Computational models for conformations of cell wall mycolates from *Mycobacterium tuberculosis*

Wilma Prinsloo

Submitted in partial fulfillment of the requirements for the degree Master of Science
in the Department of Biochemistry
Faculty of Natural and Agricultural Sciences
University of Pretoria
Pretoria

May 2008

Acknowledgements

I would like to thank the following people and organisations for their specific contributions towards the completion of my degree:

My heavenly Father, thank you for Your awesome creation and for the priviledge that I have of discovering little details in Your intricate design. May my work be to the glory of your Name.

Professor Verschoor, thank you for allowing me to take this risk, it has been a rewarding experience. I appreciate your time and wise supervision.

Professor Baird, thank you for the opportunity to have been a part of your research group at the University of Bangor. I have been enriched by the experience.

Dr. Croft, thank you for taking a chance on me and for your patience all through my steep learning curve. I have enjoyed working with you and learning from you. I am also very grateful for the kindness you showed me during my stay in Bangor.

Thank you to the BIA group at the University of Pretoria for valuable discussions. I want to give a special thanks to Simon Thanyani for the biosensor sequence for my experiments and for instructing me on the use of the biosensor with such patience, and to Mrs. Sandra van Wyngaardt for the preparation of the mycolic acids and support in all aspects of life.

Dr. Davies, thank you for your guidance and interesting conversations concerning both the practical and theoretical aspects of molecular dynamics.

Dr. Chass, thank you for your work and ideas on the quantum mechanics work on mycolic acids.

Dr. Alexandra Simperler from Accelrys and my friends Jurgens de Bruin and Mykola Rozhok, thank you for writing scripts – without you I would never have been able to finish the analysis of my results in this lifetime.

Thanks to the NRF and University of Pretoria for financial support.

I am very grateful for all my friends that made me feel welcome on arrival in Bangor and during the completion of my work. I particularly want to thank everyone at BCC for being my family while mine was far away, for giving me lifts and for all your prayers. I miss you all.

Last but not least, a big thank you to my family for their support physically, spiritually and emotionally. Thank you for all your words of encouragement and prayers through good times and hard times. I love you.

Table of Contents

Acknowledgements ii
Table of Contents iv
List of Abbreviations vii
List of Figures x
List of Tables xiii
Summary xv
Chapter 1: General Introduction 1
1.1 Introduction 1
1.2 Mycolic acids in the mycobacterial cell wall 3
1.3 *Mycobacterium tuberculosis* in the host 5
1.3.1 Entry into host cell: phagocytosis 5
1.3.2 Prevention of endosomal fusion 7
1.3.3 Mycolic acids: lipid antigens 10
1.3.3.1 Lipid antigen presentation 10
1.3.3.2 Extraction of mycolic acids 11
1.3.3.3 Presentation of mycolic acid antigens on CD1b 13
1.3.3.4 Recognition of mycolic acids by antibodies 16
1.4 Responses to free mycolic acids 19
1.5 Folding of mycolic acids 22
1.6 Modeling of mycolic acid folding 28
Chapter 2: Biosensor measurement of the interaction between mycolic acids and cholesterol 30
2.1 Introduction 30
2.1.1 Cholesterol in tuberculosis 31
2.1.1.1 Cholesterol is essential for the uptake of mycobacteria 31
2.1.1.2 Cholesterol accumulates in macrophages: foam cells 32
2.1.2 Cholesterol effects on a lipid bilayer 33
2.1.2.1 Phase separations in cell membranes 34
2.1.3 The ESPRIT biosensor 35
2.2 Hypothesis 39
2.3 Aims: 39

2.4 Materials 40
2.4.1 Consumables 40
2.4.2 Buffers and solutions 41
2.4.3 Instrumentation 41
2.5 Methods: 42
2.5.1 Preparation of liposomes containing mycolic acids or cholesterol 42
2.5.2 Measuring cholesterol accumulation 43
2.5.3 Statistics 43
2.6 Results 45
2.6.1 Transfer of technology from IASYS to ESPRIT 45
2.6.2 Optimisation for quantitative measuments 47
2.7 Discussion 54
Chapter 3: Computer modeling of mycolic acids 59
3.1 Introduction 59
3.1.1 Computational chemistry 60
3.1.1.1 Quantum mechanical models 60
3.1.1.2 Molecular mechanics models 61
3.1.1.3 Conformational searching 62
3.1.2 Statistical analyses 63
3.1.2.1 Principal Component Analysis (PCA): 64
3.1.2.2 Self-Organised Maps (SOMs) 66
3.1.3 Modeling of mycolic acids 66
3.2 Hypothesis 70
3.3 Aims 70
3.4 Methods 71
3.4.1 Quantum mechanics 71
3.4.2 Molecular dynamics 71
3.5 Results 76
3.5.1 Quantum mechanics 76
3.5.2 Molecular dynamics 80
3.5.2.1 Selection of force field 80
3.5.2.2 Equilibration of the system 81
3.5.2.3 Principal Component Analysis 82
3.5.2.4 Self-Organising Maps 86

3.5.2.5 "WUZ" analysis 90
3.6 Discussion 101
Chapter 4: Concluding Discussion 105
References 113
Appendices 122
Appendix A 122
Trans, gauche and eclipsed conformations 122
Appendix B 124
Results for semi-empirical calculations on cyclopropanes as methylene chain lengths are increased 124
Appendix C 126
Potential Energy Surface data and energy calculations 126
Appendix D 131
SOM-plots 131
Appendix E 136
Example of overlap in SOM-cluster of WUZ-groups 136

List of Abbreviations

ABC	ATP-Binding Cassette
AFM	Atomic Force Microscopy
AG	Arabinogalactan
Ag85	Antigen 85
AIDS	Acquired Immuno-Deficiency Syndrome
AmB	Amphotericin B
APC	Antigen Presenting Cell
ATP	Adenosine Triphosphate
BCG	Bacille Calmette-Guerin
°C	Degree Celcius
C	Complement
Ca	Calcium
CD	Cluster of Differentiation
CF	Cord Factor
CM	Cytoplasmic Membrane
CmaA2	Cyclopropane mycolic acid synthase 2
CR	Complement Receptor
D	Dihedral
DC	Dendritic Cell
df	degrees of freedom
DNA	Deoxyribose Nucleic Acid
E	Energy
EEA1	Early Endosomal Autoantigen 1
ELISA	Enzyme-Linked Immunosorbent Assay
EtOH	Ethanol
FAS	Fatty Acid Synthetase
g	gram
GL	Glycolipid
GMM	Glucose Monomycolate
GPI	Glycosylphosphatidylinositol
H	Hamiltonian operator

HF	Hartree Fock
HIV	Human Immuno-deficiency Virus
hma	Hydroxy Mycolic Acid gene
IL	Interleukin
INF	Interferon
IRD	Immune Reconstitution Disease
k	Pseudo first order rate constant
K	Kelvin
k_{ass}	Association constant
K_D	Dissociation equilibrium constant
k_{diss}	Dissociation constant
ℓ	litre
LAM	Lipoarabinomannan
LCAO	Linear Combination of Atomic Orbitals
ld	liquid disordered
lo	liquid ordered
LPS	Lipopolysaccharide
M	Molar
MA	Mycolic Acid
MAC	*Mycobacterium avium-intracellulare-scrofulaceum* complex
ManLAM	Mannose-capped LAM
MC	Monte Carlo
Mce	Mammalian cell entry
MD	Molecular Dynamics
MDR-TB	Multi-drug-resistant TB
ME	Methyl ester
MHC	Major Histocompatibility Complex
min.	Minute
mM	millimolar
MM	Molecular Mechanics
MeO	Methoxy
MR	Mannose Receptor
MS	Mass Spectroscopy
M. tb	*Mycobacterium tuberculosis*

nm	nanometer
NMR	Nuclear Magnetic Resonance
ns	nanosecond
ODT	Octadecanethiol
PBS	Phosphate-buffered saline
PC	Phosphatidylcholine
PcaA	Proximal cyclopropane of alpha-mycolates
PCA	Principal Component Analysis
PEHS	Potential Energy Hypersurface
PES	Potential Energy Surface
PG	Peptidoglycan
PIM	Phosphatidylinositol mannoside
Pks	Polyketide synthase
QM	Quantum Mechanic
R	Response in millidegrees
RA	Reflection-absorption
SDS	Sodium Dodecyl Sulfate
SERS	Surface-enhanced Fourier transform Raman Spectroscopy
so	solid ordered
SOM	Self-organising map
SPR	Surface Plasmon Resonance
TACO	tryptophane aspartate containing coat protein
TB	Tuberculosis
TBneg	TB negative
TBpos	TB positive
TCR	T-Cell Receptor
TDM	Trehalose dimycolate
T_m	melting (or gel-to-liquid transition) temperature
TMM	Trehalose Monomycolate
TNF	Tumor Necrosis Factor
WHO	World Health Organisation
XDR-TB	Extremely-drug resistant TB

List of Figures

Figure 1.1 *Schematic representation of the arrangement of components in the cell wall of mycobacteria..* 3

Figure 1.2 *Mycolic acid structure.* 4

Figure 1.3 *Model of the role of TACO in the survival of intracellular mycobacteria..* 8

Figure 1.4 *Model of the action of Saposins in loading glycolipids onto CD1 molecules..* 12

Figure 1.5 *A summary of the antigenic determinants of the TCR-response to natural GMM.*.....14

Figure 1.6 *A hypothetical model for the formation of a combinatorial epitope in keto-MA..* 15

Figure 1.7 *The top and side views of the crystal structure of human CD1b loaded with the mycobacterial lipid antigen glucose monomycolate.* 16

Figure 1.8 *CD1a-, CD1b-, and CD1c-restricted T cells in antimicrobial immunity.* .20

Figure 1.9 *Schematic drawings of the α-MA molecule in its monolayer at different surface pressures.*.....23

Figure 1.10 *Phase diagrams of Langmuir monolayers of α-, MeO- and keto-MA.* 27

Figure 2.1 *Theoretical phase diagram for DPPC:cholesterol..*.....34

Figure 2.2 *The SPR setup.*.....36

Figure 2.3 *A schematic representation of the ESPRIT configuration.* 37

Figure 2.4 *Plot of k_{on} at different ligate concentrations from which k_{ass} and k_{diss} can be determined.* 38

Figure 2.5 *A representative profile for the sequence of events of the experiment designed to show MA-cholesterol interaction on the ESPRIT biosensor.* 45

Figure 2.6 *The difference between accumulation from cholesterol containing liposomes (33 % cholesterol in PC) and empty (only PC) liposomes on the ESPRIT biosensor.* 46

Figure 2.7 *The average cholesterol accumulation is shown for different concentrations of cholesterol liposomes accumulating on MA-liposomes.* 49

Figure 2.8 *The decline in MA liposome immobilisation in repeated cycles of measuring cholesterol binding to MA.* 50

Figure 2.9 *Reliability of MA immobilisation on the ESPRIT biosensor with successive experiments.* 51

Figure 2.10 *The difference between cholesterol accumulation and PC accumulation on the ESPRIT biosensor for the adjusted protocol for liposome preparation and surface regeneration.* 52

Figure 2.11 *The average cholesterol accumulation shown for different concentrations of cholesterol liposomes accumulating on MA-liposomes.* 53

Figure 3.1 *Principal component analysis of a pen.* 64

Figure 3.2 *Molecular dynamics of MeO-MA.* 67

Figure 3.3 *Structures and specific stereochemistry of the MAs that were modeled.* .72

Figure 3.4 *The five points used in the analysis of MA-structures shown on a keto-MA.* 74

Figure 3.5 *A schematic representation of the seven possible W, U and Z folds.* 75

Figure 3.6 *The bond lengths and angles of the cyclopropane molecules that were measured.*.....76

Figure 3.7 *A schematic representation of the rotation seen at an imaginary frequency for disubstituted trans-cyclopropane.*.....77

Figure 3.8 *Dimethyl cis-cyclopropane showing the numbering system that was used in the PES calculations for this molecule.* 78

Figure 3.9 *A PES calculated for dimethyl cis-cyclopropan.*.....78

Figure 3.10 *The first derivative of Pressure-Potential Energy plots of MAs.* 82

Figure 3.11 *The PCA plots for α-MA.* 83

Figure 3.12 *PCA plots of α-MA, MeO-MA, MeO-ME and Keto-MA.* 84

Figure 3.13 *Self Organising Map of one representative α-MA MD simulation.*.....91

Figure 3.14 *The seven folds represented as minimised structures from α-MA.*.....96

Figure 3.15 *The sum of distributions for the bd distance of each MA-type.*.....99

Figure 3.16 *Keto-MA shown in the W-fold.* 100

List of Tables

Table 2.1 *Liposome compositions with various cholesterol concentrations.* 42

Table 2.2 *Accumulation of cholesterol liposomes on MA liposomes.* 46

Table 2.3 *A quantitative determination of cholesterol that accumulated on MA liposomes.* 48

Table 2.4 *Statistical calculations done on the quantitative data obtained for cholesterol accumulation.* 49

Table 2.5 *Quantitative binding data of cholesterol to immobilised MA on ESPRIT biosensor after optimisation of conditions.* 52

Table 2.6 *Statistical calculations done on the quantitative data obtained from the optimised experimental method for cholesterol accumulation.* 53

Table 3.1 *The parameters for the eight selected distances in a MA molecule for the different types of W, U and Z folds.* 75

Table 3.2 *The stabilised bond lengths and angles measured as the chain length was increased in a symmetrical way for cis- and trans-cyclopropane.* 76

Table 3.3 *The stabilised bond lengths and angles measured as the chain length was increased in a symmetrical way for R- and S diastereomers of α-methyl trans-cyclopropane* 79

Table 3.4 *The square of the differences of the bond lengths and angles as calculated for cis- and trans-cyclopropane with two 16-carbon chains, with the respective force fields.* 81

Table 3.5 *The ten most highly populated clusters in SOM for 8000 frames of each MA type.* 87

Table 3.6 *The seven most highly populated clusters in SOM for 5020 frames of each MA type.* 89

Table 3.7 *Frequency of occurrence of WUZ-folds during last 251 ns of molecular dynamics for Alpha-MA* 93

Table 3.8 *Frequency of occurrence of WUZ-folds during last 251 ns of molecular dynamics for MeO-MA* 93

Table 3.9 *Frequency of occurrence of WUZ-folds during last 251 ns of molecular dynamics for MeO-ME* 94

Table 3.10 *Frequency of occurrence of WUZ-folds during last 251 ns of molecular dynamics for Keto-MA* 94

Table 3.11 *The total Potential Energy of the minimised structures of the different folds in α-MA, in descending order.* 96

Summary

Literature highlights the effects of mycolic acid (MA) fine-structure on biological activity, pathogenicity, virulence and cell wall structure and permeability. Knowledge on MA-structure and how their conformations are dependent on their precise molecular composition becomes essential in exploiting these properties in drug-design and in advancing our understanding of the disease. In our group evidence for a structural or functional relationship between cholesterol and MAs have been discovered. The aim of the experimental part of this work was to study this relationship further by attempting to quantify the interaction between cholesterol and MAs in liposomes on an evanescent field biosensor. The binding profiles that were obtained could not be evaluated with kinetic software and the interaction between cholesterol and MAs was not linearly dependant on the concentration of cholesterol. However, novel insight into the interaction was gained when it was observed that cholesterol only accumulated on MA liposomes when cholesterol liposomes containing concentrations of cholesterol resulting in a suspected liquid ordered phase, were used. This is significant since it implies that cholesterol in membrane rafts of the host cell that exist in a liquid ordered phase would be able to interact with MAs under physiological conditions. The theoretical part of this work represents the first molecular modeling study in which MAs are allowed to fold with no conformational restrictions. It is proposed that MAs fold as a function of their functional groups, stereochemistry, and various chain lengths. It was also investigated whether methylation of the acid group changes conformational preferences. The effect of chain length on cyclopropane structure and the viability of systematic conformational searching in MAs were shown using quantum mechanics. Replicate molecular dynamics simulations were done for 4 ns *in vacuo* on alpha-; methoxy-; methoxy methyl ester- and keto-MAs. MAs had an open starting conformation without conformational restrictions. Results were analysed using eight distances characteristic of the conformational fold. Using these distances, W-, U- and Z-shaped folds were identified. Principal component analysis (PCA) and self-organising maps (SOMs) were used to evaluate differences and trends in MA-conformations. Quantum chemical results showed that chain length did not affect cyclopropane structure and that the systematic plotting of potential energy surfaces is an effective tool to analyse

effects of changes in geometry on the energy of the molecule and to predict favoured conformations. Remarkably, single MAs assumed W-, U- and Z-folds *in vacuo* during molecular dynamics simulations that have previously been observed in monolayers. PCA and SOM plots showed that keto-MA folded faster than other MAs. Alpha-MA showed the highest frequency of W-, U- and Z-folds. Methoxy-MA did not readily fold at its cis-cyclopropane group. Methylation of the acid group of methoxy-MA did not show remarkable differences in the conformations assumed, but almost doubled the frequency of WUZ-structures obtained as compared to non-esterified methoxy-MA. The inherent structural differences between MA-subclasses clearly affect the trends in structural folds that they assume. Molecular modeling of MAs proved to be a versatile tool for resolving structure-function relationships at the molecular level.

Chapter 1: General Introduction

1.1 Introduction

In Africa, tuberculosis (TB) has already caused higher mortality in 2002 than ever registered before for this disease. This is largely due to the high occurrence of human immunodeficiency virus (HIV) [1], which increases both the incidence of TB infection, as well as the risk of dying from TB, because HIV weakens the immune system dramatically [2]. People without HIV will have a smaller chance of getting sick when infected with TB, because their immune system is strong enough [2]. If control over TB is not improved drastically, it is estimated that between the year 2000 and 2020, 1 billion new infections will occur, 200 million people will develop tuberculosis and 35 million people will die from the disease [3].

After infection, the TB bacilli normally go dormant in a person for months, that may extend into years, and reactivation to cause active tuberculosis is typically the result of an existing state of immunosuppression [3, 4]. A minimum of six months of antibiotic treatment is needed to clear the body of the dormant mycobacteria [3, 5]. TB patients who do not comply to the treatment regime over this period, during which a combination of drugs have to be taken regularly, are a major cause of the mycobacteria becoming drug resistant [3]. A very dangerous form of drug-resistant TB is extremely-drug-resistant TB (XDR-TB) that has now become prevalent in South Africa [1]. This occurs when the mycobacteria are resistant to the two most important first-line drugs (isoniazid and rifampicin as in MDR-TB), and also to a fluoroquinolone and at least one second-line injectable agent [1]. XDR-TB therefore describes a pattern of resistance in mycobacteria for which patients have a much smaller chance of being successfully treated with existing second-line drugs [1]. Treatment results for the year 2004 reflected the effects of HIV and inadequate patient support in African countries, which spurred the World Health Organisation (WHO) to

emphasize the necessity of improving collaborative TB/HIV programs and the management of drug-resistance [1]. It is therefore evident that the management of TB is a multi-facetted challenge that needs the best of modern science to stem its tide.

TB is the most frequent cause of rapid death in HIV-infected individuals, especially in sub-Saharan Africa where HIV infection abounds [6]. There is a very big risk for patients, co-infected with TB and HIV, to develop immune reconstitution disease (IRD) if they start antiretroviral treatment within two months of antituberculosis treatment [7]. IRD presents as worsening symptoms of TB or other opportunistic infections in HIV-infected patients and is directly attributable to their improved immune responses when they start antiretroviral therapy [7]. The rapidity with which antiretroviral treatment reverses the impact of HIV in the host response to TB, possibly combined with a lack of immunoregulatory mechanisms, may lead to the life-threatening, uncontrolled tissue-damaging responses that characterise IRD [7]. Therefore it is imperative that HIV-positive patients co-infected with TB are diagnosed quickly and placed on an anti-TB treatment regime in order to bring the TB-infection under control before antiretroviral therapy is started. Early TB-diagnosis in such patients may make antiretroviral roll-out safer and curb mortality due to AIDS and TB dramatically.

The classical methods for reliable diagnosis of TB since the 1890s, including the tuberculin skin test, smear positive microscopy, sputum culture and radiology are all less sensitive in HIV-co-infected individuals [8]. An ideal diagnostic for TB in HIV populations would not depend on immunological surrogate markers that were impaired by HIV infection. It also needs to be fast, providing results within 24 hours after sampling, because it has been shown that 70% of HIV patients co-infected with XDR-TB died within 30 days after a sputum sample was collected from them [9]. This demonstrates the need for a quick and accurate diagnosis in patients with the lethal combination of HIV and TB.

Although not a single serodiagnostic assay for TB has been developed to live up to the required sensitivity and specificity in HIV burdened test populations to date, an opportunity did arise in our research group with a test to detect antibodies to mycolic acid (MA) antigens occurring in the cell wall of *Mycobacterium tuberculosis* [10].

The antibody response to MAs appeared not to be affected by the severity of AIDS in the patient, but the test was still not adequate for reliable diagnosis. Our group focussed on improving antibody detection technology as well as learning more about the immunobiology of MAs in order to implement the application of this discovery in diagnostic and other means to control the TB epidemic in Africa.

This dissertation emanates from the vision of our research group to understand the role of MAs in tuberculosis that may lead to new scientific approaches to manage the epidemic.

1.2 Mycolic acids in the mycobacterial cell wall

The mycobacterial cell wall is exceptionally rich in lipids, which constitute about 60% of its weight [11]. MAs are among these lipids, which are the major components of the mycobacterial cell envelope, and are exposed to the external environment. The lipids in the envelope will most likely make first contact with the host cells, and therefore are of special interest in the molecular knowledge of TB.

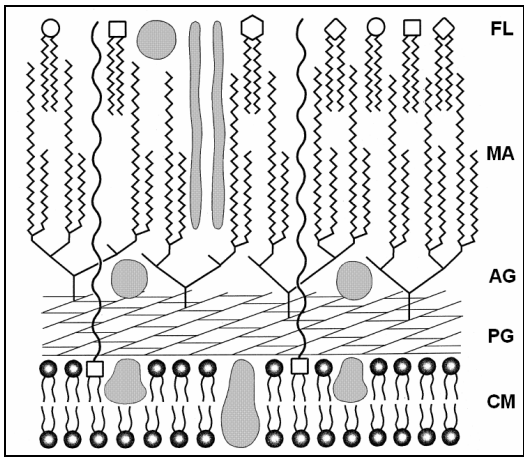

Figure 1.1 *Schematic representation of the arrangement of components in the cell wall of mycobacteria. CM, cytoplasmic membrane; PG, peptidoglycan; AG, arabinogalactan; MA, mycolic acid; FL free lipids (various species indicated by different symbols). The cell wall contains proteins indicated by shaded areas, and lipoarabinomannan, possibly linked to the membrane by a phosphatidylinositol lipid. Modified from [12].*

As shown in Figure 1.1 the inner leaflet of the mycobacterial cell wall consists of peptidoglycan linked to arabinogalactan, (both orientated perpendicularly to the cell membrane), and MAs covalently linked to arabinogalactan. This allows MAs to arrange into an ordered monolayer, contributing to the impermeability of the cell wall. The outer leaflet composed of complex free lipids interacts hydrophobically with the MAs, to produce a lipid bilayer [13].

MAs are high-molecular-weight 2-alkyl, 3-hydroxy fatty acids occurring mostly covalently bound to arabinogalactan. Some can also be found as trehalose monomycolates (TMM) and trehalose dimycolates (TDM, also known as cord factor, CF) [14].

MAs consist of two characteristic parts:

1. a saturated carboxylic acid part containing the mycolic motif, and
2. a long fatty alcohol part, the meromycolate chain (Figure 1.2)

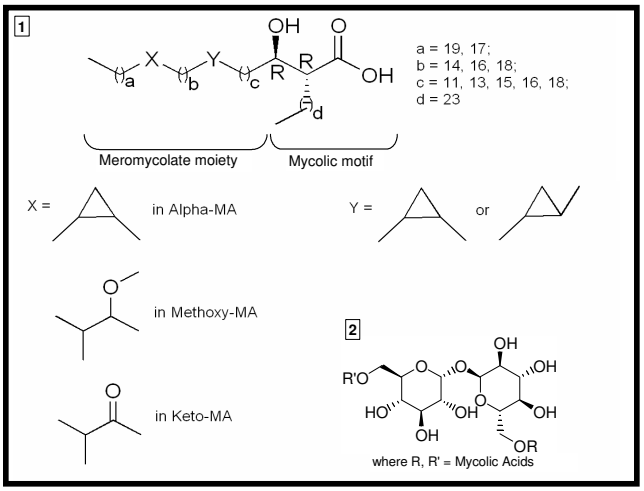

Figure 1.2 *Mycolic acid structure (1), trehalose monomycolate (2) (R = MA, R' = H) and trehalose dimycolate (2) (R, R' = MA).*

On the basis of the nature of the functional groups present in the meromycolate chain, MAs from *M. tb* are organized into three subclasses:

- α-MA with no oxygen-containing group,
- methoxy-MA (MeO-MA) with a methoxy group in the distal position and
- keto-MA, characterized by a carbonyl group in the distal position.

Mycobacteria and related genera possess many unique characteristics which are directly attributed to the presence of MAs in their cell wall, including resistance to chemical injury, low permeability to hydrophobic antibiotic substances, resistance to dehydration and the ability to persist within the environment of the host macrophage [15]. MA-containing bacteria occur in the genera *Mycobacterium*, *Nocardia*, *Rhodococcus* and *Corynebacterium* [16]. The distribution of MA-types varies within species. For example in the genus *Mycobacterium*, *M. tb* and *M. bovis* contain α-, MeO- and keto-MA subclasses while in *M. avium*, *M. intracellulare* and *M. phlei* MeO-MA is replaced by wax ester mycolates [16]. MAs also show variation in size. *Mycobacteria* have very long MAs with 60-90 carbon atoms, *Nocardia* have 46-60, *Rhodococcus* have 34-64 and *Corynebacterium* have short MAs ranging from 22-38 carbon atoms [16]. While the alpha-chain of MAs from *M. tb* is 24-carbons in length, other species such as *M. avium*, *M. leprae* and *M. phlei* have a 22-carbon alpha-chain [16].

The chain lengths and stereochemistry of MAs determine their preferred conformations and will therefore affect the way they fold and pack in the cell wall. From Langmuir monolayer studies, it was suggested that keto-MA facilitates a W-conformation in which all four alkyl chains are parallel to each other when it bends back onto itself at each functional group [17, 18]. Less compactly folded structures, such as those which only fold at one or none of its functional groups in the mero-chain, can then be described as Z-, or U-folds, respectively (personal communication with Prof. D.E. Minnikin, University of Birmingham, UK, June 2006).

1.3 *Mycobacterium tuberculosis* in the host

1.3.1 Entry into host cell: phagocytosis

Cells take up particles and solutes by pinocytosis, receptor-mediated endocytosis and phagocytosis. Pinocytosis is usually the uptake of fluids and solutes while receptor-

mediated endocytosis is the specific process through which macromolecules, viruses and small particles enter cells. Both pinocytosis and receptor-mediated endocytosis use a clathrin-based mechanism that is usually independent of actin polymerisation. In contrast, phagocytosis is the uptake of large particles into cells and its mechanism is usually independent of clathrin, but actin-dependent. Monocytes, macrophages and neutrophils are referred to as professional phagocytes because they are very effective at engulfing particles. Phagocytosis by macrophages is critical for the uptake of infectious agents and it participates in the immune response and inflammation [19]. Due to the main route of entry of mycobacteria being through the respiratory track, the macrophages in the lung are primarily involved in the initial uptake of mycobacteria.

Phagocytosis of *M. tb* makes use of different receptors on the phagocytic cell [20]. Complement receptors (CR) and complement opsonisation of mycobacteria play the most important roles in mediating phagocytosis [21]. Schlesinger *et al*. have shown that CR1, CR3 and complement component C3 account for the majority of adherence and uptake of *M. tb* to monocytes [22]. Blocking of CR1 and/or CR3 by monoclonal antibodies inhibits adherence of *M. tb* to monocytes by up to 81% and blocking of complement component C3 by 71%. Another way that mycobacteria enter macrophages was shown to occur by a mycobacteria-associated C2a cleavage of C3, resulting in C3b opsonisation of the mycobacteria and recognition by macrophages [23]. CR3 has also been shown to bind to *M. tb* that is not opsonised [24]. This nonopsonic binding was shown to be mediated by outer capsular polysaccharides of the mycobacteria and that it is likely to occur at the β-glucan site of CR3 [25].

Although CR3 can bind to nonopsonised mycobacteria, the mannose receptor (MR) is the major contributer to nonopsonic phagocytosis of *M. tb* [21]. MRs recognise and bind to the terminal mannosyl units of lipoarabinomannan (LAM) [26]. It has been shown that unlike CRs, MRs are only utilised by virulent strains of *M. tb* in macrophage phagocytosis [21] and it was suggested that there may be other subtle structural differences in LAM among strains because they all had terminal dimannosyl units [27]. Other receptors that are involved in phagocytosis of *M. tb* to lesser degrees are scavenger receptors, Fc-receptors and pattern recognition molecules such as CD14 and toll-like receptors [20, 28, 29].

Cholesterol also plays an important role in the entry of mycobacteria into host cells. Gatfield and Pieters [30] showed that a decrease in cholesterol levels of 60-70% in macrophages caused a reduction of *M. tb* uptake of ~85% and the percentage of phagocytosing cells was reduced to ~15%. Macrophages depleted of cholesterol were still able to internalise various other microorganisms, indicating that the reduction in uptake by cholesterol depletion was specific for mycobacteria. Cholesterol is involved in the regulation of membrane fluidity, and therefore it is possible that its depletion may affect other cellular functions. However, the cholesterol depleted cells only showed a slight increase in motility (consistent with cholesterol's role of providing rigidity to the cell membrane) and no other abnormalities. The state of receptors located in cholesterol rafts in the cell membrane may also affect uptake of mycobacteria, but it was shown that CR3-mediated uptake of *Lactobacillus casei* was not affected in cholesterol-depleted cells. The authors suggested that the apparent cholesterol-dependence of *M. tb* may be explained by the requirement of cholesterol to produce a stable physical interaction of the mycobacteria with the membrane, because mycobacteria showed a higher cholesterol binding capacity than other microorganisms.

In contrast to Gatfield and Pieters [30], another group [31] demonstrated that the resistance of cholesterol-depleted cells to phagocytose mycobacteria was due to an effect on receptor signalling, rather than direct binding to cholesterol. They showed that for the internalisation of *M. kansasii* by neutrophils to occur, CR3 needs to associate with proteins that are anchored in the membrane by glycosylphosphatidylinositol (GPI) in cholesterol-rich domains. In contrast, zymogen (glucan and mannan particles from the wall of *Saccharomyces cerevisiae*), opsonised or nonopsonised, could be phagocytosed by CR3 that is not associated with GPI-linked protein when it is outside of cholesterol-rich domains.

1.3.2 Prevention of endosomal fusion

It is characteristic of all pathogenic *Mycobacterium* species to survive in their host phagocytes [32]. The inhibition of phagosome maturation may secure the survival of

mycobacteria by limiting the hydrolytic capacity of the phagosome [33, 34], ensuring access to nutrients [35] and preventing contact with the antigen-presenting machinery of the host cell [36]. As shown in Figure 1.3, Ferrari *et al*. have postulated that live mycobacteria prevent phagosome-lysosome fusion by retaining TACO (for tryptophan aspartate-containing coat protein), preventing their degradation, whereas dead bacteria released TACO and were degraded [37].

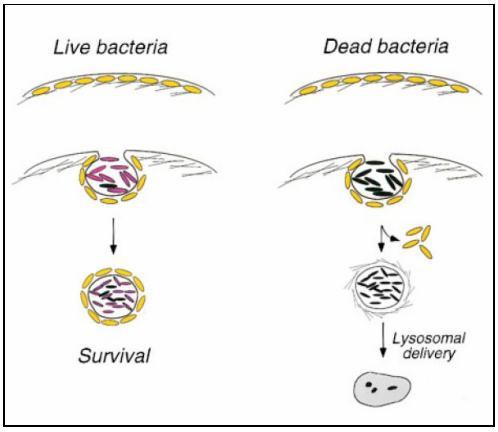

Figure 1.3 *Model of the role of TACO in the survival of intracellular mycobacteria. TACO (yellow) is recruited to the phagosome by phagocytosis of mycobacteria. TACO is actively retained at the phagosome with live (violet) mycobacteria but is rapidly released from phagosomes containing dead (black) mycobacteria, followed by lysosomal delivery of the bacilli [37].*

Recent studies have also examined the proposed role of TACO, and a number of findings oppose the role of TACO to stay associated to phagosomes in order to inhibit phagosome-lysosome fusion and to ensure the survival of mycobacteria in the host cell:

- Schüller *et al*. [36] found that TACO (also called coronin) was gathered at the site of entry of mycobacteria but was released from the phagosomal membrane shortly after internalisation of the mycobacteria. Therefore, it was shown that coronin is not responsible for the inhibition of phagosome maturation. This was consistent in both human and mouse macrophages where mycobacteria were taken up as isolated bacilli. Interestingly, coronin was retained only when phagosomes contained clumps of mycobacteria. It is possible that the theory of Ferrari *et al*. [37] was based on a similar observation in mouse macrophages (Figure 1.3).

- While mycobacteria in the phagosome of macrophage cells are able to grow and access recycling endosomes that may contain nutrients, mycobacteria in phagosomes of dendritic cells (DC) had reduced communication with the host cell recycling pathways and did not grow - even though they were not killed. Coronin was also present in DC phagosomes, and therefore coronin cannot be the molecule that ensures intracellular survival of mycobacteria as it did not protect it from bacteriostatic action in the DC [38].
- As a model for mycobacterial diseases, Solomon *et al.* [39] used *Mycobacterium marinum* (causing tuberculosis-like infections in fish and amphibians) to infect *Dictyostelium discoideum* amoeba cells which are highly phagocytic and that showed growth that resembled the growth of *M. marinum* in macrophages. They concluded that the coronin isoform of *D. discoideum* cannot mediate the establishment or maintenance of a replication vacuole for *M. marinum* because on the contrary, coronin mutants showed enhanced intracellular growth of mycobaceria. This could be a result of increased uptake of the bacteria or enhanced survival after phagocytosis by coronin mutant cells.
- Virulent mycobacteria were shown to disrupt the actin filament network in macrophages one day after infection by forming large patches instead of the small filaments that are normally found [40]. Since coronin is known to be an actin binding protein, the ability of virulent mycobacteria to disrupt the actin network does not support the continued association of coronin at the phagosome. The disruption of actin is also not involved in inhibiting maturation of phagosomes, since the authors showed that latex beads that did not disrupt actin also inhibited phagosome maturation.

Current knowledge points to the important role that Ca^{2+} plays in phagosome-lysosome fusion [33, 34]. Malik *et al.* have shown that infection of human macrophages with *M. tb* inhibits Ca^{2+}-mediated signal transduction [33]. They also determined that the inhibition of a cytosolic increase in Ca^{2+} depended on the viability of mycobacteria and on CR-mediated phagocytosis. Dead mycobacteria and the redirection of phagocytosis by opsonisation with antibodies resulted in increased cytosolic Ca^{2+} levels and increased phagosome-lysosome fusion. Later, they also

demonstrated the mechanism for the inhibition in Ca^{2+}-signaling [41] and determined the role of down-stream Ca^{2+}-dependent effector molecules [34].

Several studies show that mycobacterial components such as CF and LAM play a role in mediating phagosomal maturation arrest, but the increased phagosomal maturation seen with dead mycobacteria suggests that other factors are also involved [33]. CF has been shown to be very effective at inhibiting Ca^{2+}-induced fusion between phospholipid vesicles *in vitro* [42] although its molecular components, trehalose and MAs showed no effect. There was also no synergism between the separate MA and trehalose components, but only the covalently bound MA and trehalose, CF, showed inhibition. A more recent study [43] also proposed that CF plays an important role in intracellular trafficking and survival of mycobacteria. It was shown that the removal of surface lipids (including CF) led to enhanced trafficking of organisms to acidic compartments and decreased bacterial survival. Reconstitution with the crude extract of lipids or pure CF reversed these responses. Polystyrene beads coated with CF showed similar intracellular trafficking to normal mycobacteria and decreased fusion of phagosomes with lysosomes. Mannose-capped LAM (ManLAM, found in cell walls of virulent mycobacteria) also decreased maturation of phagosomes when coated onto latex beads and did not recruit EEA1 (early endosomal autoantigen 1) that bridges membranes for fusion [44, 45].

1.3.3 Mycolic acids: lipid antigens

1.3.3.1 Lipid antigen presentation

Here, a brief discussion on how protein antigens are produced, presented and recognized by antibodies is done in order to clarify the aspects that will be discussed next, regarding the presentation of MAs as lipid antigens and their recognition by antibodies.

The immune system makes use of MHC (major histocompatibility complex) class I and class II molecules to present antigens on the surface of APCs to T-cells [46].

MHC class I molecules bind to and display endogenous antigens that were synthesized inside the cell. MHC class II molecules recognize and display antigens that are foreign to the body, from an exogenous source. The MHC-antigen complex is recognized by a T-cell receptor (TCR). T-cells that recognize antigen displayed on MHC class I molecules display a CD8 surface marker and are cytotoxic T-cells. The T-cells that recognize antigen on MHC class II molecules normally help or induce immune responses and are called T-helper cells. They have a CD4 surface marker. MAs are lipid antigens and unlike protein antigens they are displayed by CD1 molecules and not on MHC molecules. The CD1-antigen complex is in turn recognised by T-cells without CD4 and CD8 surface markers.

When an exogenous source of antigen, for example a microorganism is phagocytosed, the phagosome fuses with the lysosome where the proteins are partially digested by enzymes [46]. MHC class II molecules are also localized to these compartments where they are loaded with a peptide from the microorganism into their antigen-presenting groove. The MHC-antigen complex is transported to the cell's surface where it is recognized by $CD4^+$-restricted T-cells. The clustering of TCRs during this interaction causes activation of intracellular cascades resulting in an activated T-cell.

B-cells are activated to produce antibodies by two main interactions [46]. The first is similar to T-cells and occurs by the recognition of soluble antigen by immunoglobulin on the B-cell's surface that leads to clustering of the receptors and activation of downstream intracellular cascades. The B-cell, being an APC, internalizes and degrades the microorganism to produce antigens that are in turn displayed on MHC class II on the B-cell's surface. The second interaction that activates B-cells involves binding of the activated T-cells to their respective antigen-MHC-II complex that is displayed on the B-cell's surface. Other cell-surface molecules and cytokines secreted by the $CD4^+$ T-cells further stimulate the B-cells to produce antibodies.

1.3.3.2 Extraction of mycolic acids

It is known that mycobacterial lipids are presented to T-cells on CD1 molecules [47]. However, it was unknown how these large hydrophobic lipids such as MAs were

loaded onto CD1 molecules that are exposed to a hydrophilic environment until 2004 when the mediator at this interface, Saposin C, was discovered [48].

Saposins A-D are derived from proteolytic cleavage of their common precursor, prosaposin, and are found mainly in lysosomes where they assist in the catabolism of glycosphingolipids by binding directly to the enzymes or the lipid substrates [49]. Saposin C was known to act as cofactor of the enzyme β-glucosidase which degrades glucosylceramide, by binding directly to the enzyme [50].

