal urine PLS-DA models showed however lower sensitivities (40-60%) despite the high classification rates (83-97%). The UPLC-MS study has also suggested possible markers, present in sub-μM concentrations in connection with FM. However, not confirming previous variations detected in another study by ^{1}H NMR (Diaz *et al* 2011), which may have resulted from the reduced number of samples considered. The results obtained for *pre*-diagnostic GDM, showed once more non predictive models in connection with the results obtained for AF. Therefore, in the case of *pre*-diagnostic GDM, it can be suggested that the search for markers should be performed using maternal plasma for instance, which may be more sensitive to the biochemical and pathophysiological alterations resulting from the onset of the disease.

Special mention is made to the limitation of the applicability of results presented, which were restricted to the small population of women who underwent amniocentesis procedure, i.e. to high risk and advanced maternal age pregnancies. More representative metabolic profiling studies should entail biofluids obtained less invasively (blood plasma and serum) or non-invasively (urine) and also consider pregnant women with a broader range of maternal ages and those at lower risk for prenatal disorders.

Taking into account the results obtained in AF profiling, ^{1}H NMR seemed to be the best suited for a first screening of AF composition. In fact, it presented an advantage over UPLC-MS and MIR in terms high throughput, reproducibility and ease of spectral interpretation, being inclusively used to aid interpretation of MIR and UPLC-MS data through SHY. MIR, which is the least expensive technique, has minimal AF preparation and possibility of high throughput. However, the MIR AF profiling procedure presented reproducibility issues which should be improved for future applications. UPLC-MS, which is the most sensitive technique, has a limited application to small batches of samples (n≈120) due to instrumental reproducibility constraints and sample preparation is more time consuming, which presents an important limitations for large scale AF profiling studies. The development of predictive methods may however, require the use of more sensitive techniques, such as UPLC-MS or other MS-based methods.

The results presented demonstrate the potential of 2nd trimester AF and maternal urine metabolic profiling for the development of new prenatal diagnosis methods for high risk pregnancies, laying the ground for future studies which should involve enlarged sample numbers in order to improve prediction models and search for biomarkers.

References

ADA: American Diabetes Association; Gestational diabetes mellitus. *Diabetes Care*, **2004**, 27(suppl 1): s88-s90.

Albert, K.; Ed. *On-line LC-NMR and related techniques*. Chichester, England, John Wiley & Sons, Ltd., 2002.

Alm, E., Torgrip, R., Åberg, K., Schuppe-Koistinen, I. and Lindberg, J.; A solution to the 1D NMR alignment problem using an extended generalized fuzzy Hough transform and mode support. *Analytical and Bioanalytical Chemistry*, **2009**, 395(1): 213-223.

Amaral, C., Gallardo, E., Rodrigues, R., Pinto Leite, R., Quelhas, D., Tomaz, C. and Cardoso, M. L.; Quantitative analysis of five sterols in amniotic fluid by GC-MS: Application to the diagnosis of cholesterol biosynthesis defects. *Journal of Chromatography B*, **2010**, 878(23): 2130-2136.

Amorini, A., Giorlandino, C., Longo, S., D'Urso, S., Mesoraca, A., Santoro, M., Picardi, M., Gullotta, S., Cignini, P., Lazzarino, D., Lazzarino, G. and Tavazzi, B.; Metabolic profile of amniotic fluid as a biochemical tool to screen for inborn errors of metabolism and fetal anomalies. *Molecular and Cellular Biochemistry*, **2012**, 359(1): 205-216.

An, Z., Chen, Y., Zhang, R., Song, Y., Sun, J., He, J., Bai, J., Dong, L., Zhan, Q. and Abliz, Z.; Integrated Ionization Approach for RRLC−MS/MS-based metabonomics: finding potential biomarkers for lung cancer. *Journal of Proteome Research*, **2010**, 9(8): 4071-4081.

Anagnostopoulos, A. K., Kolialexi, A., Mavrou, A., Vougas, K., Papantoniou, N., Antsaklis, A., Kanavakis, E., Fountoulakis, M. and Tsangaris, G. T.; Proteomic analysis of amniotic fluid in pregnancies with Klinefelter syndrome foetuses. *Journal of Proteomics*, **2010**, 73(5): 943-950.

Andrews, D.; *Electromagnetic radiation*. *Encyclopedia of spectroscopy and spectrometry*. J. C. Lindon, J. Holmes and G. Tranter. London, Academic Press, 2000, 1**:** 397-401.

Atkins, P. and de Paula, J.; *Physical Chemistry*, Oxford University Press, 2006.

Atzori, L., Antonucci, R., Barberini, L., Griffin, J. L. and Fanos, V.; Metabolomics: a new tool for the neonatologist. *Journal of Maternal-Fetal and Neonatal Medicine*, **2009**, 22(s3): 50-53.

Auray-Blais, C., Raiche, E., Gagnon, R., Berthiaume, M. and Pasquier, J.-C.; Metabolomics and preterm birth: What biomarkers in cervicovaginal secretions are predictive of high-risk pregnant women? *International Journal of Mass Spectrometry*, **2011**, 307(1–3): 33-38.

Avdulov, N. A., Chochina, S. V., Daragan, V. A., Schroeder, F., Mayo, K. H. and Wood, W. G.; Direct binding of ethanol to bovine serum albumin: a fluorescent and ^{13}C NMR multiplet relaxation study. *Biochemistry*, **1996**, 35(1): 340-347.

Azmi, J., Griffin, J. L., Shore, R. F., Holmes, E. and Nicholson, J. K.; Chemometric analysis of biofluids following toxicant induced hepatotoxicity: A metabonomic approach to distinguish the effects of 1-naphthylisothiocyanate from its products. *Xenobiotica*, **2005**, 35: 839-852.

Baggot, P. J., Eliseo, A. J. Y., DeNicola, N. G., Kalamarides, J. A. and Shoemaker, J. D.; Organic acid concentrations in amniotic fluid found in normal and down syndrome pregnancies. *Fetal Diagnosis and Therapy*, **2008**, 23: 245–248.

Balayssac, S., Delsuc, M.-A., Gilard, V., Prigent, Y. and Malet-Martino, M.; Two-dimensional DOSY experiment with excitation sculpting water suppression for the analysis of natural and biological media. *Journal of Magnetic Resonance*, **2009**, 196(1): 78-83.

Barnes, R. J., Dhanoa, M. S. and Lister, S. J.; Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Applied Spectroscopy*, **1989**, 43(5): 772-777.

Barros, A. S.; *Contribution à la sélection et la comparaison de variables caractéristiques*. PhD Thesis, Institut National Agronomique Paris-Grignon, 1999.

Barton, R. H., Nicholson, J. K., Elliott, P. and Holmes, E.; High-throughput ^{1}H NMR-based metabolic analysis of human serum and urine for large-scale epidemiological studies: validation study. *International Journal of Epidemiology*, **2008**, 37: 31-40.

Bax, A. and Davis, D. G.; MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. *Journal of Magnetic Resonance*, **1985**, 65: 355-360.

Beardsall, K., Vanhaesebrouck, S., Ogilvy-Stuart, A. L., Vanhole, C., Palmer, C. R., Ong, K., vanWeissenbruch, M., Midgley, P., Thompson, M., Thio, M., Cornette, L., Ossuetta, I., Iglesias, I., Theyskens, C., de Jong, M., Gill, B., Ahluwalia, J. S., de Zegher, F. and Dunger, D. B.; Prevalence and determinants of hyperglycemia in very low birth weight infants: cohort analyses of the NIRTURE study. *The Journal of pediatrics*, **2010**, 157(5): 715-719.e713.

Beckonert, O., Keun, H. C., Ebbels, T. M. D., Bundy, J., Holmes, E., Lindon, J. C. and Nicholson, J. K.; Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nature Protocols*, **2007**, 2(11): 2692-2703.

Bene, B. C.; Diagnosis of meconium in amniotic fluids by nuclear magnetic-resonance spectroscopy. *Physiological Chemistry and Physics*, **1980**, 12(3): 241-247.

Bene, G. J., Borcard, B., Graf, V., Hiltbrand, E., Magnin, P. and Noack, F.; Proton NMR-relaxation dispersion in meconium solutions and healthy amniotic-fluid - possible applications to medical diagnosis. *Zeitschrift Fur Naturforschung C- A Journal of Biosciences*, **1982**, 37(5-6): 394-398.

Bernini, P., Bertini, I., Luchinat, C., Nincheri, P., Staderini, S. and Turano, P.; Standard operating procedures for pre-analytical handling of blood and urine for metabolomic studies and biobanks. *Journal of Biomolecular NMR*, **2011**, 49(3): 231-243.

Bianchi, D. W., Platt, L. D., Goldberg, J. D., Abuhamad, A. Z., Sehnert, A. J. and Rava, R. P.; Genome-wide fetal aneuploidy detection by maternal plasma DNA sequencing. *Obstetrics & Gynecology*, **2012**, 119(5): 890-901.

Bock, J. L.; Metabolic profiling of amniotic fluid by proton nuclear magnetic resonance spectroscopy: correlation with fetal maturation and other clinical variables. *Clinical Chemistry*, **1994**, 40(1): 56-61.

Bodenhausen, G. and Ruben, D. J.; Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chemical Physics Letters*, **1980**, 69(1): 185-189.

Boisvert, M. R., Koski, K. G., Burns, D. H. and Skinner, C. D.; Prediction of gestational diabetes mellitus based on an analysis of amniotic fluid by capillary electrophoresis. *Biomarkers in Medicine*, **2012a**, 6(5): 645-653.

Boisvert, M. R., Koski, K. G., Burns, D. H. and Skinner, C. D.; Early prediction of macrosomia based on an analysis of second trimester amniotic fluid by capillary electrophoresis. *Biomarkers in Medicine*, **2012b**, 6(5): 655-662.

Borcard, B., Hiltbrand, E., Magnin, P., Béné, G. J., Briguet, A., Duplan, J. C., Delmau, J., Guibaud, S., Bonnet, M., Dumont, M. and Fara, J. F.; Estimating meconium (fetal feces) concentration in human amniotic fluid by nuclear magnetic resonance. *Physiological Chemistry and Physics* **1982**, 14(3): 189-192

Brindle, J. T., Antti, H., Holmes, E., Tranter, G., Nicholson, J. K., Bethell, H. W. L., Clarke, S., Schofield, P. M., McKilligin, E., Mosedale, D. E. and Grainger, D. J.; Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using ^{1}H-NMR-based metabonomics. *Nature Medicine*, **2002**, 8(12): 1439-1445.

Broadhurst, D. and Kell, D.; Statistical strategies for avoiding false discoveries in metabolomics and related experiments. *Metabolomics*, **2006**, 2(4): 171-196.

Brouns, R., Ursem, N., Lindemans, J., Hop, W., Pluijm, S., Steegers, E. and Steegers-Theunissen, R.; Polymorphisms in genes related to folate and cobalamin metabolism and the associations with complex birth defects. *Prenatal Diagnosis*, **2008**, 28: 485–493.

Bruker Avance 1D/2D manual. Fällanden, Switzerland, Bruker AG. 2003.

Buhimschi, I. A., Zhao, G., Rosenberg, V. A., Abdel-Razeq, S., Thung, S. and Buhimschi, C. S.; Multidimensional proteomics analysis of amniotic fluid to provide insight into the mechanisms of Idiopathic preterm birth. *PLoS ONE*, **2008**, 3(4): e2049.

Bujold, E., Romero, R., Kusanovic, J. P., Erez, O., Gotsch, F., Chaiworapongsa, T., Gomez, R., Espinoza, J., Vaisbuch, E., Mee Kim, Y., Edwin, S., Pisano, M., Allen, B., Podust, V. N., Dalmasso, E. A., Rutherford, J., Rogers, W., Moser, A., Yoon, B. H. and Barder, T.; Proteomic profiling of amniotic fluid in preterm labor using two-dimensional liquid separation and mass spectrometry. *The Journal of Maternal-Fetal & Neonatal Medicine*, **2008**, 21(10): 697 - 713.

Bylesjö, M., Rantalainen, M., Cloarec, O., Nicholson, J. K., Holmes, E. and Trygg, J.; OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. *Journal of Chemometrics*, **2006**, 20(8-10): 341-351.

Camacho, D., de la Fuente, A. and Mendes, P.; The origin of correlations in metabolomics data. *Metabolomics*, **2005**, 1(1): 53-63.

Carrola, J., Rocha, C. M., Barros, A. S., Gil, A. M., Goodfellow, B. J., Carreira, I. M., Bernardo, J., Gomes, A., Sousa, V., Carvalho, L. and Duarte, I. F.; Metabolic signatures of lung cancer in biofluids: NMR-based metabonomics of urine. *Journal of Proteome Research*, **2011**, 10(1): 221-230.

Castro, L. C.; *Hypertensive disorders of pregnancy*. *Essentials of obstetrics and gynecology*. N. F. Hacker, J. G. Moore and J. C. Gambone. Philadelphia, Elsevier, 2004**:** 197-207.

Caughey, A. B., Robinson, J. N. and Norwitz, E. R.; Contemporary diagnosis and management of preterm premature rupture of membranes. *Reviews in obstetrics and gynecology*, **2008**, 1(1): 11-22.

Cecconi, D., Lonardoni, F., Favretto, D., Cosmi, E., Tucci, M., Visentin, S., Cecchetto, G., Fais, P., Viel, G. and Ferrara, S. D.; Changes in amniotic fluid and umbilical cord serum proteomic profiles of foetuses with intrauterine growth retardation. *Electrophoresis*, **2011**, 32(24): 3630-3637.

Cetin, I.; Amino acid interconversions in the fetal-placental unit: the animal model and human studies in vivo. *Pediatric Research*, **2001**, 49(2): 148-154.

Cetin, I. and Alvino, G.; Intrauterine growth restriction: implications for placental metabolism and transport. A Review. *Placenta*, **2009**, 30, Supplement A: S77-S82.

Cetin, I., de Santis, M. S. N., Taricco, E., Radaelli, T., Teng, C., Ronzoni, S., Spada, E., Milani, S. and Pardi, G.; Maternal and fetal amino acid concentrations in normal pregnancies and in pregnancies with gestational diabetes mellitus. *American Journal of Obstetrics and Gynecology*, **2005**, 192(2): 610-617.

Cetin, I., Ronzoni, S., Marconi, A. M., Perugino, G., Corbetta, C., Battaglia, F. C. and Pardi, G.; Maternal concentrations and fetal-maternal concentration differences of plasma amino acids in normal and intrauterine growth-restricted pregnancies. *American Journal of Obstetrics and Gynecology*, **1996**, 174(5): 1575-1583.

Chen, E. Z., Chiu, R. W. K., Sun, H., Akolekar, R., Chan, K. C. A., Leung, T. Y., Jiang, P., Zheng, Y. W. L., Lun, F. M. F., Chan, L. Y. S., Jin, Y., Go, A. T. J. I., Lau, E. T., To, W. W. K., Leung, W. C., Tang, R. Y. K., Au-Yeung, S. K. C., Lam, H., Kung, Y. Y., Zhang, X., van Vugt, J. M. G., Minekawa, R., Tang, M. H. Y., Wang, J., Oudejans, C. B. M., Lau, T. K., Nicolaides, K. H. and Lo, Y. M. D.; Noninvasive prenatal diagnosis of fetal trisomy 18 and trisomy 13 by maternal plasma DNA sequencing. *PLoS ONE*, **2011**, 6(7): e21791.

Chen, J., Wang, W., Lv, S., Yin, P., Zhao, X., Lu, X., Zhang, F. and Xu, G.; Metabonomics study of liver cancer based on ultra performance liquid chromatography coupled to mass spectrometry with HILIC and RPLC separations. *Analytica Chimica Acta*, **2009**, 650(1): 3-9.

Chen, J., Wen, H., Liu, J., Yu, C., Zhao, X., Shi, X. and Xu, G.; Metabonomics study of the acute graft rejection in rat renal transplantation using reversed-phase liquid chromatography and hydrophilic interaction chromatography coupled with mass spectrometry. *Molecular BioSystems*, **2012**.

Chmait, R. H. and Moore, T. R.; *Obstetrical procedures*. *Essentials of obstetrics and gynecology*. N. F. Hacker, J. G. Moore and J. C. Gambone. Philadelphia, Elsevier, 2004**:** 247-255.

