DOCTOR OF PHILOSOPHY

Diverse approaches to modelling conformations of mycolic acids from Myobactrium tuberculosis

Groenewald, Wilma

Award date:
2014

Awarding institution:
Bangor University

Diverse approaches to modelling conformations of mycolic acids from *Mycobacterium tuberculosis*

PRIFYSGOL
BANGOR
UNIVERSITY

By Wilma Groenewald

A thesis submitted for the degree of
Doctor of Philosophy
in Chemistry
at Bangor University, Wales
May 2014

Acknowledgements

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

To my supervisor, Dr Croft, thank you for your continued patience and support - especially in the last stretch through late Skype sessions and numerous emails. I appreciate the exposure that I gained during this project under your guidance through attending numerous conferences and visiting other research groups and establishing valuable collaborations.

I would like to thank collaborators on this work, Prof. Dr S.J. Marrink and Dr G. Chass for the opportunity of learning from their expertise and applying it to this work. I especially enjoyed my time at Groningen University in the group of Prof. Marrink – a great environment for learning Gromacs in a friendly and helpful atmosphere.

Others that have been instrumental in the completion of this work include Dr Monica Bulacu for support in the initial development of the coarse grained model and Dr Tsjerk Wassenaar for long discussions on different ways of analysing MA conformations as well as the Insane lipid layer building script, in addition to numerous others at Groningen University; James Maskery, Jurgens de Bruin, Dr M. Rozhok and Kenneth Robins on various scripting aspects of the molecular dynamics and quantum mechanics work.

I would like to thank Professor Baird and Dr Gwenin for their patience and practical support at work while I was writing up this work.

Financial support for this work was provided by Bangor University and HPC Europa 2.

The start of this work was not merely the start of a new project for me. As well as the chance to address a research question that intrigues me, and an opportunity to learn new skills, it was also a new start in my personal life. I would like to thank my friends at BCC for their love, support and abundant prayers from settling in Bangor as a newly wed, through to the completion of this work.

A big thank you goes to my family for their support – although they live far away, I could always depend on them because they know me so well.

Last but not least, my better half, Jakobus; your love, prayers and practical support have been constant and have seen me through the highs and lows. Dankie vir *alles* my skat!

Summary

Tuberculosis (TB) has plagued the human race for centuries. Despite being largely a curable disease, over 3500 people die from the disease each day. *Mycobacterium tuberculosis*, that causes TB, is particularly resilient. This is due, in part, to its lipid-rich cell wall, containing mycolic acids (MAs) as a major component. MAs are long fatty acids with functional groups of precise stereochemistry. The MA composition in the cell wall has been shown to affect properties such as cell wall permeability and virulence. The functional group stereochemistry is key in the interaction with different components of the host immune system. It has been shown that MA structure determines its conformational preference. In this work, a selection of diverse computational methods is employed to study different aspects of MA conformation. The precise stereochemical effects of the cyclopropane functional groups in model compounds with short alkyl chains were investigated, through quantum chemistry. Whole MAs were simulated using molecular dynamics in the gas phase and in polar and nonpolar solvents, in order to investigate folding patterns of MAs as a function of their structure together with the external environment. Lastly, a coarse grained (CG) model of an Alpha-MA was produced and used to simulate a MA monolayer. The monolayer is a good representation of the cell wall packing of MAs and could be correlated with existing monolayer experimental data. Together, the results presented here confirm the functional groups as folding points in MAs. Larger systems that are closer to the natural MA environment are essential as they can be compared with experimental data and will be useful as a tool in combating TB. However, the integrity of the larger CG system relies on the flow of information from the smaller, more accurate computational methods through to the CG model, and hence these approaches should be continued in parallel in future work.

Table of Contents

List of Manuscripts

The work for this thesis is presented in the following draft manuscripts that are referred to by their Roman Numerals I – III.

I **Systematic Conformational Search for Low Energy Conformers of Cyclopropane Units From Mycolic Acids**
Wilma Groenewald, Gregory A. Chass and Anna K. Croft

II **Computational Investigation of Solvent and Structural Effects on Mycolic Acid Conformation**
Wilma Groenewald, Anna K. Croft

III **Monolayer Simulation of Coarse Grained Alpha-Mycolic Acid**
Wilma Groenewald, Anna K. Croft and Siewert-Jan Marrink

Chapter 1: Introduction

"The grim story of the past should not be forgotten; it should serve as a spur to a united effort to try to give the *coup de grâce* to this invader from which the world has suffered so severely and for so long."
H. D. Chalke[1]

Tuberculosis

For thousands of years, tuberculosis (TB) can be traced as being a part of human history; molecular traces of TB have been found in human remains that are around 9000 years old.[2] In the 18th and 19th centuries TB became particularly widespread. Claiming countless lives, TB gained the title of being "the captain among these men of death."[3]

Initial insight into TB was marked by the discovery that TB is infectious and that it is caused by a microorganism. These findings were followed by the development of the skin test for testing for TB infection and later, vaccine development. In the 20th century the first drugs against TB were discovered, followed by a sharp decline in TB incidence. Eradication of the disease seemed tangible. However, this goal has not yet been realised. Due to several factors, an explosive increase in TB occurred, starting in the early 1980s. Human immunodeficiency virus (HIV), rendering individuals more susceptible to developing TB disease, has had a major role to play.[4] The emergence of drug-resistant strains has also degraded the control over the disease.[5] In 1993 TB was declared a global emergency by the World Health Organisation.[6]

Twenty years later, TB is still not under control, and HIV coinfection and drug resistance remain major challenges. Around 1.3 million deaths were caused by TB in the year 2012.[7] In addition to multidrug resistant (MDR) strains, resistant to isoniazid and rifampicin, extensively drug resistant strains of TB (XDR-TB) have been recognised since 2006.[8] XDR-TB is resistant to all currently known TB drugs, and has been reported in over 92 countries worldwide.[7] Due to the lack of

culture facilities, drug susceptibility testing is not performed widely and, hence, drug resistant cases are overlooked. Only 5 % of new cases and 9 % of previously treated cases are tested for drug susceptibility.[7] It is estimated that only a third of drug resistant cases are currently detected.[7] Around 10 % of MDR cases are expected to be XDR-TB. The prognosis for MDR-TB is not good, with only around 50 % of cases successfully treated.[7] In addition, HIV changes the presentation of TB by less pulmonary cavity formation and lowered sputum bacillary load, making it hard to detect TB in coinfected patients.[9] A quick and accurate diagnosis is even more urgent in these cases where the life expectancy is very short and mortality rates very high.[9]

Currently, sputum smear microscopy is the most widely used method for the diagnosis of TB, but it has low sensitivity.[7] Sputum culture remains the golden standard, but takes several weeks. Testing for drug susceptibility increases the time before starting treatment even further.[7] The latest diagnostic for TB implemented and endorsed by the World Health Organisation, the Xpert MTB/RIF molecular diagnostic, is able to diagnose TB and test for rifampicin resistance within 2 hours.[7]

Tuberculosis (TB) is caused by the bacterium *Mycobacterium tuberculosis* (*M. tb*). Small droplets of mycobacteria in the air that are produced by sneezing or coughing of an infected individual, and subsequently inhaled by other people, spread the disease.[10] Once inhaled, the mycobacteria can be completely eliminated from the body of a healthy individual, or infection occurs if the mycobacteria are not killed.[11] After infection, the TB bacilli normally become dormant in a person for months, resulting in a latent infection that is not symptomatic nor infectious.[10] Only 5-10 % of people with latent infection will go on to develop active TB.[10,11] Symptoms of TB include coughing, fever, weight loss and night sweats.[10] A minimum of six months of a combination of antibiotics is needed to clear an uncomplicated TB infection.[12] Anti-TB drugs have severe side effects, often causing non-compliance, and this in turn is a major cause of the mycobacteria becoming drug resistant.

With enormous efforts made worldwide to curb TB over the last twenty years, some progress has been made; the mortality rate has decreased by 45 % since 1990 and the number of TB cases has decreased each year, although at a slow rate

of 2 % per year.[7] In 2012, the first new anti-TB drug in 40 years was approved against MDR-TB. Numerous other drugs, new vaccines and new diagnostics are also currently in development.[7]

Mycolic acids in tuberculosis

Mycolic acids (MAs) are in the outer membrane of the lipid rich cell wall of the mycobacteria, contributing to its resistance to chemicals by forming an impermeable layer.[13] The majority of MAs can be found covalently bound to arabinogalactan, but some are also found as free lipids bound to trehalose as trehalose mono- or dimycolates, as shown in Figure 1.[14]

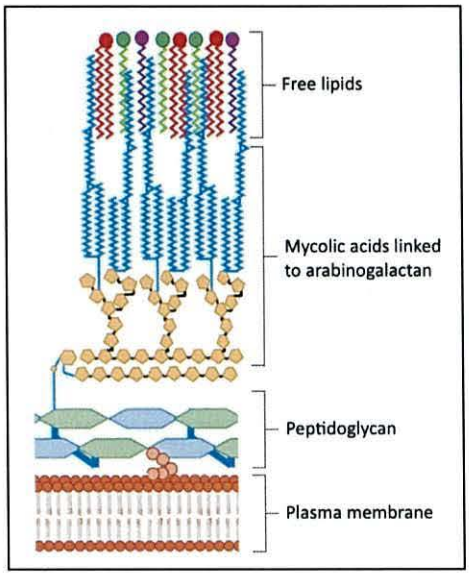


Figure 1: Schematic representation of the arrangement of components in the mycobacterial cell wall. Modified from Nguyen *et al.*[21]

MAs are very long fatty acids containing a mycolic motif (an acid group with an alpha-alkyl chain) and a long fatty alcohol part containing a proximal and distal functional group, the mero-chain (). In *M. tb* there are there three main types of MAs, determined by the distal functional group in the mero-chain (, *D*). In α-MA (AMA) the distal group is a *cis*-cyclopropane, while in the oxygenated methoxy (MMA) and keto-MAs (KMA) it is a methoxy-methyl or keto-methyl

group, respectively. AMA also has a *cis*-cyclopropane group in the proximal position (, *P*), while MMA and KMA may have either *cis*-cyclopropane or α-methyl *trans*-cyclopropane. In *M. tb* the major components, as represented in contain a *cis*-cyclopropane for MMA and an α-methyl *trans*-cyclopropane for KMA. The stereochemistries at the functional groups are very specific, and have been analysed through comparison with synthetic compounds.[15-20] It is believed that the stereochemistries of the major components of the natural MAs are as shown in .

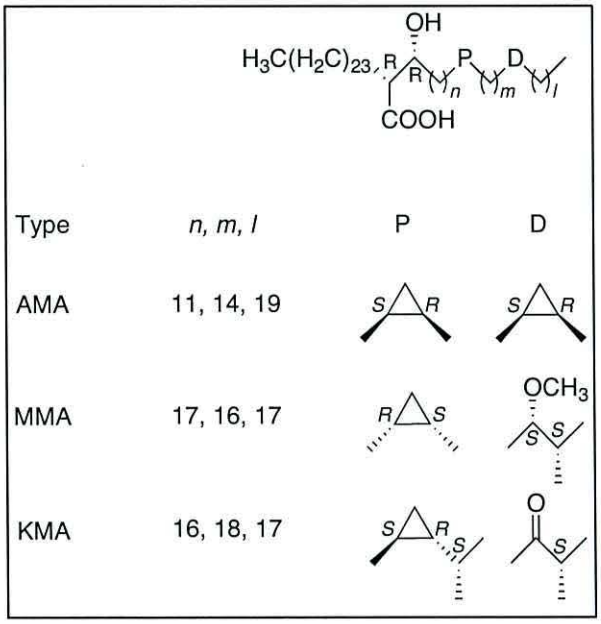


Figure 2: Chemical structures of MAs, representing main components of MAs from *M. tb*. All MAs have identical alpha chains and mycolic motif. *P* and *D* represent the proximal and distal functional groups, respectively, and *n*, *m* and *l* represent alkyl chain lengths.

Mycobacteria and related genera possess many unique characteristics that 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.[13] MA-containing bacteria occur in the genera *Mycobacterium, Nocardia, Rhodococcus* and *Corynebacterium*.[22] The distribution of MA-types varies

within species. For example, in the genus *Mycobacterium*, *M. tb* and *M. bovis* contain AMA, MMA and KMA subclasses while in *M. avium*, *M. intracellulare* and *M. phlei* MMA is replaced by wax ester mycolates.[22] MAs also show variation in size. *Mycobacteria* have very long MAs with 60-90 carbon atoms, *Nocardia* have 46-60, *Rhodococcus* 34-64 and *Corynebacterium* have short MAs ranging from 22-38 carbon atoms. 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.[22]

The widely accepted model for the outer membrane of the mycobacterial cell wall has been as shown in Figure 1, with arabinogalactan-bound MAs forming the inner layer, and free glycolipids such as trehalose mono- and dimycolates forming the outer layer of this bilayer structure. Recently, the CEMOVIS (cryo-electron microscopy of vitreous sections) technique has made it possible to view images of the mycobacterial cell wall in its fully hydrated, native state.[23,24] These studies indicated that there is a bilayer outer membrane structure, that it is composed of lipids, and that in the *Corynebacterium glutamicum* mutant without MAs, this outer membrane disappears, giving strong evidence that the outer membrane consists of MAs as a major component.[23,24]

According to Hoffman and co-workers[23] their new data obtained by CEMOVIS challenges two major aspects of the existing cell wall model, namely that it is symmetrical (not asymmetrical, as previously suggested) and it is thinner than would be needed to accommodate stretched out MAs. They therefore proposed that the inner membrane also consists of extractable lipids (not only MAs covalently bound to arabinogalactan) so that the bilayer structure is symmetrical (Figure 3, A, B). To fit MAs into the space measured for the outer membrane (~8 nm) they propose that either the MA mero-chains also span the outer layer of the outer membrane (Figure 3, A) or only their mero-chains are part of the inner leaflet, with the rest of the MAs situated below the outer membrane (Figure 3, B).[25] Zuber *et al.*[24] proposed a zipper model in which MAs not only have to bend at their functional groups, but also intercalate with the lipids in the outer layer in order to fit into the outer membrane space (Figure 3, C). It is suggested by the authors that the extended mero-chain is too long to fit into the outer membrane space when drawn to scale (as indicated in Figure 3, C). However, more recent

modelling studies indicate that the mero chain of MAs that represent the major components (as in) are not too long to fit into this space.[26] An extended conformation with only one bend at the head group was shown to be 5 nm or more, while the conformations folded at each functional group varied from 1.8-3.0 nm between the different MAs.

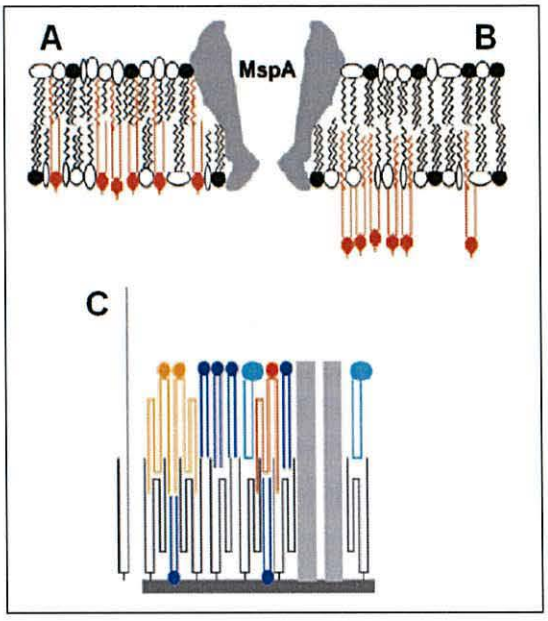


Figure 3: Theoretical models proposed for the mycobacterial outer membrane. (A) MA (red) with mero-chains stretch across the entire outer membrane, (B) MA (red) with mero-chain as part of the inner leaflet with the remainder of the MA situated below the outer membrane[23] and (C) the zipper model in which MAs bound to arabinogalactan (black) and in trehalose monomycolates (red) and dimycolates (orange) are folded and intercalated in the two leaflets.[24]

It is clear that more knowledge is needed to elucidate structural features of the cell wall, and specifically the mycolates in it. The CEMOVIS studies[23,24] provide new models for the mycobacterial cell wall and also highlight the need for a better understanding of the components of the mycobacterial cell wall at the molecular level.

In addition to these lipids being unusual and thus interesting components of mycobacterial cell walls, MAs are a key component in a TB infection.

Mycobacteria cannot grow nor survive when grown *in vitro* or *in vivo* without MAs.[27,28] 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.[29] *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. MAs have been shown to induce a similar neutrophilic airway inflammation.[29]

MAs aid in making the mycobacteria robust so that they can survive for long periods in the host.[13] MAs contribute to the impermeability of the cell wall, making it exceptionally difficult to penetrate with drugs, necessitating long drug therapy courses with multiple drugs.[13] MA profiles change according to external factors such as intracellular growth[30] and changes in temperature,[31-33] suggesting that MA structure is important in maintaining the integrity of the mycobacteria. The effect of MA composition on mycobacterial functionality and, in particular, infection has also been tested by various studies where genes are manipulated in order to eliminate certain enzymes that are involved in the MA biosynthesis pathway. These studies have shown that MA structure is key in developing a persistent infection, maintaining the cell wall permeability, virulence, and intracellular growth and survival.[30,34-37] The precise dependence of these properties on specific functional groups in MAs is elaborated further below. The dramatic dependence of the mycobacterium on the MA profile highlights that all the enzymes involved in MA biosynthesis could be excellent drug targets – even more so since these enzymes are not found in humans and therefore provide potential for greater specificity and low toxicity.[21]

MAs have been shown to be recognised by antibodies in patient serum,[38-40] and therefore provide an avenue to improved rapidity and accuracy in TB diagnosis. Using MAs as antigens is especially appealing since antibodies to MAs also occur in patients coinfected with HIV.[41] Protein antigens are usually presented on MHC class II molecules that recognise and display antigens foreign to the body from an exogenous source. Antigens on MHC class II are recognised by helper T cells with CD4 markers. These cells are very limited in HIV patients, and hence significantly reduce the immune response of coinfected individuals. MAs, being

lipids, are presented by an alternative pathway on CD1 molecules that are recognised by T-cells without CD4 or CD8 markers,[42] and therefore the antibody production against MAs is not inhibited in HIV-coinfected patients. Furthermore, both the antigenicity and inflammatory responses to MAs have been shown to be directly correlated to the MA chemical structure, as demonstrated with synthetic MAs.[39] These relationships will be explored further below.

Structure-function relationships of mycolic acids

There are now numerous studies providing evidence that the structures of MAs, as well as the balance of their composition in the cell wall, are important for the survival, virulence, antigenicity and establishment of a lethal infection of mycobacteria.[30,34-37,39,43,44] However, the structure-function relationships are complex and are only beginning to be unravelled.[45]

The organic synthesis of pure MAs has the advantage of excluding the heterogeneity in chain lengths and stereochemistry that is found in naturally extracted MAs of a single type. Therefore, these pure compounds allow for the systematic unpicking of structure-function traits. A comprehensive set of synthetic MAs have been used to study recognition by patient serum antibodies.[39] Oxygenated MAs were found to be more antigenic than AMA. Within MMA and KMA, the α-methyl *trans*-cyclopropane group in the proximal position of the mero-chain showed better antigenicity than *cis*-cyclopropane. MMA with an α-methyl *trans*-cyclopropane was the most antigenic synthetic MA tested.

Different synthetic MAs have also been used in mouse intratracheal treatments, revealing a strong dependence of the inflammatory response elicited on the MA structure.[43] Once again, the oxygenated MAs showed activity, whereas AMA was inert. MMA with a proximal *cis*-cyclopropane group caused a strong inflammatory response, which was diminished in the α-methyl *trans*-cyclopropane isomer. KMA with a *cis*-cyclopropane group afforded a mild inflammatory response that switched to an anti-inflammatory response for KMA with a proximal α-methyl *trans*-cyclopropane group. These studies on synthetic

MAs have clearly illustrated that both the proximal and distal functional groups in the mero-chain direct the host immune responses.

Another approach that has shed light on the structure-function relationships of MAs is the genetic manipulation of the mycobacteria, eliminating specific enzymes involved in the biosynthesis of MAs and thereby changing the MA composition of the organism. Without the pcaA enzyme that cyclopropanates the proximal position of AMA, a persistent, fatal infection could not be established in mice.[35] The MA composition in these mutants changed so that AMA was unsaturated at the proximal position and the KMA level increased dramatically. TDM isolated from a pcaA deficient mutant was hypoinflammatory, while the wild-type was proinflammatory. The mutant also did not induce severe granuloma formation in mice, indicating that the structure of MAs in TDM altered the host immune response.

A mutant lacking cmaA2, the enzyme that synthesizes *trans*-cyclopropanes in oxygenated MAs, was hypervirulent and induced larger granulomas than the wild type.[36] The lack of *trans*-cyclopropanes also resulted in increased inflammatory responses. These results indicate that cyclopropanes in MAs directly affect the host immune response and that the presences of *trans*-cyclopropanes curb inflammation and virulence. Complete lack of all the enzymes in cyclopropane synthesis, resulting in MAs without any cyclopropanation (only oxygenated functional groups left) resulted in dramatic weakening of the mycobacteria during infection.[34] The net effect of cyclopropanation was shown to subdue the host immune response, as was seen for *trans*-cyclopropanation. In addition, the removal of oxygenated groups over the elimination of cyclopropanes, expressing only unsaturated MAs, leads to decreased virulence and antigencity.[34] Therefore, the hyperinflammatory response observed for the mutant lacking all cyclopropanation, may be attributed, in part, to the presence of unsaturated oxygenated MAs.

The lack of all oxygenated functional groups, obtained by the elimination of the mma4 enzyme that synthesises the hydroxyl precursor to both MMA and KMA, resulted in much slower growth in mice.[44] The slow growth was proposed to be due to the less permeable cell wall, causing a decrease in uptake of essential nutrients in the mutant. Another mutant strain that only produced AMA and

MMA showed impaired *in vitro* growth at lower temperatures and severely reduced ability to grow inside macrophages.[30] It was seen that levels of AMA and KMA were increased significantly during intracellular growth, and therefore the correct MA composition, and the presence of KMA were shown to be important for intracellular growth.[30] KMA plays an important role in pellicle formation, a biofilm structure composed of cells and extracellular matrix, which makes the mycobacteria drug tolerant.[37]

The discussion above illustrates how the MA composition affects various aspects of mycobacterial host interaction and infection. It has been shown that MA structure determines the range of MA conformations[26] and hence our premise is that MA function is related to its conformation. The modelling of MAs could provide key insights into these relationships that are not yet well understood.

Experimental data on mycolic acid conformations

There is very little experimental data on mycolic acid conformations. This is due to MAs being waxy, therefore they do not form crystal structures. They are also difficult to use in varied applications for structure determination due to their extreme hydrophobicity, limiting solubility in most solvents.

MAs in the cell wall, attached to arabinogalactan, have been considered to form a monolayer, and therefore numerous monolayer experiments of MAs on water have been performed[46-54] to study MA conformational behaviour with the goal of gaining knowledge on mycobacterial cell wall properties. Figure 4 (bottom) shows a schematic representation of a monolayer experiment. The MAs are dissolved (typically in chloroform) and put onto a water layer of known surface area. Once the solvent has evaporated, lateral compression is applied. As the lateral pressure increases, the molecules become packed closer together and their hydrophobic chains will become straightened and ordered with their hydrophilic head groups in contact with the water surface. As lateral pressure is applied, the average area per molecule decreases and the pressure in the monolayer increases. The standard way of following this relationship is by plotting the surface pressure and average molecular area against each other. These plots are called isotherms (Figure 4, top). Isotherms contain additional data apart from the

average area per molecule at specific pressures; the pressure and average area at which the monolayer collapses can be obtained, as well as phase transitions,[46-50] indicating varied fluidity and degrees of order in the monolayer. The elasticity modulus can also be calculated as a measure of the rigidity of the monolayer. Villeneuve *et al.*[46,48-50] have performed monolayer experiments at various temperatures and, using data from the monolayers as listed above, were able to plot phase diagrams for MAs.

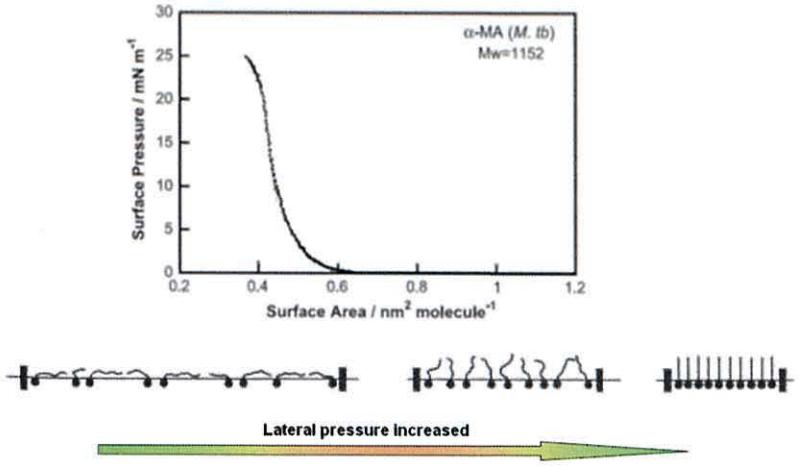


Figure 4: A general pressure-area isotherm isotherm for MAs (top, modified from Hasegawa *et al.*, 2003[51]) and schematic representation of a monolayer experiment (bottom), showing the tighter packing and straightening of alkyl chains as lateral pressure is increased.

In general, from MA monolayer experiments, it has been found that all MAs have folded conformations with larger average molecular areas at low surface pressure and temperature.[46-54] Villeneuve has proposed this to be a W-conformation, with the MA folded at each of the functional groups and the alkyl chains in parallel to each other, as in Figure 5.[49] As the surface pressure increases, the mero-chain of AMA and MMA extends upwards (Figure 5),[47-50,52-54] producing conformations with smaller surface area. KMA is an exception, since it remains compactly folded, even at high surface pressures.[47-49,52] The α-methyl *trans*-

cyclopropane group has recently been shown to contribute to the tendency for KMA to remain compactly folded by promoting folding in MAs by creating a defined hairpin bend.[47]

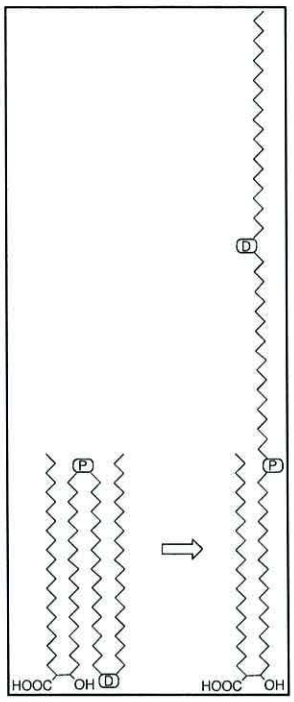


Figure 5: Schematic representation of a MA in a folded W-conformation at large surface areas and complete extension of the mero-chain on monolayer compression. *P* and *D* indicate proximal and distal mero-chain functional groups, respectively.

A variety of techniques have been applied to MA monolayers that have been transferred onto a solid support in order to study their conformations. For example, the thickness of monolayers has been studied by atomic force microscopy[48,54] and vibrations characteristic of certain conformations have been studied by different infrared[52,55] and Raman[51] spectroscopy techniques, as well as X-ray diffraction.[48] Thermodynamic data such as enthalpy of fusion[48,49] and melting temperatures [48,50] have also been reported. In studies by Villeneuve and co-workers the elasticity modulus at various temperatures and surface pressures

has also been calculated for a variety of MAs,[47-50] giving a measure of the monolayer rigidity.

MA monolayer experiments provide physical measurements and factual data on MA conformations, of which there exists very little information. In addition, the monolayers ought to be a close approximation of the arrangement of MAs in the cell wall. This data therefore provides a good model system for modelling MA conformations and a foundation for validating theoretical models of MAs. One of the drawbacks of MA monolayer experiments is that, to date, all these experiments have been formed with mixtures of compounds that were separated from mycobacteria. The only way in which properties of pure compounds will be obtained is to use pure synthesised MAs.[16-18] This would be an ideal system to model, but for the time being, it does not exist experimentally and therefore the effect of MA mixtures will have to be taken into account or modelled to the best of our ability in order to make comparisons between the systems.

Modelling of mycolic acids

Although there is a lot of experimental data on MA monolayers, and monolayers are being modelled more frequently, MA monolayers have not yet been modelled. The closest approximation of a MA monolayer that has been modelled was performed by Hong and Hopfinger in a study of the construction of the mycobacterial cell wall[56] and simulations of its permeability.[57] In order to assemble the MA monolayer part of the cell wall, initially the 17-arabinose units were modelled and extensive conformational searches performed. Pseudo-MAs were attached to eight of the 5-hydroxyl groups of the low energy conformers. Pseudo MAs consisted of a 24-carbon alpha-chain and a shortened mero-chain of the same length, with a *cis*-cyclopropyl group at the proximal position of the mero-chain in order to approximate the densely packed region of MAs in the best way. The end carbon in the mero-chain was given a high mass, in order to simulate some restricted movement as though the chain was full length. After placing the complexes together at different distances from each other and running simulations for 90 to 130 ps, it was determined that the average packing

distance of MAs from each other is approximately 7.3 Å. This spacing was subsequently used to build a monolayer containing only the pseudo-MAs attached to methanol (which was also given a heavy mass to simulate the remaining connectivity in the cell wall) at the acid group. Through 210 ps simulations and comparison with similar monolayers with lipids found more commonly in cell membranes, Minnikin's model of a closely packed MA-monolayer was found to be plausible. The flexibility of the arabinan structure was suggested to support this tight packing. In studying the permeability properties of the monolayer, the pseudo-MA layer was found to be less permeable than lipids that are found in the cell membranes. The pseudo MA layer remained tightly packed, compared with the lipids found in cell membranes that were packed more loosely.

From experimental data, different trends in pressure-area isotherms are seen, with unique phase transition and collapse pressures for the different MA types. These differences are attributable to the unique MA chemical structures, with the mero-chain accounting for the difference between MA types. For this reason it may be considered that the pseudo-MA structure used to model generalised MAs in the work of Hong and Hopfinger[56,57] may have omitted some of the key factors that determine MA monolayer properties.

Villeneuve and co-workers have published several computational studies on MAs in conjunction with monolayer experiments.[47,49,50] They have modelled MMA, KMA and deoxo-MAs (derived from KMA by reduction) from *M. bovis* BCG.[49] The major component of *cis*- and *trans*-cyclopropyl MMA, as well as two *cis*-cyclopropyl KMAs with differing chain lengths and a *trans*-cyclopropyl KMA, were modelled (deoxo was modelled as KMA, without the carbonyl group in the mero chain). All the starting conformations were set up as four parallel chains (W-conformation, Figure 5) with the carbon dihedral at the mycolic motif and the two methylene-groups on either side set to 40, 41, 92, 93 or 94°, as calculated from NMR spectra. The hydrogen of the hydroxyl group and the oxygen of the acid group were assumed to be hydrogen bonded and therefore the distance between them was set accordingly. Lastly, to speed calculations up, the alkyl chains were restricted to be straight (carbon dihedrals set to 180°) before Monte Carlo (MC) calculations were performed with varying forces on the set dihedral

in order to obtain different structures for MD. Structures obtained from MC were minimised and subjected to 20 ps MD at 320 K, with only the restrictions to the carbon dihedral and hydrogen bonding remaining. No explicit mention of solvation is made, and hence either gas phase or implicit solvent was most likely used.

The results were analysed by the five points (*a-e*) as shown in Figure 6, with *a* and *e* being the last carbon of the alpha and mero-chains, respectively, and *b*, *c* and *d*, the carbons at the functional groups in the molecule, as shown in Figure 6. The compactly folded W-conformation had all alkyl chains more than 70 % of their original stretched out length, with small intramolecular distances for *a-c*, *a-e* and *c-e* of less than double their original length and *b-d* less than 1.8 nm. According to a set of rules, the W-shaped molecules that entered MD simulation were analysed to determine whether they had either stayed in this conformation or unfolded. Unfolded conformations were classified either as "stretched-out" or "in-between". An in-between conformation only has the b-d distance increasing to the range 1.8-2.2 nm, and the remainder of distances as defined for the folded W-conformation. In the stretched-out conformation, b-d was above 2.3 nm, and *a-c* was less than twice its original value, with the remainder of the distances the same as those defined for the W-conformation.

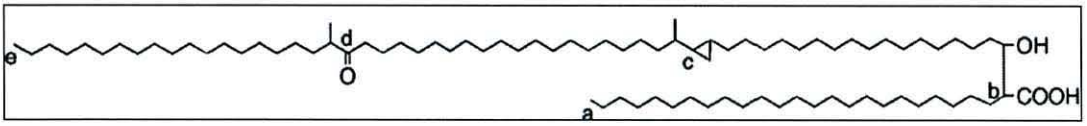


Figure 6: The points, *a-e* that were used to analyse MA conformations on a KMA structure. The points are *a:* the end carbon of the alpha alkyl chain, *b:* the carbon to which the acid group is attached, *c:* the distal carbon of the cyclopropyl group, *d:* the carbon with the carbonyl or methoxy group in KMA or MMA, respectively and *e:* the last carbon in the mero-chain. Taken from Villeneuve *et al.*[49]

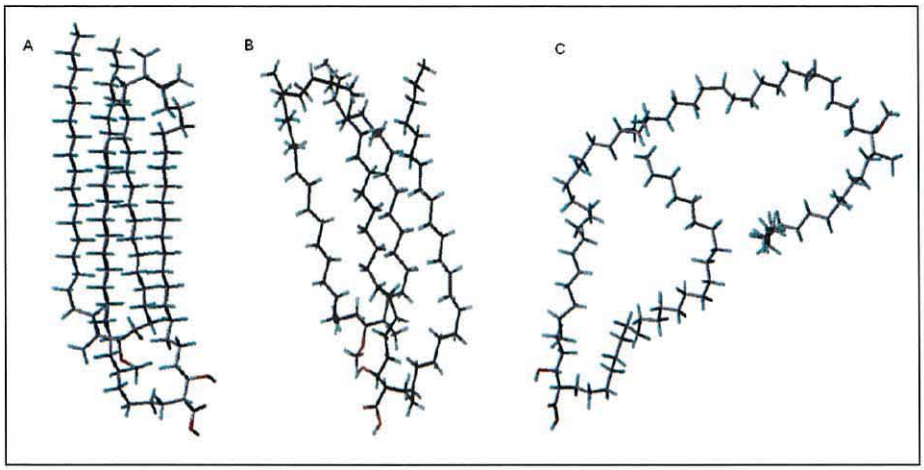


Figure 7: MA conformations shown for MMA. (A) Starting W-conformation, (B) retained W-conformation after 20 ps MD simulation and (C) stretched-out conformation after 20 ps MD simulation. Modified from Villeneuve *et al.*[49]

Overall the authors concluded that, similar to results obtained from their monolayer experiments, KMA preferred to stay in a W-conformation, while MMA often unfolded and stretched out. Deoxo-MA, having only a methyl branch in the distal position of the mero-chain, did not deviate much from its starting structure and seemed to give more proper W-conformations.

The findings were related to the nature of the functional group in the mero-chain as well as the different chain lengths in the molecules.[49] Hydrogen bonding between the carbonyl or methoxy group and the acid group was proposed to contribute to W-conformations being retained. However, this interaction is absent in deoxo-MA, and therefore it is suggested that the bulkiness of the group, and where it is situated (the relative chain lengths) may perturb or favour the W-conformation. KMAs with a longer chain length between the cyclopropane and carbonyl group in the mero-chain than between the mycolic motif and the mero-chain are energetically more stable in the W-conformation (lower minimized energies) than MMAs in which this is not true. A more compact and stable W-conformation is possible when the bulky carbonyl group is pushed towards the end, or is sticking out of the pillar-like structure.

It is clear from these findings that we can learn a great deal about MA conformations, flexibility, and the factors that play a role in determining MA conformations from such atomistic models, as long as the limitations of the model are carefully considered in its application. The limitations of this work included very short simulations times, and inconsistent numbers of replicates for different MAs.

In subsequent work, monolayer experiments were done on AMAs from different species, and were modelled accordingly.[50] Specifically, two types of AMAs were considered; α1-MAs having two *cis*-cyclopropyl groups in the mero-chain (from *M. tb*, *M. kansasii* and MAC) and α3-MAs having one *cis*-cyclopropyl group and one *cis*-double bond (from BCG and MAC). Another useful property of the MAs chosen is that they vary widely in their chain lengths, even though they have similar functional groups.

Quantum chemical DFT calculations on a part of the MA structure were published for the first time.[50] All possible conformations of the mycolate moiety were scanned and two conformers were found to be stable; the *trans* and *gauche* conformers. The *trans*-conformer was reported as slightly more stable, with a 3.43 $kJ.mol^{-1}$ (0.82 $kcal.mol^{-1}$) difference in energy between them. This is not a very high energy barrier and suggests that these structures should be readily interchangeable. The presence of a hydrogen bond is also implied in QM calculations. Therefore, MAs were built with *trans* and *gauche* conformations at the mycolate moiety and with a set distance range between the hydroxyl hydrogen and acid carbonyl group to accommodate hydrogen bonding.[50] The chains were also set as straight for MC calculations, omitting some carbons neighbouring the functional groups. The structures obtained from MC calculations were minimised and subjected to 10 fs MD to relieve any stress in the molecule, followed by 35 ps MD simulation. In these calculations only the carbon dihedral and hydrogen bonding distance at the head group were restricted.