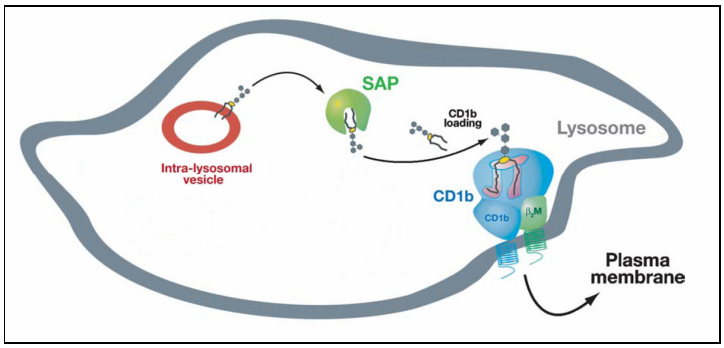

Figure 1.4 *Model of the action of Saposins in loading glycolipids onto CD1 molecules. The Saposins extract lipids or glycolipids from inner lysosomal membranes and transfer them to immunoreceptors of the CD1 family on the perimeter of the membrane. The loaded CD1 molecules then travel to the plasma membrane where they activate lipid-antigen specific T-cells. Modified from reference [51].*

Winau and co-workers used Saposin-deficient and normal fibroblasts both transfected with the human gene encoding CD1b to show the involvement of Saposins in antigen presentation [48]. It was shown that Saposin-deficient fibroblasts could not stimulate CD1b-restricted T-cells with LAM, GMM or MA. When the Saposin-deficient fibroblasts were reconstituted with soluble Saposins, it was found that only Saposin C restored T-cell activation to the *M. tb*-derived antigens. They also showed that mycobacterial lipids, CD1 molecules and Saposin C co-localised in lysosomes, and therefore it is likely that the antigens are loaded onto CD1b molecules in these acidic compartments [52]. Further they showed that Saposin C made lipid antigens accessible for loading onto CD1 molecules when it extracted LAM form liposomes and also that Saposin C specifically bound to CD1b. A model for the proposed action of Saposin in loading molecules onto CD1 is shown in Figure 1.4.

1.3.3.3 Presentation of mycolic acid antigens on CD1b

CD1 molecules are structurally related to MHC class I molecules, but differ from MHC in that they do not have structural variability. They occur on APCs, including B-cells. The CD1 genes consist of five closely linked genes of which four are expressed, encoding group 1 (CD1a, CD1b, CD1c) and group 2 molecules (CD1d). Group 1 molecules have a function similar to MHC class II molecules because they present lipid and glycoplipid antigens that are foreign to the host, while most CD1d molecules present self antigens, although they also present lipids from parasites [46]. Cd1b ligands include the mycobacterial lipids; MAs, GMM, LAM, PIM (phosphatidyl-inositol mannosides) and sulfoglycolipids [47].

Porcelli's research group has done numerous studies on the structural requirements of GMM (Figure 1.5) as lipid antigen for recognition by CD1b-restricted T-cells [52-54]. These studies assist in plotting the way in which antigens are presented on CD1b by determining which parts of the antigens are recognised by the T-cell receptor (TCR). This group have determined numerous structural requirements for the recognition of GMM by CD1b-restricted T-cells, as summarised in Figure 1.5. They hypothesised that this antigen may be restricted to mycobacteria that have parasitized host cells efficiently, since the GMM molecule is comprised of host glucose and mycobacterial MA [54].

The recognition of GMM by T-cells was not specific to the hydrophobic part of MAs. GMM isolated from different species (differing in acyl chain length, meromycolate functional groups, double bonds and cyclopropane rings) as well as a synthetic GMM with a short 32-carbon MA all produced good T-cell responses [53]. T-cell responses were, however, sensitive to the MA structure adjacent to glucose. T-cells did not respond to non-mycolyl glycolipids and neither to a GMM analog that lacked an alpha-chain, nor to GMM with removed or derivatised β-hydroxyl group [53]. The linkage of MA at carbon-6 of glucose and the R,R-stereochemistry at carbon 2 and 3 in the MA was also crucial for a T-cell response [54].

The carbohydrate moiety showed very specific recognition; neither free MA nor MA glycosylated with glycerol, trehalose, arabinose, or phospholipids elicited a T-cell response [53, 54]. Sensitivity for the carbohydrate group was so high, that T-cells were able to discriminate between stereoisomers of glucose (mannose and galactose) varying only in the orientation of a hydroxyl group on the pyranose ring [53]. It is suggested that the lipid antigens are not chemically or enzymatically cleaved in APCs, since APCs were not able to cleave TDM (CF) into GMM and it did not elicit a T-cell response [52].

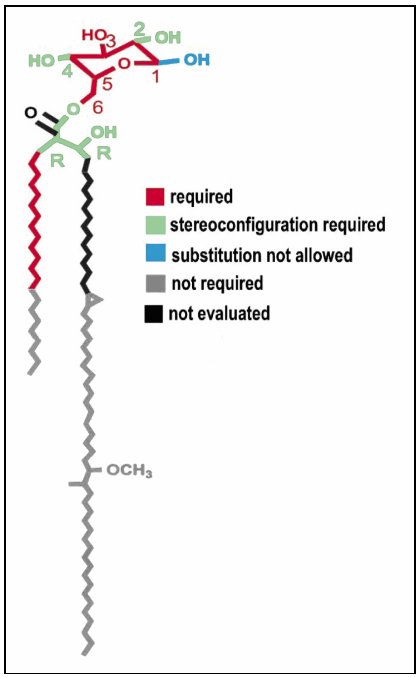

Figure 1.5 *A summary of the antigenic determinants of the TCR-response to natural GMM [54]. Analogs lacking the specific chemical element (red) and stereochemistry (green) were not recognised by the specific TCR but analogs lacking the distal elements of the mycolate alkyl chains (grey) were recognised. Substituting the anomeric carbon (blue) with glucose (to form TMM) or GMM (to form TDM) abolished recognition.*

The converse was also shown when a T-cell line that specifically recognised free MAs did not recognise GMM or TDM [55]. The recognition of free MAs indicated that an oxygenated function in the mero-chain is necessary for recognition of the CD1b-MA complex by the specific T-cell line. The free carboxyl group was also important in T-cell stimulation and the MA-methyl esters did not stimulate T-cells. From these findings the authors proposed that in oxygenated MAs, the two

hydrophilic groups form a combinatorial epitope by folding of the alkyl chain when bound to CD1b, as demonstrated schematically in Figure 1.6. This hydrophilic combinatorial epitope is exposed for recognition by the TCR while the long alkyl chains are hidden inside the CD1b molecule [55].

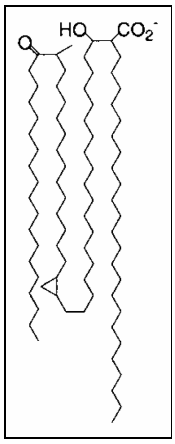

Figure 1.6 *A hypothetical model for the formation of a combinatorial epitope in keto-MA. The polar functional groups are brought together when the alkyl chains are bound by CD1b [54].*

The crystal structure of CD1b complexed with GMM showed that the binding groove is partitioned into deep hydrophobic channels where the long alkyl chains fit, leaving the polar moieties accessible at the surface for recognition by T-cell receptors (Figure 1.7) [56]. The GMM used for the determination of this crystal structure, isolated from *Nocardia farcinica*, had an 8-carbon alpha chain and a 58-carbon mero-chain, containing two cis-double bonds. Once the model had been obtained, the authors wanted to examine how CD1b was able to accommodate lipids containing many different functional groups at variable positions in the mero-chain. Using computer models the lipids were inserted into the binding groove and their functional groups were at various points in the channels, but there were no stereochemical clashes. This model is in agreement with the experimental data discussed previously, except that it could not provide any mechanism by which different functional groups in the mero-chain would influence TCR recognition as reported by Grant *et al.* [55].

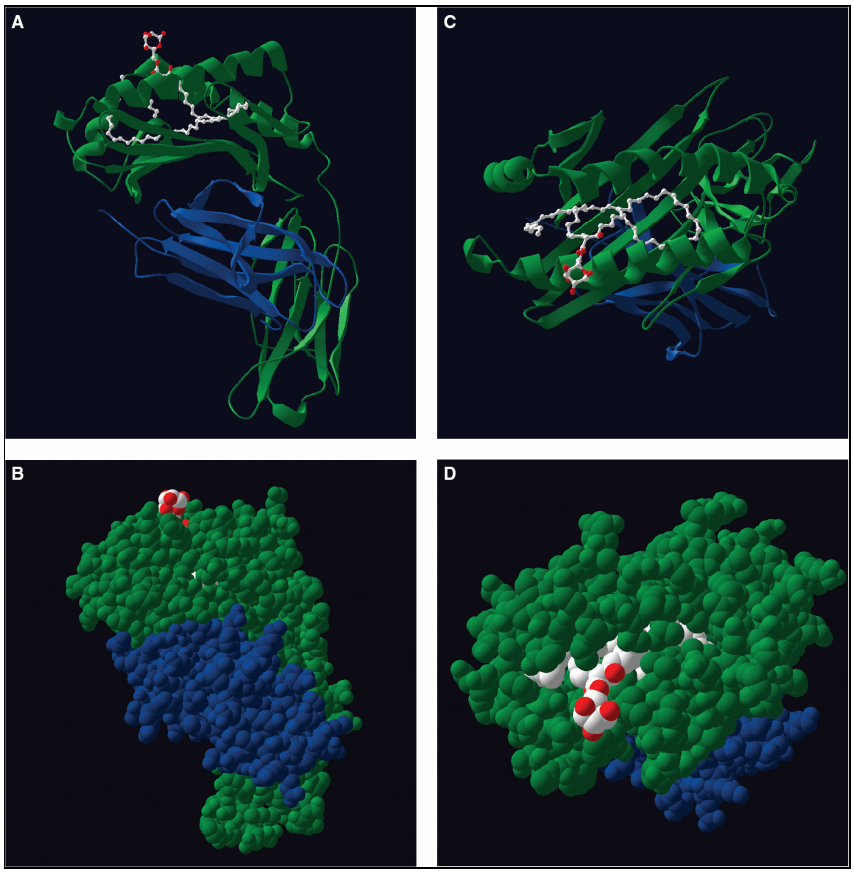

Figure 1.7 *The top and side views of the crystal structure of human CD1b loaded with the mycobacterial lipid antigen glucose monomycolate. The CD1 and β2m polypeptide chains are shown in green and blue, respectively. The lipid is shown with carbons drawn in white, oxygens in red and nitrogens in blue. A) Side view. CD1 and β2m chains are shown as ribbons. The lipid antigen is shown as a ball and stick model. B) Side view, shown as a space-filling model of the van der Waals surfaces. C) Top view of the ribbon model (i.e. looking down onto the α-helices from the presumed perspective of the T-cell receptor). D) Top view, shown as a space-filling model [56, 57].*

1.3.3.4 Recognition of mycolic acids by antibodies

For a long time it was believed that the immune response against *M. tb* was only a cell mediated response, since it was thought that intracellular pathogens mainly elicited cell-mediated immunity and they were not accessible to antibodies. Even more unexpected than finding that antibodies against *M. tb* protein antigens existed, was finding antibodies to glycolipids and lypoglycans [58]. It is likely that the presentation of lipid-antigens on CD1-molecules aid in the formation of antibodies against such antigens, but the mechanism is still unknown.

MAs are extremely hydrophobic molecules in comparison to hydrophilic antigens such as proteins or carbohydrates. Yet, it has been shown that anti-MA- and anti-cord factor antibodies come about in TB-positive (TBpos) patients [10, 59]. The epitopes of these antigens that are recognised by antibodies present useful hints in unravelling structural folds of the antigens.

Pan *et al.* [59] have shown that the antibodies produced in TB-infected patients not only recognised CF exceptionally well, but that they alsorecognised MA-subclasses. This was demonstrated by the higher reactivity of TBpos patient serum to *M. tb* CF and its lower reactivity to *M. avium* CF. The only difference between CF from these two species is in the fine structure of their MAs. *M. tb* possesses α-, MeO- and keto--MAs whereas *M. avium* possesses α-, keto- and diester-MA subclasses. The free MA methyl esters (ME) isolated from *M. tb* were also recognised by TBpos patient sera, and here it was seen that MeO-ME reacted most prominently out of the three sub-types. From the species-specific recognition of CF and the difference seen between subclasses of methyl esters, it seems likely that the distal functional group in the mero-chain of MA at least forms part of the epitope that was recognised by antibodies.

From our research group, Deysel [60] confirmed the recognition of CF, the natural heterogeneous pool of MAs and the separated α- and MeO-MAs by TBpos patient sera. However, no binding to MA-MEs occurred. Her work also showed for the first time which synthetic MAs are antigenic. This is significant because these compounds are chemically pure, which cannot be obtained from natural MAs in terms of their alkyl chain lengths and stereochemistries. Antibody recognition of the synthetic MAs showed that the mycolic motif as well as the mero-chain plays a role in the antigenicity of MAs. With regards to the mycolic motif, a protected α-MA methylated at its acid hydroxyl group and acetylated at its β-hydroxyl group was not antigenic. From the oxygenated MAs tested, namely keto-, MeO-, and hydroxy-MAs, MeO-MA was the most active. The diastereomer with an S,S stereochemistry at the distal methoxy-methyl group and an S,R stereochemistry at the proximal cis-cyclopropane, showed the highest activity and was proposed to be one of the antigenic

components of natural MAs. It will also serve as an excellent antigen in the serodiagnosis of TB.

Deysel [60] made another discovery regarding the antibody responses of TBpos and TB-negative (TBneg) patient sera to *M. tb* lipids. TBpos patient serum reacts with both MA and CF in ELISA, while TBneg patient serum also reacts with MA, but shows no reactivity with CF. The reactivity seen against MA in TBneg patients was explained by a cross-reactivity of antibody binding between MA and cholesterol. All humans have antibodies against cholesterol [61], and HIV patients have elevated anti-cholesterol antibody levels [62]. Therefore, if MA folded in such a way as to become recognised by anti-cholesterol antibodies in TBneg patients, then the cross-reactivity is explained and the knowledge is gained that MAs in CF fold differently to free MAs. MAs in CF have lost their sterol-fold, since there is no recognition of CF by TBneg patient serum. Therefore, the antibodies which bind to CF are different, and have a different epitope (probably containing trehalose) than those which bind to MA.

Other observations made in earlier experimental work by our research group have also lead to a structural or functional relationship between cholesterol and MAs. In summary, biosensor experiments showed that MA attracted cholesterol, but this attraction was abrogated when MA-MEs were used. Furthermore, Amphotericin B (AmB) which binds to sterols such as cholesterol also bound to MAs [63].

If we apply the knowledge from viral protein antigens [64] to MAs, we can state that the epitope that anti-MA antibodies bind to disappears, or is hidden when MA is bound to trehalose in CF. It is therefore a cryptotope of MA that is recognized by anti-MA antibodies and anti-cholesterol antibodies. In CF, a new epitope, or neotope forms which is only recognized by anti-CF antibodies and not by anti-MA- or anti-cholesterol antibodies anymore (as was shown by the lack of reactivity against CF by TBneg patient serum).

The sizes of epitopes on protein antigens that are recognised by antibodies vary greatly. They have been shown to have an approximate surface area ranging from 70-200 nm^2 [65-67]. In Langmuir-Blodgett studies the surface area of a MA is between 0.48-0.8 nm^2 in the upright position assumed on a hydrophilic surface, with variations

in the MA-type and surface pressure of the experiment [17, 68, 69]. Therefore, a MA is much smaller than a single protein epitope. Even with a trehalose added and the molecule's carbon chains maximally extended, it seems unlikely that an antibody will bind to one MA alone, regardless of which face of the molecule served as epitope. Most likely, a cluster of MA molecules are recognized by a single anti-MA antibody, such that the packing of MAs together will influence the antibody binding. It is reasonable to assume that the fine structure of the MA will determine how the MAs pack to constitute an antigenic determinant and that a combination of different MAs will make up a different epitope to that which can be composed from a homogeneous MA population.

The size difference between proteins and MAs poses the question of how appropriate the terms used for protein epitopes are for lipid epitopes, since several lipid molecules will be necessary to compare to the size of a single protein epitope. It may therefore be more correct to refer to the different forms, or shapes of the total structure of lipid antigens, rather than trying to single out epitopes. Perhaps the structure of MA recognized by anti-MA and anti-cholesterol antibodies could better be referred to as the ster*oid*, or more specifically cholester*oid* morphology (with *–oid* indicating a particular shape, form or resemblance) of a cluster of MAs constituting an antigenic determinant. In the body, MAs can certainly be expected to be presented in context of a phospholipid complex, such as in a membrane or in a lipoprotein, which further complicates the interpretation of ELISA results where MAs are recognised as precipitates on polystyrene.

1.4 Responses to free mycolic acids

The mycobacterial lipids GMM and free MA presented on CD1b stimulated Cd1b-restricted T-cells to secrete perforin (which lyse the APC) and granulysin (which has direct antimicrobial activity). It also stimulates them to secrete IFN-γ and TNF-α which activate the anti-microbial functions of macrophages (Figure 1.8) [70].

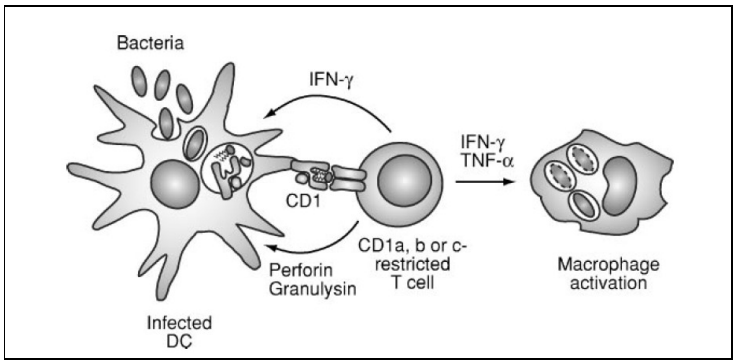

Figure 1.8 *CD1a-, CD1b-, and CD1c-restricted T cells in antimicrobial immunity. Dendritic cells that are infected with intracellular bacteria present foreign bacterial lipid antigens on the cell surface bound to CD1 molecules. CD1-restricted T cells that are specific for the foreign microbial lipids are stimulated to carry out effector functions, including the secretion of cytolytic granules containing perforin and granulysin, and the production of IFN-γ and TNF-α [70].*

Korf *et al*. [71] showed that MAs are able to express certain features of an *M. tb* infection. Administration of MAs in liposome carriers into the mouse lung or peritoneum induced a macrophage morphotype similar to the foamy macrophage derivatives observed in tuberculous granulomas. This induced foam cell morphotype showed an increased cell size accompanied by multiple vacuoles, intracellular accumulation of lipids (including cholesterol), and entry into mitosis. Unlike an *M. tb* infection, the MA liposomes did not interfere with phagosome maturation and acidification. These observations were seen in both peritoneal macrophages and alveolar macrophages – the natural targets of the mycobacteria in the host. Liposomes containing CF and whole heat-killed *M. tb* bacilli also induced the formation of foam cells, indicating that the MAs, which are common to all these systems, are the most likely candidates that contribute to this characteristic cellular response.

M. tb infection also recruits inflammatory cells, neutrophils, monocytes and lymphocytes, which accumulate around the site of infection in the lungs due to the production of inflammatory cytokines TNF-α, IL-12 and IFN-γ. MAs were able to induce a similar neutrophilic airway inflammation, with IL-12 production. MA also conditioned macrophages to produce IFN-γ on exposure to a secondary innate stimulus, such as LPS. The MA-induced IL-12 production and macrophage-derived IFN-γ promote the cellular and innate defenses. Even at high concentrations, MA was

unable to induce TNF-α production, whereas CF (which was less effective in inducing airway inflammation with neutrophillic inflammation and IL-6 and IL-12 production) was very effective in this unique property. MA-induced foamy macrophages were also able to produce MPO, an abundant heme enzyme that generates reactive oxidants, implicated in mycobacterial killing, when they were exposed to heat-killed *M. tb*. This confirms that MAs condition macrophages for enhanced cellular defenses [71].

Immune inflammatory disorders, such as allergic disorders, are driven by IL-4 and IL-13 cytokines, which have the opposite regulatory activities of IL-12 and IFN-γ cytokines which are induced by MAs. Realising the relevance of MAs to such disorders led to the testing of MA stimulatory functions in an experimental model of asthma in mice [72]. The initial response of MA inserted into the lungs of mice was to produce an influx of neutrophills. Only later, after the neutrophillic inflammatory response was cleared, did the MA tolerogenic function become apparent. It was shown that a single intratracheal insertion of MA in mice that were sensitized to the model allergen OVA, prevented an allergic airway inflammation from developing when the mice were exposed to both a primary and secondary set of nebulized OVA airway exposures. When MA-conditioned peritoneal macrophages were intratracheally instilled into allergen-sensitized mice and the protective activity was once again seen, it became clear that the protection against allergen-induced eosinophilia was not coupled to neutrophillic inflammation, and is a direct result of the MA-conditioned macrophages. The delayed appearance of the tolerogenic function, which coincides with the formation of foamy macrophages, also suggests that these cells are involved in generating the tolerant condition in airways. It was, however, shown that it is not the mere formation of foamy macrophages that promote the differentiation of dendritic cells into tolerogenic cells, since the MHC haplotypes of transferred macrophages had to match for the tolerogenic function to manifest. This indicates that it is the presentation of antigen by the macrophages that is directly responsible for the tolerogenic effect.

1.5 Folding of mycolic acids

In order to understand the role of MAs in the outer cell wall, their interaction with one another was explored by means of monolayer films. Hasegawa and co-workers have done successive studies on the monolayer properties of MAs in a Langmuir-Blodgett chamber. In these experiments an insoluble, single layer of MAs are formed on an aqueous surface, which can be compressed by applying lateral pressure to examine how the ordering and packing of the molecules change. They started by looking at monolayers of only α-MA (from *M. avium-M. intracellulare* complex) in 2000 [68], and they showed that the conformation of the α-MA molecule in its monolayer changed greatly when the surface pressure changed. The thickness changes of the monolayer were directly detected by atomic force microscopy (AFM), which verified the changes of the molecular cross-section measurements by the surface pressure-area isotherm. AFM measurements were done on Langmuir-Blodgett films, in which the monolayer is transferred onto a solid film.

A proposed schematic representation of α-MA conformations based on monolayer thickness profiles by AFM at different surface pressures is shown in Figure 1.9. In Figure 1.9a an image of the molecule with alkyl chains lying flat on the mica surface at low surface pressures is shown. Figure 1.9b shows a partly upright conformation in which the longer alkyl chain is folded at least once. Its larger portion is mostly in a disordered state where the carbons in the acyl chain cannot be in the lowest energy trans-conformation but instead the chains will have more gauche conformations with additional conformational stress in the molecule (see Appendix A on alkyl chain conformations). Nonetheless, the lower part of the longer chain is ordered and parallel to the shorter alkyl chain, since this arrangement provides a more stable form that owes to the hydrophobic interaction with the shorter alkyl chain.

The schematic pictures illustrated in Figure 1.9c and d are the possible molecular models in the liquid-expanded state monolayer. In these models, the second cyclopropyl group situated further away from the carboxyl group forms a turning point to make the longer alkyl chain folded. They differ slightly: in Figure 1.9c, the cyclopropyl group is located at the turning point, while the second carbon in the

cyclopropyl group is the folding point in Figure 1.9d. In both cases, the terminal end of the folded longer chain is in the vicinity of the terminal end of the shorter chain, so that most parts of the chains take a parallel arrangement with the aid of hydrophobic interactions.

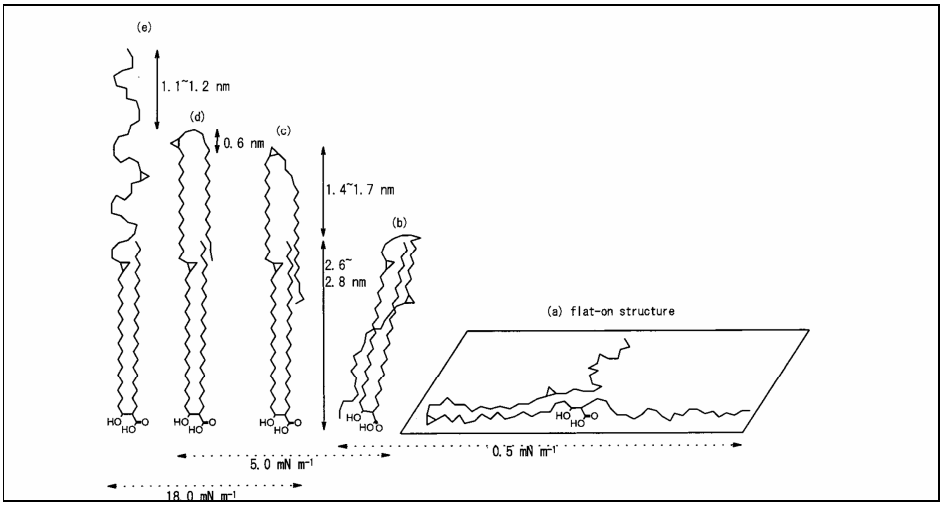

Figure 1.9 *Schematic drawings of the α-MA molecule in its monolayer at different surface pressures. (a) 0.5, (b) 5.0, and (c) 18.0 mNm^{-1} [68].*

Upon further compression of the monolayer, the turning point (kink) of the longer alkyl chain where the chain folding takes place moves further away toward the tail and then a new stable conformation is formed (Figure 1.9e). In the new conformation, the longer chain has no folding. Of particular note is that a half or a larger part of the extended longer chain is not supported by the parallel-juxtaposed shorter alkyl chain. Therefore, the portion of the longer alkyl chain above the top end of the shorter alkyl chain may be loosely packed in the double chain space. Consequently, the longer chain is likely to be rich in gauche conformations above the first cyclopropyl group. Figure 1.9e follows in which the total chain length of the molecule is longer than that in Figure 1.9d by 1.1-1.2 nm.

These suggested structures are supported by the surface area per molecule found at the different surface pressures. At low surface pressures, the molecular area was approximately 1.0 nm^2.molecule^{-1}, representing the unfolded state of the gas-phase. Upon compression, the surface area is reduced to 0.7 nm^2.molecule^{-1} and at the highest pressure where the solid phase was formed, the surface area was

0.48 nm^2molecule^{-1}. This is between the area which is expected for a double and triple chain (0.4 and 0.6 nm^2.molecule^{-1}, respectively), suggesting a transition between the triple and double-chain folds at this pressure. Therefore, the fully extended α-MA molecule is possible only when the molecule is laterally compressed by high pressure, which makes the monolayer take a solid state. The authors suggest that the position and stereochemistry of the cyclopropyl groups in the longer chain and also the hydrophobic interaction between the parallel sections of short and long chains are the crucial factors that determine the various possible conformations [68].

In a follow-up study, Hasegawa and colleagues revealed the difference in conformational states between α-MA and keto-MA from *M. tb* in their model [69]. The keto-MA molecule differs from α-MA structure only at the distal functional group in the meromycolate chain. With this group being a carbonyl group, it is expected that the keto group has a large dipole moment, and is definitely more hydrophilic than α-MA. Combined with infrared reflection-absorption (RA) spectrometry in this study it was shown that keto-MA molecules have a different response of conformational changes during the monolayer compression from that of α-MA in the Langmuir-Blodgett technique.

It was found that the long hydrocarbon chain in the keto-MA molecule remains folded, even under high surface pressure, and never unfolds as observed with α-MA [69]. The keto-MA formed an extremely stable monolayer and had a very large limiting molecular area of 0.77 nm^2.molecule^{-1} which the authors related to a triple chain fold in which the mero-chain is doubly folded, as in Figure 1.9b and c. However, in the conformation of a three-chain fold the hydrophilic carbonyl group which is likely to be drawn to the hydrophilic surface, is on the opposite end of the mycolic motif. It may be that this surface area represents a more favourable four-chain (or W) conformation where the carbonyl group is in close proximity to the mycolic motif at the hydrophilic surface, as suggested by Villeneuve [17, 18].

Analysis of the RA-spectra at low and high pressures of α-MA monolayers showed that the acyl-chains are more ordered (with a high trans-conformer percentage) at higher surface pressures when hydrocarbon chains are extended fully. In contrast, the folded keto-MA had large conformational stresses (a high percentage of gauche

conformers). This was explained by the observation that keto-MA never fully extended and stayed in a folded conformation. It also showed that the high surface pressure caused large molecular stress in the molecule, especially where it folded. These results demonstrate that a small difference in the structure between α-MA and keto-MA causes great differences in their conformational changes in response to changes in the external pressure.

In a subsequent study [73], the combination of surface-enhanced Fourier transform Raman spectroscopy (SERS) and Mass Spectrometry (MS), were used to determine at which points in the chains MAs are likely to fold or 'kink'. The good correlation between SERS and MS results indicate that the folding is present at the exact position of the cyclopropane and/or keto groups in α- and keto-MAs from *M. tb* and *M. kansasii*. Because certain vibrational modes specific for an all-trans conformation in acyl-chains were seen in their SERS experiments, it was concluded that the alkyl chains in the molecules have an ordered conformation.

In a study where the monolayer films of MA from five different species (*M. tb*, *M. tb Canetti*, *M. bovis* BCG Tokyo, *M. avium intracellulare* complex (MAC), and *M. kansasii*) were compared, the trends in limiting molecular areas held across the different strains in relation to the composition of their different subclasses of MA were realised [74]. It was found that MeO-MA had a limiting molecular area between that of α- and keto-MA. The authors stated that keto-MA had an area corresponding to a triple-chain folded conformation, and α-MA a double-chain. These differences can mainly be attributed to their particular functional groups in the mero-chains. Within the group of α-MAs tested, MAC was the only species with a shorter 23-carbon alpha-chain, where the others had a 25-carbon chain. MAC also had two different types of α-MA; one that had a cyclopropane group in its mero-chain (MAC1) as in α-MAs of other mycobacterial species, and another which had a double bond in a ratio of 4 : 1 for trans : cis conformations (MAC 2). MAC 1 gave the unexpected results of having a larger limiting area than the other α-MAs. With the only difference being the shorter alpha-chain, it was deduced that the extension of the mero-chain is governed by the intramolecular interaction through the alpha chain, and that the 23-carbon chain does not have enough molecular aggregation control to support the extension of the mero-chain at low pressures. MAC 2 also showed a

higher surface area, which is most likely due to the disorder that its double bond in the mero-chain causes.

These results allowed Hasegawa and Leblanc to formulate a useful model to describe the isotherm behaviours of MAs by considering two parts of the molecules [74]. They propose that at low surface pressures the mero-chain in all MAs is folded and that the unfolding process is influenced by two major factors when an external pressure is applied to the monolayer. Firstly, unfolding is dependent and dominated by the functional groups in the distal part of the mero chain. They say that the molecular unfolding of keto-MA would be inhibited by the strong molecular interactive force, represented as a molecular friction, by its polar carbonyl group. The other factor which becomes prominent in α-MAs without a polar group in the mero-chain is the length of the alpha-chain of MA. A slight difference in length (C23 to C25) largely influences the aggregation properties, and would also influence the stabilization of the unfolded mero-chain.

This theory explains why the keto group with its large dipole (high molecular friction) does not unfold. α-MA represents the opposite, being the most hydrophobic MA. Therefore, MeO-MA with an intermediate polarity may perhaps take a combination of double- and triple-chain conformations.

In all the studies by Hasegawa and co-workers mentioned above, experiments were done at 25 °C. In 2005, the group of Villeneuve *et al.* [17] examined the temperature effect on Langmuir monolayer packing of α-, MeO-, and keto-MA from *M. tb* in the 10-46 °C temperature range. In doing this, they were able to obtain the phase diagrams of the MAs shown in Figure 1.10.

The phase diagrams of the different MAs are very different from one another. The α-MA film has a wide liquid condensed phase. MeO-MA produces a liquid condensed phase, together with a multi-phase region over a wide range of surface pressures and temperatures. Keto-MA forms the most stable monolayer. It is the only MA in which a solid condensed phase is seen, and it did not show any liquid expanded or multi-phase regions. The exceptional rigidity of keto-MA in monolayers over a wide temperature range suggests that its keto group is also in contact with the hydrophilic

surface, and gives the molecule a four-chained W-conformation. In the biological environment of MAs, the keto group could be in interaction with the arabinogalactan layer or with the 3-hydroxy group of MAs (inter- or intramolecularly).

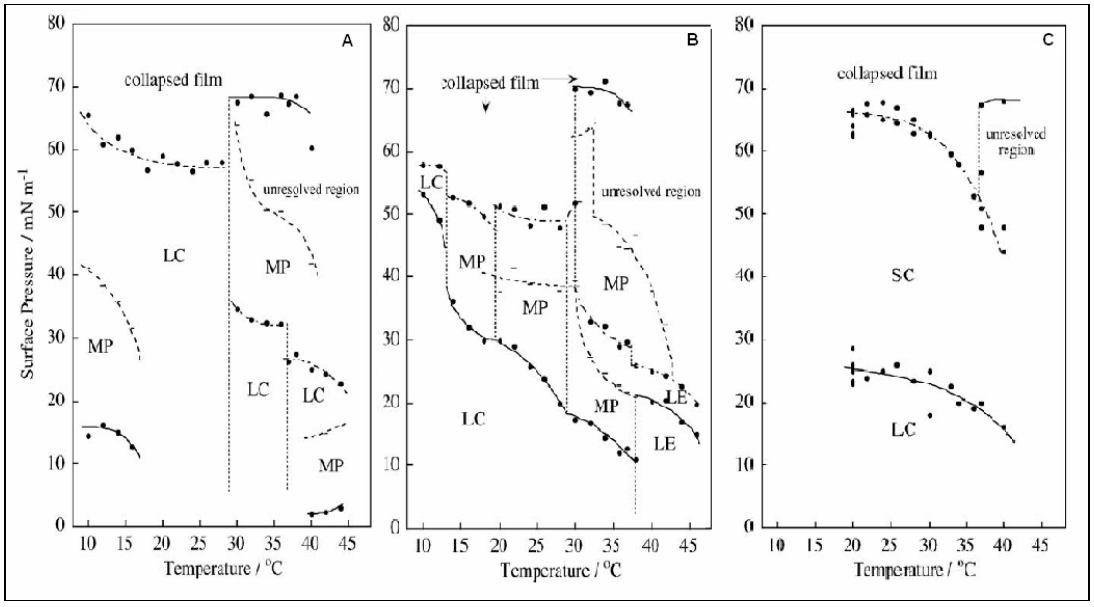

Figure 1.10 *Phase diagrams of Langmuir monolayers of α-, MeO- and keto-MA (C) The diagrams, (A) α-, (B) MeO- and (C) keto-MA, were prepared by plotting phase-transition pressures (points on the solid lines) and film collapse pressures (points on the dotted-dashed lines) of each Langmuir monolayer isotherm against temperature. The vertical dotted lines represent the temperatures at which phase transitions took place. LC: liquid condensed state, LE: liquid expanded state, SC: solid condensed state, MP: multi-phase [17].*

The group of Villeneuve *et al*. also did a combined Langmuir and molecular dynamics study [18] on MeO-, keto-, and deoxo-MA (derived by reducing keto-MA) from the non-pathogenic mycobacteria *M. bovis* BCG. This study was also done at a range of temperatures for each MA-subclass. Phase diagrams, similar to the previous ones were obtained. The deoxo-MA film was very unstable and collapsed easily. In agreement with the work of Hasegawa *et al*. [68, 69, 73, 74] all the MAs were in a folded conformation at low pressures and on compression the MeO- and deoxo-MAs took on an extended conformation. Keto-MA was proposed to stay in a four-chain folded W-conformation. The oxygenated MAs from *M. bovis* BCG have much less trans-cyclopropanes than those of *M. tb*. The ratio between cis- and trans-cyclopropane contents are 1 : 1.03 and 1 : 0.22 in MeO-MA from *M. bovis* BCG and *M. tb* respectively, and 1 : 0.33 and 1 : 2.7 for Keto-MA from *M. bovis* BCG and

M. tb, respectively. There does not seem to be any difference in the folding by cis- and trans-cyclopropanes, but the enthalpy values of MeO-MA for changing from a four-chain conformation into an extended one, was mostly smaller for *M. tb* than for *M. bovis* BCG MeO-MA. Therefore it does seem that the higher ratio of trans-isomers stabilised the four-chain conformation. Their results were confirmed by the molecular dynamics study, which will be discussed separately in the next section.

The significance of these monolayer studies to the knowledge and understanding of MA structure is considerable. The only disadvantage in these studies has been the fact that all of the MAs are isolated from the mycobacteria and are therefore heterogeneous and not chemically pure substances.

It is at this point where we recognize the value of the chemical synthesis of MAs. The stereo-controlled synthesis of such large and unsoluble molecules is not trivial, but the benefit we have of using chemically pure substances in experiments, is essential. Combined with this, theoretical chemistry, such as computer models, sheds light on the molecular aspects of the system, which can help to elucidate experimental observations.

1.6 Modeling of mycolic acid folding

Until very recently, no in-depth molecular modeling studies have been done on MAs. Some studies have been done on the structure of the cell wall of *M. tb* [75, 76] and its permeability [77], but in these studies MAs are represented only by a generalized structure, and no assumptions on their individual conformations can be made.

To date, the one and only molecular dynamics study that has been done on a detailed level of MA structure was published in April 2007, performed in conjunction with a Langmuir monolayer study of the oxygenated MAs of *M. bovis* BCG, mentioned above [18].

In this study it was seen that the α-, keto-, and deoxo-MAs showed differences in their likelihood of unfolding, or remaining in a compact, four-chain folded starting

conformation when subjected to 20 ps of molecular dynamics (MD). Keto-MA normally stayed in its original four-chain conformation, while MeO-MA often gave stretched-out structures. These results correlate well with the experimental observations. Even though deoxo-MA assumed stretched-out conformations, like MeO-MA in monolayer studies, it mostly stayed in the four-chain conformation during MD simulations.

The authors give a possible explanation for these results in terms of the bulkiness of the distal functional group in the mero-chain, and the lengths of the alkyl chains in the molecule that favour or oppose MAs to stay in a compactly folded four-chain conformation. The type of information that we can obtain at the molecular level to help explain the experimentally observed trends is what adds value to such theoretical studies of MAs. Being at the dawn of exploration in this new field, the challenge is to create virtual systems which can relate to experimental observations.

Chapter 2: Biosensor measurement of the interaction between mycolic acids and cholesterol

---

2.1 Introduction

In our group, several pieces of evidence for a structural or functional relationship between cholesterol and MAs have been discovered. As discussed in section 1.3.3.4, antigenic cross-reactivity between MAs and cholesterol was observed. It was shown that free MAs which may fold into a sterol-fold may be recognised by anti-cholesterol antibodies in TBneg patients and that MAs in CF fold differently and are not recognised [60, 78]. A methyl ester of a synthetic α-MA that is also acetylated at the β-hydroxy group was not recognised. It was also shown that MAs and cholesterol interact with one another [78]. Liposomes containing MAs accumulated cholesterol on an IASYS affinity biosensor. The synthetic, protected α-MA could not accumulate cholesterol.

More support for the structural or functional relationship between cholesterol and MAs was indicated in our group when it was shown that AmB (an antifungal macrolide which binds to sterols such as cholesterol and ergosterol) also binds to MAs in liposomes [78]. The interaction of MA and AmB was shown to be specific, because there was no binding to empty liposomes, and also no binding to the protected α-MA. When AmB was modified at its amino group, binding to MA was abrogated. This modification also led to cholesterol not being recognised by AmB anymore. This was taken to suggest that at least part of MAs fold in such a way as to mimic cholesterol's sterol fold.

The current work aims to study the relation between cholesterol and MAs further by attempting to quantify the interaction between cholesterol and MAs in liposomes on an evanescent field biosensor.

2.1.1 Cholesterol in tuberculosis

2.1.1.1 Cholesterol is essential for the uptake of mycobacteria

As discussed in chapter 1 (section 1.3.1), cholesterol plays an important role in the uptake of mycobacteria in host cells. Peyron *et al.* [31] showed that opsonised mycobacteria were phagocytosed unaffected by cholesterol-depleted cells, but the non-opsonic uptake was dependent on cholesterol, as was the case in the work of Gatfield and Pieters [30]. Gatfield and Pieters proposed that the decreased uptake of mycobacteria by cholesterol-depleted cells was due to a direct interaction with cholesterol, demonstrated by the high cholesterol-binding capacity of mycobacteria. On the other hand, Peyron *et al.* [31] believed that this decreased uptake was due to the inability of CR3 to associate with GPI-receptors that are known to occur in cholesterol-rich domains in the membrane. This was supported by their observations that cholesterol depletion, the removal of GPI-anchored proteins, as well as the disruption of the interaction between CR3 and GPI-anchored proteins decreased the uptake of mycobacteria.