Cho, C.-K. J., Smith, C. R. and Diamandis, E. P.; Amniotic Fluid proteome analysis from down syndrome pregnancies for biomarker discovery. *Journal of Proteome Research*, **2010**, 9(7): 3574-3582.

Chong, I.-G. and Jun, C.-H., 78, 103–112.; Performance of some variable selection methods when multicollinearity is present. *Chemometrics and Intelligent Laboratory Systems*, **2005**, 78: 103–112.

Claridge, T. D. W.; *High-Resolution NMR techniques in organic chemistry*. Oxford, Elsevier Science, 2009.

Clifton, M. S., Joe, B. N., Zektzer, A. S., Kurhanewicz, J., Vigneron, D. B., Coakley, F. V., Nobuhara, K. K. and Swanson, M. G.; Feasibility of magnetic resonance spectroscopy for evaluating fetal lung maturity. *Journal of Pediatric Surgery*, **2006**, 41(4): 768-773.

Cloarec, O., Campbell, A., Tseng, L. h., Braumann, U., Spraul, M., Scarfe, G., Weaver, R. and Nicholson, J. K.; Virtual chromatographic resolution enhancement in cryoflow LC-NMR experiments via statistical total correlation spectroscopy. *Analytical Chemistry*, **2007**, 79(9): 3304-3311.

Cloarec, O., Dumas, M. E., Craig, A., Barton, R. H., Trygg, J., Hudson, J., Blancher, C., Gauguier, D., Lindon, J. C., Holmes, E. and Nicholson, J.; Statistical total correlation spectroscopy: an exploratory approach for latent biomarker identification from metabolic ^{1}H NMR data sets. *Analytical Chemistry*, **2005**, 77(5): 1282-1289.

Coen, M., Hong, Y. S., Cloarec, O., Rhode, C. M., Reily, M. D., Robertson, D. G., Holmes, E., Lindon, J. C. and Nicholson, J. K.; Heteronuclear ^{1}H-^{31}P Statistical total correlation NMR spectroscopy of intact liver for metabolic biomarker assignment: application to galactosamine-induced hepatotoxicity. *Analytical Chemistry*, **2007**, 79(23): 8956-8966.

Cohn, B., Joe, B., Zhao, S., Kornak, J., Zhang, V., Iman, R., Kurhanewicz, J., Vahidi, K., Yu, J., Caughey, A. and Swanson, M.; Quantitative metabolic profiles of 2nd and 3rd trimester human amniotic fluid using ^{1}H HR-MAS spectroscopy. *Magnetic Resonance Materials in Physics, Biology and Medicine*, **2009**, 22(6): 343-352.

Cohn, B. R., Fukuchi, E. Y., Joe, B. N., Swanson, M. G., Kurhanewicz, J., Yu, J. and Caughey, A. B.; Calculation of gestational age in late second and third trimesters by ex vivo magnetic resonance spectroscopy of amniotic fluid. *American Journal of Obstetrics and Gynecology*, **2010**, 203(1): 76.e71-76.e10.

Conde-Agudelo, A., Papageorghiou, A. T., Kennedy, S. H. and Villar, J.; Novel biomarkers for the prediction of the spontaneous preterm birth phenotype: a systematic review and meta-analysis. *BJOG: An International Journal of Obstetrics & Gynaecology*, **2011**, 118(9): 1042-1054.

Corcoran, O. and Spraul, M.; LC-NMR-MS in drug discovery. *Drug Discovery Today*, **2003**, 8(14): 624-631.

Crockford, D. J., Holmes, E., Lindon, J. C., Plumb, R. S., Zirah, S., Bruce, S. J., Rainville, P., Stumpf, C. L. and Nicholson, J. K.; Statistical heterospectroscopy, an approach to the integrated analysis of NMR and UPLC-MS data sets: application in metabonomic toxicology studies. *Analytical Chemistry*, **2006**, 78(2): 363-371.

Crockford, D. J., Maher, A. D., Ahmadi, K. R., Barrett, A., Plumb, R. S., Wilson, I. D. and Nicholson, J. K.; ^{1}H NMR and UPLC-MS^{E} statistical heterospectroscopy: characterization of drug metabolites (xenometabolome) in epidemiological studies. *Analytical Chemistry*, **2008**, 80(18): 6835-6844.

Crutchfield, C. A. and Harris, D. J.; Molecular mass estimation by PFG NMR spectroscopy. *Journal of Magnetic Resonance*, **2007**, 185: 179–182.

Daykin, C. A., Corcoran, O., Hansen, S. H., Bjornsdottir, I., Cornett, C., Connor, S. C., Lindon, J. C. and Nicholson, J. K.; Application of directly coupled HPLC NMR to separation and characterization of lipoproteins from human serum. *Analytical Chemistry*, **2001**, 73(6): 1084-1090.

Daykin, C. A., Foxall, P. J. D., Connor, S. C., Lindon, J. C. and Nicholson, J. K.; The comparison of plasma deproteinization methods for the detection of low-molecular-weight metabolites by ^{1}H nuclear magnetic resonance spectroscopy. *Analytical Biochemistry*, **2002**, 304(2): 220-230.

Dear, G. J., Plumb, R. S., Sweatman, B. C., Parry, P. S., Roberts, A. D., Lindon, J. C., Nicholson, J. K. and Ismail, I. M.; Use of directly coupled ion-exchange liquid chromatography-mass spectrometry and liquid chromatography-nuclear magnetic resonance spectroscopy as a strategy for polar metabolite identification. *Journal of Chromatography B: Biomedical Sciences and Applications*, **2000**, 748(1): 295-309.

Deprez, S., Sweatman, B. C., Connor, S. C., Haselden, J. N. and Waterfield, C. J.; Optimisation of collection, storage and preparation of rat plasma for ^{1}H NMR spectroscopic analysis in toxicology studies to determine inherent variation in biochemical profiles. *Journal of Pharmaceutical and Biomedical Analysis*, **2002**, 30(4): 1297-1310.

Dessì, A., Atzori, L., Noto, A., Adriaan Visser, G. H., Gazzolo, D., Zanardo, V., Barberini, L., Puddu, M., Ottonello, G., Atzei, A., Magistris, A. D., Lussu, M., Murgia, F. and Fanos, V.; Metabolomics in newborns with intrauterine growth retardation (IUGR): urine reveals markers of metabolic syndrome. *Journal of Maternal-Fetal and Neonatal Medicine*, **2011**, 24(S2): 35-39.

Di Iorio, R., Marinoni, E., Coacci, F., La Torre, R. and Cosmi, E. V.; Amniotic fluid nitric oxide and uteroplacental blood flow in pregnancy complicated by intrauterine growth retardation. *BJOG: An International Journal of Obstetrics & Gynaecology*, **1997**, 104(10): 1134-1139.

Diaz, S. O., Pinto, J., Graça, G., Duarte, I. F., Barros, A. S., Galhano, E., Pita, C., Almeida, M. d. C., Goodfellow, B. J., Carreira, I. M. and Gil, A. M.; Metabolic biomarkers of prenatal disorders: an exploratory NMR metabonomics study of 2nd trimester maternal urine and blood plasma. *Journal of Proteome Research*, **2011**, 10(8): 3732-3742.

Dieterle, F., Ross, A., Schlotterbeck, G. and Senn, H.; Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in ^{1}H NMR metabonomics. *Analytical Chemistry*, **2006**, 78(13): 4281-4290.

Ding, C., Chiu, R. W. K., Lau, T. K., Leung, T. N., Chan, L. C., Chan, A. Y. Y., Charoenkwan, P., Ng, I. S. L., Law, H.-y., Ma, E. S. K., Xu, X., Wanapirak, C., Sanguansermsri, T., Liao, C., Ai, M. A. T. J., Chui, D. H. K., Cantor, C. R. and Lo, Y. M. D.; MS analysis of single-nucleotide differences in circulating nucleic acids: Application to noninvasive prenatal diagnosis. *Proceedings of the National Academy of Sciences of the United States of America*, **2004**, 101(29): 10762-10767.

Divers, W. A., Wilkes, M. M., Babaknia, A., Hill, L. M., Quilligan, E. J. and Yen, S. S.; Amniotic fluid catecholamines and metabolites in intrauterine growth retardation. *American Journal of Obstetrics and Gynecology*, **1981**, 141(6): 608-610.

Dong, H., Zhang, A., Sun, H., Wang, H., Lu, X., Wang, M., Ni, B. and Wang, X.; Ingenuity pathways analysis of urine metabolomics phenotypes toxicity of Chuanwu in Wistar rats by UPLC-Q-TOF-HDMS coupled with pattern recognition methods. *Molecular BioSystems*, **2012**, 8(4): 1206-1221.

Dong, J., Cheng, K.-K., Xu, J., Chen, Z. and Griffin, J. L.; Group aggregating normalization method for the preprocessing of NMR-based metabolomic data. *Chemometrics and Intelligent Laboratory Systems*, **2011**, 108(2): 123-132.

Duarte, I. F., Godejohann, M., Braumann, U., Spraul, M. and Gil, A. M.; Application of NMR spectroscopy and LC-NMR/MS to the identification of carbohydrates in beer. *Journal of Agricultural and Food Chemistry*, **2003**, 51(17): 4847-4852.

Duarte, I. F., Goodfellow, B. J., Barros, A., Jones, J. G., Barosa, C., Diogo, L., Garcia, P. and Gil, A. M.; Metabolic characterisation of plasma in juveniles with glycogen storage disease type 1a (GSD1a) by high-resolution ^{1}H NMR spectroscopy. *NMR in Biomedicine*, **2007**, 20(4): 401-412.

Duarte, I. F., Legido-Quigley, C., Parker, D. A., Swann, J. R., Spraul, M., Braumann, U., Gil, A. M., Holmes, E., Nicholson, J. K., Murphy, G. M., Vilca-Melendez, H., Heaton, N. and Lindon, J. C.; Identification of metabolites in human hepatic bile using 800 MHz ^{1}H NMR spectroscopy, HPLC-NMR/MS and UPLC-MS. *Molecular BioSystems*, **2009**, 5(2): 180-190.

Duarte, I. F., Stanley, E. G., Holmes, E., Lindon, J. C., Gil, A. M., Tang, H., Ferdinand, R., McKee, C. G., Nicholson, J. K., Vilca-Melendez, H., Heaton, N. and Murphy, G. M.; Metabolic assessment of human liver transplants from biopsy samples at the donor and recipient stages using high-resolution magic angle spinning ^{1}H NMR spectroscopy. *Analytical Chemistry*, **2005**, 77(17): 5570-5578.

Dunn, W. B., Brown, M., Worton, S. A., Crocker, I. P., Broadhurst, D., Horgan, R., Kenny, L. C., Baker, P. N., Kell, D. B. and Heazell, A. E. P.; Changes in the metabolic footprint of placental explant-conditioned culture medium identifies metabolic disturbances related to hypoxia and pre-eclampsia. *Placenta*, **2009**, 30(11): 974-980.

Dunn, W. B., Wilson, I. D., Nicholls, A. W. and Broadhurst, D.; The importance of experimental design and QC samples in large-scale and MS-driven untargeted metabolomic studies of humans. *Bioanalysis*, **2012**, 4(18): 2249-2264.

Ebbels, T.; *Non-linear Methods for the analysis of metabolic profiles*. *The handbook of metabonomics and metabolomics*. J. C. Lindon, J. K. Nicholson and E. Holmes. Oxford, Elsevier, 2007**:** 203-226.

Ebbels, T. M. D. and Cavill, R.; Bioinformatic methods in NMR-based metabolic profiling. *Progress in Nuclear Magnetic Resonance Spectroscopy*, **2009**, 55(4): 361-374.

Ellis, D. I., Dunn, W. B., Griffin, J. L., Allwood, J. W. and Goodacre, R.; Metabolic fingerprinting as a diagnostic tool. *Pharmacogenomics*, **2007**, 8(9): 1243-1266.

Ellis, D. I. and Goodacre, R.; Metabolic fingerprinting in disease diagnosis: biomedical applications of infrared and Raman spectroscopy. *Analyst*, **2006**, 131(8): 875-885.

Evans, R., Powers, R., Ness, R., Cropcho, L., Daftary, A., Harger, G., Vergona, R. and Finegold, D.; Maternal and fetal amino acid concentrations and fetal outcomes during pre-eclampsia. *Reproduction*, **2003**, 125(6): 785-790.

Farina, A., Rapacchia, G., Freni Sterrantino, A., Pula, G., Morano, D. and Rizzo, N.; Prospective evaluation of ultrasound and biochemical-based multivariable models for the prediction of late pre-eclampsia. *Prenatal Diagnosis*, **2011**, 31(12): 1147-1152.

Favretto, D., Cosmi, E., Ragazzi, E., Visentin, S., Tucci, M., Fais, P., Cecchetto, G., Zanardo, V., Viel, G. and Ferrara, S.; Cord blood metabolomic profiling in intrauterine growth restriction. *Analytical and Bioanalytical Chemistry*, **2012**, 402(3): 1109-1121.

Fawcett, T.; An introduction to ROC analysis. *Pattern Recognition Letters*, **2006**, 27(8): 882-891.

Fenton, B. W., Lin, C.-S., Seydel, F. and Macedonia, C.; Lecithin can be detected by volume-selected proton MR spectroscopy using a 1·5T whole body scanner: a potentially non-invasive method for the prenatal assessment of fetal lung maturity. *Prenatal Diagnosis*, **1998**, 18(12): 1263-1266.

Fiehn, O.; Metabolomics – the link between genotypes and phenotypes. *Plant Molecular Biology*, **2002**, 48(1): 155-171.

Foxall, P. J. D., Spraul, M., Farrant, R. D., Lindon, L. C., Neild, G. H. and Nicholson, J. K.; 750 MHz ^{1}H-NMR spectroscopy of human blood plasma. *Journal of Pharmaceutical and Biomedical Analysis*, **1993**, 11(4-5): 267-276.

Gambone, J. C., Moore, J. G. and Koos, B. J.; *Common medical and surgical conditions complicating pregnancy*. *Essentials of obstetrics and gynecology*. N. F. Hacker, J. G. Moore and J. C. Gambone. Philadelphia, Elsevier, 2004**:** 216-246.

Gao, T., Zablith, N., Burns, D. H., Koski, K. G. and Skinner, C. D.; Identification and quantitation of human amniotic fluid components using capillary zone electrophoresis. *Analytical Biochemistry*, **2009**, 388(1): 155-157.

Gao, T., Zablith, N. R., Burns, D. H., Skinner, C. D. and Koski, K. G.; Second trimester amniotic fluid transferrin and uric acid predict infant birth outcomes. *Prenatal Diagnosis*, **2008**, 28: 810–814.

Garcia-Perez, I., Earll, M. E., Angulo, S., Barbas, C. and Legido-Quigley, C.; Chemometric analysis of urine fingerprints acquired by liquid chromatography-mass spectrometry and capillary electrophoresis: Application to the schistosomiasis mouse model. *Electrophoresis*, **2010**, 31(14): 2349-2355.

Georgiou, H., Lappas, M., Georgiou, G., Marita, A., Bryant, V., Hiscock, R., Permezel, M., Khalil, Z. and Rice, G.; Screening for biomarkers predictive of gestational diabetes mellitus. *Acta Diabetologica*, **2008**, 45(3): 157-165.

Ghanta, S., Mitchell, M. E., Ames, M., Hidestrand, M., Simpson, P., Goetsch, M., Thilly, W. G., Struble, C. A. and Tomita-Mitchell, A.; Non-invasive prenatal detection of trisomy 21 using tandem single nucleotide polymorphisms. *PLoS ONE*, **2010**, 5(10): e13184.

Gika, H. G., Theodoridis, G. A. and Wilson, I. D.; Hydrophilic interaction and reversed-phase ultra-performance liquid chromatography TOF-MS for metabonomic analysis of Zucker rat urine. *Journal of Separation Science*, **2008**, 31(9): 1598-1608.

Gil, A. M., Duarte, I. F., Godejohann, M., Braumann, U., Maraschin, M. and Spraul, M.; Characterization of the aromatic composition of some liquid foods by nuclear magnetic resonance spectrometry and liquid chromatography with nuclear magnetic resonance and mass spectrometric detection. *Analytica Chimica Acta*, **2003**, 488(1): 35-51.

Glew, R. H., Melah, G., El-Nafaty, A. I., Brandt, Y., Morris, D. and VanderJagt, D. J.; Plasma and urinary free amino acid concentrations in preeclamptic women in northern Nigeria. *Clinica Chimica Acta*, **2004**, 342(1–2): 179-185.