Since the MAs were all of varied methylene chain lengths, the points that were used to analyse conformations were adjusted slightly in order to be relative so that *a* and *e* were defined relative to the distance between the head group and the

proximal functional group and the distance between the proximal and distal functional groups, respectively (see supporting information for more detail). Shapes were defined as W or various extended shapes; extended (eS), folding extended (Se) or fully extended (Sf) shapes, as indicated in Figure 8.

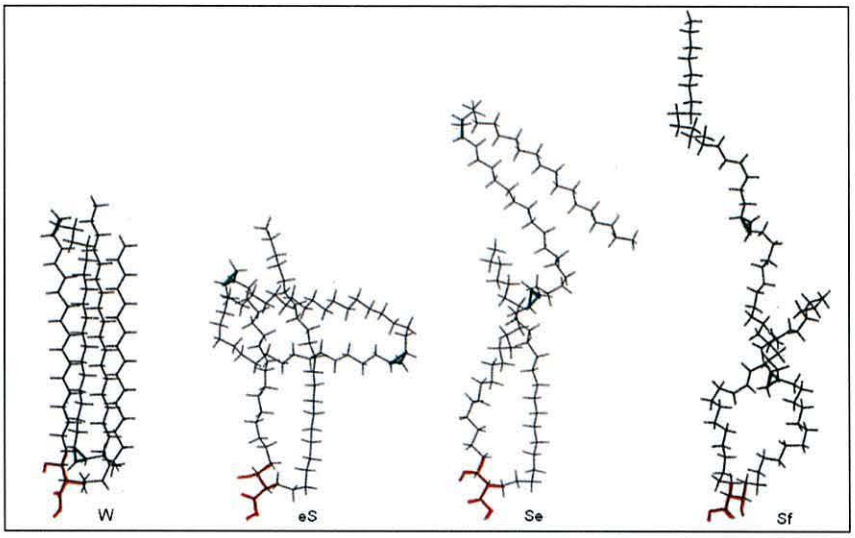


Figure 8: Representative AMA conformations in the starting W-conformation and extended eS, Se and Sf conformation. Modified from Villeneuve *et al.*[50]

The results of the MD simulations are intriguing. All AMAs modelled, with *trans* or *gauche* conformations at the mycolic motif stayed in a W-conformation after 35 ps MD when no restrictions were applied to the molecule. However, once restrictions were applied to the carbon dihedral and to the hydrogen bonding distance in the mycolic motif – with the goal of simulating either lateral pressure, intra-molecular hydrogen bonding or ester linkage to arabinosyl units in the cell wall - it was found that the *trans* conformers still mostly retained their W-shapes, while all the *gauche* conformers changed their shapes to extended ones (even if they ended up in W-shapes again by the end of the simulation).

The percentage of each fold at different time intervals is shown for the different AMAs. The disadvantage of looking at these time steps is that they are discontinuous. It is likely that there will be conformational changes that were missed between them. The authors suggest that α3-MAs with one double bond

are slower to unfold than α1-MAs with two cyclopropanes having identical methylene chain lengths and that this may be due to a more energetically stable W-conformation of α3-MAs.

Another trend that is mentioned is that AMAs from *M. tb* tend to unfold more quickly than those from *M. kansasii* and MAC. Within this set the trend seems to be that MAs with shorter *n*-length chains in the mero-chain unfold quicker than those with longer *n*-chains. This trend brings an interesting difference between AMAs and oxygenated MAs to light. In a previous study on MMA and KMA, it was indicated that those MAs with smaller *n* than *m* values were more stable in a W-shape, and didn't unfold readily.[49] In the present study, the opposite is observed for AMAs.[50] The difference in trends may lie in the intrinsic compositional differences, rather than the chains lengths only, since the distal functional group of AMAs in the mero-chain is not polar and bulky as in oxygenated MAs and may have different preferred conformations.

The components of extended shapes, indicated by the ratio of fully folded to folding-extended shapes, stays relatively constant for each MA after the early stages.[50] The preferred extended shapes seem to be determined by the *m-l* chain lengths in α1-MAs with two cyclopropyl groups. Those with *m-l* lengths that differ most (14-19) prefer the fully extended shapes, while those with more similar *m-l* lengths (16-17 and 14-17) prefer the folding extended shape, perhaps because these matching chain lengths can form a more tightly packed W-shape to resist unfolding. Comparing *M. tb* α1-MAs, the MA with the shorter *n* chain of 11 seems to access fully extended folds less than the MA with *n* of 13. It is suggested that the interaction between the long terminal chains may be exaggerated when *n* is shorter. In comparing α1- and α3-MAs, it was concluded that the double bond does not form a definite bending point as the cyclopropyl group does.

One other factor that is likely to influence MA conformations, but was not addressed by the authors, is the difference in length of the alpha-chain (having either 22 or 24 carbons). It seems apparent that the five MAs with longer 24-carbon alpha-chains generally have lower percentages of W-shapes than the MAs with 22-carbon alpha-chains. The effect of the alpha-chain was proposed to

support the extension of MAs in monolayer experiments, and MAs with the shorter chain collapsed sooner.[53] This experimental observation seems to be true in these MD results too, although they were not modelled under monolayer conditions, but as single molecules.

Recently, the increased capacity of α-methyl *trans*-cyclopropane to facilitate MA folding, as compared to *cis*-cyclopropane, was demonstrated both in monolayer experiments and through simulations.[47] Short-chain model molecules containing *cis*- or α-methyl *trans*-cyclopropane with butyl alkyl chains were shaped into a hairpin U-shape and open-chain conformations and energy calculations performed using B3LYP/6-31G(d). For the *cis*-cyclopropane, the open-chain conformation was 7.61 $kJ.mol^{-1}$ (1.82 $kcal.mol^{-1}$) lower in energy than the U-shaped one, whereas for α-methyl *trans*-cyclopropane the U-shaped conformation was more stable, being 3.81 $kJ.mol^{-1}$ (0.91 $kcal.mol^{-1}$) lower in energy than the open-chain conformation. Molecular dynamics simulation of small model compounds with alkyl chains 3-5 carbons long for 11 ps showed that the distances between the third carbons from the ring were closer for α-methyl *trans*-cyclopropane containing molecules than those with *cis*-cyclopropane. The data shows that α-methyl *trans*-cyclopropane facilitates tighter U-shaped conformations than *cis*-cyclopropane.

Molecular dynamics simulations on the main components of MAs from *M. tb* have also been performed, as shown in Figure 9.[26] From *in vacuo* simulations of AMA, MMA and KMA, it was seen that MAs were able to fold into reproducible groupings, with subtle conformational differences between the three different MA types. The main differences between this approach and the modelling performed by Villeneuve and co-workers are 1) longer timescales were used for improved sampling, 2) no restrictions in the simulations were included and 3) the simulations started from open, unfolded conformations with initial straight alkyl chains and bends, as defined by the precise stereochemistry of the functional groups. Twenty replicate MD runs were done on each MA, starting from the same structure but with different random initial velocities, consequently giving different trajectories. Each simulation was 4 ns long, resulting in 400 frames per simulation.

Analysis on MA conformations was done using eight defined distances, as defined by Villeneuve with five specific points in the molecule (Figure 6),[49] by using principal component analysis (PCA), self organizing maps (SOM, a clustering method) and WUZ-fold analysis.[26] PCA was performed on all frames of the simulations allowing the molecular path of folding to be followed with such a plot. From the variation in the twenty replicates for each MA, it was concluded that the surface energy was explored well with this approach. KMA was seen to show a diffuse pattern of conformer distributions, with fewer unfolded conformations, unlike AMA and MMA. From PCA plots as well as SOM it was concluded that KMA folded into compact conformations faster than other MAs.

To analyse MA conformations more specifically, criteria were derived to define all the possible folds of MAs folding at their functional groups with extended parallel alkyl chains. As the MA folded at all of its functional groups with adjacent and straight chains was defined as a W-shape in 2-dimesional space, U and Z folds, where MAs bend at one and two functional groups, respectively, were also defined with the prefix *a* in aU and aZ and *e* in eU and eZ used to describe the chain that is extended, terminating in *a* or *e* and *s* referring to symmetry in the shape, as found in sU and sZ. These conformations are shown both schematically and specifically for AMA in Figure 9.

AMA was found to be the most flexible with the highest frequency of WUZ-folds (7.93 %) and it was the only MA that assumed all seven possible WUZ conformations. MMA and KMA showed reduced numbers of WUZ-conformations (1.35 % and 3.01 %, respectively). The variation in the frequency of the different WUZ-folds for the different MAs distinguishes them based on their conformational preferences. Therefore, the chemical structure determines the conformations of MAs.

It is significant that these folds, as observed in monolayer experiments, were also detected for single molecules, without any lateral pressure or packing effects, as found in a monolayer. This supports the findings of monolayer studies by showing that these are feasible MA conformations.

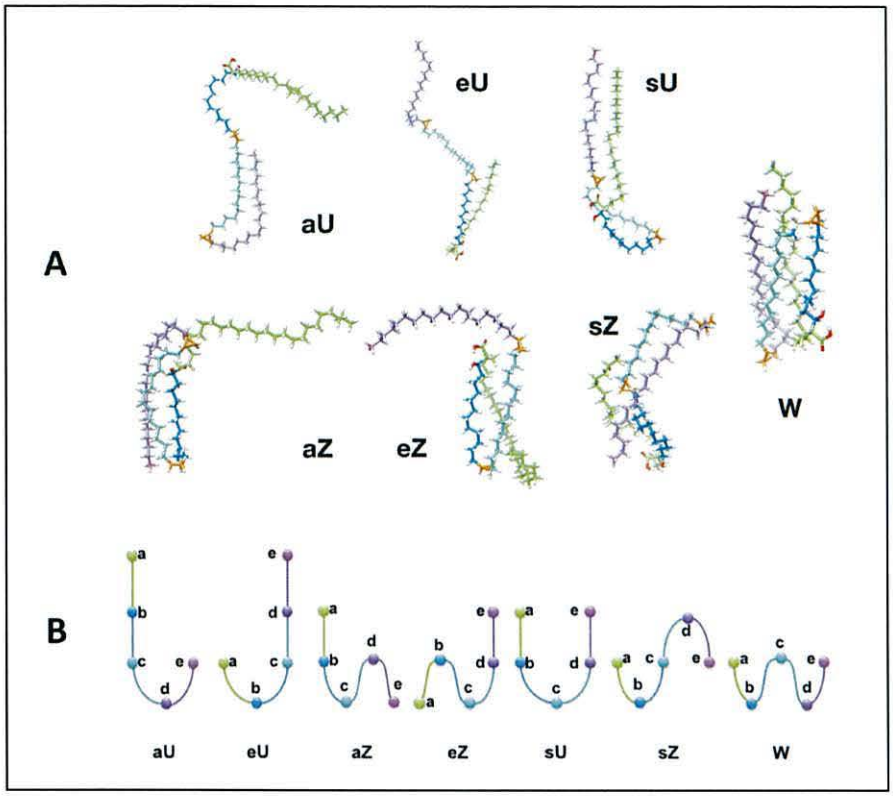


Figure 9: The seven WUZ folds represented as minimised structures from AMA (A) and schematically (B). Taken from Groenewald *et al.*[26]

MA models based on MAs from *M. bovis* BCG, are very similar to the main components of *M. tb*, with slightly different chain length distributions. The *cis*-cyclopropyl MMA modelled by Villeneuve[49] however, is the same as that from *M. tb*, described above (provided that stereochemistries around the cyclopropyl groups are as stated in Villeneuve *et al.*,[49] because this is not in agreement with the drawn structures). It is complicated to compare these studies because of the vastly different approaches taken in each.

Although this study lacked the kind of variation in starting conformations as was provided by MC calculations in the work of Villeneuve,[49] it was shown that the random starting velocities provided good variation and that the potential energy surface was explored well for each MA.[26] Apart from the timescale and starting structures, which were very different in these studies, differences may also be

caused by different forcefields and software that were used, and therefore comparisons should be made with careful consideration of these variations.
Lacking from all the above work on MA modelling is the inclusion of external factors on MA folding, for example the presence of more MAs in order to study packing effects that would be relevant in the cell wall environment. Additionally, explicit solvent with differing polarity would inform the effect of these external changes in environment on MA conformations. Detailed insight into the functional groups and the reasoning behind their specific stereochemistry will also improve the understanding of the effect of the functional group on MA folding. Together this data holds the potential to improve connections of MA structure to the functions of the different MAs in survival, infection and cell wall permeability.

Rationale

In the 19th century, despite the huge amount of suffering caused by TB at the time, TB obtained a romantic, almost glamorous reputation.[1,3] People suffering with TB were described in an idyllic way as having clean, fair or pale skin and delicate rosy complexion. It was seen as a beautiful, tragic death and hence became a popular disease for heroes in arts and literature of the 19th century.[1]
The picture painted by TB today is certainly not romantic. At the current rate, the goal of eliminating TB is not yet in sight. Despite the progress made, it is still estimated that out of the 8.6 million TB cases in 2012, roughly 3 million have been missed, either because they were not diagnosed, or they were diagnosed but not reported.[10] The diagnosis and treatment of drug resistant strains of TB are enormous challenges that are inhibiting progress.[10] In addition, HIV testing among TB patients and appropriate preventative treatment is not yet implemented sufficiently, leading to high mortality rates.[10]
A better understanding of TB at the molecular level will ensure better control over the disease. As discussed above, MAs play a key role in TB infection and MA structure modulates the host immune response. Hence, an improved

knowledge of MA conformational behaviour on a molecular level is urgently called for. To highlight the main points, MA models can be:

- used to predict minimal antigenic structures for use in serodiagnosis
- exploited in drug design
- used to advance understanding of the mycobacterial cell wall; how MAs fit into the cell wall and affect its permeability and fluidity
- used to elucidate specific structure-function relationships of MAs, especially related to their interaction with the human immune system.

The objectives of this work are to explore the conformational flexibility and conformational preferences of MAs using different levels of computational theory, namely

a) Quantum mechanics: to study small fragments of MAs in detail by doing systematic conformational searching
b) Atomistic molecular dynamics: to study the conformations assumed by single MAs in different solvents and
c) Coarse graining: to study the packing effects of MAs using a monolayer system

In this way, the strengths of the various computational theories are harnessed to investigate different aspects of MA conformations. Background information of the different techniques can be found in Chapter 2. Quantum mechanics, delivering the highest accuracy, but also being the most computationally intensive, is appropriate for an in-depth study of functional group stereochemistry by examining small fragments containing the functional groups. The effect of cyclopropane stereochemistry is specifically probed for its ability to form a folding point in MAs. This work is introduced in Chapter 3 and presented in Manuscript I.

Atomistic molecular dynamics is sufficiently fast and accurate for the simulation of single MAs in explicit solvent. Using this methodology, the effect of intrinsic factors (namely the chemical structure of MAs) and external factors (such as polar and nonpolar solvents) on MA conformations can be elucidated to further increase our understanding of MA conformational behaviour. This part of the work is presented in Chapter 4 and Manuscript II.

The coarse grained approach, using beads to represent groups of atoms, has the lowest accuracy, but due to the fewer interactions that need to be calculated in such a model, is capable of simulations of large systems for longer timescales. Hence, a coarse grained model lends itself well to modelling numerous MAs together in a monolayer. This model is the most realistic depiction of MAs in accordance with their natural environment presented in this work and provides the essential advantage of comparison with existing experimental data. It is the product of work done on a smaller scale in quantum mechanics and molecular dynamics that provided data to feed into the coarse grained model, and also provides a basis for further developing accurate models of other MAs and the remaining components of the mycobacterial cell wall. This last component of the work can be found in Chapter 5 and Manuscript III.

References

(1) Chalke, H. D. *Med. Hist.* **1962**, *6*, 301.

(2) Hershkovitz, I.; Donoghue, H. D.; Minnikin, D. E.; Besra, G. S.; Lee, O.-C.; Gernaey, A. M.; Galili, E.; Eshed, V.; Greenblatt, C. L.; Lemma, E.; Bar-Gal, G. K.; Spigelman, M. *PLoS One* **2008**, *3*, e3426.

(3) Daniel, T. M. *Resp. Med.* **2006**, *100*, 1862.

(4) Glynn, J. R. *Br. Med. Bull.* **1998**, *54*, 579.

(5) Frieden, T. R.; Sterling, T.; Pablos-Mendez, A.; Kilburn, J. O.; Cauthen, J. M.; Dooley, S. W. *N. Engl.J.Med.* **1993**, *328*, 521.

(6) Dolin, P. J.; Raviglione, M. C.; Koch, A. *A review of current epidemiological data and estimation of future tuberculosis incidence and mortality*, World Health Organization, 1993.

(7) WHO *Global tuberculosis report*, **2013**.

(8) Gandhi, N. R.; Nunn, P.; Dheda, K.; Schaaf, H. S.; Zignol, M.; van Soolingen, D.; Jensen, P.; Bayona, J. *The Lancet* **2010**, *375*, 1830.

(9) Hargreaves, N. J.; Kadzakumanja, O.; Whitty, C. J. M.; Salaniponi, F. M. L.; Harries, A. D.; Squire, S. B. *Int.J.Tuberc.Lung Dis.* **2001**, *5*, 847.

(10) WHO *Countdown to 2015. Global tuberculosis report supplement* ,**2013**.

(11) O'Garra, A.; Redford, P. S.; McNab, F. W.; Bloom, C. I.; Wilkinson, R. J.; Berry, M. P. R. *Annu. Rev. Immunol.* **2013**, *31*, 475.

(12) Fox, W.; Ellard, G. A.; Mitchison, D. A. *Int. J. Tuberc. Lung Dis.* **1999**, *3*, S231.

(13) Barry, C. E.; Lee, R. E.; Mdluli, K.; Sampson, A. E.; Schroeder, B. G.; Slayden, R. A.; Yuan, Y. *Prog. Lipid Res.* **1998**, *2/3*, 143.

(14) Sekanka, G.; Baird, M.; Minnikin, D.; Grooten, J. *Expert Opin. Ther. Patents* **2007**, *17*, 315.

(15) Al Dulayymi, J. R.; Baird, M. S.; Roberts, E. *J. Chem. Soc., Chem. Commun.* **2003**, 228.

(16) Al Dulayymi, J. R.; Baird, M. S.; Roberts, E. *Tetrahedron* **2005**, *61*, 11939.

(17) Al Dulayymi, J. R.; Baird, M. S.; Roberts, E.; Deysel, M.; Verschoor, J. *Tetrahedron* **2007**, *63*, 2571.

(18) Al Dulayymi, J. R.; Baird, M. S.; Roberts, E.; Minnikin, D. E. *Tetrahedron* **2006**, *62*, 11867.

(19) Al Dulayymi, J. R.; Baird, M. S.; Mohammed, H.; Roberts, E.; Clegg, W. *Tetrahedron* **2006**, *62*, 4851.

(20) Koza, G.; Baird, M. S. *Tetrahedron Lett.* **2007**, *48*, 2165.

(21) Nguyen, L.; Pieters, J. *Annu. Rev. Pharmacol. Toxicol.* **2009**, *49*, 427.

(22) Minnikin, D. E.; Goodfellow, M. *Soc. Appl. Microbiol. Symp. Ser.* **1980**, *8*, 189.

(23) Hoffmann, C.; Leis, A.; Niederweis, M.; Plitzko, J. M.; Engelhardt, H. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 3963.

(24) Zuber, B.; Chami, M.; Houssin, C.; Dubochet, J.; Griffiths, G.; Daffè, M. *J. Bacteriol.* **2008**, *190*, 5672.

(25) Hoffmann, C.; Leis, A.; Niederweis, M.; Plitsko, J.; Engelhardt, H. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 3963.

(26) Groenewald, W.; Baird, M. S.; Verschoor, J. A.; Minnikin, D. E.; Croft, A. K. *Chem. Phys. Lipids* **2014**, *180*, *15*.

(27) Portevin, D.; de Sousa-D'Auria, C.; Houssin, C.; Grimaldi, C.; Chami, M.; Daffé, M.; Guilhot, C. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 314.

(28) Vilchèze, C.; Morbidoni, H. R.; Weisbrod, T. R.; Iwamoto, H.; Kuo, M.; Sacchettini, J. C.; Jacobs, W., R. Jr. *J. Bacteriol.* **2000**, *182*, 4059.

(29) Korf, J.; Stoltz, A.; Verschoor, J.; De Baetselier, P.; Grooten, J. *Eur. J. Immunol.* **2005**, *2005*, 890.

(30) Yuan, Y.; Zhu, Y.; Crane, D. D.; Barry, C. E. *Mol. Microbiol.* **1998**, *29*, 1449.

(31) Liu, J.; Barry, C. E.; Besra, G. S.; Nikaido, H. *J. Biol. Chem.* **1996**, *271*.

(32) Toriyama, S.; Yano, I.; Masui, M.; Kusunose, E.; Kusunose, M.; Akimori, N. *J. Bioch.* **1980**, *88*, 221.

(33) Alibaud, L.; Alahari, A.; Trivelli, X.; Ojha, A. K.; Hatfull, G. F.; Guerardel, Y.; Kremer, L. *J. Biol. Chem.* **2010**, *285*, 21698.

(34) Barkan, D.; Hedhli, D.; Yan, H.-G.; Huygen, K.; Glickman, M. S. *Infect. Immun.* **2012**, *80*, 1958.

(35) Glickman, M. S.; Cox, J. S.; Jacobs, W. R. *Mol. Cell* **2000**, *5*, 717.

(36) Rao, V.; Gao, F.; Chen, B.; Jacobs, W. R.; Glickman, M. S. *J. Clin. Invest.* **2006**, *116*, 1660.

(37) Sambandan, D.; Dao, D. N.; Weinrick, B. C.; Vilcheze, C.; Gurcha, S. S.; Ojha, A.; Kremer, L.; Besra, G. S.; Hatfull, G. F.; Jacons, W. R. J. *mBio* **2013**, *4*, e00222.

(38) Pan, J.; Fujiwara, N.; Oka, S.; Maekura, R.; Ogura, T.; Yano., I. *Microbiol. Immunol.* **1999**, *43*, 863.

(39) Beukes, M.; Lemmer, Y.; Deysel, M.; Al Dulayymi, J. R.; Baird, M. S.; Koza, G.; Iglesias, M. M.; Rowles, R. R.; Theunissen, C.; Grooten, J.; Toschi, G.; Roberts, V. V.; Pilcher, L.; Van Wyngaardt, S.; Mathebula, N.; Balogun, M.; Stoltz, A. C.; Verschoor, J. A. *Chem. Phys. Lipids* **2010**, *163*, 800.

(40) Thanyani, S. T.; Roberts, V.; Siko, D. G.; Vrey, P.; Verschoor, J. A. *J. Immunol. Methods* **2008**, *332*, 61.

(41) Schleicher, G. K.; Feldman, C.; Vermaak, Y.; Verschoor, J. A. *Clin. Chem. Lab. Med.* **2002**, *40*, 882.

(42) Beckman, E. V.; Porcelli, S. A.; Morita, C. T.; Nehar, S. M.; Furlong, S. T.; Brenner, M. B. *Nature (London, U. K.)* **1994**, *372*, 691.

(43) Vander Beken, S.; Al Dulayymi, J. R.; Naessens, T.; Koza, G.; Maza-Iglesias, M.; Rowles, R.; Theunissen, C.; De Medts, J.; Lanckacker, E.; Baird, M. S.; Grooten, J. *Eur. J. Immunol.* **2011**, *41*, 450.

(44) Dubnau, E.; Chan, J.; Raynaud, C.; Mohan, V. P.; Lanèelle, M.-A.; Yu, K.; Quèmard, A.; Smith, I.; Daffè, M. *Mol. Microbiol.* **2000**, *36*, 630.

(45) Verschoor, J.; Baird, M. S.; Grooten, J. *Prog. Lipid Res.* **2012**, *51*, 325.

(46) Villeneuve, M. *Characteristic Conformational Behaviors of Representative Mycolic Acids in the Interfacial Monolayer*. In Cardona, P-J., Ed. *Understanding Tuberculosis - Deciphering the Secret Life of the Bacilli*. InTech 2012, p 317.

(47) Villeneuve, M.; Kawai, M.; Horiuchi, K.; Watanabe, M.; Aoyagi, Y.; Hitotsuyanagi, Y.; Takeya, K.; Gouda, H.; Hirono, S.; Minnikin, D. E. *Microbiology* **2013**, *159*, 2405.

(48) Villeneuve, M.; Kawai, M.; Kanashima, H.; Watanabe, M.; Minnikin, D. E.; Nakahara, H. *Biochim. Biophys. Acta, Biomembr.* **2005**, *1715*, 71.

(49) Villeneuve, M.; Kawai, M.; Watanabe, M.; Aoyagi, Y.; Hitotsuyanagi, Y.; Takeya, K.; Gouda, H.; Hirono, S.; Minnikin, D. E.; Nakahara, H. *Biochim. Biophys. Acta, Biomembr.* **2007**, *1768*, 1717.

(50) Villeneuve, M.; Kawai, M.; Watanabe, M.; Aoyagi, Y.; Hitotsuyanagi, Y.; Takeya, K.; Gouda, H.; Hirono, S.; Minnikin, D. E.; Nakahara, H. *Chem. Phys. Lipids* **2010**, *163*, 569.

(51) Hasegawa, T. *Biopolymers* **2004**, *73*, 457.

(52) Hasegawa, T.; Amino, S.; Kitamura, S.; Matsumoto, L.; Katada, S.; Nishijow, J. *Langmuir* **2003**, *19*, 105.

(53) Hasegawa, T.; Leblanc, R. M. *Biochim. Biophys. Acta.* **2003**, *1617*, 89.

(54) Hasegawa, T.; Nishijo, J.; Watanabe, M.; Funayama, K.; Imae, T. *Langmuir* **2000**, *16*, 7325.

(55) Hasegawa, T.; Nishijo, J.; J., U.; Watanabe, M. In *Two-Dimensional Correlation Spectroscopy*; AIP Conferernce Proceedings: Kobe-Sanda (Japan), 2000; Vol. 503.

(56) Hong, X.; Hopfinger, A. J. *Biomacromolecules* **2004**, *5*, 1052.

(57) Hong, X.; Hopfinger, A. J. *Biomacromolecules* **2004**, *5*, 1066.

Supporting Information

Definition of points a-e from Villleneuve *et al.* 2010

Points *a-e* were adjusted as follows: *a* was moved to be n+2 carbons away from the carboxyl-bearing group carbon *b*, with n being the chain length between the mycolic motif and the proximal functional group in the mero-chain. Points *b*, *c* and *d* are the same as defined in Figure 6 at each functional group, with *d* in this case being the distal carbon in the double bond or cyclopropyl group. Point *e* was moved to be at the m^{th} carbon from *d*, m being the methylene chain length between the proximal and distal functional groups in the mero-chain, and *f* is labelled as the last carbon in the mero-chain. The rules for defining different structures are as follows:

- W-shape (W):
    - b/d < b/c and
    - a/c < b/c and
    - c/e < c/d or d/e and
    - a/e < a/b
- Either shape (e):
    - b/c < b/d < b/c × 1.3 or 1.4 or 1.5 (for n = 17, 15, 11, respectively) and
    - a/c < b/c and
    - c/e < c/d
- Extended shape (eS):
    - c/e > c/d and
    - c/f > c/d

- Folding extended shape (Se):
    - a/c < b/c and
    - b/d > b/c × 1.3 or 1.4 or 1.5 (for n = 17, 15, 11, respectively) and
    - c/e > c/d and/or
    - c/f > c/d
- Fully extended shape (Sf):
    - c/e > c/d and
    - c/f > c/d

Chapter 2: Methodology Background

"To the man who only has a hammer, everything he encounters begins to look like a nail."
Abraham Maslow

Introduction

The application of different methodologies to the conformational study of MAs is key to this thesis. There are several reasons for applying more than one theory to a single research question. There has been very little work done on the modelling of MAs, and therefore there is no established and proven methodology to be applied. Numerous theories exist currently. Each has its own strengths and weaknesses and vary in their accuracy and cost, as illustrated in Figure 10. Therefore different system sizes and detail can be handled well by different theories. Applying each theory to perform in its preferred niche ensures high quality data and information can flow from one methodology into another (see Figure 10) to give a full and detailed overall picture.

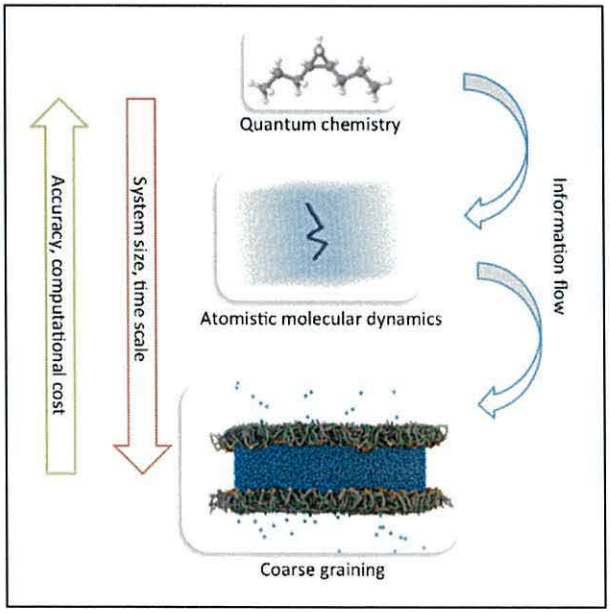


Figure 10: An illustration of the system size and accuracy of the computational methods utilised, as well as the flow of information between them.

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.[1] 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.[2]

Quantum Mechanics

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.[2,3] 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.[1]

Hartree-Fock (HF) models are *ab initio* models that attempt to solve the Schrödinger equation without any empirical information. 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.[2] Different QM models differ in these approximations, and therefore different models have different capability, reliability, and cost in terms of computational time and power.[2]

Since the nuclei move slowly in comparison to electrons, the first approximation to be made to the Schrödinger equation is to assume that, relative to the electrons, the nuclei do not move. Further, it is assumed that electrons move independently of each other. 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. The equations need to be solved iteratively with the Self Consistent Field (SCF) Method, and therefore are sometimes referred to as SCF.

The HF model is lacking in the treatment of the correlation of electron movements. Each electron only sees an average distribution of all the other electrons. In the assumptions made, the major correlation between electrons of the same spin is included (exchange correlation), and therefore it is the remaining correlation between electrons of opposite spin that is absent. Methods that attempt to rectify this are referred to as electron correlation methods, or post-SCF methods. One such method that was used here is second order Møller-Plesset (MP2).

MP2 draws on Møller-Plesset perturbation theory to add electron correlation to HF theory. The familiar HF theory is built on by adding a small, external perturbation, as performed in perturbation theory. This perturbation accounts for the missing electron correlation in HF theory. MP2 scales as N^5, with N the number of basis functions.[4] In perturbation theory, the added perturbation needs to be sufficiently small, and the fact that electron-electron repulsion has a significant effect on the total energy may affect the accuracy of MP2.[4]

Density Functional Theory

In Density functional theory (DFT), molecules are seen as electrons interacting with each other and an external potential (attraction to the nucleus).[4] Hohenberg and Kohn, proved the basis for the theory, showing that there is a one-to-one correspondence between the electron density and the energy of a system. This

implies that only three coordinates are required with no dependence on the number of electrons, compared to the wave function for a system with N electrons containing $3N$ coordinates.[1] DFT is very popular because it can improve on the results of HF with very little added cost.

A functional is a function of a function. In DFT, a functional is a function of the electron density, which in turn is a function of atomic coordinates. The electronic energy is divided into components that are computed separately, namely kinetic energy (from movement of electrons), electron-nuclear attraction and repulsion between nuclei, Coulomb self-interaction of electron density (electron-electron repulsion), and an exchange-correlation term. The latter includes correlation of individual electron motion and exchange energy arising from the antisymmetry of the quantum mechanical wave function. The challenge in DFT is calculating the exchange and correlation components. Different methods exist:

- Local density approximation (LDA) is the simplest approximation and is based only on the values of electron spin densities,
- Generalised gradient approximation (GGA) also includes the gradient of the densities, in addition to the values of electron spin densities.
- Hybrid functionals include a fraction of the exact exchange energy calculated from HF theory. The exchange functional is therefore linear combination of HF, local, and gradient-corrected exchange terms. The exchange functional is combined with a local and/or gradient-corrected functional.[5]

In reality, the difference between the exact kinetic energy and the energy calculated by assuming no interaction is absorbed into the exchange-correlation term.

ωB97X-D[6] is a new, long-range-corrected hybrid DFT functional based on the previous ωB97X, and is one of a number of newer DFT methods providing improved results for complex biological systems. It contains 100 % long-range HF exact exchange and a small fraction (approximately 22 %) short-range exact exchange, which has been shown to give improved results for covalent systems as part of the precursor ωB97X functional.[7] It also contains a modified B97

exchange density functional for short-range interaction and the B97 correlation density functional, which are both generalised gradient approximations. Lastly, it contains empirical atom-atom dispersion corrections. The dispersion is added to the DFT Kohn-Sham energy and therefore contributes very little cost and complexity. The dispersion term replaces some of the medium- and long-range electron correlation effects in the gradient-corrected DF using damped atom-pair dispersion corrections. The dispersion energy is dependant on the distance between the two atoms and the dispersion coefficient. The dispersion coefficient is derived, in this case, from the London dispersion formula and is calculated based on atomic ionisation potentials and static dipole polarisabilities, and is hence a less empirical approach than other methods.[8]

The parameters for the hybrid functional were optimised on a large, diverse training set in order to give good performance over a range of applications. With the focus on improving long-range van der Waals interactions, the S22 data set for non-covalent interactions carried ten times more weight than other data. ωB97X-D has been shown to give very good results, with an overall accuracy of 8.37 $kJ.mol^{-1}$ (2 $kcal.mol^{-1}$), relative to experimental data. Good non-covalent interactions with a small mean absolute error of 0.92 $kJ.mol^{-1}$ (0.22 $kcal.mol^{-1}$) have been obtained and therefore ωB97X-D is recommended for use in systems where non-covalent interactions are expected to be significant. The good overall accuracy and reproducibility of non-covalent interactions were the reasons it was selected as the DF method of choice in this thesis. The limitations of ωB97X-D include the fact that dispersion is only added as an empirical term and therefore dispersion effects do not influence the Kohn-Sham orbitals. Although ωB97X-D is free from long-range self-interaction, it still has some self-interaction at short-range.

Molecular mechanics

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.[1] 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 a 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.[1] Each term describes the energy required for distorting a molecule from some idealised form (ideal bond lengths and angles etc.) and reflects the strain inherent to a real molecule, compared to an idealised form.[2]

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 modelled. Parameterisation encourages the development of models that 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.[1] 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.[1]

For the MD simulations described in this thesis, the optimised potentials for liquid simulations all-atom (OPLS-AA)[9,10] forcefield was used. As the name

suggests, this forcefield has been optimised for simulation of liquids. The focus in parameterisation for this forcefield was on non-bonded parameters that were developed to reproduce thermodynamic and structural properties of pure organic liquids,[9,10] but have since expanded to include a range of organic molecules. Bond stretching and angle bending parameters have been mostly acquired from the AMBER-AA forcefield. Torsional parameters have been obtained by fitting to *ab intio* rotational energy profiles obtained at RHF/6-31G* level of theory.[9-11] Torsional energies showed deviations of 0.84 $kJ.mol^{-1}$ (0.2 $kcal.mol^{-1}$) on average.[9] The average error in heat of vaporisation for hydrocarbons is 2%, while the average error for densities is 3%.[9]

One of the main advantages of forcefield methods is that parameters for small building blocks will be generally applicable to larger molecules made up of several building blocks. However, in contrast to this principle, it is quite common for additional parameters to be derived for specific molecules or classes of molecules to get a better description for them.[12,13] Deriving parameters specific to each molecule would be pointless, but whether existing, general parameters are adequate remains subject to each researcher. The balance between the two extremes should most heavily depend on what outcome hopes to be achieved. Additionally, the availability of experimental data will facilitate the optimisation of forcefield parameters to reproduce the experimental data as precisely as possible, whereas the lack thereof limits the degree to which parameters can be optimised for specific molecules.

The OPLS-AA forcefield is optimised for liquid simulations and is therefore not the most widely used forcefield for simulating lipids, as these are usually modelled in a monolayer or bilayer setup.[14] The GROMOS and CHARMM27 forcefields have been more popular for lipid simulations, with adjusted non-bonded parameters for better reproduction of phase behaviour in lipid membranes.[14] The OPLS-AA forcefield has required re-parameterisation for better description of lipid molecules. It has been found that the adjustment of non-bonded parameters[12,15] and partial charges,[12] as well as the inclusion of longer alkanes in the parameterisation[12] gave improved lipid properties such as area per lipid and lipid phases.