In 1993, a protein was discovered that also played an important role in the uptake of mycobacteria in host cells [79]. Mammalian cell entry protein 1A (Mce1A) has conferred *E. coli* and polystyrene beads coated with the protein with the ability to invade non-phagocytic mammalian cells [79, 80]. The uptake of mycobacteria by non-phagocytic cells is suggested to be correlated with the pathogenicity of the organism since the uptake is mediated by the organism itself and promotes invasion of host cells. Similarly, the uptake of coated polystyrene beads into cells is correlated with the pathogenicity of the material that the bead was coated with. Mce2A, the homolog of Mce1A with the highest identity (67%) is not able to promote uptake of Mce2A-coated polystyrene beads into mammalian cells [80].

The reason for the mammalian cell invasion phenotype associated with Mce1A has been suggested to be associated with its structure [81]. The predicted molecular structure of Mce1A showed that it was similar in structure to Colicin N [81]. Colicin N is a bacterial toxin that acts by forming a pore in the lipid bilayer, allegedly binding

at cholesterol-rich lipid rafts, like other bacterial toxins [82]. Therefore, due to Mce1A having a similar structure to Colicin N, Mce1A may have the ability to bind to cholesterol or raft associated molecules and form pores for the transport of solutes [81]. Lipid rafts have been implicated as pathogen portals, providing an ideal clustering mechanism that could trigger cell-signaling events that are required in host-pathogen interactions [83].

This suggested mechanism is also in agreement with another study by Joshi *et al*. [84] where it was suggested that Mce proteins may play a role in transmembrane transport due to their similarity in sequence and predicted secondary structure with the transmembrane permease subunits of ABC (ATP-binding cassette) transporters. The ABC superfamily of ATP-powered membrane transporters is diverse, including several hundred different transport proteins found in organisms ranging from bacteria to humans [85]. Each ABC transporter is specific for a single substrate or group of related substrates that could be ions, sugars, amino acids, peptides, proteins, phospholipids, cholesterol or other molecules [85]. Mce3A and Mce4A homologs of Mce1A are relatively similar to each other and their partial molecular structures were indicated to have lipid bound conformations [81]. This conformation implies that Mce3A and Mce4A may be involved in the transport of lipids across membranes [84]. The direction of transport and the substrates are not known, but possibilities include the export of virulence factors such as CF, MA and LAM or the import of cholesterol and fatty acids as a source of nutrition for the Mycobacterium.

2.1.1.2 Cholesterol accumulates in macrophages: foam cells

At sites where bacilli have lodged in the lungs, activated macrophages collect around the bacteria and some macrophages fuse together to form large multinucleated giant cells. Lymphocytes and macrophages then collect around the giant cells to wall off the infected area from the surrounding tissue. This localised collection of inflammatory cells is called a granuloma, or tubercle [86]. Foam cells (macrophages that have accumulated neutral lipids of which a major component is cholesterol [71]) occupy the outer layer of the granuloma. Foam cells are thought to get their morphology from scavenging dead host cells and bacilli [87].

Korf *et al*. [71] showed that MAs administered intraperitoneally or into the airways of mice induced a foam cell morphotype in macrophages. This observation indicates an active involvement of MAs in inducing the foam cell morphotype. The cells were filled with lipid droplets of which a major component was cholesterol. Similar responses were seen when CF or killed *M. tb* were administered. By speculation of the functional significance of this mycobacterial-induced host response, the authors suggested that the active lipid import into macrophages may be due to metabolic needs of the mycobacteria in a nutrient-poor phagosome. This proposition is supported by the fact that 51 genes that were specifically expressed by a *Rhodococcus* species grown on cholesterol, and that were shown to function in the catabolism of cholesterol, were conserved in *M. tb* [88].

2.1.2 Cholesterol effects on a lipid bilayer

Lipid composition in a bilayer or membrane affects its rigidity. Lipids with a low T_m (melting temperature or gel-to-liquid transition temperature) are normally unsaturated phosphatidylcholines (PC). The double bonds in their chains prevent them from packing very tightly and rigidly. Saturated PC and sphingolipids are able to pack tightly and will only transform into the liquid phase at higher temperatures, therefore they have a high T_m [85].

Figure 2.1 is a theoretical diagram showing the phase changes of a lipid bilayer at different temperatures and different concentrations of cholesterol. Below 40 °C, low concentrations of cholesterol result in a solid ordered (*so*) state in a lipid bilayer, where the lipids are locked into lattice positions for long times [89]. At a physiological temperature the more cholesterol is added, the less ordered the system becomes. Moving from left to right below 40 °C in Figure 2.1 the phases start at *so* for low cholesterol concentrations to a combination of *so* and liquid ordered (*lo*) as the cholesterol concentration gets higher. It is significant that above ~38 mol% of cholesterol (corresponding to ~26% by mass), only the *lo* phase exists at any temperature. In this phase, the acyl chains of lipids are orientationally ordered (as

opposed to the liquid disordered phase) while the lipids can still diffuse in the plane of the membrane and rotate around the axis perpendicular to the membrane [89].

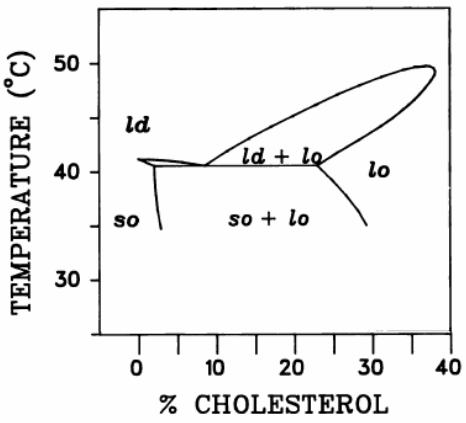

Figure 2.1 *Theoretical phase diagram for DPPC:cholesterol. Lines represent phase boundaries. Cholesterol concentrations are expressed as mol%. Phases are: ld, liquid disordered; so, solid ordered; lo, liquid ordered [89, 90].*

2.1.2.1 Phase separations in cell membranes

It was generally accepted that lipids in cell membranes exist in a *ld* phase because of the phospholipids in the membrane with low T_m. However, membranes are not composed entirely of phospholipids, but also of sphingolipids with high T_m and cholesterol that also affect the membrane phase. A number of studies have now shown that different phases coexist in lipid monolayers, bilayers, liposomes and eukaryotic cell membranes [91, 92].

In cell membranes lateral organisation is brought about by the clustering of specific lipids and proteins such as sphingolipids, cholesterol and glycosylphosphatidylinositol-linked proteins with an affinity for an ordered environment, forming condensed domains referred to as rafts [92, 93]. These rafts have been characterised experimentally as being resistant to solubilisation by non-ionic detergents [93]. Gaus *et al*. showed the existence of phase separations in macrophage membranes using Laurdan, a fluorescent probe in the membrane that undergoes a 50 nm red shift as membranes undergo a phase transition from the gel to fluid phase [92]. In both human and mouse macrophages *lo* and *ld* phases coexisted.

The filopodia and cell-to-cell contacts were abundant in the condensed, possibly raft-like domains in the *lo* phase. The authors also showed that these condensed domains corresponded with detergent insoluble rafts by examining the segregation of raft markers and comparing the fluorescence of Laurdan in the rafts to that of extracted rafts. The phases in macrophage membranes are also affected by cholesterol content and temperature, becoming more fluid as the cholesterol concentration decreases and temperature increases.

2.1.3 The ESPRIT biosensor

A biosensor in general, is an analytical device that brings an immobilised biological sensing material together with a transducer in order to produce a measurable signal that is proportional to the amount of analyte. The three main parts of a biosensor include [94, 95]:

1. A biorecognition element that recognises and binds the analyte of interest.
2. A transducer which converts the interactions between biomolecules into a quantifiable signal.
3. A readout system with which the interaction output can easily be displayed and stored.

The Autolab ESPRIT biosensor that was used here is an optical biosensor that measures biomolecular interactions in real-time and its transducer is based on surface plasmon resonance (SPR). SPR is a physical process that occurs when light hits a metal under a special angle position during total internal reflection conditions (Figure 2.2). In the SPR-setup there is a thin layer of gold on a glass disk which is on top of the hemi-cylinder. Between the hemi-cylinder and the disk is a thin layer of oil. The hemi-cylinder, oil, and glass disk have all got the same refractive index, so that the laser light will not bend on passing these layers before it reaches the gold layer. When there is a medium with a higher dielectric constant than gold on top of the gold (the sample) and the photons hit the gold layer, the free electrons in the gold will fluctuate. This electron fluctuation gives charge fluctuations in the metal. Since the metal layer is very thin, the charge fluctuations occur only at the surface where they

cause an electromagnetic surface wave, called surface plasma oscillations. SPR occurs when the incoming light is monochromatic and p-polarised (the electric vector component is parallel to the plane of incidence).

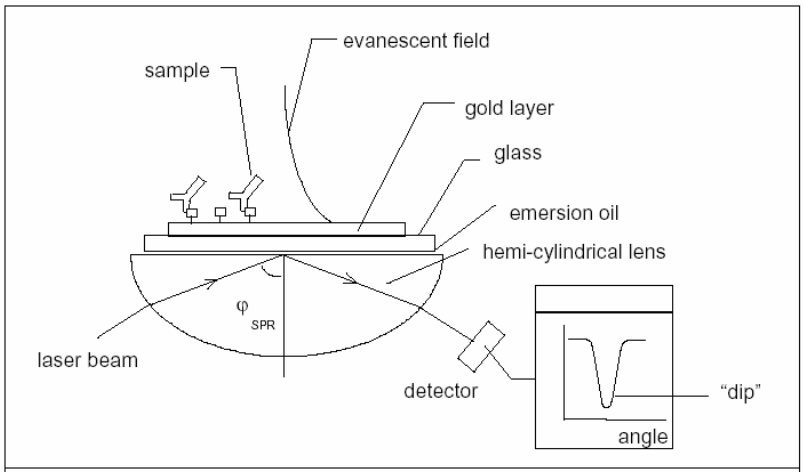

Figure 2.2 *The SPR setup. The hemi-cylinder, oil and glass disk have the same refractive index so that the light beam is not bent during its total internal reflection when it reaches the gold surface. The evanescent field decreases exponentially away from the gold surface [96].*

SPR is detected by measuring the intensity of reflected light. When SPR occurs, there will be almost no reflected light to detect by the detector, giving a sharp 'dip' in the light intensity when plotting reflectivity against angle of incidence at the specific SPR angle (Figure 2.2). The position of the SPR angle depends on the refractive index of the substance with a low refractive index at the sensing surface. The refractive index of the sensor surface changes upon binding of macromolecules to the surface, and therefore the SPR wave will change, and the angle will change accordingly. There is a linear relationship between the amount of bound material and the shift in SPR angle, and therefore the SPR shift in millidegrees is used as a response unit to quantify the binding of macromolecules to the surface. The penetration depth of the evanescent wave is 300-400 nm and therefore determines the size of macromolecules that can be studied. For particles larger than 400 nm, it will only be possible to study the binding qualitatively because the whole particle cannot be measured totally and therefore its signal is not linearly related to the amount of bound particles [95].

The advantages of evanescent field optical biosensors are that they can measure the interaction between biomolecules in real-time, without them being fluorescently

labelled and requiring only small amounts of sample. The ESPRIT has a modular setup (Figure 2.3) including an autosampler, for flexible design of experiments and rapid analysis of the interaction plots by the kinetic evaluation software.

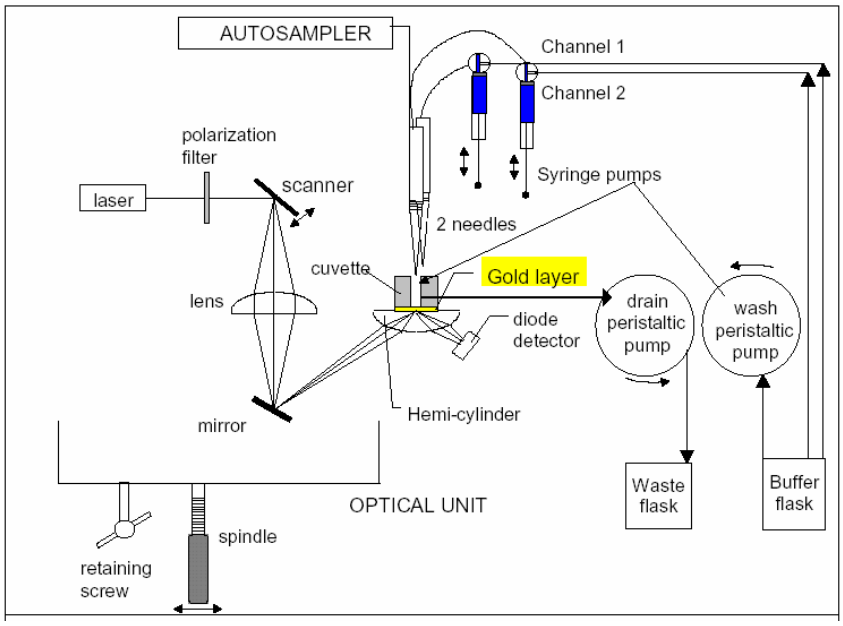

Figure 2.3 *A schematic representation of the ESPRIT configuration. Taken from [95].*

The dissociation equilibrium constant, K_D, which gives a measure of the affinity of the ligand for the ligate can be obtained from kinetic evaluation of interaction plots. The response of the biosensor is related to the mass of bound ligate and is given by the following equation:

$$R_t = (R_{eq} - R_0)(1 - e^{-k_{on}t}) + R_0$$

Where R_t = response at time t

R_0 = initial response

R_{eq} = the response at equilibrium / maximum

k_{on} = pseudo first-order rate constant

The responses are determined experimentally, which only leaves k_{on} to be determined for each concentration. When k_{on} for each concentration has been obtained and plotted against the concentration of ligate as in Figure 2.4, the dissociation constant

(k_{diss}) can be obtained from the y-axis, and the association constant, (k_{ass}) from the slope of the line. K_D can then be determined from k_{ass} and k_{diss} by $K_D = k_{diss}/k_{ass}$.

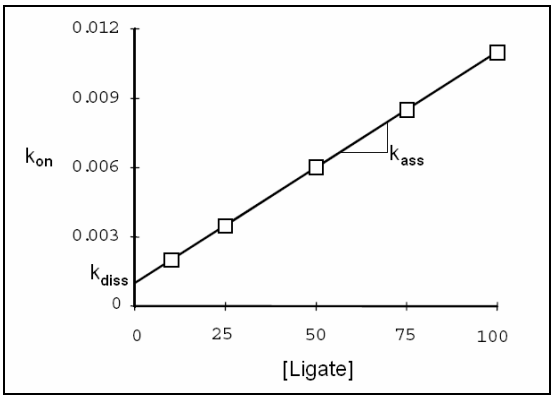

Figure 2.4 *Plot of k_{on} at different ligate concentrations from which k_{ass} and k_{diss} can be determined. k_{diss} can be read from the y-intercept and k_{ass} from the slope of the line.*

2.2 Hypothesis

If the interaction between cholesterol and MA can be observed on the ESPRIT biosensor, it will be possible to measure the interaction quantitatively with different concentrations of cholesterol liposomes in order to derive a dissociation equilibrium constant (K_D) for the MA-cholesterol interaction.

2.3 Aims:

1) To transfer technology from IASYS to ESPRIT (show qualitative interaction between MA and cholesterol on the ESPRIT biosensor as has previously been shown on the IASYS biosensor).
2) To do quantitative experiments to analyse data with the kinetic evaluation software in order to derive an equilibrium constant for the MA-cholesterol interaction.

2.4 Materials

2.4.1 Consumables

Acetic acid	Saarchem, Gauteng, SA
Amber vials	Separations Pty Ltd, Randburg, RSA
Chloroform	Merck, Darmstadt, Germany
Cholesterol	Sigma Chemical Co., St Louis, USA
Ethanol	BDH, Gauteng, RSA, Analytical grade
Gold disks	Metrohm, Gauten, SA
Isopropanol	Saarchem, Gauteng, SA
Membrane filters:	
0.22 µm, hydrophilic	Sartorius AG, Goettingen, Germany
Multi-well plates	Bibby Sterilin Ltd, Stone, UK
Mycolic acids	Mycobacterial mycolic acids were isolated from a culture of *M. tuberculosis* H37Rv (American Type Culture Collection 27294) as described by Goodrum *et al.* (2001).
Octadecanethiol (ODT)	Sigma, St. Louis, USA
Phosphatidyl choline	Sigma Chemical Co., St Louis, USA, (99% pure)
Potassium chloride (KCl)	Merck, Darmstadt, Germany
Potassium dihydrogen phosphate (KH_2PO_4)	Merck, Darmstadt, Germany
Sodium azide	Sigma Chemical Co., St Louis, USA
Sodium bicarbonate ($NaHCO_3$)	Saarchem, Gauten, SA
Sodium chloride (NaCl)	Saarchem, Gauteng, RSA
Sodium dodecyl sulfate (SDS)	Merck, Gauteng, SA
di-Sodium hydrogen phosphate (Na_2HPO_4)	Merck, Darmstadt, Germany
Sodium hydroxide (NaOH)	Saarchem, Gauteng, SA
Ethylene diamine tetra acetic acid	

(EDTA)	Merck, Darmstadt, Germany
Urea	Sigma-Aldrich, St. Louis, USA

2.4.2 Buffers and solutions

PBS-Azide EDTA buffer (PBS/AE): 8.0 g NaCl, 0.2 g KCl, 0.2 g KH_2PO_4, 1.05 g Na_2HPO_4, 0.38 g EDTA and 0.25 g sodium azide per 1 ℓ double distilled, deionized water, adjusted to pH 7.4. The buffer was degassed by allowing helium gas to flow through it before it was used in an experiment.

Isopropanol/NaOH solution: Isopropanol and 1 M or 50 mM aqueous NaOH mixed together in a 2:3 (v/v) ratio.

0.9% Saline: 0.9 g NaCl in 100 mℓ double distilled deionised water.

0.5% SDS: 5 g SDS in 1 ℓ double distilled deionised water.

6 M Urea: 108.11 g urea in 300 mℓ double distilled deionised water.

1% Acetic acid: 10 mℓ Acetic acid (18 mol/ℓ) to 990 mℓ double distilled deionised water.

0.2 M $NaHCO_3$: 16.802 g $NaHCO_3$ in 1 ℓ double distilled deionised water.

1 M NaOH: 40.00 g NaOH pellets in 1 ℓ double distilled deionised water.

2.4.3 Instrumentation

The measurement of the binding properties between MA and cholesterol was done on an Autolab ESPRIT instrument, supplied by Eco Chemie B.V. (Utrecht, The Netherlands).

For preparation of liposomes a Branson sonifier, model B30 (Branson Sonic Power Co., USA) was used.

2.5 Methods:

2.5.1 Preparation of liposomes containing mycolic acids or cholesterol

For the preparation of MA containing liposomes, phosphatidyl choline (PC) stock solution (90 µl, 100 mg/mℓ chloroform) was added to an amber glass vial containing MA (1 mg). The sample was dried under a stream of N_2 gas at 85 ºC. Saline (2 mℓ) was then added and the sample was heated on a heatblock for 20 min. at 85 ºC, with vortexing every 5 min. The sample was then sonified (30 cycles, output 4 or 5 for 2 min.), aliquoted into 0.2 mℓ per vial, lyophilized and stored at -70 ºC until use. For the preparation of cholesterol containing liposomes, a concentration range was made by varying the ratio of PC:Cholesterol, and keeping the total lipid content 9 mg. PC stock solution (amount as indicated in Table 2.1, 100 mg/mℓ chloroform) was added to an amber glass vial containing cholesterol (amount as indicated in Table 2.1, 50 mg/mℓ chloroform). The sample was dried, suspended in 2 mℓ saline, sonified and stored, as was done for the MA containing liposomes. The "empty" liposomes were prepared similarly, but containing PC (90 µl, 100 mg/mℓ chloroform) only. Before use, all liposomes were reconstituted with PBS/AE (2 mℓ). The liposomes were placed on a heatblock for 20 min. at 85 ºC, vortexed every 5 min., and sonified as above. New liposomes were prepared for each day's experiments.

Table 2.1 *Liposome compositions with various cholesterol concentrations.*

%Cholesterol liposomes	Cholesterol stock (µl)	PC stock (µl)
10%	18	81
17.78%	32	74
25.56%	46	67
33.33%	60	60
41.11%	74	53

2.5.2 Measuring cholesterol accumulation

The gold disk was prepared by incubating it in a solution of octadecanethiol (0.143 g) in absolute ethanol (50 mℓ) overnight. The needles were washed before experiments to avoid contamination with other students' work which involved serum samples. This was done in a sequential way with 0.5% sodium dodecylsulfate (SDS), 6 M urea, 1% acetic acid, 0.2 M sodium hydrogen carbonate ($NaHCO_3$) and double distilled, deionised water, with a short sequence which only washed the needles. Before each set of experiments, regeneration was done with an isopropanol/NaOH solution (2:3 v/v) and absolute ethanol. Between experiments regeneration with absolute ethanol was done to dehydrate the surface. The regeneration steps are followed by buffer washes so that the ethanol is removed gradually, and the surface remains degassed for the start of an experiment. The sequence that was designed (by T.S. Thanyani) to measure the MA-cholesterol interaction, was made to resemble previous work where this interaction was shown on an IASYS Affinity biosensor. All samples were added by the needles from a pre-programmed position on a multi-well plate. The samples were added to the multi-well plate manually by pipette in a slight excess to avoid air bubbles coming into the system when the needles aspirate the sample from the well. The sequence started by adjusting the baseline to zero and then MAs were added and immobilised for 20 minutes, after which the excess was removed by buffer washes. Cholesterol liposomes of a specific concentration were then added and allowed to associate for 10 minutes. Once again the excess was removed by buffer washes, followed by a 5 minute pause to observe dissociation. Regeneration was then done by isopropanol/NaOH (1 M or 50 mM, as stated) solution followed by absolute ethanol. The amount of cholesterol accumulated was measured by subtracting the binding signal (measured in millidegrees) after the buffer washes following MA immobilisation from the signal after 5 minutes of dissociation of the cholesterol liposomes.

2.5.3 Statistics

The Student's t-test was done using the Analysis ToolPak in Microsoft Excel. If the calculated t-value was less than the tabulated value at a 95 % confidence level, it was

assumed that there was no difference between the two samples. To determine whether the test assuming equal or unequal variances should be used, the larger variance was divided by the smaller variance, and the ratio was compared to the value of the F-distribution corresponding to the numerator (degrees of freedom between the groups, always two groups, therefore df = 2 – 1 = 1) and denominator (degrees of freedom for all the samples, ie. df = $n_1 + n_2 - 2$). If the calculated variance-ratio is less than the tabulated value, there was assumed to be no difference between the two variances and the t-test assuming equal variances was applied to the data. Outliers were determined with the Grubb's test, where G was calculated with the suspect value included in the mean and standard deviation, and compared to tabulated critical values for G. If the calculated G-value was larger than the tabulated value, the datum point could be rejected [97].

2.6 Results

2.6.1 Transfer of technology from IASYS to ESPRIT

The interaction between MA and cholesterol had only been demonstrated on the IASYS affinity biosensor in a qualitative way, and therefore it was necessary to first determine whether it is possible to also show cholesterol accumulation onto MA liposomes using the ESPRIT biosensor before attempting any quantitative measurements.

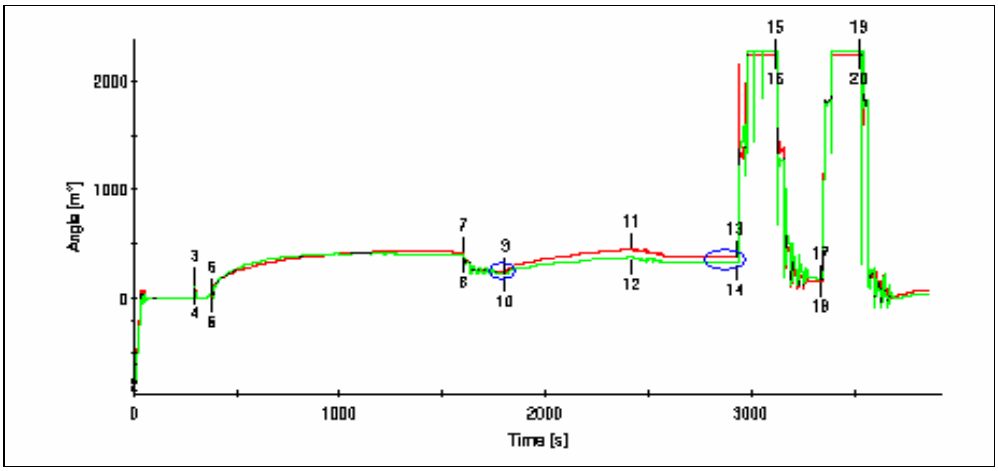

Figure 2.5 *A representative profile for the sequence of events of the experiment designed to show MA-cholesterol interaction on the ESPRIT biosensor. The red and green lines represent the response of the two cells of the cuvette. After adjusting the baseline to zero (3, 4), MA liposomes were immobilised for 20 minutes (5, 6), and the excess washed away by buffer washes (7, 8). Then cholesterol liposomes were added (9, 10), allowed to accumulate for 10 minutes and the excess washed away with buffer washes (11, 12). This was followed by a 5 minute pause in order to observe dissociation. Regeneration was done with isopropanol and NaOH (1 M or 50 mM, as stated) in a 2:3 (v/v) solution (13, 14) followed by buffer washes (15, 16). Then regeneration was done with absolute EtOH (17, 18) followed by buffer washes (19, 20). The amount of cholesterol accumulated was measured by subtracting the binding signal after removal of excess MA liposomes (line running into 9, 10, circled in blue) from the binding signal after 5 minutes of dissociation of the cholesterol liposomes (line running into 13, 14, circled in blue).*

A representative profile of such an experiment is shown in Figure 2.5. Duplicate experiments are run in the two cells of the cuvette and their responses are represented

by the red and green lines. After cholesterol liposomes have been allowed to bind to the immobilised MA liposomes, the amount of cholesterol that accumulated was measured by the difference in response (measured in millidegrees) after MA liposome immobilisation (circled in blue at 9, 10) and after cholesterol binding (circled in blue at 13, 14).

In order to compare the binding of the liposomes with those of a more experienced student, two different sets of liposomes were used in the first set of experiments.

Table 2.2 *Accumulation of cholesterol liposomes on MA liposomes. Transfer of technology from the IASYS biosensor to the ESPRIT biosensor.*

	Liposomes		MA immobilisation		Chol binding (angle m°)		PC binding (angle m°)	
Exp	PhD student	This study	Red	Green	Red	Green	Red	Green
1.1	x		254.86	239.76	127.47	81.57	NA	NA
1.2	x		277.36	284.92	156.38	133.84	NA	NA
1.3		x	232.26	221.95	146.06	100.89	NA	NA
2.1	x		319.94	310.17	NA	NA	54.3	53.73
2.2	x		290.18	301.17	NA	NA	63.36	40.83
2.3		x	218.7	221.26	NA	NA	21.83	24.82

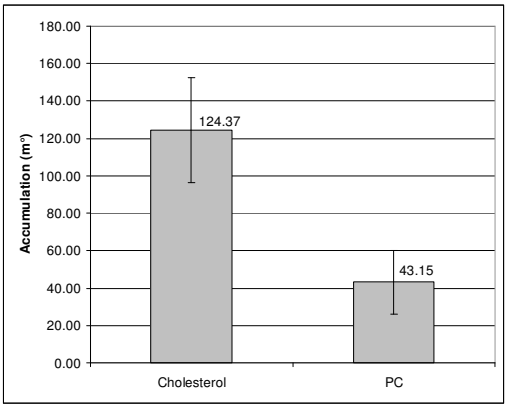

Figure 2.6 *The difference between accumulation from cholesterol containing liposomes (33 % cholesterol in PC) and empty (only PC) liposomes on the ESPRIT biosensor. Error bars show standard deviation (includes variation in liposomes prepared by different people, different PC batches and variation between the two cells of the biosensor). n = 6 for each set.*

MA liposomes were shown to accumulate cholesterol on the ESPRIT biosensor, as was shown previously on the IASYS biosensor (Table 2.2). The difference in accumulation between cholesterol liposomes and empty liposomes was significantly different at the 95% confidence level (Figure 2.6). The binding of my own liposomes

compared well to those prepared by the PhD student (Simon Thanyani, University of Pretoria, South Africa).

2.6.2 Optimisation for quantitative measuments

Knowing that it was possible to detect the accumulation of cholesterol on MA liposomes qualitatively, the next step was to investigate the quantitative measurement of this interaction. To do this, experiments had to be done at different concentrations of cholesterol to show that the response is dependent on the concentration (refer back to Figure 2.4). It is known that the lipids in membranes and bilayers can show phases in which the lipids can be in an ordered, or less ordered state. This is influenced by the lipid composition as well as the temperature and cholesterol content (section 2.1.2). It was uncertain what the effects of these different states would have on the quantitative measurement of the interaction between MA and cholesterol in liposomes.

The cholesterol liposomes that were used previously contained 33.33% cholesterol; that is a 1:2 ratio of cholesterol to PC by mass. A range of cholesterol concentrations below and above this was used. Cholesterol above ~40-60% (depending on the lipids in the system) is not miscible and starts to form cholesterol crystals. Therefore very high concentrations of cholesterol should be avoided. The lower limit should be as low as possible, but ensuring that the aliquots can still be made with good accuracy. Therefore, the range was made up from 10-41.11% cholesterol. This range covered both the ordered and less ordered phases of liposomes.

It was noted that there was a percentage of residual ethanol in the cuvette at the start of the experiment, which may have affected the whole experiment. Just prior to an experiment, regeneration with absolute ethanol was done in order to dehydrate the surface of the disk for the experiment to follow. This was followed by buffer washes so that there was 50 µl buffer in the cuvette at the start of an experiment. In order to keep the surface from being exposed to gas (air), half of the ethanol (50 µl) was removed by the needles and replaced by 50 µl buffer. The regeneration sequence consisted of five ethanol washes followed by five buffer washes. Therefore, the first

buffer wash diluted the ethanol to 50%, the second to 25%, the third to 12.5%, the fourth to 6.25%, and the last to 3.125% which is diluted to 1.56% at the start of the experiment by the addition of another 50 µl of buffer.

Since the interaction between MA and cholesterol is already compounded by the use of liposomes, it will be favourable to reduce the effect of another factor, namely ethanol. To do this, the sequence for regeneration done before each experiment, was changed to include eight buffer washes which decreased the ethanol concentration to 0.4% after regeneration, and to 0.2% at the start of the experiment. The lower percentage of ethanol led to lower responses on the biosensor (compare response at 33.33% cholesterol, Table 2.3 to the response in Table 2.2 and Figure 2.6). The first set of results with a range of cholesterol concentrations was obtained with this slight modification in the sequence and the results are shown in Table 2.3 and Figure 2.7.

From this set of results a region of high (33.33% and 41.11%) and low (10.0%, 17.78% and 25.56%) cholesterol accumulation could be seen (Figure 2.7). Results of statistical analysis in Table 2.4 confirm this. It shows that there were no significant differences between the accumulation of the three possible pairs of concentrations in the low cholesterol accumulating group. In the group of high cholesterol accumulation the accumulation at 33.33% and 41.11% also did not differ significantly from wach other. However, all six combinations of percentage pairs between the 'low' and 'high' cholesterol accumulation regions showed statistically significant differences in their responses.

Table 2.3 *A quantitative determination of cholesterol that accumulated on MA liposomes.*

%Chol	Response (millidegrees)						Average	Standard deviation
	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6		
10.00	25.61	42.69	15.55	16.21	12.29	NA	22.47	12.34
17.78	25.47	NA	15.69	13.19	16.97	17.50	17.76	4.62
25.56	32.00	43.16	15.46	14.77	11.82	10.18	21.23	13.29
33.33	90.92	NA	64.54	77.52	34.42	27.59	59.00	27.31
41.11	59.61	97.24	70.98	89.47	38.58	40.35	66.04	24.50

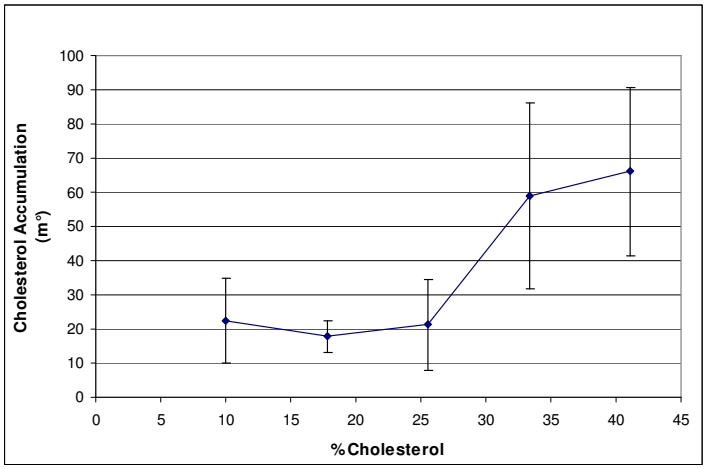

Figure 2.7 *The average cholesterol accumulation is shown for different concentrations of cholesterol liposomes accumulating on MA-liposomes. Liposomes were prepared at the start of the day and were sonified at 30 cycles and an output of 4%. Regeneration with the NaOH/isopropanol solution was made with 1 M NaOH. Error bars show standard deviation (includes variation in pipetting during aliquoting, liposomes prepared and experiments conducted on different days and variation between the two cells of the biosensor). For 25.56% and 41.11% n=6 and for 10.00%, 17.78% and 33.33% n=5.*

Table 2.4 *Statistical calculations done on the quantitative data obtained for cholesterol accumulation.*

	10%	17.78%	25.56%	33.33%
17.78%	x			
25.56%	x	x		
33.33%	√	√	√	
41.11%	√	√	√	x

The results for the student's t-tests are shown, with the two data sets that are being compared in each instance represented by the respective column and row in which the result appears. The result "x" represents no statistically significant difference and "√" represents a statistically significant difference between the two data sets at the 95% confidence level.

Sometimes it was observed that the order of the experiments had an effect on the amount of accumulation that was seen. For one particular set of these results, the apparent order-dependence of responses was seen to be due to a constant decrease in the amount of MA liposomes immobilized on the surface (Figure 2.8). Clearly, with a decrease in MA immobilization parallel to a decrease in cholesterol concentration (experiments were done from high to low cholesterol concentrations) in these

experiments, it will not be possible to discern whether a lower response was due to less cholesterol, or less MA present.

As shown in Figure 2.8, MA immobilisation was below 100 millidegrees at the sixth experiment. The seventh experiment was done after the disk was taken out and wiped with lens paper in order to remove any residue that may have been left on the surface from previous experiments. It was seen that the MA immobilisation increased from experiment six to seven, but it had not resumed to where it started in experiment 1. Therefore, MA immobilisation decreased due to some lipids that are not removed from the surface in a preceding experiment, but it may also be that the 1 M NaOH used during regeneration removed the ODT layer from the disk, enabling less liposomes to immobilize on the hydrophobic surface. The electrochemistry results of another member of our research group (Nsovo Mathebula, University of Pretoria) confirmed that this was likely.

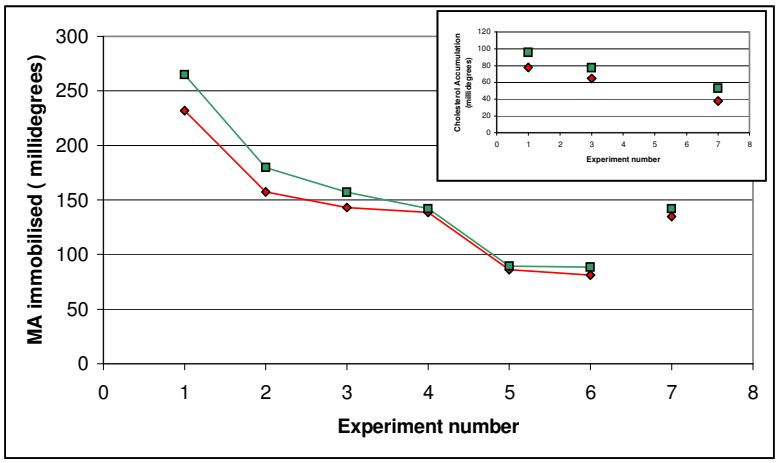

Figure 2.8 *The decline in MA liposome immobilisation in repeated cycles of measuring cholesterol binding to MA. For the seventh experiment, the disk was removed, wiped with lens-paper and replaced. The insert shows that the cholesterol binding response also decreased with successive repititions at the same cholesterol concentration (33.33%) for experiments 1, 3 and 7.*

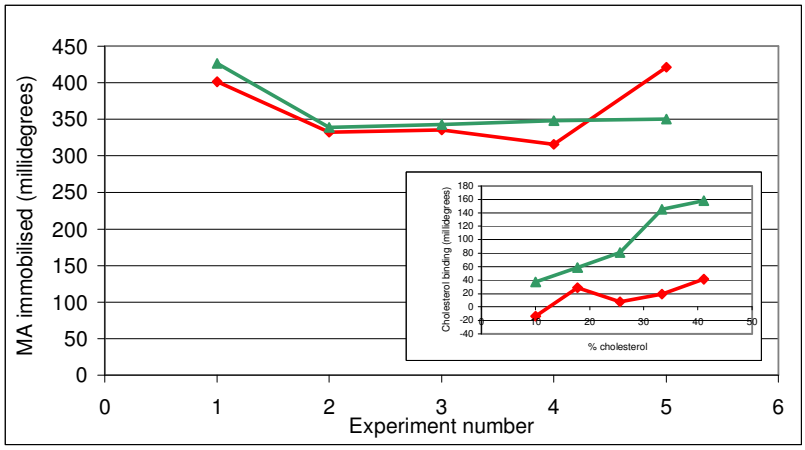

Figure 2.9 *Reliability of MA immobilisation on the ESPRIT biosensor with successive experiments. 50 mM NaOH was used for regeneration and the surface was wiped with lense paper after each cycle. The insert shows that the cholesterol binding response showed a general increasing trend with an increase in cholesterol concentration.*

From this point on, 50 mM aqueous NaOH was used for regeneration, together with isopropanol. The gold disk was also wiped with lens-paper before each experiment in order to remove any lipids that were left after regeneration. This proved to provide a more reliable level of MA immobilisation in successive experiments, as shown in Figure 2.9.

A new set of liposomes were prepared in order to repeat the range of cholesterol concentrations interacting with MA with the same, optimised method each time, in order to have data that can be analysed quantitatively. The disk was wiped before each experiment, except the first which follows directly after isopropanol/NaOH (50 mM) and ethanol regeneration. The liposomes were sonified at a higher output (5%) because this was the method used for doing experiments on the ESPRIT biosensor by other students, while the 4% output was the optimised method used to prepare liposomes for the cholesterol-MA interaction on the IASYS Affinity biosensor. Since it has been found that MA liposomes are more stable than cholesterol liposomes, the cholesterol liposomes were prepared just prior to an experiment. In this way they were used within an hour after they were prepared, whereas previously the cholesterol liposomes used in the last experiment may have been used up to six or more hours after its preparation. The order of experiments was done from the lowest to the highest cholesterol percentage.

When the next set of experiments was started with the modifications as stated above, it was noted that the MA immobilization was significantly slower. The very first set of experiments was repeated to find out whether the difference between accumulation of cholesterol liposomes and empty liposomes would still be evident. There was a notable difference in response between the graph in Figure 2.10 and Figure 2.6, where the method still included a higher ethanol concentration to initiate the liposome immobilisation step. However, the results in Figure 2.10 show that there is still a statistically significant difference between cholesterol accumulation and PC accumulation despite lower responses.