Godejohann, M.; Hydrophilic interaction chromatography coupled to nuclear magnetic resonance spectroscopy and mass spectroscopy-A new approach for the separation and identification of extremely polar analytes in bodyfluids. *Journal of Chromatography A*, **2007**, 1156(1-2): 87-93.

Goldenberg, R. L., Culhane, J. F., Iams, J. D. and Romero, R.; Epidemiology and causes of preterm birth. *The Lancet*, **2008**, 371(9606): 75-84.

Goldenberg, R. L., Goepfert, A. R. and Ramsey, P. S.; Biochemical markers for the prediction of preterm birth. *American Journal of Obstetrics and Gynecology*, **2005**, 192(5): Supplement 1: S36-S46.

Graça, G., Duarte, I. F., Barros, A. S., Goodfellow, B. J., Diaz, S., Carreira, I. M., Couceiro, A. B., Galhano, E. and Gil, A. M.; ^{1}H NMR based metabonomics of human amniotic fluid for the metabolic characterization of fetus malformations. *Journal of Proteome Research*, **2009**, 8(8): 4144-4150.

Graça, G., Duarte, I. F., Barros, A. S., Goodfellow, B. J., Diaz, S. O., Pinto, J., Carreira, I. M., Galhano, E. l., Pita, C. and Gil, A. M.; Impact of prenatal disorders on the metabolic profile of second trimester amniotic fluid: a nuclear magnetic resonance metabonomic study. *Journal of Proteome Research*, **2010**, 9(11): 6016-6024.

Graça, G., Duarte, I. F., Goodfellow, B. J., Barros, A. S., Carreira, I. M., Couceiro, A. B., Spraul, M. and Gil, A. M.; Potential of NMR spectroscopy for the study of human amniotic fluid. *Analytical Chemistry*, **2007**, 79(21): 8367-8375.

Graça, G., Duarte, I. F., Goodfellow, B. J., Carreira, I. M., Couceiro, A. B., do Rosario Domingues, M., Spraul, M., Li-Hong, T. and Gil, A. M.; Metabolite profiling of human amniotic fluid by hyphenated nuclear magnetic resonance spectroscopy. *Analytical Chemistry*, **2008**, 80(15): 6085-6092.

Graça, G., Goodfellow, B. J., Barros, A. S., Diaz, S., Duarte, I. F., Spagou, K., Veselkov, K., Want, E. J., Lindon, J. C., Carreira, I. M., Galhano, E., Pita, C. and Gil, A. M.; UPLC-MS metabolic profiling of second trimester amniotic fluid and maternal urine and comparison with NMR spectral profiling for the identification of pregnancy disorder biomarkers. *Molecular BioSystems*, **2012**, 8: 1243-1254.

Graça, G., Moreira, A. S., Correia, A. J. V., Goodfellow, B. J., Barros, A. S., Duarte, I. F., Carreira, I. M., Galhano, E., Pita, C., Almeida, M. d. C. and Gil, A. M.; Mid-infrared (MIR) metabolic fingerprinting of amniotic fluid: A possible avenue for early diagnosis of prenatal disorders? *Analytica Chimica Acta*, **2013**, 764: 24-31.

Graça, L. M.; *Desenvolvimento e processo maturativo do feto*. *Medicina materno-fetal*. Lisboa, Lidel, 2005a**:** 16-33.

Graça, L. M.; *Parto pré-termo*. *Medicina materno-fetal*. L. M. Graça. Lisboa, Lidel, 2005b**:** 387-409.

Graça, L. M.; *Ruptura permatura de membranas*. *Medicina materno-fetal*. L. M. Graça. Lisboa, Lidel, 2005c**:** 410-420.

Graça, L. M. and Barros, J. J. S.; *Placenta e anexos fetais*. *Medicina materno-fetal*. Lisboa, Lidel, 2005**:** 16-33.

Gracie, S., Penne, C., Ekman-Ordeberg, G., Lye, S., McManaman, J., Williams, S., Palmer, L., Kelley, M., Menon, R. and Gravett, M.; An integrated systems biology approach to the study of preterm birth using "-omic" technology - a guideline for research. *BMC Pregnancy and Childbirth*, **2012**, 11(71): doi:10.1186/1471-2393-1111-1171.

Griffin, J. L. and Vidal-Puig, A.; Current challenges in metabolomics for diabetes research: a vital functional genomic tool or just a ploy for gaining funding? *Physiological Genomics*, **2008**, 34(1): 1-5.

Griffiths, W. J., Wang, Y., Karu, K., Samuel, E., McDonnell, S., Hornshaw, M. and Shackleton, C.; Potential of sterol analysis by liquid chromatography-tandem mass spectrometry for the prenatal diagnosis of Smith-Lemli-Opitz syndrome. *Clinical Chemistry*, **2008**, 54(8): 1317-1324.

Groenen, P. M. W., Engelke, U. F., Wevers, R. A., Hendriks, J. C. M., Eskes, T., Merkus, H. and Steegers-Theunissen, R. P. M.; High-resolution ^{1}H NMR spectroscopy of amniotic fluids from spina bifida fetuses and controls. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, **2004**, 112(1): 16-23.

Groenen, P. M. W., Wevers, R. A., Janssen, F. S. M., Tuerlings, J. H. A. M., Merkus, J. M. W. M. and Steegers-Theunissen, R. P. M.; Are myo-inositol, glucose and zinc concentrations in amniotic fluid of fetuses with spina bifida different from controls? *Early Human Development*, **2003**, 71: 1-8.

Gross, T. L., Sokol, R. J., Wilson, M. V., Kuhnert, P. M. and Hirsch, V.; Amniotic fluid phosphatidylglycerol: a potentially useful predictor of intrauterine growth retardation. *American Journal of Obstetrics and Gynecology*, **1982**, 140(3): 277-281.

Haddow, J. E., Palomaki, G. E., Knight, G. J., Williams, J., Pulkkinen, A., Canick, J. A., Saller, D. N. and Bowers, G. B.; Prenatal screening for Down's syndrome with use of maternal serum markers. *New England Journal of Medicine*, **1992**, 327(9): 588-593.

Harwood, L. M. and Claridge, T. D. W.; *Introduction to organic spectroscopy*, Oxford University Press, 1999.

Heazell, A. E. P., Bernatavicius, G., Warrander, L., Brown, M. C. and Dunn, W. B.; A Metabolomic approach identifies differences in maternal serum in third trimester pregnancies that end in poor perinatal outcome. *Reproductive Sciences*, **2012**, 19(8): 863-875.

Heazell, A. E. P., Brown, M., Worton, S. A. and Dunn, W. B.; Review: The effects of oxygen on normal and pre-eclamptic placental tissue – insights from metabolomics. *Placenta*, **2011**, 32, Supplement 2: S119-S124.

Hicks, R. P., Young, J. K. and Moskau, D.; Magnetization transfer via isotropic mixing: An introduction to the HOHAHA experiment. *Concepts in Magnetic Resonance*, **1994**, 6(2): 115-130.

Hobel, C. J.; *Obstetrical complications: preterm labor, PROM, IUGR, postterm pregnancy, and IUFD*. *Essentials of obstetrics and gynecology*. N. F. Hacker, J. G. Moore and J. C. Gambone. Philadelphia, Elsevier, 2004**:** 167-182.

Hod, M., Jovanovic, L., Di Renzo, G. C., Leiva, A. d. and Langer, O.; *Textbook of diabetes and pregnancy*. London, Informa Healthcare, 2008.

Holmes, E., Loo, R. L., Stamler, J., Bictash, M., Yap, I. K. S., Chan, Q., Ebbels, T., De Iorio, M., Brown, I. J., Veselkov, K. A., Daviglus, M. L., Kesteloot, H., Ueshima, H., Zhao, L. C., Nicholson, J. K. and Elliott, P.; Human metabolic phenotype diversity and its association with diet and blood pressure. *Nature*, **2008**, 453(7193): 396-400.

Hore, P. J.; *NMR Principles*. *Encyclopedia of spectroscopy and spectrometry*. J. C. Lindon, J. Holmes and G. Tranter. London, Academic Press, 2000, 2**:** 1545-1553.

Horgan, R. P., Broadhurst, D. I., Dunn, W. B., Brown, M., Heazell, A. E. P., Kell, D. B., Baker, P. N. and Kenny, L. C.; Changes in the metabolic footprint of placental explant-conditioned medium cultured in different oxygen tensions from placentas of small for gestational age and normal pregnancies. *Placenta*, **2010**, 31(10): 893-901.

Horgan, R. P., Clancy, O. H., Myers, J. E. and Baker, P. N.; An overview of proteomic and metabolomic technologies and their application to pregnancy research. *BJOG: An International Journal of Obstetrics & Gynaecology*, **2009**, 116(2): 173-181.

Hsu, C.-S., Chou, S.-Y., Liang, S.-J., Chang, C.-Y., Yeh, C.-L. and Yeh, S.-L.; Effect of glutamine on cell adhesion molecule expression and leukocyte transmigration in endothelial cells stimulated by preeclamptic plasma. *Nutrition*, **2005**, 21(11–12): 1134-1140.

Jaroszewski, J.; Hyphenated NMR methods in natural products research, Part 2: HPLC-SPE-NMR and other new trends in NMR hyphenation. *Planta Medica*, **2005**, (9): 795-802.

Jauniaux, E. and Gulbis, B.; Fluid compartments of the embryonic environment. *Human Reproduction Update*, **2000**, 6(3): 268-278.

Jauniaux, E., Gulbis, B. and Gerloo, E.; Free amino acids in human fetal liver and fluids at 12–17 weeks of gestation. *Human Reproduction*, **1999**, 14(6): 1638-1641.

Jauniaux, E., Hempstock, J., Teng, C., Battaglia, F. C. and Burton, G. J.; Polyol concentrations in the fluid compartments of the human conceptus during the first trimester of pregnancy: maintenance of redox potential in a low oxygen environment. *Journal of Clinical Endocrinology & Metabolism*, **2005**, 90(2): 1171-1175.

Joe, B. N., Vahidi, K., Zektzer, A., Chen, M.-H., Clifton, M. S., Butler, T., Keshari, K., Kurhanewicz, J., Coakley, F. and Swanson, M. G.; 1H HR-MAS spectroscopy for quantitative measurement of choline concentration in amniotic fluid as a marker of fetal lung maturity: Inter- and intraobserver reproducibility study. *Journal of Magnetic Resonance Imaging*, **2008**, 28(6): 1540-1545.

Johnson, C. S.; Diffusion ordered nuclear magnetic resonance spectroscopy: principles and applications. *Progress in Nuclear Magnetic Resonance Spectroscopy*, **1999**, 34(3): 203-256.

Joong Shin, P., Oh, K.-J., Norwitz, E. R., Han, J.-S., Choi, H.-J., Hyo Suk, S., Yoon Dan, K., Park, C.-W., Byoung Jae, K., Jong Kwan, J. and Hee Chul, S.; Identification of proteomic biomarkers of preeclampsia in amniotic fluid using SELDI-TOF mass spectrometry. *Reproductive Sciences*, **2008**, 15(5): 457-468.

Kadrofske, M. M., Parimi, P. S., Gruca, L. L. and Kalhan, S. C.; Effect of intravenous amino acids on glutamine and protein kinetics in low-birth-weight preterm infants during the immediate neonatal period. *American Journal of Physiology - Endocrinology And Metabolism*, **2006**, 290(4): E622-E630.

Kang, S.-M., Park, J.-C., Shin, M.-J., Lee, H., Oh, J., Ryu, D. H., Hwang, G.-S. and Chung, J. H.; 1H nuclear magnetic resonance based metabolic urinary profiling of patients with ischemic heart failure. *Clinical Biochemistry*, **2011**, 44(4): 293-299.

Katajamaa, M., Miettinen, J. and Orešič, M.; MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data. *Bioinformatics*, **2006**, 22(5): 634-636.

Kebarle, P. and Verkerk, U. H.; Electrospray: from ions in solution to ions in the gas phase, what we know now. *Mass Spectrometry Reviews*, **2009**, 28(6): 898-917.

Keeler, J.; *Understanding NMR spectroscopy*. Cambridge, John Wiley & Sons, Ltd, 2002.

Keller, P.; *Basic principles of MR imaging*. Milwaukee, GE Medical Systems, 1988.

Kenny, L., Dunn, W., Ellis, D., Myers, J., Baker, P. and Kell, D.; Novel biomarkers for pre-eclampsia detected using metabolomics and machine learning. *Metabolomics*, **2005**, 1(3): 227-234.

Kenny, L. C., Broadhurst, D., Brown, M., Dunn, W. B., Redman, C. W. G., Kell, D. B. and Baker, P. N.; Detection and identification of novel metabolomic biomarkers in preeclampsia. *Reproductive Sciences*, **2008**, 15(6): 591-597.

Kenny, L. C., Broadhurst, D. I., Dunn, W., Brown, M., North, R. A., McCowan, L., Roberts, C., Cooper, G. J. S., Kell, D. B., Baker, P. N. and on behalf of the Screening for Pregnancy Endpoints Consortium; Robust early pregnancy prediction of later preeclampsia using metabolomic biomarkers. *Hypertension*, **2010**, 56(4): 741-749.

Kerssenbaum, R.; *DOSY and diffusion by NMR*. Rheinstetten, Germany, Bruker BioSpin GmbH, 2006.

Keun, H. C., Athersuch, T. J., Beckonert, O., Wang, Y., Saric, J., Shockcor, J. P., Lindon, J. C., Wilson, I. D., Holmes, E. and Nicholson, J. K.; Heteronuclear ^{19}F–^{1}H statistical total correlation spectroscopy as a tool in drug metabolism: study of flucloxacillin biotransformation. *Analytical Chemistry*, **2008**, 80(4): 1073-1079.

Khaustova, S., Davydov, I., Trushkin, E., Shkurnikov, M., Mueller, R., Backhaus, J. and Tonevitsky, A.; Application of mid-infrared molecular spectroscopy for assessment of biochemical parameters of blood serum. *Bulletin of Experimental Biology and Medicine*, **2009**, 148(6): 943-947.

Kim, D. H., Vahidi, K., Caughey, A. B., Coakley, F. V., Vigneron, D. B., Kurhanewicz, J., Mow, B. and Joe, B. N.; In vivo 1H magnetic resonance spectroscopy of amniotic fluid and fetal lung at 1.5 T: Technical challenges. *Journal of Magnetic Resonance Imaging*, **2008**, 28(4): 1033-1038.

Kind, T., Tolstikov, V., Fiehn, O. and Weiss, R. H.; A comprehensive urinary metabolomic approach for identifying kidney cancer. *Analytical Biochemistry*, **2007**, 363(2): 185-195.

Kirshon, B., Mari, G. and Moise, K. J., Jr.; Indomethacin therapy in the treatment of symptomatic polyhydramnios. *Obstetrics & Gynecology*, **1990**, 75(2): 202-205.

Kobayashi, H., Kobayashi, M., Arai, T., Kikuchi, M. and Okino, H.; Study on a new index of fetal respiratory function: a novel optical analysis of the amniotic fluid by the vacuum-dried infrared attenuated total reflection (ATR) method. *Tokai Journal Experimental Clinical Medicine*, **1990**, 15(6): 425-433.

Kohl, S., Klein, M., Hochrein, J., Oefner, P., Spang, R. and Gronwald, W.; State-of-the art data normalization methods improve NMR-based metabolomic analysis. *Metabolomics*, **2012**, 8: 146-160.

Kolialexi, A., Tounta, G., Mavrou, A. and Tsangaris, G. T.; Proteomic analysis of amniotic fluid for the diagnosis of fetal aneuploidies. *Expert Review of Proteomics*, **2011**, 8(2): 175-185.

Kumps, A., Vamos, E., Mardens, Y., Abramowicz, M., Genin, J. and Duez, P.; Assessment of an electron-impact GC-MS method for organic acids and glycine conjugates in amniotic fluid. *Journal of Inherited Metabolic Disease*, **2004**, 27(5): 567-579.

Lauridsen, M., Hansen, S. H., Jaroszewski, J. W. and Cornett, C.; Human urine as test material in ^{1}H NMR-based metabonomics: recommendations for sample preparation and storage. *Analytical Chemistry*, **2007**, 79(3): 1181-1186.