While MAs are similar to membrane lipids in that they are hydrophobic in nature with a hydrophilic head-group, they are also different since MAs are much longer than membrane lipids with different functional groups in the hydrocarbon chains. By using the OPLS-AA forcefield to model MAs in different solvents we are relying on the general applicability of parameters for building blocks to be valid for larger molecules. Since MAs are lipids, they do not form crystal structures, and there is little structural data available besides indirect measurements from studies such as monolayers. After obtaining conformational models for MAs that correlate well with experimental findings by using the Compass forcefield without any additional parameterisation in the gas phase,[16] simulating the molecules in solvents was the next step forward. Therefore the OPLS-AA forcefield seemed appropriate for our simulations of single MAs in solvents. However, the forcefield was missing one angle- and several dihedral parameters for cyclopropane. These parameters were obtained from QM calculations using a force-matching approach.[17]

Coarse grained modelling

In order to approach a system size and simulation length scale to examine biological systems in a more realistic setting from which collective phenomena (such as phase behaviour and packing effects in this case) can be studied, simpler models need to be considered.[18] The detailed structure and dynamics obtained from short atomistic simulations is often less relevant for large system sizes where the collective phenomena are dictated by a few key properties like the hydrophobic and hydrophilic natures of the molecule.[18] Coarse graining (CG) implies a simplified molecular representation in which all the atomistic detail is not included. Pseudo-atoms are used that represent more than one atom. In such a simplified model with fewer degrees of freedom there are fewer interactions to calculate and the potential energy surface is smoother due to the coarse scale. Therefore this approach makes it feasible to simulate larger systems for longer time scales and to observe collective phenomena, which are not currently possible using *ab initio* methods.

The essence of a CG model that enables larger systems and longer timescales to be modelled, namely that there is a loss of chemical detail, is ironically also its limitation in predictive power. The paradox in all forcefield methods therefore, is that the model can be improved by adding more detail, with the very advantage of forcefields being that parameters of the building blocks can be generally applied to a wide range of molecules.[19] Therefore it is important to consider the most appropriate way of implementing a CG model in order to retain the degrees of freedom that will reproduce the fundamental physics of the system.[18,20]

In this work, the MARTINI CG forcefield[21] has been used. This forcefield has been in existence for about a decade and has been reviewed in detail elsewhere.[19,22] The underlying principles, strengths and limitations will be outlined here.

The MARTINI forcefield is based on a four-to-one mapping where an average of four heavy atoms (and attached hydrogens) are represented by one interacting site, called a coarse grained bead. There are four main types of beads; polar (P), nonpolar (N), apolar (C) and charged (Q) which are further subdivided by number or letter into 18 types in total to improve accuracy. The numbers range from one to five with one indicating a low polarity and five a high polarity, whereas the letters indicate hydrogen-bonding capabilities with d = donor, a = acceptor, da = donor and acceptor and 0 = no hydrogen bonding. Standard beads have a mass of 72 amu and a size of 0.47 nm. Exceptions to the above for standard beads are ions and ring-like molecules, for which details can be found elsewhere.[21]

Nonbonded interactions are modelled with a 12-6 Lennard-Jones (LJ) potential for which the particle types involved determine the strength of the interaction. LJ interactions are in the shifted form from 0.9 nm to 1.2 nm, where it is cut off. Charged group interactions are modelled with a Coulombic energy function that is also in the shifted form from 0 nm to 1.2 nm, where it is cut off.

Bonded interactions are modelled by potential energy functions commonly used in forcefields. To parameterise bonded interactions, atomistic simulations are often utilised. For this, the atomistic simulation will be mapped to the CG structure by identifying the centre of mass of the atoms corresponding to the

mapped CG (MCG) bead. Then distribution functions for the mapped and true CG simulations are obtained and compared, with the CG parameters being adjusted until the MCG distributions are represented well. The model is dependent on the simulation parameters that it was parameterised for, namely a timestep of 20-40 fs, the frequency of neighbour list updates, cut-off radii for non-bonded potentials and the form of the shift function.

The MARTINI forcefield has been systematically parameterised using thermodynamic data such as the water/oil partitioning of the building blocks. This approach makes the forcefield more widely applicable to a variety of molecules with the idea that the properties of individual beads that have been carefully parameterised are transferable to larger molecules. The simplicity of the model, in being able to string together basic building blocks to create complex molecules, is one of its best attributes, in addition to its speed and reasonable accuracy for a large variety of biological systems.

Some limitations need to be kept in mind when deriving conclusions from MARTINI CG simulations. The fluid phase is represented well, but the solid and gas phases appear too stable in comparison to the fluid phase. There is an inherent loss of entropy, which will affect the enthalpy/entropy balance and its dependence on temperature. There are also currently no long-range electrostatic interactions and implicit screening implies that the interaction strength of polar substances is under-estimated in a low dielectric medium such as a non-polarisable solvent. Due to the loss of the fine-grained degrees of freedom, the CG potential energy surface is smoother and results in more sampling, speeding up the kinetics of the system. This is generally the advantage of CG simulations, but the speed-up is not constant, nor easily predictable. An average speed-up factor of around 4 has been found for MARTINI.

References

(1) Jensen, F. *Introduction to Computational Chemistry;* John Wiley and Sons Ltd: Chichester, 1999.

(2) Hehre, W. J. *A Guide to Molecular Mechanics and Quantum Chemical Calculations;* United States of America: Wavefunction Inc., 2003.

(3) Grant, H.; Richards, W. G. *Computational Chemistry;* Oxford University Press: New York, 1995.

(4) Cramer, C. J. *Essentials of Computational Chemistry: Theories and Models;* 2nd ed.; John Wiley & Sons Ltd: Chichester, 2008.

(5) Foresman, J. B.; Frisch, E. *Exploring Chemistry with Electronic Structure Methods;* 2nd ed.; Gaussian Inc: Pittsburgh, 1996.

(6) Chai, J.-D.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6515.

(7) Chai, J.; Head-Gordon, M. *J. Chem. Phys.* **2008**, *128*, 084106.

(8) Grimme, S. *J. Comput. Chem.* **2006**, *27*, 1787.

(9) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.

(10) George Kaminski, E. M. D., Tooru Matsui, William L. Jorgensen *J. Phys. Chem.* **1994**, *98*, 13077.

(11) Damm, W.; Frontera, A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Comput. Chem.* **1997**, *18*, 1955.

(12) Siu, S. W. I.; Pluhackova, K.; Böckmann, R. A. *J. Chem. Theor. Comput.* **2012**, *8*, 1459.

(13) Kony, D.; Damm, W.; Stoll, S.; Van Gunsteren, W. F. *J. Comput. Chem.* **2002**, *23*, 1416.

(14) Mackerell, A. D. *J. Comput. Chem.* **2004**, *25*, 1584.

(15) Shinoda, W.; Susumu, O. *J. Mol. Liq.* **2001**, *90*, 95.

(16) Groenewald, W.; Baird, M. S.; Verschoor, J. A.; Minnikin, D. E.; Croft, A. K. *Chem. Phys. Lipids* **2014**, *180*, *15*.

(17) Wang, L. P.; Van Voorhis, T. *J. Chem. Phys.* **2010**, *133*.

(18) Muller, M.; Katsov, K.; Schick, M. *Phys. Rep.* **2006**, 113.

(19) Marrink, S. J.; Tieleman, D. P. *Chem. Soc. Rev.* **2013**, *42*, 6801.

(20) Kamerlin, S. C. L.; Vicatos, S.; Dryga, A.; Warshel, A. *Annnu.Rev. Phys. Chem.* **2011**, *62*, 41.

(21) Marrink, S.; Risselada, H.; Yefimov, S.; Tieleman, D.; de Vries, A. *J. Phys. Chem. B* **2007**, *111*, 7812.

(22) Periole, S.; Marrink, S. J. *Methods Mol. Biol. (N. Y.)* **2013**, *924*, 533.

Chapter 3: Introduction to Manuscript I

"Do not trust atoms, they make up everything."
Unknown

Cyclopropane is the smallest cyclic carbon ring that can be formed. A wide range of cyclopropane-containing compounds is known to date.[1,2] Nature abounds in all kinds of cyclopropanes, ranging from simple molecules with single cyclopropanes to very complex frameworks containing multiple cyclopropanes. Organic chemistry has developed to the extent that even the most complex cyclopropane-containing compounds can be synthesised today.[1-3] It is important to be able to synthesise these compounds, as they have been shown to possess a myriad of desirable properties that can be developed and used.[1]

Genetic modification of *M. tb* strains has made it possible to study mycobacteria without certain enzymes in the mycolic acid (MA) biosynthetic pathway. In this way a significant quantity of data has been obtained on the effects of cyclopropanation of MAs on viability, cell-wall integrity, and virulence of these bacteria, as discussed in chapter 1.[4-8]

Cyclopropanes have triggered interest in theoretical studies due to their unusual structure and apparent strain,[9-12] and intriguing conformations.[13-15] Various *ab initio* quantum mechanics (QM), coupled cluster and density functional theories (DFT) have been used to model these cyclopropyl-containing compounds. Specific to MAs, cyclopropane fragments with with short 4-5-carbon alkyl chains with *cis*-and α-methyl *trans*-cyclopropanes were modelled using B3LYP/6-31G(*d*) level of theory.[16] Molecules were pre-folded into either a hairpin U-shape or open-chain straight conformations and the potential energies were compared. It was found that, for *cis*-cyclopropane, the open-chain conformation was 7.61 kJ.mol^{-1} (1.82 kcal.mol^{-1}) lower in energy than the U-shape, whereas for α-methyl *trans*-cyclopropane the U-shape conformer was 3.81 kJ.mol^{-1} (0.91 kcal.mol^{-1}) lower in energy than the open-chain conformer.[16] These results

showed that the α-methyl *trans*-cyclopropane facilitates folding at the cyclopropane group more than *cis*-cyclopropane.

In the work presented here, small model compounds with propyl chains were also used. The effect of cyclopropane stereochemistry on the folding at the cyclopropane group was investigated. *Trans*-cyclopropane was studied without an α-methyl group, as well as both diastereomers of α-methyl *trans*-cyclopropane. It was found that the natural homologue with *S*-stereochemistry at the α-methyl group (relative to the fixed cyclopropane stereochemistry) promoted tighter folding with the alkyl chains in close proximity to each other. This finding supports observations from monolayer studies that show that MAs with an α-methyl *trans*-cyclopropane group fold into a tighter W-conformation.

In addition, the effect of interaction between the alkyl chains was investigated. The use of ωB97XD, a density functional including dispersion, was analysed, but found to over-estimate inter-chain interactions. No stabilising contribution of interaction between the chains was observed when using MP2 level of theory, however, increased chain lengths may highlight these interactions better.

Although only the cyclopropane stereochemistry was examined here, the work can be expanded to the methoxy-methyl, keto-methyl and mycolate moeieties and will be likely to shed additional light onto preferred MA conformations around these functional groups.

References

(1) Salaun, J.; Baird, M. S. *Curr. Med. Chem.* **1995**, *2*.

(2) de Meijere, A.; Kozhushkov, S. I.; Schill, H. *Chem. Rev.* **2006**, *106*, 4926.

(3) Chen, D. Y. K.; Pouwer, R. H.; Richard, J. A. *Chem. Soc. Rev.* **2012**, *41*, 4631.

(4) Glickman, M. S.; Cox, J. S.; Jacobs, W. R. *Mol. Cell* **2000**, *5*, 717.

(5) Glickman, M. S.; Cahill, S. M.; Jacobs, W. R. *J. Biol. Chem.* **2001**, *276*, 2228.

(6) Rao, V.; Fujiwara, N.; Porcelli, S. A.; Glickman, M. S. *J. Exp. Med.* **2005**, *201*, 535.

(7) Rao, V.; Gao, F.; Chen, B.; Jacobs, W. R.; Glickman, M. S. *J. Clin. Invest.* **2006**, *116*, 1660.

(8) Barkan, D.; Hedhli, D.; Yan, H.-G.; Huygen, K.; Glickman, M. S. *Infect. Immun.* **2012**, *80*, 1958.

(9) Cremer, D.; Gauss, J. *J. Am. Chem. Soc.* **1986**, *108*, 7467.

(10) Schleyer, P. V.; Williams, J. E.; Blanchard, K. R. *J. Am. Chem. Soc.* **1970**, *92*, 2377.

(11) Gauss, J.; Cremer, D.; Stanton, J. F. *J. Phys. Chem. A* **2000**, *104*, 1319.

(12) Novak, I. *Tetrahedron Lett.* **2010**, *51*, 2920.

(13) Greenberg, A.; Liebman, J. F. *J. Am. Chem. Soc.* **1981**, *103*, 44.

(14) Stolevik, R.; Bakken, P. *J. Mol. Struct.* **1989**, *197*, 137.

(15) Aleman, C. *J. Phys. Chem. B* **1997**, *101*, 5046.

(16) Villeneuve, M.; Kawai, M.; Horiuchi, K.; Watanabe, M.; Aoyagi, Y.; Hitotsuyanagi, Y.; Takeya, K.; Gouda, H.; Hirono, S.; Minnikin, D. E. *Microbiology* **2013**, *159*, 2405.

Chapter 4: Introduction to Manuscript II

"I tell my students to try to know molecules, so well that when they have some question involving molecules, they can ask themselves, What would I do if I were that molecule?"
G. Wald

In this part of the work, presented in manuscript II, atomistic molecular dynamics (MD) is performed on single mycolic acids (MAs). The three main components of the three classes of MAs from *Mycobacterium tuberculosis* are modelled. Additionally a control molecule without functional groups in the mero-chain is modelled to aid in the elucidation of the effects of the functional groups on folding. Each molecule is modelled in the gas phase, in water (a polar solvent) and in hexane (a nonpolar solvent). MA conformations are explored in terms of their chemical composition as well as the external effect of the solvent (or lack thereof).

This work is complementary to previous MD on MAs. In comparison to the work done by Villeneuve and co-workers,[1,2] this work differs in the following ways:

- No restrictions were placed on MA conformations in the simulations.
- The simulations started from a variety of conformations, taken from different frames of an initial simulation, in contrast to initial structures that were folded into a W-conformation.[1,2] Our approach harnesses the flexibility of MAs in order to improve sampling of the potential energy surface, whereas the approach by Villeneuve and co-workers is a measure of the molecule's preference for the W-conformation it was in, or for unfolding.

In comparison to all previous modelling on MA, this study is unique in the following ways:

- Longer timescales are used for improved sampling. Twenty replicate MD runs of 10 ns each were performed for each molecule, providing a

significant extension of timescale from a picosecond simulation length used by others[1,2] and more than doubled on previous simulation lengths of 4 ns.[3]

- A control molecule without any functional groups in the mero-chain is included.
- Polar and nonpolar solvents are included explicitly.

Although the systems modelled here are artificial and cannot directly relate to the biological context, it is efficient at illustrating the effect of functional groups with specific stereochemistry on MA conformations due to the fact that whole molecules and absolute stereochemistry can be accommodated. Simulations with explicit solvent are much more computationally intensive, and therefore the system size can be kept relatively small by only including single molecules. The solvents have been chosen to represent both polar and nonpolar environments as approximations of the aqueous extracellular matrix in the biofilm-like pellicle surroundings, and the nonpolar cell wall context, respectively.

This work can be expanded in future by altering stereochemistry of MA functional groups and modifying alkyl chain lengths to produce series of chemical structures, correlated with their conformational data. In addition, simulations can be scaled up to include more MAs in which MA interactions can be monitored. The immune-active compound trehalose dimycolate, abundant in the mycobacterial cell wall, provides the additional dimensions of two MAs (of several different combinations) interacting with one another as well as the effect of the trehalose sugar on the conformations expected.

Modelling all possible isomers of alpha mycolic acid

An example of the further expansion of the MD approach described in manuscript II would be to model all the possible combinations of cyclopropane isomers for alpha-MA (AMA, Figure 11). The stereochemistry for AMA in natural MAs has been confirmed to be as shown for AMA1 in Figure 11. The effect of the altered stereochemistries on overall MA conformation – if any – would be noteworthy, as it not only points out effects of individual functional

group orientations, but also their stereochemistry relative to each other and to the head group.

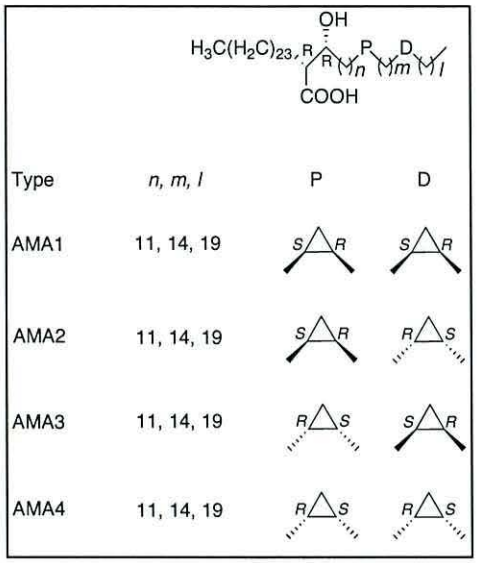


Figure 11: Chemical structures of all possible AMAs isomers. All MAs are identical except with respect to the cyclopropane orientations. Proximal and distal cyclopropane groups are represented by *P* and *D*, respectively.

A preliminary study has indicated that the conformational difference between these AMA isomers is smaller than that detected between the different MAs. However, subtle differences have been noted between isomers. For example, in Table 1, WUZ-conformations obtained for the various AMA isomers look very similar. However, on closer inspection it can be seen that in water, AMA2 shows a slightly increased tendency for the W-conformation (13.2 %) as compared to the natural isomer (9.0 %). The change in stereochemistry at the distal group may increase the folding at the distal cyclopropane, as indicated by the slight increase in aU and sZ conformations that fold at this point. This is accompanied with a correlation in lowered aZ conformations for AMA2 in comparison to AMA1, presumably due to the lack of stretching at the distal cyclopropane. AMA3, differing from AMA1 at the proximal cyclopropane stereochemistry, shows increased folding at the distal cyclopropane with a coupled extension at the proximal cyclopropane in water, with a higher percentage of aU conformations (1.6 % versus 0.2 %). The decrease in aZ-conformations from 5.4 % in water for

AMA1 to 1.0 % for AMA3 also suggests a decreased folding at the altered proximal cyclopropane group. With the change in both the proximal and distal cyclopropane groups decreasing the percentage of aZ conformations, it may be expected that for AMA4, in which both cyclopropanes have altered stereochemistry, this will be reduced even further. Unexpectedly the aZ conformations for AMA4 in water are 2.4 %. The stereochemistry of the cyclopropanes in relation to each other is the same in AMA1 and AMA4, which has restored some of AMA1 properties in AMA4, although AMA4 still differs in respect of the cyclopropane stereochemistry in relation to the head group.

These preliminary results again establish the cyclopropane group as a bending point in AMA. It is clear that the WUZ-conformations do not account for the majority of conformations of the single AMAs. More light is shed on MA conformations other than WUZ conformations in manuscript II.

Table 1: The percentage of WUZ-conformers obtained for each AMA isomer modelled in vacuum, water and hexane.

MA	Solvent	W	aZ	eZ	sZ	eU	sU	aU	Total
	Vacuum	7.9	1.6	0.2	0.5	0.0	1.0	0.3	11.5
AMA1	Water	9.0	5.4	0.0	0.3	0.1	0.0	0.2	15.0
	Hexane	0.0	0.5	0.1	0.3	2.7	1.0	5.8	10.4
	Vacuum	7.9	2.5	0.8	1.4	0.4	1.2	0.9	15.1
AMA2	Water	13.2	2.6	0.1	0.0	0.4	0.4	0.4	17.1
	Hexane	0.0	0.3	0.2	0.4	3.1	0.7	6.0	10.6
	Vacuum	9.4	0.8	0.2	0.6	0.0	0.4	0.1	11.5
AMA3	Water	8.4	1.0	0.1	0.2	0.6	0.2	1.6	12.2
	Hexane	0.0	0.3	0.1	0.4	2.8	0.8	5.9	10.3
	Vacuum	8.2	1.0	0.2	0.5	0.0	0.6	0.0	10.6
AMA4	Water	10.9	2.4	0.1	0.3	0.0	0.1	1.1	14.9
	Hexane	0.0	0.3	0.1	0.5	2.9	0.7	6.7	11.2

Tabu-Search algorithms for the determination of the global minimum for mycolic acids

A different approach was used for the conformational searching of MAs, specifically aimed at finding the global minimum. This approach is valuable for

conformational searching of large molecules with many degrees of freedom, in which a systematic approach (such as the approach presented in manuscript I) is not feasible. CAST (Conformational Analysis and Search Tool)[5] utilises a Tabu-Search algorithm for which a simplified scheme is shown in Figure 12. The global minimum is determined by using local optimisation methods (for example, from point 1 to 2 and point 3 to 4 in Figure 12) and modest ascent (for example point 2 to 3 in Figure 12).

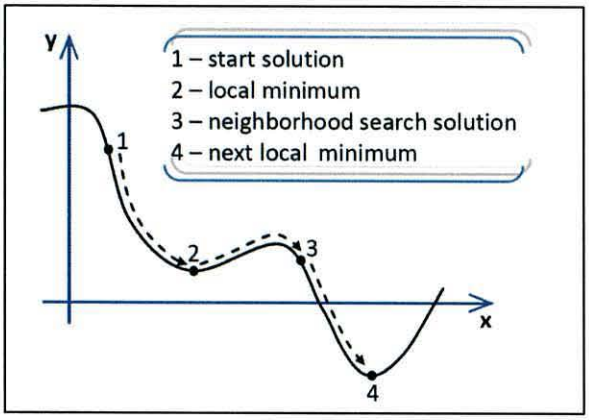


Figure 12: Scheme of Tabu-Search based methods. Taken from the CAST user guide.[4]

All MAs had a W-conformation (folding at each of its functional groups) as the lowest energy conformation from the Tabu-search. KMA only had W-conformations as the final global minimum for each of three replicated conformational searches. AMA and MMA included other structures as the global minimum from separate conformational searches such as the sZ and sU shapes (symmetrical shapes bending at two or one functional group, respectively), as shown for the three search results for AMA in Figure 13. However, for AMA and MMA the W-conformation still had the lowest energy. The control molecule without any functional groups in the mero-chain had a lowest energy conformer in the sZ shape. The results from the Tabu-Search establishes W as the lowest energy conformer identified up to this point, implying the functional groups in the MA as bending points in the molecule. It also confirmed that close proximity of the alkyl chains in parallel contributes to stabilisation, resulting in low energy conformations.

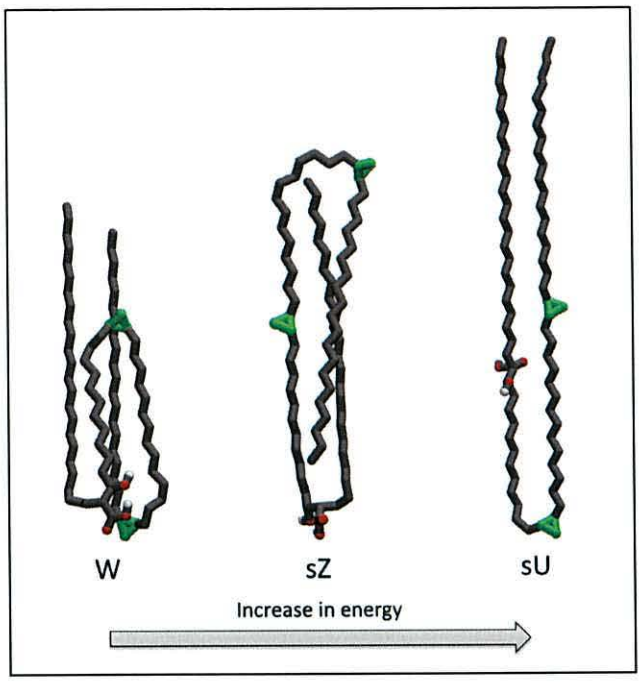


Figure 13: Conformations obtained as global minima for AMA from three separate conformational searches using CAST.

Acknowledgements

Eleri Jones is acknowledged for her contribution to the work on the AMA isomers in her final year undergraduate project, as is Christoph Grebner for his help with the implementation of Tabu-search algorithms.

References

(1) Villeneuve, M.; Kawai, M.; Watanabe, M.; Aoyagi, Y.; Hitotsuyanagi, Y.; Takeya, K.; Gouda, H.; Hirono, S.; Minnikin, D. E.; Nakahara, H. *Biochim. Biophys. Acta, Biomembr.* **2007**, *1768*, 1717.

(2) Villeneuve, M.; Kawai, M.; Watanabe, M.; Aoyagi, Y.; Hitotsuyanagi, Y.; Takeya, K.; Gouda, H.; Hirono, S.; Minnikin, D. E.; Nakahara, H. *Chem. Phys. Lipids* **2010**, *163*, 569.

(3) Groenewald, W.; Baird, M. S.; Verschoor, J. A.; Minnikin, D. E.; Croft, A. K. *Chem. Phys. Lipids* **2014**, *180*, *15*.

(4) Grebner, C. *CAST Conformational Analysis and Search Tool User Guide*. **2011**.

(5) Grebner, C.; Becker, J.; Stepanenko, S.; Engels, B. *J. Comp. Chem.* **2011**, *32*, 2245.

Chapter 5: Introduction to Manuscript III

"Make everything as simple as possible, but not simpler"
Albert Einstein

It has been shown that atomistic detail is not always required in order to have a good model that maintains the fundamental physical characteristics of the system.[1-3] However, it does mean that the exact stereochemistry of mycolic acid (MA) functional groups – which have been shown to be functionally important[4,5] - cannot be included explicitly in a coarse grained model. Parameterisation of the bonded interactions of the coarse grain (CG) model will include the stereochemical effects indirectly, but it remains to be determined whether this allows for the conformational characteristics to be well represented by the CG model.

The MARTINI forcefield originally started with work on lipids and has since shown good representation of many lipid systems.[6,7] MAs are different from usual membrane lipids that are most commonly modelled. The core of the MARTINI model, the simplicity of stringing together well-parameterised building blocks to create more complex molecules, therefore seems attractive in attempting to model different lipid structures.

The wide applicability and topologies available for MARTINI are also attractive in that this work can be expanded to more complete cell wall models for mycobacteria consisting of many complex structures that need to be better understood.

References

(1) Kamerlin, S. C. L.; Vicatos, S.; Dryga, A.; Warshel, A. *Annu. Rev. Phys. Chem.* **2011**, *62*, 41.

(2) Levitt, M.; Warshel, A. *Nature (London, U. K.)* **1975**, *253*, 694.

(3) Taketomi, H.; Ueda, Y.; Gõ, N. *Int. J. Pept. Res.* **1975**, *7*, 445.

(4) Beukes, M.; Lemmer, Y.; Deysel, M.; Al Dulayymi, J. R.; Baird, M. S.; Koza, G.; Iglesias, M. M.; Rowles, R. R.; Theunissen, C.; Grooten, J.; Toschi, G.; Roberts, V. V.; Pilcher, L.; Van Wyngaardt, S.; Mathebula, N.; Balogun, M.; Stoltz, A. C.; Verschoor, J. A. *Chem. Phys. Lipids* **2010**, *163*, 800.

(5) Vander Beken, S.; Al Dulayymi, J. R.; Naessens, T.; Koza, G.; Maza-Iglesias, M.; Rowles, R.; Theunissen, C.; De Medts, J.; Lanckacker, E.; Baird, M. S.; Grooten, J. *Eur. J. Immunol.* **2011**, *41*, 450.

(6) Marrink, S. J.; Tieleman, D. P. *Chem. Soc. Rev.* **2013**, *42*, 6801.

(7) Periole, S.; Marrink, S. J. *Methods Mol. Biol.* **2013**, *924*, 533.

Chapter 6: Concluding Discussion

Different aspects of mycolic acid (MA) conformation were investigated in this work by utilising different levels of computational theory. Each method utilised differs in its accuracy, computational cost and hence, the feasible system size. The objective in the use of quantum mechanics was to study small fragments of MAs in detail by doing systematic conformational searching, while the conformations assumed by single MAs in different solvents were studied using atomistic molecular dynamics. A coarse grained approach was utilised to examine large systems containing numerous MAs.

Quantum mechanics, the most accurate theory used in this work, has shed light on the energetically favoured conformations of small fragments of MAs containing cyclopropanes with varied stereochemistry. It was found that *cis*-cyclopropane had some conformers that were not viable due to the alkyl chains intersecting, while *trans*-cyclopropane and α-methyl *trans*-cyclopropane alkyl chains did not intersect. For all cyclopropane molecules studied it was found that it is energetically favourable for the dihedral angles of the alkyl chains adjacent to the cyclopropane group to be in either the *trans* or *gauche* orientation. The α-methyl group with *S*-stereochemistry for *trans*-cyclopropane facilitates a tighter hairpin fold with the alkyl chains in close proximity to each other, compared to the methyl group in the *R*-stereochemistry. The *S*-stereochemistry has been shown to be the isomer found in natural MAs with *trans*-cyclopropane, and hence this finding suggests that MA folding will be facilitated by the presence of the α-methyl *trans*-cyclopropane group.

In the molecular dynamics part of this work, keto-MA (KMA) was shown to prefer more folded conformations than alpha- and methoxy-MA (AMA and MMA, respectively), even in hexane where predominantly unfolded conformations were found. It showed a particular tendency to fold at its α-methyl *trans*-cyclopropane group. Conformations for all MAs were mostly

focused around more folded conformations in the gas phase and in water. The distribution of conformations varied for each MA structure, as shown by the different percentages of WUZ-conformations. All MAs had the highest percentage of W-conformations out of all the possible WUZ-conformations, and the W-conformation was shown to be the most stable. KMA had the highest percentage of WUZ-conformations, while the backbone control molecule without any functional groups in the mero-chain (BMA) had the least. The lack of folding in BMA underlines the natural tendency for the functional groups to act as bending points in the alkyl chains. Although straight alkyl chains in the *trans*-orientation were energetically favoured, such conformations are a minor component of single MA conformations. New conformations with numerous bends and twists in the alkyl chains were shown to be representative of the majority of conformations assumed in simulations of single MAs. These newly defined conformations may be of relevance to MAs in the mycobacterial biofilm.

In the coarse grain (CG) simulation of AMA, ordering of the alkyl chains was notable when MAs were tightly packed. The monolayer is thought to be a good depiction of the packing of MAs in the cell wall. A limitation of the CG model is that it does not incorporate the absolute stereochemistry of the MA functional groups, which seems key in interactions with components of the immune system.[1,2] Comparison with monolayer data at different temperatures showed accurate trends in the change of surface tension with temperature, as well as phase changes and the monolayer collapse point. Therefore the CG model seems to capture the vital characteristics of MA conformational behaviour. This approach therefore lends itself to permeability studies for applications in drug design, as well as an improved understanding of the mycobacterial cell wall through modelling its numerous complex components with a CG approach.

Each of the computational methods used in this study have focused on MA conformations on a different scale. Each method was utilised in balance determined by its accuracy and computational cost. In this way, information from the smallest, but most accurate theoretical system – the quantum mechanics calculations - could feed into the MD simulations in the form of cyclopropane parameters. On the atomistic scale, single molecules in solvent provided a feasible system for investigating the conformational behaviour of single MAs.

The atomistic conformations of single MAs, in turn, served as reference angle and bond distributions for the parameterisation of the CG model.

Diverse approaches to the investigation of MA conformations are presented in this work. It is envisaged that expansion of the CG model, with a view to creating a model of the complete, complex mycobacterial cell wall, will be a powerful weapon at the forefront of the battle against tuberculosis. Yet, each component of the work presented here is unique in its contribution and needs to be maintained in future work in order to complement and inform the other approaches. As demonstrated here, diverse approaches to the same research question have resulted in a flow of information and a synergism that produces a deeper understanding of the finer details, as well as an expansion to larger systems that are closer to the natural setting.

References

(1) Vander Beken, S.; Al Dulayymi, J. R.; Naessens, T.; Koza, G.; Maza-Iglesias, M.; Rowles, R.; Theunissen, C.; De Medts, J.; Lanckacker, E.; Baird, M. S.; Grooten, J. *Eur. J. Immunol.* **2011**, *41*, 450.

(2) Beukes, M.; Lemmer, Y.; Deysel, M.; Al Dulayymi, J. R.; Baird, M. S.; Koza, G.; Iglesias, M. M.; Rowles, R. R.; Theunissen, C.; Grooten, J.; Toschi, G.; Roberts, V. V.; Pilcher, L.; Van Wyngaardt, S.; Mathebula, N.; Balogun, M.; Stoltz, A. C.; Verschoor, J. A. *Chem. Phys. Lipids* **2010**, *163*, 800.

Manuscript I

Systematic Conformational Search for Low Energy Conformers of Cyclopropane Units From Mycolic Acids

Wilma Groenewald,[†] Gregory A. Chass[‡] and Anna K. Croft[†a]*

[†]*School of Chemistry, Bangor University, Bangor, Gwynedd, LL57 2UW, United Kingdom*

[‡]*School of Biological and Chemical Sciences, Queen Mary, University of London, Mile End Road, London E1 4NS*

[a]*Current address Department of Chemical and Environmental Engineering, University of Nottingham, University Park, Nottingham NG7 2RD, United Kingdom*

Abstract

In mycolic acids (MAs), the unusually long fatty acids found in *Mycobacterium tuberculosis*, *cis*- and *trans*-cyclopropanation have been shown to modulate the immune response. MAs with different structures have different effects on the immune system, with the structure-function relationships only beginning to be unravelled. In this work model compounds with short dipropyl alkyl chains and various stereochemistries are used to sample the conformational space of the alkyl chains adjacent to the cyclopropanes. An assessment of different theories reveals that Hartree-Fock theory is not sufficient to model dispersion, while density functional theory including dispersion (ωB97XD) over-estimates dispersion interaction between the alkyl chains. MP2 is shown to be the most accurate, and an analysis of the conformers for each molecule at this level afforded relative energies within 0 - 36.8 kJ.mol^{-1}. For all molecules considered, it was found that low energy conformers have χ^1 and $\chi^{1'}$ (the dihedral angles adjacent to the ring) in *trans*- or *gauche*-conformations. The stereochemistry of the adjacent methyl group in α-methyl *trans*-cyclopropane favours conformations that bring the alkyl chains closer together with the methyl group outside for the naturally occurring *S*-orientation, while low energy conformers

with *R*-orientation have the alkyl chains separated by the methyl group. These results support folding at the cyclopropane groups in MAs, with the *S*-orientation of the α-methyl group in α-methyl *trans*-cyclopropane promoting hairpin folding.

Introduction

In cell wall lipids, cyclopropanes are commonly found in gram-positive and gram-negative prokaryotes, but only occur in very few eukaryotes.[1] The role of cyclopropane in membrane lipids is still uncertain, but has been suggested to aid in late stage survival of prokaryotic organisms, due to its association with adverse conditions and later stages of growth. Effects on membrane fluidity have been examined, but remain controversial due to contradicting results.[2-4] Cyclopropanated lipids in a membrane result in an increased surface area per molecule, making the membrane less dense, due to the bulky cyclopropane group preventing tight packing.[2,5]

Mycolic acids (MAs) are one of the major lipid components in the *Mycobacterium tuberculosis* (*M. tb*) cell wall and create an impermeable barrier around it. The MA layer contributes to the robustness of the mycobacterium and makes it impenetrable to drugs, necessitating treatment regimes of more than 6 months. MAs can be up to 90 Carbons long and contain *cis*- and α-methyl *trans*-cyclopropanes, as shown in Figure 1 at position X. Alpha-MAs (AMAs) also have a cyclopropane at position Y, and both are *cis*-cylcopropanes. On the other hand, Methoxy- and Keto-MAs (MMA and KMA, respectively) may have either *cis*- or α-methyl *trans*-cyclopropanes at position X, with methoxy and keto-methyl groups at position Y, respectively. Recent work has shed light on the role that the cyclopropanes in MAs play in cell wall permeability, virulence, and modulating the host immune response.

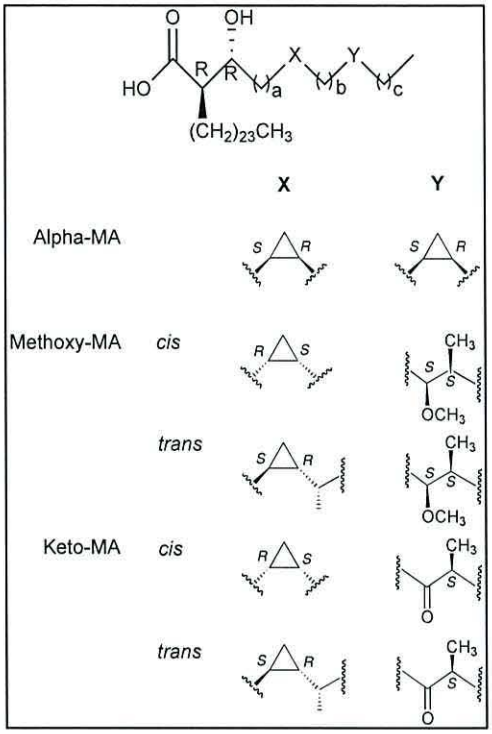


Figure 1: Structure of the main components of mycolic acids in *M. tb* with proximal and distal functional groups X and Y, respectively.

Illustrating the importance of cyclopropanation to the biological activities associated with MAs, a mutant strain of *M. tb* lacking pcaA, the enzyme that synthesises the proximal cyclopropane group of AMAs (Figure 1, position X), has been examined. Lack of this enzyme results in unsaturation in AMAs at the proximal position, as well as an abundance of KMA. The altered MA composition changes the course and outcome of a TB infection dramatically; in mice, growth was attenuated initially and lacked persistence and fatality in later stages.[6,7] Trehalose dimycolates (sugar esters of MAs) extracted from the mutant strain also showed less inflammation in macrophages and less severe granulomas in mice, as compared to the wild type.[7] In contrast, lack of cmaA2, which *trans*-cyclopropanates oxygenated MAs, resulted in hypervirulence and more pronounced inflammatory responses compared to the wild-type,[8,9] implying that *trans*-cyclopropanation in MAs curbs inflammation and virulence

during infection. Glickman and coworkers have recently created a mutant strain without any cyclopropanation,[10] which showed attenuation in the first week of infection in mice, but progressed to show hypervirulence and strong antigen-specific immune responses. Therefore, the overall effect of cyclopropanation, as noted before for *trans*-cyclopropanation, was to suppress the immune response to *M. tb*. These results suggest that the precise composition of cell wall mycolates requires a fine balance to result in a successful infection and that cyclopropanation is an immune modulating tool in this process.