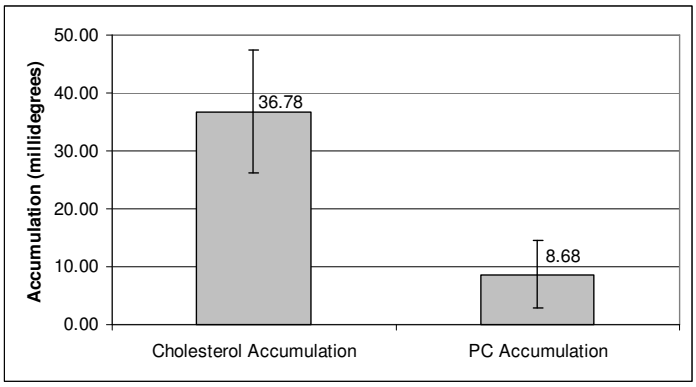

Figure 2.10 *The difference between cholesterol accumulation and PC accumulation on the ESPRIT biosensor for the adjusted protocol for liposome preparation and surface regeneration. For the cholesterol data n=6, and for the PC data n=5. Error bars show standard deviation (includes variation from the two cells of the biosensor).*

The next set of results with a range of cholesterol concentrations with optimised conditions is shown in Table 2.5 and Figure 2.11.

Table 2.5 *Quantitative binding data of cholesterol to immobilised MA on ESPRIT biosensor after optimisation of conditions.*

% Chol	Response (millidegrees)						Average	Standard deviation
	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6		
10.00	32.59	24.87	1.24	5.20	11.43	17.31	15.44	11.91
17.78	66.02	69.96	-2.62	-4.46	18.33	20.03	29.45	36.33
25.56	33.09	27.65	7.37	12.74	20.20	25.12	21.03	9.61
33.33	164.98	153.29	42.12	50.23	58.63	55.82	87.51	55.89
41.11	180.45	204.09	51.77	57.16	78.63	84.76	109.48	65.75

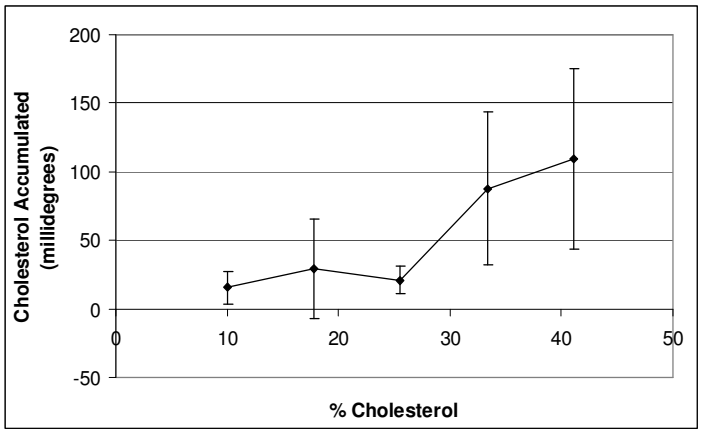

Figure 2.11 *The average cholesterol accumulation shown for different concentrations of cholesterol liposomes accumulating on MA-liposomes. The adjusted protocol for liposome preparation and surface regeneration was used. Error bars show standard deviation (includes variation in pipetting during aliquoting, liposomes prepared and experiments conducted on different days and variation between the two cells of the biosensor). n = 6 for each cholesterol concentration.*

Statistical analysis (Table 2.6) once again confirmed the high and low cholesterol accumulating regions that differ significantly from each other in response, as seen in Figure 2.11, just as in the previous set of results.

Table 2.6 *Statistical calculations done on the quantitative data obtained from the optimised experimental method for cholesterol accumulation.*

	10%	17.78%	25.56%	33.33%
17.78%	x			
25.56%	x	x		
33.33%	√	√	√	
41.11%	√	√	√	x

The results for the student's t-tests are shown, with the two data sets that are being compared in each instance represented by the respective column and row in which the result appears. The result "x" represents no statistically significant difference and "√" represents a statistically significant difference between the two data sets at the 95% confidence level.

The profiles obtained from quantitative experiments could not be fitted to monophasic or biphasic association and dissociation curves. The distribution of points around the x-axis of the residual plot was not random, indicating that the monophasic and biphasic models did not fit the data correctly. Therefore no kinetic rate constants could be calculated in order to obtain the dissociation equilibrium constant for the

interaction between MA and cholesterol. For this reason the data could not be analysed by the kinetic evaluation software.

2.7 Discussion

Initial attempts to immobilise MA liposomes on the ESPRIT biosensor surface in order to measure the attraction of cholesterol to MA demonstrated that a number of issues would first need to be resolved which were not relevant to the protocol previously used by other authors [98, 99] on the IASYS biosensor. One of these issues was the residual ethanol present after the regeneration protocol, which was unique to the ESPRIT biosensor. It was favourable to eliminate or reduce the amount of residual ethanol in order to have a better quantification of the interaction between MAs and cholesterol. Once the ethanol concentration at the start of an experiment was reduced from 1.56% to 0.2% the average response at a cholesterol concentration of 33% decreased from 124.37 to 59.00 and 87.51 millidegrees (Figure 2.6, Table 2.3 and Table 2.5, respectively).

Ethanol is a very small amphiphilic molecule. It has been suggested to interact strongly with PC headgroups in MD studies of bilayers with a preference for occupying the area close to the bilayer interface with the aqueous layer [100]. One such MD study [100] showed that the addition of ethanol to the system caused disruption of the regular parallel alignment of the PC headgroups in the bilayer due to strong interaction between the polar ethanol molecules and the lipid phosphate groups. Therefore, the lipid headgroups are likely to have an orientation toward the aqueous layer in the presence of ethanol. This behaviour was noted at both high (1:1) and low (1:8 molar ratio of ethanol:lipid) concentrations of ethanol. This is an important result which suggests that ethanol is likely to influence interfacial dynamics of a bilayer. A molecular explanation given for faster dynamics at the hydration layer of water molecules in the presence of ethanol was the easy exchange of phosphate-water hydrogen bonds by energetically equivalent phosphate-ethanol bonds. Another study by these authors [101] showed that the rigid hydration layer of a lipid bilayer is decreased in the presence of ethanol due to this easy exchange of hydrogen bonds. The dynamics of lipid molecules showed an increase in mobility both in the plane of

the bilayer and out of the plane of the bilayer in the presence of ethanol. This resulted from the increase in the volume of the simulation cell and the increased probability for the headgroups to be orientated towards the aqueous layer [100].

In an experimental design to test the effect that ethanol has on lipid bilayer mechanical properties and lipid bilayer surface area, a micropipette aspiration approach was used on unilamellar vesicles [102]. This is an accurate and direct video microscopy technique where vesicles were mechanically stressed by micropipette aspiration. The bilayers of vesicles were seen to laterally expand upon exposure to aqueous streams containing ethanol, so that the area per molecule was 9% larger in 20% (v/v) water/ethanol solution than in water. This is estimated to be accompanied by an 8% decrease in bilayer thickness. Significant changes in mechanical properties were observed when the vesicles were exposed to an ethanol-containing solution. The authors concluded that ethanol is able to alter mechanical and structural properties of a bilayer.

From these literature examples, it is therefore expected that liposomes exposed to ethanol in biosensor experiments would also show altered mechanical and structural properties. From the experimental and theoretical findings discussed, it can be inferred that the liposomes exposed to ethanol would increase in size and decrease in thickness. The parallel arrangement of the lipid head groups will also be disrupted and be orientated toward the aqueous layer, with the lipids having more mobility. The presence of ethanol may therefore lead to slightly amplified signals during an experiment on the ESPRIT biosensor with MA- and cholesterol liposomes by a) exposing more of the MA due to the decreased thickness of the lipid layer and the increased size of the liposomes and b) by increasing the dynamics at the aqueous interface to increase the overall contact between MA and cholesterol in order for accumulation to occur.

For the quantitative experiments, a range of cholesterol concentrations were chosen so that cholesterol would still be soluble at the highest concentration and the lower limit would stretch to the lowest concentration possible to prepare the liposomes with good accuracy. The range went from 10.0% to 41.11% cholesterol by mass. The 41.11% liposomes had a more visible white suspension in the PBS/AE buffer on preparation,

but after sonification the solution was clear, indicating that the cholesterol did not form crystals and was still soluble at this concentration.

The final quantitative data obtained by using the optimised method is shown in Table 2.5 and Figure 2.11. In Figure 2.11 a region of low cholesterol accumulation (10 to 25.56% cholesterol) and a region of high cholesterol accumulation (33.33 and 41.11% cholesterol) can be distinguished despite the large deviations in biosensor responses. Large deviations have also been observed by other students working on the ESPRIT biosensor, and work is continuing to further improve the transfer of technology from the IASYS to the ESPRIT biosensor, in order to obtain more reliable results with smaller variations.

When referring back to the theoretical phase diagrams for DPPC:cholesterol in Figure 2.1 [89, 90], the high and low regions of cholesterol accumulation can be correlated to the phases of the liposomes due to their cholesterol concentrations. The biosensor experiments were done at ~23 °C. When the bilayer phases are followed at this temperature on Figure 2.1, it can be seen that at very low cholesterol concentrations the bilayer is in the *so* phase that changes to a combination of *so* and *lo* phases, after which only the *lo* phase exists, as the cholesterol concentration is increased. The ~25 mol% at which the *lo* state starts corresponds to ~16% cholesterol by mass, and the point at which only the *lo* phase exists at all temperatures at ~38 mol% corresponds to ~26% cholesterol by mass. The phase boundary is not drawn in at the experimental temperature of interest, but by extrapolating from the higher temperatures it is possible to estimate that it would be between 16 an 26% cholesterol (by mass), and most likely closer to 26% at the end of the graph if the phase transition line continues down into the right-hand corner. This coincides with the separation between the low and high regions of cholesterol accumulation. The liposomes that showed low cholesterol accumulation had less than 26% cholesterol while the liposomes that showed higher cholesterol accumulation had more than 26% cholesterol and were therefore likely to be in the *lo* phase (see Figure 2.1). Therefore, the liposome lipids need to be in the *lo* state for cholesterol to interact with MA.

The concentration of cholesterol in a red blood cell membrane is ~16% by mass (26 mol%) [85], but it is important to keep in mind that lipids in eukaryotic cells tend to

separate into different phases through the clustering of lipids that have a high affinity for an ordered state, forming rafts. The rafts, rich in cholesterol and sphingolipids, have a *lo* phase that is in coexistence with the *ld* phase in the rest of the membrane. This is insightful, since it implies that cholesterol in membrane rafts would be able to interact with MAs under physiological conditions. As mentioned before (section 2.1.1.1), lipid rafts act as pathogen portals that provide an ideal clustering mechanism that could trigger cell-signaling events that are required in host-pathogen interactions [83].

Kinetic evaluation could not be performed on the data to calculate a dissociation equilibrium constant, K_D, for the interaction between MA and cholesterol. Although the response between MA and cholesterol was not linearly dependent on cholesterol concentration, it has given novel insight into the circumstances under which the interaction will occur.

The transfer of technology from the IASYS to the ESPRIT biosensor was successful, and a clear difference in the accumulation seen from adding empty liposomes or cholesterol liposomes to MA liposomes was established. This was the first biosensor study to attempt to quantitatively measure the interaction between MA and cholesterol. The lipid environment in these experiments did have a significant effect on the measurement of the interaction. Apart from the fact that the kinetic constants derived from such experiments would be influenced by the lipid environment, the phases of the lipids also changed during the course of increasing the cholesterol concentrations in the liposomes with significant consequences. Nevertheless, in the biosensor approach this interaction could only be measured on account of the lipid environment which suspended the cholesterol and MA molecules.

Langmuir monolayer and Langmuir Blodgett studies have an experimental design in which the lipids are observed as an insoluble monolayer. Thus, in these experiments there is an opportunity to work with pure MA and cholesterol with no added 'solvent' molecules to add complexity to the lipid layer. The effect of cholesterol on the packing of MA could be analysed in a quantitative way in these monolayers by increasing the ratio of cholesterol in the monolayer and observing changes in the area per molecules at different surface pressures and temperatures. From these studies it is

possible to obtain the different phases of the system as the temperature, cholesterol concentrations and surface pressure are varied. However, the attraction between the two molecules cannot be measured in monolayer experiments.

Although the lipidic system has added some unprecedented findings during the attempt to quantitatively measure the MA-cholesterol interaction, it has certainly also added value to the study. It has been determined that the interaction between MA and cholesterol occurred when the lipids surrounding cholesterol are in the *lo* phase and that it is therefore plausible for MAs to interact with host cholesterol in the membrane rafts of host cells. This may be of significance in mycobacterial cell entry.

Chapter 3: Computer modeling of mycolic acids

3.1 Introduction

To discover the biological potential contained in the fine-structure and assembly of MAs that constitute the *M. tb* cell wall, the best of investigative biology, organic and analytical chemistry and computational technology is required. From our research group, Stoltz, Korf and co-workers [71, 72] discovered novel biological effects of free MAs on the cells and physiology of experimental animals in a collaboration with University of Gent. Siko [98], Vrey [103] and Thanyani *et al.* [99] exploited wave guide biosensor technology, a most modern tool of analytical chemistry, to reveal the potential of MAs as antigens in serodiagnosis in collaboration with Erasmus Medical Centre, Rotterdam in The Netherlands. Deysel [60] explored the stereo-controlled chemical synthesis of MAs of various types in her collaboration with University of Bangor, UK and contributed one of several synthetic MeO-MAs that was tested for biological activity as an antigen against sera of TB patients. An important discovery by Benadie *et al.* [78] demonstrated the cholesteroid nature of MA. The implication of a cholesteroid nature of MA is so vast in terms of its potential to unravel the infectious strategy of *M. tb* and manage the MA antigen better in its application as an antigen in TB serodiagnosis (chapter 1), that it was followed up in this dissertation to attempt to provide a quantitative understanding of the phenomenon (chapter 2). This chapter continues to incorporate advanced cyber-technology to develop a tool to assist in the analysis of structure-function relationships of MAs in tuberculosis. MAs are known to contribute to the impermeablility of the mycobacterial cell wall by forming an ordered monolayer. The W-, U- and Z-folds that have been proposed for MAs in such monolayers in Langmuir monolayer experiments are therefore of special interest, and will be investigated further here (section 1.2).

3.1.1 Computational chemistry

Theoretical chemistry is the science where mathematical methods are combined with the fundamental laws of physics to study processes of chemical relevance. Computational chemistry is a branch of theoretical chemistry which is focused on obtaining results relevant to chemical problems by using theoretical chemistry [104]. Different models for calculating molecular structure and energetics have been developed. They fall into two main categories namely quantum mechanical (QM) models and molecular mechanics (MM) models [105].

3.1.1.1 Quantum mechanical models

The underlying mathematical model that is used in QM treats molecules as a collection of nuclei and electrons. The equations that are used ultimately stem from the Schrödinger equation, shown in its simplest form as:

$$H\Psi = E\Psi$$

where H represents the Hamiltonian operator, which operates on a mathematical function - the wave function Ψ, and E represents the energy of the system [105, 106]. When solutions of this equation are generated without reference to experimental data, the methods are called *ab initio* (meaning 'from the beginning' in Latin) in contrast to semi-empirical models [104].

In reality, the Schrödinger equation has only been solved for the hydrogen atom, a one-electron system, and for any system larger than this, approximations to the Schrödinger equation need to be made [105]. Different QM models differ in these approximations, and therefore different models have different capability, reliability, and cost in terms of computational time and power [105].

Because the nuclei move slowly in comparison to electrons, the first approximation to be made to the Schrödinger equation is to assume that the nuclei do not move. This is known as the Born-Oppenheimer approximation, and the result is the electronic

Schrödinger equation. Further, the Hartree-Fock approximation assumes that electrons move independently of each other and this leads to a set of coupled differential equations, the Hartree-Fock equations. The Linear Combination of Atomic Orbitals (LCAO) approximation transforms these differential equations into algebraic equations. These approximations lead to a set of equations called the Roothaan-Hall equations and the methods resulting from the Roothaan-Hall equations are the Hartree-Fock (HF) models. HF models are *ab initio* models that attempt to solve the Schrödinger equation without any empirical information.

Semi-empirical calculations follow directly from HF-models, with further simplification, such as only treating valence electrons and treating the inner electrons as part of the nucleus. Further approximations are introduced to simplify the calculation more and to incorporate empirical parameters. Parametrization involves adjusting the constants used in these methods so that the result of the HF calculation fits experimental data as closely as possible [105]. This can be seen as an advantage of semi-empirical calculations, since the HF results lack electron correlation, and therefore cannot reproduce the experimental results that do require them. In this way, electron correlation has been implicitly taken into account in the parametrization. Semi-empirical calculations will give the best performance for systems for which much experimental information is already available. On the contrary, they cannot predict unknown and new compound types [104].

3.1.1.2 Molecular mechanics models

MM models are typically used in molecular dynamics (MD) simulations. The fundamental difference between QM and MM models is that in MM calculations, the underlying mathematical model that is used describes molecules in terms of atoms and bonds, whereas the mathematical QM models make no explicit reference to bonding [104]. The core principle in MM models is that molecules are composed of units that are structurally similar and behave similarly in different molecules. This principle is implemented into the force field which treats atoms as certain types, depending on their atomic number and the type of bonding they are involved in. The force field describes the potential energy surface of entire classes of molecules with

reasonable accuracy. MM methods are therefore often referred to as force field methods. The force field energy (E_{FF}) is written as a sum of terms:

$$E_{FF} = E_{str} + E_{bend} + E_{tors} + E_{vdw} + E_{el} + E_{cross}$$

where E_{str} is the energy function for stretching a bond between two atoms, E_{bend} is the energy required for bending an angle, E_{tors} is the torsional energy for rotation around a bond, E_{vdw} (van der Waals) and E_{el} (electrostatic) describe the non-bonded interactions, and E_{cross} is the cross term which describes the coupling between the first three terms [104]. Each term describes the energy required for distorting a molecule from some idealized form (ideal bond lengths and angles etc.) and reflects the strain inherent to a real molecule, compared to an idealised form [105].

The main advantage of MM methods is the increased speed at which calculations can be performed, which as a result allows large systems to be modeled. Parametrization encourages the development of models which reproduce known experimental data accurately. As is the case with semi-empirical methods, MM methods also perform well for classes of molecules for which plenty of empirical data exists, but not for unusual molecules [104]. A variety of force fields exist. Class I force fields are those that were designed primarily for the use with large systems, such as proteins and DNA and were kept as simple as possible with no cross-terms. Class II force fields aim to reproduce small to medium-sized molecules to a high degree of accuracy and they will use more complex terms as well as a number of cross terms [104].

3.1.1.3 Conformational searching

Conformational searching methods fall into three main areas [105]:

- *Systematic* methods which involve systematic rotation around bonds
- *Monte Carlo* and *MD* techniques, which randomly sample conformational space, and

- *Genetic algorithms* which randomly mutate populations of conformers in search of survivors (the chemically and structurally legitimate conformers).

A *systematic search* on a molecule with N single bonds and a 'step size' of 360°/M, would need to examine M^N conformers. For a molecule with three single bonds and a step size of 120° (M = 3), this leads to 27 conformers, and for a molecule of eight single bonds it becomes over 6500 conformers that need to be considered. It is therefore only practical to use systematic methods for systems with a few degrees of freedom [105]. Being a very thorough method for conformational searching, it would be desirable to apply it to MAs, especially since there is limited information on the conformations of MAs on the theoretical level. In this chapter a way in which this approach may be made more practical for large systems such as MAs will briefly be introduced, as proof of principle.

MD offers sufficiently rapid calculations for molecules containing 100-200 or more atoms. This makes it possible to do conformational searching of larger molecules. It may be the only option to survey the full conformational energy surface for larger systems. Conformational analysis is definitely one of the most important applications of MD [105].

In this work the bioinformatics approach of using *algorithms* to mutate conformers in search of survivors has not been used, but bioinformatics approaches were applied in the statistical analyses as explained below.

3.1.2 Statistical analyses

To identify trends in large data sets, statistical methods are often employed. The field of bioinformatics has recently expanded to allow the analyses of large amounts of data resulting from microarray experiments. Microarray experiments characteristically produce very large data sets. Therefore, the programs (such as GeneSight) and analyses used for these data sets are capable of handling large amounts of data. Principal Component Analysis (PCA) and Self-Organised Maps (SOMs) are two such

tools that have been applied in analysing the large amounts of data. These methods are also suitable to identify trends in the folding of MAs since a large quantity of data needs to be analysed.

3.1.2.1 Principal Component Analysis (PCA):

Microarray data will often contain information from thousands of genes and/or samples. In mathematical terms, it is highly dimensional. We can only observe up to three dimensions, and therefore the dimensionality of the data needs to be reduced in order to visualise it and more readily interpret the patterns underlying the data. The purpose of PCA is to project a high-dimensional space onto a lower dimensional space. The reduction of dimensionality can be imagined as the projection that will be seen on a piece of paper when this piece of paper is placed behind a cloud of points in 3D-space. The three-dimensional groups of points have now been projected onto a two-dimensional space. The angle at which the high-dimensionality of space is viewed at can be chosen so that as much of the variability of the original data is captured in the lower dimensional projection of it [107].

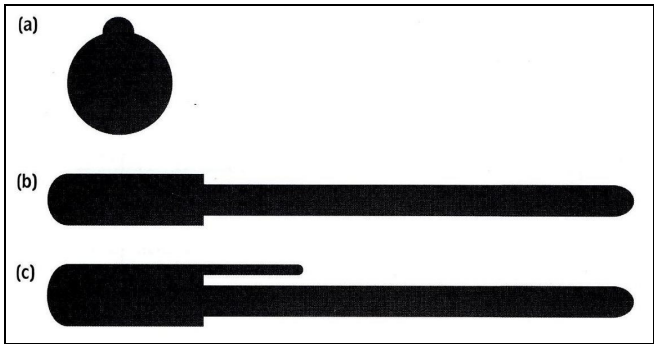

Figure 3.1 *Principal component analysis of a pen. (a) An end-on view, (b) a sideways view with the clip hidden and (c) a sideways view showing the clip [106].*

As an example of the principal component analysis of a pen in Figure 3.1, the different ways of constructing a two-dimensional view of a pen that exists in three dimensions is shown. In (a) the shape is not recognisable as a pen when an end-on view is taken, and therefore no principal components have been resolved. In (b) the pen is seen from the side, but with the clip still hidden. In this view, the first principal

component has been determined by finding the angle from which the most variability is seen in the shape of the pen; it is a long object, and it is wider at one end. Finally in (c) when the clip can be seen, the object is recognisable as a pen. Here, the second principal component has been resolved from an angle that is perpendicular to the first axis, and contains the most of the remaining variability. Clearly, some information is lost; for example, we cannot see whether the pen is a round or a square one [107].

In PCA of microarray data, firstly a variance-covariance matrix is constructed in which the variability of each gene is captured, as well as its co-variation with every other gene. This array is used to identify a new variable, a vector that is a linear combination of the genes and contains the maximum amount of variance. This is the choice of the projection line or the first principal component, and is called an eigenvector. For an n x n matrix, n eigenvectors with their corresponding eigenvalues exist. Next, the eigenvector that is orthogonal to the first, and that maximises the remaining variability, is found. This is the second principal component. This process can be repeated to find as many principal components as are of interest [107, 108].

In this work numerical data that has been collected from certain distances in a MA molecule was used. These distances were descriptive of the entire molecular conformation and could be relevantly applied in this work. Nevertheless, one drawback is that the distances are 2-dimensional and their directions (the third dimension) could not be taken into account. Molecular modeling software can perform PCA analysis on 3-dimensional structures using the coordinates of atoms in the molecule. This would be the ultimate progress in future analysis of this data and the 3-dimensional results can be compared to the present 2-dimensional results.

Because PCA involves the reduction of dimensionality, information will be lost (as explained above). Therefore it is important to check how well the variance of the data is presented on the 2D axis. This can be done by:

- Checking that the analysis makes biological sense, i.e. neighbouring time points are close to each other.

- Check how much of the total variance is accounted for in the projection by summing up the percentages of variance accounted for by each of the eigenvectors in a 2D plot [107].

3.1.2.2 Self-Organised Maps (SOMs)

SOM is one of the clustering techniques that implement an unsupervised competitive learning algorithm. In this type of machine-learning, the training is entirely data-driven and needs no further information [109]. It only requires the number of clusters to be chosen in advance by the user and then organises the clusters so that they relate to each other in a spatial topology on a square grid. The training is done by one sample vector being chosen randomly from the input data set and the most similar vector to it is then moved closer to the input vector together with its neighbours. This process is repeated for the whole dataset. One of the disadvantages of SOM is that slightly different results may be obtained every time it is repeated, since the genes are first allocated to clusters randomly [107].

Choosing the size of the grid or the number of clusters may require some prior testing. If there is a natural number of clusters, this size can be used. To check, the profiles should be similar within a cluster and different between clusters. As with PCA, the clustering should show biological relevance [107].

3.1.3 Modeling of mycolic acids

The first ever detailed molecular dynamics (MD) study of MAs by Villeneuve *et al.* [18] reinforced the results obtained experimentally in Langmuir monolayer studies, namely that keto-MA has a preference for staying in a W-fold, while MeO-MA unfolded most of the time. Additionally, it was possible to make some inferences from the MD as to why and how these findings are established in the experiments on a molecular level. This is a good example of the application and value that theoretical studies can lend us in terms of exploring and discovering the structural features of MAs.

In the MD study on MAs [18], their approach was to give the MAs a four-chain folded starting conformation. This was accomplished by arranging the molecule into the proposed fold and restraining the carbons in the long chains and certain torsions around the acid group to keep them in this conformation during MC calculations. It was then determined whether each MA had a preference for staying in this compact conformation, or for unfolding into a stretched-out conformation by running a 20 ps MD simulation with only the restrictions for the torsions and hydrogen-bonding around the acid group remaining (Figure 3.2). This was done with various isomers of keto- and MeO-MA from *Mycobacterium bovis* BCG, as well as with a deoxo-MA, derived from the reduction of keto-MAs.

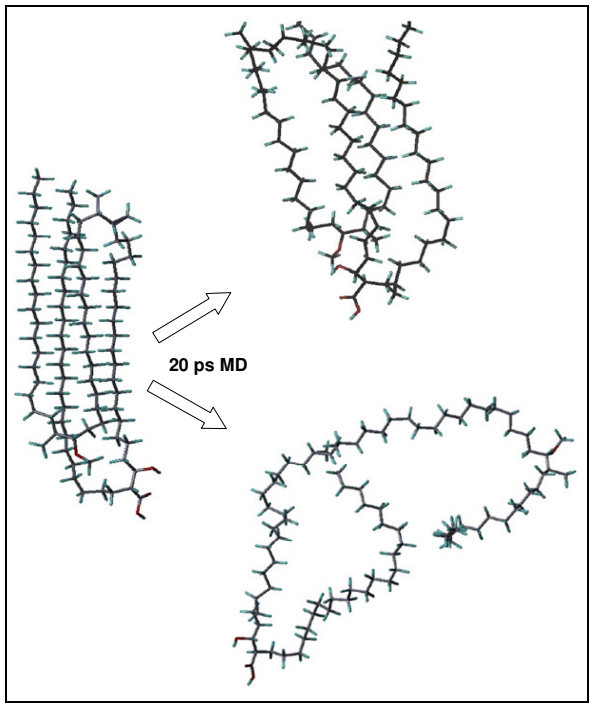

Figure 3.2 *Molecular dynamics of MeO-MA. A four-chain structure (left) maintains a four-chain fold (top right), or becomes stretched out (bottom right) after 20ps molecular dynamics [18].*

The potential energy surface was explored for a comparatively short period in MD of 20 ps. The incorporation of differences in torsions around the carboxyl group, multiple starting structures generated from MC calculations, and even structures taken from different stages of the minimisation afforded more variation in starting structures in order to ensure a reasonable sampling of the PES. The force field that was used in

MD, MMFF94, is a second generation force field developed by Merck, to be applicable to a broad range of molecules.

In the Villeneuve MD study on MAs [18] the experimental observation that keto-MA is very stable in the four-chain fold was supported. The molecular detail available in computer models of MAs has brought to light that the bulkiness of the distal functional group in the mero-chain and the hydrocarbon chain lengths may have significant effects on the ways in which MAs fold.

In the current study of MA structure, a different approach to the previous study has been taken by giving an unrestricted, non-specific starting structure to each MA subjected to MD. This was an extended, open structure, in contrast to the compactly folded four-chain starting structure in the study of Villeneuve *et al*. Then, during the 4 ns MD run, it was determined what folds the MAs take on. In this way, the possible conformations of an isolated molecule were explored and information obtained on the molecular flexibility and conformational fluctuations.

It was significant that the MAs did display two-, three- and four-chain folding, as has been described in the literature [13], without any structural predisposition or restrictions to their structures, and without any surrounding molecules or pressure in the system. To ensure that the potential energy surface (PES) was sampled well, a comparatively long (4 ns) MD run was performed and replicated 20 times for each MA-type. The same starting structure was used in all the replicates of a certain MA, with random velocities from the Boltzmann distribution for starting velocities to ensure that different areas of the PES were sampled. Calculations were done at room temperature (25 °C) with the Compass force field. Although this force field did not treat the cyclopropane carbons differently, the MD study was combined with QM calculations to both verify that the force field that most accurately represents the cyclopropane was used, and look in more detail at the effect of chain length on cyclopropane structure. Systematic calculations were done to create a PES for dimethyl cyclopropane as proof of the principle that this extremely powerful technique may be viable for MAs, despite their numerous degrees of freedom. The further implementation of this technique carries the potential to describe the complete PES of a MA molecule with a mathematical function, providing a means of accurately

describing the effect of every torsion angle in the molecule and therefore predicting conformations under certain conditions.

3.2 Hypothesis

The way in which MAs fold is dependent on their composition in terms of:

a) The type of oxygenated group in the distal position in the meromycolate chain (keto, MeO, or none)
b) The presence of a cyclopropane ring (cis or trans)
c) The state of the carboxylic acid group
d) The various chain lengths

3.3 Aims

1) Quantum Mechanics:
    - To explore the effect of methylene chain lengths on cis- and trans-cyclopropane ring structure.
    - To systematically calculate the different conformations of dimethyl cis-cyclopropane and to plot its PES.

2) Molecular Dynamics:

To run 4 ns of MD *in vacuo* with:

- α-MA
- MeO-MA
- MeO methyl ester (MeO-ME) and
- Keto-MA

doing 20 replicates of each MA type, in order to analyse the folding of the molecules by Principal Component Analysis, Cluster Analysis (Self Organising Maps) and to determine if W, U or Z-folds do occur and if so, any preferences that might exist.

3.4 Methods

3.4.1 Quantum mechanics

For QM calculations, Spartan '02 version 1.0.8 was used. An equilibrium geometry calculation at the ground state was run with semi-empirical PM3 level of theory on cis- and trans-cyclopropanes. Angles and bond lengths were measured as the methylene chain lengths were extended.

For the PES calculations of dimethyl cis-cyclopropane, shell-scripts were used to construct the input files for points at 30 degree intervals starting from D14,D15 = 0,0 to generate points on the PES. At each point, D14 and D15 were restrained to their associated values, while relaxing all other 3N-6 degrees of intramolecular freedom. The calculations were run in Gaussian [110] at the B3LYP/6-31G(d,p) level of theory, in vacuum. By systematically going through all the combinations of D14 and D15 in the 0 – 90 degree range (0, 30, 60 and 90 degrees), a 4 x 4 = 16 point surface was created. Since the molecule was symmetrical, every successive 90 degree rotation was degenerate to the first set of combinations. Therefore the first 16 point surface was replicated to create a 13 x 13 = 169 point surface from 0 - 360 degrees for D14 and D15. The energies were converted to relative energies in kJ/mol. The potential energy plot was created in Axum 5.0 by plotting the energy of the molecule as a function of D14 and D15.

3.4.2 Molecular dynamics

The MAs that were modeled were α-, MeO-, and keto-MA as well as the methyl-ester of MeO-MA. The MAs had the specific chain lengths and stereocentres as shown in Figure 3.3 below [17, 111-113]. These MAs have been shown to be representative of the components of the natural MAs [111-113]. Each of these MAs have been, or are in the process of being chemically synthesised by the group of Professor M.S. Baird at the University of Bangor, UK [111-113].

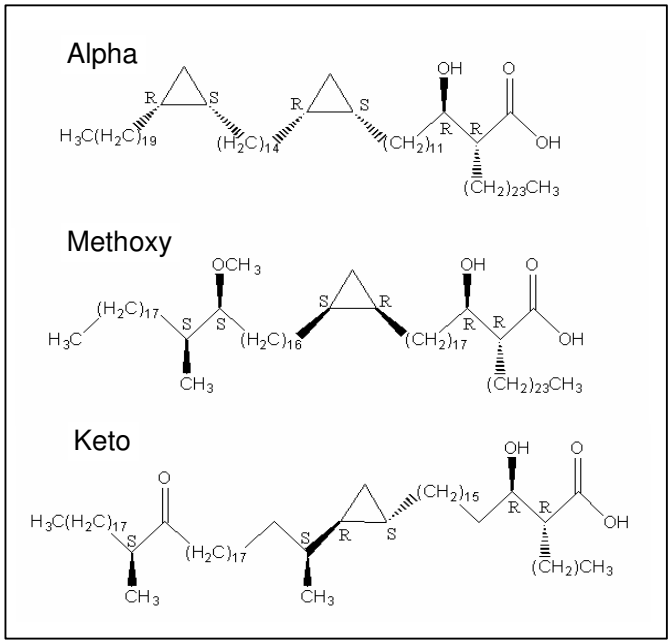

Figure 3.3 *Structures and specific stereochemistry of the MAs that were modeled. The fourth structure was the methyl-ester of MeO-MA [17, 111-113].*

A library of virtual fragments was first built, from which the MAs were constructed. The fragments contained the functional groups with their required stereochemistry and hydrocarbon chains of various lengths. Once a MA was built from the fragments, the torsion angles were changed to 180° where the different fragments were linked.

MAs were then minimised using the Discover module of the Accelrys Materials Studio 4.1 software package. The potential energy hyper-surface (PEHS) is a multi-dimensional surface which describes the energy as a function of all the degrees of freedom in a molecule. The minima (valleys) represent an arrangement of the atoms which result in a stable structure, where the first derivative of the energy (or the slope) is zero. The minimisation was done using the Smart Minimizer method at the ultra-fine convergence level with a maximum of 20,000 iterations. In this setup the calculation starts with a steepest descent method with a convergence criterion (slope) of 1000 kcal/mol/Å, and then moves on to a Conjugate Gradient method using the Fletcher-Reeves algorithm with a convergence criterion of 10.0 kcal/mol/Å. The last method usually used is the Newton method with a convergence criterion of 0.0001 kcal/mol/Å, but because the maximum number of atoms able to be treated

easily by this method was 200, it was never applied to the minimisation of the MAs. None of the four minimisations ever converged to a minimum after 20,000 iterations, but the highest maximum derivative was only 0.023 kcal/mol/Å (α-MA). Since the maximum derivative was close to zero and it was not expected that the open starting structures will be an energy minimum, the 20,000 steps of minimisation were considered sufficient.

The minimised structures were placed into the centre of a 100 x 100 x 100 Å vacuum cube. This is large enough to ensure that the MAs will not interact with another MA when periodic boundary conditions are applied. From this, 20 replicate MD runs were then set up for each MA using an NVT ensemble (number of molecules, volume and temperature stay constant) at room temperature (298K), with the initial velocities being random velocities generated from the Boltzmann distribution. The MD time was 4 ns with a 1.0 fs time-step, resulting in 4,000,000 steps calculated with the velocity verlet method. The frame output was saved for every 10,000th step, resulting in 400 frames per MD simulation. The Compass force field was used in all calculations. For non-bonded energies the van der Waals energies were calculated with an atom-based method using a cut-off of 12.50 Å and Coulomb energies were calculated using the Ewald method. By using a cutoff of ±7.9 for the first derivative of the Pressure-Potential Energy Fluctuation plot, it was determined that from the 150th frame (1.5 ns), every MD run had equilibrated satisfactorily. For the analyses, the Length Distribution module from the Structural Discover Analysis suite was used.

Five points (a-e) were used to analyse the fold of the molecule [18]. As shown in Figure 3.4, (a) was the last carbon in the 2-alkyl chain, (b) was the carbon on which the carboxyl group was, (c) was the distal carbon of the cyclopropane ring, (d) was the carbon on which the keto- or methoxy group were and in the case of α-MA it was the distal carbon of the cyclopropane ring, and (e) was the end carbon of the meromycolate chain. Eight distances (ab, bc, cd, de, ac, ae, ce, bd) were used to describe the MA fold [18]. The first four are the chain lengths of the four hydrocarbon chains in the MA molecule separated by functional groups, and the last four are distances between functional groups and/or the ends of the molecule. In order to retrieve these distances from the MD runs, Perl scripts (coded by Dr. Alexandra Simperler, Accelrys) were used to find the atom numbers, and then to

output the distance between two of the atoms for every frame. These distances were used in three types of analyses.

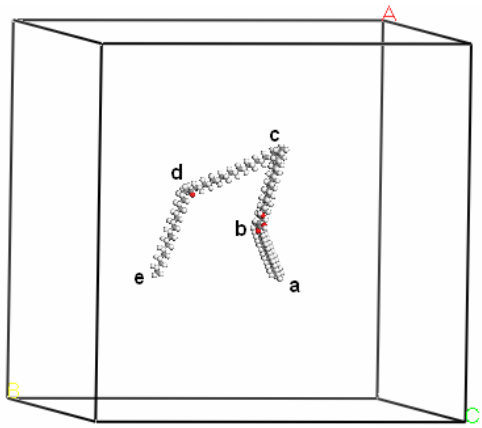

Figure 3.4 *The five points used in the analysis of MA-structures shown on a keto-MA. a = last carbon in the 2-alkyl chain, b = carbon on which the carboxyl group occurs, c = distal carbon of the cyclopropane ring, d = carbon on which the keto- or methoxy group (for keto- and MeO-MA, respectively) occurs, or the distal carbon of the cyclopropane group (α-MA), and e = end carbon of the meromycolate chain [18].*

MA key distances were used to define the seven possible W, U and Z- folds that are shown schematically in Figure 3.5. The parameters by which these folds were defined originated from the W-fold description, as defined by Villeneuve *et al.* [18], whereby each of the four chains are stretched to 70% or more of their fully elongated length, and have a b-d distance of 18 Å. The other parameters used by Villeneuve *et al.* [18] could not be used, since they were derived from folded starting structures. The parameters were then customised so that a W-shape would have a closer b-d distance of 10 Å so that the chains are close enough to allow inter-chain interaction. The remainder of the parameters for the W, U, and Z-folds shown schematically in Figure 3.5 were completed as shown in Table 3.1 below. To include conformations that still represented the W, U, or Z-folds distinctly, but did not have all of their chain lengths elongated enough to fall into the first class, the two end chains (ab and de), were allowed to be 50% or more of their original stretched-out length. For each MD run, all 400 frames were labelled as one of these seven folds if they met the criteria for that fold. This was done with a Python script (coded by Jurgens de Bruin, University of Pretoria).