Lavee, M., Goldman, S., Daniel-Spiegel, E. and Shalev, E.; Matrix metalloproteinase-2 is elevated in midtrimester amniotic fluid prior to the development of preeclampsia. *Reproductive Biology and Endocrinology*, **2009**, 7(85): doi:10.1186/1477-7827-1187-1185.

Le Moyec, L., Muller, F., Eugene, M. and Spraul, M.; Proton magnetic resonance spectroscopy of human amniotic fluids sampled at 17-18 weeks of pregnancy in cases of decreased digestive enzyme activities and detected cystic fibrosis. *Clinical Biochemistry*, **1994**, 27(6): 475-483.

Lind, T.; *Amniotic fluid*. *Geigy scientific tables*. C. Lentner. Basle, Switzerland, Ciba-Geigy, 1981, 1**:** 197-211.

Lindon, J. C., Holmes, E., Bollard, M. E., Stanley, E. G. and Nicholson, J. K.; Metabonomics technologies and their applications in physiological monitoring, drug safety assessment and disease diagnosis. *Biomarkers*, **2004**, 9(1): 1-31.

Lindon, J. C., Holmes, E. and Nicholson, J. K.; Metabonomics techniques and applications to pharmaceutical research & development. *Pharmaceutical Research*, **2006**, 23(6): 1075 -1088.

Lindon, J. C. and Nicholson, J. K.; Analytical technologies for metabonomics and metabolomics, and multi-omic information recovery. *Trac-Trends in Analytical Chemistry*, **2008a**, 27(3): 194-204.

Lindon, J. C. and Nicholson, J. K.; Spectroscopic and statistical techniques for information recovery in metabonomics and metabolomics. *Annual Review of Analytical Chemistry*, **2008b**, 1(1): 45-69.

Lindon, J. C., Nicholson, J. K. and Everett, J. R.; NMR spectroscopy of biofluids. *Annual Reports On NMR Spectroscopy*, **1999**, 38: 1-88.

Lindon, J. C., Nicholson, J. K., Holmes, E. and Everett, J. R.; Metabonomics: metabolic processes studied by NMR spectroscopy of biofluids. *Concepts in Magnetic Resonance*, **2000a**, 12(5): 289-320.

Lindon, J. C., Nicholson, J. K. and Wilson, I. D.; Directly coupled HPLC-NMR and HPLC-NMR-MS in pharmaceutical research and development. *Journal of Chromatography B: Biomedical Sciences and Applications*, **2000b**, 748(1): 233-258.

Liu, K.-Z., Dembinski, T., C. and Mantsch, H., H. ; Prediction of RDS from amniotic fluid analysis: a comparison of the prognostic value of TLC and infra-red spectroscopy. *Prenatal Diagnosis*, **1998a**, 18(12): 1267-1275.

Liu, K.-Z., Dembinski, T. C. and Mantsch, H. H.; Rapid determination of fetal lung maturity from infrared spectra of amniotic fluid. *American Journal of Obstetrics and Gynecology*, **1998b**, 178(2): 234-241.

Liu, K.-Z. and Mantsch, H. H.; Simultaneous quantitation from infrared spectra of glucose concentrations, lactate concentrations, and lecithin/sphingomyelin ratios in amniotic fluid. *American Journal of Obstetrics and Gynecology*, **1999**, 180(3): 696-702.

Liu, K.-Z., Shaw, R. A., Dembinski, T. C., Reid, G. J., Ying, S. L. and Mantsch, H. H.; Comparison of infrared spectroscopic and fluorescence depolarization assays for fetal lung maturity. *American Journal of Obstetrics and Gynecology*, **2000**, 183(1): 181-187.

Liu, K.-Z., Shaw, R. A., Man, A., Dembinski, T. C. and Mantsch, H. H.; Reagent-free, simultaneous determination of serum cholesterol in HDL and LDL by infrared spectroscopy. *Clinical Chemistry*, **2002**, 48(3): 499-506.

Liu, K. Z., Ahmed, M. K., Dembinski, T. C. and Mantsch, H. H.; Prediction of fetal lung maturity from near-infrared spectra of amniotic fluid. *International Journal of Gynecology & Obstetrics*, **1997**, 57(2): 161-168.

Longini, M., Perrone, S., Kenanidis, A., Vezzosi, P., Marzocchi, B., Petraglia, F., Centini, G. and Buonocore, G.; Isoprostanes in amniotic fluid: a predictive marker for fetal growth restriction in pregnancy. *Free Radical Biology and Medicine*, **2005**, 38(11): 1537-1541.

Lu, M. C. and Hobel, C. J.; *Antepartum care: preconception and prenatal care, genetic evaluation and teratology, and antenatal fetal assessment*. *Essentials of obstetrics and gynecology*. N. F. Hacker, J. G. Moore and J. C. Gambone. Philadelphia, Elsevier, 2004**:** 83-103.

Ludwig, C. and Viant, M. R.; Two-dimensional J-resolved NMR spectroscopy: review of a key methodology in the metabolomics toolbox. *Phytochemical Analysis*, **2010**, 21(1): 22-32.

Maher, A. D., Cysique, L. A., Brew, B. J. and Rae, C. D.; Statistical integration of ^{1}H NMR and MRS data from different biofluids and tissues enhances recovery of biological information from individuals with HIV-1 infection. *Journal of Proteome Research*, **2011**, 10(4): 1737-1745.

Majer, S., Bauer, M., Magnet, E., Strele, A., Giegerl, E., Eder, M., Lang, U. and Pertl, B.; Maternal urine for prenatal diagnosis - an analysis of cell-free fetal DNA in maternal urine and plasma in the third trimester. *Prenatal Diagnosis*, **2007**, 27(13): 1219-1223.

Mann, H. B. and Whitney, D. R.; On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics* **1947**, 18(1): 50-60.

Marconi, A. M. and Paolini, C. L.; Nutrient transport across the intrauterine growth-restricted placenta. *Seminars in perinatology*, **2008**, 32(3): 178-181.

Marcos, J., Craig, W. Y., Palomaki, G. E., Kloza, E. M., Haddow, J. E., Roberson, M., Bradley, L. A. and Shackleton, C. H. L.; Maternal urine and serum steroid measurements to identify steroid sulfatase deficiency (STSD) in second trimester pregnancies. *Prenatal Diagnosis*, **2009**, 29(8): 771-780.

Martin, T. C., Moecks, J., Belooussov, A., Cawthraw, S., Dolenko, B., Eiden, M., von Frese, J., Kohler, W., Schmitt, J., Somorjai, R., Udelhoven, T., Verzakov, S. and Petrich, W.; Classification of signatures of bovine spongiform encephalopathy in serum using infrared spectroscopy. *Analyst*, **2004**, 129(10): 897-901.

Masson, P., Spagou, K., Nicholson, J. K. and Want, E. J.; Technical and biological variation in UPLC−MS-based untargeted metabolic profiling of liver extracts: application in an experimental toxicity study on galactosamine. *Analytical Chemistry*, **2011**, 83(3): 1116-1123.

Mavrou, A., Anagnostopoulos, A. K., Kolialexi, A., Vougas, K., Papantoniou, N., Antsaklis, A., Fountoulakis, M. and Tsangaris, G. T.; Proteomic analysis of amniotic fluid in pregnancies with Turner syndrome fetuses. *Journal of Proteome Research*, **2008**, 7(5): 1862-1866.

Mayer, M. and Meyer, B.; Characterization of ligand binding by saturation transfer difference NMR Spectroscopy. *Angewandte Chemie International Edition*, **1999**, 38(12): 1784-1788.

McGowan, P. E., Lawrie, W. C., Reglinski, J., Spickett, C. M., Wilson, R., Walker, J. J., Wisdom, S. and Maclean, M. A.; ^{1}H NMR as a non-invasive probe of amniotic fluid in insulin dependant diabetes mellitus. *Journal of Perinatal Medicine*, **1999**, 27(5): 404-408.

McGowan, P. E., Reglinski, J., Wilson, R., Walker, J. J., Wisdom, S. and McKillop, J. H.; Quantitative ^{1}H NMR analysis of amniotic-fluid. *Journal of Pharmaceutical and Biomedical Analysis*, **1993**, 11(8): 629-632.

Meiboom, S. and Gill, D.; Modified spin-echo method for measuring nuclear relaxation times. *Review of Scientific Instruments*, **1958**, 29(8): 688-691.

Metzger, B. E. and Coustan, D. R.; Proceedings of the fourth international work-shop-conference on gestational diabetes mellitus. *Diabetes Care*, **1998**, 21(Suppl. 2): B1–B167.

Miall, L., Rudolf, M. and Levene, M. I.; *Paediatrics at a glance*. Oxford, Blackwell Science, 2003.

Minai, L., Martinovic, J., Chretien, D., Dumez, F., Razavi, F., Munnich, A. and Rötig, A.; Mitochondrial respiratory chain complex assembly and function during human fetal development. *Molecular Genetics and Metabolism*, **2008**, 94(1): 120-126.

Moore, K. L. and Persaud, T. V. N.; Eds. *The developing human - Clinically oriented embryology*. Philadelphia, Saunders, 2008.

Moore, R. M., Mansour, J. M., Redline, R. W., Mercer, B. M. and Moore, J. J.; The physiology of fetal membrane rupture: insight gained from the determination of physical properties. *Placenta*, **2006**, 27(11–12): 1037-1051.

Murray, R. K., Granner, D. K., Mayes, P. A. and Rodwell, V. W.; *Harper's illustrated biochemistry*. New York, Lange Medical Books/McGraw-Hill, 2003.

Nath, A. K., Krauthammer, M., Li, P., Davidov, E., Butler, L. C., Copel, J., Katajamaa, M., Oresic, M., Buhimschi, I., Buhimschi, C., Snyder, M. and Madri, J. A.; Proteomic-based detection of a protein cluster dysregulated during cardiovascular development identifies biomarkers of congenital heart defects. *PLoS ONE*, **2009**, 4(1): e4221.

Nelson, T. R., Gillies, R. J., Powell, D. A., Schrader, M. C., Manchesters, D. K. and Pretorius, D. H.; High resolution proton NMR spectroscopy of human amniotic fluid. *Prenatal Diagnosis*, **1987**, 7(5): 363-372.

Ng, K. J., Andresen, B. D., Bianchine, J. R., Iams, J. D., O'Shaugnessy, R. W., Stempel, L. E. and Zuspan, F. P.; Capillary gas chromatographic--mass spectrometric profiles of trimethylsilyl derivatives of organic acids from amniotic fluids of different gestational age. *Journal of Chromatography B: Biomedical Sciences and Applications*, **1982**, 228: 43-50.

Nicholson, J. K., Lindon, J. C. and Holmes, E.; 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica*, **1999**, 29(11): 1181-1189.

Nilsson, M., Duarte, I. F., Almeida, C., Delgadillo, I., Goodfellow, B. J., Gil, A. M. and Morris, G. A.; High-resolution NMR and diffusion-ordered spectroscopy of Port wine. *Journal of Agricultural and Food Chemistry*, **2004**, 52(12): 3736-3743.

Noda, I.; Two-dimensional infrared spectroscopy. *Journal of the American Chemical Society*, **1989**, 111(21): 8116-8118.

Odibo, A. O., Goetzinger, K. R., Odibo, L., Cahill, A. G., Macones, G. A., Nelson, D. M. and Dietzen, D. J.; First-trimester prediction of preeclampsia using metabolomic biomarkers: a discovery phase study. *Prenatal Diagnosis*, **2011**, 31(10): 990-994.

Ogihara, T., Kim, H. S., Hirano, K., Imanishi, M., Ogihara, H., Tamai, H., Okamoto, R. and Mino, M.; Oxidation products of uric acid and ascorbic acid in preterm infants with chronic lung disease. *Neonatology*, **1998**, 73(1): 24-33.

Oh, K. J., Park, J. S., Norwitz, E. R., Kim, S. M., Kim, B. J., Park, C.-W., Jun, J. K. and Syn, H. C.; Proteomic biomarkers in second trimester amniotic fluid that identify women who are destined to develop preeclampsia. *Reproductive Sciences*, **2012**, 19(7): 694-703.

Ottolenghi, C., Abermil, N., Lescoat, A., Aupetit, J., Beaugendre, O., Morichon-Delvallez, N., Ricquier, D., Chadefaux-Vekemans, B. and Rabier, D.; Gestational age-related reference values for amniotic fluid organic acids. *Prenatal Diagnosis*, **2010**, 30(1): 43-48.

Paine, M. A., Scioscia, M., Williams, P. J., Gumaa, K., Rodeck, C. H. and Rademacher, T. W.; Urinary inositol phosphoglycan P-type as a marker for prediction of preeclampsia and novel implications for the pathophysiology of this disorder. *Hypertension in Pregnancy*, **2010**, 29(04): 375-384.

Pearce, J. M., Krone, J. T., Pappas, A. A. and Komoroski, R. A.; Analysis of saturated phosphatidylcholine in amniotic fluid by ^{31}P NMR. *Magnetic Resonance in Medicine*, **1993**, 30(4): 476-484.

Pearce, J. M., Shifman, M. A., Pappas, A. A. and Komoroski, R. A.; Analysis of phospholipids in human amniotic fluid by ^{31}P NMR. *Magnetic Resonance in Medicine*, **1991**, 21(1): 107-116.

Petrich, W., Lewandrowski, K. B., Muhlestein, J. B., Hammond, M. E. H., Januzzi, J. L., Lewandrowski, E. L., Pearson, R. R., Dolenko, B., Fruh, J., Haass, M., Hirschl, M. M., Kohler, W., Mischler, R., Mocks, J., Ordonez-Llanos, J., Quarder, O., Somorjai, R., Staib, A., Sylven, C., Werner, G. and Zerback, R.; Potential of mid-infrared spectroscopy to aid the triage of patients with acute chest pain. *Analyst*, **2009**, 134(6): 1092-1098.

Power, K. M., Sanchez-Galan, J. E., Luskey, G. W., Koski, K. G. and Burns, D. H.; Use of near-infrared spectroscopic analysis of second trimester amniotic fluid to assess preterm births. *Journal of Pregnancy*, **2011**: doi:10.1155/2011/980985.

Prasad, A. N., Malinger, G. and Lerman-Sagie, T.; Primary disorders of metabolism and disturbed fetal brain development. *Clinics in Perinatology*, **2009**, 36(3): 621–638.

Pretsch, E., Clerc, T., Seibl, J. and Simon, W.; *Tables of Spectral data for structure determination of organic compounds*. Berlin, Springer-Verlag, 1989.

Raijmakers, M. T. M., Dechend, R. and Poston, L.; Oxidative stress and preeclampsia. Rationale for antioxidant clinical trials. *Hypertension*, **2004**, 44: 374-380.

Ramón Y Cajal, C. L., Martínez, R. O., Naveira, E. C. and Martínez, M.; Amino acids in amniotic fluid in the 15th–16th weeks of gestation and preterm labor. *Journal of Maternal-Fetal and Neonatal Medicine*, **2007**, 20(3): 225-231.

Reynolds, W. F.; *NMR pulse sequences*. *Encyclopedia of spectroscopy and spectrometry*. J. C. Lindon, J. Holmes and G. Tranter. London, Academic Press, 2000, 2**:** 1554-1567.

Rezzi, S., Ramadan, Z., Fay, L. B. and Kochhar, S.; Nutritional metabonomics: Applications and perspectives. *Journal of Proteome Research*, **2007**, 6(2): 513-525.

Rinaldi, P. L.; *Two-Dimensional NMR, Methods*. *Encyclopedia of spectroscopy and spectrometry*. J. C. Lindon, J. Holmes and G. Tranter. London, Academic Press, 2000, 3**:** 2370-2381.

Rocha, S. M., Caldeira, M., Carrola, J., Santos, M., Cruz, N. and Duarte, I. F.; Exploring the human urine metabolomic potentialities by comprehensive two-dimensional gas chromatography coupled to time of flight mass spectrometry. *Journal of Chromatography A*, **2012**, 1252(0): 155-163.

Roman, A., Koklanaris, N., Paidas, M., Mulholland, J., Levitz, M. and Rebarber, A.; "Blind" vaginal fetal fibronectin as a predictor of spontaneous preterm delivery. *Obstetrics and Gynecology*, **2005**, 105(2): 285-289.

Romero, R., Espinoza, J., Rogers, W. T., Moser, A., Kae nien, J., Pedro kusanovic, J., Gotsch, F., Erez, O., Gomez, R., Edwin, S. and Hassan, S. S.; Proteomic analysis of amniotic fluid to identify women with preterm labor and intra-amniotic inflammation/infection: The use of a novel computational method to analyze mass spectrometric profiling. *The Journal of Maternal-Fetal & Neonatal Medicine*, **2008**, 21(6): 367-387.