Studies using synthetic MAs have been published,[11,12] with the advantage of utilising pure compounds in contrast to naturally extracted MAs, which consist of heterogeneous mixtures of alkyl chain lengths and functional group stereochemistries. Synthetic compounds also allow for systematic unpicking of structure-function relationships in MAs. Using a range of structurally diverse synthetic MAs to measure antibody recognition in TB patient sera it was shown that, in general, oxygenated MAs are more antigenic than AMA, and that *trans*-cyclopropanation in oxygenated MAs improved antigenicity as compared to the *cis*-isomers.[11] The airway inflammatory pattern of MAs in mice showed a dependence on proximal and distal cyclopropane groups and their specific stereochemistry.[12] While AMA was inert, *cis*-cyclopropanated MMA and KMA showed intermediate inflammatory responses. The *trans*-isomer of MMA lost most of its inflammatory action, while the KMA *trans*-cyclopropane isomer switched to an anti-inflammatory action that could counteract inflammation. In addition to being very important immune modulators of the host, cyclopropanation of MAs influences interaction with key components of the host immune system.

MAs are very long, flexible molecules and their conformations involved in the cell wall or in immunological responses are not known. The functional groups have been proposed to form bending points in the MA molecule, and recent cryo-electron microscopy data has shown that some MAs are too long to fit into the outer membrane without folding.[13,14] Zuber and co-workers[13] have proposed a zipper-model in which MAs, covalently attached to arabinogalactan in the cell

wall, not only need to fold at each of their functional groups, but also intercalate with lipids in the outer leaflet in order to fit into the 7 - 8 nm space available.

We have shown that single MAs simulated *in vacuo* spontaneously assume conformations in which the molecules are folded at the functional groups, primarily driven by van der Waals interactions between the alkyl chains.[15] Villeneuve *et al.*[16,17] have modelled MAs folded at each of the functional groups, forming a "W"-shape in two dimensions (Figure 2) and found that factors like relative chain lengths and mero-chain functional groups may affect the tendency of the lipids to stay folded, or to unfold. In particular, the α-methyl *trans*-cyclopropane group has been shown to promote folding more than the *cis*-cyclopropane group.[18] Alkyl chains in MAs need to be in close proximity and parallel to each other for such folding to occur. We propose that the cyclopropane group plays an important part in directing the folding of the alkyl chains adjacent to it, and therefore that the stereochemistry is important. Additionally, van der Waals interactions - specifically London dispersion forces involving induced dipoles between the non-polar alkyl chains - play a pivotal role in making such folds feasible.

In this work the cyclopropane group is specifically investigated as a folding point in MAs by using systematic conformational searching for low energy conformers in models with contracted, adjacent alkyl chains. Quantum mechanics (QM) calculations were used to systematically search through the conformational space of short-chain cyclopropanyl derivatives **1-4** (Figure 3). QM *ab initio* computations provide structural detail at the atomic level. Searching and optimising all possible alkyl chain conformers has allowed us to identify the stable, low-energy conformations. The differences between *cis*- and *trans*-isomers are highlighted, as well as the changes that an adjacent methyl group in α-methyl *trans*-cyclopropane-containing MAs (**3** and **4**) brings about.

Figure 2: Schematic representation of MA in a W-fold in 2 dimensions, with proximal and distal functional groups at X and Y, respectively.

Methods

Systematic conformational searching was performed on the alkyl chains of four model systems containing cyclopropane (Figure 3, **1-4**), namely *cis*-dipropylcyclopropane (**1**), *trans*-dipropylcyclopropane (**2**) as well as the *S* and *R* diasteriomers of α-methyl *trans*-dipropylcylopropane molecules. *Cis*-cyclopropane, as illustrated in structure **1**, is found in all three MA types in *M. tb*, whereas *trans*-cyclopropanes are only found in oxygenated MAs. It is thought that α-methyl *trans*-cyclopropane with the α-methyl group in the *S*-orientation (**3**) is found in nature.[19,20] Therefore the opposite diasteriomer, **4**, as well as a *trans*-cyclopropane without a methyl group, **2**, were also included in order to examine the effect of the methyl group and its orientation on conformation.

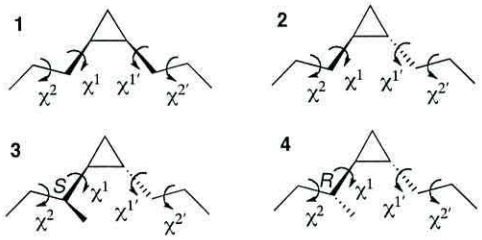


Figure 3: Cyclopropane fragments with labelled dihedrals that were used in the systematic conformational search. (**1**) (1*R*,2*S*)-1,2-dipropylcyclopropane, (**2**) (1*S*,2*S*)-1,2-dipropylcyclopropane, (**3**) (1*R*,2*S*)-1-((*S*)-*sec*-butyl)-2-propylcyclopropane, (**4**) (1*R*,2*S*)-1-((*R*)-*sec*-butyl)-2-propylcyclopropane.

Conformational search strategy

A modular numbering system for the cyclopropane fragments was used, based on similar modular numbering systems used for phospholipids and triglycerides.[21,22] In this system, numbering is started at the cyclopropane (Figure 4), followed by the carbon atoms in the chains and their hydrogen atoms. The numbering is done symmetrically so that the last two hydrogens can easily be substituted for carbons in order to extend the alkyl chains simultaneously. Here, the benefit of using a modular numbering system is that all degrees of freedom are defined in a similar, systematic way using a Z-matrix (an example of the Z-matrix can be seen in the supporting information). This allows for the automated generation of all the input files because the dihedral definitions and their dependence on each other (as illustrated in Figure 5) are known and constant for each molecule and are therefore easily adjusted in the Z-matrix in order to generate all the conformers. All dihedrals in the alkyl chains were varied except the terminal ones, which result in methyl group rotation. Each alkyl chain dihedral that was varied (χ^1, $\chi^{1'}$, χ^2, $\chi^{2'}$, Figure 3) has three predicted rotamers, namely *anti* (*a*), *gauche*$^+$ (g^+) and *gauche*$^-$ (g^-) as shown in Figure 5. The total number of possible conformations is the combination of the individual rotamers: three conformations per alkyl chain dihedral (a, g^+, g^-) for each of χ^1, $\chi^{1'}$, χ^2 and $\chi^{2'}$, giving 81 total conformations. The number of possible conformers increases by a factor of 9 for each additional carbon in the alkyl chains. This results in a steep exponential increase in the number of possible conformers and hence the feasible limit for calculation is reached quickly. The feasible limit is dependent on the level of theory used, computational resources and time available. Dipropylated molecules were used here, as they are the longest chains that are practical within the remit of this study.

Computational methods

All conformers were generated using a python script to modify the Z-matrix. The original Z-matrix that was used for input was obtained from a geometry-optimised structure at the relevant level of theory. Calculations were performed with Gaussian 09[23] using either RHF/*3-21g 6d*, MP2/6-31+G(*d,p*) with frozen core, ωB97XD/6-31+G(*d,p*) referred to as DFT-D, or CCSD/6-311+G(*d,p*) theoretical

methods, as stated. The calculations were given a maximum of 500 cycles and, in the case of no convergence within that period, either the number of cycles was increased or a slight modification in the starting structure was made in order to reach a stationary point. The structures were allowed to relax fully, and therefore the end conformation may differ significantly from the starting conformation. It should be noted that the complete conformational searches on molecules **1** and **2** include symmetrical molecules and therefore a number of conformers are identical structures within the limits of the accuracy of the optimisations.

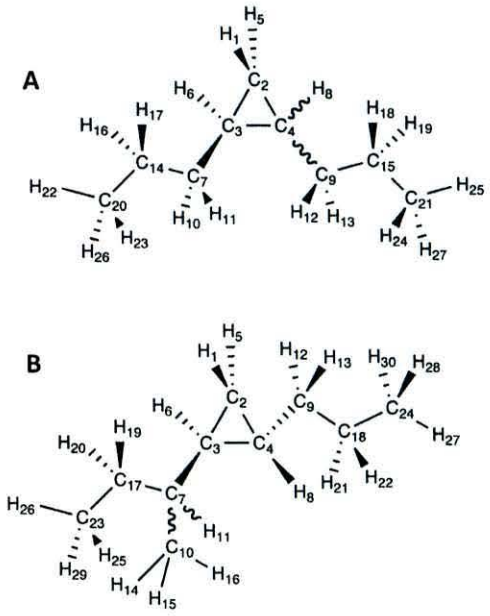


Figure 4: Modular numbering system for *cis*- and *trans*-diasteriomers of 1,2-dipropylcyclopropane (A) and the *R* and *S*-diasteriomers of the α-methyl *trans*-cyclopropyl molecules (B).

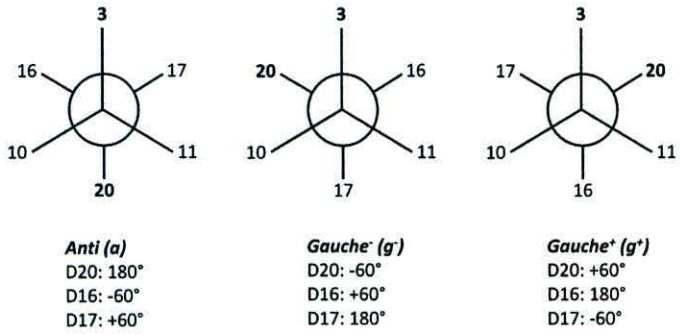


Figure 5: Newman projections of dihedral angle 20 (D20) for *cis*-1,2-dipropylcyclopropane at +60°, -60° and 180° and its relation to D16 and D17. In the Z-matrix D20 is defined by atoms C3, C7 (front carbon in Newman projection) C14 (rear carbon in Newman projection) and C20. D16 and D17 were defined similarly with atoms 16 and 17, respectively.

Results

Assessment study

An assessment study was carried out on *cis*-dipropylcyclopropane **1** using different theoretical methods, namely HF, MP2 and DFT with dispersion correction (DFT-D). From the resulting lowest energy conformers for **1**, as calculated with the HF approach, starting conformers with chains close to each other (Figure 6, A, D and G) were seen to open up (the chains moving away form each other, Figure 6, B, E and H). The starting conformers shown in Figure 6, which were among the resulting eight lowest energy conformers using HF, give a different end conformer when using MP2 theory than when using the HF method. In the MP2 calculations the chains did not open as much as when using HF (Figure 6, C, F and I). This may be due to electron correlation, which is included in MP2 calculations but not in HF. In addition, the relative energies of all conformers differ, when comparing the HF to the MP2 results. Changes in order of energy between different levels of theory has also been observed by others.[24,25] The difference in relative energies is most noticeable for the three conformers that give different end structures, shown in Figure 6; they are among the lowest energy conformers for the HF results with relative energies ranging

from 0.002 to 0.22 kJ.mol^{-1}, but are among the high-energy conformers in MP2 results, being 15.1 to 29.44 kJ.mol^{-1} higher in energy than the lowest MP2 conformers. Although some interaction is seen between the chains when using MP2 level of theory, this interaction does not serve to stabilise the structures, as indicated by the higher relative energy.

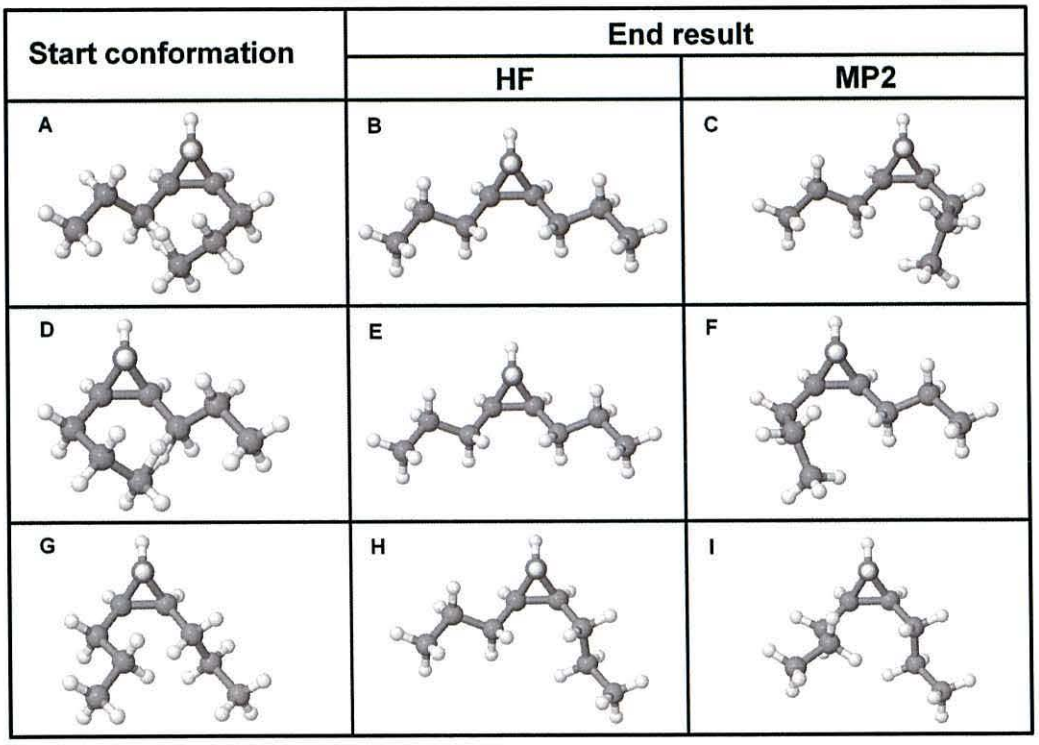


Figure 6: Different results obtained with RHF/3-21g 6d (B, E and H) and MP2/6-31+G(d,p) (C, F, and I) theories from the same starting structures, namely, (A) "++a+", (B) "---a" and (C) "–aaa" for $\chi^1, \chi^{1'}, \chi^2, \chi^{2'}$, respectively.

Since it is likely that there will be London dispersion forces involved if any favourable inter-chain interaction is seen, it was deemed necessary to include these in the conformational searches. MP2 is much more costly in computational time than HF and therefore DFT-D was tested as a quicker method that also

includes a treatment of dispersion. Although the dispersion coefficient of the dispersion term that is added on in DFT-D has a very theoretical basis, being derived from the London dispersion formula and calculated using ionisation potentials and static dipole polarisabilities, the functional itself is parameterised on a training set of empirical data, and therefore it was uncertain how well this model would perform on the uncommon cyclopropane system.

When the conformers that result in the ten lowest energy structures using MP2 theory were also optimised using DFT-D, noticeable differences were seen in two of the final structures (Figure 7). Here, the MP2 conformations (Figure 7, B and E) show more separation between the alkyl chains than the DFT-D conformations (Figure 7, C and F). The HF results for these two conformers were similar to the MP2 results. These results raise the question whether MP2 is in fact underestimating the inter-chain dispersion-interactions because it is giving the same results as HF theory, which neglects these effects, or whether the DFT-D dispersion protocol is in fact over-compensating and affording an interaction stronger than it is in reality.

The training- and test-set results for non-covalent interaction for ωB97XD,[26] as determined by comparing to intermolecular distances obtained with MP2, show small mean average errors and negative mean signed errors. ωB97X (without empirical dispersion) shows larger mean average errors and positive mean signed errors, which may indicate an underestimation of intermolecular distances in some non-covalently bound complexes (relative to MP2) due to the added dispersion correction in ωB97XD, although it gives the best overall description of such non covalent interactions.[26] For transition metal and metal compounds, the general overestimation of all internuclear distances was decreased when using functionals including dispersion correction and ωB97XD gave more compact structures that gave the best correlation with experiment.[27,28] Accurate results have also been obtained with ωB97XD for π -conjugated oligomers[29] and strained four-membered rings.[30]

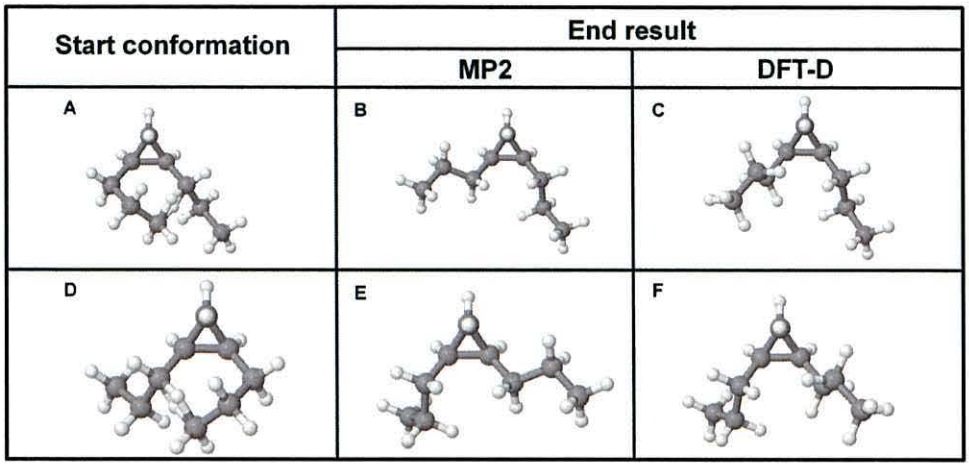


Figure 7: Different results obtained with MP2/6-31+G(d,p) (B and E) and ωB97XD/6-31+G(d,p) (C and F) level of theory from the same starting structures, namely (A) "-a-a" and (B) "a+-+" for $\chi^1, \chi^{1'}, \chi^2, \chi^{2'}$, respectively.

Unlike DFT, MP2 shows a dependence on basis set size to describe dispersion interactions since the individual functions cannot be polarised individually. As the basis set increases and becomes more flexible, polarisation emerges.[31] This effect often leads to under-estimation of dispersion forces with low basis sets and over-estimation with large basis sets.[32] MP2 has shown pleasing results with non-bonded interactions such as hydrogen bonding,[33] but showed a trend in over estimating π -stacking[31-33] and did not show accurate results for aliphatic-aliphatic interactions.[32] However, small to medium size basis sets can give good results for geometries and stabilisation energies due to the cancellation of errors in the over-estimation of stabilisation energy due to the basis set size and under-estimation due to neglect of higher orders of correlation energy contributions.[32] On calculations of relative energies of benzene isomers (of which several contain cyclopropane rings), MP2 was found to correlate well with CCSD(T).[34]

In order to determine which result is more accurate, single point (SP) calculations were performed with MP2/6-31+G(*d,p*), ωB97XD/6-31+G(*d,p*) and a more accurate coupled cluster theory CCSD/6-311+G(*d,p*) on the geometries previously obtained from MP2 and DFT-D. The SP energies of the different

structures at the same level of theory provide a scale on which to compare the two different outcomes relative to each other. SP energies (Table 1) revealed that the difference in energy of the two conformations obtained for MP2, DFT-D and CCSD for each molecule differ significantly (> 9 kJ.mol^{-1}) for each method used, with the MP2 geometries consistently lower in energy. The SP energies show that the long-range interactions in DFT-D are in fact over estimated, as they do not stabilise the structure more as compared to the MP2 result showing less correlation effect. For this reason the purely theoretical MP2 method was chosen here, despite its longer computation times.

Table 1: Relative single point energies for two *cis*-dipropylcyclopropane conformers with different theories.

Molecule Geometry		Single point ΔE (kJ.mol^{-1})		
		MP2[a]	DFT-D[b]	CCSD[c]
-a-a	MP2	0	0	0
	ωB97X-D	13.0	9.9	11.4
a+-+	MP2	0	0	0
	ωB97X-D	14.1	10.5	12.4

[a] MP2/6-31+G(*d*,*p*), [b]ωB97XD/6-31+G(*d*,*p*), [c]CCSD/6-311+G(*d*,*p*)

The different theories show different distributions of energies, as shown in Figure 8. Between HF and MP2, the difference is in the higher energy distributions. With HF theory most conformers have relative energies below 10 kJ.mol^{-1}. MP2 is the only method that has a range of energies extending to more than 30 kJ.mol^{-1}, with exception of the 20 – 25 kJ.mol^{-1} range. DFT has a more equal distribution between the different energy ranges. This could indicate that inclusion of dispersion in DFT-D has resulted in a "smoother" and "flatter" potential energy surface (PES), with the smaller differences in high and low energy conformers allowing for a more equal distribution among them. It has been noted before that with HF theory and a small basis set, the PES is not flat enough.[24,25] Thus, the HF PES may be an over-simplification due to the lack of correlation effects. The MP2 PES with correlation included by perturbation

theory seems to be between the HF and DFT-D PESs, with more conformers at higher energies than HF, but also more conformers at the lowest energy as compared to DFT-D.

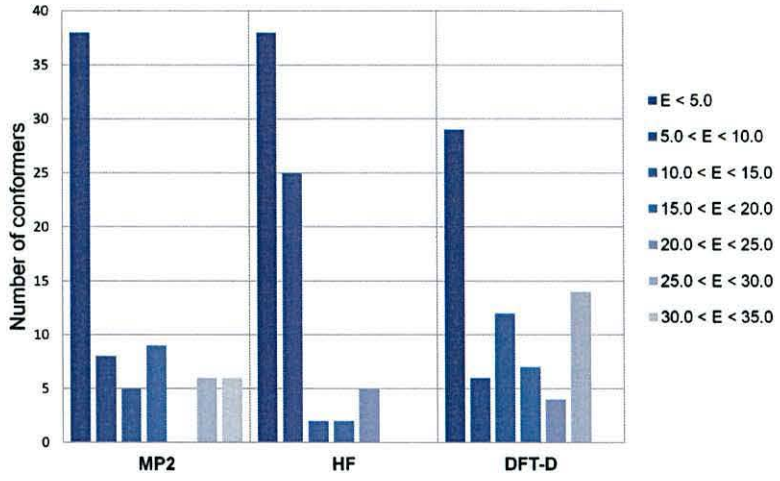


Figure 8: Comparison of energy distribution among the end results of the 72 conformers of cis-dipropylcyclopropane with different theories. Relative energies (E) in $kJ.mol^{-1}$.

Conformational Search Results

In order to refer to certain alkyl chain conformations, the terms upwards, inwards, downwards and sideways have been used, as depicted in Figure 9. The cyclopropyl CH(R)-CH(R) bond is used as reference point. The green dashed line passes through the CH(R)-CH(R) bond and when the alkyl chains are below this line, they are pointing downwards (right-hand side) or sideways (left-hand side). In other words, the alkyl chains extend in a direction that is opposite to the CH_2 of the cyclopropyl and away from the central CH(R)-CH(R) bond of the cyclopropane ring and these alkyl chain conformations result in low-energy conformers. When alkyl chains are above the green dashed line, they point upwards and these conformers have high-energies, as indicated by the shaded region. Alkyl chains pointing towards each other and towards the centre of the CH(R)-CH(R) bond so that their carbon atoms are close to, or inside the shaded

area between the magenta dotted lines will also result in high-energy conformers, and this is referred to as pointing inwards.

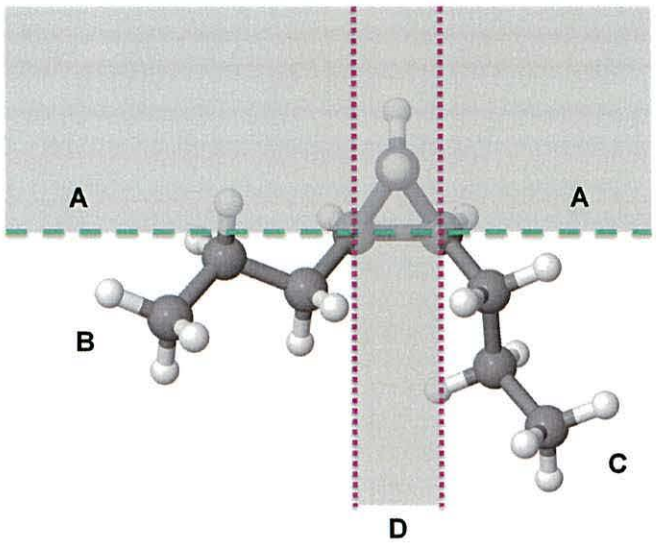


Figure 9: Representation of the terms used to describe different alkyl chain conformations. Shaded areas are higher in energy. High-energy "upward" conformation (A), low-energy "sideways" and "down" conformations (B and C, respectively) and "inward" pointing chain conformations (D).

Cis-dipropylcyclopropane

For the conformations calculated, *cis*-dipropylcyclopropane **1** experienced severe steric constraints that limited the number of accessible conformations. This was not seen for any of the *trans*-cyclopropanes **2-4**. The conformers that were not viable for the *cis*-isomer had the two chains intersecting with $\chi^1 = -60°$ and $\chi^{1'} = +60°$, which brings the chains toward each other in front of the cyclopropane ring, pointing inwards so that the distance between the two central carbons in the propyl chains is only ~0.14 nm and the hydrogen atoms on them overlap in the starting structure. Subsequently, nine conformers were inaccessible, resulting in 72 (out of a total of 81) viable structures. The resulting conformers ranged in energy from 0 to 32.2 $kJ.mol^{-1}$. (All conformers found with their

relative energies are listed in Table S1). The low-energy conformers are those with alkyl chains pointing downwards and/or sideways from the ring (Figure 10, A and B). This corresponds to three main conformer distributions were: ($\chi^1 \approx 90°$, $\chi^{1'} \approx 150°$), ($\chi^1 \approx -150°$, $\chi^{1'} \approx -90°$) and ($\chi^1 \approx 90°$, $\chi^{1'} \approx -90°$), as shown in Figure 11, A. There are also a few low-energy conformers at ($\chi^1 \approx -150$ to $-170°$, $\chi^{1'} \approx 150°$ to $170°$), ($\chi^1 \approx -160°$, $\chi^{1'} \approx -170°$) and ($\chi^1 \approx 170°$, $\chi^{1'} \approx 160°$). In contrast, high-energy conformers have either one, or both chains pointing upwards and/or inwards, (Figure 10, C and D). These conformers occur in two main conformer distributions with ($\chi^1 \approx -40°$, $\chi^{1'} \approx -90°$) and ($\chi^1 \approx 90°$, $\chi^{1'} \approx 40°$), as can be seen in Figure 11, A. There are also a few high-energy conformers with χ^1 and $\chi^{1'}$ values in close vicinity to 0°, implying that they are not far off from an eclipsed conformation at ($\chi^1 \approx -150°$ to $-180°$, $\chi^{1'} \approx -10$ to $40°$), ($\chi^1 \approx -40°$, $\chi^{1'} \approx 150°$), ($\chi^1 \approx 15°$, $\chi^{1'} \approx -90°$), ($\chi^1 \approx 8°$, $\chi^{1'} \approx 180°$) and ($\chi^1 \approx 90°$, $\chi^{1'} \approx -15°$). In contrast to χ^1 and $\chi^{1'}$, dihedrals χ^2 and $\chi^{2'}$ do not show a clear distinction between high and low energy conformers (Figure 12, A) with mostly staggered dihedrals obtained at $\chi^2 \approx \chi^{2'} \approx 180°$, 60°, -60° overall. There are only a minority of high-energy conformers that are outliers and occur at $\chi^2 \approx -100°$ or $\chi^{2'} \approx 100°$.

Trans-cyclopropanated molecules

Trans-dipropylcyclopropane **2** has 81 viable conformers and they range from 0 to 36.3 kJ.mol^{-1} in energy. (All conformers found with their relative energies are listed in Table S2). The lowest energy conformers, similar to those of **1**, have alkyl chains pointing downwards and sideways from the ring, while the high-energy conformers have alkyl chains pointing upwards and inwards (Figure 13). Likewise, high and low-energy conformers are distinct in χ^1 and $\chi^{1'}$ dihedral distributions. The low-energy conformers are spread in four main groups at ($\chi^1 \approx -150°$, $\chi^{1'} \approx 90°$), ($\chi^1 \approx -150°$, $\chi^{1'} \approx -150°$), ($\chi^1 \approx 90°$, $\chi^{1'} \approx 90°$), and ($\chi^1 \approx 90°$, $\chi^{1'} \approx -150°$) with a limited number in other conformations (Figure 11, B). On the other hand, high-energy conformers are found with ($\chi^1 \approx -150°$, $\chi^{1'} \approx -60°$), ($\chi^1 \approx -60°$, $\chi^{1'} \approx -60°$), ($\chi^1 \approx -60°$, $\chi^{1'} \approx -150°$), ($\chi^1 \approx -60°$, $\chi^{1'} \approx 90°$), and ($\chi^1 \approx 90°$, $\chi^{1'} \approx -60°$). There is no clear distinction in χ^2 and $\chi^{2'}$ preferences, which adopt mostly staggered conformations with $\chi^2 \approx \chi^{2'} \approx 180°$, 60°, -60° (Figure 12, B).

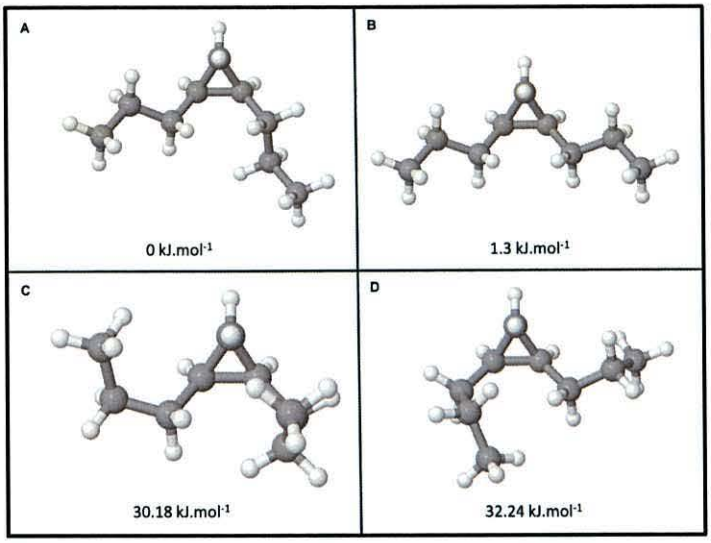


Figure 10: Low-energy (A and B, from "a-a-" and "+-aa" starting structures, respectively) and high-energy conformers (C and D, from "++--" and "----" starting structures, respectively) of **1** with their relative energies.

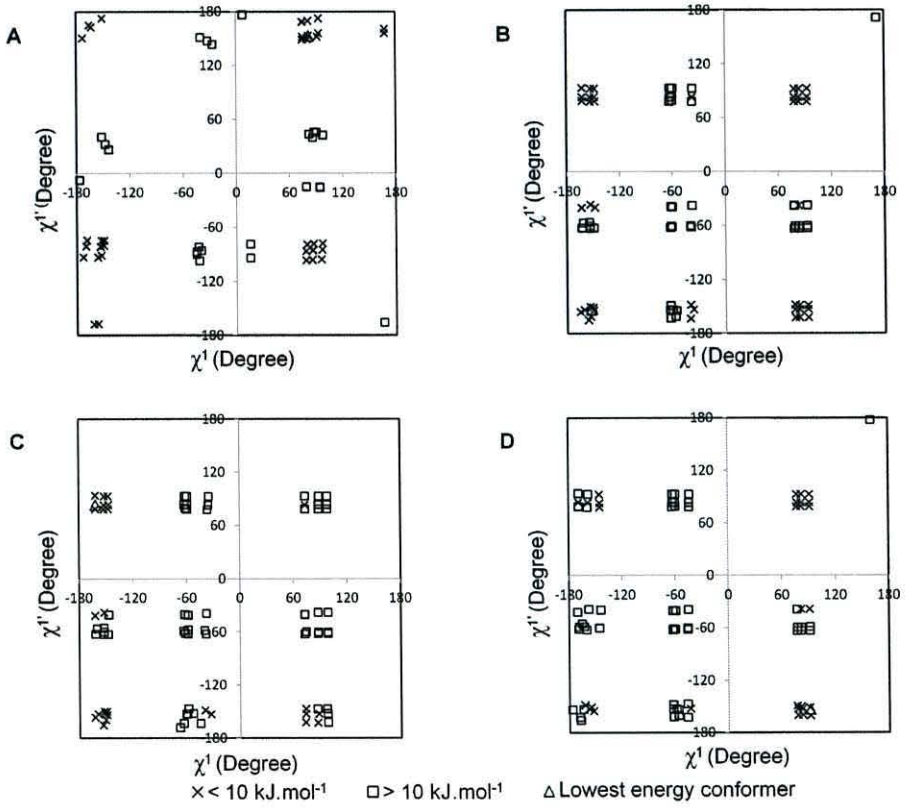


Figure 11: Dihedral distributions of χ^1 and $\chi^{1'}$ for **1-4** in panels A-D, respectively.

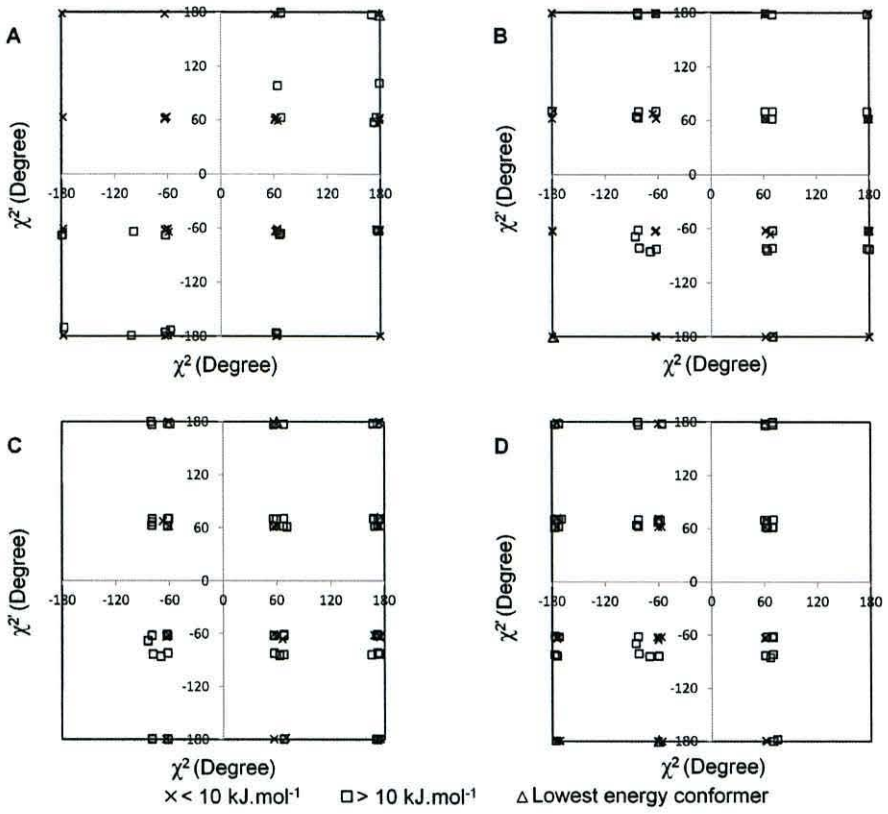


Figure 12: Dihedral distributions of χ^2 and $\chi^{2'}$ for **1-4** in panels A-D, respectively.

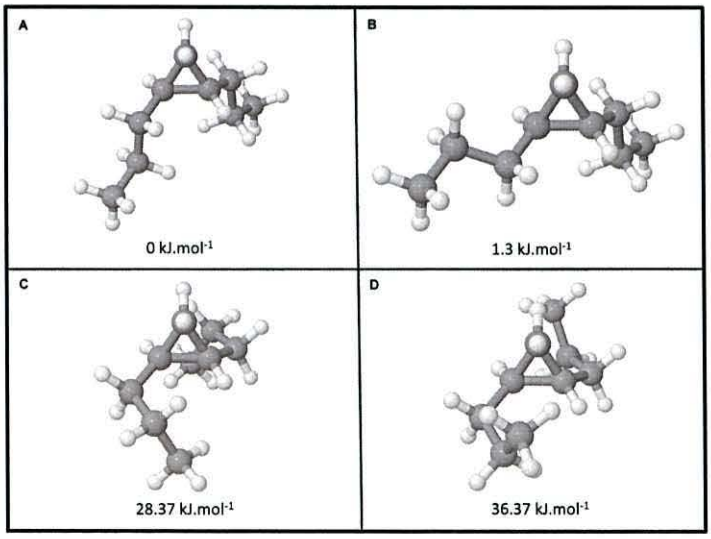


Figure 13: Low-energy (A and B, from "aaaa" and "+aaa" starting structures, respectively) and high-energy (C and D, from "----" and "--++" starting structures, respectively) conformers of **2** with their relative energies.

The results obtained for the α-methyl *trans*-cyclopropanes (**3** and **4**) show great similarity to those for *trans*-cyclopropane (**2**). Both of the epimers each have 81 viable conformers ranging from 0 - 36.0 kJ.mol^{-1} and 0 - 36.8 kJ.mol^{-1} for **3** and **4**, respectively (all conformers found with their relative energies are listed in Table S3 and Table S4). They also show a similar distribution in dihedral angles to **2** (Figure 11 and Figure 12, panels C and D, respectively). Additionally, for **3** and **4** the general trend that low-energy conformers are those with alkyl chains pointing downwards and sideways, while high-energy conformers have alkyl chains pointing upwards and inwards, holds (Figure 14 and Figure 15). However, the methyl group orientations do add some variations to conformational preferences, as is evident in the lowest energy conformer of **4** (Figure 15 A), which is not in line with the above-mentioned trend.