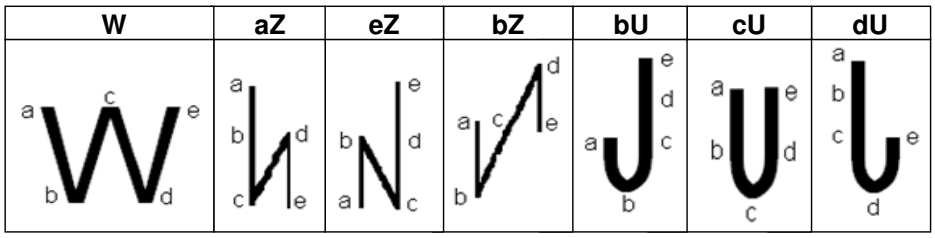

Figure 3.5 *A schematic representation of the seven possible W, U and Z folds.*

Table 3.1 *The parameters for the eight selected distances in a MA molecule for the different types of W, U and Z folds.*

Name	Percentage of the maximally extended chain length				Distance between points (Å)			
	ab	*bc*	*cd*	*de*	*ac*	*ae*	*ce*	*bd*
W	> 50	> 70	> 70	> 50	< 20	< 25	< 20	< 10
aZ	> 50	> 70	> 70	> 50	> 20	> 25	< 20	< 10
eZ	> 50	> 70	> 70	> 50	< 20	> 25	> 20	< 10
bZ	> 50	> 70	> 70	> 50	< 20	< 25	< 20	> 18
bU	> 50	> 70	> 70	> 50	< 20	> 25	> 20	> 18
cU	> 50	> 70	> 70	> 50	> 25	< 25	> 25	< 10
dU	> 50	> 70	> 70	> 50	> 20	> 25	< 18	> 18

The distances were also subjected to Principal Component Analysis (PCA) that was done on:

i. All 8000 available frames for each MA type (400 frames from each run of which there are 20 replicates)
ii. 5020 frames for each MA type (the last 251 frames of each MD run representing the equilibrated part of the run).

Finally, the distances were clustered into groups using Self Organising Maps (SOMs). The 400 frames of each run were grouped into a maximum of 16 groups, and groups (i) and (ii) described above for each MA type were also grouped into a maximum of 49 and 25 groups respectively. Both PCA and SOM were done using GeneSight4.1.

3.5 Results

3.5.1 Quantum mechanics

In order to study the effect of the chain lengths on the cyclopropane ring, semi-empirical calculations were done with lengthening of the hydrocarbon chains, with the bond lengths and angles shown in Figure 3.6 measured in each case.

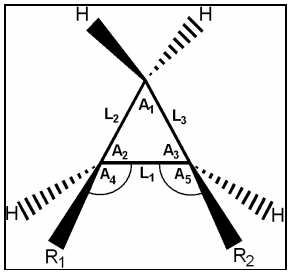

Figure 3.6 *The bond lengths and angles of the cyclopropane molecules that were measured.*

Table 3.2 *The stabilised bond lengths and angles measured as the chain length was increased in a symmetrical way for cis- and trans-cyclopropane.*

	Internal Angle (degrees)			Bond length (Angstroms)			Outer angle (degrees)		Energy (kcal/mol)
	A_1	A_2	A_3	L_1	L_2	L_3	A_4	A_5	
cis	60.5	59.8	59.8	1.513	1.502	1.502	121.0	121.0	-85.357*
trans	60.4	59.8	59.8	1.510	1.502	1.502	119.5	119.5	-85.604*

*the energies refer to the cyclopropane with a total of 18 carbons in its chains because from 20 carbons trans-cyclopropane had a negative frequency (refer to Appendix B).

Cis-cyclopropane was unaffected by the chain length from a symmetrical short 2-carbon chain until a long 16-carbon chain on each side. The bond lengths and angles stabilised to those shown in Table 3.2 for both cis- and trans-cyclopropanes with a 9-carbon methylene chain on either side. From a chain of 10-carbons and longer on each side, symmetrical trans-cyclopropane always had one negative frequency. Equilibrium structures are characterised by having only positive (real) frequencies, while transition states are characterised by having one negative (imaginary) frequency [105]. However, the imaginary frequency had a very small magnitude (-24 cm^{-1}),

which usually corresponds to torsional or related motions, and therefore this structure may represent a very low energy transition structure [105].

These results show that the lowest energy conformation is not always directly accessible in these semi-empirical calculations. An equilibrium conformation calculation which randomly changes the conformation in order to find the lowest energy conformer was performed on trans-cyclopropane with two 4-carbon chains (Figure 3.7B). The results of the calculation afforded a conformation in which the hydrocarbon chains move towards each other to a position where they could possibly interact (Figure 3.7B). The animation of the negative frequency in the symmetrical trans-cyclopropanes with longer chains is a rotation of the chains, as indicated by the arrows in Figure 3.7A (as was expected from a low frequency). The small rotation of the chains towards each other suggests that it may be likely for the chains to interact in a more stable conformation, with no imaginary frequencies.

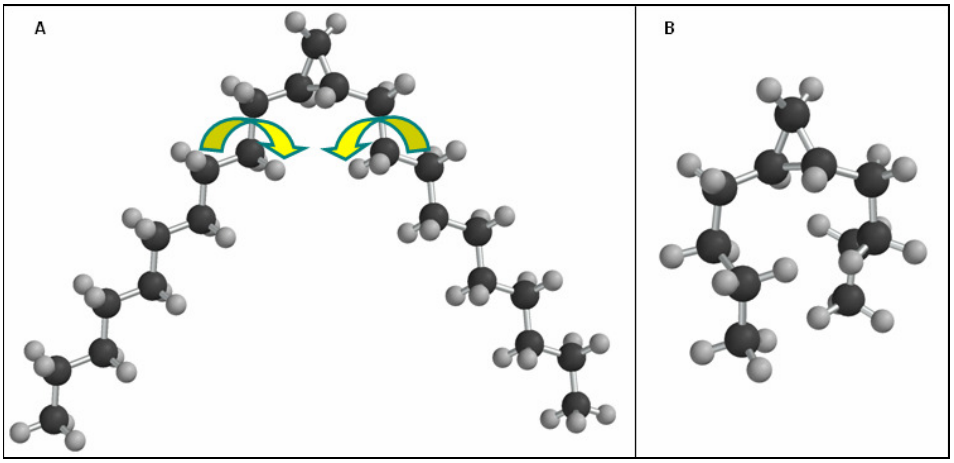

Figure 3.7 *A schematic representation of the rotation seen at an imaginary frequency for disubstituted trans-cyclopropane. This was seen at a frequency of -24 cm^{-1} in (1R, 2R)-1,2-didecylcyclopropane (A). The equilibrium conformer result of (1R, 2R)-1,2-dibutylcyclopropane, showing possible inter-chain interaction (B).*

In order to find the low-energy conformations, a systematic approach is necessary. As mentioned before, it is not practical to do systematic conformational searches on large molecules such as MAs with numerous degrees of freedom. However, the use of an automated numerical approach may make this task more efficient. As a proof of this principle, a PES of dimethyl cis-cyclopropane was systematically calculated by

rotating the methyl groups by 30°. The atoms were numbered in a systematic way [114], as shown in Figure 3.8, in order to allow automation in future work when the hydrocarbon chains will be extended. The PES was calculated as a function of two dihedral angles, D14 and D15. When considering four atoms connected in the order A-X-Y-B, the torsion angle or dihedral angle is the angle between the A-X and B-Y bonds, viewed along the X-Y bond. D14 was defined by atoms 14, 7, 3 and 2 and D15 by 15, 9, 4 and 3 (Figure 3.8).

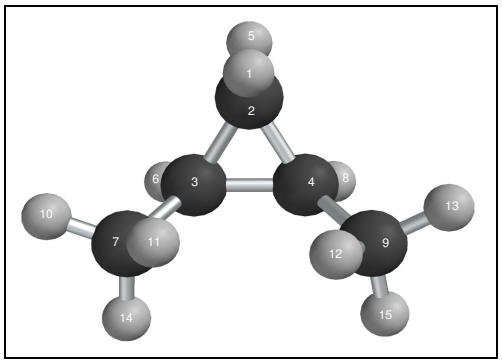

Figure 3.8 *Dimethyl cis-cyclopropane showing the numbering system that was used in the PES calculations for this molecule. D14 defined by atoms 14, 7, 3 and 2 and D15 by 15, 9, 4 and 3.*

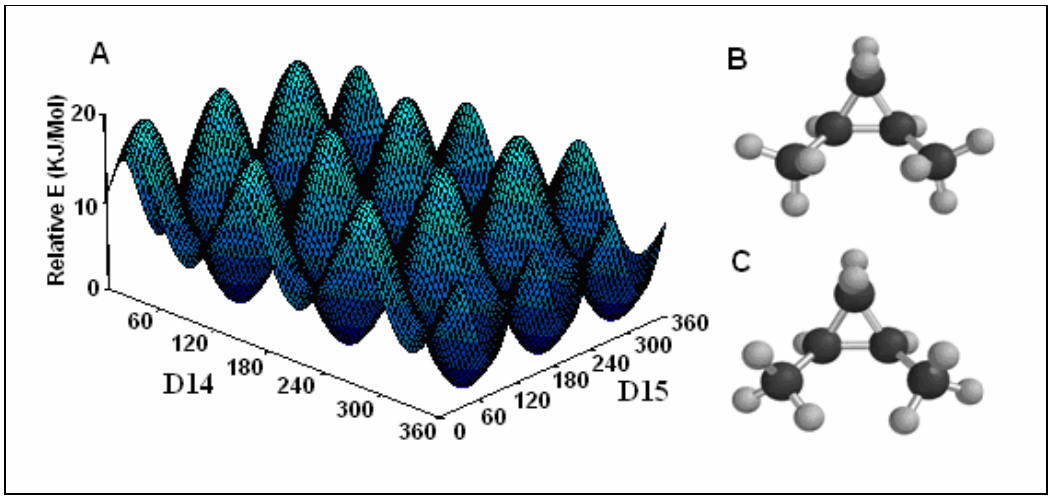

Figure 3.9 *A PES calculated for dimethyl cis-cyclopropan. (A) 30° rotations in dihedral angles D14 and D15. The low (B) and high (C) energy conformations that correlate to the high and low areas of the PES. (For raw data and calculations refer to Appendix C.)*

In Figure 3.9 the energy is plotted as a function of D14 and D15, resulting in a 3-dimensional PES. The energy minima (valleys) occur where both D14 and D15 are

equal to 90°, 210° and 330° (corresponding to structure B), and the energy maxima (peaks) where both D14 and D15 are equal to 30°, 150° and 270° (corresponding to structure C).

This work can be extended to PEHSs involving four dihedral angles (when chain length is 2 carbons), six dihedral angles (when chain length is 3 carbons) and so on, until the limit of current computational resources is reached. When this limit is reached, further in-depth analyses of the extensive data-sets will provide insight as to the reasons for the exact chain lengths and associated interactions between them. What can be done with confidence is to use the qualitative representations of these highly dimensional systems to generate trends and predict where conformational minima will lie on the PEHSs of these systems. Effectively these serve as a means to provide an idea of what the preferred conformations are as well as some information of the energetics and dynamics associated with interchange between conformers.

Table 3.3 *The stabilised bond lengths and angles measured as the chain length was increased in a symmetrical way for R- and S diastereomers of α-methyl trans-cyclopropane*

	Internal Angle (degrees)			Bond length (Angstroms)			Outer angle (degrees)		Energy (kcal/mol)
	A_1	A_2	A_3	L_1	L_2	L_3	A_4	A_5	
S	60.3	59.8	59.8	1.510	1.502	1.502	119.2	119.5	-166.659*
R	60.3	59.8	59.8	1.510	1.502	1.502	119.2	119.6	-164.981*

* The energies refer to the cyclopropane with a total of 32 carbons in its chains (refer to Appendix B).

Equilibrium geometries were also calculated for trans-cyclopropane with an α-methyl group in the S or R stereochemistry with increasing hydrocarbon chain lengths (Table 3.3). In nature, trans-cyclopropane is always accompanied by an α-methyl group. With an α-methyl group, all trans-cyclopropane calculations did not have any imaginary frequencies anymore, in contrast to the structures of the trans-cyclopropanes without the α-methyl group with a carbon-chain of 10 or more on each side. It was seen that between the two diastereomers S was slightly more stable than R. The S-diastereomer is believed to be that found in nature [115]. Similar to cis-cyclopropane (Table 3.2), it was seen that the key structural parameters obtained for α-methyl trans-cyclopropane (S or R) are unaffected by the chain length.

In summary of the findings for cis- and trans-cyclopropanes (Table 3.2 and Table 3.3), it was found that the key structural parameters of the cyclopropane rings are not affected by the lengths of the carbon chains attached to the ring. This is understandable, since the angle of the chains with the cyclopropane rings (Figure 3.6 A4 and A5) did not change very much (Table 3.2 and Table 3.3). For trans-cyclopropane, the α-methyl group stabilsed the symmetric structure so that there were no more negative frequencies, as was seen without the α-methyl group. The α-methyl trans-cyclopropanes (S or R) were slightly more stable than corresponding cis-cyclopropane, and cis-cyclopropane had a slightly longer L1 bond length, with correspondingly wider A1 angle opposite to it. This is thought to be due to steric reasons with the two bulky methyl-groups next to each other in cis-cyclopropane. Finally, PES/Hs are an effective tool to analyse the effect of changes in the geometry on the energy of the molecule and to predict what the favoured conformations will be.

3.5.2 Molecular dynamics

3.5.2.1 Selection of force field

The cyclopropane group with its strained ring calls for careful consideration, as this is also a point where folding is likely to occur in the molecule. Most commonly used force fields do not include parameters for such infrequently used groups. From four available force fields in the Materials Studio software (Table 3.4), each force field was applied to a cis- and trans-cyclopropane with two 16-carbon chains, where the bond lengths and angles have stabilized to constant values in the semi-empirical calculations (Table 3.2). The same bond lengths and angles of the cyclopropane ring (Figure 3.6) were measured after each force field was applied. The square of the difference between these values and the semi-empirically-obtained values were calculated in order to find the force field that resulted in the cyclopropane bond lengths and angles closest to those obtained in the semi-empirical calculations. Therefore, the angle and bond length squared difference average closest to zero represents the structure which has bond lengths and angles with the least variation compared with the semi-empirical values (Table 3.4). It was found that the Compass

force field had the closest fit for both cis- and trans-cyclopropanes and was subsequently used in all MD calculations.

Table 3.4 *The square of the differences of the bond lengths and angles as calculated for cis- and trans-cyclopropane with two 16-carbon chains, with the respective force fields.*

Forcefield used	cis/ trans	Internal Angle (degrees)			Bond length (Angstroms)			Average†
		A_1	A_2	A_3	L_1	L_2	L_3	
Compass	trans	0.591	0.151	0.151	0.043	0.178	0.178	0.215
Compass	cis	0.005	0.004	0.031	0.256	0.205	0.149	0.108
PCFF	trans	0.101	0.009	0.050	0.691	0.260	0.186	0.216
PCFF	cis	0.023	0.050	0.009	0.398	0.186	0.260	0.155
CVFF	trans	0.201	0.051	0.051	2.899	4.180	4.180	1.927
CVFF	cis	0.036	0.000	0.000	3.376	3.965	3.965	1.891
Amber	trans	0.537	0.137	0.137	1.548	3.421	3.421	1.533
Amber	cis	0.188	0.018	0.018	2.099	3.186	3.186	1.449

†The average represents the average of square differences of angles and bond lengths.

3.5.2.2 Equilibration of the system

In order to see if the 4 ns MD runs had reached convergence, a Pressure-Potential Energy Fluctuation graph was plotted. From this it was evident that the MD runs needed approximately 1.5 ns to equilibrate. A cut-off value of ±7.938 was assigned for the first derivative of the Pressure-Potential Energy function Figure 3.10. This led to the last 2.51 ns of the MD simulation to be used as the equilibrated part of the run (frames 150-400).

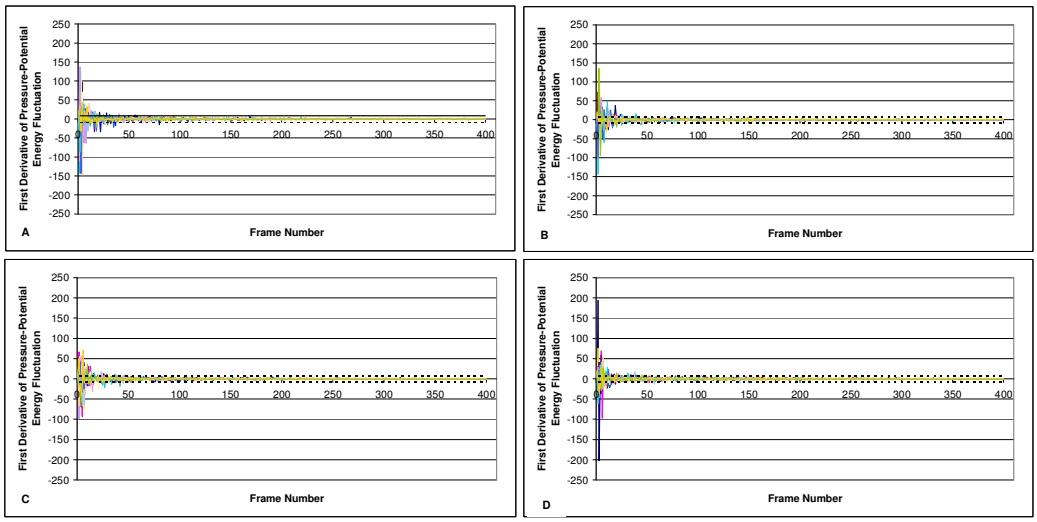

Figure 3.10 *The first derivative of Pressure-Potential Energy plots of MAs. Twenty replicates of α-MA (A), MeO-MA (B), MeO-ME (C) and Keto-MA (D) were plotted with the ±7.9 cut-off shown as dashed lines.*

3.5.2.3 Principal Component Analysis

It is believed that the PES was explored well for each MA-type. Figure 3.11 (A) shows the PCA plot of one representative α-MA run of the 20 replicates that were performed. The plot makes chemical sense because related structures are close to each other in time sequence (frame order) and therefore the 'path' taken by the time series through the principal component space can be traced as conformational rearrangement takes place. This path starts at the right-hand side and moves to the left where the time-points are densely populated. A representation of the variation and different pathways included by the random starting velocities from the Boltzmann-distribution in the simulations can be seen in the comparison of the PCA plot for one representative α-MA (A) to that of all 20 replicates for that class (B) in Figure 3.11. The variance of the data in both plots is a good representation since 88.67% and 80.60% of the total variance for (A) and (B) respectively is displayed on the 2D-projection of the data.

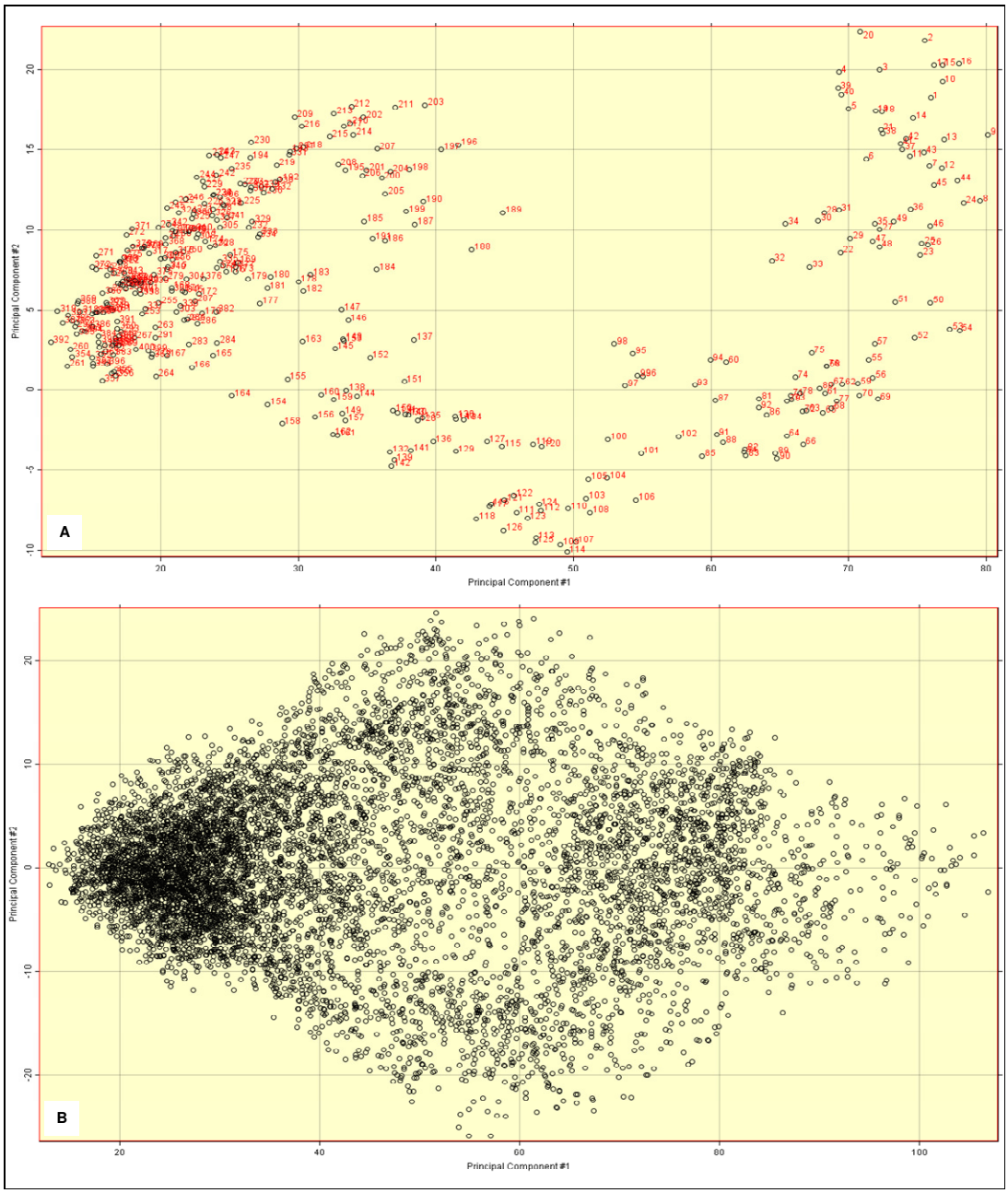

Figure 3.11 *The PCA plots for α-MA. (A) one representative MD simulation of α-MA and (B) all of the 20 replicate MD simulations for α-MA. Each of the plots includes all 400 frames of the simulations. In (A) 88.67% (80.35% + 8.32%) and in (B) 80.60% (69.62% + 10.98%) of the total variance is accounted for by the two principal components. Original PCA reports can be found in Supplementary Data\PCA_reports.*

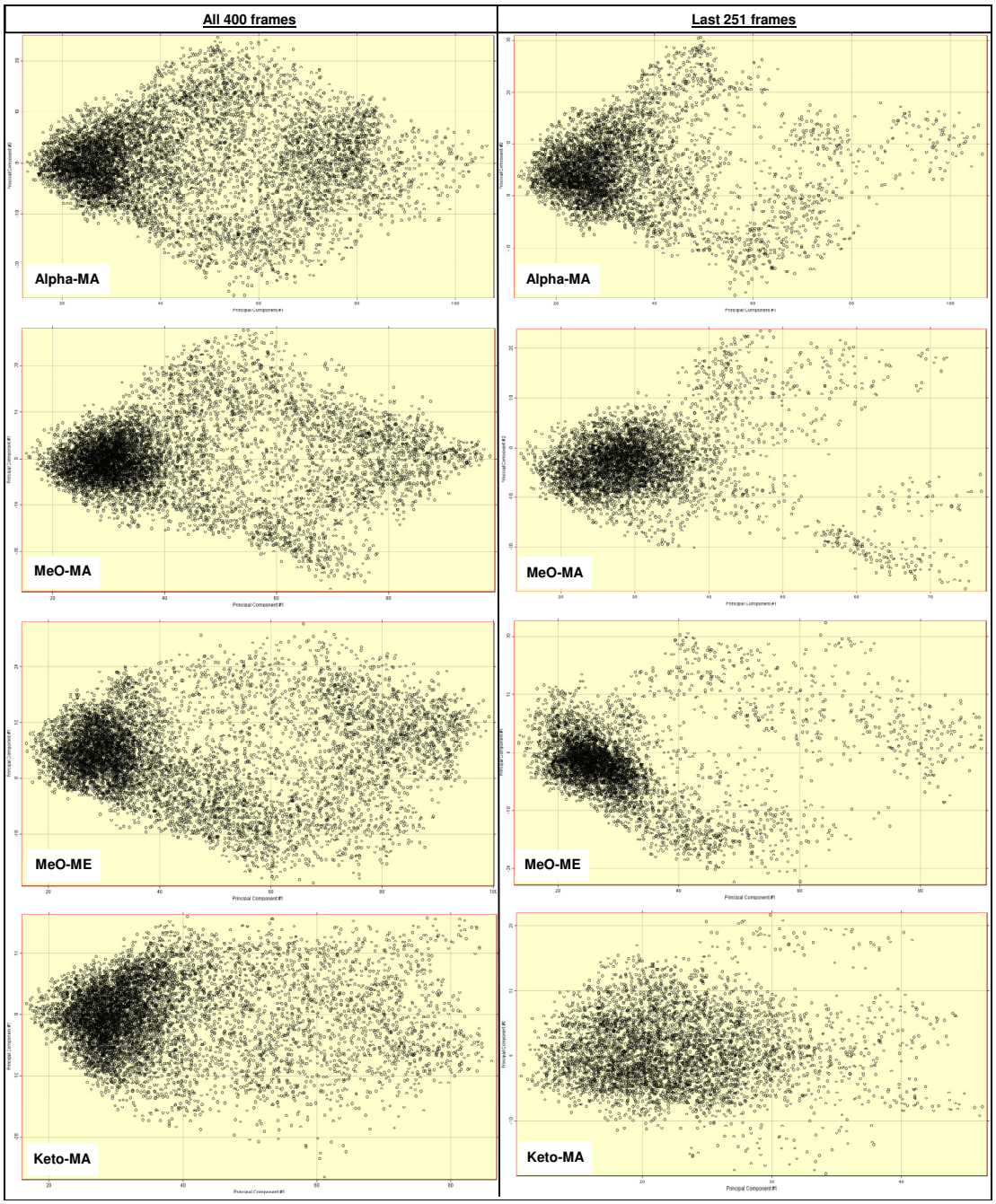

Figure 3.12 *PCA plots of α-MA, MeO-MA, MeO-ME and Keto-MA. For all 20 replicates of each MA all 400 frames (left) and the last 251 equilibrated frames (right) were used to draw up the plots. The percentage of the total variance accounted for in each plot is 80.60% (69.62% + 10.98%, Alpha, left), 73.94% (60.82% + 13.12%, Alpha, right), 75.93% (64,58% + 11.35%, MeO, left), 56.99% (38.62% + 18.37%, MeO, right), 79.34% (69.94% + 9.29%, MeO-ME, left), 67.93% (55.23% + 12.69%, MeO-ME, right), 66.51% (56.94% + 9.58%, Keto, left), 41.20% (22.00% + 19.20%, Keto, right). Original PCA reports can be found in Supplementary Data\PCA_reports.*

When considering all 8000 frames of each MA type (ie. 400 frames of the 20 replicates for each type, Figure 3.12 left hand side), it is seen that each of the four types show a group of structures toward the left-hand side that is very densely populated. In MeO-MA, there are less densely populated regions that lead to the dense region diagonally up- and downwards, whereas in MeO-ME there is only a predominant group leading diagonally upwards to the densely populated group. This indicates that within the structural pathways sampled, two pathways were generally followed to the final group of structures in MeO-MA, but only one pathway was frequently followed in MeO-ME. In α-MA the same trend is seen as for MeO-MA, but with the downward diagonal being the more frequently used path, compared to the diagonal going upwards. Keto-MA does not show any significantly populated regions around its densely populated region. This indicates that within the structural pathways sampled there is no significant preference for pathways to get to the densely populated region of structures, in contrast to the other MAs. α-, MeO- and MeO-ME also show a less densely populated region on the right-hand side of the plot that is nearly absent in keto-MA.

In the latter part of the MD run that is well equilibrated, ie. plots containing 5020 frames of each MA type representative of conformational snapshots taken only from 1.5 – 4 ns of the MD run (Figure 3.12 right hand side), the trends that are described above for the full MD runs still hold. The densely populated region to the left in each plot is still the most densely populated region for each MA type. The less-dense regions that lead into this region are still visible, but are now very lightly populated. The less densely populated regions to the right in α-, MeO- and MeO-ME plots are absent in the second set of plots on the right hand side in Figure 3.12. This indicates that this region represents the structural folds of the MAs at the start of the MD runs, in other words, the more open structures. It appears that keto-MA folded differently compared to the other types of MAs, since it did not have any significant population on the right-hand-side of the plot when all 400 frames were considered and therefore does not show any significant number of frames with relatively open folds. Instead, there is a wide, lightly populated area on the right hand side that is seen (other than the one dense area) that disappears when only the last 251 frames are taken into account as well as the dense region itself that becomes less dense when only the equilibrated part of the run is considered. This seemingly less dense population in

Keto-MA's plot for the last 251 frames is in fact representing the dense region from Keto-MA's plot for all 400 frames stretching from approximately -10 to 10 on the y-axis and from 0 to 40 on the x-axis when the difference in scale of the two plots is taken into consideration. This indicates that the folds assumed by keto-MAs within the first 1.5 ns of the MD run do not have common folds or pathways as in the other MAs, but that all of the pathways taken lead to the very densely populated region of folds very quickly.

The percentage of the total variation that is displayed by the 2D-plots also confirms the observations made above. The variance displayed in the plots where all 400 frames were included for Alpha-, MeO-, MeO-ME, and Keto-MA (80.60%, 75.93%, 79.24% and 66.51% respectively) is more than that displayed in the plots where only the last 251 frames were included (73.94%, 56.99%, 67.93% an 41.21% respectively). This once again shows that by removing the more open, less folded structures from the first 1.5 ns of the MD simulation, the population that is left is a population of folded structures, having less structural variation. The smaller variation of keto-MA when all 400 frames are taken into account is representative of the lack of a significant population of open, less folded structures, as seen in the other MAs. In the Keto-MA plot with the last 251 frames the very small variation is typical of the one dense population consisting of a number of similarly folded structures. This once again confirms that keto-MA folded much faster than the other MAs, having only one major populated region of folded structures in the last 2.51 ns of its simulations.

3.5.2.4 Self-Organising Maps

When comparing the SOM plots of the four subclasses containing all 8000 representative frames for each class, some general trends can be seen. Focus will be kept on the ten groups that are the most highly populated in each SOM out of a total of 49 groups (Table 3.5).

Table 3.5 *The ten most highly populated clusters in SOM for 8000 frames of each MA type. These are shown in descending order of cluster size with their conformations indicated.*

MA-type	Cluster †	Number of frames ‡	Conformation
Alpha	35	521	F
	47	450	F
	5	380	I
	1	331	O
	7	302	I
	8	291	O
	49	262	F
	45	260	F
	21	240	F
	4	224	O / I
Methoxy	1	486	O
	38	480	F
	26	417	F
	21	264	I
	49	257	F
	15	254	O
	43	238	F
	2	234	O
	7	229	I
	35 or 45	220	F
Methoxy-ester	47	401	F
	1	382	I
	5	353	O
	8	338	I
	31	333	F
	7	320	O
	43	275	F
	6	270	O
	49	233	F
	29	221	F
Keto	1	456	O
	45	416	F
	21	367	F
	35	367	F
	44	329	I / F
	6	280	I
	19	262	I
	15	255	I
	32	252	F
	31	250	I / F

O = open, I = intermediately folded, F = folded. When any two conformations are shown, there is a combination of the two conformations.

† The cluster number refers to the position of the cluster in the SOM grid (refer to appendix D for the original SOM-plots from which the data in this table has been obtained).

‡ The number of frames in each cluster is labelled as the number of genes (indicated in brackets next to the cluster number) by default as for microarray data that is normally used by this program on the SOM plots.

Structures out of the ten most highly populated groups from each class represent 38.4 to 40.7% of all the structures and mostly represent relatively compact structures of which the ac, ae, ce and bd distances are ~20 Å or less. From these highly populated groups, the most open structures (those with the highest peaks at ac, ae, ce and bd distances) are slightly favoured over intermediately folded structures by MeO-MA and MeO-ME. Table 3.5 shows that α-MA has a relatively even distribution between open and intermediately folded structures, while keto-MA has hardly any representatives of the very open structures amongst its highest occupied groups, and an almost equal amount of intermediately folded structures to the most folded ones. From this it was learnt that in all of the MA subclasses, the most compactly folded groups of structures are the most abundant in the ten most highly populated groups. This shows the tendency of the MAs to fold up into compact structures, even when starting from open, unfolded structures. This is reinforced by the dense population of folded structures on the left-hand side of each PCA plot as discussed above.

α-MA, MeO-MA, and MeO-ME all had a significant number of highly populated groups that represent open structures, showing that they spend a significant amount of time in open, unfolded conformations before they entangle into more tightly folded conformations. The pathways that were commonly followed to the folded group of structures in α-, MeO- and MeO-ME PCA plots were representative of these groups of structures. Keto-MA, on the other hand has a considerable number of highly populated groups that represent intermediately folded conformations. Once again this shows that keto-MA folds more rapidly than the other three MAs, and therefore has more structures that are half-way folded and hardly any that are in the open, unfolded states.

Next, SOM plots of 5020 frames for each subclass representing the last 251 frames (2.51 ns) of each 4 ns MD simulation were considered. These structures represent the types of groups that the molecules fall into in the equilibrated part of the simulation. The seven groups that are the most highly populated in each SOM out of a total of 25 groups (representing 37.5 to 42.4% of structures) are discussed.

Table 3.6 *The seven most highly populated clusters in SOM for 5020 frames of each MA type. These are shown in descending order of cluster size with their conformations indicated.*

MA-type	Cluster †	Number of frames ‡	Conformation
Alpha	17	368	F
	1	329	F
	6	285	F
	16	284	F
	2	281	F
	24	264	O / I
	22	261	I
Methoxy	24	411	I / F
	23	305	F
	3	314	I
	25	288	I / F
	21	283	F
	18	260	F
	22	258	F
Methoxy-ester	18	332	F
	11	302	I
	21	289	I
	14	278	F
	4	276	F
	23	274	F
	25	259	F
Keto	1	323	F
	3	313	F
	16	272	F
	11	259	I
	2	257	F
	12	233	F
	18	221	F

O = open, I = intermediately folded, F = folded. When any two conformations are shown, there is a combination of the two conformations.

† The cluster number refers to the position of the cluster in the SOM grid (refer to appendix D for the original SOM-plots from which the data in this table has been obtained).

‡ The number of frames in each cluster is labelled as the number of genes (indicated in brackets next to the cluster number) by default as for microarray data that is normally used by this program on the SOM plots.

When the seven most highly populated groups are looked at in this set of results as summarised in Table 3.6, a 20-30% and 70-80% representation of the intermediately folded and more compactly folded groups is seen respectively, for all the MA subtypes. Except for α-MA (with 7%), there are no open and unfolded groups of structures in the equilibrated parts of the simulations. This is expected since the first part of the simulation is excluded in these SOM's. Therefore, once the first 150 frames of each simulation were removed, keto-MA shows a similar distribution in its

folded states to the rest of the MAs, with most of them being compactly folded, and some intermediately folded.

3.5.2.5 "WUZ" analysis

Each frame was also analysed for W-, U- and Z-folds that have been described in literature [13, 17] and in more detail by Prof. D.E. Minnikin, University of Birmingham, UK (personal communication, July 2006). The seven possibilities of W-, U- and Z are shown schematically in Figure 3.5. The fold in each frame was identified according to a set of parameters that describes each of the seven different W, U, and Z folds with eight distances (Table 3.1). The number of frames (or duration) that these folds appear in varied from 1 or 2 frames up to 50 or more frames (ie.10 ps to more than 0.5 ns). These may be continuous stretches, but may also be interrupted by frames that are not identified as W's, U's or Z's. These unidentified frames between stretches of W's, U's or Z's very often did not have a methylene chain(s) that was stretched out enough, or one of the other parameters fell just outside of the fold boundaries.

Grouping structures into W-, U- and Z-folds is a biased grouping method. Known folds were identified and parameters were largely obtained from representative folds within the simulations. For this reason WUZ-groups were compared to SOM-groups which were obtained in a non-biased manner based on data-driven machine learning. The grouping (or clustering) by the 'WUZ'-analysis and the SOMs correspond well (comparisons of the original SOM-report and WUZ-data for all simulations can be found in Supplementary Data\SOM_WUZ_groups). As an example, one of the α-MA simulation results is shown in Figure 3.13. This example was chosen because it had a high frequency of 'WUZ'-structures identified in the WUZ-analysis and therefore it served as a good example to make a comparison between 'WUZ' and- SOM-groups.

As can be seen in Figure 3.13, the different W, U, and Z-folds were also grouped separately in the SOM clusters. In fact, there is often more than one SOM cluster for one type of fold, indicating that there are subclasses within the currently defined folds. It will be the undertaking of future studies to unravel the differences within the classes

and to assign them to subclasses. The number of frames (genes) in a cluster is commonly more than the number of W-, U- or Z-folds that match with the structures in that cluster. Most often these extra structures in a cluster are the structures in between multiple frames of a certain fold (as grouped by the WUZ analysis) that are not classified as that fold, even though they may still be recognised as such, as described above.

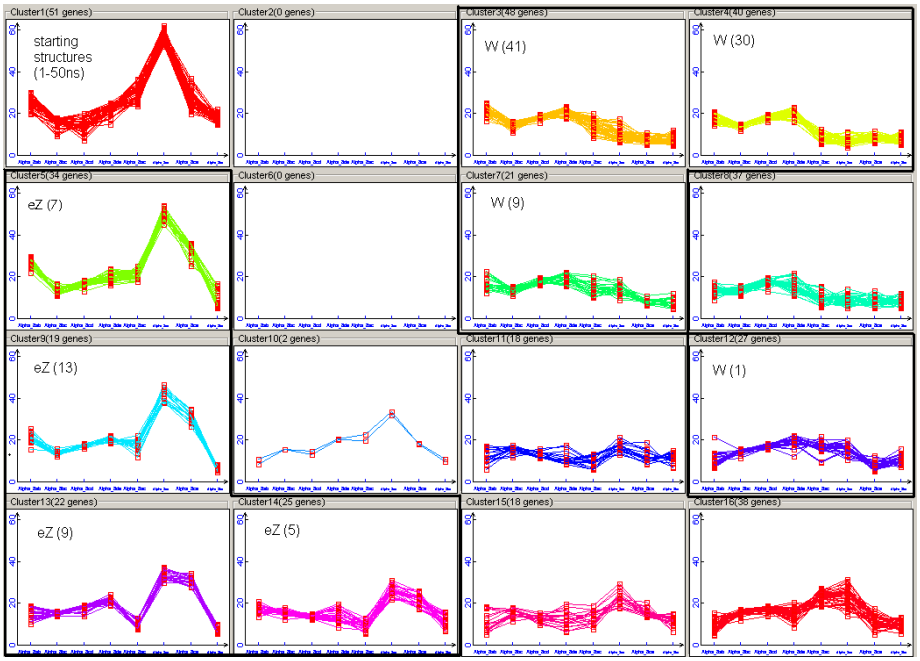

Figure 3.13 *Self Organising Map of one representative α-MA MD simulation. Each line represents the length (Å, y-axis) of each of the eight selected distances (ab, bc, cd, de, ac, ae, ce, bd, x-axis) for one frame. The number of frames in each cluster is labelled as the number of genes (indicated in brackets next to the cluster number) by default as for microarray data that is normally used by this program. The annotation of the WUZ-analysis is shown in the top left-hand corner of each cluster, indicating the conformation and the number of frames with that conformation. The comparison of the original SOM-report and WUZ-data can be found in Supplementary Data\SOM_WUZ_groups\Alpha\Alpha-8.xls.*

In the example in Figure 3.13 there was no overlap where two WUZ-groups fell into the same SOM-cluster. On the infrequent occasion that this did occur, there was only ever a maximum of two structures (frames) from a different fold included in the same SOM cluster consisting of another WUZ-fold in the majority. These were also closely related structures which overlapped because they were on the borderline of the parameters for that certain fold. In each of α- and keto-MA as well as MeO-ME there

was only ever one such occurrence in all the replicates. An example from α-MA is discussed in Appendix E.