Romero, R., Mazaki-Tovi, S., Vaisbuch, E., Kusanovic, J. P., Chaiworapongsa, T., Gomez, R., Nien, J. K., Yoon, B. H., Mazor, M., Luo, J., Banks, D., Ryals, J. and Beecher, C.; Metabolomics in premature labor: a novel approach to identify patients at risk for preterm delivery. *Journal of Maternal-Fetal and Neonatal Medicine*, **2010**, 23(12): 1344-1359.

Šašić, S., Muszynski, A. and Ozaki, Y.; A New Possibility of the Generalized two-dimensional correlation spectroscopy. 1. Sample−sample correlation spectroscopy. *The Journal of Physical Chemistry A*, **2000**, 104(27): 6380-6387.

Saude, E. and Sykes, B.; Urine stability for metabolomic studies: effects of preparation and storage. *Metabolomics*, **2007**, 3(1): 19-27.

Savorani, F., Tomasi, G. and Engelsen, S. B.; icoshift: A versatile tool for the rapid alignment of 1D NMR spectra. *Journal of Magnetic Resonance*, **2010**, 202(2): 190-202.

Scioscia, M., Gumaa, K., Whitten, M., Selvaggi, L. E., Rodeck, C. H. and Rademacher, T. W.; Inositol phosphoglycan P-type in healthy and preeclamptic pregnancies. *Journal of Reproductive Immunology*, **2007a**, 76(1–2): 85-90.

Scioscia, M., Kunjara, S., Gumaa, K., McLean, P., Rodeck, C. H. and Rademacher, T. W.; Urinary excretion of inositol phosphoglycan P-type in gestational diabetes mellitus. *Diabetic Medicine*, **2007b**, 24(11): 1300-1304.

Shao, J.; Linear model selection by cross-validation. *Journal of the American Statistical Association*, **1993**, 88(422): 486-494.

Shestopalov, A., Shkurat, T., Mikashinovich, Z., Kryzhanovskaya, I., Bogacheva, M., Lomteva, S., Prokof'ev, V. and Gus'kov, E.; Biological functions of allantoin. *Biology Bulletin*, **2006**, 33(5): 437-440.

Sims, C. J., Burholt, D. R., Fujito, D. T., Ketterer, D. M. and Wilkinson, D. A.; Proton magnetic-resonance spectroscopic examination of human amniotic-fluid from trisomy-21 pregnancies. *American Journal of Human Genetics*, **1993a**, 53(3): 1458-1458 Suppl. S

Sims, C. J., Fujito, D. T., Burholt, D. R., Dadok, J., Giles, H. R. and Wilkinson, D. A.; Quantification of human amniotic-fluid constituents by high-resolution proton nuclear-magnetic-resonance (NMR) spectroscopy. *Prenatal Diagnosis*, **1993b**, 13(6): 473-480.

Sims, C. J., Fujito, D. T., Burholt, D. R., Dadok, J. and Wilkinson, D. A.; Comparison of metabolite levels in second and third trimester human amniotic fluid samples using proton magnetic resonance spectroscopy. *Journal of Maternal-Fetal Investigation*, **1996**, 6(2): 62-66.

Sinclair, A. J., Viant, M. R., Ball, A. K., Burdon, M. A., Walker, E. A., Stewart, P. M., Rauz, S. and Young, S. P.; NMR-based metabolomic analysis of cerebrospinal fluid and serum in neurological diseases – a diagnostic tool? *NMR in Biomedicine*, **2010**, 23(2): 123-132.

Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R. and Siuzdak, G.; XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and Identification. *Analytical Chemistry*, **2006**, 78(3): 779-787.

Smith, L. M., Maher, A. D., Cloarec, O., Rantalainen, M., Tang, H., Elliott, P., Stamler, J., Lindon, J. C., Holmes, E. and Nicholson, J. K.; Statistical correlation and projection methods for improved information recovery from diffusion-edited NMR spectra of biological samples. *Analytical Chemistry*, **2007**, 79(15): 5682-5689.

Snoep, J. and Westerhoff, H.; *From isolation to integration, a systems biology approach for building the silicon cell systems biology*. L. Alberghina and H. V. Westerhoff. Berlin, Springer, 2005, 13**:** 13-30.

Solanky, K. S., Bailey, N. J. C., Holmes, E., Lindon, J. C., Davis, A. L., Mulder, T. P. J., Van Duynhoven, J. P. M. and Nicholson, J. K.; NMR-based metabonomic studies on the biochemical effects of epicatechin in the rat. *Journal of Agricultural and Food Chemistry*, **2003**, 51(14): 4139-4145.

Sommer, K. R., Hill, R. M. and Horning, M. G.; Identification and quantification of drugs in human amniotic fluid. *Research Communications in Chemical Pathology & Pharmacology*, **1975**, 12(3): 583-595.

Spencer, K.; Point-of-care screening for chromosomal anomalies in the first trimester of pregnancy. *Clinical Chemistry*, **2002**, 48(3): 403-404.

Spragg, R.; *IR spectrometers*. *Encyclopedia of spectroscopy and spectrometry*. J. C. Lindon, J. Holmes and G. Tranter. London, Academic Press, 2000, 2**:** 1048-1057.

Spraul, M., Freund, A. S., Nast, R. E., Withers, R. S., Maas, W. E. and Corcoran, O.; Advancing NMR sensitivity for LC-NMR-MS using a cryoflow probe: application to the analysis of acetaminophen metabolites in urine. *Analytical Chemistry*, **2003**, 75(6): 1536-1541.

Star, J., Canick, J. A., Palomaki, G. E., Saller, D. N., Sung, C. J., Tumber, M. B. and Coustan, D. R.; The Relationship Between Second-trimester amniotic fluid insulin and glucose levels and subsequent gestational diabetes. *Prenatal Diagnosis*, **1997**, 17(2): 149-154.

Stein, P. T., Scholl, T. O., Schluter, M. D., Leskiw, M. J., Chen, X., Spur, B. W. and Rodriguez, A.; Oxidative stress early in pregnancy and pregnancy outcome. *Free Radical Research*, **2008**, 42(10): 841-848.

Stewart, C. J., Iles, R. K. and Perrett, D.; The analysis of human amniotic fluid using capillary electrophoresis. *Electrophoresis*, **2001**, 22(6): 1136-1142.

Stuart, E. L., Evans, G. S., Lin, Y. S. and Powers, H. J.; Reduced collagen and ascorbic acid concentrations and increased proteolytic susceptibility with prelabor fetal membrane rupture in women. *Biology of Reproduction*, **2005**, 72(1): 230-235.

Tautenhahn, R., Bottcher, C. and Neumann, S.; Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinformatics*, **2008**, 9(504): doi:10.1186/1471-2105-1189-1504.

Tavares, P.; *Fundamentos de genética média*. *Medicina materno-fetal*. L. M. Graça. Lisboa, Lidel, 2005.

Tea, I., Le Gall, G., Küster, A., Guignard, N., Alexandre–Gouabau, M.-C., Darmaun, D. and Robins, R. J.; ^{1}H-NMR-based metabolic profiling of maternal and umbilical cord blood indicates altered materno-foetal nutrient exchange in preterm infants. *PLoS ONE*, **2012**, 7(1): e29947.

Teahan, O., Gamble, S., Holmes, E., Waxman, J., Nicholson, J. K., Bevan, C. and Keun, H. C.; Impact of analytical bias in metabonomic studies of human blood serum and plasma. *Analytical Chemistry*, **2006**, 78(13): 4307-4318.

Thadikkaran, L., Crettaz, D., Siegenthaler, M. A., Gallot, D., Sapin, V., Iozzo, R. V., Queloz, P.-A., Schneider, P. and Tissot, J.-D.; The role of proteomics in the assessment of premature rupture of fetal membranes. *Clinica Chimica Acta*, **2005**, 360(1-2): 27-36.

Thomas, B., Gruca, L. L., Bennett, C., Parimi, P. S., Hanson, R. W. and Kalhan, S. C.; Metabolism of methionine in the newborn infant: response to the parenteral and enteral administration of nutrients. *Pediatric Research*, **2008**, 64(4): 381-386.

Tisi, D. K., Burns, D. H., Luskey, G. W. and Koski, K. G.; Fetal exposure to altered amniotic fluid glucose, insulin, and insulin-like growth factor–binding protein 1 occurs before screening for gestational diabetes mellitus. *Diabetes Care*, **2011**, 34(1): 139-144.

Tolun, A. A., Zhang, H., Il'yasova, D., Sztáray, J., Young, S. P. and Millington, D. S.; Allantoin in human urine quantified by ultra-performance liquid chromatography-tandem mass spectrometry. *Analytical Biochemistry*, **2010**, 402(2): 191-193.

Toot, P. J. and Lu, J. K. H.; *Female reproductive physiology*. *Essentials of obstetrics and gynecology*. N. F. Hacker, J. G. Moore and J. C. Gambone. Philadelphia, Elsevier, 2004**:** 33-45.

Trygg, J., Holmes, E. and Lundstedt, T.; Chemometrics in metabonomics. *Journal of Proteome Research*, **2007**, 6(2): 469-479.

Trygg, J. and Lundstedt, T.; *Chemometrics techniques for metabonomics*. *The handbook of metabonomics and metabolomics*. J. C. Lindon, J. K. Nicholson and E. Holmes. Oxford, Elsevier, 2007**:** 171-199.

Trygg, J. and Wold, S.; Orthogonal projections to latent structures (O-PLS). *Jounal of Chemometrics*, **2002**, 16(3): 119-128.

Tuma, P., Samcova, E. and Andelova, K.; Determination of free amino acids and related compounds in amniotic fluid by capillary electrophoresis with contactless conductivity detection. *Journal of Chromatography B*, **2006**, 839(1-2): 12-18.

Turner, E., Brewster, J. A., Simpson, N. A. B., Walker, J. J. and Fisher, J.; Plasma from women with preeclampsia has a low lipid and ketone body content—a nuclear magnetic resonance study. *Hypertension in Pregnancy*, **2007**, 26(3): 329-342.

Turner, E., Brewster, J. A., Simpson, N. A. B., Walker, J. J. and Fisher, J.; Aromatic amino acid biomarkers of preeclampsia - A nuclear magnetic resonance investigation. *Hypertension in Pregnancy*, **2008**, 27(3): 225-235.

Turner, E., Brewster, J. A., Simpson, N. A. B., Walker, J. J. and Fisher, J.; Imidazole-based erythrocyte markers of oxidative stress in preeclampsia—an NMR investigation. *Reproductive Sciences*, **2009**, 16(11): 1040-1051.

Turner, J. A.; Diagnosis and management of pre-eclampsia: an update. *International Journal of Women's Health*, **2010**, 2(1): 327-337.

Ulrich, E. L., Akutsu, H., Doreleijers, J. F., Harano, Y., Ioannidis, Y. E., Lin, J., Livny, M., Mading, S., Maziuk, D., Miller, Z., Nakatani, E., Schulte, C. F., Tolmie, D. E., Kent Wenger, R., Yao, H. and Markley, J. L.; BioMagResBank. *Nucleic Acids Research*, **2008**, 36(suppl 1): D402-D408.

Underwood, M. A., Gilbert, W. M. and Sherman, M. P.; Amniotic Fluid: Not just fetal urine anymore. *Journal of Perinatology*, **2005**, 25(5): 341-348.

van der Schoor, S. R., Wattimena, D. L., Huijmans, J., Vermes, A. and van Goudoever, J. B.; The gut takes nearly all: threonine kinetics in infants. *The American Journal of Clinical Nutrition*, **2007**, 86(4): 1132-1138.

Van Eijk, H. M. H., DeJong, C. H. C., Deutz, N. E. P. and Soeters, P. B.; Influence of storage conditions on normal plasma amino-acid concentrations. *Clinical Nutrition*, **1994**, 13(6): 374-380.

Vandenberghe, G., Mensink, I., Twisk, J. W. R., Blankenstein, M. A., Heijboer, A. C. and van Vugt, J. M. G.; First trimester screening for intra-uterine growth restriction and early-onset pre-eclampsia. *Prenatal Diagnosis*, **2011**, 31(10): 955-961.

Vascotto, C., Salzano, A. M., D'Ambrosio, C., Fruscalzo, A., Marchesoni, D., di Loreto, C., Scaloni, A., Tell, G. and Quadrifoglio, F.; Oxidized transthyretin in amniotic fluid as an early marker of preeclampsia. *Journal of Proteome Research*, **2007**, 6(1): 160-170.

Veselkov, K. A., Lindon, J. C., Ebbels, T. M. D., Crockford, D., Volynkin, V. V., Holmes, E., Davies, D. B. and Nicholson, J. K.; Recursive segment-wise peak alignment of biological ^{1}H NMR spectra for improved metabolic biomarker recovery. *Analytical Chemistry*, **2009**, 81(1): 56-66.

Veselkov, K. A., Vingara, L. K., Masson, P., Robinette, S. L., Want, E., Li, J. V., Barton, R. H., Boursier-Neyret, C., Walther, B., Ebbels, T. M., Pelczer, I. n., Holmes, E., Lindon, J. C. and Nicholson, J. K.; Optimized preprocessing of ultra-performance liquid chromatography/mass spectrometry urinary metabolic profiles for improved information recovery. *Analytical Chemistry*, **2011**, 83(15): 5864–5872.

Viegas, A., Manso, J., Nobrega, F. L. and Cabrita, E. J.; Saturation-transfer difference (STD) NMR: A simple and fast method for ligand screening and characterization of protein binding. *Journal of Chemical Education*, **2011**, 88(7): 990-994.

Vina, J. R., Palacin, M., Puertes, I. R., Hernandez, R. and Vina, J.; Role of the gamma-glutamyl cycle in the regulation of amino acid translocation. *American Journal of Physiology - Endocrinology And Metabolism*, **1989**, 257(6): E916-E922.

von Dadelszen, P., Payne, B., Li, J., Ansermino, J. M., Pipkin, F. B., Côté, A.-M., Douglas, M. J., Gruslin, A., Hutcheon, J. A., Joseph, K. S., Kyle, P. M., Lee, T., Loughna, P., Menzies, J. M., Merialdi, M., Millman, A. L., Moore, M. P., Moutquin, J.-M., Ouellet, A. B., Smith, G. N., Walker, J. J., Walley, K. R., Walters, B. N., Widmer, M., Lee, S. K., Russell, J. A. and Magee, L. A.; Prediction of adverse maternal outcomes in pre-eclampsia: development and validation of the fullPIERS model. *The Lancet*, **2011**, 377(9761): 219-227.

von Kleist-Retzow, J.-C., Cormier-Daire, V., Viot, G., Goldenberg, A., Mardach, B., Amiel, J., Saada, P., Dumez, Y., Brunelle, F., Saudubray, J.-M., Chrétien, D., Rötig, A., Rustin, P., Munnich, A. and De Lonlay, P.; Antenatal manifestations of mitochondrial respiratory chain deficiency. *The Journal of pediatrics*, **2003**, 143(2): 208-212.

Vu, T. N., Valkenborg, D., Smets, K., Verwaest, K. A., Dommisse, R., Lemière, F., Verschoren, A., Goethals, B. and Laukens, K.; An integrated workflow for robust alignment and simplified quantitative analysis of NMR spectrometry data. *BMC Bioinformatics*, **2011**, 12(405): doi:10.1186/1471-2105-1112-1405.

Walker, J. J.; Pre-eclampsia. *The Lancet*, **2000**, 356(9237): 1260-1265.

Wall, P. D., Pressman, E. K. and Woods, J. R.; Preterm premature rupture of the membranes and antioxidants: the free radical connection. *Journal of Perinatal Medicine*, **2002**, 30(6): 447.

Wang, C.-N., Chen, J. Y.-S., Sabu, S., Chang, Y.-L., Chang, S.-D., Kao, C.-C., Peng, H.-H., Chueh, H.-Y., Chao, A.-S., Cheng, P.-J., Lee, Y.-S., Chi, L.-M. and Wang, T.-H.; Elevated amniotic fluid F_2-isoprostane: A potential predictive marker for preeclampsia. *Free Radical Biology and Medicine*, **2011a**, 50(9): 1124-1130.