Further investigation shows that whereas **2** had four main groups of low-energy conformers, **3** has two main groups with χ^1 and $\chi^{1'}$ configurations of *a*+ and *aa* where $\chi^1 \approx -150°$ and $\chi^{1'} \approx 90°$ or -150° and **4** has two main groups with χ^1 and $\chi^{1'}$ configurations of ++ and +*a* where $\chi^1 \approx 90°$ and $\chi^{1'} \approx 90°$ or -150° (Figure 11, C and D). Hence, only two groups are favoured in low energy conformers for **3** and **4** while the remaining two groups that were equally favourable for **2** have become more populated with high-energy conformers for **3** and **4**. For **3**, the two groups that are favoured have $\chi^1 \approx -150°$ (in the *trans*-configuration), which results in the methyl group pointing away from the cyclopropane CH(R)-CH(R) central bond, resulting in stable, low-energy conformers (Figure 14, A and B and Figure 16 *S*, *trans*). The g^+-conformers of χ^1 for **3** on the other hand, result in the methyl group pointing towards the central CH(R)-CH(R) bond of the cyclopropane ring (Figure 16 *S*, *gauche*$^+$). This orientation is less favourable than the *trans*-configuration, as shown in Figure 11, C where there are very few low-energy conformers at $\chi^1 \approx 90°$. For **4**, the low-energy conformers that are favoured lie at $\chi^1 \approx 90°$ (in g^+-configuration), which results in lower-energy conformers in which the methyl group is pointing away from the central CH(R)-CH(R) cyclopropane bond (Figure 16 *R*, *gauche*$^+$) whereas the *trans*-configuration results in the methyl group

pointing towards it (Figure 16 *R, trans*). Accordingly, low-energy conformers for **4** are found mostly at $\chi^1 \approx 90°$, and not $\chi^1 \approx -150°$ (Figure 11, D).

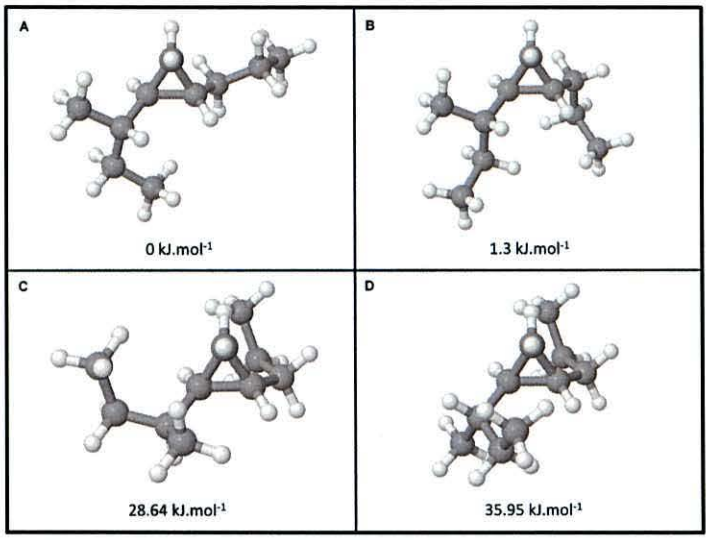


Figure 14: Low-energy (A and B, from "a++a" and "aaa-" starting structures, respectively) and high-energy (C and D, from "+--+" and "--++" starting structures, respectively) conformers of **3** with their relative energies.

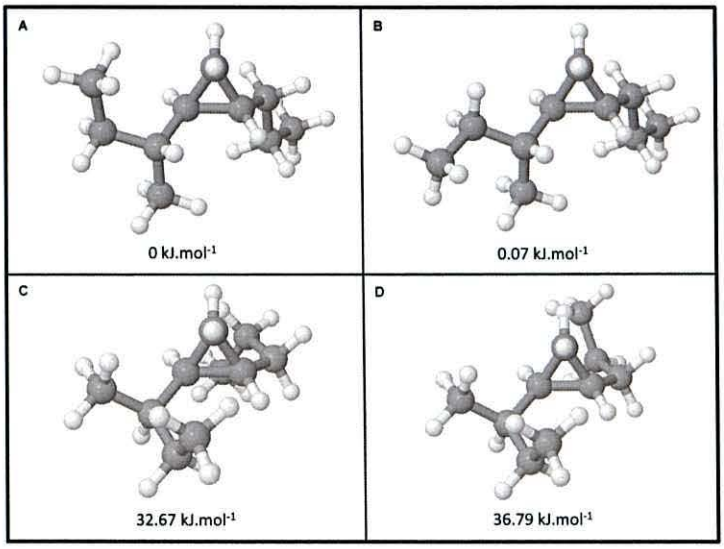


Figure 15: Low-energy (A and B, from "+a-a" and "+aaa" starting structures, respectively) and high-energy (C and D, from "--+-" and "--++" starting structures, respectively) conformers of **4** with their relative energies.

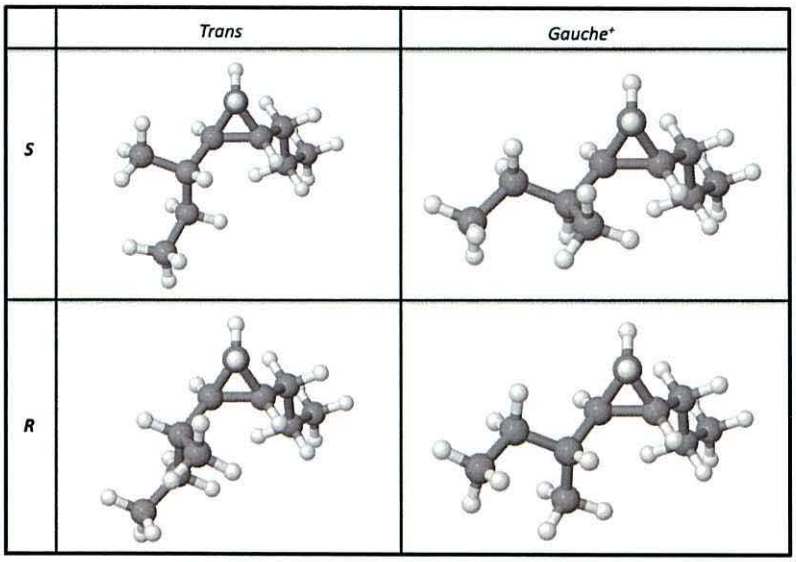


Figure 16: The methyl group orientations in *trans*- and *gauche*$^+$-conformers for **3** (top) and **4** (bottom).

The notable difference in low-energy conformers for **3** and **4** (Figure 14 and Figure 15) is that the methyl group orientation in **3** results in the methyl group pointing sideways, allowing the alkyl chains to be in close proximity to each other, while the methyl group in **4** is positioned between the alkyl chain and the cyclopropane and therefore causes the chains to be far away from each other. In **3**, compact folding at the cyclopropane group is promoted with the methyl group positioned on the outside of the hairpin bend, allowing the alkyl chains to align in parallel, but for **4** compact folding seems unlikely as low-energy conformers have alkyl chains separated by the methyl group. For both **3** and **4** it is energetically unfavourable for the methyl group to point towards the central cyclopropane bond and these conformers have energies of > 6 kJ.mol^{-1} above the lowest energy conformers.

For **3** and **4**, dihedrals χ^2 and $\chi^{2'}$ are also spread among *a*, g^+ and g^-configurations (Figure 12, C and D) with no clear distinction between high and low-energy conformers, as is the case for **2** (Figure 12, B). One slight adaptation in the

methylated molecules is the loss of an approximate symmetrical distribution of χ^2 between values close to +180° and -180°, as in **2**. For **3** and **4** the dihedral χ^2 deviates from 180° slightly more than for **2**, and opens up by approximately 5 to 7° more on the side of the methyl group, resulting in opposite trends in **3** and **4** as a result of their opposite stereochemistry at the α-methyl group.

Discussion

MAs are key to the survival of *M. tb* and changes in MA profiles in response to changes in environment indicate that the different classes of MAs have different functions.[35] The different MA classes, in addition to whether they contain *cis*- or α-methyl *trans*-cyclopropanes, have been shown to affect antigenicity and inflammatory properties.[11,12] Cyclopropanation is a trait of pathogenic mycobacteria[36] and it is hypothesised here that they play an important role in determining how MAs fold – which in turn translates to their function. The recent cryo-electron microscopy results showing a 7 - 8 nm space in which MAs sit as a symmetrical bilayer in the cell wall suggest that MAs are folded in order to fit into this space.[13] Such folding at each functional group into a W-fold involves the alkyl chains adjacent to the functional groups running closely in parallel. Hence, the preferred alkyl chain conformations of short chain cyclopropanes and the role of dispersion interactions and stereochemistry in the folding of these molecules was investigated in this work.

From the assessment study it was seen that the relative energies of the conformers change, depending on the theory employed. This is not unexpected as the PES is likely to differ between the different theories, and others have also seen this effect.[24,31,34,37] It is proposed that the HF PES, not accounting for dispersion forces in any way, is not flat enough.[24,25] Energy distributions of conformers suggest that the PES is flatter for MP2 but flattest for DFT-D (Figure 8). A greater dispersion effect was seen with the alkyl chain interaction for DFT-D than for MP2 (Figure 7), and according to SP energies at CCSD level of theory, the DFT-D geometries were less stable (higher in energy, Table 1) suggesting that

these calculations overestimate the dispersion effects for this system. Therefore, MP2 was used for examining conformational preferences of cyclopropane fragments **1-4**.

The proximity of the alkyl chains in *cis*-dipropylcyclopropane **1** resulted in 9 conformers that were not viable due to the chains intersecting. Cyclopropanes with alkyl chains on opposite sides of the ring in the *trans*-orientation **2-4**, did not have alkyl chains intersecting, although with longer alkyl chains it is expected that the number of viable conformers will be reduced from the total amount possible in both *cis*- and *trans*-orientations. In general, for **1-4**, those conformers with chains pointing upwards or inwards (Figure 9, A and D) are higher in energy than those with chains pointing sideways and down (Figure 9, B and C). For all of the molecules low-energy conformers were found where χ^1, $\chi^{1'} \approx +/-90°$ and $+/-150°$ which correspond to *gauche* and *trans*-conformations, in agreement with previous findings for cyclopropane and cyclopropene rings with various substituents.[38,39]

For **1** the lowest energy conformer was at $\chi^1 = 90°$, $\chi^{1'} = 150°$ in *gauche* and *trans*-conformations respectively, while for **2** the lowest energy conformer was in the *trans*-conformation at χ^1, $\chi^{1'} = -150°$. χ^2 and $\chi^{2'}$ for **1-4** were distributed among *trans*- and *gauche*-conformations. The lowest energy conformers for **3** and **4** were at $\chi^1 = -150°$, $\chi^{1'} = 90°$ and $\chi^1 = 90°$, $\chi^{1'} = -150°$, respectively. The orientation of the α-methyl group skewed the preference for χ^1 to the *trans*-conformation in the *S*-diastereomer **3** and to the *gauche* configuration for the *R*-diastereomer **4**. The arrangement of these low-energy geometries imply that the *S*-diastereomer, which is the configuration found in *M. tb* cell walls, promotes the close parallel alignment that is required for the proposed W-fold with the α-methyl group situated outside, while the methyl group points down in a way that separates the two alkyl chains from each other in the non-natural *R*-diastereomer. This result correlates with the recent observation that the α-methyl *trans*-cyclopropane promotes tight hairpin U-shaped folding in MAs.[18] Villeneuve *et al.*[18] have noted that α-methyl *trans*-cyclopropane model compounds with 4 or 5 carbons in the alkyl chains fold with the alkyl chains in closer proximity than *cis*-cyclopropanes

and that U-shaped conformers were energetically favoured compared to extended ones. It can be postulated that the methyl groups adjacent to the oxygenated functional groups in MAs will also promote such folding. Therefore, the oxygenated MAs are more likely to fold at least once in the mero-chain, forming more rigid condensed structures.

For MAs to fold in the proposed W-fold, the cyclopropane group is an important folding point at which the alkyl chains are expected to come in close proximity and align in parallel, providing ample opportunity for stabilisation by London dispersion forces. Stabilising effects due to dispersion interactions in the alkyl chains were not seen in these short-chain cyclopropane fragments with MP2 calculations. It remains to be seen whether dispersion interactions will be more dominant as the alkyl chains are extended. For all model compounds **1-4**, the lowest energy conformers in *gauche* or *trans*-conformations suggest that the W-fold is plausible since in this geometry the cyclopropane is the bending point, as in Figure 2. The folding in whole MAs is also dependent on the relative alkyl chain lengths between the functional groups,[16-18] which will affect the position of the cyclopropane in a W-fold and hence determine the orientation of the adjacent alkyl chains. It is conceivable that the chains will extend laterally to the cyclopropane in a linear conformation if one of the adjacent alkyl chains is significantly shorter so that the cyclopropane cannot form the bending point. However, other MA folds with the alkyl chains adjacent to the cyclopropane pointing upwards (Figure 9, A) are high in energy and not favourable.

Although MAs are flexible molecules with no rigid geometry, it has been shown here that MA conformations are likely to be influenced by low-energy preferred conformations around the cyclopropane groups. The presence and stereochemistry of cyclopropane groups in the *M. tb* cell wall has been shown to affect cell wall permeability, alter virulence in infection and modify interactions with key components of the immune response. AMA, the most abundant MA in *M. tb* cell walls containing two *cis*-cyclopropane groups, inherently has the adjacent alkyl chains of the cyclopropanes in close proximity to each other. AMA has been shown to be able to stretch the mero-chain out completely when packed tightly[17,40-42] and to unfold during simulation,[16,17] showing the most

flexibility in adopting a range of different conformations.[15] This flexibility correlates with the smallest difference in energy between high- and low-energy structures observed here for the *cis*-cyclopropane. This flexibility and accessibility of different conformers at cyclopropane groups in AMA may indicate a contribution of fluidity to the cell wall and hence affect viability of the organism and thus warrant successful infection, but does not stimulate interactions with the immune system – as shown by the lack of antigenic and inflammatory properties. On the other hand, KMA with the lowest abundance in the *M. tb* cell wall, mostly contains α-methyl *trans*-cyclopropanes and has been shown to stay folded in the W-fold, even under high lateral pressure where AMA and MMA unfold.[16,42]

The current results suggest that α-methyl *trans*-cyclopropane with *S* stereochemistry, as in **3**, contributes to the stabilisation of the W-fold by bringing the alkyl chains together with the methyl group on the outside. These findings indicate that the α-methyl *trans*-cyclopropane group may predispose MA to the W-fold, as shown by the rigidity of KMA with α-methyl *trans*-cyclopropane. The lack of antigenicity and inflammatory responses of AMA indicate that the α-methyl *trans*-cyclopropane group occurring in MMA and KMA may facilitate these interactions with the immune system. However, it is evident that the structure-function relationships of MAs are more complex, involving both the proximal and distal functional groups of the mero-chain, and their relation to each other. Whole MAs need to be studied to gain further insight into these associations.

Conclusions

The results of the current study confirm that the cyclopropane and α-methyl *trans*-cyclopropane groups provide a bending point in the MA molecule. Favourable interactions between propyl chains were not observed, but it remains to be seen whether longer alkyl chains will show such effects. It was seen, however, that it is most favourable for the chains to point downwards and away

from the central CH(R)-CH(R) bond of the cyclopropane ring and not toward or above it. This agrees with the folding proposed for MAs in the W-conformation, in which the cyclopropane provides a point of folding in the molecule with the ring forming the turning point of the bend with the chains pointing downwards from it in a hairpin U-shape. It is also possible that the bend may occur in the alkyl chain either side of the ring, in which case the chains will be stretched out in line with the cyclopropyl CH(R)-CH(R) bond. This folding however, will not promote tight packing. In nature, the *trans*-isomer is methylated. The α-methyl group may make some conformers energetically more favourable due to the added steric hindrance of the methyl group, yet the added hindrance is not as drastically restrictive as the *cis*-isomer in which a number of conformers are not viable due to the proximity of the chains to each other. What seems clear is that it is energetically preferable for the methyl group to point away from the ring in both methylated epimers. This preference automatically brings the alkyl chains from the cyclopropane closer to each other when the stereochemistry of the α-methyl group is *S* (as is found in nature) but it forces a distance between them in the *R*-orientation. This finding supports the notion that bending is likely to occur at the cyclopropane group and that the alkyl chains fold to be adjacent and parallel to each other. It remains to be seen how these orientations affect folding in the complete MA molecule, as other factors, such as the relative stereochemistry of functional groups to each other, as well as alkyl chain lengths may promote or disrupt tight packing of all the alkyl chains in parallel.

References

(1) Grogan, D. W.; Cronan, J. E. *Microbiol. Mol. Biol. Rev.* **1997**, *61*, 429.

(2) Pandit, K. R.; Klauda, J. B. *Biochim. Biophys. Acta, Biomembr.* **2012**, *1818*, 1205.

(3) Loffhagen, N.; Haertig, C.; Geyer, W.; Voyevoda, M.; Harms, H. *Eng. Life Sci.* **2007**, *7*, 67.

(4) Silvius, J. R.; McElhaney, R. N. *Chem. Phys. Lipids* **1979**, *25*, 125.

(5) Perly, B.; Smith, I. C. P.; Jarrell, H. C. *Biochemistry* **1985**, *24*, 1055.

(6) Glickman, M. S.; Cox, J. S.; Jacobs, W. R. *Mol. Cell* **2000**, *5*, 717.

(7) Rao, V.; Fujiwara, N.; Porcelli, S. A.; Glickman, M. S. *J. Exp. Med.* **2005**, *201*, 535.

(8) Glickman, M. S.; Cahill, S. M.; Jacobs, W. R. *J. Biol. Chem.* **2001**, *276*, 2228.

(9) Rao, V.; Gao, F.; Chen, B.; Jacobs, W. R.; Glickman, M. S. *J. Clin. Invest.* **2006**, *116*, 1660.

(10) Barkan, D.; Hedhli, D.; Yan, H.-G.; Huygen, K.; Glickman, M. S. *Infect. Immun.* **2012**, *80*, 1958.

(11) Beukes, M.; Lemmer, Y.; Deysel, M.; Al Dulayymi, J. R.; Baird, M. S.; Koza, G.; Iglesias, M. M.; Rowles, R. R.; Theunissen, C.; Grooten, J.; Toschi, G.; Roberts, V. V.; Pilcher, L.; Van Wyngaardt, S.; Mathebula, N.; Balogun, M.; Stoltz, A. C.; Verschoor, J. A. *Chem. Phys. Lipids* **2010**, *163*, 800.

(12) Vander Beken, S.; Al Dulayymi, J. R.; Naessens, T.; Koza, G.; Maza-Iglesias, M.; Rowles, R.; Theunissen, C.; De Medts, J.; Lanckacker, E.; Baird, M. S.; Grooten, J. *Eur. J. Immunol.* **2011**, *41*, 450.

(13) Zuber, B.; Chami, M.; Houssin, C.; Dubochet, J.; Griffiths, G.; Daffè, M. *J. Bacteriol.* **2008**, *190*, 5672.

(14) Hoffmann, C.; Leis, A.; Niederweis, M.; Plitzko, J. M.; Engelhardt, H. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 3963.

(15) Groenewald, W.; Baird, M. S.; Verschoor, J. A.; Minnikin, D. E.; Croft, A. K. *Chem. Phys. Lipids* **2014**, *180*, *15*.

(16) Villeneuve, M.; Kawai, M.; Watanabe, M.; Aoyagi, Y.; Hitotsuyanagi, Y.; Takeya, K.; Gouda, H.; Hirono, S.; Minnikin, D. E.; Nakahara, H. *Biochim. Biophys. Acta, Biomembr.* **2007**, *1768*, 1717.

(17) Villeneuve, M.; Kawai, M.; Watanabe, M.; Aoyagi, Y.; Hitotsuyanagi, Y.; Takeya, K.; Gouda, H.; Hirono, S.; Minnikin, D. E.; Nakahara, H. *Chem. Phys. Lipids* **2010**, *163*, 569.

(18) Villeneuve, M.; Kawai, M.; Horiuchi, K.; Watanabe, M.; Aoyagi, Y.; Hitotsuyanagi, Y.; Takeya, K.; Gouda, H.; Hirono, S.; Minnikin, D. E. *Microbiology* **2013**, *159*, 2405.

(19) Al Dulayymi, J. R.; Baird, M. S.; Roberts, E.; Minnikin, D. E. *Tetrahedron* **2006**, *62*, 11867.

(20) Al-Dulayymi, J. R.; Baird, M. S.; Mohammed, H.; Roberts, E.; Clegg, W. *Tetrahedron* **2006**, *62*, 4851.

(21) Law, J.; Chass, G.; Torday, L.; Varro, A.; Papp, J. *Theochem (J. Mol. Struct.)* **2002**, *609*, 1.

(22) Law, J.; Koo, J.; Setiadi, D.; Chass, G.; Viskolcz, B.; Csizmadia, I. *Theochem (J. Mol. Struct.)* **2003**, *666*, 445.

(23) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; J. A. Montgomery, J.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J.; Gaussian, Inc.: Wallingford CT, 2010.

(24) Aleman, C. *J. Phys. Chem. B* **1997**, *101*, 5046.

(25) Headgordon, T.; Headgordon, M.; Frisch, M. J.; Brooks, C. L.; Pople, J. A. *J. Am. Chem. Soc.* **1991**, *113*, 5989.

(26) Chai, J.-D.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6515.

(27) Minenkov, Y.; Singstad, A.; Occhipinti, G.; Jensen, V. R. *Dalton Trans.* **2012**, *41*, 5526.

(28) Karamanis, P.; Pouchan, C. *Int. J. Quantum Chem.* **2011**, *111*, 788.

(29) Salzner, U.; Aydin, A. *J. Chem. Theor. Comput.* **2011**, *7*, 2568.

(30) Smith, S. A.; Hand, K. E.; Love, M. L.; Hill, G.; Magers, D. H. *J. Comput. Chem.* **2013**, *34*, 558.

(31) Marianski, M.; Oliva, A.; Dannenberg, J. J. *J. Phys. Chem. A* **2012**, *116*, 8100.

(32) Riley, K. E.; Platts, J. A.; Rezac, J.; Hobza, P.; Hill, J. G. *J. Phys. Chem. A* **2012**, *116*, 415.

(33) Riley, K. E.; Hobza, P. *J. Phys. Chem. A* **2007**, *111*, 8257.

(34) Dinadayalane, T. C.; Priyakumar, U. D.; Sastry, G. N. *J. Phys. Chem. A* **2004**, *108*, 11433.

(35) Yuan, Y.; Zhu, Y.; Crane, D. D.; Barry, C. E. *Mol. Microbiol.* **1998**, *29*, 1449.

(36) George, K. M.; Yuan, Y.; Sherman, D. R.; Barry, C. E. *J. Biol. Chem.* **1995**, *270*, 27292.

(37) Holroyd, L. F.; van Mourik, T. *Chem. Phys. Lett.* **2007**, *442*, 42.

(38) Stolevik, R.; Bakken, P. *J. Mol. Struct.* **1989**, *197*, 137.

(39) Greenberg, A.; Liebman, J. F. *J. Am. Chem. Soc.* **1981**, *103*, 44.

(40) Villeneuve, M.; Kawai, M.; Kanashima, H.; Watanabe, M.; Minnikin, D. E.; Nakahara, H. *Biochim. Biophys. Acta, Biomembr.* **2005**, *1715*, 71.

(41) Hasegawa, T.; Nishijo, J.; Watanabe, M.; Funayama, K.; Imae, T. *Langmuir* **2000**, *16*, 7325.

(42) Hasegawa, T.; Amino, S.; Kitamura, S.; Matsumoto, L.; Katada, S.; Nishijow, J. *Langmuir* **2003**, *19*, 105.

Supporting Information

Systematic Conformational Search for Low Energy Conformers of Cyclopropane Units From Mycolic Acids

W. Groenewald, Gregory A. Chass and A. K. Croft

Contents

Z-matrix for *cis*-1,2-dipropylcyclopropane (**1**)

H
C 1 B2
C 2 B3 1 A3
C 2 B4 1 A4 3 D4
H 2 B5 1 A5 3 D5
H 3 B6 2 A6 4 D6
C 3 B7 2 A7 4 D7
H 4 B8 2 A8 5 D8
C 4 B9 2 A9 3 D9
H 7 B10 3 A10 2 D10
H 7 B11 3 A11 2 D11
H 9 B12 4 A12 2 D12
H 9 B13 4 A13 2 D13
C 7 B14 3 A14 2 D14
C 9 B15 4 A15 2 D15
H 14 B16 7 A16 3 D16
H 14 B17 7 A17 3 D17
H 15 B18 9 A18 4 D18
H 15 B19 9 A19 4 D19
C 14 B20 7 A20 3 D20
C 15 B21 9 A21 4 D21
H 20 B22 14 A22 7 D22
H 20 B23 14 A23 7 D23
H 21 B24 15 A24 9 D24
H 21 B25 15 A25 9 D25
H 20 B26 14 A26 7 D26
H 21 B27 15 A27 9 D27

Bonds:

B2 1.08452194
B3 1.51037741
B4 1.51039398
B5 1.08178104
B6 1.08630566
B7 1.51043409
B8 1.08630848
B9 1.51042116
B10 1.093305
B11 1.09611585
B12 1.09330967
B13 1.09611355
B14 1.52829165
B15 1.52829971
B16 1.09355757
B17 1.09399184
B18 1.09355785
B19 1.09399316
B20 1.52583559
B21 1.52583088
B22 1.09130216
B23 1.09041648
B24 1.09130332
B25 1.09041668
B26 1.09122484
B27 1.09122486

Angles:

A3 116.53822622
A4 116.53928491
A5 114.60972857
A6 115.02483466
A7 120.61224839
A8 115.02179521
A9 120.60854621
A10 109.97199835
A11 109.74423033
A12 109.97038086
A13 109.74874868
A14 112.26765799
A15 112.26368027
A16 108.77178594
A17 109.06966552
A18 108.77195054
A19 109.06829463
A20 112.66076101
A21 112.66053967
A22 110.76114659
A23 111.48305657
A24 110.76083856
A25 111.48231948
A26 110.819476
A27 110.81997768

Dihedrals:

D4 -67.94098541
D5 146.02646313
D6 -105.05742423
D7 113.00895232
D8 -3.78979949
D9 -113.01346527
D10 -60.000
D11 180.000
D12 180.000
D13 -60.000
D14 60.000
D15 60.000
D16 180.000
D17 -60.000
D18 -60.000
D19 180.000
D20 60.000
D21 60.000
D22 59.82271458
D23 179.99415128
D24 -59.84720854
D25 179.98209641
D26 -59.76419523
D27 59.74034208

Dihedrals that are modified to generate various conformers grouped together.

Table S1: Relative energies[1] for all *cis*-dipropylcyclopropane conformers obtained at MP2/6-31+G(*d,p*) level of theory.

Conformer[2]		Dihedral angles for end conformers				Relative Energy
Start	End	χ^{1}	$\chi^{1'}$	χ^{2}	$\chi^{2'}$	ΔE (kJ.mol^{-1})
-a-a	+aaa	81.5	153.0	179.0	177.8	0.000
a-aa	a-aa	-153.1	-81.5	-177.8	-179.1	0.001
+aaa	+aaa	81.5	153.0	179.1	177.8	0.002
a+a+	a-aa	-153.1	-81.5	-177.9	-179.1	0.002
--aa	+-aa	85.7	-85.6	179.5	-179.5	1.332
++aa	+-aa	85.8	-85.6	179.6	-179.5	1.332
+-aa	+-aa	85.7	-85.7	179.6	-179.6	1.333
+a+a	+a+a	74.9	152.1	60.9	177.7	1.655
a-a-	a-a-	-152.1	-74.9	-177.7	-60.9	1.656
+aa+	+aa+	80.5	149.2	179.1	60.1	1.808
a--a	a--a	-149.2	-80.5	-60.1	-179.1	1.808
a+-+	a--a	-149.1	-80.3	-60.1	-179.0	1.809
-a-+	+aa+	80.4	149.2	179.1	60.1	1.809
aaaa	aaaa	-164.9	164.9	-178.8	178.8	2.392
+a-a	+a-a	93.3	156.1	-63.0	178.0	2.461
a-a+	a-a+	-156.1	-93.3	-178.0	63.0	2.461
+aa-	+aa-	81.2	169.6	179.2	-63.4	2.568
a-+a	a-+a	-169.6	-81.3	63.4	-179.2	2.569
---+	+-a+	85.0	-96.2	179.5	62.7	3.043
+--a	+--a	96.2	-85.2	-62.7	-179.5	3.046
+-a+	+-a+	85.2	-96.2	179.5	62.7	3.046
+-+a	+-+a	78.2	-86.8	61.8	-179.8	3.332
+-a-	+-a-	86.8	-78.2	179.8	-61.8	3.332
+a++	+a++	74.7	148.8	60.9	60.1	3.449
a---	a---	-148.8	-74.7	-60.1	-60.9	3.449
a+--	a---	-148.8	-74.6	-60.1	-60.8	3.450
-a++	+a++	74.6	148.7	60.8	60.1	3.451
aa-a	aa-a	-150.4	172.5	-56.2	-176.8	3.793
aaa+	aaa+	-172.5	150.4	176.8	56.1	3.793
-a+-	+a+-	74.7	168.4	61.1	-63.4	4.392
a++-	a-+-	-168.4	-74.6	63.5	-60.9	4.394
a-+-	a-+-	-168.4	-74.6	63.4	-61.0	4.396
+a+-	+a+-	74.6	168.4	61.0	-63.4	4.396
+a-+	+a-+	91.4	152.0	-63.3	60.9	4.573
a--+	a--+	-152.0	-91.4	-60.9	63.3	4.575
+--+	+--+	95.5	-95.5	-62.7	62.7	4.781
+---	+---	96.8	-78.0	-62.6	-61.6	4.982
+-++	+-++	78.0	-96.8	61.6	62.6	4.982
+a--	+a--	93.1	172.5	-62.8	-63.1	5.266

[1]Relative to the +aaa end conformer of –a-a.

[2]Labeled as $\chi^{1}\chi^{1'}\chi^{2}\chi^{2'}$

Table S1: continued.

Conformer[2]		Dihedral angles for end conformers				Relative Energy
Start	End	χ^{1}	$\chi^{1'}$	χ^{2}	$\chi^{2'}$	ΔE (kJ.mol^{-1})
a-++	a-++	-172.5	-93.1	63.1	62.8	5.266
+-+-	+-+-	79.2	-79.2	62.2	-62.2	5.484
aa-+	aa-+	-162.8	162.8	-62.5	62.5	7.267
aa+a	aa+a	167.7	160.6	64.4	177.6	7.974
aaa-	aaa-	-160.6	-167.7	-177.6	-64.4	7.975
aa--	aa--	-156.0	-167.6	-58.9	-64.2	9.769
aa++	aa++	167.6	156.0	64.2	58.9	9.770
++-+	+--a	94.2	-16.1	-63.0	-175.7	14.224
--a+	+-a+	15.5	-94.1	175.8	63.0	14.227
++-a	+--a	94.1	-15.4	-63.0	-175.8	14.231
--a-	+-a-	15.2	-78.9	176.0	-61.8	14.416
+++a	+-+a	78.9	-15.3	61.9	-175.9	14.420
-aaa	-aaa	-32.1	147.2	170.3	177.3	15.101
a+aa	a+aa	-147.2	32.3	-177.4	-170.3	15.101
a++a	a-+a	-176.3	-7.6	63.3	-177.5	15.880
-aa-	+aa-	7.6	176.3	177.5	-63.3	15.880
-a--	+aa-	7.3	176.4	177.5	-63.2	15.882
a+++	a-+a	-176.2	-7.6	63.2	-177.4	15.883
-aa+	-aa+	-26.6	143.3	173.4	57.4	16.659
a+-a	a+-a	-143.3	26.6	-57.4	-173.4	16.659
aa+-	aa+-	166.0	-166.0	63.6	-63.6	18.890
-a+a	-a+a	-40.6	151.2	67.9	178.9	26.877
a+a-	a+a-	-151.2	40.6	-178.8	-67.9	26.879
--+a	--+a	-45.6	-90.2	68.1	179.6	28.914
++a-	++a-	90.2	45.6	-179.6	-68.1	28.914
++a+	++a+	86.0	39.8	179.4	101.2	29.441
---a	---a	-40.2	-86.0	-100.9	-179.4	29.443
--++	--++	-42.6	-97.3	67.7	62.6	30.186
++--	++--	97.3	42.6	-62.6	-67.7	30.187
--+-	--+-	-45.6	-87.5	67.6	-65.9	31.965
+++-	+++-	87.5	45.6	65.9	-67.6	31.965
----	----	-43.2	-81.9	-98.2	-63.9	32.246
++++	++++	81.9	43.2	63.9	98.2	32.246

Table S2: Relative energies[1] for all *trans*-dipropylcyclopropane conformers obtained at MP2/6-31+G(*d,p*) level of theory.

Conformer[2]		Dihedral angles for end conformers				Relative Energy
Start	End	χ^{1}	$\chi^{1'}$	χ^{2}	$\chi^{2'}$	ΔE (kJ.mol^{-1})
aaaa	aaaa	-153.0	-153.0	-179.5	-179.5	0.000
aa-a	aa-a	-150.0	-153.9	-63.7	-179.6	1.036
aaa-	aaa-	-153.9	-150.0	-179.6	-63.7	1.036
+aaa	+aaa	83.0	-153.3	-180.0	-180.0	1.304
a+aa	a+aa	-153.3	83.0	-180.0	-180.0	1.304
++aa	++aa	83.3	83.3	179.8	179.8	1.870
aa--	aa--	-150.6	-150.6	-63.7	-63.7	2.065
+aa+	+aa+	83.2	-162.3	179.9	62.0	2.132
a++a	a++a	-162.3	83.2	62.0	179.9	2.132
a+a+	a+a+	-153.3	78.2	-179.9	62.0	2.674
+a+a	+a+a	78.2	-153.3	62.0	-179.9	2.675
+aa-	+aa-	83.0	-148.3	180.0	-62.0	2.713
a+-a	a+-a	-148.3	83.3	-62.0	-180.0	2.715
+++a	+++a	78.4	83.3	62.0	179.7	3.265
++a+	++a+	83.3	78.4	179.7	62.0	3.265
+a++	+a++	78.7	-162.3	61.9	61.8	3.393
a+++	a+++	-162.3	78.7	61.8	61.9	3.393
+a-a	+a-a	92.6	-152.9	-62.4	-179.9	3.742
a+a-	a+a-	-152.9	92.6	-179.9	-62.4	3.743
aa+a	aa+a	-160.8	-153.6	70.5	-178.4	3.772
aaa+	aaa+	-153.6	-160.8	-178.4	70.5	3.772
+a+-	+a+-	78.4	-148.1	62.1	-61.8	4.091
a+-+	a+-+	-148.1	78.4	-61.8	62.1	4.092
++-a	++-a	92.7	83.5	-62.4	180.0	4.186
++a-	++a-	83.5	92.7	180.0	-62.4	4.186
+a-+	+a-+	93.2	-162.3	-62.1	61.9	4.588
a++-	a++-	-162.3	93.2	61.9	-62.1	4.588
++++	++++	78.5	78.5	62.1	62.1	4.611
+a--	+a--	92.6	-148.3	-62.5	-62.3	5.152
a+--	a+--	-148.3	92.6	-62.3	-62.5	5.152
aa-+	aa-+	-156.0	-165.0	-65.9	67.5	5.482
aa+-	aa+-	-165.0	-156.0	67.5	-65.9	5.485
++-+	++-+	92.7	78.5	-62.3	62.0	5.532
+++-	+++-	78.5	92.7	62.0	-62.3	5.532
++--	++--	92.9	92.9	-62.2	-62.2	6.368
-aaa	-aaa	-36.6	-153.3	179.9	-179.9	8.649
a-aa	a-aa	-153.3	-36.6	-179.9	179.9	8.649
-+aa	-+aa	-37.2	82.8	179.2	-179.6	9.210
+-aa	+-aa	82.8	-37.2	-179.7	179.1	9.212
a-+a	a-+a	-163.2	-40.2	61.5	178.0	9.513
-aa+	-aa+	-40.3	-163.2	178.0	61.5	9.514

[1]Relative to the +aaa end conformer of –a-a.

[2]Labeled as $\chi^{1}\chi^{1'}\chi^{2}\chi^{2'}$

Table S2: continued.