Table 3.7 *Frequency of occurrence of WUZ-folds during last 251 ns of molecular dynamics for Alpha-MA*

Fold	Frequency of occurrence in frame 150 to 400 for each replicate†																				Total frequency‡
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
W		13.94			0.40			32.27	6.37	0.80			9.16	3.19	0.40						3.33
aZ				20.72					0.40							7.57			0.80		1.47
bZ																					0
eZ																					0
bU							10.36									3.98	6.77				1.06
cU									0.40											3.19	0.18
dU			13.55			3.98								0.40							0.90

† The frequency of each fold is represented as a percentage out of 251 frames

‡The total frequency represented as a percentage of 5020 frames (251 x 20)

Table 3.8 *Frequency of occurrence of WUZ-folds during last 251 ns of molecular dynamics for MeO-MA*

Fold	Frequency of occurrence in frame 150 to 400 for each replicate†																				Total frequency‡
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
W					2.39									9.16					0.40		0.60
aZ																					0
bZ		11.95																			0.60
eZ																					0
bU											0.40										0.02
cU																					0
dU	1.59																6.77				0.42

† The frequency of each fold is represented as a percentage out of 251 frames

‡The total frequency represented as a percentage of 5020 frames (251 x 20)

Table 3.9 *Frequency of occurrence of WUZ-folds during last 251 ns of molecular dynamics for MeO-ME*

Fold	Frequency of occurrence in frame 150 to 400 for each replicate†																				Total frequency‡
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
W							17.53	7.57	0.40			1.99									1.37
aZ																				1.20	0.06
bZ															3.98				0.80		0.24
eZ								0.80													0.04
bU													1.99								0.10
cU																					0
dU													5.18						19.12	0.40	1.24

† The frequency of each fold is represented as a percentage out of 251 frames

‡The total frequency represented as a percentage of 5020 frames (251 x 20)

Table 3.10 *Frequency of occurrence of WUZ-folds during last 251 ns of molecular dynamics for Keto-MA*

Fold	Frequency of occurrence in frame 150 to 400 for each replicate†																				Total frequency‡
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
W		1.20		1.20	10.36			8.37			0.40		3.59	0.40		0.40	0.40				1.31
aZ																					0
bZ	8.37																				0.42
eZ																					0
bU									0.80												0.04
cU																					0
dU																					0

† The frequency of each fold is represented as a percentage out of 251 frames

‡The total frequency represented as a percentage of 5020 frames (251 x 20)

From the SOM-plots α-MA may be described as the MA that folded into a compact structure the slowest. However, in WUZ-analysis α-MA had the highest frequency of WUZ-structures (Table 3.7). Although the frequencies presented in tables 3.7 – 3.10 seem low, it is significant to see even such low frequencies under the simulation conditions, namely with no restrictions applied, no external pressure involved and no packing effects from surrounding molecules involved.

Unlike the other MAs, WUZ-structures appeared in each of the 20 replicates of α-MA, but still in a remarkable 16 replicates when only the equilibrated part of the simulations were taken into account (Table 3.7). From the data in Table 3.7 it is clear that a W-fold was most favoured, followed by aZ- and bU-folds during the equilibrated part of the simulation. The folds that show zero frequencies, namely bZ and eZ, did arise during the first part of the simulation that is not considered here. Therefore, on the whole α-MA assumed all of the seven folds within its 20 replicates indicating that α-MA is flexible and bends with ease at its acid- and cyclopropane-groups.

A comparison of the minimised energies of the seven WUZ-folds from Figure 3.14 is shown in Table 3.11. It appeared as if the structures were stabilised by the degree of van der Waals interactions that occurred between the methylene chains. The W-conformation with all four methylene chains in close proximity and parallel to each other facilitates van der Waals interactions between all the chains and results in the most stable conformation. From the different U-conformations cU is more stable than both bU and dU, having the lowest potential energy (Table 3.11). Although they all bend only at one functional group, cU bends roughly in the middle of the molecule which makes it possible for all the methylene chains in the molecule to interact with an adjacent chain by means of van der Waals interactions. The three Z-conformations, each folding at two functional groups, showed energies that varied from the highest (least stable) to almost the lowest (most stable). From these three conformations, bZ is the only one where the molecule does not bend at group c (one of the cyclopropane groups), and it results in the least stable conformation. Between aZ and eZ, eZ has the most stable conformation which may be attributed to the long alpha-chain being in close proximity to adjacent methylene-chains, facilitating van der Waals interactions to stabilise the molecule. In contrast, the long alpha-chain is

extended and isolated in the aZ-conformation. An eZ-conformation could originate from bU and the relative lack of stable eZ-conformations in α-MA may be due to the lack of bU-conformations (in comparison to the abundance of dU-conformations) in the first 149 frames in α-MA simulations that are not considered in Table 3.7. This concept will be considered in more detail below.

Table 3.11 *The total Potential Energy of the minimised structures of the different folds in α-MA, in descending order.*

Conformation	Total Potential Energy (kcal/mol)
W	-138.8
eZ	-116.6
cU	-113.2
aZ	-94.3
bU	-89.0
dU	-83.6
bZ	-69.6

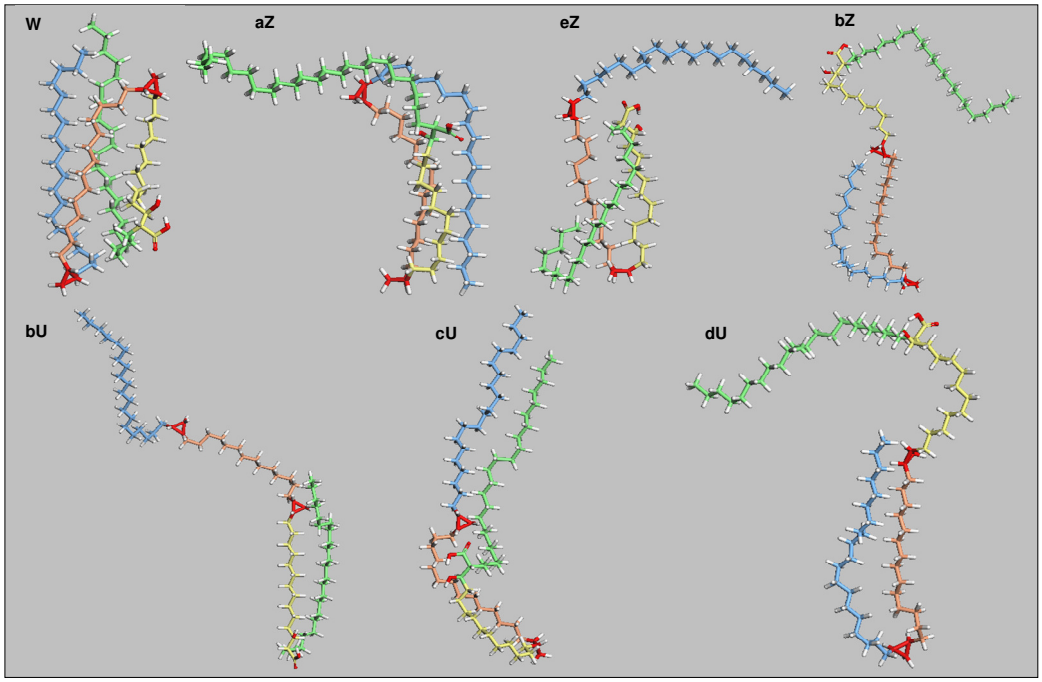

Figure 3.14 *The seven folds represented as minimised structures from α-MA. The molecules are colour-coded for clarity; the oxygen atoms in the mycolic motif and in the distal functional group of the mero-chain, as well as all cyclopropanes are red; the methylene chain between a and b is green, between b and c is yellow; between c and d is orange and between d and e is blue.*

MeO-MA is succeeded by its methylated analogue and ranks in third place for its frequency of WUZ-structures when the equilibrated part of simulations are taken into

account (Table 3.8). If the whole 4 ns were taken into account, 11 replicates showed occurrences of WUZ-folds, but as Table 3.8 indicates, only 7 replicates showed WUZ-folds during the equilibrated part of the simulations. W-, bZ- and dU-folds were most frequent with bU in the minority. The remaining three folds, aZ, eZ and cU were never come across in MeO-MA folding. Interestingly the structures that occurred, W; bZ; dU and bU, were structures that bent at groups b and d (the carboxyl- and methoxy-groups respectively). Out of the four folds which require bending at the cis-cyclopropane at group c (W, aZ, eZ, cU), only one occurs in all the MeO-MA replicates, namely the W-fold (refer to Figure 3.5). This may indicate a preferred bending of MeO-MA at functional groups other than its cyclopropane group.

Similar to MeO-MA, MeO-ME had WUZ-structures in 12 and 8 replicates when taking the entire 4 ns and only the equilibrated part of the simulation into consideration, respectively. W and dU were most frequently found in MeO-ME folding (Table 3.9). MeO-ME had nearly double the frequency of WUZ-structures compared to MeO-MA and only cU was not observed at all in the 20 replicates of MeO-ME. Therefore, methylation of MeO-MA increased its propensity to assume WUZ-structural folds and MeO-ME vaguely showed more flexibility at its cyclopropane group than MeO-MA, folding into structures that fold at functional groups b, c and d, although at low frequencies.

Similar to MeO-MA, keto-MA also showed a lower frequency of WUZ-structures occurring. In Table 3.10, 11 out of the 20 replicates showed WUZ-folds, which differs only slightly from the 14 replicates that showed WUZ-folds during the complete 4 ns simulations. During the equilibrated part of the simulations W occurred most frequently, bZ less and bU the least (Table 3.10). When the entire 4 ns was taken into account, eZ and cU structures did arise during the starting frames of the simulations, leaving aZ and dU folds that were never come across in any of the keto-MA replicates. The four WUZ-folds that require folding at position d, its ketone group, are W, aZ, bZ and dU, of which aZ and dU are never found in keto-MA folding. This may indicate a preference of keto-MA molecules to start folding on the opposite side of the molecule, bending initially at the acid and trans-cyclopropane groups (b and c) rather than at its keto functional group (d).

On average roughly one (MeO-MA, MeO-ME and keto-MA) or two (α-MA) types of WUZ-folds can be seen in any one MD run. Therefore, it was not feasible to uncover whether there were certain trends / patterns in MA folding. Logically, MAs would be expected to fold into U-shapes, leading to Z-shapes, and finally to the more folded W-shapes. In α-MA, the first 149 frames of all the 20 replicates taken together had 286 frames classified as WUZ-folds, out of which more than half were dU-folds. MeO-MA and MeO-ME did not have many WUZ-structures (less than 100) in the beginning of their simulations, but once again at least half of these were dU-folds. The first 149 frames in all 20 keto-MA replicates had 152 WUZ-structures. Contrary to the other MAs, approximately half of these structures were bU's. It was mentioned before that dU-shapes never occurred anywhere in all 20 keto-MA replicates. Here again keto-MA differs from the other MA-classes. Perhaps the long alpha-chain (ab) being drawn in towards the bc chain in the bU-fold inclines keto-MAs to fold into compact structures faster than those MAs that were more likely to fold at the other extremity of the molecule first (via dU).

When Villeneuve *et al*. [18] observed that keto-MA had the propensity to stay in a W-fold, whereas MeO-MA more often unfolded into an open structure, they suggested that hydrogen bonding between the keto/methoxy-group and the acid group and the methylene chain lengths influenced this tendency. They assumed that there will be intra-molecular hydrogen bonding between the 3-hydroxy group and the carbonyl-group of the carboxyl group. In their MD simulations the distance between these H- and O-atoms was restricted to 2.4±0.6 Å. For MAs of all subclasses, even MeO-ME, this hydrogen-bond is possible. However, in the present study no restrictions were put on this hydrogen bond, and it was seen that the groups rotated freely.

For oxygenated MAs there is also another possibility of hydrogen-bonding involving the keto or methoxy functional group in the mero-chain. Figure 3.15 shows distributions of the bd distances for all the MAs over 4 ns simulations. The b and d groups (carboxyl and keto / methoxy respectively) are the groups that are able to hydrogen-bond in oxygenated MAs and therefore bd distances are a good indication of the likelihood of hydrogen-bonding between these groups. Figure 3.15 clearly

indicates that the bd-distance in keto-MA had a much higher probability of being below 10 Å than the other MAs. Therefore hydrogen-bonding between these groups could be driving the folding of keto-MA in a way that brings these groups together faster and results in more folded structures more quickly.

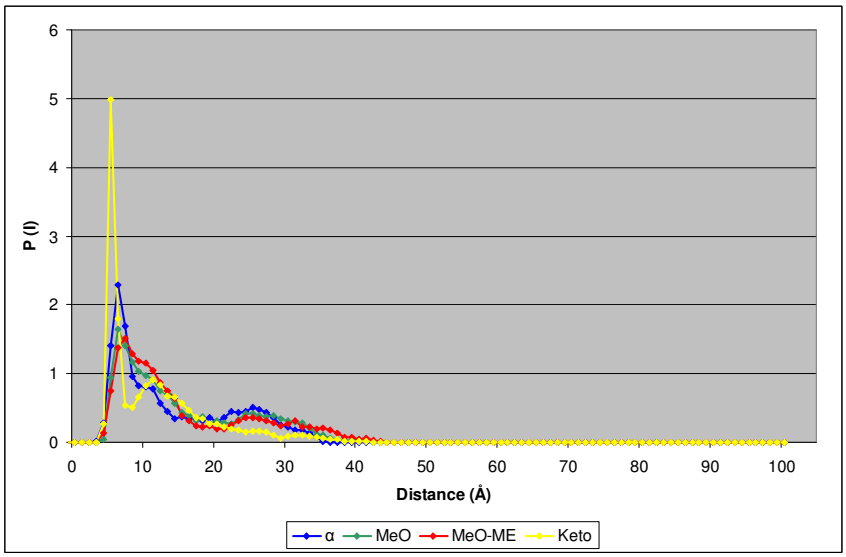

Figure 3.15 *The sum of distributions for the bd distance of each MA-type. A total of 17 replicates were studied. (Distance distributions of 3 simulations that could not be included were confirmed to be due to a technical problem in MS 4.1 software that could not be solved to date. Therefore, the same replicate numbers were left out in all the MA-classes so that the plot can still be a true and unbiased representation of the distributions.)*

Figure 3.16 shows an example of the two types of hydrogen-bonds discussed above, occurring in keto-MA. Further analysis and quantification of this interaction is necessary to make any further conclusions about the role of hydrogen bonding in the folding of MAs.

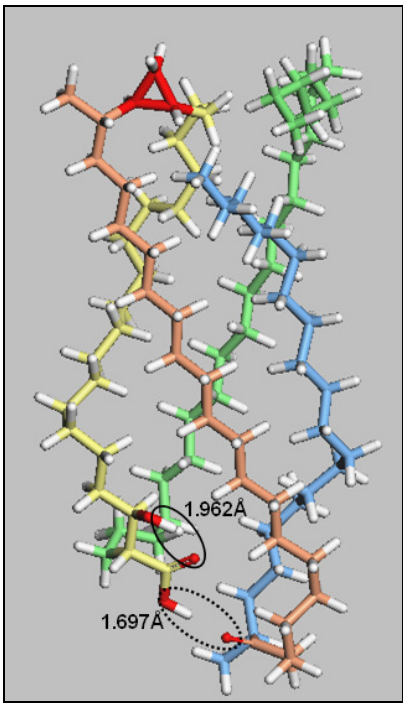

Figure 3.16 *Keto-MA shown in the W-fold. O-H Distances are shown between the ketone oxygen and the carboxyl-group hydrogen (circled in dashed black line) and the carboxyl-group oxygen and the 3'-OH hydrogen (circled in solid black line).*

Villeneuve *et al.* [18] also proposed that when the chain length between the acid and cyclopropane group (n) is shorter than that between the cyclopropane and keto/methoxy group (m), the molecule can maintain the W-fold better. When n is shorter than m the bulky polar group (methoxy or keto) sticks out of the column-like W-structure and allows interaction with the water surface in Langmuir monolayer experiments. It also allows for tight and ordered packing of the methylene chains, which would be disturbed when the bulky group is inside the column-structure in the case of n being larger than m.

During these simulations that were performed in a vacuum, the only interactions that could govern the folding of MAs are van der Waals interactions between the methylene chains (that can be stabilised as hydrophobic forces in an aqueous environment) and intra-molecular hydrogen bonding. From the MAs from *M. tb* that were used in the present study both α- and keto-MA had n shorter than m chain. It was evident from observing MA-folding that methylene chain lengths may allow certain folds to exist or not, or affect how comfortably the MA assumes that fold.

However, it would be inaccurate to make conclusions on the precise effect of methylene chain lengths on MA-folding here since the chain lengths were not varied on one MA-type. A direct comparison can also not be made with the theory of Villeneuve *et al.* [18] because in this case folding occurred in vacuum without any lateral pressure or surrounding molecules – contrary to Langmuir monolayers.

In *M. tb* only keto-MA is found as a trans-cyclopropane stereoisomer in the majority. Therefore in this study, only keto-MA had a trans-cyclopropane group. The different, faster folding observed in keto-MA is presumed to be a function of its entire composition. For that reason it is not possible to tell whether its folding properties can be attributed to the ketone or trans-cyclopropane functional groups or the combination of methylene chain lengths alone. Therefore it was not possible to draw any conclusions on the effect of cis- or trans-cyclopropane groups alone on MA-folding, but stereochemistry may contribute to the diverse behaviour of keto-MA.

3.6 Discussion

Computational modeling is a key tool in deciphering the structure-function relationships of MAs. Here, one of the first attempts to model MAs has been presented. However, it is essential to keep in mind that computational modeling of MAs is still in its preliminary stages. As is the case in any theoretical work, the limitations of the approach that was used here should not be overlooked.

Unlike their natural environment in the cell wall surrounded by other MAs and covalently attached to trehalose or arabinogalactan, single MA molecules were modeled in a vacuum. This is also not directly comparable to experimental environments such as monolayers and in liposomes where hydrophilic surfaces and hydrophobic interactions with surrounding MAs feature. This study cannot shed any light on the packing effects of MAs on each other.

MAs were allowed to fold at room temperature only, which may limit the extent to which the available conformational space is explored. Higher temperatures will give more kinetic energy to the molecules, enabling them to cross high-energy barriers,

which may lead to lower energy conformations of interest. At higher temperatures the molecules may also fold faster due to their higher kinetic energy and either assume more conformations or equilibrate faster.

The most important limitation to bear in mind is the intrinsic chemical composition of each MA. The structure of each MA was chosen to represent the highest percentage of content for each MA-subclass of *M. tb*, as it has been determined by NMR and mass spectrometric analysis [17]. The methylene chain lengths, cyclopropane stereochemistry and distal functional group in the mero-chain differ in each MA (except for MeO-MA and MeO-ME that only differed by one methyl group at the acid group). Therefore, conclusions on the folding of MAs can only be made safely when the intrinsic differences of each MA-subclass are referred to as a whole. No effects of different methylene chain lengths, cyclopropane stereochemistry or MA-type were tested on MA-structures in which no other variable were changed, and therefore it is only possible to comment on the possible effects of these combined structural features on MA-folding.

The stereochemistry at carbons 2 and 3 was R, R in each case [116] and it was assumed that the acid was not ionised. In the mycobacterial cell wall the acid group is always covalently linked to arabinogalactan or trehalose and in liposomes in an experimental set-up it is possible that the hydrophobic environment may prevent its ionisation at a physiological pH. If the group is in fact ionised at this pH, electrostatic interactions will also be involved in the folding and packing of MAs.

Despite these limitations, exciting properties of MA-folding have been discovered. On the whole the most significant discovery has been the occurrence of WUZ-folds in all classes of MAs without any conformational restrictions or any lateral pressure applied to single molecules in a vacuum. This reinforces findings of monolayer studies because it is definitely plausible for MAs to assume two-, three-, or four-chain folds when they are under lateral pressure if they can assume them as single molecules in a vacuum.

Models of single MAs are of interest because single MAs are presumably presented on CD1b [48], they are known to express certain features of an *M. tb* infection [71,

72] and have different abilities to stimulate foam-cell formation and neutrophillic inflammation (Van der Beken *et al*., manuscript in preparation). Although MAs are always surrounded by other MAs or lipids, the vacuum used in these simulations at least represents an environment of low dielectric constant, similar to hydrophobic pockets that form when lipids are in contact with water.

From PCA plots it was clear that the PES of each MA was explored well, despite each replicate having the same starting structure and the simulations being run at room temperature. Therefore the random starting velocities on each atom provided ample variation in the simulations. For a single MA molecule to fold *in vacuo*, van der Waals interactions and intermolecular hydrogen bonding were the only interactions moulding the PES, and therefore it was not expected to have high energy barriers. Despite this, clear pathways to folding can be observed for certain classes.

It has been shown that keto-MA folding diverged from the other MA-classes. Keto-MA folded faster and it started folding from the alpha-chain side of the molecule, in contrast to other MAs. Intermolecular hydrogen bonding between the ketone group and the hydroxyl groups on carbons 1 and 3 may account for faster folding in keto-MA, as is suggested by a higher probability of small bd distances in keto-MA (Figure 3.15). The precise methylene chain lengths relative to each other, as well as the trans-cyclopropane group may influence keto-MA to fold quickly.

α-MA folded the slowest and had a considerably higher frequency of WUZ-structures than any of the other MAs. MeO-MA appeared to be reluctant to fold at the cis-cyclopropane group. Methylation caused MeO-ME to have a higher propensity to assume WUZ-structural folds and demonstrate slightly more flexibility at the cyclopropane group than MeO-MA. These results provide evidence that the inherent structural differences between MA-subclasses provide different structural trends.

The first structural models of MAs from *M. bovis* BCG by Villeneuve *et al*. [18] were the first to show that the folding of MAs into four-chain W-structures was plausible. The current work adds to this by showing that MAs can fold into W-, U- and Z-folds from a random, open starting structure without any conformational restrictions on 20 replicates for each MA. This work has also included α-MA and a methylated-MA for

the first time. Where Villeneuve's group added variation to 20 ps simulations in starting structures and dihedral angles, the time allowed for the exploration of conformational space was extended 200-fold in this study by increasing the duration of the simulations to 4 ns.

Even though this is a preliminary study on MA-modeling, the benefits in using this technique to uncover the structure-function relationships of MAs are already clear. Future studies will build on this pilot study. In quantum mechanics the aim is to zoom in on the fine structural details that affect MA conformations and describe the PEHSs as mathematical functions. In contrast, the general direction of future molecular mechanics will be to expand the current system to include more than one MA-molecule in order to unravel what the packing effects are and to inspect solvent effects such as adding MAs to a lipid bilayer. In any future MD simulations it should be a priority to improve the theoretical system to be a close experimental resemblance of nature. The theoretical work can be complemented by empirical data obtained in a wet laboratory to provide a set of parameters for the simulations that will allow a closer application of theory to the understanding of MAs as they occur in nature. The adaptable chemical synthesis of MAs and their structural analysis with instruments such as a 900 MHz NMR machine and the latest mass spectrometers now available should rapidly accelerate progress in this research field.

Chapter 4: Concluding discussion

Virtual models of MAs allow us to make structural predictions on any number of structures that can be used in conjunction with experimental findings. It can also give molecular explanations for experimental findings and guide future experimental studies. Models can be designed to closely represent experimental conditions, but are also not limited to naturally occurring substances or time-consuming and costly chemical synthesis of a particular structure. The current scope for MA-models includes the prediction of minimal antigenic structures of MAs that are (a) recognised by antibodies for use in the serodiagnosis of TB and (b) alleviates asthma in mice. These predictions will serve to guide the synthesis of minimal antigenic structures so that the highest possible probability of success can be expected at the minimal experimental cost. Whether the whole MA structure is needed for its biological activity, or whether it is principally acting as a scaffold for positioning certain groups together, remains to be resolved.

All MAs are extremely similar; they are very long, branched fatty acids. Between subclasses there are two differences, namely the two functional groups in the mero-chain resulting in α-, MeO- and keto-MAs with cis- or trans-cyclopropanes in *M. tb*. Would it be possible that such slight differences will have no significance whatsoever, or may it be that each MA subclass is equipped with a specific function and / or biological activity? To answer this, much can be learned from the biosynthesis of MAs and the biological effects of particular genes of the biosynthesis pathway.

The biosynthesis of MAs starts with a carbon chain produced by fatty acid synthetase I (FAS I) multienzyme complex. A chain of ~26 carbons forms the alpha chain and a chain of ~20 carbons is elongated in the disaggregated FAS II system to produce the long mero-chain. The alkyl chain from FAS I will be passed through several FAS II modules. Each module contains several enzymes and adds one additional building block to the growing chain [117]. Some FAS II modules will elongate the chain by 2-carbon units and others will introduce unsaturations. It is believed that the precursor

mero-chain for all MA subclasses consists of the long alkyl-chain with two cis-double bonds at the positions of the proximal and distal functional groups [117].

It is only once the proximal and distal functional groups of the mero-chain are added that different enzymes, specific to each subclass of MA are involved in the biosynthesis [117]. PcaA and MmaA2 are cyclopropane synthases that introduce the proximal and distal cyclopropanes for the α-mero-chain, respectively. The precursor for oxygenated MAs is produced by MmaA4, which adds a secondary alcohol- and methyl group to the distal cis-unsaturation. Methyl transferase MmaA3 adds the methyl group to the secondary hydroxyl group in the distal position to form the methoxy mero-chain while it is proposed that the carbonyl group is formed by oxidation to give the keto mero-chain. The proximal cis-unsaturation is converted to a cis-cyclopropane by MmaA2 to form cis-cyclopropanated MeO- and keto-MAs. For trans-cyclopropane oxygenated MAs, the proximal cis-unsaturation is converted to a trans-unsaturation by MmaA1 and CmaA2 converts it to trans-cyclopropane.

Investing energy into different enzymes for creating different MA subclasses would be logical if the bacterium benefits from each MA subclass. Exploitation of the mycobacterial genome has made it possible to examine the effects of eliminating some of these enzymes. The subsequent effects on pathogenesis (to be discussed) as well as the fact that FAS II enzymes are exclusive to the bacterium and not found in humans render these enzymes very attractive drug targets.

PcaA, the gene that codes for PcaA (for proximal cyclopropane of alpha-mycolates) was shown to be important in ensuring a chronic and persistent infection in mice [118]. Wild type *M. bovis* BCG and a BCG *pcaA* mutant did not show any difference in the first 6 weeks of infection in mice. However, after 6 weeks the wild type organisms persisted at constant levels while the pcaA mutant was progressively eliminated. Interestingly in *M. tb*, the *pcaA* mutant strain grew faster than the wild type during the first 3 weeks of infection, but it was also eliminated from the animal readily after that. This steady elimination had the result that mice infected with the mutant strain were all alive and healthy at day 219 of infection, while all the animals infected with the wild-type have succumbed to the infection at this stage.

In a subsequent study [119], the cyclopropane cis- or trans-stereochemistries were shown to have pathogenetic and immunomodulatory functions. Infection with the wild type, and a *cmaA2* (cyclopropane mycolic acid synthase 2) mutant which lacked any trans-cyclopropanated MAs, showed similar initial infection, but the mutant strain was hypervirulent and induced larger granulomas than the wild type. This property was transferable to the wild type by substituting it with the extractable lipids (largely CF) from the mutant strain. The purified CF lacking trans-cyclopropanes was 5-fold more potent in stimulating macrophages than CF from the wild type. Therefore *cmaA2*-dependent trans-cyclopropanation of oxygenated MAs in CF suppressed *M. tb*-induced inflammation and virulence.

Strains lacking a functional *hma* gene (coding for MmaA4, renamed *hma* for hydroxymycolic acid) only produced α-MA, confirming the hydroxyl-MA as precursor for MeO- and keto-MAs [120]. Cell walls lacking oxygenated MAs would be expected to be less fluid, and in accordance they were found to be less permeable to hydrophobic and hydrophilic substances. Mycobacterial growth without functional *hma* was also attenuated in mice, suggesting that oxygenated MAs are important for pathogenicity in mice.

MmaA3, the enzyme that methylates the hydroxyl group at the distal position of the mero-chain in order to form MeO-MA, was experimentally overproduced in *M. tb* H37Rv, *M. bovis* BCG *Pasteur* strain (normally producing only α- and keto-MAs) and the *M. bovis* BCG *Connaught* strain (producing MAs identical to *M. tb*) [121]. It was seen that keto-MA was totally replaced by MeO-MA, again confirming the hydroxylated mero-chain as a common precursor for MeO- and keto-MAs. The effect was most dramatic in the BCG *Pasteur* strain, producing α- and MeO-MA on overproduction of MmaA3 in contrast to normally producing only α- and keto-MA. In spite of the dramatic shift in the chemical nature of the MAs produced the overall organisation of the cell walls where keto-MA was substituted with MeO-MA showed no difference viewed under microscopy, in phase transition temperatures, in cell-associated and secreted lipids and proteins, and permeability. Growth at normal temperatures was unaffected, but at lower temperatures it was severely reduced. The altered cell wall composition on overproduction of MmaA3 substantially reduced the successful entry and replication inside a macrophage-like cell-line. The observation

that the wild type *M. tb* showed enhanced keto-MA production within macrophages further supports the notion that keto-MA is essential for growth inside the host and that an appropriate MA-composition is a prerequisite for intracellular growth.

Originally it was thought that one common precursor for the mero-chain was synthesised, namely a long, doubly cis-unsaturated alkyl chain. It was proposed that this chain differentiated into an α-, MeO-, or keto mero-chain when it was modified by the cyclopropane synthases [117, 118]. In contrast to the initial theory, recent evidence [122] has suggested that the mero-chain was modified as it was elongated. According to the initial theory, the mero-chain of different subclasses possibly start folding uniquely once the mero-chain has been modified by the cyclopropane synthases, but prior to this, there is one common precursor having a particular fold, as moulded by the FAS II enzymes. The more recent theory implies that mero-chains of all subclasses have a similar structural fold, as moulded by the FAS II enzymes, despite the introduction of functional groups as the mero-chain is formed. In the last reaction of the synthesis of MAs, the two components, the alpha-chain (formed by FAS I) and the long mero-chain (FAS I and FAS II), are added together by the enzyme Pks13 (Polyketide synthase 13) in a Claisen-type condensation reaction. Therefore, whether the mero-chains are modified as they are formed, or only after a doubly unsaturated precursor-chain has been formed, it seems likely that MAs of different subclasses (or at least their precursors) will be able to assume similar folds that are easily accessible during their synthesis because Pks13 has not been shown to be subclass specific. In agreement with this, the flexibility, as well as the ability of MAs of different subtypes to assume similar conformational folds *in vacuo*, has been demonstrated in this work.

After MAs have been synthesised, they are transported to the outside of the cell where CF is formed and the MAs are linked to the cell-wall arabinogalactan. Much of this process is still vague, but the role of Antigen 85 (Ag85) complex, a major protein component of the *M. tb* cell wall, has been elucidated. Ag85 consists of three proteins; Ag85A, B and C. They have all been shown to have mycolyltransferase activity in which a MA from the 6-hydroxyl group of one TMM is transferred to the 6'-hydroxyl group of another TMM to produce TDM (CF) and trehalose [123]. They also attach MAs to arabinogalactan in the cell wall [124, 125].

A non-functional mutant Ag85C in *M. tb* resulted in 40% less MA linked to the cell wall [125]. The bacteria were still able to grow both extra- and intracellularly despite this decrease in cell wall mycolates. The cell wall of this strain showed altered permeability, which, as calculated, could be due to the MAs not being enough to form an impermeable monolayer around the cell, while the parent strain had more than adequate amounts of MA to do so. The proportions of α-, MeO- and keto-MA were the same as in the parent strain and therefore, Ag85C was not specific to any one subclass of MA although it reduced the amount of product formed [125]. In contrast, *M. tb* with non-functional mutant Ag85A or B genes, showed normally mycolylated walls [126]. The growth of a mutant Ag85A strain, however, showed impeded growth in a macrophage cell line and had an increased dependence on nutrients, while a mutant Ag85B strain showed no apparent differences to the wild type [127].

The overexpression of Ag85A, B or C in an Ag85-inactivated mutant strain all showed restoration of the cell wall, which confirmed the involvement of all three enzymes in mycolylation of the cell wall components [126]. None of the three Ag85 proteins are specific for any arabinogalactan sites to attach MAs to [126]. Ag85C was not specific for any mycolate subclass, but it is not yet known whether Ag85A and B are. Since these proteins are all very closely related (68-80% in amino acid identity) with the same catalytic triad of residues conserved in their active sites, it is not expected that Ag85A and B will show any subclass specificity. This once again proposes that α-, MeO- and keto-MAs fold in a similar fashion so that they are all able to fit into the same active site of Ag85 proteins.

This does raise the question of the partial redundance seen in the three Ag85 complex proteins. The decrease in MA levels in the cell wall in an Ag85C mutant suggests that perhaps Ag85C is able to compensate for Ag85A and B in a way that they cannot fully compensate for Ag85C. This implies a partial, rather than a total redundance. However, this is an issue that still needs to be explored to be better understood.

The crystal structure of Ag85C has been obtained both with and without a covalent inhibitor [128]. These structures have already revealed structural features with which we can propose ways in which the MAs could possibly fit into the enzyme. Ag85

proteins are good candidates for investigating the structures of MAs in by docking models of MAs and / or CFs into the experimentally-determined crystal structures. Their specificity, or lack thereof for different subclasses of MAs and their combinations in CF may also be proven in this way. It is possible that Ag85 proteins play a part in the release and presentation of MAs to the host immune system by removing MAs from CF and / or arabinogalactan in the phagosome to be transported to lysosomes. If mycobacteria are stripped from their complex free lipids and reconstituted with Deuterium and Carbon-13-labelled synthetic CF, it would be possible to identify the pathway of these lipids and the proteins they interact with. Fujita *et al.* showed that the subclass and molecular species composition of MAs in CF affected biological activities such as toxicity, thymic atrophy and granulomaticity [129]. Therefore, an additional advantage of using synthetic CF is that the subclasses of MAs attached to trehalose can be selected. If MAs presented on CD1b were derived from CF, it will also be possible to determine at which point they may be dissociated into free MAs.

As discussed in chapter 1, antibodies in TBpos patients recognise CF and free MAs. The antibodies bound to the separate MA subclasses in different degrees. Antibodies therefore seem to recognise a polar part of MA that is modified by the mero-chain into a novel or cross-reactive antigen. Precisely how antibodies against MAs come about is not yet understood. Presentation of the lipids to T-cells on CD1b could play a key role, but how B-cells are in turn stimulated to produce antibodies is unknown. As mentioned in chapter 1, T-cells specifically recognised the polar moieties, while the alpha and mero-chains' interactions with CD1 were not sensitive to changes. Consequently, as far as our knowledge currently stretches, it is presumed that all subclasses of MAs are displayed by a universal CD1b protein. This implies that all subclasses have to fit the same mould and have to be able to access similar conformations with relative ease.

The pattern of antibody binding in TBneg patients brought the realisation that free MAs fold differently to those attached to trehalose in CF, since free MAs were recognised by TBneg patient antibodies, but CF was not. It was observed that anti-cholesterol antibodies in TBneg patients were able to recognise free MAs, suggesting that MAs assume a cholesteroid-fold. All humans have anti-cholesterol antibodies,

which may make serodiagnostic assays which make use of MA-antigens less sensitive due to cross-reaction of anti-cholesterol antibodies with MAs. The interference of anti-cholesterol antibodies can be removed by detecting low-affinity anti-MA antibody binding in such assays, as is measured in serodiagnosis using the evanescent field biosensor [99].

Cholesterol is central in the uptake of mycobacteria into host cells (section 2.1.1.1). Mce1A may interact with cholesterol in lipid rafts to advance entry into the host cell. Mce4A and Mce3A may anchor in the host cell membrane and transport cholesterol into the cell for access by the mycobacteria as nutrient supply. Gatfield and Pieters showed that mycobacteria have a much higher cholesterol-binding capacity than other microorganisms [30]. It has been shown in our group that this phenomenon may be due to an interaction between MAs in the cell wall with cholesterol since cholesterol accumulated onto MA-liposomes [78]. By attempting to measure this interaction quantitatively it was shown in this work that the interaction is exclusive to cholesterol concentrations in which the liposome phospholipids are in the *lo* phase. This is significant given that lipids in membrane rafts are in the *lo* phase and that rafts are rich in receptor proteins that are involved in the uptake of mycobacteria into host cells.

Direct evidence in favour of a structural and functional relatedness between MA and cholesterol was found in the discovery that AmB, specific for recognising cholesterol, also recognised free MAs [78]. If free MAs fold and pack into a cholesteroid-fold that is recognised by anti-cholesterol antibodies and binds to AmB, it will also be possible to test whether isolated subclasses fold into this cholesteroid-fold and whether CF loses the ability to be recognised by AmB. With molecular modeling, the most abundant structural folds of MAs in CF can be compared to those of free MAs to elucidate the molecular basis for the differences of antibody recognition observed experimentally.

This project has shown for the first time that the interaction between MA and cholesterol is dependent on the phase of the lipids in the liposomes. It also includes the first MD simulations and comparison between MA-subclasses without any restrictions on any atoms. The MD simulations have shown that all three MA

subclasses can freely assume W-, U- and Z-conformations *in vacuo*, without any packing effects or external pressure applied. Different trends in MA-folding were observed that occur as a result of the unique molecular composition and structure of each MA-subclass that was modeled.

α-MA folded slowly relative to other MA-types, but demonstrated itself to be the most flexible by being the only MA-type to assume all seven possible WUZ-conformations at least once. MeO-MA seemed unwilling to bend at its cis-cyclopropane group and its methyl ester did not show a great deal of difference although it had a higher number of WUZ-conformations. Keto-MA folded very differently to the other MA-types. It folded very quickly, which may be attributed to hydrogen-bonding between its keto-group and hydroxyl-groups at carbons 1 and 3. Keto-MA also seemed to start folding by drawing in its long alpha methylene chain first, while the other MAs preferred to fold by starting from the mero-chain on the opposite end of the molecule.

Current literature highlights the effects of MA fine-structure on biological activity, pathogenicity, virulence and cell wall permeability. Therefore, knowledge on the structure of MAs and how their conformations are dependent on their precise molecular composition becomes essential in order to exploit these properties in drug-design and in advancing our understanding of the disease. Molecular modeling is the versatile tool that allows the structure-function relationship of MAs to be resolved at the molecular level.

"Molecular knowledge of a bacterium and its host is essential for the control of a disease." Jose A. Ainsa [130].