Wang, T.-H., Chao, A.-S., Chen, J.-K., Chao, A., Chang, Y.-L., Cheng, P.-J., Chang, S.-D. and Wang, H.-S.; Network analyses of differentially expressed proteins in amniotic fluid supernatant associated with abnormal human karyotypes. *Fertility and sterility*, **2009**, 92(1): 96-107.

Wang, T., Zhou, R., Zhang, L., Wang, Y., Song, C. p., Lin, W., Niu, X., Lin, Y. and Hu, H.; Proteins in leaked amniotic fluid as biomarkers diagnostic for prelabor rupture of membranes. *Proteomics – Clinical Applications*, **2011b**, 5(7-8): 415-421.

Wang, Y., Holmes, E., Nicholson, J. K., Cloarec, O., Chollet, J., Tanner, M., Singer, B. H. and Utzinger, J.; Metabonomic investigations in mice infected with Schistosoma mansoni: An approach for biomarker identification. *Proceedings of the National Academy of Sciences of the United States of America*, **2004**, 101(34): 12676-12681.

Want, E. J., Nordström, A., Morita, H. and Siuzdak, G.; From exogenous to endogenous: The inevitable imprint of mass spectrometry in metabolomics. *Journal of Proteome Research*, **2006**, 6(2): 459-468.

Want, E. J., Wilson, I. D., Gika, H., Theodoridis, G., Plumb, R. S., Shockcor, J., Holmes, E. and Nicholson, J. K.; Global metabolic profiling procedures for urine using UPLC-MS. *Nature Protocols*, **2010**, 5(6): 1005-1018.

Wapner, R. J. and Toy, E.; *Chorionic villus sampling*. *Sonography in obstetrics & gynecology: principles and practice*. A. Fleischer, E. Toy, W. Lee, F. Manning and R. Romero. New York, McGraw-Hill, 2011.

Westerhuis, J., Hoefsloot, H., Smit, S., Vis, D., Smilde, A., van Velzen, E., van Duijnhoven, J. and van Dorsten, F.; Assessment of PLSDA cross validation. *Metabolomics*, **2008**, 4(1): 81-89.

Wiklund, S., Johansson, E., Sjostrom, L., Mellerowicz, E. J., Edlund, U., Shockcor, J. P., Gottfries, J., Moritz, T. and Trygg, J.; Visualization of GC/TOF-MS-based metabolomics data for identification of biochemically interesting compounds using OPLS class models. *Analytical Chemistry*, **2007**, 80(1): 115-122.

Wilson, I. D., Nicholson, J. K., Castro-Perez, J., Granger, J. H., Johnson, K. A., Smith, B. W. and Plumb, R. S.; High Resolution "Ultra Performance" Liquid chromatography coupled to oa-TOF mass spectrometry as a tool for differential metabolic pathway profiling in functional genomic studies. *Journal of Proteome Research*, **2005a**, 4(2): 591-598.

Wilson, I. D., Plumb, R., Granger, J., Major, H., Williams, R. and Lenz, E. M.; HPLC-MS-based methods for the study of metabonomics. *Journal of Chromatography B*, **2005b**, 817(1): 67-76.

Wishart, D. S., Knox, C., Guo, A. C., Eisner, R., Young, N., Gautam, B., Hau, D. D., Psychogios, N., Dong, E., Bouatra, S., Mandal, R., Sinelnikov, I., Xia, J., Jia, L., Cruz, J. A., Lim, E., Sobsey, C. A., Shrivastava, S., Huang, P., Liu, P., Fang, L., Peng, J., Fradette, R., Cheng, D., Tzur, D., Clements, M., Lewis, A., De Souza, A., Zuniga, A., Dawe, M., Xiong, Y., Clive, D., Greiner, R., Nazyrova, A., Shaykhutdinov, R., Li, L., Vogel, H. J. and Forsythe, I.; HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Research*, **2009**, 37(suppl 1): D603-D610.

Witkin, S. S., Skupski, D., Herway, C., Rudge, M. V. C., Saito, F. and Harris, M.; Fatty acid composition of mid-trimester amniotic fluid in women of different ethnicities. *Journal of Maternal-Fetal and Neonatal Medicine*, **2012**, 25(6): 818-821.

Wold, S., Sjöström, M. and Eriksson, L.; PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, **2001**, 58(2): 109-130.

Wolfender, J.-L., Queiroz, E. F. and Hostettmann, K.; Phytochemistry in the microgram domain - a LC-NMR perspective. *Magnetic Resonance in Chemistry*, **2005**, 43(9): 697-709.

Wu, D. H., Chen, A. D. and Johnson, C. S.; An Improved Diffusion-Ordered Spectroscopy Experiment Incorporating Bipolar-Gradient Pulses. *Journal of Magnetic Resonance, Series A*, **1995**, 115(2): 260-264.

Yang, Y. J., Lee, J., Choi, M. H. and Chung, B. C.; Direct determination of estriol 3- and 16-glucuronides in pregnancy urine by column-switching liquid chromatography with electrospray tandem mass spectrometry. *Biomedical Chromatography*, **2003**, 17(4): 219-225.

Yap, I. K. S., Angley, M., Veselkov, K. A., Holmes, E., Lindon, J. C. and Nicholson, J. K.; Urinary metabolic phenotyping differentiates children with autism from their unaffected siblings and age-matched controls. *Journal of Proteome Research*, **2010**, 9(6): 2996-3004.

Yaron, Y., Hamby, D. D., O'Brien, J. E., Critchfield, G., Leon, J., Ayoub, M., Johnson, M. P. and Evans, M. I.; Combination of elevated maternal serum alpha-fetoprotein (MSAFP) and low estriol is highly predictive of anencephaly. *American Journal of Medical Genetics*, **1998**, 75(3): 297-299.

Zeisel, S. H.; Choline: critical role during fetal development and dietary requirements in adults. *Annual Review of Nutrition*, **2006**, 26(1): 229-250.

Websites:

Bacino, C. A.; (2012) "Approach to congenital malformations" Retrieved November 8, from www.uptodate.com (Wolters Kluer Health) 2012.

Brown, H. L.; (2007 November 2007) "Stages of development of the fetus." Retrieved in July 2011, from http://www.merckmanuals.com/home/sec22/ch254/ch254b.html (Merck Sharp & Dohme Corp.) 2011.

National Institutes of Health. National Human Genome Research Institute. "Talking Glossary of Genetic Terms" Retrieved in September 8, 2012, from http://www.genome.gov/glossary/?id=114

Mestrelab Research. "Automatic DOSY processing". Retrieved in July 2012, from http://mestrelab.com/resources/dosy/

Matrix Science. "Quadrupole Ion Trap MS" Retrieved in November 2012, from http://www.matrixscience.com/help/ion_trap_main_help.html

Annexes

Annex I - Informed signed consent and questionnaire.

Annex II - ^{1}H NMR pulse programs.

Annex III - UPLC-MS reproducibility evaluation.

Annex IV - MIR reproducibility evaluation.

Annex V – Full list of AF compounds identified by high resolution NMR and LC-NMR/MS.

Annex VI - Lists of UPLC-MS unassigned features detected in AF.

Annex VII - Lists of UPLC-MS unassigned features detected in maternal urine.

Annex I - Informed signed consent and questionnaire.

Maternidade Bissaya Barreto

universidade de aveiro 30 anos a projectar futuros

Universidade de Coimbra

Coloque aqui a
Etiqueta de identificação da consulta

CONFIDENCIAL

PEDIDO DE AUTORIZAÇÃO DA GRÁVIDA PARA RECOLHA DE SANGUE, URINA E EXCEDENTE DO LÍQUIDO AMNIÓTICO

O sangue recolhido destina-se a análise **confidencial** no âmbito de um projecto de investigação sobre:

"DIAGNÓSTICO PRECOCE DE DESORDENS NA SAÚDE DA GRÁVIDA E DO FETO POR MÉTODOS ESPECTROSCÓPICOS"

Esta investigação não trará quaisquer consequências para a dadora de sangue, podendo no entanto levar a importantes avanços na área de detecção precoce de doenças do feto e da mãe.

Autorizo a recolha de sangue, aceitando que a minha amostra de sangue fica sob a responsabilidade da Universidade de Aveiro, Universidade de Coimbra e Maternidade Bissaya Barreto que conduzirão a sua análise em total confidencialidade, e abdicando de quaisquer direitos a potenciais benefícios, materiais ou não, resultantes da investigação em curso.

P-LA_____/____
P-S _____/____
P-U _____/____

Motivo: _____

Assinatura:……Data: ……

✂----------

Se respondeu afirmativamente, agradecemos que responda às seguintes questões:

1- Quantas semanas tem de gravidez neste momento?.....

2- Toma (ou tomou) alguma medicação especial durante a sua gravidez? Se sim, por favor especifique qual e para que efeito:……..……

3- Pratica exercício físico durante a gravidez? Se sim, especifique:

a) o tipo de exercício…….
b) frequência com que o pratica……
c) período da gravidez em que o praticou:……

4- Tem cuidados específicos com a sua alimentação durante a gravidez? Se sim, especifique quais……..……

5- Tem antecedentes de Pré-Eclampsia, Diabetes Gestacional ou Outros (em caso afirmativo especifique).....

6 – Indique-nos por favor:

a)seu peso actual.....
b) peso anterior à gravidez.....
c) a sua altura.....

P-LA_____/____
P-S _____/____
P-U _____/____

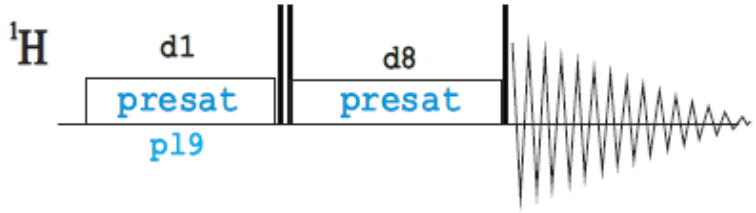
Motivo: _____

Muito obrigada pelo seu tempo e sua colaboração!

Pela Equipa de Investigação,
Maternidade Bissaya Barreto
Universidade de Aveiro

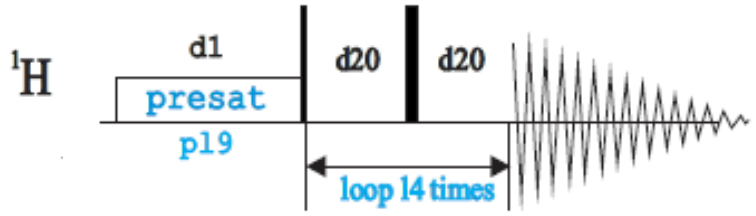
Annex II - [1]H NMR pulse programs.

Pulse program *noesy1d* from Bruker library[1]:



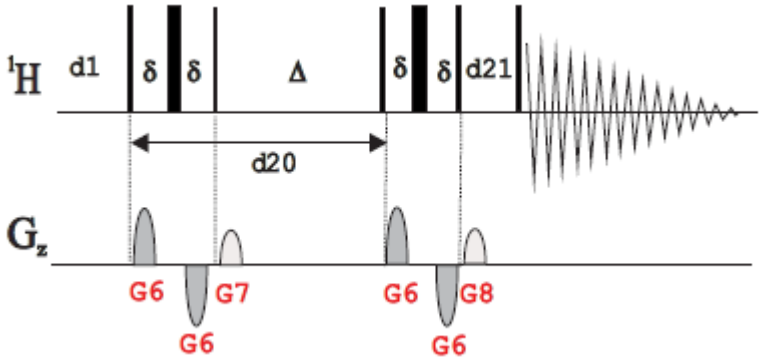
pl9 : power level for presaturation; d1: relaxation delay; d8: mixing time.

Pulse program *cpmgpr* from Bruker library[1]:



pl9 : power level for presaturation; d1: relaxation delay; d20: fixed echo time to allow elimination of diffusion and J-mod. effects; l4: loop for T_2 filter.

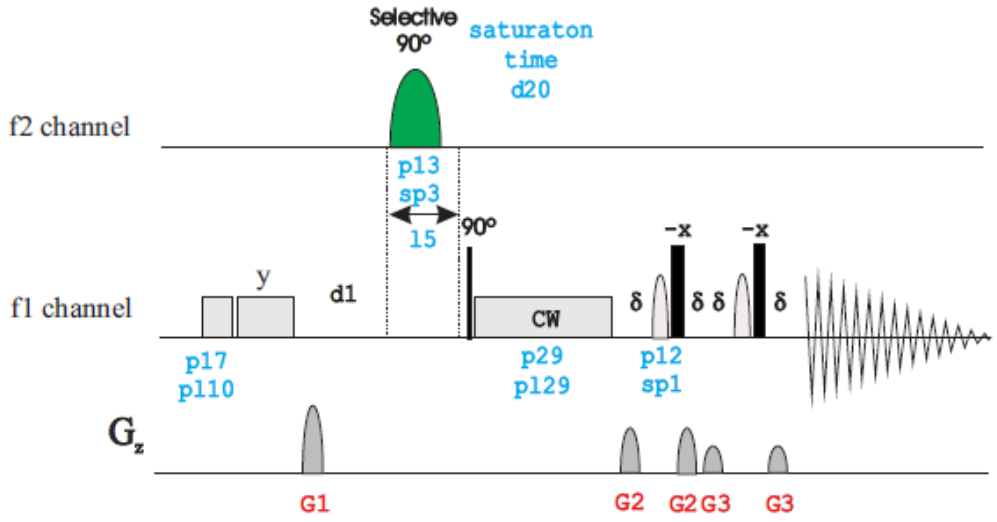
Pulse program *ledbpgp2s1dpr* from Bruker library[1]:



d1: relaxation delay; d20: diffusion time; d21 eddy current delay; G6, G7, G8: z-gradients. A pulse for presaturation (with pl9 power) is also applied during d1 (not shown in the scheme).

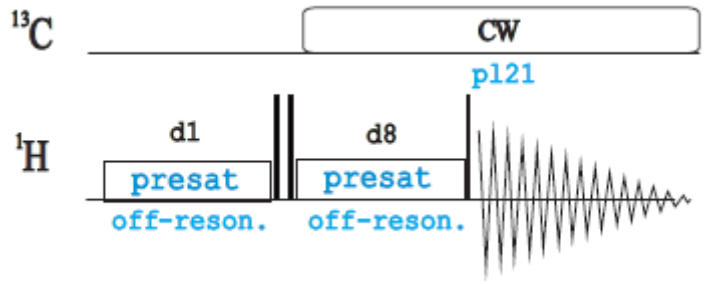
Pulse program *stddiffesgp.3* from Bruker library[1,2]:

[1] Parella, T.; *Pulse Program Catalogue: Volume I: 1D & 2D NMR Experiments. NMRGuide4.1 – Topspin 2.0*. Germany, Bruker BioSpin GmbH, 2006.



p17: f1 channel - trim pulse [2.5msec]; pl10: f1 channel - power level for TOCSY-spinlock; p29: f1 channel - trim pulse; pl29: f1 channel - power level for trim pulse; p12: f1 channel - 180 degree shaped pulse (Squa100.1000) [2msec]; sp1 : f1 channel - shaped pulse 180 degree; ;p13: f2 channel - shaped pulse for saturation [50msec]; d20: saturation time; l5: loop for saturation (p13*l5: saturation time); G1, G2, G3: z-gradients. The on- and off-resonance irradiated frequencies are stored in a list file (fq2list).

Pulse program *lc1pncwps* from Bruker library[1]:

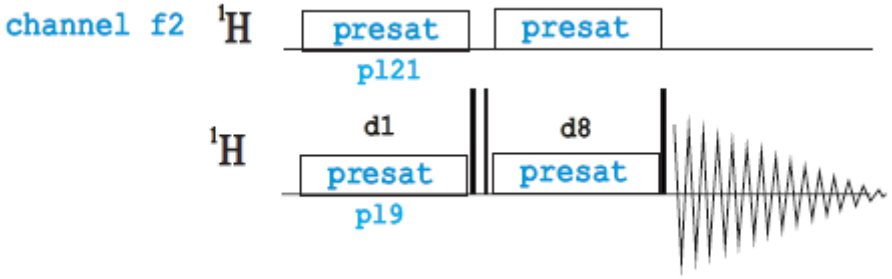


d1: relaxation delay; d8: mixing time; p121: f2 channel - power level for cw-decoupling; presaturation during relaxation delay and mixing time using shaped pulse for off-resonance presaturation.