Conformer[2]		Dihedral angles for end conformers				Relative Energy
Start	End	χ^{1}	$\chi^{1'}$	χ^{2}	$\chi^{2'}$	ΔE (kJ.mol^{-1})
-aa-	-aa-	-39.6	-147.9	178.5	-61.2	9.650
a--a	a--a	-148.0	-39.6	-61.2	178.5	9.650
+-+a	+-+a	78.1	-37.5	62.1	179.0	10.455
-+a+	-+a+	-37.5	78.0	179.0	62.1	10.456
+--a	+--a	92.6	-36.9	-62.2	179.1	11.314
-+a-	-+a-	-37.0	92.6	179.1	-62.2	11.316
a-a-	a-a-	-153.7	-59.9	179.8	-82.9	14.248
-a-a	-a-a	-60.1	-153.7	-82.8	179.8	14.248
-a-+	-a-+	-56.9	-161.1	-84.3	64.1	14.927
a-+-	a-+-	-161.1	-56.9	64.1	-84.3	14.927
+-a-	+-a-	83.0	-60.2	180.0	-82.3	15.121
-+-a	-+-a	-60.2	83.0	-82.3	179.9	15.122
-+-+	-+-+	-60.0	79.2	-82.2	62.6	16.230
+-+-	+-+-	79.2	-60.0	62.6	-82.2	16.230
-a--	-a--	-56.0	-154.2	-85.5	-69.0	16.360
a---	a---	-154.2	-56.0	-69.0	-85.5	16.360
--aa	--aa	-37.8	-37.8	178.2	178.2	17.391
-+--	-+--	-59.8	92.9	-82.7	-62.1	17.439
+---	+---	92.9	-59.8	-62.1	-82.7	17.439
-a+a	-a+a	-62.8	-153.7	70.2	-179.9	18.268
a-a+	a-a+	-153.7	-62.8	-179.9	70.2	18.268
aa++	aa++	171.4	171.4	62.2	62.2	18.431
-a++	-a++	-62.3	-162.8	69.9	61.8	18.900
a-++	a-++	-162.8	-62.3	61.8	69.9	18.900
-++a	-++a	-62.7	82.9	70.2	-180.0	18.935
+-a+	+-a+	82.9	-62.7	-180.0	70.2	18.935
-a+-	-a+-	-62.6	-148.9	70.2	-62.2	19.103
a--+	a--+	-148.9	-62.6	-62.2	70.2	19.104
-+++	-+++	-62.8	78.1	69.9	61.9	20.076
+-++	+-++	78.1	-62.8	61.9	69.9	20.076
-++-	-++-	-62.6	92.9	70.0	-62.2	21.147
+--+	+--+	92.9	-62.7	-62.2	70.0	21.150
---a	---a	-60.1	-39.5	-82.2	177.3	23.042
--a-	--a-	-39.5	-60.1	177.3	-82.2	23.042
--+a	--+a	-61.9	-39.0	70.2	177.9	26.819
--a+	--a+	-39.0	-61.9	177.9	70.2	26.819
----	----	-61.0	-61.0	-81.5	-81.5	28.370
---+	---+	-60.5	-61.2	-81.8	69.8	32.412
--+-	--+-	-61.2	-60.5	69.8	-81.8	32.414
--++	--++	-62.2	-62.2	70.2	70.2	36.377

Table S3: Relative energies[1] for all *trans*-dipropylcyclopropane (*S*) conformers obtained at MP2/6-31+G(*d,p*) level of theory.

Conformer[2]		Dihedral angles for end conformers				Relative Energy
Start	End	χ^{1}	$\chi^{1'}$	χ^{2}	$\chi^{2'}$	ΔE (kJ.mol^{-1})
a++a	a++a	-161.3	83.4	59.9	179.9	0.000
aaaa	aaaa	-151.6	-152.7	175.6	-179.4	0.298
aa++	a+++	-161.3	78.6	59.8	61.8	1.149
a+++	a+++	-161.3	78.7	59.8	61.8	1.149
aaa-	aaa-	-152.4	-150.0	175.3	-64.0	1.314
a+aa	a+aa	-151.8	83.0	174.7	179.9	1.568
aa-a	aa-a	-149.1	-153.9	-63.7	-179.6	2.032
aa+a	aa+a	-159.0	-153.3	69.5	-178.7	2.095
a++-	a++-	-161.2	93.4	59.9	-61.9	2.295
a+a+	a+a+	-151.8	78.2	174.8	61.8	2.837
aa--	aa--	-149.4	-150.6	-63.7	-63.7	3.045
a+-a	a+-a	-147.0	83.2	-60.6	179.9	3.639
a+a-	a+a-	-151.7	92.7	174.8	-62.2	3.842
aa+-	aa+-	-162.5	-156.8	66.1	-66.3	3.941
aaa+	aaa+	-150.7	-159.6	176.1	72.2	4.110
a+-+	a+-+	-147.0	78.4	-60.5	61.9	4.946
a+--	a+--	-147.1	92.7	-61.0	-62.3	5.945
aa-+	aa-+	-153.7	-165.4	-67.0	67.5	6.402
a-+a	a-+a	-162.3	-41.9	59.5	177.0	6.960
+a+a	+a+a	73.0	-152.7	56.6	-179.7	7.776
+a++	+a++	72.7	-162.3	56.5	61.2	7.949
a-aa	a-aa	-152.0	-38.2	175.1	178.9	8.634
+a+-	+a+-	73.0	-147.2	56.7	-60.9	8.689
+++a	+++a	72.8	82.9	56.5	179.8	8.778
-aaa	-aaa	-33.2	-153.2	173.4	-179.1	9.211
+aaa	+aaa	87.6	-152.6	172.7	-179.8	9.242
-aa-	-aa-	-40.2	-148.7	168.7	-61.9	9.577
+aa+	+aa+	86.2	-162.5	172.3	61.7	9.680
++++	++++	73.1	78.5	56.5	62.2	10.006
++aa	++aa	87.9	83.1	172.7	-180.0	10.187
+aa-	+aa-	85.8	-147.1	172.2	-60.9	10.249
a--a	a--a	-146.7	-41.1	-59.1	177.6	10.288
-+aa	-+aa	-35.7	82.7	170.7	-179.7	10.442
+++-	+++-	73.0	92.8	56.5	-62.1	10.830
+a-a	+a-a	97.6	-152.5	-62.2	-179.9	10.913
+a-+	+a-+	98.0	-162.3	-61.9	61.8	11.342
++a+	++a+	87.8	78.6	172.7	62.3	11.470
-+a+	-+a+	-36.9	77.7	169.5	61.9	11.594
++-a	++-a	97.7	83.4	-62.1	-180.0	11.681
a-+-	a-+-	-160.0	-56.5	62.7	-84.6	11.963
+a--	+a--	96.5	-147.1	-62.6	-61.0	12.016

[1]Relative to the +aaa end conformer of –a-a.

[2]Labeled as $\chi^{1}\chi^{1'}\chi^{2}\chi^{2'}$

Table S3: continued.

Conformer[2]		Dihedral angles for end conformers				Relative Energy
Start	End	χ^{1}	$\chi^{1'}$	χ^{2}	$\chi^{2'}$	ΔE (kJ.mol^{-1})
++a-	++a-	88.2	92.8	172.9	-62.2	12.274
-+a-	-+a-	-35.2	92.5	170.9	-62.2	12.558
++-+	++-+	97.8	78.5	-62.1	62.1	12.934
a-a-	a-a-	-152.1	-60.0	174.8	-82.9	13.693
++--	++--	97.8	92.8	-62.1	-62.1	13.697
-aa+	-aa+	-45.4	-163.6	168.8	69.4	14.284
a-++	a-++	-161.8	-62.4	59.8	69.9	15.964
+-+a	+-+a	72.3	-40.5	56.5	176.5	16.225
-a-a	-a-a	-61.0	-152.9	-80.1	179.7	16.921
a---	a---	-152.3	-55.7	-69.6	-85.8	16.976
-a+a	-a+a	-60.2	-153.1	67.8	-179.6	17.735
a-a+	a-a+	-152.4	-63.0	175.1	70.2	17.871
+-aa	+-aa	87.1	-38.1	172.9	177.7	18.012
-a+-	-a+-	-58.8	-147.1	67.4	-60.8	18.067
-+-a	-+-a	-62.3	83.2	-79.1	-179.9	18.462
--aa	--aa	-38.2	-39.5	167.7	177.8	18.477
-a--	-a--	-53.8	-152.4	-83.5	-68.1	19.080
-++a	-++a	-59.5	83.0	67.5	-179.8	19.081
+--a	+--a	98.1	-38.0	-62.0	177.8	19.214
-a-+	-a-+	-63.9	-163.6	-78.8	66.3	19.474
a--+	a--+	-147.7	-62.6	-60.9	70.1	19.489
-+-+	-+-+	-61.6	79.5	-79.3	62.7	19.544
-+++	-+++	-59.7	78.1	67.4	62.0	20.174
-+--	-+--	-62.1	92.9	-79.3	-62.0	20.774
-a++	-a++	-68.0	-168.2	71.1	61.2	21.125
-++-	-++-	-59.7	92.8	67.5	-62.2	21.279
+-+-	+-+-	73.8	-59.9	56.8	-82.0	21.457
+-a-	+-a-	87.5	-60.3	172.7	-82.0	23.354
--a-	--a-	-40.5	-58.3	165.9	-83.8	23.781
+---	+---	98.3	-60.4	-61.9	-82.0	24.735
+-++	+-++	72.0	-62.4	56.3	70.0	25.272
---a	---a	-62.3	-40.5	-79.0	176.8	26.266
--+a	--+a	-58.8	-41.3	67.6	176.9	26.849
+-a+	+-a+	87.2	-61.9	173.0	70.1	27.247
--a+	--a+	-38.6	-62.5	168.0	70.5	27.547
+--+	+--+	98.3	-61.9	-61.8	70.0	28.642
----	----	-63.5	-59.1	-78.2	-83.2	31.326
--+-	--+-	-58.7	-57.7	67.5	-83.9	32.143
---+	---+	-62.0	-61.5	-79.1	70.0	35.237
--++	--++	-59.1	-62.5	67.6	70.4	35.958

Table S4: Relative energies[1] for all *trans*-dipropylcyclopropane (*R*) conformers obtained at MP2/6-31+G(*d,p*) level of theory.

Conformer[2]		Dihedral angles for end conformers				Relative Energy
Start	End	χ^{1}	$\chi^{1'}$	χ^{2}	$\chi^{2'}$	ΔE (kJ.mol^{-1})
+a-a	+a-a	91.3	-152.9	-60.3	-179.4	0.000
+aaa	+aaa	82.2	-152.9	-174.3	-179.3	0.074
++-a	++-a	91.6	83.4	-60.2	-180.0	1.142
+aa-	+aa-	81.4	-150.1	-174.4	-64.0	1.259
++aa	++aa	82.3	83.2	-174.6	179.8	1.261
+a--	+a--	90.8	-151.3	-60.4	-65.2	1.298
+a+a	+a+a	77.7	-152.7	60.9	-179.3	2.135
++-+	++-+	91.6	78.4	-60.2	61.9	2.420
++a+	++a+	82.2	78.5	-174.6	62.0	2.586
+a+-	+a+-	76.8	-149.5	60.6	-63.5	3.266
++--	++--	91.5	92.8	-60.1	-62.1	3.310
+++a	+++a	77.5	83.2	60.6	179.7	3.381
++a-	++a-	82.3	92.7	-174.5	-62.3	3.509
+aa+	+aa+	82.9	-159.8	-173.5	72.1	4.109
+a-+	+a-+	91.7	-160.6	-59.7	71.6	4.280
++++	++++	77.4	78.5	60.8	62.0	4.685
+++-	+++-	77.4	92.7	60.6	-62.2	5.612
+a++	+a++	77.8	-160.4	61.8	70.8	5.938
aa-a	aa-a	-152.6	-155.6	-61.3	-179.4	6.412
aa--	aa--	-155.3	-151.0	-61.7	-62.6	7.156
a+-a	a+-a	-144.9	82.8	-56.4	-179.9	7.496
aaaa	aaaa	-158.1	-153.1	-171.1	-179.1	7.649
+--a	+--a	91.3	-38.5	-60.2	178.4	8.001
aaa-	aaa-	-162.3	-148.7	-172.4	-62.1	8.061
+-aa	+-aa	81.8	-38.5	-174.5	178.7	8.394
a+-+	a+-+	-144.8	77.9	-56.2	61.9	8.718
a+aa	a+aa	-158.1	82.9	-172.2	-179.8	9.135
a++a	a++a	-168.7	83.4	62.0	-179.8	9.574
a+--	a+--	-145.2	92.5	-56.7	-62.2	9.795
-aaa	-aaa	-43.1	-152.8	-175.3	-179.9	9.905
-aa+	-aa+	-46.4	-162.8	-177.3	61.4	10.318
a+a+	a+a+	-158.8	78.0	-172.4	61.9	10.361
+-+a	+-+a	77.2	-38.8	60.7	178.4	10.438
a+++	a+++	-168.6	78.7	61.7	62.0	10.656
-aa-	-aa-	-46.3	-147.4	-177.0	-60.9	10.771
-+aa	-+aa	-44.1	82.9	-176.1	-179.7	10.811
a+a-	a+a-	-157.8	92.7	-172.2	-62.2	11.392
aa+a	aa+a	-163.8	-153.3	73.9	-178.1	11.564
aa-+	aa-+	-167.1	-166.0	-60.6	67.4	11.843
a++-	a++-	-168.6	93.4	62.3	-61.9	11.846
-+a+	-+a+	-43.8	78.2	-176.0	62.2	11.988

[1]Relative to the +aaa end conformer of –a-a.

[2]Labeled as $\chi^{1}\chi^{1'}\chi^{2}\chi^{2'}$

Table S4: continued.

Conformer[2]		Dihedral angles for end conformers				Relative Energy
Start	End	χ^{1}	$\chi^{1'}$	χ^{2}	$\chi^{2'}$	ΔE (kJ.mol^{-1})
-a-a	-a-a	-59.0	-153.2	-83.4	179.8	12.150
-a-+	-a-+	-56.8	-160.5	-84.6	64.1	12.226
aa+-	aa+-	-176.7	-154.0	68.6	-61.6	12.619
aaa+	aaa+	-168.0	-162.8	-169.2	70.9	12.711
-+a-	-+a-	-43.5	92.7	-176.0	-62.1	12.824
-+-a	-+-a	-59.6	83.0	-82.6	179.9	13.279
+---	+---	91.5	-58.6	-60.0	-83.7	13.517
+-a-	+-a-	82.0	-59.4	-174.8	-83.2	13.591
-a--	-a--	-56.7	-153.9	-85.2	-69.4	14.216
-+-+	-+-+	-59.4	79.1	-82.6	62.5	14.361
a--a	a--a	-143.9	-39.7	-55.7	177.7	14.616
-+--	-+--	-59.4	92.8	-82.9	-62.0	15.477
+-+-	+-+-	77.5	-59.6	60.7	-83.0	15.605
a-aa	a-aa	-157.4	-38.9	-172.0	177.9	16.931
a-+a	a-+a	-170.2	-42.2	61.2	176.3	17.158
+--+	+--+	91.5	-62.9	-60.0	69.9	17.345
+-a+	+-a+	81.8	-63.2	-174.9	70.3	17.561
--aa	--aa	-45.1	-39.1	-176.9	177.2	18.794
-a+a	-a+a	-62.6	-153.1	69.5	-179.9	18.865
-a++	-a++	-62.3	-162.5	69.2	61.5	19.018
+-++	+-++	77.1	-63.4	60.2	70.0	19.528
-a+-	-a+-	-62.5	-148.2	69.6	-61.9	19.672
-++a	-++a	-62.7	82.9	69.6	179.9	19.699
-+++	-+++	-62.8	78.1	69.4	61.9	20.799
---a	---a	-59.4	-40.5	-82.6	176.5	20.822
-++-	-++-	-62.7	92.7	69.5	-62.1	21.812
a-a-	a-a-	-162.5	-58.6	-173.9	-83.7	22.206
a-+-	a-+-	-165.0	-55.4	66.3	-85.2	22.450
aa++	aa++	160.6	177.6	63.6	61.4	22.532
a---	a---	-168.4	-59.0	-70.0	-84.0	22.634
--a-	--a-	-45.3	-60.1	-177.3	-82.1	24.120
a--+	a--+	-145.6	-60.2	-57.8	69.5	24.289
----	----	-60.3	-61.9	-82.0	-80.9	25.950
a-a+	a-a+	-159.8	-62.4	-172.6	70.3	26.097
a-++	a-++	-168.9	-61.3	62.6	69.9	26.278
--+a	--+a	-62.4	-40.5	69.9	177.0	27.365
--a+	--a+	-46.0	-61.7	-177.2	70.2	27.946
---+	---+	-59.6	-61.5	-82.4	69.9	30.085
--+-	--+-	-61.9	-60.9	69.5	-81.6	32.672
--++	--++	-62.6	-62.5	69.8	70.3	36.790

Manuscript II

Computational Investigation of Solvent and Structural Effects on Mycolic Acid Conformation

Wilma Groenewald,[†] *Anna K. Croft*[*†a]

[†]*School of Chemistry, Bangor University, Bangor, Gwynedd, LL57 2UW, United Kingdom*

[a]*Current address Department of Chemical and Environmental Engineering, University of Nottingham, University Park, Nottingham NG7 2RD, United Kingdom*

Abstract

Despite being largely a curable disease, Tuberculosis is still claiming thousands of lives every day. The persistence of *Mycobacterium tuberculosis* is partly due to its lipid rich cell wall, of which mycolic acids (MAs) are a major component. The different types of MAs have differences in effects on the immune system, in recognition by antibodies and in conformational behaviour. The functional groups have been coined as bending points in the molecule. Folding at each of the functional groups, the MA has a W-shape in two dimensions. We have shown before through *in vacuo* simulation that seven different conformations with folding at various functional groups classified as W, U and Z-conformations are possible. Here we expand on previous work by increasing the simulation timescale and including explicit solvents. Each MA showed a unique distribution of conformations. Out of all possible WUZ-conformations, all MAs preferred the W-conformation, being the energetically favoured conformation. A control backbone-MA without any mero-chain functional groups showed significantly less folding in the mero-chain. KMA showed the highest percentage of WUZ-conformations and, in particular, a tendency to fold at its α-methyl *trans*-cyclopropane group, in agreement with recent results by Villeneuve *et al.* MAs fold similarly in vacuum and water with a majority of folded conformations around the W-conformation, although molecules are more flexible in vacuum than in water. On the other hand, unfolded conformations with a

disperse distribution are common in hexane for all MAs except KMA. Newly defined globular conformations, consisting of high-energy, folded conformations that may be relevant in biofilms are identified.

Introduction

It is estimated that in 2012 1.3 million people died from Tuberculosis (TB), amounting to more than 3500 deaths per day.[1] This is shocking in view of the fact that TB is largely a curable disease, although treatments are prolonged and require multiple drugs. The organism causing TB in humans, *Mycobacterium tuberculosis*, is particularly resilient, in part due to a lipid-rich cell wall. Mycolic acids (MAs) are major components of the mycobacterial cell wall.[2,3]

MAs are 2-alkyl-3-hydroxy fatty acids that can be up to 80 carbons long. MAs mostly occur covalently bound to arabinogalactan, but also exist as trehalose mono- and dimycolates.[3,4] In *M. tb* there are three main classes of MAs. MAs vary at the proximal (*P*) and distal (*D*) functional groups in the long mero-chain, as well as the chain lengths between the groups, as shown in Figure 1. Alpha-MA (AMA) has *cis*-cyclopropane groups at both *P* and *D*. Oxygenated MAs have a methoxy or keto group with adjacent methyl group at D. The oxygenated MAs exist with either a *cis*- or *trans*-methyl cyclopropane group at *P*. In *M. tb* methoxy-MA (MMA) occurs mostly with *cis*-cyclopropane, whereas keto-MAs (KMA) generally have *trans*-methyl cyclopropane groups.[5] The absolute stereochemistries of the functional groups have been proposed as shown in Figure 1, obtained through comparison of natural compounds with synthetic compounds.[4,6-11]

The fact that MA compositional profiles change in response to the bacterial environment and growth[12] suggests that the cell wall MA composition is important for cell wall function. Furthermore, the change of MA functional groups through genetic mutations has shown varied cell wall permeability, growth and virulence as a result of altered MA composition and structure. Pathogenicity in mice seems to depend on oxygenated MAs.[13] Oxygenated MAs

also play a central role in foam cell formation, possibly facilitating long-term survival in the host.[14] AMA and KMA were increased during intracellular growth and without KMA growth inside cells was significantly decreased,[12] suggesting an important role for KMA in intracellular survival of the mycobacteria. Cyclopropanation is not to be overlooked in MA structure-function relationships. The lack of a *cis*-cyclopropane group in *M. bovis* BCG mutants resulted in a non-lethal infection, in contrast to the wild type.[15] *Trans*-cyclopropanation of oxygenated MAs subdued the inflammatory and virulence effects of *M. tb*.[16] A mutant strain lacking all cyclopropanation showed hypervirulence and strong immune responses.[17] Therefore the effect of cyclopropanation overall is also to curb the immune response to *M. tb* infection. The inflammatory pattern and antibody recognition were also shown to vary depending on the specific MA structure, as determined with synthetic MAs.[18,19] This data implies that the underlying MA structure (individually or as sugar-ester derivatives) steers the physical properties and conformational behaviour of the molecule, and hence affects its biological role.

As described above, MAs are required for the survival and viability of the mycobacteria and can alter the outcome of an infection. They are central to the host immune response against the organism and have been shown to be good antigens for use in serodiagnosis of TB.[20-23] However, fundamental details such as how MAs are arranged in the cell wall and how they interact with immune components, are yet to be determined. Cryo-electron microscopy results suggest that the outer bilayer of the mycobacterial cell wall is 7 - 8 nm wide and consists of a symmetrical bilayer structure.[24] This would imply that longer MAs would need to fold in order to fit into this space. Zuber and co-workers have suggested that MAs fold at each of their functional groups, forming a W-conformation, and intercalate with lipids in the opposite leaflet in a zipper model. Interactions with components of the host immune system, such as antibody binding are likely to involve a macrostructure consisting of numerous MAs. Therefore, knowledge on the preferred conformations of single MAs will shed light on these areas.

Numerous studies on MA monolayers have been performed,[25-33] which serve as a close approximation of MAs in the cell wall that are thought to arrange in a

monolayer, linked to arabinogalactan.[34] These studies have shown that MA conformations change as the molecules are packed closer together at higher lateral pressures. It is thought that all MAs have folded conformations with up to four chains arranged in parallel, as in the W-conformation, occupying a large surface area at low lateral pressures.[25-29] The long mero-chain in AMA and MMA extends completely as the surface area is decreased. In contrast, KMA forms a very rigid monolayer and the molecules stay in a folded W-conformation, even when high lateral pressure is applied.

Villeneuve and co-workers have performed short simulations on various MAs in a W-conformation, studying their preference for staying in the folded conformation or unfolding.[26-28] Their finding supported monolayer results; AMA and MMA tended to unfold, while KMA mostly remained folded.[26] They have found that MAs with alkyl chains between the functional groups of similar length fold into more tightly-packed W-conformations and unfold more slowly than MAs in which the alkyl chain lengths differ. The presence of a double bond favoured an energetically more stable W-conformation, as compared to *cis*-cyclopropane.[27] An α-methyl *trans*-cyclopropane group within the molecule was also found to promote folding of an alkyl chain as compared to the *cis*-isomer.[29]

We have previously shown through unconstrained simulation *in vacuo* that MAs spontaneously fold into reproducible conformational groupings.[35] Clear differences in conformational preferences between MA classes suggest that the underlying chemical composition steers MA conformation, with KMA showing very different trends to AMA and MMA. Seven possible WUZ-folds were defined as all the possible conformations of MAs that fold at all or some of their functional groups, with two, three or four alkyl chains in parallel. It is significant that the MAs folded into the WUZ-conformations spontaneously, without any solvent or neighbouring MA molecules to aid in the folding. This suggests that it is likely for MAs to fold at their functional groups, with stabilisation from van der Waals interaction between the parallel alkyl chains. However, conformations of MAs defined by WUZ only account for a small percentage of MA conformations. Therefore, in this work we aim to define other MA

conformations that have not been described to date. The effect of different solvents on the single MA conformations is also examined, thereby providing a more complete picture of both the intrinsic and external factors that affect MA conformational preferences.

Methods

Molecules

Four different MAs (Figure 1) were modelled, namely AMA, MMA, KMA, and a backbone MA (BMA) as control MA containing no functional groups in the mero-chain. Molecules were built using WebMO[36] and Aten[37] graphical interfaces and were numbered serially from the alpha chain through to the mero-chain. The OPLS all atom (AA) forcefield (with additional parameters for cyclopropane) was applied in Aten and the residue topology parameter (.rtp) file exported. Subsequently, topologies were created using the Gromacs simulation package[38-41] and each molecule was placed in the centre of a 10 × 10 × 10 nm box and minimized with a steepest descent algorithm with a maximum stepsize of 0.01 nm, a maximum number of 200 000 steps and a tolerance of 10 $kJ.mol^{-1}.nm^{-1}$.

Cyclopropane parameters for OPLS all-atom forcefield

Six bonded parameters for cyclopropane, listed in Table 1 and 2 and were obtained from Dr Lee-Ping Wang using hybrid ensembles for force matching, as detailed elsewhere.[42] The fit to QM data is shown in supplementary figures 1-6.

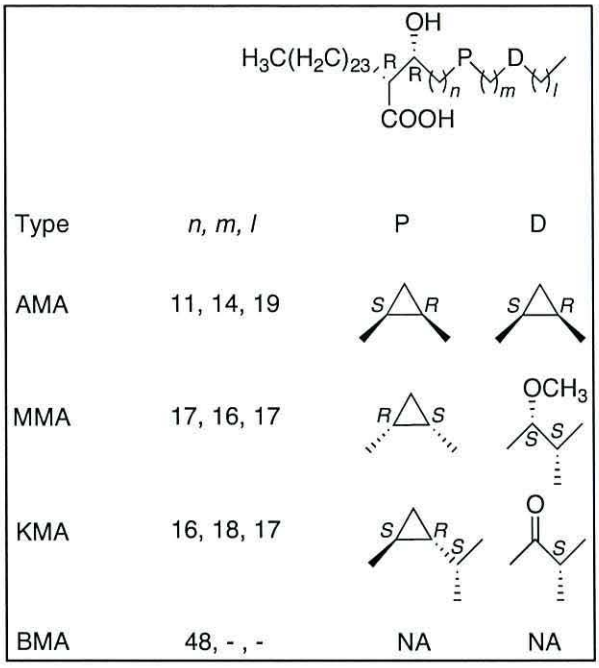


Figure 1: Chemical structures of MAs modelled, representing main components of MAs from *M. tb*. All MAs have identical alpha chains and mycolic motif. *P* and *D* represent the proximal and distal functional groups, respectively, and *n*, *m* and *l* represent alkyl chain lengths. The backbone MA (BMA) serves as a control and does not have any mero-chain functional groups.

Table 1: Angle (degrees) and force constant ($kJ.mol^{-1}.rad^{-2}$) for cyclopropane angle parameter.

Angle	th0	cth
CT-CT-CY	115.090	612.010

Table 2: Torsional fourier coefficients (kJ.mol^{-1}) obtained for cyclopropane.

Dihedral	V_1	V_2	V_3	V_4
CT-CT-CT-CY	1.00000	-22.00200	24.05800	-20.89300
CT-CT-CY-CY	1.00000	14.17700	-20.14000	-9.78690
CY-CT-CT-HC	1.00000	-32.91200	25.29700	2.47730
CT-CT-CY-HC	1.00000	1.43810	21.88000	12.80200
HC-CT-CY-HC	1.00000	9.99550	11.63500	-14.48900

Simulation details

For all simulations, unless stated otherwise, a timestep of 1 fs was used and the neighbour list was updated every 10 fs. Van der Waals interactions were modelled by using a shift function between 0.8 and 0.9 nm and electrostatic interactions were modelled by using PME. V-rescale temperature coupling was used at 300K with a time constant of 0.5 ps and no constraints were applied. The equation of motion was integrated using a leap-frog algorithm. All simulations were performed with Gromacs version 4.5.4.

Vacuum simulations were performed with a NVT ensemble.

Simulations in water were performed by filling the simulation box with TIP4P water. The number of water molecules for the different mycolic acid boxes ranged from 32951 to 32972. Equilibration was performed with position restraints by applying a force of 1000 kJ.mol^{-1}.nm^{-2} in the x, y and z-directions on all carbon and oxygen atoms of the MA. The system was equilibrated by performing 100 ps NVT, followed by 50 ps NPT using Berendsen pressure coupling to scale the box in an efficient way at the beginning of the simulations, and lastly 100 ps NPT using Parinello-Rahman pressure coupling in order to yield the correct ensemble. In both of the NPT-ensembles isotropic pressure coupling at a pressure of 1 bar was used with a 1 ps time constant. Production simulations for MAs in water also used the Parinello-Rahman setup and no constraints were applied.

In order to use hexane as a solvent, a hexane solvent box was built and equilibrated before addition to MA simulation boxes. This was done by building a single hexane and obtaining its topology parameter file as described before. Then a 3.6 × 3.6 × 3.6 nm box was filled with 216 hexane molecules, minimized (as described before) and equilibrated at 300 K using a NVT ensemble for 5 ns. The energy plots for this equilibration are shown in Supplementary figure S7. The density of the hexane box was 584.201 $kg.m^{-3}$, which is approximately 11% lower than the experimental density for hexane. Simulation boxes containing MAs were filled with hexane using the equilibrated hexane box. The number of hexane molecules ranged from 32951 to 32972 for the different MAs. Equilibration was performed with position restraints by applying a force of 1000 $kJ.mol^{-1}.nm^{-2}$ in the x, y and z-directions on all carbon and oxygen atoms. The system was equilibrated by firstly increasing the timestep from 10^{-7} ps to 10^{-3} ps by running respective simulations of 100000 steps at 0 K with a timestep of 10^{-7} ps, followed by 10^{-6} ps timestep, 10^{-5} ps timestep, 10^{-4} ps timestep, and finally 10^{-3} ps timestep with Berendsen pressure coupling using a time constant of 1 ps. Secondly, using a 10^{-3} ps timestep and Berendsen pressure coupling as above, the temperature was increased in 50 K increments from 0 K to 250 K with consecutive 100 ps simulations and a 1 ns simulation at 300 K. Production simulations with hexane as solvent were run using Parinello-Rahman pressure coupling at a pressure of 1 bar and a time constant of 1 ps, a timestep of 2 fs and with all bonds constrained.

For each MA in each of the three different environments, 20 simulations were run for 10 ns each. Starting conformations for the 20 replicate production simulations for each system were varied in order to sample the potential energy surface well. Taking the starting frames 250 ps apart from an initial 5 ns simulation in vacuum and water systems and 240 ps apart for those in hexane ensured a variety of starting conformations. For each simulation 1001 frames were written (consisting of one frame every 10 ps as well as the starting frame at 0 ps) and used in subsequent analysis.

Analysis

The backbone carbon chain consists of all the consecutively linked carbons along the length of the MA chain and excludes any non-carbon atoms, the carbon of the acid group, as well as the CH_2 carbons of cyclopropyl groups and adjacent methyl groups. A straight reference MA of each type had all its backbone carbon atom dihedrals set to 180 °.

Five points (a-e, Figure 2) were used to analyse the fold of the molecule[26] with (a) the last carbon in the 2-alkyl chain, (b) the carbon on which the carboxyl group is attached, (c) the distal carbon of the cyclopropane ring, (d) the carbon to which the keto- or methoxy group is attached and in the case of AMA the distal carbon of the cyclopropane ring, and (e) the end carbon of the meromycolate chain.

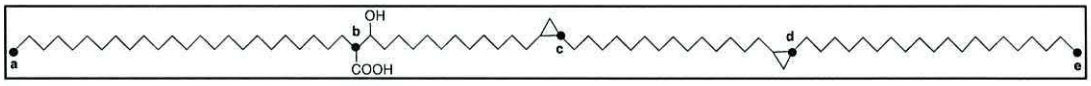


Figure 2: Points a – e that were used in analysis are indicated on AMA with dots and the relevant letters.

MA key distances were used to define the seven possible W, U and Z-conformations that describe conformers that bend at the functional groups.[35] These conformations are shown schematically in Figure 3 and the criteria are outlined in Table 3. Prefixes "a" and "e" describe conformations in which the a- and e-terminating chains are extended, while "s" refers to symmetry. 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.

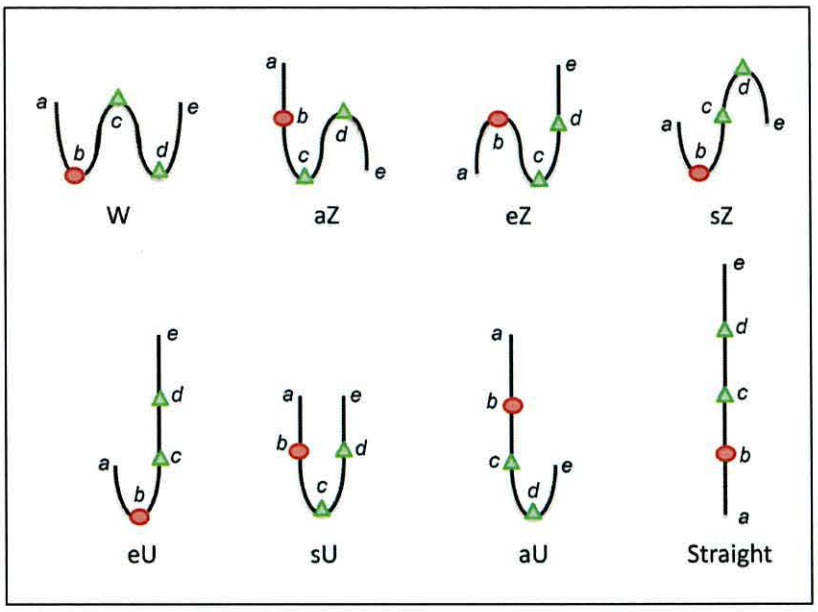


Figure 3: A schematic representation of the seven possible WUZ-conformations and a straight conformations for AMA. The acid head group is indicated by a red dot and the cyclopropane groups by green triangles.

Principal component plots were produced by using the seven WUZ folds and a straight conformer for each MA. BMA was plotted onto the map for AMA for direct comparison, since they share an identical backbone. The WUZ structures are minimised average structures, hand-picked from the simulations to represent the most ideal conformers. For MMA and KMA sZ was adapted from the average W-conformation by adjusting the angles around the cyclopropane group, *c*, to 180° due to the small percentage of sZ conformers available and their shape not conforming well to the ideal sZ shape.

Table 3: The chain lengths and intramolecular distances for the seven possible WUZ-conformations.

Fold	Percentage of extended length				Distance between points (nm)			
	ab	bc	cd	de	ac	ae	ce	bd
W	> 50	> 70	> 70	> 50	< 2.0	< 2.5	< 2.0	< 1.0
aZ	> 50	> 70	> 70	> 50	> 2.0	> 2.5	< 2.0	< 1.0
eZ	> 50	>70	> 70	> 50	< 2.0	> 2.5	> 2.0	< 1.0
sZ	> 50	> 70	> 70	> 50	< 2.0	< 2.5	< 2.0	> 1.8
eU	> 50	> 70	>70	> 50	< 2.0	> 2.5	> 2.0	> 1.8
sU	> 50	> 70	> 70	> 50	> 2.5	< 2.5	> 2.5	< 1.0
aU	> 50	> 70	> 70	> 50	> 2.0	> 2.5	< 1.8	> 1.8

Results and Discussion

Cyclopropane parameters

With the new angle and dihedral parameters obtained for the cyclopropane unit (as listed in Table 1 and 2) low energy conformers for cyclopropane are approximated well, compared to the reference quantum mechanics data. However, there is room for improving the overall fit of the forcefield data to the quantum mechanics data (see Supplementary figure S1-6). Although the parameterization is not the focus of this work, the addition of the angle parameter, which is well described, is an improvement to describing the cyclopropane with this forcefield.

Equilibration

As shown by the energy, temperature and pressure plots in Supplementary figure S8-10, all simulations reached thermal equilibration early on, as simulations in water and hexane underwent preceding equilibration steps and there are very few degrees of freedom to equilibrate in the vacuum simulations. Radius of gyration was checked as a measure that gives an indication of the shape of the molecule at each time. These plots, shown in Supplementary figure S11, show convergence in the molecular shape after 6 ns for most of the replicate simulations when simulated in water. Convergence is not seen in vacuum nor hexane, with hexane showing the most variation. Based on this observation, MA conformations from the last 4 ns of simulations were used to compare to the first 6 ns.

Defined WUZ MA conformations

Each frame from the simulations was analysed for the seven possible WUZ-folds according to their chain lengths and intramolecular distances, as defined in Table 3 and shown schematically in Figure 3. The seven possible WUZ conformations for AMA are shown in Figure 4. A W-conformation represents bending at each functional group with four parallel alkyl chains. The various Z-conformers fold at two of the functional groups while U-conformers only fold at one functional group.

As shown in Table 4, KMA had the highest percentage of WUZ-conformations in vacuum, water and hexane (48.0 %, 27.7 % and 29.4 %, respectively). Both AMA and MMA show fewer WUZ-conformations (11.5 %, 15.0 % and 10.4 % for AMA and 12.8 %, 6.5 % and 6.9 % for MMA in vacuum, water and hexane, respectively) with the backbone control molecule without mero-chain functional groups, BMA the least (3.7 %, 1.2 % and 5.9 % in vacuum, water and hexane, respectively). The large percentage of WUZ-conformers for KMA clearly distinguishes it from the other MAs and may indicate divergent function for this MA class. The decreased WUZ-conformations for BMA suggests that the mero-

chain functional groups do influence how MAs fold and various functional groups may steer conformations in unique ways dependent on the chemical structure of each molecule.

For each of the molecules modelled, the highest percentage of WUZ-conformations occurs in vacuum, indicating that the MAs access more WUZ-conformations when their movement is not restricted. For each MA-type modelled, the most compactly folded W-conformation occurs most often when simulated in vacuum or in water. In contrast, more open U-conformations are mostly found in hexane.