References

1. WHO, Global Tuberculosis Control 2007, in *WHO Report 2007*. 2007, WHO.
2. Underwood, J.C.E., General and Systematic Pathology. 4th ed. 2004: Elsevier Limited. 182-184, 596-597, 336-338.
3. Tuberculosis Fact Sheet No. 104. 2002, WHO: Geneva.
4. Damm, D.D.; Allen, C.M.; Bouquot, J.E., Oral & Maxillofacial Pathology. 2nd ed. 2002: Saunders. 173-176.
5. Rang, H.P.; Dale, M.M.; Ritter, J.M.; Moore, P.K., Pharmacology. 5 ed. 2003, Loanhead: Elsevier Limited. 649-651.
6. Mukadi, Y.D.; Maher, D.; Harries, A., Tuberculosis case fatality rates in high HIV prevalence populations in sub-Saharan Africa. AIDS, 2001(15): p. 143-152.
7. Lawn, S.D.; Bekker, L.-G.; Miller, R.F., Immune reconstitution disease associated with mycobacterial infections in HIV-infected individuals receiving antiretrovirals. The Lancet Infectious Diseases, 2005. **5**: p. 361-373.
8. Wood, R., Challenges of TB diagnosis and treatment in South Africa: Roche Symposium, 3rd South African AIDS Conference, Durban, 5-8 June 2007. The Southern African Journal of HIV Medicine, 2007. **27**: p. 44-48.
9. Gandhi, N.R.; Moll, A.; Sturm, A.W.; Pawinski, R.; Govender, T.; Lalloo, U.; Zeller, K.; Andrews, J.; Friedland, G., Extensively drug-resistant tuberculosis as a cause of death in patients co-infected with tuberculosis and HIV in a rural area of South Africa. The Lancet, 2006(368): p. 1575-1580.
10. Schleicher, G.K.; Feldman, C.; Vermaak, Y.; Verschoor, J.A., Prevalence of Anti-mycolic Acid Antibodies in Patients with Pulmonary Tuberculosis Co-infected with HIV. Clinical Chemistry and Laboratory Medicine, 2002. **40**(9): p. 882-887.
11. Chimote, G.; Banerjee, R., Lung surfactant dysfunction in tuberculosis: Effect of mycobacterial tubercular lipids on dipalmitoylphosphatidylcholine surface activity. Colloids and Surfaces, 2005. **45**: p. 215-223.
12. Lambert, P.A., Cellular impermeability and uptake of biocides and antibiotics in Gram-positive bacteria and mycobacteria. Journal of Applied Microbiology Symposium Supplement, 2002. **92**: p. 46S-54S.
13. Sekanka, G.; Baird, M.; Minnikin, D.; Grooten, J., Mycolic acids for the control of tuberculosis. Expert Opin. Ther. Patents, 2007. **17**(3): p. 315-331.
14. Minnikin, D.E.; Kremer, L.; Dover, L.G.; Besra, G.S., The methyl-branched fortifications of *Mycobacterium tuberculosis.* Chemistry and Biology, 2002. **9**(5): p. 545-553.
15. Barry, C.E.; Lee, R.E.; Mdluli, K.; Sampson, A.E.; Schroeder, B.G.; Slayden, R.A.; Yuan, Y., Mycolic Acids: Structure, Biosynthesis and Physiological Functions. Progress in Lipid Research, 1998. **2/3**(37): p. 143-179.
16. Minnikin, D.E.; Goodfellow, M., Lipid Composition in the Classification and Identification of Acid-fast Bacteria. The Society for Applied Bacteriology Symposium Series, 1980(8): p. 189-256.
17. Villeneuve, M.; Kawai, M.; Kanashima, H.; Watanabe, M.; Minnikin, D.E.; Nakahara, H., Temperature dependence of the Langmuir monolayer packing of mycolic acids from Mycobacterium tuberculosis. Biochimica et Biophysica Acta, 2005. **1715**: p. 71-80.

18. Villeneuve, M.; Kawai, M.; Watanabe, M.; Aoyagi, Y.; Hitotsuyanagi, Y.; Takeya, K.; Gouda, H.; Hirono, S.; Minnikin, D.E.; Nakahara, H., Conformational behavior of oxygenated mycobacterial mycolic acids from *Mycobacterium bovis* BCG. Biochimica et Biophysica Acta, 2007. **1768**: p. 1717-1726.
19. Aderem, A.; Underhill, D.M., Mechanisms of phagocytosis in macrophages. Annual Reviews in Immunology, 1999. **17**: p. 593-623.
20. van Crevel, R.; Ottenhoff, T.H.M.; Van der Meer, J.W.M., Innate Immunity to *Mycobacterium tuberculosis*. Clinical Microbiology Reviews, 2002. **15**(2): p. 294-308.
21. Schlesinger, L.S., Macrophage Phagocytosis of Virulent but Not Attenuated Strains of *Mycobacterium tuberculosis* Is Mediated by Mannose Receptors in Addition to Complement Receptors. The Journal of Immunology, 1993. **150**(7): p. 2920-2930.
22. Schlesinger, L.S.; Bellinger-Kawahara, C.G.; Payne, N.R.; Horwitz, M.A., Phagocytosis of *Mycobacterium tuberculosis* is mediated by human monocyte complement derived receptors and complement component C3. The Journal of Immunology, 1990. **144**(7): p. 2771-2780.
23. Schorey, J.S.; Carroll, M.C.; Brown, E.J., A Macrophage Invasion Mechanism of Pathogenic Mycobacteria. Science, 1997. **277**: p. 1091-1093.
24. Cywes, C.; Godenir, N.L.; Hoppe, H.C.; Scholle, R.R.; Steyn, L.M.; Kirsch, R.E.; Ehlers, M.W., Nonopsonic Binding of *Mycobacterium tuberculosis* to Human Complement Receptor Type 3 Expressed in Chinese Hamster Ovary Cells. Infection and Immunity, 1996. **64**(12): p. 5373-5383.
25. Cywes, C.; Hoppe, H.C.; Daffé, M.; Ehlers, M.R.W., Nonopsonic Binding of *Mycobacterium tuberculosis* to Complement Receptor Type 3 Is Mediated by Capsular Polyscaccharides and Is Strain Dependent. Infection and Immunity, 1997. **65**(10): p. 4258-4266.
26. Schlesinger, L.S.; Hull, S.R.; Kaufman, T.M., Binding of the Terminal Mannosyl Units of Lipoarabinomannan from a Virulent Strain of *Mycobacterium tuberculosis* to Human Macrophages. The Journal of Immunology, 1994. **152**: p. 4070-4079.
27. Schlesinger, L.S.; Kaufman, T.M.; Iyer, S.; Hull, S.R.; Marchiando, L.K., Differences in Mannose Receptor-Mediated Uptake of Lipoarabinomannan from Virulent and Attenuated Strains of *Mycobacterium tuberculosis* by Human Macrophages. The Journal of Immunology, 1996. **157**: p. 4568-4575.
28. Ernst, J.D., Macrophage Receptors for *Mycobacterium tuberculosis*. Infection and Immunity, 1998. **66**(4): p. 1277-1281.
29. Amstrong, J.A.; Hart, P.D., Response of cultured macrophages to *Mycobacterium tuberculosis*, with observations on fusion of lysosomes with phagosomes. The Journal of Experimental Medicine, 1971. **134**(3): p. 713-740.
30. Gatfield, J.; Pieters, J., Essential Role for Cholesterol in Entry of Mycobacteria into Macrophages. Science, 2000. **288**: p. 1647-1650.
31. Peyron, P.; Bordier, C.; N'Diaye, E.-N.; Maridonneau-Parini, I., Nonopsonic Phagocytosis of *Mycobacterium kansasii* by Human Neutrophils Depends on Cholesterol and Is Mediated by CR3 Associated with Glycosylphosphatidylinositol-Anchored Proteins. The Journal of Immunology, 2000. **165**: p. 5186-5191.

References

32. Russell, D.G.; Mwandumba, H.C.; Rhoades, E.E., *Mycobacterium* and the coat of many lipids. The Journal of Cell Biology, 2002. **158**(3): p. 421-426.
33. Malik, Z.A.; Denning, G.M.; Kusner, D.J., Inhibition of Ca^{2+} Signaling by *Mycobacterium tuberculosis* Is Associated with Reduced Phagosome-Lysosome Fusion and Increased Survival within Human Macrophages. The Journal of Experimental Medicine, 2000. **191**(2): p. 287-302.
34. Malik, Z.A.; Iyer, S.S.; Kusner, D.J., *Mycobacterium tuberculosis* Phagosomes Exhibit Altered Calmodulin-Dependent Signal Transduction: Contribution to Inhibition of Phagosome-Lysosome Fusion and Intracellular Survival in Human Macrophages. The Journal of Immunology, 2001. **166**: p. 3392-3401.
35. Russell, D.G.; Dant, J.; Sturgill-Koszycki, S., *Mycobacterium avium*- and *Mycobacterium tuberculosis*- Containing Vacuoles Are Dynamic, Fusion-Competent Vesicles That Are Accessible to Glycosphingolipids from the Host Cell Plasmalemma. The Journal of Immunology, 1996. **156**: p. 4764-4773.
36. Schüller, S.; Neefjes, J.; Ottenhof, T.; Thole, J.; Young, D., Coronin is involved in uptake of *Mycobacterium bovis* BCG in human macrophages but not in phagosome maintenance. Cellular Microbiology, 2001. **3**(12): p. 785-793.
37. Ferrari, G.; Langen, H.; Naito, M.; Pieters, J., A Coat Protein on Phagosomes Involved in the Intracellular Survival of Mycobacteria. Cell, 1999. **97**: p. 435-447.
38. Tailleux, L.; Neyrolles, O.; Honoré-Bouakline, S.; Perret, E.; Sanchez, F.; Abastado, J.-P.; Lagrange, P.H.; Gluckman, J.C.; Rosenzwajg, M.; Herrmann, J.-L., Constrained Intracellular Survival of *Mycobacterium tuberculosis* in Human Dendritic Cells. The Journal of Immunology, 2003. **170**: p. 1939-1948.
39. Solomon, J.M.; Leung, G.S.; Isberg, R.R., Intracellular Replication of *Mycobacterium marinum* within *Dictyostelium discoideum*: Efficient Replication in the Absence of Host Coronin. Infection and Immunity, 2003. **71**(6): p. 3578-3586.
40. Guérin, I.; Chastellier, C., Pathogenic Mycobacteria Disrupt the Macrophage Actin Filament Network. Infection and Immunity, 2000. **68**(5): p. 2655-2662.
41. Malik, Z.A.; Thompson, C.R.; Hashimi, S.; Porter, B.; Iyer, S.S.; Kusner, D.J., Cutting Edge: *Mycobacterium tuberculosis* Blocks Ca^{2+} Signaling and Phagosome Maturation in Human Macrophages Via Specific Inhibition of Sphingosine Kinase. The Journal of Immunology, 2003. **170**: p. 2811-2815.
42. Spargo, B.J.; Crowe, L.M.; Ioneda, T.; Beaman, B.L.; Crowe, J.H., Cord factor (α,α-trehalose 6,6'-dimycolate) inhibits fusion between phospholipid vesicles. Proceedings of the National Academy of Science of the USA, 1991. **88**: p. 737-740.
43. Indrigo, J.; Hunter, R.L.; Actor, J.K., Cord facor trehalose 6,6'dimycolate (TDM) mediates trafficking events during mycobacterial infection of murine macrophages. Microbiology, 2003. **149**: p. 2049-2059.
44. Fratti, R.A.; Backer, J.M.; Gruenberg, J.; Corvera, S.; Deretic, V., Role of phosphatidylinositol 3-kinase and Rab5 effectors in phagosomal biogenesis and mycobacterial phagosome maturation arrest. The Journal of Cell Biology, 2001. **154**(3): p. 631-644.
45. Fratti, R.A.; Chua, J.; Vergen, I.; Deretic, V., *Mycobacterium tuberculosis* glycosylated phosphatidylinositol causes phagosome maturation arrest.

References

Proceedings of the National Academy of Science of the USA, 2003. **100**(9): p. 5437-5442.

46. Roitt, I.; Brostoff, J.; Male, D., Immunology. 6th ed. 2002: Mosby. 65-141.
47. Kolter, T.; Winau, F.; Svhaible, U.E.; Leippe, M.; Sandhoff, K., Lipid-binding Proteins in Membrane Digestion, Antigen Presentation, and Antimicrobial Defense. Journal of Biological Chemistry, 2005. **280**(50): p. 41125-41128.
48. Winau, F.; Schwierzeck, V.; Hurwitz, R.; Remmel, N.; Sieling, P.A.; Modlin, R.; Porcelli, S.A.; Brinkmann, V.; Sugita, M.; Sandhoff, K.; Kaufmann, S.H.E.; Schaible, U.E., Sapocin C is required for lipid presentation by human CD1b. Nature Immunology, 2004. **5**(2): p. 169-174.
49. Munford, R.S.; Sheppard, P.O.; O'Hara, P.J., Saposin-like proteins (SAPLIP) carry out diverse functions on a common backbone structure. Journal of Lipid Research, 1995. **36**: p. 1653-1663.
50. Ho, M.W.; O'Brien, J.S., Gaucher's Disease: Deficiency of 'Acid' β-Glucosidase and Reconstitution of Enzyme Activity *In Vitro*. Proceedings of the National Academy of Science of the USA, 1971. **68**(11): p. 2810-2813.
51. Kolter, T.; Sandhoff, K., Principles of Lysosomal Membrane Digestion: Stimulation of Sphingolipid Degradaion by Sphingolipid Activator Proteins and Anionic Lysosomal Lipids. Annual Review of Cell and Developmental Biology, 2005. **21**: p. 81-103.
52. Moody, D.B.; Reinhold, B.B.; Reinhold, V.N.; Besra, G.S.; Porcelli, S.A., Uptake and processing of glycosylated mycolates for presentation to CD1b-restricted cells. Immunology Letters, 1999. **65**: p. 85-91.
53. Moody, D.B.; Reinhold, B.B.; Guy, M.R.; Beckman, E.M.; Frederique, D.E.; Furlong, S.T.; Ye, S.; Reinhold, V.N.; Sieling, P.A.; Modlin, R.L.; Besra, G.S.; Porcelli, S.A., Structural Requirements for Glycolipid Antigen Recognition by CD1b-Restricted T Cells. Science, 1997. **278**: p. 183-186.
54. Moody, D.B.; Guy, M.R.; Grant, E.; Cheng, T.; Brenner, M.B.; Besra, G.S.; Porcelli, S.A., CD1b-mediated T Cell Recognition of a Glycolipid Antigen Generated from Mycobacterial Lipid and Host Carbohydrate during Infection. Journal of Experimental Medicine, 2000. **192**(7): p. 965-976.
55. Grant, E.P.; Beckman, E.M.; Behar, S.M.; Degano, M.; Frederique, D.; Besra, G.S.; Wilson, I.A.; Porcelli, S.A.; Furlong, S.T.; Brenner, M.B., Fine Specificity of TCR Complementarity-Determining Region Residues and Lipid Antigen Hydrophilic Moieties in the Recognition of a CD1-Lipid Complex. The Journal of Immunology, 2002. **168**: p. 3933-3940.
56. Batuwangala, T.; Sheperd, D.; Gadola, S.D.; Gibson, K.J.C.; Zaccai, N.R.; Fersht, A.R.; Besra, G.S.; Vincenzo, C.; Jones, E.Y., The Crystal Structure of Human CD1b with Bound Bacterial Glycolipid. The Journal of Immunology, 2004. **172**: p. 2382-2388.
57. Gumperz, J.E., The Ins and Outs of CD1 Molecules: Bringing Lipids Under Immunological Surveillance. Traffic, 2006. **7**: p. 2-13.
58. Lang, M.L.; Glatman-Freedman, A., Do CD1-Restricted T Cells Contribute to Antibody-Mediated Immunity against *Mycobacterium tuberculosis*? Infection and Immunity, 2006. **74**(2): p. 803-809.
59. Pan, J.; Fujiwara, N.; Oka, S.; Maekura, R.; Ogura, T.; Yano, I., Anti-Cord Factor (Trehalose 6,6' -Dimycolate) IgG Antibody in Tuberculosis Patients Recognizes Mycolic Acid Subclasses. Microbiology and Immunology, 1999. **43**(9): p. 863-869.

60. Deysel, M.S.M., Structure-function relationships of mycolic acids in tuberculosis. PhD thesis, in *Faculty of Natural and Agricultural Sciences Department of Biochemistry*. 2007, University of Pretoria: South Africa.
61. Biro, A.; Cervenak, L.; Balogh, A.; Lorincz, A.; Uray, K.; Horvath, A.; Romics, L.; Matko, J.; Fust, G.; Laszlo, G., Novel anti-cholesterol monoclonal immunoglobulin G antibodies as probes and potential modulators of membrane raft-dependent immune functions. Journal of Lipid Research, 2007. **48**: p. 19-29.
62. Fust, G.; Beck, Z.; Banhegui, D.; Kocsis, J.; Biro, A.; Prohaszka, Z., Antibodies against heat shock proteins and cholesterol in HIV infection. Molecular Immunology, 2005. **42**: p. 79-85.
63. Benadie, Y., Amphotericin B as a mycolic acid specific targeting agent in tuberculosis. MSc thesis, in *Faculty of Natural and Agricultural Sciences Department of Biochemistry*. 2007, University of Pretoria: South Africa.
64. Van Regenmortel, M.H.V., The Antigenicity of Tobacco Mosaic Virus. Philosophical Transactions: Biological Sciences, 1999. **354**(1383): p. 559-568.
65. Enshell-Seijffers, D.; Denisov, D.; Groisman, B.; Smelyanski, L.; Meyuhas, R.; Gross, G.; Denisova, G.; Gershoni, J.M., The Mapping and Reconstitution of a Conformatinal Discontinuous B-cell Epitope of HIV-1. Journal of Molecular Biology, 2003. **334**: p. 87-101.
66. Chakrabarti, P.; Janin, J., Dissecting Protein-Protein Recognition Sites. Proteins: Structure, Function, and Genetics, 2002. **47**: p. 334-343.
67. Lo Conte, L.; Chothia, C.; Janin, J., The Atomic Structure of Protein-Protein Recognition Sites. Journal of Molecular Biology, 1999. **285**: p. 2177-2198.
68. Hasegawa, T.; Nishijo, J.; Watanabe, M.; Funayama, K.; Imae, T., Conformational characterization of α-mycolic acid in a monolayer film by the Langmuir-Blodgett technique and atomic force microscopy. Langmuir, 2000. **16**: p. 7325-7330.
69. Hasegawa, T.; Amino, S.; Kitamura, S.; Matsumoto, L.; Katada, S.; Nishijo, J., Study of the molecular conformation of alpha- and keto-mycolic acid monolayers by the Langmuir-Blodgett technique and fourier transform infrared reflection-absorption spectroscopy. Langmuir, 2003. **19**: p. 105-109.
70. Brigl, M.; Brener, M.B., CD1: Antigen Presentation and T Cell Function. Annual Reviews in Immunology, 2004. **2004**(22): p. 817-890.
71. Korf, J.; Stoltz, A.; Verschoor, J.; De Baetselier, P.; Grooten, J., The *Mycobacterium tuberculosis* cell wall component mycolic acid elicits pathogen-associated host innate immune responses. European Journal of Immunology, 2005. **35**: p. 890-900.
72. Korf, J.E.; Pynaert, G.; Tournoy, K.; Boonefaes, T.; Van Oosterhout, A.; Ginneberge, D.; Haegeman, A.; Verschoor, J.A.; De Baetselier, P.; Grooten, J., Macrophage Reprogramming by Mycolic Acid Promotes a Tolerogenic Response in Experimental Asthma. American Journal of Respiratory and Critical Care Medicine, 2006. **174**: p. 152-160.
73. Hasegawa, T., Structural Analysis of Biological Aliphatic Compounds Using Surface-Enhanced Fourier Transform Raman Spectroscopy. Biopolymers, 2003. **73**: p. 457-462.
74. Hasegawa, T.; Leblanc, R.M., Aggregation properties of mycolic acid molecules in monolayer films: a comparative study of compounds from various acid-fast bacterial species. Biochimica et Biophysica Acta, 2003. **1617**: p. 89-95.

References

75. Hong, X.; Hopfinger, A.J., Construction, molecular modeling, and simulation of Mycobacterium tuberculosis cell walls. Biomacromolecules, 2004. **5**: p. 1052-1065.
76. Dmitriev, B.A.; Ehlers, S.; Rietschel, E.T.; Brennan, P.J., Molecular mechanics of the mycobacterial cell wall: From horizontal layers to vertical scaffolds. International Journal of Medical Microbiology, 2000. **290**: p. 251-258.
77. Hong, X.; Hopfinger, A.J., Molecular modeling and simulation of Mycobacterium tuberculosis cell wall permeability. Biomacromolecules, 2004. **5**: p. 1066-1077.
78. Benadie, Y.; Deysel, M.; Siko, D.G.R.; Roberts, V.V.; Van Wyngaardt, S.; Thanyani, S.T.; Sekanka, G.; Ten Bokum, A.M.C.; Collett, L.A.; Grooten, J.; Baird, M.S.; Verschoor, J.A., Cholesteroid nature of free mycolic acids from *M. tuberculosis*. Chemistry and Physics of Lipids, 2008. **152**: p. 95-103.
79. Arruda, S.; Bomfim, G.; Knights, R.; Huima-Byron, T.; Riley, L.W., Cloning of an M. tuberculosis DNA fragment associated with entry and survival inside cells. Science, 1993. **261**(5127): p. 1454-1457.
80. Chitale, S.; Ehrt, S.; Kawamura, I.; Fujimura, T.; Shimono, N.; Anand, N.; Lu, S.; Cohen-Gould, L.; Riley, L.W., Recombinant *Mycobacterium tuberculosis* protein associated with mammalian cell entry. Cellular Microbiology, 2001. **3**(4): p. 247-254.
81. Mitra, D.; Saha, B.; Das, S.; Wiker, H.G.; Das, A.K., Correlating sequential homology of Mce1A, Mce2A, Mce3A and Mce4A with their possible functions in mammalian cell entry of *Mycobacterium tuberculosis* performing homology modeling. Tuberculosis, 2005. **85**: p. 337-345.
82. Fivaz, M.; Vilbois, F.; Thurnheer, S.; Pasquali, C.; Abrami, L.; Bickel, P.E.; Parton, R.G.; Vand Der Goot, F.G., Differential sorting and fate of endocytosed GPI-anchored proteins. The EMBO Journal, 2002. **21**(15): p. 3989-4000.
83. Rosenberger, C.M.; Brumell, J.H.; Finlay, B.B., Microbial pathogenesis: Lipid rafts as pathogen portals. Current Biology, 2000. **10**(22): p. 823-825.
84. Joshi, S.M.; Pandey, A.K.; Capite, N.; Fortune, S.M.; Rubin, E.J.; Sassetti, C.M., Characterization of mycobacterial virulence genes through genetic interaction mapping. Proceedings of the National Academy of Science of the USA, 2006. **103**(31): p. 11760-11765.
85. Lodish, H.; Berk, A.; Matsudaira, P.; JKaiser, C.A.; Krieger, M.; Scott, M.P.; Zipursky, S.L.; Darnell, J., Molecular Cell Biology. 5th ed. 2004: W. H. Freeman and Company. 147-192.
86. Nester, E.W.; Anderson, D.G.; Roberts, C.E.; Pearsall, N.N.; Nester, M.T., Microbiology, A Human Perspective. 2001, New York: McGraw-Hill. 569-570.
87. Cardona, P.-J., RUTI: A new chance to shorten the treatment of latent tuberculosis infection. Tuberculosis, 2006. **86**: p. 273-289.
88. Van der Geize, R.; Yam, K.; Heuser, T.; Wilbrink, M.H.; Hara, H.; Anderton, M.C.; Sim, E.; Dijkhuizen, L.; Davies, J.E.; Mohn, W.W.; Eltis, L.D., A gene cluster encoding cholesterol catabolism in a soil actinomycete provides insight into *Mycobacterium tuberculosis* survival in macrophages. Proceedings of the National Academy of Science of the USA, 2007. **104**(6): p. 1947-1952.
89. Thewalt, J.L.; Bloom, M., Phosphatidylcholine: cholesterol phase diagrams. Biophysical Journal, 1992. **63**: p. 1176-1181.

References

90. Ipsen, J.H.; Mouritsen, O.G.; Zuckermann, M.J., Theory of thermal anomalies in the specific heat of lipid bilayers containing cholesterol. Biophysical Journal, 1989. **56**: p. 661-667.
91. Dietrich, C.; Bagatolli, L.A.; Volovyk, Z.N.; Thompson, N.L.; Levi, M.; Jacobson, K.; Gratton, E., Lipid Raft Reconstitution in Model Membranes. Biophysical Journal, 2001. **80**(3): p. 1417-1428.
92. Gaus, K.; Gratton, E.; Kable, E.P.W.; Jones, A.S.; Gelissen, I.; Kritharides, L.; Jessup, W., Visualizing lipid structure and raft domains in living cells with two-photon microscopy. Proceedings of the National Academy of Science of the USA, 2003. **100**(26): p. 15554-15559.
93. Brown, D.A.; London, E., Functions of lipid rafts in biological membranes. Annual Review of Cell and Developmental Biology, 1998. **14**: p. 111-136.
94. Healy, D.A.; Hayes, C.J.; Leonard, P.; McKenna, L.; O'Kennedy, R., Biosensor developments: application to prostrate-specific antigen detection. TRENDS in Biotechnology, 2007. **25**(3): p. 125-131.
95. Ecochemie, Autolab ESPRIT / SPRINGLE User Manual. Version 4.2, 2006: p. 119-124.
96. Echochemie, Autolab Application note: Autolab SPR systems with a flexible range in refractive index to broaden the application range. (www.ecochemie.nl).
97. Miller, J.N.; Miller, J.C., Statistics and Chemometrics for Analytical Chemistry. 2005, Great Britain: Pearson Education Limited. 39-58, 251-262.
98. Siko, D.G.R., Mycobacterial mycolic acids as immunoregulatory lipid antigens in the resistance to tuberculosis. PhD thesis, in *Faculty of Natural and Agricultural Sciences Department of Biochemistry*. 2002, University of Pretoria: South Africa.
99. Thanyani, S.T.; Roberts, V.; Siko, D.G.R.; Vrey, P.; Verschoor, J.A., A novel application of affinity biosensor technology to detect antibodies to mycolic acid in tuberculosis patients. Journal of Immunological Methods, 2008. **332**(1-2): p. 61-72.
100. Chanda, J.; Bandyopadhyay, S., Perturbation of phospholipid bilayer properties by ethanol at a high concentration. Langmuir, 2006. **22**: p. 3775-3781.
101. Chanda, J.; Chakraborty, S.; Bandyopadhyay, S., Sensitivity of hydrogen bond lifetime dynamics to the presence of ethanol at the interface of a phospholipid bilayer. Journal of Physical Chemistry, 2006. **110**: p. 3791-3797.
102. Ly, H.V.; Block, D.E.; Longo, M.L., Interfacial tension effect of ethanol on lipid bilayer rigidity, stability, and area/molecule: a micropipet aspiration approach. Langmuir, 2002. **18**: p. 8988-8995.
103. Vrey, P.J., Lipid ligand - protein receptor interactions characterised by a resonant mirror biosensor. MSc thesis, in *Faculty of Natural and Agricultural Sciences Department of Biochemistry*. 2004, University of Pretoria: South Africa.
104. Jensen, F., Introduction to Computational Chemistry. 1999, England: John Wiley and Sons Ltd.
105. Hehre, W.J., A Guide to Molecular Mechanics and Quantum Chemical Calculations. 2003, United States of America: Wavefunction Inc. 1-60, 393-407.
106. Grant, H.; Richards, W.G., Computational Chemistry. 1995: The Bath Press. 1-61.

107. Stekel, D., Microarray Bioinformatics. 2003: Cambridge University Press. 151-180.
108. BioDiscovery, GeneSight User's Manual. 2002.
109. Sturn, A., Cluster Analysis for Large Scale Gene Expression Studies. 2000: Rockville, Maryland, USA.
110. M. J. Frisch, G.W.T., H. B. Schlegel, G. E. Scuseria,; M. A. Robb, J.R.C., J. A. Montgomery, Jr., T. Vreven,; K. N. Kudin, J.C.B., J. M. Millam, S. S. Iyengar, J. Tomasi,; V. Barone, B.M., M. Cossi, G. Scalmani, N. Rega,; G. A. Petersson, H.N., M. Hada, M. Ehara, K. Toyota,; R. Fukuda, J.H., M. Ishida, T. Nakajima, Y. Honda, O. Kitao,; H. Nakai, M.K., X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross,; C. Adamo, J.J., R. Gomperts, R. E. Stratmann, O. Yazyev,; A. J. Austin, R.C., C. Pomelli, J. W. Ochterski, P. Y. Ayala,; K. Morokuma, G.A.V., P. Salvador, J. J. Dannenberg,; V. G. Zakrzewski, S.D., A. D. Daniels, M. C. Strain,; O. Farkas, D.K.M., A. D. Rabuck, K. Raghavachari,; J. B. Foresman, J.V.O., Q. Cui, A. G. Baboul, S. Clifford,; J. Cioslowski, B.B.S., G. Liu, A. Liashenko, P. Piskorz,; I. Komaromi, R.L.M., D. J. Fox, T. Keith, M. A. Al-Laham,; C. Y. Peng, A.N., M. Challacombe, P. M. W. Gill,; B. Johnson, W.C., M. W. Wong, C. Gonzalez, and J. A. Pople., Gaussian. 2003, Gaussian, Inc.: Pittsburgh PA.
111. Al Dulayymi, J.R.; Baird, M.S.; Roberts, E., The synthesis of a single enantiomer of a major α-mycolic acid of *M. tuberculosis.* Tetrahedron, 2005. **61**(2005): p. 11939-11951.
112. Al Dulayymi, J.R.; M.S., B.; E., R.; Deysel, M.; Verschoor, J., The first synthesis of single enantiomers of the major methoxymycolic acid of *Mycobacterium tuberculosis.* Tetrahedron, 2007. **63**: p. 2571-2592.
113. Koza, G.; Baird, M.S., The first synthesis of single enantiomer of a ketomycolic acid. Tetrahedron Letters, 2007. **48**: p. 2165-2169.
114. Chass, G.A.; Sahai, M.A.; Law, J.M.S.; Lovas, S.; Farkas, Ö.; Perczel, A.; Rivail, J.; Csizmadia, I.G., Toward a Computed Peptide Structure Database: The Role of a Universal Atomic Numbering System of Amino Acids in Peptides and Internal Hierarchy of Database. International Journal of Quantum Chemistry, 2002. **90**: p. 933-368.
115. Al Dulayymi, J.R.; Baird, M.S.; Roberts, E.; Minnikin, D.E., The synthesis of single enantiomers of meromycolic acids from mycobacterial wax esters. Tetrahedron, 2006. **62**: p. 11867-11880.
116. Asselineau, C.; Tocanne, G., Stereochemistry of mycolic acids. Bulletin Societe Chimique de France, 1970. **4**: p. 1455-1459.
117. Takayama, K.; Wang, C.; Besra, G.S., Pathway to Synthesis and Processing of Mycolic Acids in *Mycobacterium tuberculosis*. Clinical Microbiology Reviews, 2005. **18**(1): p. 81-101.
118. Glickman, M.S.; Cox, J.S.; Jacobs Jr, W.R., A Novel Mycolic Acid Cyclopropane Synthetase Is Required for Cording, Persistence, and Virulence of *Mycobacterium tuberculosis.* Molecular Cell, 2000. **5**: p. 717-727.
119. Rao, V.; Gao, F.; Chen, B.; Jacobs Jr, W.R.; Glickman, M.S., *Trans*-cyclopropanation of mycolic acids on trehalose dimycolate suppresses *Mycobacterium tuberculosis*-induced inflammation and virulence. The Journal of Clinical Investigation, 2006. **116**(6): p. 1660-1667.
120. Dubnau, E.; Chan, J.; Raynaud, C.; Mohan, V.P.; Lanéelle, M.; Yu, K.; Quémard, A.; Smith, I.; Daffé, M., Oxygenated mycolic acids are necessary

for virulence of *Mycobacterium tuberculosis* in mice. Molecular Microbiology, 2000. **36**(3): p. 630-637.

121. Yuan, Y.; Zhu, Y.; Crane, D.D.; Barry III, C.E., The effect of oxygenated mycolic acid composition on cell wall function and macrophage growth in *Mycobacterium tuberculosis.* Molecular Microbiology, 1998. **29**(6): p. 1449-1458.
122. Veyron-Churlet, R.; Bigot, S.; Guerrini, O.; Verdoux, S.; Malaga, W.; Daffé, M.; Zerbib, D., The Biosynthesis of Mycolic Acids in *Mycobacterium tuberculosis* Relies on Multiple Specialized Elongation Complexes Interconnected by Specific Protein-Protein Interactions. Journal of Molecular Biology, 2005. **353**: p. 847-859.
123. Belisle, J.T.; Vissa, V.D.; Sievert, T.; Takayama, K.; Brennan, P.J.; Besra, G.S., Role of the major antigen of *Mycobacterium tuberculosis* in cell wall biogenesis. Science, 1997. **276**: p. 1420-1422.
124. Puech, V.; Bayan, N.; Salim, K.; Leblon, G.; Daffé, M., Characterization of the *in vivo* acceptors of the mycoloyl residues transferred by the corynebacterial PS1 and the related mycobacterial antigens 85. Molecular Microbiology, 2000. **35**(5): p. 1026-1041.
125. Jackson, M.; Raynaud, C.; Lanéelle, M.; Guilhot, C.; Laurent-Winter, C.; Ensergueix, D.; Gicquel, B.; Daffé, M., Inactivation of the antigen 85C gene profoundly affects mycolate content and alters the permeability of the *Mycobacterium tuberculosis* cell envelope. Molecular Microbiology, 1999. **31**(5): p. 1573-1587.
126. Puech, V.; Guilhot, C.; Perez, E.; Tropis, M.; Armitage, L.Y.; Gicquel, B.; Daffé, M., Evidence for a partial redundancy of the fibronectin-binding proteins for the transfer of mycoloyl residues onto the cell wall arabinogalactan termini of *Mycobacterium tuberculosis.* Molecular Microbiology, 2002. **44**(4): p. 1109-1122.
127. Armitage, L.Y.; Jagannath, C.; Wanger, A.R.; Norris, S.J., Disruption of the Genes Encoding Antigen 85A an Antigen 85B of *Mycobacterium tuberculosis* H37Rv: Effect on Growth in Culture and in Macrophages. Infection and Immunity, 2000. **68**(2): p. 767-778.
128. Ronning, D.R.; Klabunde, T.; Besra, G.S.; Vissa, V.D.; Belisle, J.T.; Sacchettini, J.C., Crystal structure of the secreted form of antigen 85C reveals potential targets for mycobacterial drugs and vaccines. Nature Structural Biology, 2000. **7**(2): p. 141-146.
129. Fujita, Y.; Okamoto, Y.; Uenishi, Y.; Sunagawa, M.; Uchiyama, T.; Yano, I., Molecular and supra-molecular structure related differences in toxicity and granulomatogenic activity of mycobacterial cord factor in mice. Microbial Pathogenesis, 2007. **43**: p. 10-21.
130. Ainsa, J.A.; Martín, C.; Gicquel, B., Molecular approaches to tuberculosis. Molecular Microbiology, 2001. **42**(2): p. 561-570.

Appendices

Appendix A

Trans, gauche and eclipsed conformations

The acyl chains of MAs can adopt several conformations that are determined by the dihedral angles in the chain.

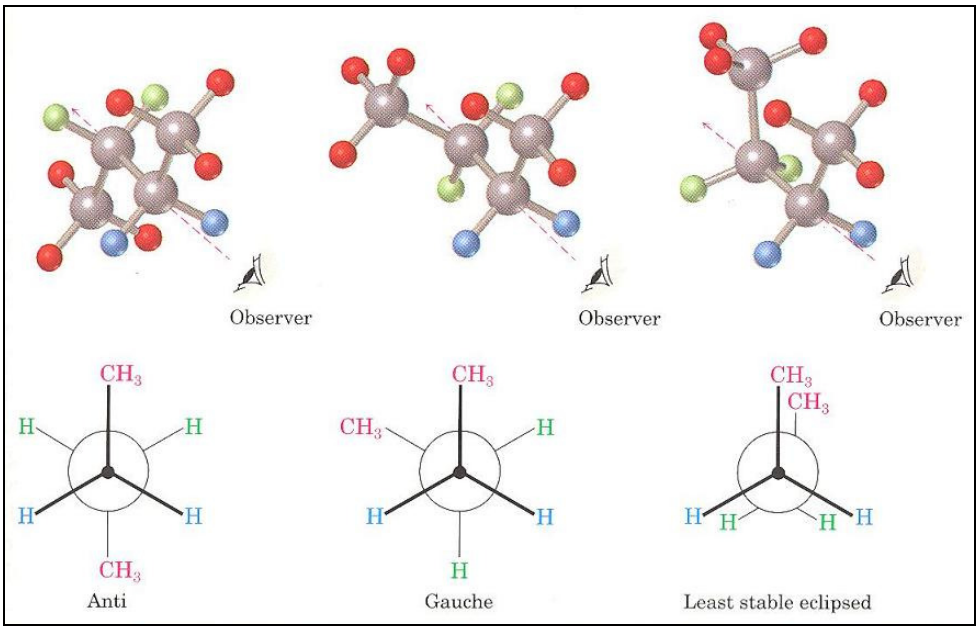

Figure A.1 *Different conformations for butane at different rotations about the C2-C3 bond* [†].

[†] McMurry, J. Organic Chemistry. 5th ed. 2000: Brooks/Cole. 116-117.

When considering four atoms connected in the order A-X-Y-B, the torsion angle or dihedral angle is the angle between the A-X and B-Y bonds, viewed along the X-Y bond. This angle is shown by the 'observer' in figure A.1 where an example for butane is used. To relate butane to the long acyl-chains of MAs, the two methyl-groups can be substituted with a CH_2R and CH_2R' group, where R and R' represent the rest of the acyl chain.

As shown for the butane example[†], the anti-conformation has the methyl-groups separated maximally by 180 degrees. This is the lowest energy conformation for butane. The trans-conformation is also stable and energetically favourable in long alkyl chains, forming an ordered zig-zag pattern in the chain. When the dihedral angle between the methyl groups of butane are decreased as in the gauche-conformation with a dihedral angle of 60 degrees, there is increasingly more conformational strain in the molecule due to steric hindrance when the methyl-groups are in close proximity. This conformation is higher in energy than the trans-conformation and less stable. The least energetically favourable conformation occurs when the two methyl-groups in butane eclipse, as shown in figure A.1.