[1] Parella, T.; *Pulse Program Catalogue: Volume I: 1D & 2D NMR Experiments. NMRGuide4.1 – Topspin 2.0*. Germany, Bruker BioSpin GmbH, 2006.

[2] The STD experiment is protected by international patents owned by: Alepharma, Raamfeld 67, 22397 Hamburg, Germany.

Pulse program *lc1pnf2* from Bruker library[1]:



d1: relaxation delay; d8: mixing time; double presaturation during relaxation delay and mixing time using f1 and f2 chanels: pl9 (power level for f1 channel) and p121 (power level for f2 channel).

[1] Parella, T.; *Pulse Program Catalogue: Volume I: 1D & 2D NMR Experiments. NMRGuide4.1 – Topspin 2.0*. Germany, Bruker BioSpin GmbH, 2006.

Annex III - UPLC-MS reproducibility evaluation

The reproducibility of AF samples in UPLC-MS was evaluated. For this reason a quality control (QC) AF sample was injected between each 10 samples of each analytical batch. The reproducibility was evaluated by PCA of the QC samples (n=10) together with the remaining samples run in UPLC-MS profiling study (n=82). The QCs distribution in the scores scatter plots of PCA was inspected for each data set using after application of the scaling transformations Pareto scaling, unit variance (UV) scaling and logarithm transformation, which were used to minimize the dominance of higher intensity signals over less intense ones (Figure A1).

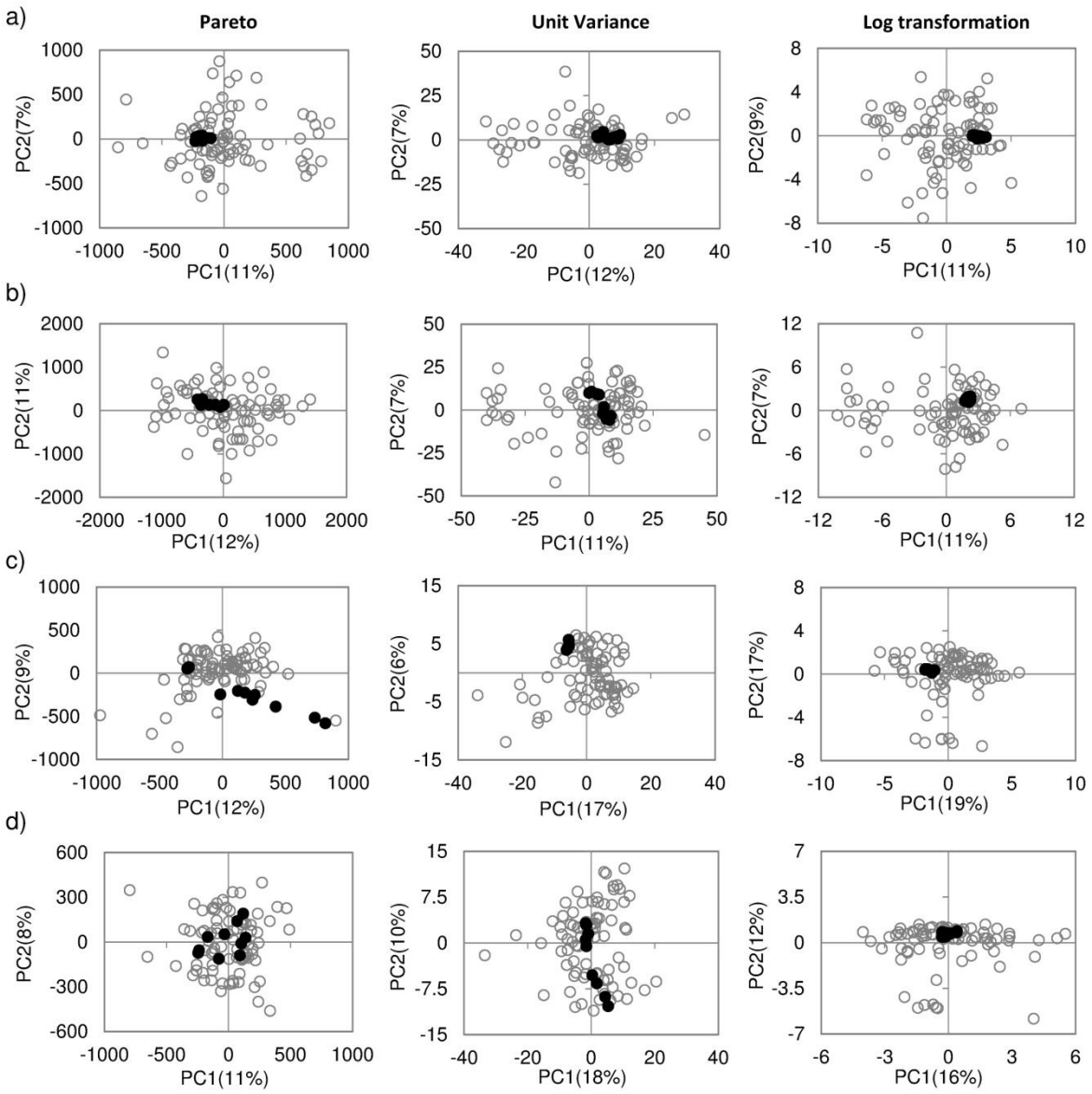


Figure A1 - PCA scores scatter plots of AF UPLC-MS data sets using Pareto scaling (left), Unit variance scaling (centre) and Log transformation (right): a) HSS/ESI$^+$, b) HSS/ESI$^-$, c) HILIC/ESI$^+$, d) HILIC/ESI$^-$. QCs (●,n=10) and study samples (○, n =82).

As shown in the PCA scores scatter plots of QCs and study samples, the QCs samples were clustered together for HSS/ESI$^+$ data set in the 3 scaling transformations applied, with only a small drift noticed in PC1 of UV scaling (Figure A1a, centre). In the HSS/ESI$^-$ data set, the QCs show good clustering only when Log transformation was employed (Figure A1b, right). For HILIC data sets, HILIC/ESI$^+$ showed good QC clustering in UV and Log transformed data (Figure A1c), whereas HILIC/ESI$^-$ only showed good clustering in the Log transformed data set (Figure A1d).

The results demonstrate that better analytical stability seems to be achieved in HSS/ESI$^+$ data sets, compared to data obtained using a HILIC column. This probably resulted from a poorer performance of the HILIC column. Therefore all Log transformed data sets and HSS/ESI$^+$ (Pareto scaled) and HILIC/ESI$^+$ (UV scaled) were used in further UPLC-MS multivariate analysis (MVA).

To evaluate the reproducibility of the UPLC-MS maternal urine data sets, a PCA analysis of QC and study samples was also performed using the same approach used for UPLC-MS AF data sets (Figure A2, next page).

Good clustering of QC samples was obtained in the PCA scores plots from all data sets of HSS/ESI$^+$ (Figure A2a), Pareto and UV scaled data sets of HSS/ESI$^-$ (Figure A2b) as well as all HILIC/ESI$^-$ data sets (Figure A2d), which is indicative of reproducibility of those data sets.

However, no clustering of the HILIC/ESI$^+$ data sets was observed (Figure A2c), which may indicate an experimental drift resulting from the acquisition of this data set. Moreover, the logarithmic transformation also resulted in a small drift in the QCs of the HSS/ESI$^-$ data set (Figure A2c, right). The explanation for such drift is not clear since good clustering was observed for the Pareto and UV scaling of the same data set (Figure A2b).

The maternal urine data sets which registered a good reproducibility using the different scaling procedures were considered for further control *versus* disease MVA.

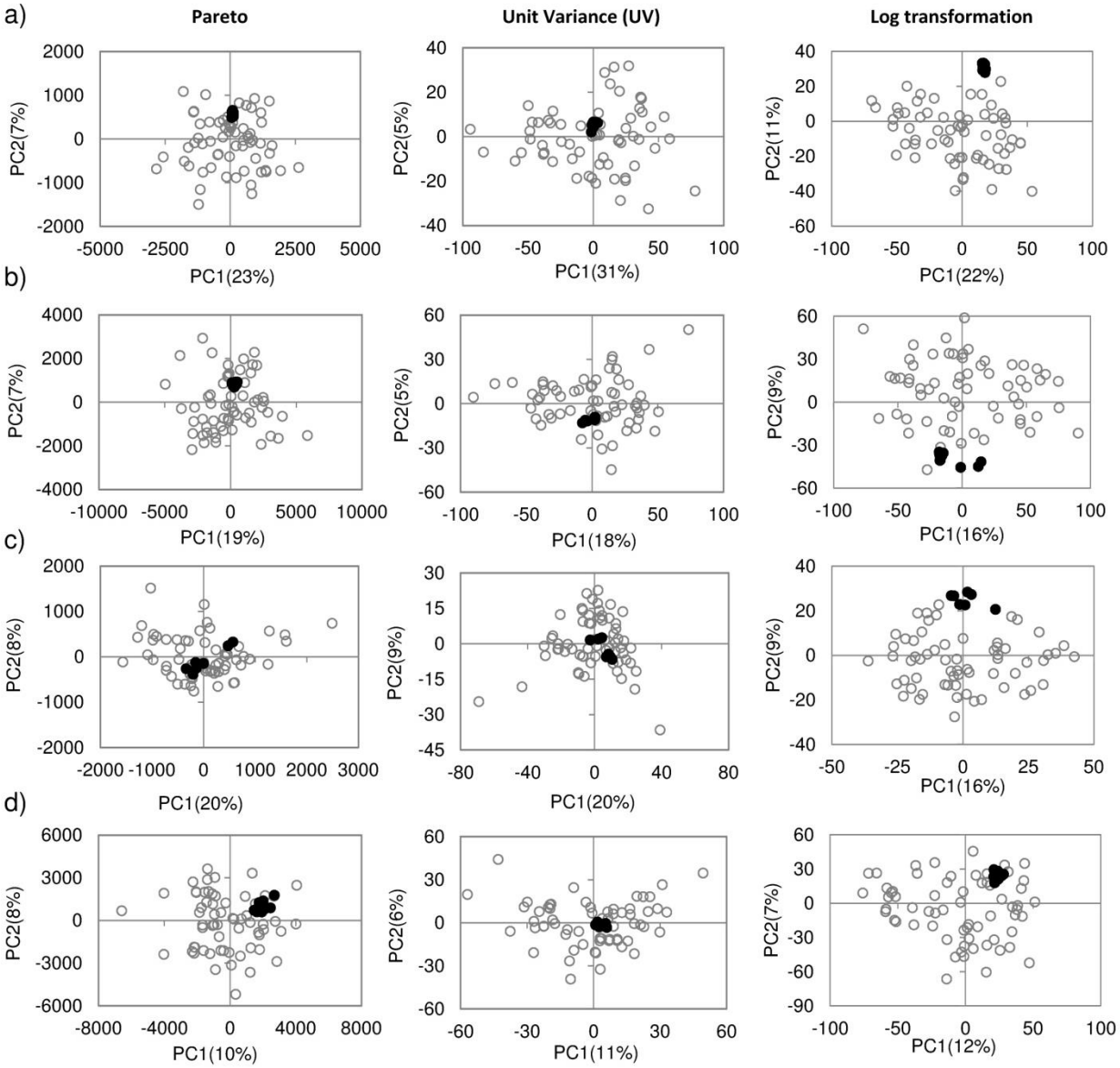


Figure A2 - PCA scores scatter plots of maternal urine UPLC-MS data sets using Pareto scaling (left), Unit variance scaling (centre) and Log transformation (right): a) HSS/ESI$^+$, b) HSS/ESI$^-$, c) HILIC/ESI$^+$, d) HILIC/ESI$^-$. QCs (●, n=10) and study samples (○, n =64).

Annex IV - MIR reproducibility evaluation

A typical sequence of MIR spectra of AF recorded as a function of sample drying time is shown in Figure A3. Sample drying was necessary to remove the excess of water that contributed to the broad band absorbing around 3300 and 1700 cm^{-1} (Figure A3).

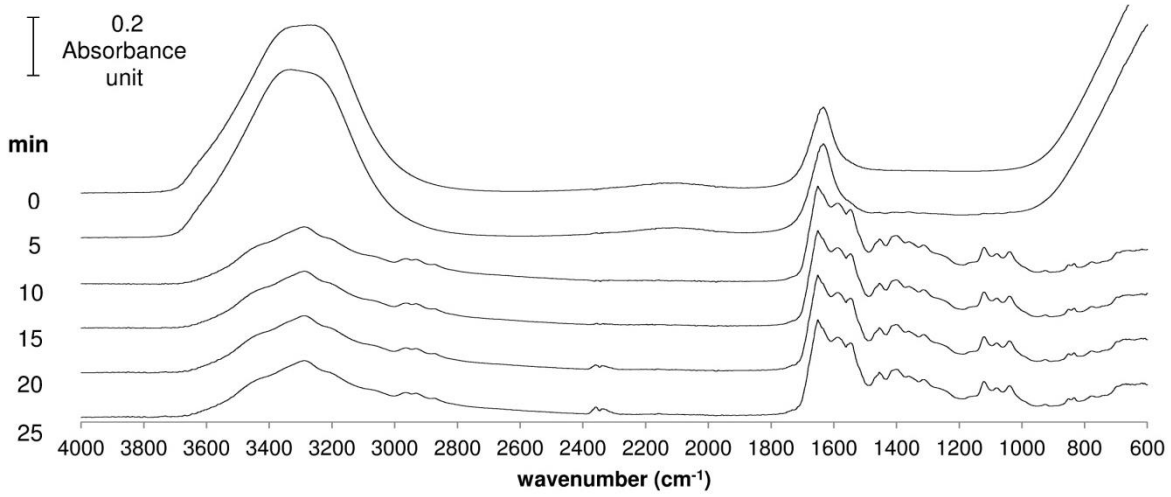


Figure A3 - ATR - MIR spectra of human amniotic fluid recorded as a function of drying time.

After 25 min, a spectrum with enhanced signal to noise ratio particularly in the fingerprint region (1800-600 cm^{-1}) was obtained (Figure A3), for which proteins, lipids, organic acids (mainly lactate), amino acids, and glucose contribute importantly.

In order to evaluate spectral reproducibility, the MIR spectra of a quality control (QC) sample were repeatedly acquired in the course of the study. PCA of the full spectra (Figure A4a) showed slightly less QC reproducibility, compared to that observed when narrower spectral ranges were considered (Figure A4b-e). The lower reproducibility of the full spectrum may be due to the relative variability of the O-H stretching (ν) vibration at 3300 cm^{-1} which possibly reflected small deviations in bound water content. Some drift was also present in QCs of 3000-2800 cm^{-1} region along PC1 (Figure A4b), also possibly reflecting the contribution of the nearby water band at 3300 cm^{-1}. Smaller QC data point dispersion was obtained for the 1200-900 cm^{-1} (Figure A4d) and 1800-1200 cm^{-1} regions (Figure A4c). Notably, second derivative transformation of the original spectra, which was used in the transformation of some spectral regions in order to correct baseline differences and increase resolution, resulted in slightly lesser QC dispersion when applied to 1200-900 cm^{-1} spectral region (Figure A4e).