Table 4: The percentage of WUZ-conformers obtained for each molecule modelled in vacuum, water and hexane.

MA	Solvent	W	aZ	eZ	sZ	eU	sU	aU	Total %
AMA	Vacuum	7.9	1.6	0.2	0.5	0.0	1.0	0.3	11.5
	Water	9.0	5.4	0.0	0.3	0.1	0.0	0.2	15.0
	Hexane	0.0	0.5	0.1	0.3	2.7	1.0	5.8	10.4
MEO	Vacuum	10.3	0.7	0.7	0.0	0.2	0.9	0.0	12.8
	Water	4.4	0.1	1.4	0.0	0.5	0.0	0.0	6.5
	Hexane	0.1	0.5	0.4	0.1	1.8	1.7	2.3	6.9
KMA	Vacuum	39.5	7.3	0.6	0.0	0.0	0.4	0.2	48.0
	Water	19.2	7.7	0.0	0.0	0.0	0.3	0.5	27.7
	Hexane	0.2	7.3	1.1	0.3	0.9	16.4	3.3	29.4
BMA	Vacuum	0.7	0.3	0.2	0.7	0.4	1.0	0.4	3.7
	Water	0.3	0.3	0.1	0.1	0.3	0.0	0.2	1.2
	Hexane	0.0	0.0	0.1	0.1	3.8	0.2	1.7	5.9

- Calculated as a percentage of the total number of frames for the 20 simulations of each MA

Relative energies show that the W-conformer has the lowest energy for all MA-classes (Table 5), suggesting that van der Waals interaction between the alkyl chains contributes to the stabilisation of this conformation. The different MAs vary in the remaining energy distributions of conformers, but in general U-conformers are higher in energy than Z-conformers with more alkyl chain interaction. The curious difference for KMA is that sU has a much lower relative

energy (58.0 kJ.mol^{-1}) compared to AMA and MMA (> 152.0 kJ.mol^{-1}). KMA in hexane has a high percentage of frames with sU conformation, which is much larger than for any conformers for any MAs in hexane (Table 4). The sU conformation requires a hairpin fold at the α-methyl *trans*-cyclopropane group. This group has recently been described to facilitate folding in MAs.[29] KMA is the only MA with an α-methyl *trans*-cyclopropane group modelled here. KMA shows clearly distinct conformational preferences to AMA and MMA. It has the highest percentage of W- and sU-conformations, which reinforce the α-methyl *trans*-cyclopropane group as a good hairpin folding point, facilitating MA folding. Additionally, the conformations requiring extension at the α-methyl *trans*-cyclopropane group (sZ, eU and aU) appear infrequently.

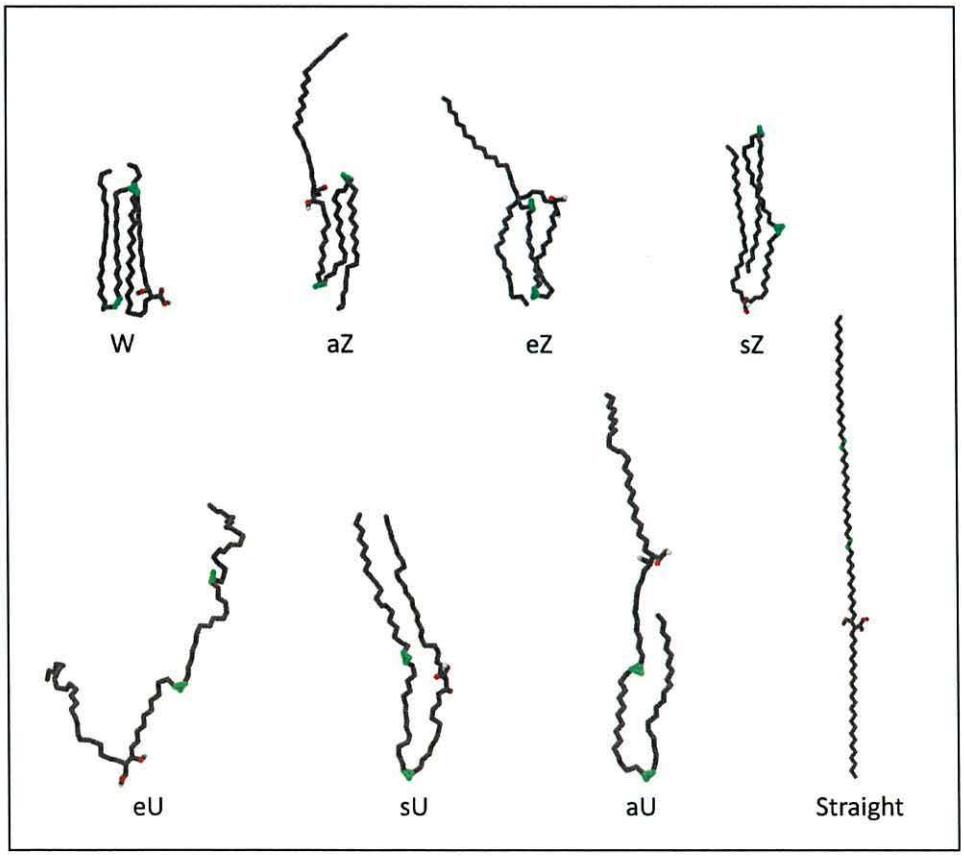


Figure 4: The average structures of the seven possible WUZ-conformations for AMA.

Table 5: Relative energies (kJ.mol^{-1}) for WUZ-conformers for each MA modelled.

	W	aZ	eZ	sZ	eU	sU	aU	Straight
AMA	0	137.6	161.0	32.5	582.7	158.3	149.5	147.6
MMA	0	425.8	83.4	115.2	252.8	152.2	279.2	266.1
KMA	0	48.1	124.4	127.1	511.8	58.0	150.0	164.9

BMA, consisting of an identical carbon backbone to AMA, but lacking any mero-chain functional groups, was modelled as a control to study the effect of the functional groups on chain folding. In vacuum the WUZ-conformations are quite similar to AMA, at low frequencies, with the exception of W and aZ conformers which are significantly lower for BMA. In water, with solvent restriction on movement, the same trend is amplified with much lower WUZ conformers for BMA. However, in hexane AMA also has generally low frequencies of WUZ-conformers and BMA conformations are therefore similar, except for a higher percentage of eU conformers for BMA (3.8 %) compared to AMA (2.7 %). With folding at the acid head group in the eU conformer, this result shows that the presence of the remaining functional group, namely the head group, in BMA, facilitates folding at this functional group.

The WUZ-conformations are defined based on two dimensional intramolecular distances and chain extensions. This is a limitation, as the cutoffs mean that molecules that resemble WUZ-conformations well are sometimes excluded. On the contrary, due to the two-dimensionality of the definitions, conformations that do not resemble the correct shape can be included in WUZ-defined folds. Extending the model to include additional rules for the distances *ad* and *be* can help to eliminate this. Translating the conformations into three-dimensionally defined folds is expected to improve the definition of the conformations, although it may still remain challenging to determine cutoffs, as it is evident that closely grouped conformations can differ significantly (as shown in the next section). It was also noted that there exist subtypes for each defined conformer. These subtypes differ in the arrangement of the parallel chains next to each other,

as they may intertwine in different ways. Hence, in future it is possible to go into a lot more detail in defining MA conformations by defining these subtypes.

In general the percentages of WUZ-conformations found here are higher than previously found,[35] most notably for KMA showing the highest total percentages in all environments simulated. This is largely attributable to the difference in forcefield used, but may also be affected by simulation length and hence sampling of the potential surface. It remains significant that MAs can fold in such structured conformations as are found in ordered compressed monolayers, even when in isolation without any restrictions or lateral packing effects. Nevertheless, it is apparent that there are a large percentage of conformations that are not defined by WUZ-conformers and this is addressed in the next section.

Undefined MA conformations

As shown above, WUZ-conformations comprising hairpin bends at various or all functional groups with straight alkyl chains in parallel and in close proximity to each other, only account for a fraction of the conformations sampled in simulations of single MAs. In order to get a unified picture of the spread of all the conformations, the carbon backbone of average WUZ-structures combined from extracted frames were used to map out the two principal coordinates for each MA. These principal component plots are shown in Supplementary figure S12 and Figure 5, with the position for the average structures for W (circle), aZ (square), eZ (diamond), sZ (triangle), eU (cross) sU (upside-down triangle) and aU (plus) shown with the corresponding symbols on the plot. In addition, a completely straight extended conformation with carbon backbone dihedrals of 180° is represented by a star. The plot for each MA is unique, as the axes are vectors that display the maximum amount of variance for each molecule. Only BMA is plotted on the same set of axes as AMA, as they share an identical carbon backbone for direct comparison.

At the extremes of the x-axis for AMA and BMA for the first eigenvector are an extended straight conformation at low values and an sU-like conformation at the

higher end. Therefore this axes represents the degree of extension along the backbone of the MA. The second eigenvector axis has those conformations with the chain terminating with "e" extended at lower values (such as eZ and eU) and the higher values represent conformations with the chain terminating with "a" being extended (such as aZ and aU). For both MMA and KMA, the axes for the first eigenvector also correlate with the extension of the backbone of the MA, but with lower values corresponding to an sU-like conformation and higher values a straight, extended conformation, opposite to AMA. Again for MMA and KMA, the axis for the second eigenvector separates those conformations with the chain terminating with "a" being extended from those terminating with "e" extended.

All frames from the simulations were projected onto this space in Supplementary figure S12 and it can be seen that the potential energy surface of each molecule was sampled well when the projection of conformers from all 20 simulations, sampling a large area, is compared to the projection of single simulation conformers, which sample only a small portion of the conformational space, as shown in Supplementary figure S12. Therefore the approach taken here with nanosecond timescale and multiple simulations with varied starting conformations has served its purpose in improving the sampling of the conformational space. As highlighted, in water, MA conformers tended to show convergence toward the latter half of the simulation time. Therefore there seems to be conformational difference between the starting and final conformations, especially in water. When the conformers from the last 4 ns of each simulation only are projected on the principal component plot as shown in Figure 5, it can be seen that for vacuum and water simulations the conformers are grouped more specifically at folded conformations such as W and sZ. While the more extended conformers are not present in the last 4 ns in vacuum and water, the distribution of conformers in hexane remains diffuse and spread out over extended conformers even in the last 4 ns. Therefore, for clarity, the plots with conformers from the last 4 ns are shown here, with the most populated conformations more clearly visible.

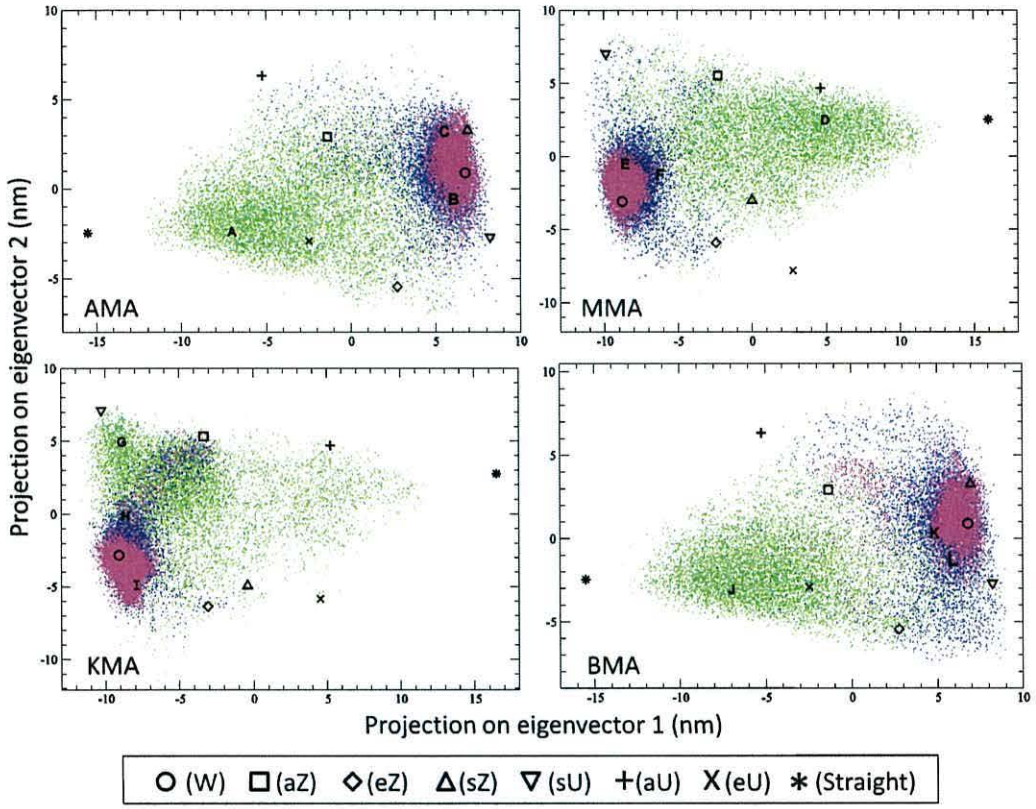


Figure 5: Principal component plots for all molecules modelled in vacuum (blue), water (magenta) and hexane (green). Frames for the last 4 ns of each simulation are shown.

From the principal component plots in Figure 5 it is evident that the WUZ-defined conformations are positioned peripherally to all the conformations represented on the plot. By extracting conformers from defined parts of each principal component plot and obtaining average structures, new conformations different to WUZ-conformations could be defined. The letters A-L in Figure 5 represent an example of newly-defined conformers. For each of the MAs, except KMA, a new conformation representative of a large proportion of conformers occurring in hexane was defined. This conformation shown in Figure 6, A is representative for those indicated by A, D and J in Figure 5. This new conformation is unfolded with a slight bend along the length of the molecule and various kinks in the alkyl chain. In addition, new conformers that are compactly folded representative of those surrounding the W-conformation (Figure 5 B and

C) are shown in Figure 6 as B and C. These structures are globular in shape with numerous bends, twists and kinks in the alkyl chains. The chains weave between each other in diverse patterns, making these conformations hard to define in two dimensions. Similar globular conformations were found for E, F, H, I, K and L. These new conformers tend to have a high ratio of *gauche*-orientation of the alkyl chains and therefore tend to have higher energies (data not shown).

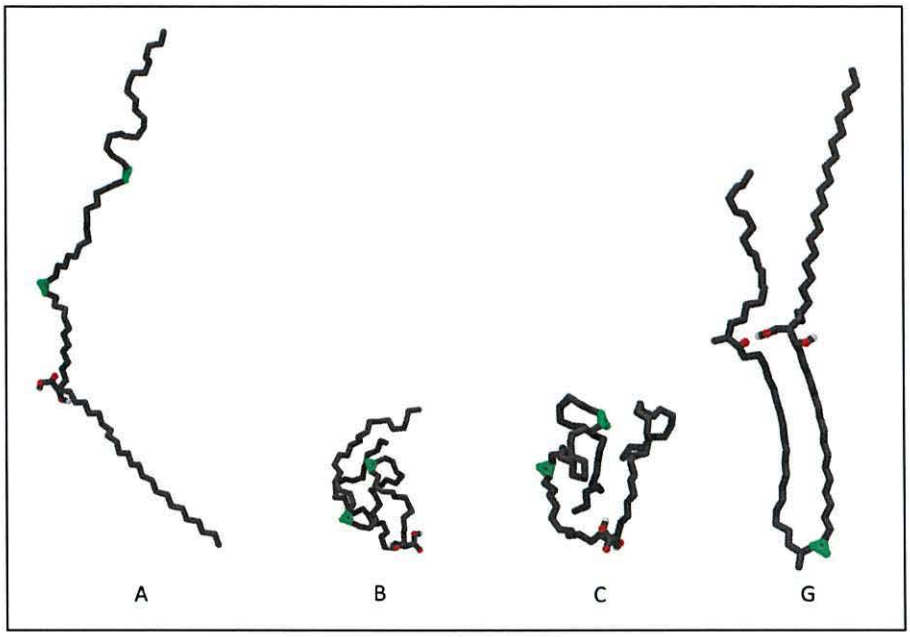


Figure 6: Newly defined conformations for MAs. Structure A was obtained from open conformations of AMA in hexane, structure B and C were obtained from folded conformations of AMA in water, and G was obtained from folded conformations of KMA in hexane.

KMA showed a different distribution in hexane for which a new conformer was obtained to represent one of the main conformational groupings for KMA in hexane. This conformation, shown in Figure 5 and 6 as G, closely resembles an sU-conformation, with a hairpin bend at the cyclopropane group with additional kinks in the chains where the polar head group and mero-chain keto-group are in close proximity. It is possible that conformations of this type are stabilised by hydrogen bonding and that this influences the conformational distribution seen for KMA. However, compared to MMA, which also has an oxygenated group at the distal functional group, this skewing effect of conformational distribution is

not seen. Granted that the methoxy group is less polar than the keto group and that the methyl group may hinder hydrogen bonding in MMA, this shows that it is a combination of factors that result in the conformational preferences of each MA.

The principal component plots show that the majority of conformations found for single MAs do not have extended alkyl chains in the *trans*-orientation, as is found in defined WUZ-conformations, but rather constitute bent alkyl chains with high *gauche*-content. When the alkyl chains of the MAs align closely and in parallel, the *trans*-orientation is promoted, as seen by the highly ordered W-conformations in which all four alkyl chains are parallel and very straight. In MA monolayers and in the cell wall arrangement, the packing of MAs close to each other is likely to promote ordering and *trans*-orientation of the MA alkyl chains. Therefore it is not likely that the globular conformations of MAs will be prevalent in these settings. However, at high surface areas in monolayer experiments, where molecules are not tightly packed, conformations with bent and twisted alkyl chains will be more predominant, especially if the molecules are not folded into the W-conformation at these surface areas, as shown in coarse grained simulations of AMA (as shown in manuscript III of this thesis). Free MAs occurring in the mycobacterial biofilm[43] are also likely to occur in more globular conformations.

This work has shed light on various aspects involved in steering and modifying MA folding. As we have shown before, the unique conformational preferences of each MA show a dependence on the underlying chemical structure.[35] In particular, the functional groups in the MA facilitate folding. This role of functionality is evidenced by the lack of WUZ-conformations for the BMA molecule lacking mero-chain functional groups.

AMA accounts for about half of the MA content in the *M.* tb cell wall and was shown in previous work to be the most flexible, with the highest percentage of WUZ-conformations. Being the only AMA without an oxygenated mero-functional group, it also showed low immune activity and antigenicity.[18,19] Here, the flexibility of AMA is confirmed with all possible WUZ-conformations

represented within the three environments simulated. The high percentage of AMA in the cell wall implies that it is key in determining cell wall fluidity and permeability properties. MMA occurs naturally with either *cis*- or *trans*-cyclopropanation. MMA is the most antigenic, with the *trans*-cyclopropanation being more antigenic than *cis*-cylopropanation. Here, the *cis*-cyclopropane-containing MMA was modelled as it represents the majority as found in *M.* tb. The *cis*-isomer is extremely immune active, eliciting a distinct inflammatory response, whereas the *trans*-isomer has largely lost this activity.[19] KMA showed the highest percentage of WUZ-conformations in this work. The difference in the percentage of WUZ-conformations to previous work is presumably due to extended timescales in the current work and different forcefields that were used.

In correspondence with monolayer experiments[25,26,29,31,33] and previous simulation results,[26,29,35] KMA was found to have a preference for more folded conformations as compared to AMA and MMA. The conformational rigidity of KMA, as shown by its folded conformations even at high lateral pressures in monolayers, is expected to alter the cell wall properties, as it does not show the flexibility and fluidity of AMA. The presence of KMA is key in intracellular survival.[12] KMA is also essential for mycobacterial pellicle formation, as well as conferring drug tolerance to the mycobacterium.[44] Only the *trans*-cyclopropane KMA was modelled here, which shows higher antigenicity than the *cis*-isomer.[18] However, the *cis*-isomer elicits a strong inflammatory response, while the *trans*-isomer showed an anti-inflammatory response.[19] KMA is the only MA modelled here with a *trans*-cyclopropyl group. KMA shows a high propensity for folding at this group, and a low percentage of conformations that require extension at this group such as sZ and aU. This trend is consistent with the recent observation that the *trans*-cyclopropyl group, with its adjacent methyl group, facilitates folding into a hairpin bend, more than *cis*-cyclopropane.[29]

Solvent effects on MA conformations are addressed here for the first time with explicit solvation. MA conformations in vacuum and in water were remarkably similar for all molecules modelled, with preference for folded conformations in the region of the W-conformation on the principal component plots. Therefore vacuum simulations at much lower cost are a good approximation of MA

conformations in water. In contrast, MA conformations were dispersed among less compactly folded conformations when simulated in hexane for AMA, MMA and BMA. KMA showed a preference for more compactly folded conformations resembling an sU conformation.

Conclusion

This work has shed light on several factors that influence the conformations of MAs. In particular, the folding at the functional groups with extended parallel chains into defined WUZ-folds was determined in vacuum, water and hexane. Utilising the WUZ definitions as primary folding classifications, all MAs afforded W-conformations as the highest percentage, consistent with the W-fold having the lowest energy and thus being the most stable conformer. The lack of folding into WUZ-conformations in the control molecule, BMA, lacking mero-chain functional groups, shows that the functional groups are crucial in creating folding points in the molecule. The unique distribution of conformers obtained for each molecule illustrates that the chemical composition determines the conformational preferences of the MA. In particular, the α-methyl *trans*-cyclopropane group of KMA creates a definite folding point, so that KMA shows an overall preference for more compactly folded conformations, relative to AMA and MMA. This is in agreement with recent results in monolayer experiments and modelling where the *trans*-cyclopropane group is suggested to facilitate folding more than the *cis*-cyclopropane group.[29]

The explicit solvation of MAs showed that the conformational distributions in vacuum and water simulations were very similar, mostly assuming folded conformations around the W-conformation on the principal component plot, although molecules are more flexible to fold and unfold in vacuum. In hexane, mostly open conformations are obtained in a more disperse fashion. KMA is the exception, with a preference for folded conformations resembling sU, even in hexane.

New conformations were identified with alkyl chains that are bent and twisted at various points, creating complex patterns of intertwining chains. These folds tend to be higher in energy due to the higher *gauche*-content in the alkyl chains. These more globular conformations of MAs are in the majority for single molecules and may be of relevance to free MAs occurring in biofilms. However, in the cell wall, the close packing of MAs is expected to facilitate ordering with *trans*-orientation in the alkyl chains.

Acknowledgements

This work was completed with the use of HPC Wales facilities and the Cardiff cluster, Merlin. The authors would like to thank Jurgens de Bruin and James Maskery for encoding python scripts used in this work. Motoko Watanabe, Marion Turner and professor David E. Minnikin are acknowledged for their contribution to the development of the WUZ folding concept. Professor David E. Minnikin, professor Mark S. Baird and professor Jan A. Verschoor are acknowledged for discussions on the modeling of the WUZ folding of MAs.

References

(1) WHO *Global tuberculosis report*, 2013.

(2) Minnikin, D. E.; Goodfellow, M. *Soc. Appl. Microbiol. Symp. Ser.* **1980**, *8*, 189.

(3) Minnikin, D. E.; Kremer, L.; Dover, L. G.; Besra, G. S. *Chem. Biol.* **2002**, *9*, 545.

(4) Verschoor, J.; Baird, M. S.; Grooten, J. *Prog.Lipid Res.* **2012**, *51*, 325.

(5) Watanabe, M.; Aoyagi, Y.; Ridell, M.; Minnikin, D. E. *Microbiology* **2001**, *147*, 1825.

(6) Al Dulayymi, J. R.; Baird, M. S.; Roberts, E. *J. Chem. Soc., Chem. Commun.* **2003**, 228.

(7) Al Dulayymi, J. R.; Baird, M. S.; Roberts, E. *Tetrahedron* **2005**, *61*, 11939.

(8) Al Dulayymi, J. R.; Baird, M. S.; Roberts, E.; Minnikin, D. E. *Tetrahedron* **2006**, *62*, 11867.

(9) Al-Dulayymi, J. R.; Baird, M. S.; Mohammed, H.; Roberts, E.; Clegg, W. *Tetrahedron* **2006**, *62*, 4851.

(10) Al Dulayymi, J. R.; Baird, M. S.; Roberts, E.; Deysel, M.; Verschoor, J. *Tetrahedron* **2007**, *63*, 2571.

(11) Koza, G.; Baird, M. S. *Tetrahedron Lett.* **2007**, *48*, 2165.

(12) Yuan, Y.; Zhu, Y.; Crane, D. D.; Barry, C. E. *Mol. Microbiol.* **1998**, *29*, 1449.

(13) Dubnau, E.; Chan, J.; Raynaud, C.; Mohan, V. P.; Lanèelle, M.-A.; Yu, K.; Quèmard, A.; Smith, I.; Daffè, M. *Mol. Microbiol.* **2000**, *36*, 630.

(14) Peyron, P.; Vaubourgeix, J.; Poquet, Y.; Levillain, F.; Botanch, C.; Bardou, F.; Daffè, M.; Emile, J.-F.; Marcou, B.; Cardona, P.-J.; de Chastellier, C.; Altare, F. *PLoS Pathog.* **2008**, *4*, e1000204.

(15) Glickman, M. S.; Cox, J. S.; Jacobs, W. R. *Mol. Cell* **2000**, *5*, 717.

(16) Rao, V.; Gao, F.; Chen, B.; Jacobs, W. R.; Glickman, M. S. *J. Clin. Invest.* **2006**, *116*, 1660.

(17) Barkan, D.; Hedhli, D.; Yan, H.-G.; Huygen, K.; Glickman, M. S. *Infect. Immun.* **2012**, *80*, 1958.

(18) Beukes, M.; Lemmer, Y.; Deysel, M.; Al Dulayymi, J. R.; Baird, M. S.; Koza, G.; Iglesias, M. M.; Rowles, R. R.; Theunissen, C.; Grooten, J.; Toschi, G.; Roberts, V. V.; Pilcher, L.; Van Wyngaardt, S.; Mathebula, N.; Balogun, M.; Stoltz, A. C.; Verschoor, J. A. *Chem. Phys. Lipids* **2010**, *163*, 800.

(19) Vander Beken, S.; Al Dulayymi, J. R.; Naessens, T.; Koza, G.; Maza-Iglesias, M.; Rowles, R.; Theunissen, C.; De Medts, J.; Lanckacker, E.; Baird, M. S.; Grooten, J. *Eur. J. Immunol.* **2011**, *41*, 450.

(20) Pan, J.; Fujiwara, N.; Oka, S.; Maekura, R.; Ogura, T.; Yano., I. *Microbiol. Immunol.* **1999**, *43*, 863.

(21) Schleicher, G. K.; Feldman, C.; Vermaak, Y.; Verschoor, J. A. *Clin. Chem. Lab. Med.* **2002**, *40*, 882.

(22) Mathebula, N. S.; Pillay, J.; Toschi, G.; Verschoor, J. A.; Ozoemena, K. I. *Chem. Commun.* **2009**, 3345.

(23) Thanyani, S. T.; Roberts, V.; Siko, D. G.; Vrey, P.; Verschoor, J. A. *J. Immunol. Methods* **2008**, *332*, 61.

(24) Zuber, B.; Chami, M.; Houssin, C.; Dubochet, J.; Griffiths, G.; Daffè, M. *J. Bacteriol.* **2008**, *190*, 5672.

(25) Villeneuve, M.; Kawai, M.; Kanashima, H.; Watanabe, M.; Minnikin, D. E.; Nakahara, H. *Biochim. Biophys. Acta, Biomembr.* **2005**, *1715*, 71.

(26) Villeneuve, M.; Kawai, M.; Watanabe, M.; Aoyagi, Y.; Hitotsuyanagi, Y.; Takeya, K.; Gouda, H.; Hirono, S.; Minnikin, D. E.; Nakahara, H. *Biochim. Biophys. Acta, Biomembr.* **2007**, *1768*, 1717.

(27) Villeneuve, M.; Kawai, M.; Watanabe, M.; Aoyagi, Y.; Hitotsuyanagi, Y.; Takeya, K.; Gouda, H.; Hirono, S.; Minnikin, D. E.; Nakahara, H. *Chem. Phys. Lipids* **2010**, *163*, 569.

(28) Villeneuve, M. *Characteristic Conformational Behaviors of Representative Mycolic Acids in the Interfacial Monolayer*. In Cardona, P-J., Ed. *Understanding Tuberculosis - Deciphering the Secret Life of the Bacilli*. InTech 2012, p 317.

(29) Villeneuve, M.; Kawai, M.; Horiuchi, K.; Watanabe, M.; Aoyagi, Y.; Hitotsuyanagi, Y.; Takeya, K.; Gouda, H.; Hirono, S.; Minnikin, D. E. *Microbiology* **2013**, *159*, 2405.

(30) Hasegawa, T.; Nishijo, J.; Watanabe, M.; Funayama, K.; Imae, T. *Langmuir* **2000**, *16*, 7325.

(31) Hasegawa, T.; Amino, S.; Kitamura, S.; Matsumoto, L.; Katada, S.; Nishijow, J. *Langmuir* **2003**, *19*, 105.

(32) Hasegawa, T.; Leblanc, R. M. *Biochim. Biophys. Acta* **2003**, *1617*, 89.

(33) Hasegawa, T. *Biopolymers* **2004**, *73*, 457.

(34) Sekanka, G.; Baird, M. S.; Minnikin, D. E.; Grooten, J. *Expert Opin. Ther. Pat.* **2007**, *17*, 315.

(35) Groenewald, W.; Baird, M. S.; Verschoor, J. A.; Minnikin, D. E.; Croft, A. K. *Chem. Phys. Lipids* **2014**, *180*, *15*.

(36) Tucker, D. W. A.; Version 10.1.002e ed.; WebMO, LLC: Queens University of Charlotte, Charlotte, NC, 2004.

(37) Youngs, T. *J. Comput. Chem.* **2010**, *31*, 639.

(38) Apol, E.; Apostolov, R.; Berendsen, H. J. C.; Buuren, A. v.; Bjelkmar, P.; Drunen, R. v.; Feenstra, A.; Groenhof, G.; Kasson, P.; Larsson, P.; Meulenhoff, P.; Murtola, T.; Pall, S.; Pronk, S.; Schulz, R.; Shirts, M.; Sijbers, A.; Tieleman, P.; Hess, B.; Spoel, D. v. d.; Lindahl, E.; VERSION 4.5.4 ed. University of Groningen, The Netherlands, 1991-2000.

(39) Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Model.* **2001**, *7*, 306.

(40) Van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A.; Berendsen, H. *J. Comput. Chem.* **2005**, *26*, 1701.

(41) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theor. Comput.* **2008**, *4*, 435.

(42) Wang, L. P.; Van Voorhis, T. *J. Chem. Phys.* **2010**, *133*.

(43) Ojha, A. K.; Baughn, A. D.; Sambandan, D.; Hsu, T.; Trivelli, X.; Guerardel, Y.; Alahari, A.; Kremer, L.; Jacobs Jr., W. R.; Hatfull, G. F. *Mol. Microbiol.* **2008**, *69*, 164.

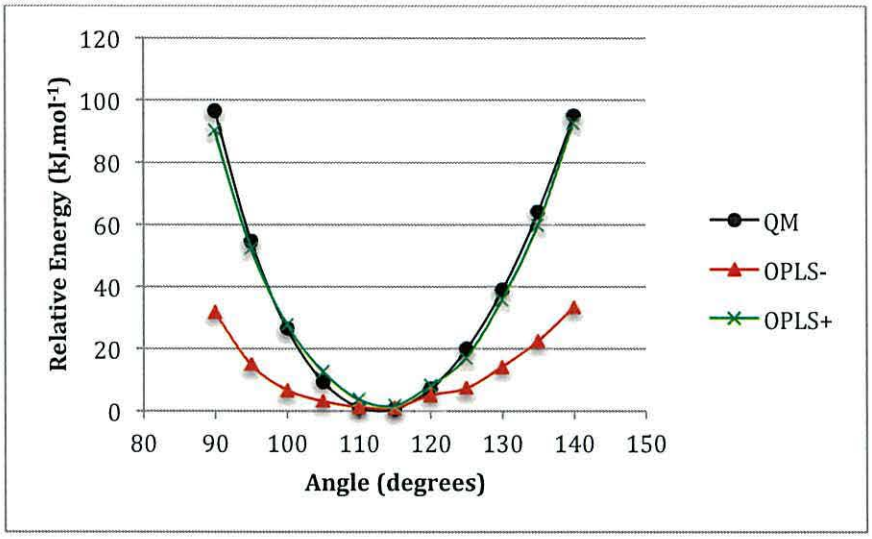
(44) Sambandan, D.; Dao, D. N.; Weinrick, B. C.; Vilcheze, C.; Gurcha, S. S.; Ojha, A.; Kremer, L.; Besra, G. S.; Hatfull, G. F.; Jacons, W. R. J. *mBioio* **2013**, *4*, e00222.

Supporting Information

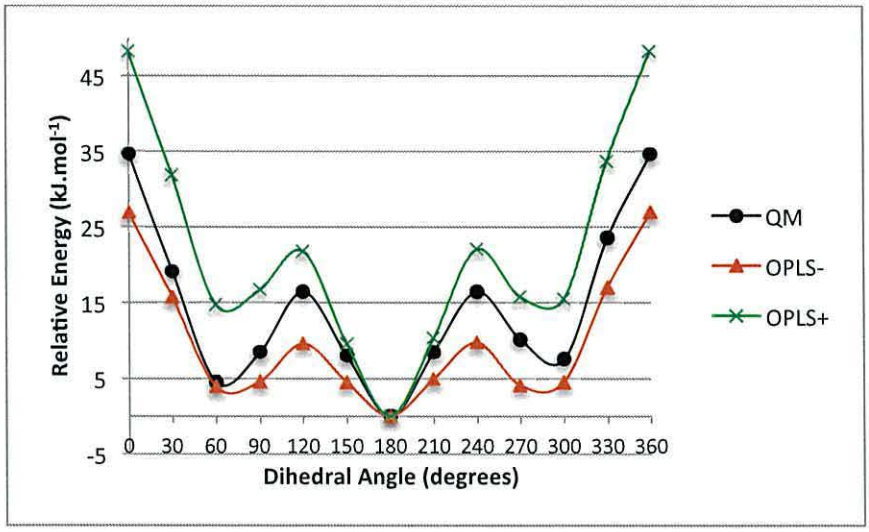
Computational Investigation of Solvent and Structural Effects on Mycolic Acid Conformation

W. Groenewald and A. K. Croft

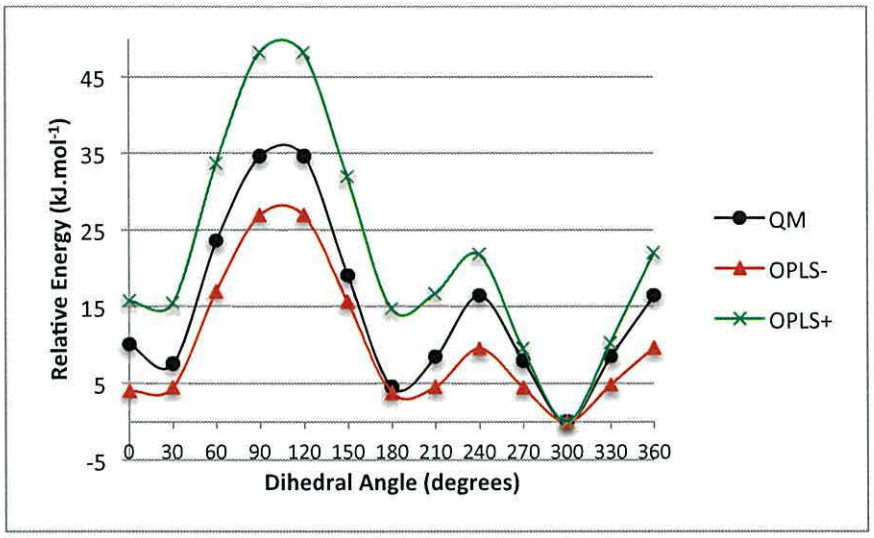
Contents



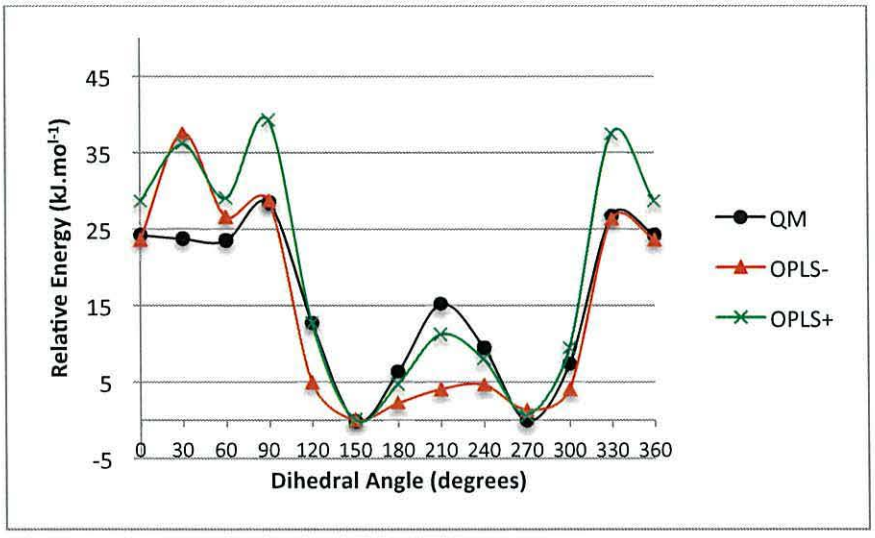
Supplementary figure S1: Optimised potential for the CT-CT-CY angle using a hybrid ensemble, with OPLS- the initial guess and OPLS+ after fitting to the QM potential.