Appendix B

Results for semi-empirical calculations on cyclopropanes as methylene chain lengths are increased

Table B.1 *The bond lengths and angles measured as the chain lengths were increased in a symmetrical way for cis- and trans-cyclopropane.*

Total Nr carbons in chains	cis/ trans	Internal Angle (degrees)			Bond length (Angstroms)			Outer angle (degrees)		Energy (kcal/mol)	Negative frequency? Y/N
		A_1	A_2	A_3	L_1	L_2	L_3	A_4	A_5		
0		60	60	60	1.499	1.499	1.499			16.269	N
1		59.7	60.1	60.1	1.504	1.504	1.497	119.7		8.553	N
2	cis	60.5	59.8	59.8	1.513	1.503	1.503	120.7	120.7	1.416	N
4	cis	60.5	59.8	59.8	1.513	1.502	1.502	121.0	121.0	-9.363	N
6	cis	60.5	59.8	59.8	1.513	1.502	1.502	121.0	121.0	-20.275	N
8	cis	60.5	59.8	59.8	1.513	1.502	1.502	121.0	121.0	-31.125	N
10	cis	60.5	59.8	59.8	1.513	1.502	1.502	121.0	121.0	-41.977	N
12	cis	60.5	59.8	59.8	1.513	1.502	1.502	121.0	121.0	-52.823	N
14	cis	60.5	59.8	59.8	1.513	1.502	1.502	121.0	121.0	-63.668	N
16	cis	60.5	59.8	59.8	1.513	1.502	1.502	121.0	121.0	-74.512	N
18	cis	60.5	59.8	59.8	1.513	1.502	1.502	121.0	121.0	-85.357	N
20	cis	60.5	59.8	59.8	1.513	1.502	1.502	121.0	121.0	-96.200	N
22	cis	60.5	59.8	59.8	1.513	1.502	1.502	121.0	121.0	-107.044	N
24	cis	60.5	59.8	59.8	1.513	1.502	1.502	121.0	121.0	-117.888	N
26	cis	60.5	59.8	59.8	1.513	1.502	1.502	121.0	121.0	-128.731	N
28	cis	60.5	59.8	59.8	1.513	1.502	1.502	121.0	121.0	-139.575	N
30	cis	60.5	59.8	59.8	1.513	1.502	1.502	121.0	121.0	-150.418	N
32	cis	60.5	59.8	59.8	1.513	1.502	1.502	121.0	121.0	-161.261	N
2	trans	60.3	59.8	59.8	1.510	1.503	1.503	119.6	119.6	0.857	N
4	trans	60.3	59.8	59.8	1.510	1.502	1.502	119.6	119.6	-9.546	N
6	trans	60.4	59.8	59.8	1.510	1.502	1.502	119.5	119.5	-20.534	N
8	trans	60.4	59.8	59.8	1.510	1.502	1.502	119.5	119.5	-31.378	N
10	trans	60.4	59.8	59.8	1.510	1.502	1.502	119.5	119.5	-42.228	N
12	trans	60.4	59.8	59.8	1.510	1.502	1.502	119.5	119.5	-53.073	N
14	trans	60.4	59.8	59.8	1.510	1.502	1.502	119.5	119.5	-63.917	N
16	trans	60.4	59.8	59.8	1.510	1.502	1.502	119.5	119.5	-74.760	N
18	trans	60.4	59.8	59.8	1.510	1.502	1.502	119.5	119.5	-85.604	N
20	trans										Y
22	trans										Y
24	trans										Y
26	trans										Y
28	trans										Y
30	trans										Y
32	trans										Y

Table B.2 *The bond lengths and angles measured as the chain lengths were increased in a symmetrical way for R- and S- diastereomers of α-methyl trans-cyclopropane.*

Total Nr carbons in chains	S/R	Internal Angle (degrees)			Bond length (Angstroms)			Outer angle (degrees)		Energy (kcal/mol)	Negative frequency? Y/N
		A_1	A_2	A_3	L_1	L_2	L_3	A_4	A_5		
6	S	60.3	59.8	59.8	1.510	1.502	1.502	119.4	119.6	-25.489	N
8	S	60.3	59.8	59.8	1.510	1.502	1.502	119.3	119.5	-36.537	N
10	S	60.3	59.8	59.8	1.510	1.502	1.502	119.3	119.5	-47.385	N
12	S	60.3	59.8	59.8	1.510	1.502	1.502	119.3	119.5	-58.233	N
14	S	60.3	59.8	59.8	1.510	1.502	1.502	119.3	119.5	-69.076	N
16	S	60.3	59.8	59.8	1.510	1.502	1.502	119.2	119.5	-79.92	N
18	S	60.3	59.8	59.8	1.510	1.502	1.502	119.3	119.5	-90.766	N
20	S	60.3	59.8	59.8	1.510	1.502	1.502	119.2	119.5	-101.606	N
22	S	60.4	59.8	59.8	1.510	1.502	1.502	119.1	119.4	-112.393	N
24	S	60.3	59.8	59.8	1.510	1.502	1.502	119.2	119.5	-123.292	N
26	S	60.3	59.8	59.8	1.510	1.502	1.502	119.2	119.5	-134.135	N
28	S	60.3	59.8	59.8	1.510	1.502	1.502	119.2	119.5	-144.978	N
30	S	60.3	59.8	59.8	1.510	1.502	1.502	119.2	119.5	-155.82	N
32	S	60.3	59.8	59.8	1.510	1.502	1.502	119.2	119.5	-166.659	N
6	R	60.3	59.8	59.8	1.510	1.502	1.502	121.7	119.2	-24.013	N
8	R	60.3	59.8	59.8	1.510	1.502	1.502	120.8	119.9	-35.353	N
10	R	60.3	59.8	59.8	1.510	1.502	1.502	120.7	119.9	-46.201	N
12	R	60.3	59.8	59.8	1.510	1.502	1.502	120.7	119.9	-57.049	N
14	R	60.3	59.8	59.8	1.510	1.502	1.502	120.7	119.9	-67.895	N
16	R	60.3	59.8	59.8	1.510	1.502	1.502	120.7	119.9	-78.738	N
18	R	60.3	59.8	59.8	1.510	1.502	1.502	120.7	119.9	-89.581	N
20	R	60.3	59.8	59.8	1.510	1.502	1.502	120.7	119.9	-100.424	N
22	R	60.3	59.8	59.8	1.510	1.502	1.502	120.7	119.9	-111.268	N
24	R	60.3	59.8	59.8	1.510	1.502	1.502	120.7	119.9	-122.109	N
26	R	60.3	59.8	59.8	1.510	1.502	1.502	120.7	119.9	-132.952	N
28	R	60.3	59.8	59.8	1.510	1.502	1.502	120.7	119.9	-143.794	N
30	R	60.3	59.8	59.8	1.510	1.502	1.502	121.6	119.2	-154.139	N
32	R	60.3	59.8	59.8	1.510	1.502	1.502	121.6	119.2	-164.981	N

Appendix C

Potential Energy Surface data and energy calculations

Relative Energies: 194.0424 Hartrees (lowest energy value) added to each energy value.

Energy (kcal/Mol) = [Energy (Hartrees)] . (627.5095)

Energy (kJ/Mol) = [Energy (kcal/Mol)] . (4.184)

D14	D15	Energy Hartrees (AU)	Relative Energies kcal/Mol	kJ/Mol
0	0	-194.04	2.55	10.68
30	0	-194.04	3.99	16.69
60	0	-194.04	2.58	10.80
90	0	-194.04	1.42	5.96
120	0	-194.04	2.55	10.68
150	0	-194.04	3.99	16.69
180	0	-194.04	2.58	10.80
210	0	-194.04	1.42	5.96
240	0	-194.04	2.55	10.68
270	0	-194.04	3.99	16.69
300	0	-194.04	2.58	10.80
330	0	-194.04	1.42	5.96
360	0	-194.04	2.55	10.68
0	30	-194.04	3.30	13.80
30	30	-194.04	4.61	19.31
60	30	-194.04	2.99	12.51
90	30	-194.04	1.87	7.84
120	30	-194.04	3.30	13.80
150	30	-194.04	4.61	19.31
180	30	-194.04	2.99	12.51
210	30	-194.04	1.87	7.84
240	30	-194.04	3.30	13.80
270	30	-194.04	4.61	19.31
300	30	-194.04	2.99	12.51
330	30	-194.04	1.87	7.84
360	30	-194.04	3.30	13.80
0	60	-194.04	1.62	6.77
30	60	-194.04	2.67	11.17
60	60	-194.04	1.30	5.42
90	60	-194.04	0.41	1.73
120	60	-194.04	1.62	6.77
150	60	-194.04	2.67	11.17
180	60	-194.04	1.30	5.42
210	60	-194.04	0.41	1.73
240	60	-194.04	1.62	6.77
270	60	-194.04	2.67	11.17
300	60	-194.04	1.30	5.42
330	60	-194.04	0.41	1.73
360	60	-194.04	1.62	6.77

Continued…

D14	D15	Energy Hartrees (AU)	Relative Energies kcal/Mol	kJ/Mol
0	90	-194.04	1.01	4.23
30	90	-194.04	2.16	9.05
60	90	-194.04	0.96	4.02
90	90	-194.04	0.00	0.00
120	90	-194.04	1.01	4.23
150	90	-194.04	2.16	9.05
180	90	-194.04	0.96	4.02
210	90	-194.04	0.00	0.00
240	90	-194.04	1.01	4.23
270	90	-194.04	2.16	9.05
300	90	-194.04	0.96	4.02
330	90	-194.04	0.00	0.00
360	90	-194.04	1.01	4.23
0	120	-194.04	2.55	10.68
30	120	-194.04	3.99	16.69
60	120	-194.04	2.58	10.80
90	120	-194.04	1.42	5.96
120	120	-194.04	2.55	10.68
150	120	-194.04	3.99	16.69
180	120	-194.04	2.58	10.80
210	120	-194.04	1.42	5.96
240	120	-194.04	2.55	10.68
270	120	-194.04	3.99	16.69
300	120	-194.04	2.58	10.80
330	120	-194.04	1.42	5.96
360	120	-194.04	2.55	10.68
0	150	-194.04	3.30	13.80
30	150	-194.04	4.61	19.31
60	150	-194.04	2.99	12.51
90	150	-194.04	1.87	7.84
120	150	-194.04	3.30	13.80
150	150	-194.04	4.61	19.31
180	150	-194.04	2.99	12.51
210	150	-194.04	1.87	7.84
240	150	-194.04	3.30	13.80
270	150	-194.04	4.61	19.31
300	150	-194.04	2.99	12.51
330	150	-194.04	1.87	7.84
360	150	-194.04	3.30	13.80

Continued…

D14	D15	Energy Hartrees (AU)	Relative Energies kcal/Mol	kJ/Mol
0	180	-194.04	1.62	6.77
30	180	-194.04	2.67	11.17
60	180	-194.04	1.30	5.42
90	180	-194.04	0.41	1.73
120	180	-194.04	1.62	6.77
150	180	-194.04	2.67	11.17
180	180	-194.04	1.30	5.42
210	180	-194.04	0.41	1.73
240	180	-194.04	1.62	6.77
270	180	-194.04	2.67	11.17
300	180	-194.04	1.30	5.42
330	180	-194.04	0.41	1.73
360	180	-194.04	1.62	6.77
0	210	-194.04	1.01	4.23
30	210	-194.04	2.16	9.05
60	210	-194.04	0.96	4.02
90	210	-194.04	0.00	0.00
120	210	-194.04	1.01	4.23
150	210	-194.04	2.16	9.05
180	210	-194.04	0.96	4.02
210	210	-194.04	0.00	0.00
240	210	-194.04	1.01	4.23
270	210	-194.04	2.16	9.05
300	210	-194.04	0.96	4.02
330	210	-194.04	0.00	0.00
360	210	-194.04	1.01	4.23
0	240	-194.04	2.55	10.68
30	240	-194.04	3.99	16.69
60	240	-194.04	2.58	10.80
90	240	-194.04	1.42	5.96
120	240	-194.04	2.55	10.68
150	240	-194.04	3.99	16.69
180	240	-194.04	2.58	10.80
210	240	-194.04	1.42	5.96
240	240	-194.04	2.55	10.68
270	240	-194.04	3.99	16.69
300	240	-194.04	2.58	10.80
330	240	-194.04	1.42	5.96
360	240	-194.04	2.55	10.68

Continued…

D14	D15	Energy Hartrees (AU)	Relative Energies kcal/Mol	kJ/Mol
0	270	-194.04	3.30	13.80
30	270	-194.04	4.61	19.31
60	270	-194.04	2.99	12.51
90	270	-194.04	1.87	7.84
120	270	-194.04	3.30	13.80
150	270	-194.04	4.61	19.31
180	270	-194.04	2.99	12.51
210	270	-194.04	1.87	7.84
240	270	-194.04	3.30	13.80
270	270	-194.04	4.61	19.31
300	270	-194.04	2.99	12.51
330	270	-194.04	1.87	7.84
360	270	-194.04	3.30	13.80
0	300	-194.04	1.62	6.77
30	300	-194.04	2.67	11.17
60	300	-194.04	1.30	5.42
90	300	-194.04	0.41	1.73
120	300	-194.04	1.62	6.77
150	300	-194.04	2.67	11.17
180	300	-194.04	1.30	5.42
210	300	-194.04	0.41	1.73
240	300	-194.04	1.62	6.77
270	300	-194.04	2.67	11.17
300	300	-194.04	1.30	5.42
330	300	-194.04	0.41	1.73
360	300	-194.04	1.62	6.77
0	330	-194.04	1.01	4.23
30	330	-194.04	2.16	9.05
60	330	-194.04	0.96	4.02
90	330	-194.04	0.00	0.00
120	330	-194.04	1.01	4.23
150	330	-194.04	2.16	9.05
180	330	-194.04	0.96	4.02
210	330	-194.04	0.00	0.00
240	330	-194.04	1.01	4.23
270	330	-194.04	2.16	9.05
300	330	-194.04	0.96	4.02
330	330	-194.04	0.00	0.00
360	330	-194.04	1.01	4.23

Continued…

D14	D15	Energy Hartrees (AU)	Relative Energies kcal/Mol	kJ/Mol
0	360	-194.04	2.55	10.68
30	360	-194.04	3.99	16.69
60	360	-194.04	2.58	10.80
90	360	-194.04	1.42	5.96
120	360	-194.04	2.55	10.68
150	360	-194.04	3.99	16.69
180	360	-194.04	2.58	10.80
210	360	-194.04	1.42	5.96
240	360	-194.04	2.55	10.68
270	360	-194.04	3.99	16.69
300	360	-194.04	2.58	10.80
330	360	-194.04	1.42	5.96
360	360	-194.04	2.55	10.68

Appendix D

SOM-plots

For each plot:

- Each line represents the length (Å, y-axis) of each of the eight selected distances (ab, bc, cd, de, ac, ae, ce, bd, x-axis) for one frame.
- The number of frames is labelled as the number of genes (indicated in brackets next to the cluster number) by default as for microarray data that is normally used by this program.

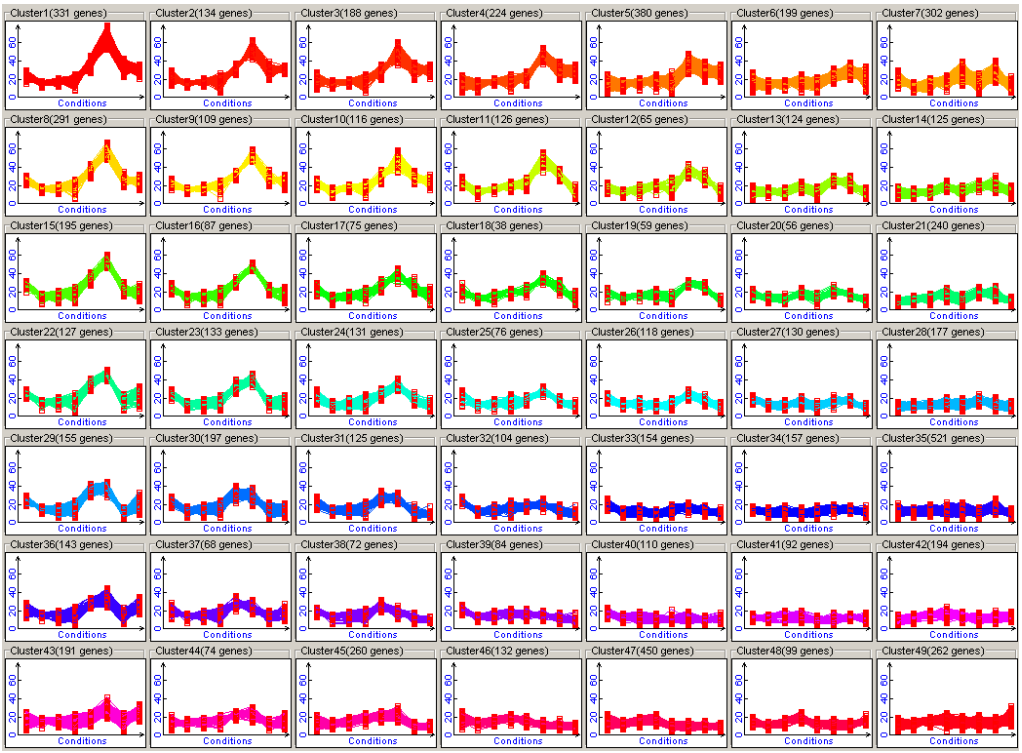

Figure D.1 *SOM plot of all 8000 frames for α-MA.*

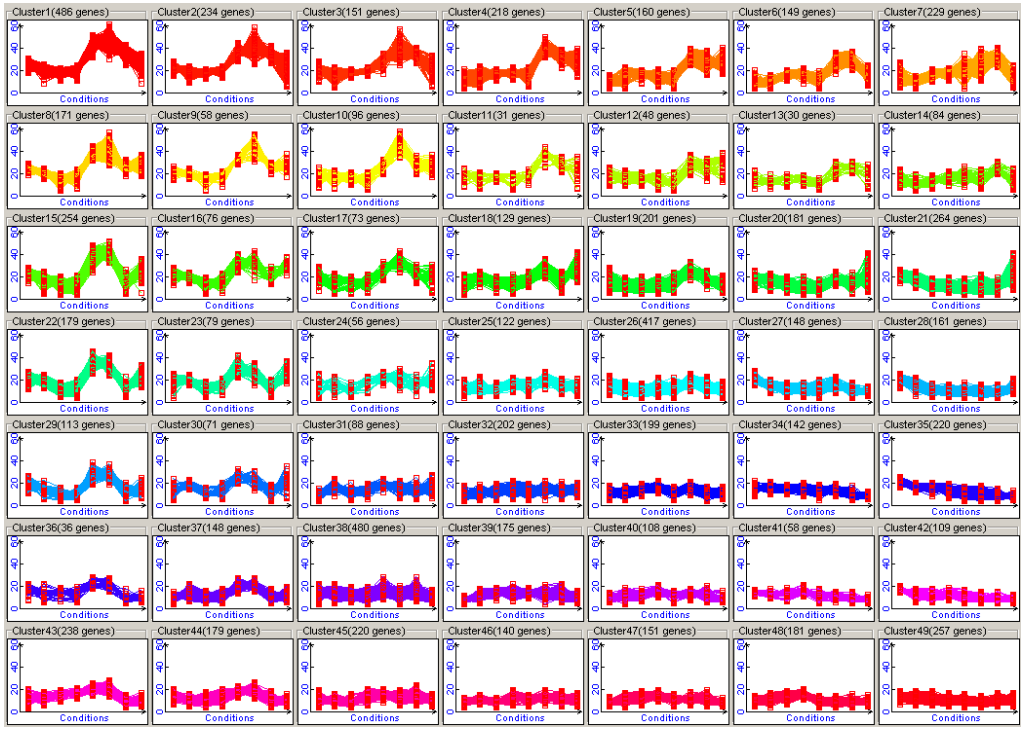

Figure D.2 *SOM plot of all 8000 frames for MeO-MA.*

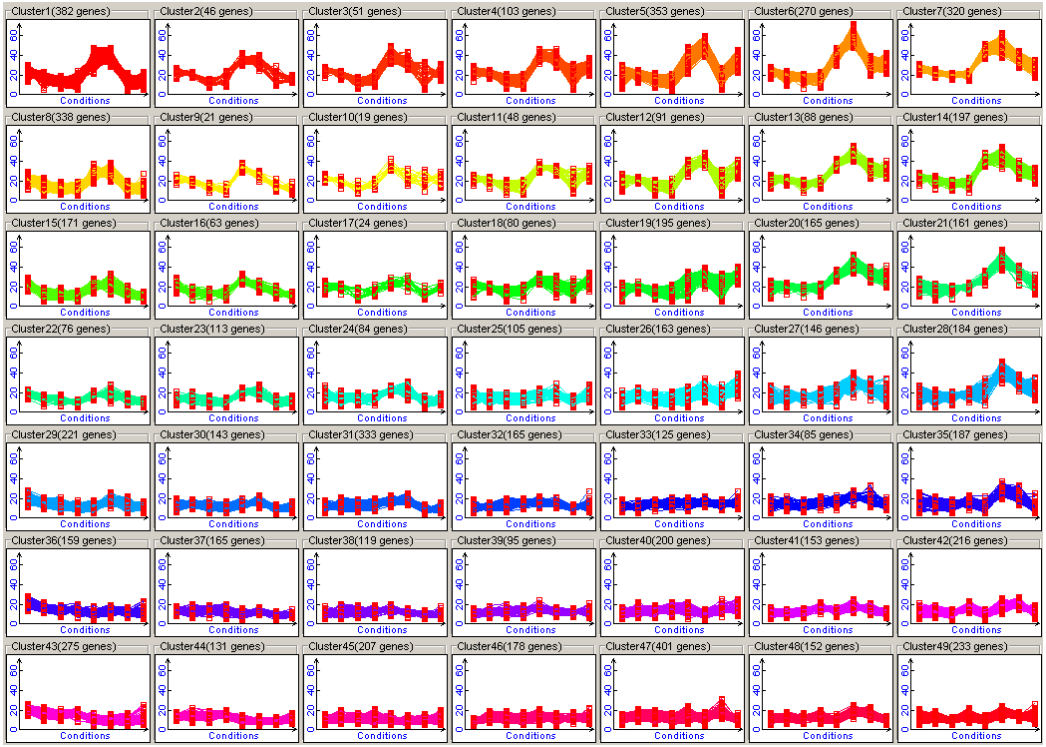

Figure D.3 *SOM plot of all 8000 frames for MeO-ME.*

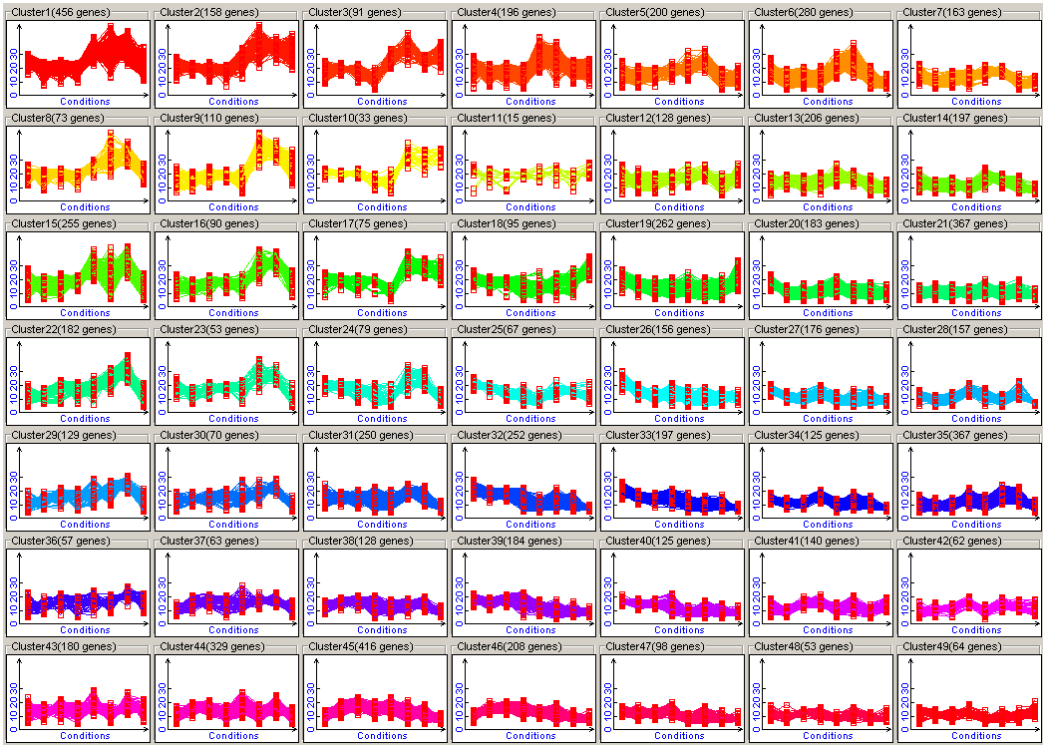

Figure D.4 *SOM plot of all 8000 frames for keto-MA.*

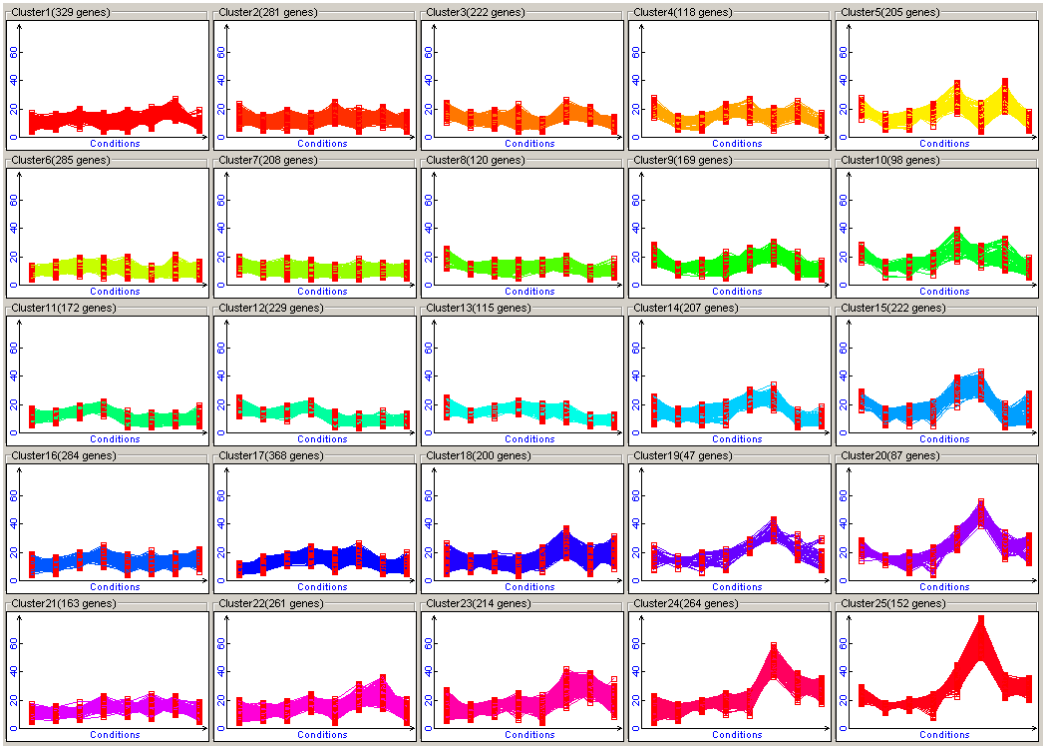

Figure D.5 *SOM plot of 5020 frames (the last 251 frames for each replicate) for α-MA.*

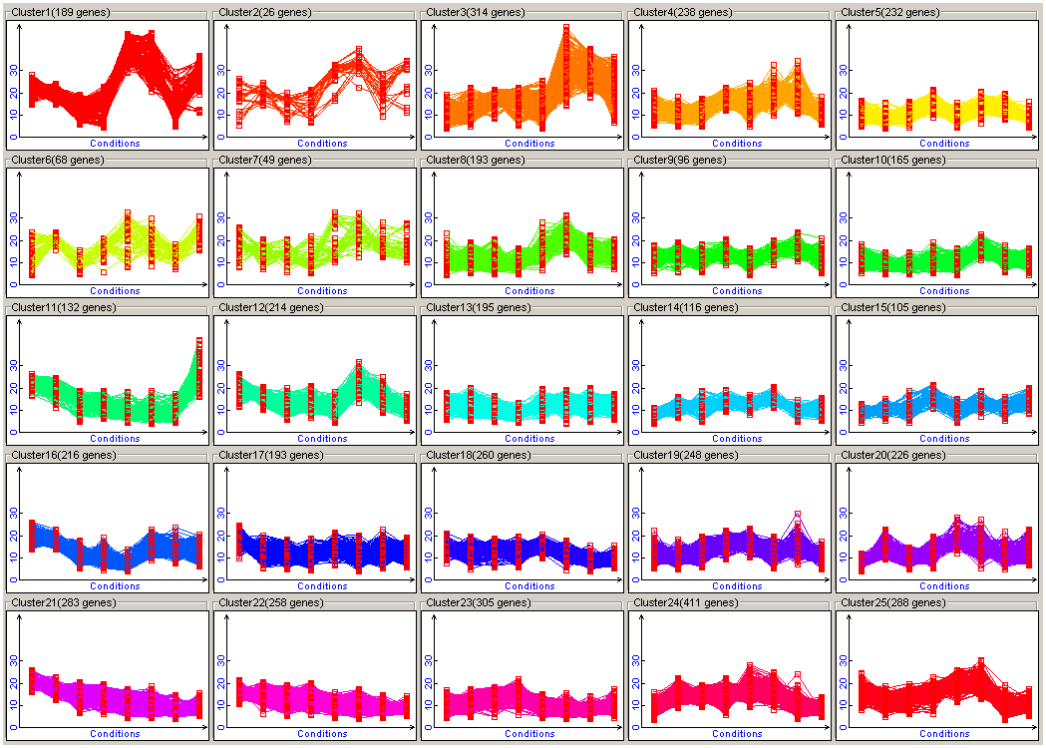

Figure D.6 *SOM plot of 5020 frames (the last 251 frames for each replicate) for MeO-MA.*

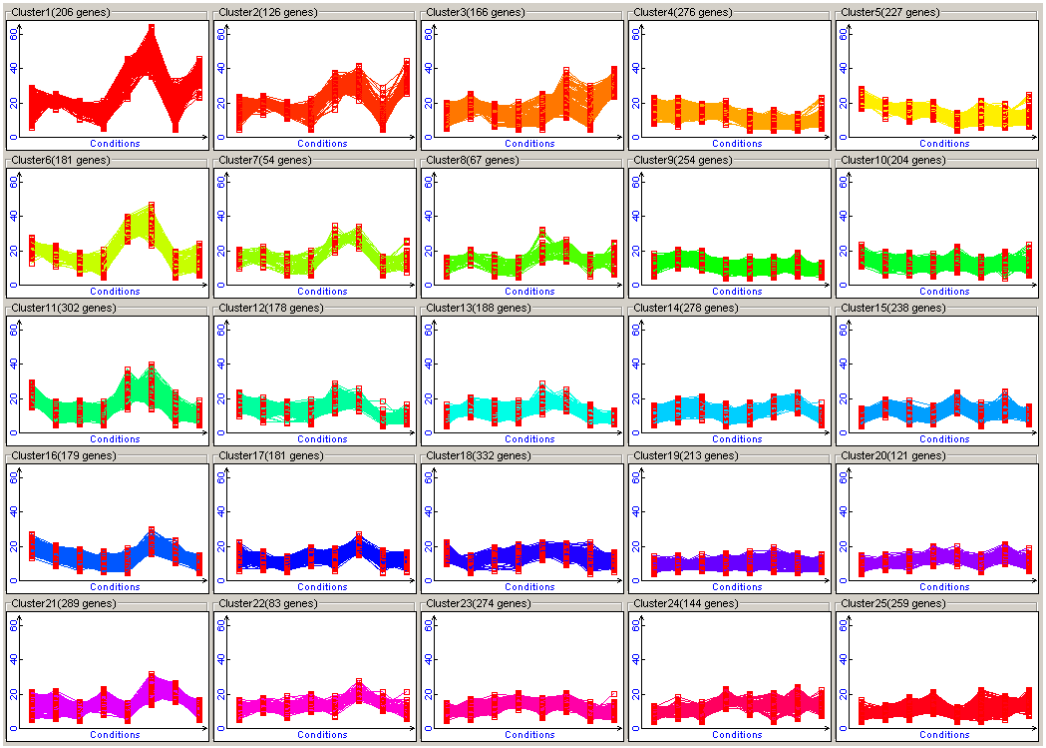

Figure D.7 *SOM plot of 5020 frames (the last 251 frames for each replicate) for MeO-ME.*

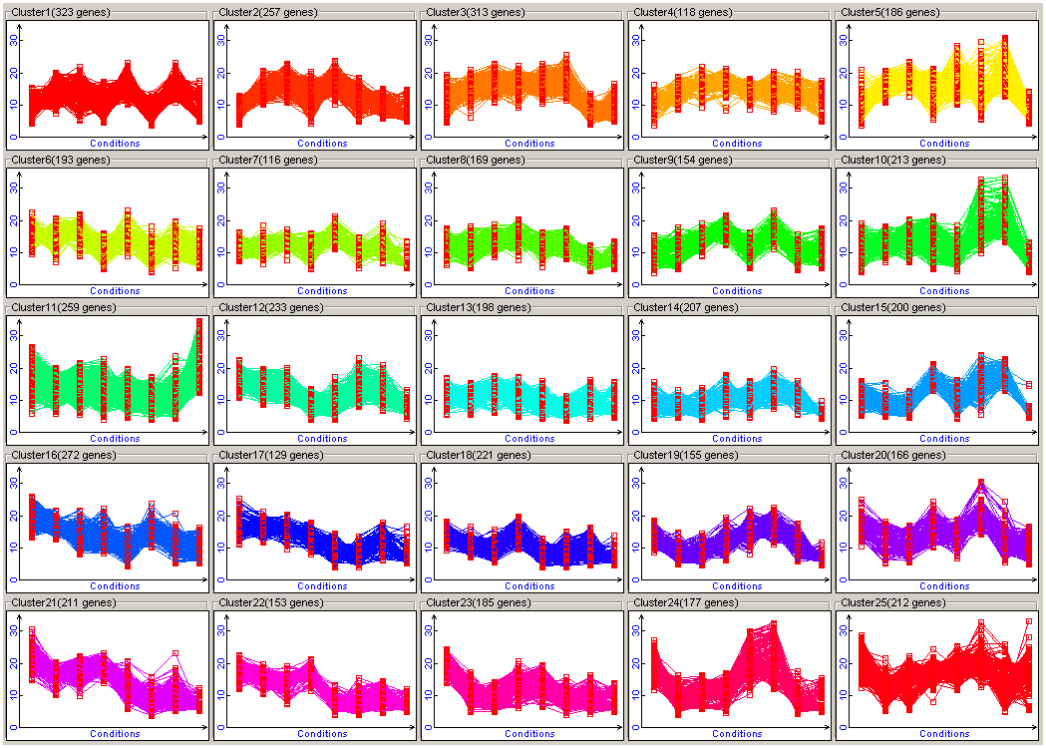

Figure D.8 *SOM plot of 5020 frames (the last 251 frames for each replicate) for keto-MA.*

Appendix E

Example of overlap in SOM-cluster of WUZ-groups

Figure E.1 shows an example from α-MA where two different WUZ-structures were grouped into the same SOM-cluster. The overlap was also the very minimal of one aZ frame grouped together with 10 W frames in cluster 15 (figure E.1). This is the only example where any overlap occurred in all of the 20 replicates for α-MA. In the remaining MA-types, keto-MA and MeO-ME also had only one such occurrence in all of their replicates involving only 2 overlapping frames. (The original SOM-report and WUZ-data for keto-MA and MeO-MA can be found in Supplementary Data\SOM_WUZ_groups\Keto\Keto_17.xls and Supplementary Data\SOM_WUZ_groups\m_meth\m_meth_8.xls).

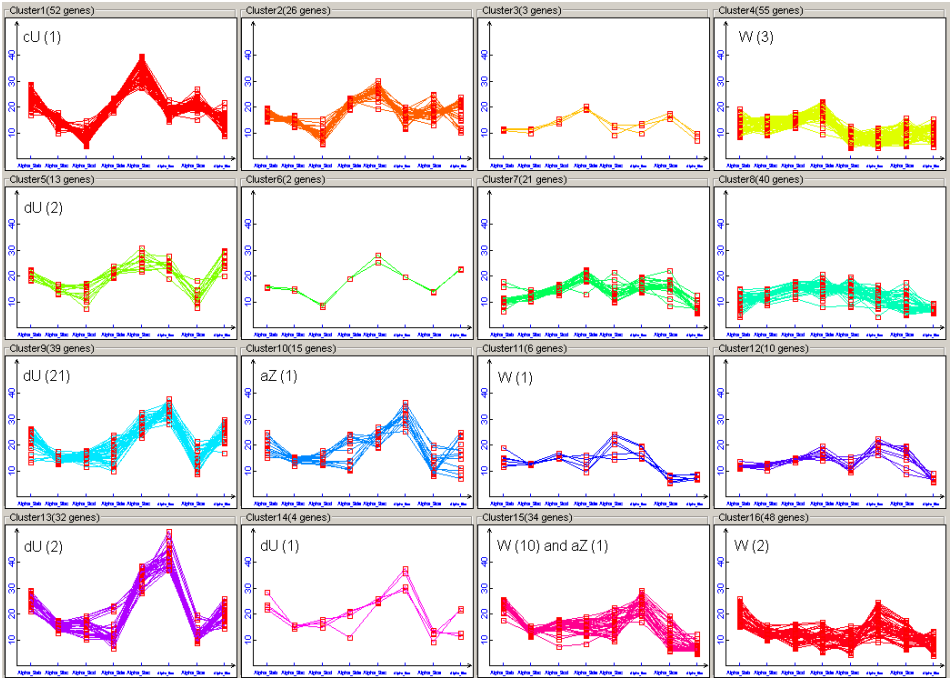

Figure E.1 *Self Organising Map of one representative α-MA MD simulation showing the unlikely event of an SOM-cluster that overlaps with two WUZ-folds (cluster 15). Each line represents one frame. The number of frames is labelled as the number of genes (indicated in brackets next to the cluster number) by default as for microarray data that is normally used by this program. The annotation of the WUZ-analysis is shown in the top left-hand corner of each cluster, indicating the conformation and the number of frames with that conformation. The comparison of the original SOM-report and WUZ-data can be found in Supplementary Data\SOM_WUZ_groups\Alpha\Alpha-9.xls.*

As can be seen in figure E.2, the two folds are very similar, which is to be expected from two consecutive frames. In (A) the fold was of aZ-type, whereas in (B) it was classed as a W-fold. The aZ in (A) does in fact resemble a W-fold more than an aZ-fold. This is due to the two-dimensionality of the distances that are used here to evaluate the types of folds. The ac distance is only slightly larger than 20 Å (21.2 Å) in (A) which means it is likely to look similar to the W-fold in the next frame (B), which has an ac distance just smaller than 20 Å (18.38 Å). Similarly, ae is only slightly larger than 25 Å (25.96 Å) in (A) and just smaller than 25 Å (24.56 Å) in (B), refer to parameters, Table 3.1. However, the factor that contributes most to this definition of the aZ-fold appearing so similar to the W-fold, is that the direction of the ab chain is upward, as in a W-fold. If this ac distance was 21.2 Å but in the downwards direction, it would give the expected aZ-form, although it still is on the border-line of the parameters. It is very seldom that the combination of all eight parameters will result in such ambiguity. As mentioned before, this occurrence was extremely rare. Nevertheless, it highlighted the necessity of studying the three-dimensional clustering and PCA analyses of MAs with their coordinates (3D) as an improvement on the 2D-distances.

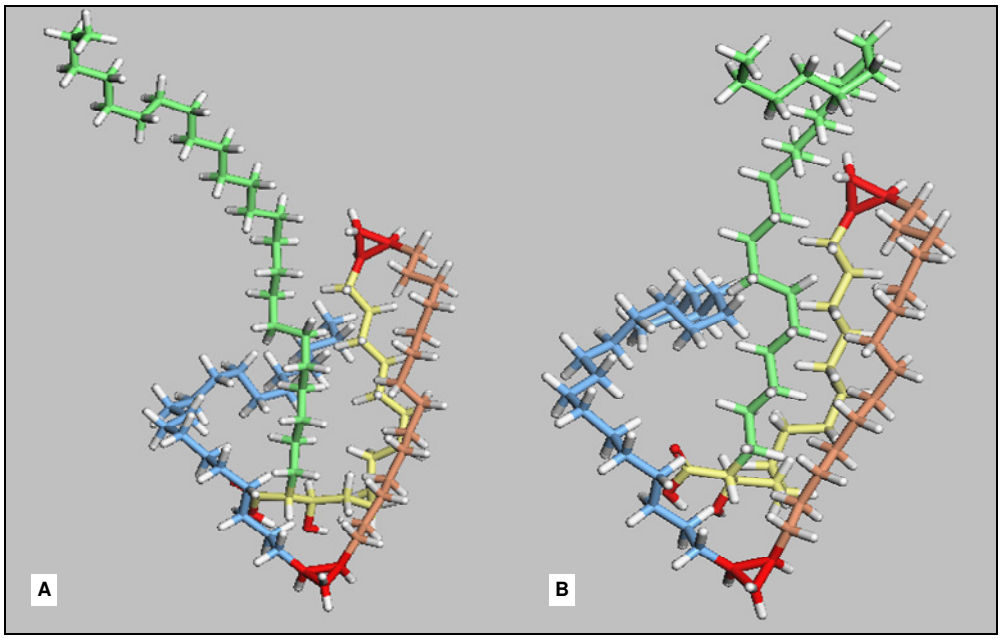

Figure E.2 *The two α-MA structures classed as different folds, but grouped into the same cluster (figure E.1). (A) was classed as an aZ-fold and it was followed by (B) in the next frame that was classed as a W-fold.*