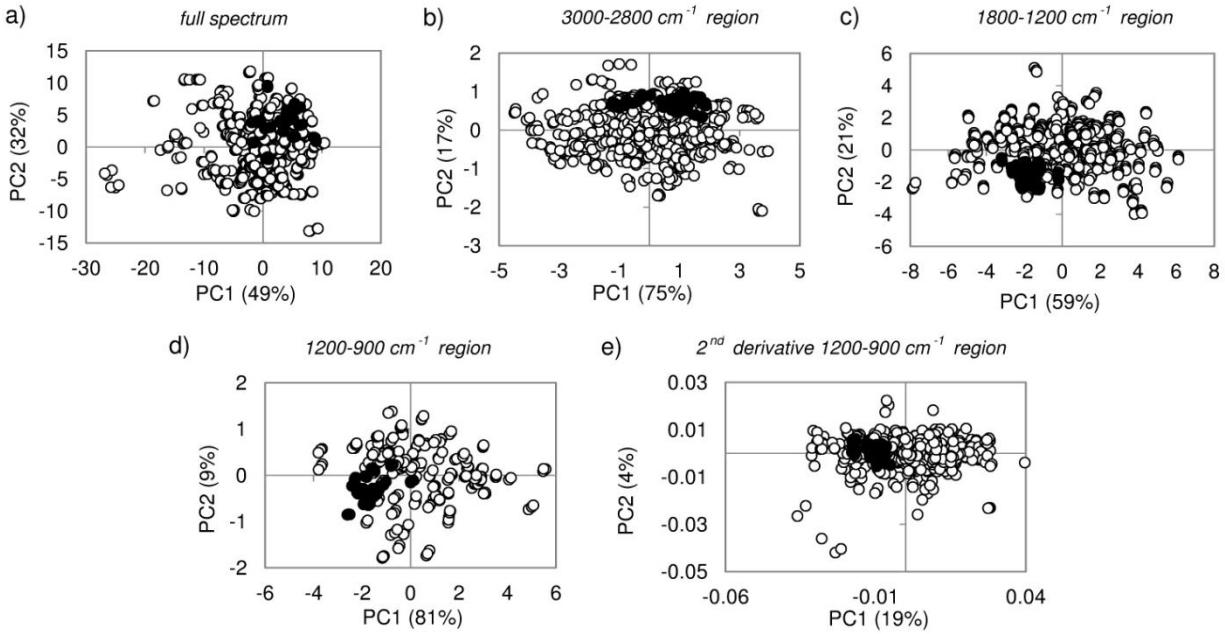


Figure A4 - PCA scores scatter plots of MIR spectra of AF of study samples (○, n=198) and QC sample aliquots (●, n=20) of spectral regions: a) full spectra; b) 3000-2800 cm^{-1}; c) 1800-1200cm^{-1}; d) 1200-900 cm^{-1}; e) 1200-900 cm^{-1} region after 2nd derivative transformation. Three spectral replicates were used per sample.

These results indicate that spectral stability and therefore a better reproducibility was obtained particularly when considering in the 1200-900 cm^{-1} and 1800-1200 cm^{-1} spectral regions, thus encouraging their use for MVA. However, second derivative transformation of data sets of 3000-2800 and 1800-1200 cm^{-1} regions (not shown) gave results similar to that observed in Figure A4a, therefore these transformed data sets were not considered. Despite the drift noted in the PCA scores scatter plot of region 3000-2800 cm^{-1} (Figure A4b), this spectral region was also tested in MVA, due to the important contributions of several absorbing species in this particular spectral region.

Annex V – Full list of AF compounds identified by high resolution NMR and LC-NMR/MS

Table A1 – Full list of AF compounds identified in this work by high resolution NMR and LC-NMR/MS. *newly assigned by NMR; ** detected for the first time in AF; ***partially assigned.

amino acids and derivatives		
Alanine	Glutamate	N,N-dimethylglycine
Acetylcholine*	Glutamine	Ornithine*
α-aminobutyrate (ABA)*	Glycine	Pyroglutamate*
Arginine*	Histidine	Taurine*
Aspartate*	Hydroxyproline*	Threonine
Betaine	Isoleucine	Proline
Carnitine	Indoxyl sulfate	Phenylalanine
Citrulline*	Leucine	Tyrosine
Choline	Lysine	Tryptophan*
Creatine	Methionine	Valine
Creatinine	N-acetylalanine**	
Cystine*	N-acetylglycine**	
organic acids		
Acetate	β-hydroxybutyrate	Malate*
Acetone*	α-hydroxyisobutyrate*	3-methyl-2-oxovalerate**
Acetoacetate*	α-hydroxyisovalerate*	α-oxoisovalerate
Citrate	Hydroxypropionate*	2-oxoleucine**
Formate	Isovalerate**	Oxaloacetate*
Fumarate*	α-ketoglutarate*	Pyruvate
α-hydroxybutyrate	Lactate	Succinate
sugars		
Glucose	Fructose*	Pentose*
Mannose*	Hexose disaccharide***	
Myo-inositol	Methoxylated oligosaccharide***	
sugar alcohols		
Sorbitol*	Xylitol**	
Proteins*		
Protein system A (glycoprotein)	Protein system B	Protein system C
other metabolites		
Caffeine*	Trimethylamine-N-oxide (TMAO)*	Ethanolamine*
Hippurate*	Paraxanthine*	Ethanol*
Urea*	4-amino-hippurate*	

Annex VI - Lists of UPLC-MS unassigned features detected in AF

Table A2 - Unassigned significant features obtained from OPLS-DA controls *vs* FM models of 2nd trimester AF UPLC-MS analysis.

Amniotic Fluid - Fetal Malformations (FM)								
UPLC column	**Detection mode**	**m/z**	**RT(min.)**	**Feature intensity**	**Model scaling/ transformation**	**direction of variation**	**magnitude of variation**	***p*-value**
HSS	ESI+	125.97	0.4	low	Pareto	↑	2.8	1.01×10^{-05}
HSS	ESI+	143.99	0.4	low	Pareto	↑	2.7	1.61×10^{-05}
HSS	ESI+	185.02	0.4	low	Pareto	↑	2.3	5.61×10^{-07}
HSS	ESI+	167.01	0.4	high	Pareto	↑	2.4	4.85×10^{-07}
HSS	ESI+	122.53	0.4	low	Log	↑	2.7	5.61×10^{-07}
HSS	ESI+	185.02	0.4	low	Log	↑	2.3	2.02×10^{-05}
HSS	ESI+	239.96	0.4	low	Log	↑	2.7	2.52×10^{-05}
HSS	ESI+	242.00	0.4	low	Pareto	↑	1.5	5.36×10^{-05}
HSS	ESI+	507.23	0.5	low	Log	↓	3.7	8.96×10^{-05}
HSS	ESI+	288.92	0.5	low	Log	↑	2.1	3.10×10^{-04}
HSS	ESI+	248.08	0.8	very low	Log	↑	1.8	4.85×10^{-04}
HSS	ESI+	209.05	1.0	low	Pareto	↑	2.0	2.25×10^{-05}
HSS	ESI+	245.07	1.0	low	Pareto	↑	1.8	1.80×10^{-05}
HSS	ESI+	267.06	1.0	low	Pareto	↑	1.7	5.49×10^{-06}
HSS	ESI+	281.14	1.2	low	Log	↓	1.9	2.02×10^{-05}
HSS	ESI+	137.03	1.4	low	Pareto	↑	2.4	4.84×10^{-06}
HSS	ESI+	170.07	1.7	low	Log	↓	1.7	3.10×10^{-04}
HSS	ESI+	153.03	1.7	low	Pareto	↑	1.7	3.40×10^{-04}
HSS	ESI+	749.29	2.6	low	Log	↑	2.8	4.44×10^{-04}
HSS	ESI+	387.17	2.6	low	Log	↑	2.1	8.93×10^{-06}
HSS	ESI+	328.14	2.9	low	Log	↓	1.5	2.83×10^{-03}
HSS	ESI+	310.13	2.9	low	Log	↓	1.6	1.14×10^{-06}
HSS	ESI+	410.18	3.4	low	Log	↓	2.3	2.59×10^{-04}
HSS	ESI+	819.35	3.4	low	Log	↓	3.5	2.91×10^{-06}
HSS	ESI+	261.14	3.5	high	Log	↓	2.4	2.67×10^{-07}
HSS	ESI+	363.22	5.7	low	Log	↑	1.7	7.43×10^{-04}
HSS	ESI+	541.27	6.7	very low	Log	↑	2.0	1.43×10^{-05}
HSS	ESI+	285.26	6.7	low	Pareto	↑	2.3	4.34×10^{-05}
HSS	ESI-	402.93	0.4	low	Log	↑	2.2	3.90×10^{-05}
HSS	ESI-	516.90	0.4	low	Log	↑	2.2	7.31×10^{-05}
HSS	ESI-	548.85	0.4	low	Log	↑	2.6	3.50×10^{-05}
HSS	ESI-	494.86	0.5	low	Log	↑	2.4	1.67×10^{-07}
HSS	ESI-	597.30	1.9	low	Log	↓	2.1	1.96×10^{-06}
HSS	ESI-	626.25	1.9	low	Log	↓	1.9	4.84×10^{-06}
HSS	ESI-	624.26	2.7	low	Log	↑	2.7	9.90×10^{-05}
HSS	ESI-	817.31	3.4	low	Log	↓	3.8	4.84×10^{-06}
HSS	ESI-	407.22	5.6	low	Log	↑	2.5	4.79×10^{-09}
HSS	ESI-	564.29	6.7	low	Log	↑	2.6	3.14×10^{-05}
HILIC	ESI+	752.56	0.8	very low	UV	↑	1.6	2.55×10^{-04}
HILIC	ESI+	123.06	0.9	very low	UV	↓	1.3	9.02×10^{-04}
HILIC	ESI+	718.58	4.0	low	UV	↑	2.3	1.80×10^{-06}
HILIC	ESI+	701.56	4.2	low	UV	↑	2.5	3.49×10^{-03}
HILIC	ESI+	260.19	4.9	low	UV	↓	2.0	4.01×10^{-05}

Table A2 (*cont.*)

Amniotic Fluid - Fetal Malformations (FM)								
UPLC column	**Detection mode**	**m/z**	**RT(min.)**	**feature intensity**	**Model scaling/ transformation**	**direction of variation**	**magnitude of variation**	***p*-value**
HILIC	ESI+	177.07	4.9	low	UV	↓	1.8	8.85×10^{-07}
HILIC	ESI+	397.13	4.9	low	Log	↓	2.7	3.07×10^{-07}
HILIC	ESI+	367.09	5.2	low	UV	↓	1.4	1.16×10^{-07}
HILIC	ESI+	299.10	5.2	very low	UV	↓	1.6	4.06×10^{-08}
HILIC	ESI+	438.16	5.2	low	UV	↓	2.0	2.07×10^{-06}
HILIC	ESI+	543.15	5.4	low	Log	↓	1.5	2.79×10^{-02}
HILIC	ESI+	521.17	5.4	low	Log	↓	1.5	5.87×10^{-02}
HILIC	ESI+	277.06	5.4	low	Log	↓	1.8	4.66×10^{-03}
HILIC	ESI+	500.19	5.4	low	Log	↓	2.2	8.12×10^{-03}
HILIC	ESI+	176.04	5.6	low	UV	↓	1.4	3.72×10^{-04}
HILIC	ESI+	154.06	5.7	low	UV	↓	1.3	1.73×10^{-04}
HILIC	ESI+	307.11	5.7	very low	UV	↓	1.5	1.57×10^{-04}
HILIC	ESI-	465.31	0.6	high	Log	↑	2.3	1.96×10^{-06}
HILIC	ESI-	413.20	0.7	low	Log	↑	2.1	1.67×10^{-07}
HILIC	ESI-	385.17	0.7	low	Log	↑	2.0	7.02×10^{-06}
HILIC	ESI-	401.16	0.7	low	Log	↑	2.7	8.10×10^{-05}
HILIC	ESI-	198.07	5.0	low	Log	↓	1.9	2.29×10^{-07}
HILIC	ESI-	497.17	5.4	low	Log	↓	1.7	1.89×10^{-02}
HILIC	ESI-	237.09	5.4	high	Log	↓	1.6	2.66×10^{-02}

Table A3 - Unassigned significant features obtained from OPLS-DA controls *vs pre*-PTD models of 2nd trimester AF UPLC-MS analysis.

Amniotic Fluid - pre-Premature Delivery (PTD)								
UPLC column	**Detection mode**	**m/z**	**RT(min.)**	**feature intensity**	**Model scaling/ transformation**	**Direction of variation**	**magnitude of variation**	***p*-value**
HSS	ESI+	331.12	0.6	low	Log	↑	1.8	6.10×10^{-03}
HSS	ESI+	353.10	0.6	very low	Log	↑	1.9	1.58×10^{-02}
HSS	ESI+	309.16	0.6	low	Log	↑	1.7	1.93×10^{-02}
HSS	ESI+	226.12	0.7	low	Log	↑	1.7	4.69×10^{-02}
HSS	ESI-	597.09	0.7	low	Log	↑	1.8	3.39×10^{-03}
HSS	ESI+	192.03	0.8	low	Pareto	↓	1.4	1.04×10^{-03}
HSS	ESI+	222.04	1.0	low	Log	↓	2.0	6.83×10^{-03}
HSS	ESI+	390.98	1.3	low	Pareto	↑	1.3	5.45×10^{-03}
HSS	ESI+	247.14	1.4	high	Log	↓	2.2	1.37×10^{-03}
HSS	ESI+	202.18	2.0	high	Log	↑	1.6	3.64×10^{-04}
HSS	ESI+	749.29	2.6	very low	Log	↑	1.9	2.33×10^{-02}
HSS	ESI+	247.16	2.9	low	Pareto	↑	1.3	6.10×10^{-03}
HSS	ESI-	371.06	3.1	low	Log	↓	1.9	2.63×10^{-04}
HSS	ESI-	248.06	3.4	low	Log	↓	2.9	6.72×10^{-04}
HSS	ESI+	316.21	3.9	low	Log	↑	1.2	3.06×10^{-02}
HSS	ESI-	920.87	4.2	high	Log	↑	2.7	1.05×10^{-02}
HSS	ESI+	361.20	5.7	very low	Log	↑	1.5	1.59×10^{-04}
HILIC	ESI+	397.20	0.7	low	UV	↓	1.2	4.92×10^{-02}
HILIC	ESI+	136.08	0.7	low	UV	↓	1.2	1.18×10^{-02}
HILIC	ESI+	147.07	0.7	low	UV	↓	1.2	1.18×10^{-02}
HILIC	ESI+	756.55	4.0	very low	Log	↑	1.4	4.74×10^{-03}

Table A3 *(cont.)*

UPLC column	Detection mode	m/z	RT(min.)	feature intensity	Model scaling/ transformation	Direction of variation	magnitude of variation	*p*-value
HILIC	ESI+	812.68	4.3	very low	UV	↑	1.7	9.49×10^{-03}
HILIC	ESI+	785.66	4.4	very low	Log	↑	1.8	6.01×10^{-03}
HILIC	ESI+	787.67	4.4	very low	Log	↑	1.4	1.46×10^{-02}
HILIC	ESI+	260.19	4.9	low	UV	↓	1.8	4.19×10^{-03}
HILIC	ESI+	169.99	4.9	low	UV	↓	1.4	4.19×10^{-03}
HILIC	ESI+	246.17	5.3	low	UV	↓	1.6	1.27×10^{-03}
HILIC	ESI+	527.25	7.2	low	UV	↓	1.7	6.01×10^{-03}

Annex VII - Lists of UPLC-MS unassigned features detected in maternal urine

Table A4 - Unassigned significant features obtained from PLS-DA controls *vs* FM models of 2nd trimester maternal urine UPLC-MS profiling analysis.

Maternal Urine - Fetal Malformations (FM)								
UPLC column	Detection mode	m/z	RT(min.)	feature intensity	model scaling	direction of variation	magnitude of variation	*p*-value
HSS	ESI-	203.07	0.63	very low	UV	↑	1.5	7.39×10^{-03}
HSS	ESI-	895.34	3.25	very low	UV	↑	1.5	3.11×10^{-03}
HSS	ESI-	239.07	3.3	very low	UV	↑	1.4	2.73×10^{-03}
HSS	ESI-	507.07	3.37	very low	UV	↑	2.7	1.37×10^{-03}
HILIC	ESI-	109.04	0.55	low	UV	↓	2.8	2.02×10^{-04}
HILIC	ESI-	879.36	0.69	very low	UV	↑	3.7	1.69×10^{-04}
HILIC	ESI-	769.35	0.70	very low	UV	↑	3.2	4.00×10^{-04}
HILIC	ESI-	799.39	0.70	very low	UV	↑	4.9	5.69×10^{-06}
HILIC	ESI-	813.38	0.71	low	UV	↑	4.0	1.09×10^{-06}
HILIC	ESI-	443.19	0.71	low	UV	↑	1.8	2.39×10^{-03}
HILIC	ESI-	829.38	0.71	low	UV	↑	3.7	1.96×10^{-06}
HILIC	ESI-	771.35	0.72	very low	UV	↑	2.6	6.47×10^{-04}
HILIC	ESI-	385.17	0.72	high	UV	↑	1.9	1.16×10^{-02}
HILIC	ESI-	857.42	0.72	very low	UV	↑	4.3	7.30×10^{-06}
HILIC	ESI-	859.42	0.74	very low	UV	↑	3.4	6.60×10^{-05}
HILIC	ESI-	281.08	1.35	very high	Pareto	↑	1.6	1.60×10^{-02}
HILIC	ESI-	617.21	5.16	low	UV	↑	1.8	4.02×10^{-03}