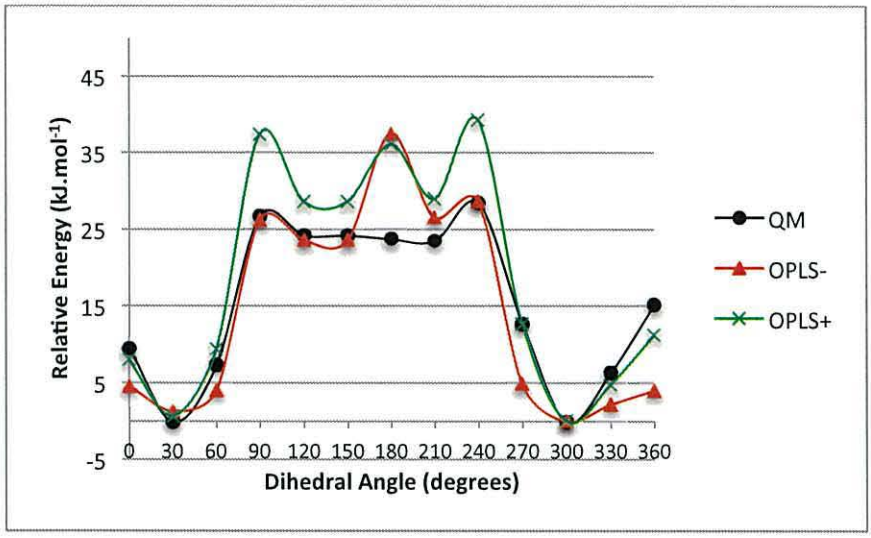
Supplementary figure S2: Optimised potential for the CT-CT-CT-CY dihedral angle using a hybrid ensemble with OPLS- the initial guess and OPLS+ after fitting to the QM potential.



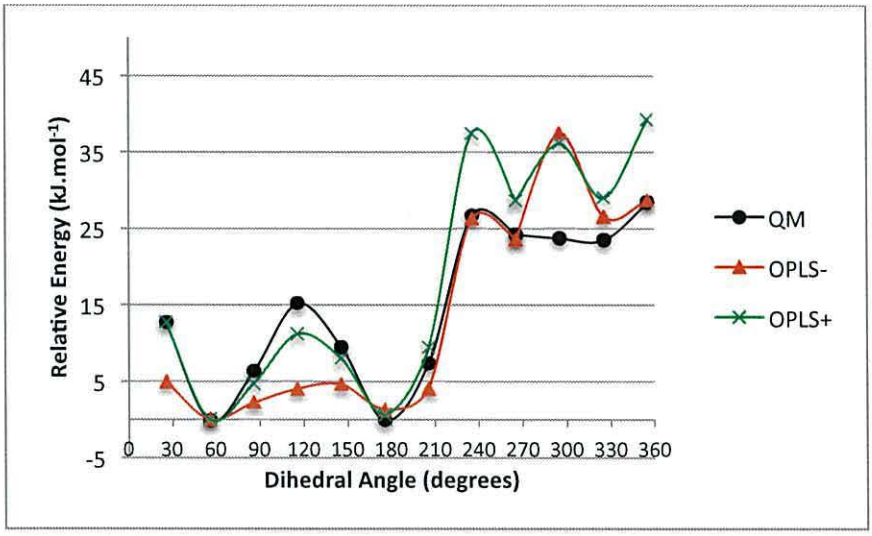
Supplementary figure S3: Optimised potential for the CT-CT-CT-HC dihedral angle using a hybrid ensemble with OPLS- the initial guess and OPLS+ after fitting to the QM potential.



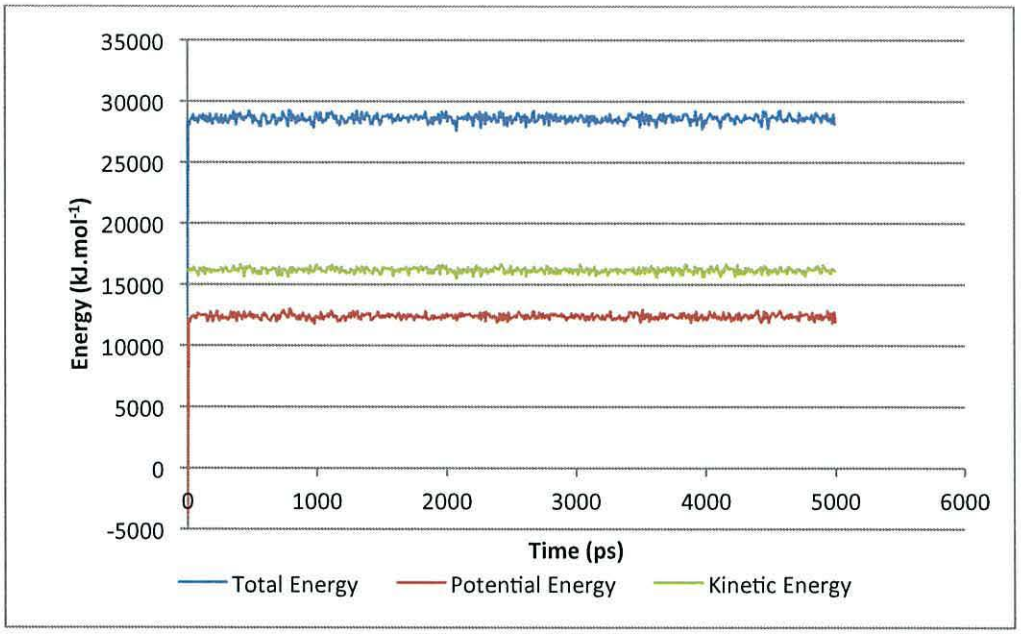
Supplementary figure S4: Optimised potential for the CT-CT-CY-CY dihedral angle using a hybrid ensemble with OPLS- the initial guess and OPLS+ after fitting to the QM potential.



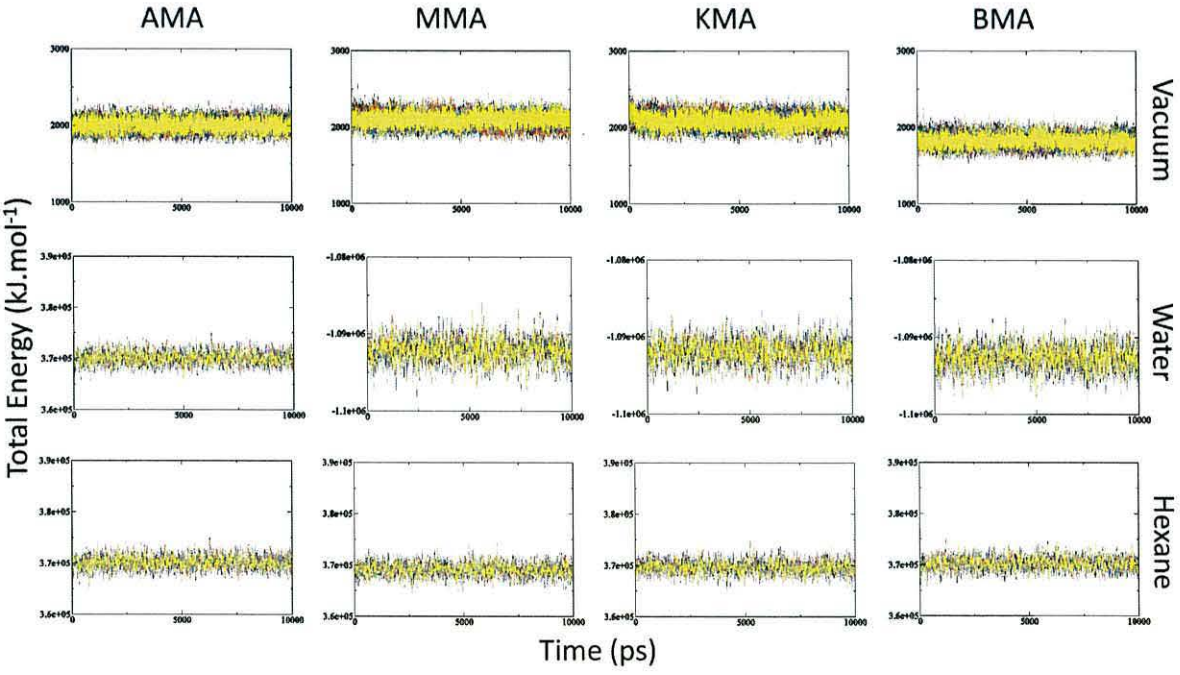
Supplementary figure S5: Optimised potential for the CT-CT-CY-HC dihedral angle using a hybrid ensemble with OPLS- the initial guess and OPLS+ after fitting to the QM potential.



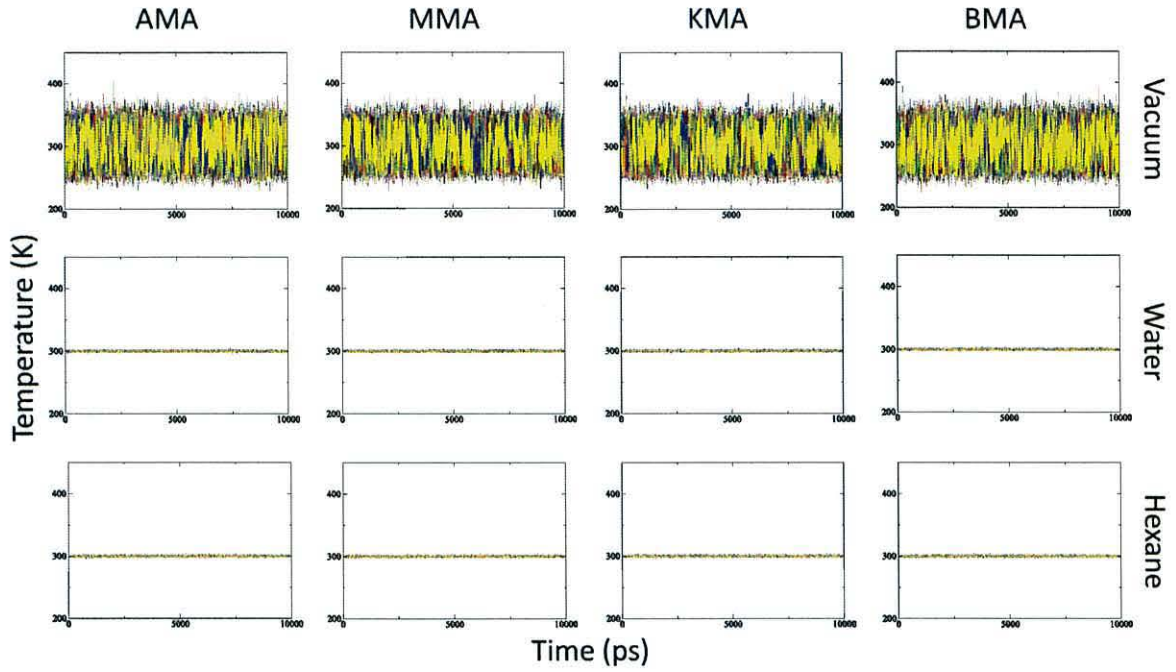
Supplementary figure S6: Optimised potential for the HC-CT-CY-HC dihedral angle using a hybrid ensemble with OPLS- the initial guess and OPLS+ after fitting to the QM potential.



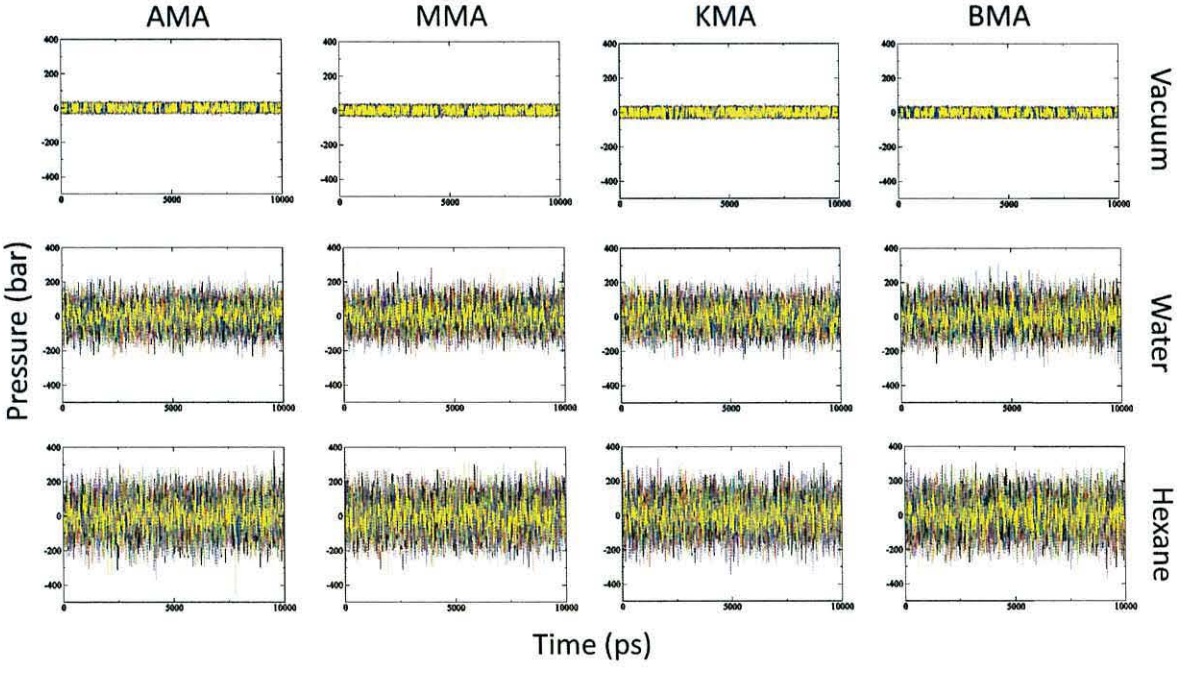
Supplementary figure S7: Total energy, potential energy and kinetic energy plots for the 5 ns NVT equilibration of the hexane solvent box.



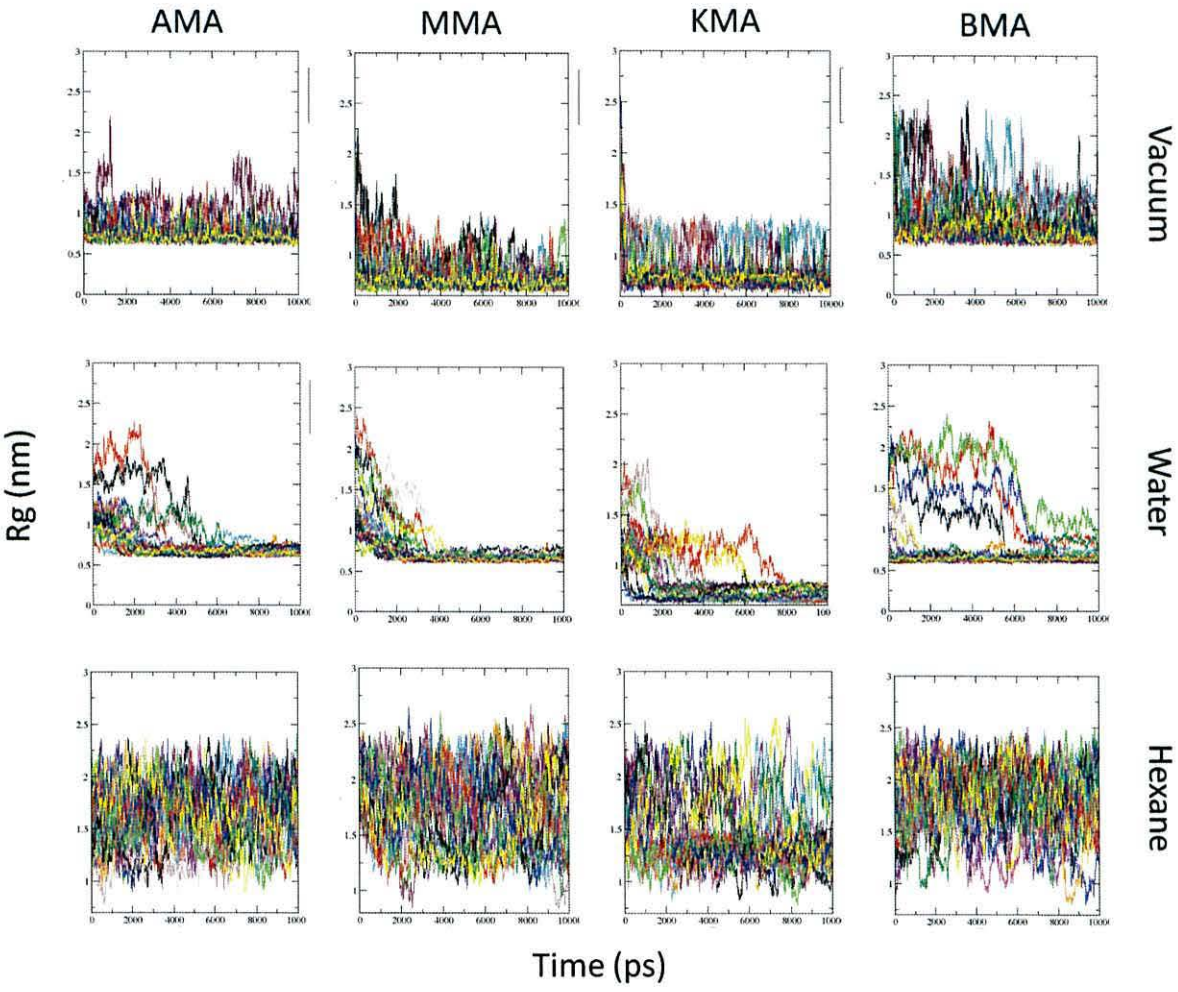
Supplementary figure S8: Total energy plots for all molecules studied in vacuum, water and hexane



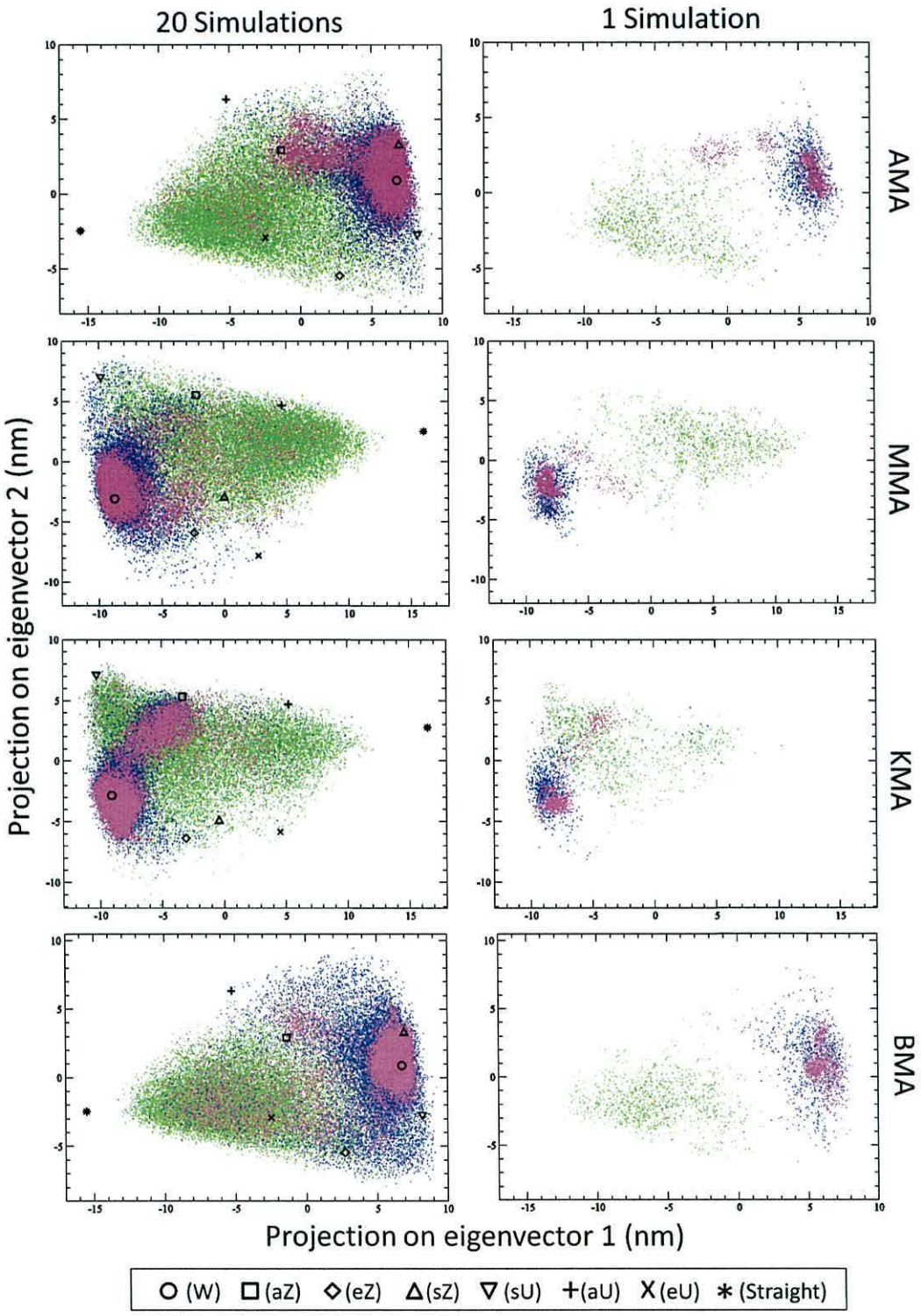
Supplementary figure S9: Temperature plots for all molecules studied in vacuum, water and hexane.



Supplementary figure S10: Pressure plots for all molecules studied in vacuum, water and hexane.



Supplementary figure S11: Radius of gyration plots for all molecules studied in vacuum, water and hexane.



Supplementary figure S12: Principal component plots for all molecules modelled in vacuum (blue), water (magenta) and hexane (green). Frames for all frames of all twenty replicate simulations (left) and frames for a single simulation from each solvent (right) are shown.

Manuscript III

Monolayer Simulation of Coarse Grained Alpha-Mycolic Acid

Wilma Groenewald,[†] Anna K. Croft[†a] and Siewert-Jan Marrink[‡]*

[†]*School of Chemistry, Bangor University, Bangor, Gwynedd, LL57 2UW, United Kingdom*

[‡]*Groningen Biomolecular Sciences and Biotechnology Institute and Zernike Institute for Advanced Materials, University of Groningen, Nijenborgh 7, 9747 AG Groningen, The Netherlands*

[a]*Current address Department of Chemical and Environmental Engineering, University of Nottingham, University Park, Nottingham NG7 2RD, United Kingdom*

Abstract

Mycobacterium tuberculosis, the organism responsible for TB infection in humans, causes up to 3800 deaths per day. The mycobacterium is inherently resilient against host defences and anti-TB drugs. This persistence is attributed partly to the presence of lipids, such as mycolic acids (MAs), which make the cell wall impermeable. A coarse grained model for a representative alpha-MA (AMA) from *Mycobacterium tuberculosis* using the MARTINI force field is presented. The model is used to simulate monolayers of different sizes; a small monolayer consisting of 220 MAs and a large monolayer consisting of 1972 MAs. The monolayers were simulated at various temperatures (283 K, 300 K and 317 K) in order to compare with existing experimental monolayer data for AMA. The coarse grained model cannot include the explicit stereochemistry of the functional groups in the MA. Despite this, the model could replicate key features of experimental monolayers such as the changes in surface tension with temperature, phase changes and the collapse point. The monolayers were found to be in the liquid expanded and liquid condensed phases, in correlation with experimental data. By studying the conformation of MAs in the simulated monolayers, it was found that AMA did not fold into the W-conformation at large surface areas, but was only folded at the head group to give a U-shape. At large surface areas the distance between points in the two chains was as large as

3.64 nm. On monolayer compression, the MA chains came closer together with the distance between the points in the two chains decreasing to 1.76 nm as the molecules were packed compactly and formed an ordered monolayer before collapsing.

Introduction

In 2012, 1.3 million people died from tuberculosis (TB),[1] amounting to over 3500 deaths per day, which is shocking in view of the fact that TB is largely a curable disease. There are numerous factors that complicate the control of TB. These include co-infection with human immunodeficiency virus (HIV) as well as development of drug resistant and extremely drug resistant strains.[1] In addition, long treatment regimes (at least 6 months) on a combination of drugs with severe side-effects leads to patient non-compliance. The sensitivity of some diagnostics (especially in HIV-co-infected patients) is low and the long time required for confirmed diagnosis by culture (2–6 weeks) plus an additional 4–6 weeks for resistance determination by bacterial culture[2,3] complicates matters further.

Mycobacterium tuberculosis, the organism causing TB infection in humans, is inherently resilient against the host defences and anti-TB drugs. This is due, in part, to its lipid-rich cell wall making it impermeable to currently employed drug substances. Mycolic acids (MAs) are β–hydroxy acids of up to 90 carbons long and make up part of the outer layer of the cell wall. MAs are covalently attached to arabinogalactan where they are thought to form an impermeable monolayer arrangement, but they also occur as part of the outermost layer of unbound lipids as sugar-esters.

There are three main types of MAs in *M. tb,* as illustrated in Figure 1A. All three types share a common 24-carbon alpha chain and β–hydroxyl acid with *R,R* stereochemistry. The distinction between the three types of MA is in the functional groups in the mero-chain (Figure 1A X,Y). Alpha-MA (AMA), the most abundant MA type in *M. tb,* has *cis*-cyclopropane groups at both the proximal and distal functional groups (Figure 1A X and Y, respectively) whereas methoxy- and keto-MAs may have either *cis*- or α-methyl trans-cyclopropane at

the proximal position (Figure 1A, X) and a methoxy-methyl or keto-methyl group in the distal position, respectively (Figure 1A, Y).[4]

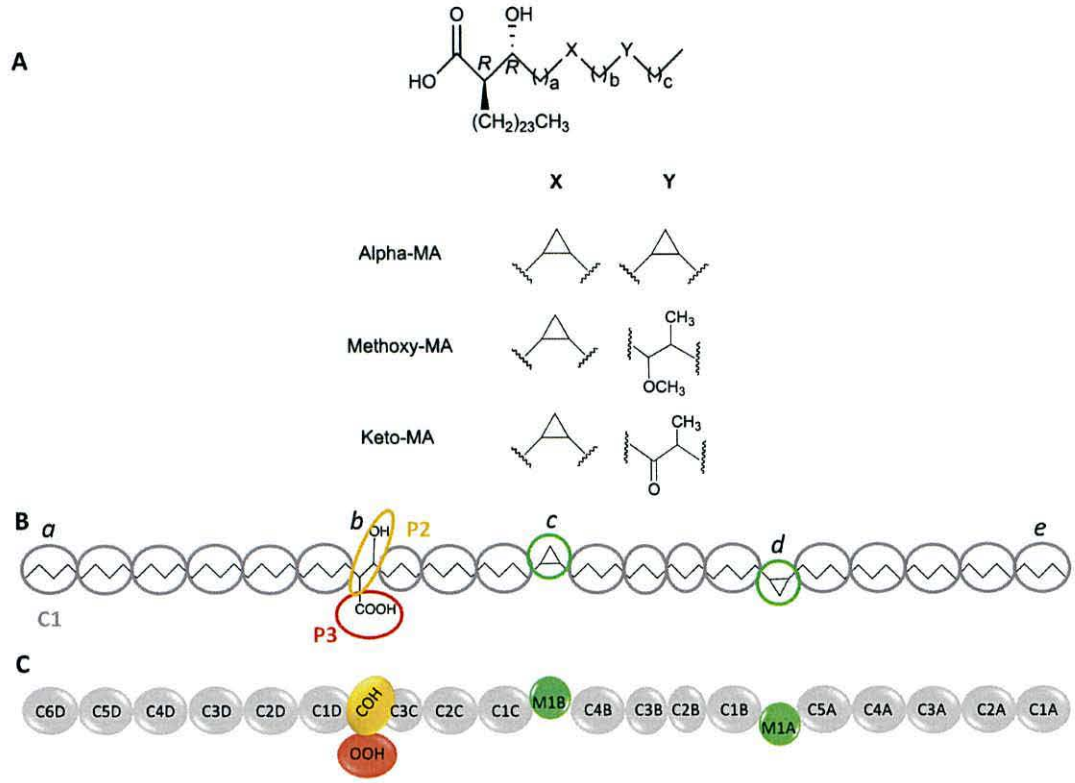


Figure 1: The general structure of MAs found in *M. tb* (A), coarse grained beads, and bead types mapped onto the all-atom structure of AMA (B) and the names of coarse grained beads for AMA (C).

The precise structural and functional roles of the different MA types are only beginning to be unravelled. It has been shown that different MA types have different effects on the immune system[5-10] and that changing the MA composition alters the outcome of an infection as well as the cell wall permeability.[11-17] Despite these advances, it is still uncertain how MAs are arranged in the cell wall. Zuber *et al.*[18] established that the bilayer of the outer membrane in which MAs are found is 7–8 nm thick. Some of the longer MAs would not be able to fit into this space without folding or bending at some point. Therefore the authors proposed a zipper model, in which MAs would not only have to fold at each of

the functional groups, but also intercalate with lipids in the outer leaflet in order to fit into this space. When the MA is folded at all three functional groups, it creates a W-shape when mapped out in two dimensions, as shown in Figure 2 (W). We have recently defined a range of other folds accessible for MAs from *M. tb* that are based on folding at the functional groups.[19]

A monolayer of MAs on water has been suggested as a suitable model for the tightly packed MA layer bound to arabinogalactan in the cell wall. Several such monolayer experiments have been performed[20-27] and some include a range of different temperatures.[24-27] The general trend in all these experiments is for MAs to have a folded conformation at large surface areas and to unfold as the monolayer is compressed, until the mero-chain is fully extended (with the exception of keto-MA, which stays folded, even at high lateral pressures).[20-27]

Figure 2 shows a schematic of the unfolding that is expected in a MA monolayer, starting from a folded W-conformation, which has a large surface area with four parallel chains. Assuming that the head group will stay in contact with the water, an extension at the proximal cyclopropane group, *c*, seems the most rational option on compression. This results in a shape that is folded at the head group (*b*) and the distal cyclopropane group (*d*), which will be referred to as sZ.[19] The Z-fold represents a molecule folded at two functional groups with three chains in parallel, and sZ has symmetry in its two-dimensional folding representation. Although an aZ fold, with *a* at its extended chain terminus, is also a possible conformation it is not included here as it is assumed that the molecule has to bend at *b* in order for the head group to be in contact with the water interface. It is also possible for the molecule to extend at the distal cyclopropane group, staying folded at group *b* and *c* (eZ, with its extended chain terminating with *e*) as indicated by the dashed arrow in Figure 2. However, this is thought to be unlikely, as the mero-chain would not have space to extend past the water interface.

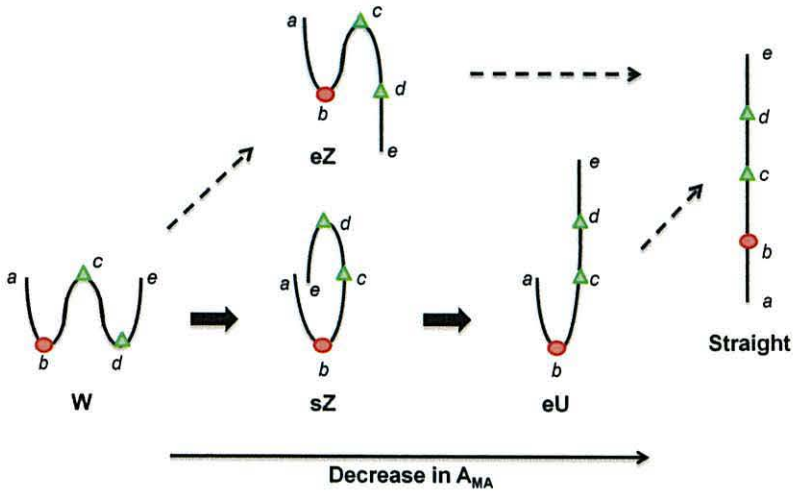


Figure 2: Schematic representation of MA conformations during monolayer compression. Points *a*-e are defined as the chain termini of the alpha- and mero-chains (*a* and *e*, respectively) and the functional groups where bending is postulated to occur; the polar head group (b) and the proximal and distal mero-functional groups (*c* and *d*, cyclopropanes for AMA). Starting from a folded W-shape at large surface areas, the molecule is thought to unfold into Z- and U-shapes that occupy smaller surface areas on monolayer compression. Dashed arrows indicate routes of unfolding that are less likely to occur.

On further compression it is expected that the mero-chain will extend fully so that the molecule now has a U–shape, only bending at one functional group, having two parallel chains. In a monolayer, eU is the only sensible U-shape, bending at the head-group with *e* at the extended chain terminus. The eU conformation only has two parallel chains, hence occupying a smaller surface area. Any further extension in the molecule would lead to a straight MA, which is fully stretched out without bending at any point. A straight conformation is not expected in a monolayer while the head group is assumed to be at the water interface.

Due to the very long carbon chains, MAs are extremely hydrophobic and hence insoluble in the majority of solvents. Being waxy, they also do not crystalise, making it very difficult to get structural data experimentally. For this reason we

explore MA conformations and packing effects through simulations, and the monolayer data available from experiments makes this an ideal system to model. Previous modelling of MAs is largely on single molecules.[19,25,26] These models give valuable insight to the preference of different MAs for the W-conformation, and it has recently been shown that an α-methyl *trans*-cyclopropane group in oxygenated MAs promotes tight folding into a W-conformation more than a *cis*-cylcopropane group.[27] It has also been shown that MA structure determines its conformational behaviour.[19] Although single molecules in vacuum were seen to adopt folded conformations such as the W-conformation spontaneously,[19] single molecules are not an accurate representation of MAs in their natural cell wall environment. MAs have also been modelled in cell wall components in which MAs were arranged as a monolayer.[28] However, in these models, MAs were truncated to have mero-chains of the same length as the alpha-chains, and hence these models could not show how MAs would fold and pack in the cell wall environment.

Here we apply the existing MARTINI[29,30] coarse grained force field to create a model for AMA, the most abundant MA in the *M. tb* cell wall. The model is used to construct AMA monolayers of different size, which are simulated at various temperatures. The MARTINI model was originally designed for the modelling of lipids and has been used to successfully model various lipid systems.[30] It has also been shown to reproduce lipid monolayer properties well, although quantitative reproduction of experimental isotherms remains challenging.[30,31] The coarse grained model has the advantage of having fewer interactions to calculate, making it more feasible to simulate large systems for longer timescales. However, in the coarse grained model the explicit stereochemistry of the functional groups in the MA will be lost and it remains to be seen how this affects the accuracy of the model.

Methods

The model

A mapping of 3 to 4 heavy atoms to one CG bead was used (Figure 1, B). All beads were standard MARTINI C1 (grey), P2 (orange) and P3 (red) types (Figure 1, B and C). Cyclopropane beads, M1A and M1B (green) are also C1 beads. Only the bonded interactions were adjusted in order to resemble reference bond and angle distributions from mapped atomistic simulations as accurately as possible.

Parameterisation of bonded interactions

CG simulations in vacuum were run using a leap-frog stochastic dynamics integrator. Other simulation parameters in the CG simulations were standard for MARTINI, as outlined below. Bonded parameters were adjusted so that angle and bond distributions from 100 ns CG simulations resembled the reference mapped atomistic distributions from simulations performed in vacuum, water and hexane. Butane was used as CG solvent to compare to the hexane atomistic solvent simulations. Details of atomistic simulations are given below.

Systems simulated

The monolayer setup consisted of a central slab of water with an AMA monolayer on either side, separated from periodic images in the z-direction by vacuum. Monolayers were constructed using the membrane-generating program Insane (courtesy of Tsjerk A. Wassenaar, University of Gröningen, manuscript in preparation). MAs were inserted into the monolayer in an approximate W-fold. Two monolayer setups were used. The small monolayer system contained 110 MAs in each monolayer (220 in total) and 4073 water beads. The box size was initially 12 × 10.39 × 50 nm, which leads to an area per MA (A_{MA}) of 1.134 nm^2. The large monolayer system contained 986 MAs in each monolayer (1972 in total) and 39169 water beads. The box was 34 × 29.4 × 200 nm, which gave an initial A_{MA} = 1.02 nm^2.

The small monolayer was equilibrated at constant volume and temperature for 1 μs. Then the monolayer was compressed and allowed to equilibrate in a stepwise manner by alternating compression simulations with constant volume and temperature simulations, in order to allow the MAs to rearrange. (Continuous compressions were also performed with notable difference only at small A_{MA}, Supplementary figure S4). The small monolayer was simulated at three different temperatures; 283 K, 300 K and 317 K. At 317 K, the monolayer was distorted after the first 1 μs, with the MAs aggregating and open patches of water forming. Therefore, simulations at 317 K were started from $A_{MA} = 1.02$ nm^2 of which the starting structure was obtained from the 300 K simulation after the first compression.

Compression was performed to reduce the surface area by 10 %, followed by a simulation at constant volume and temperature. Compressions were performed by setting the surface tension to -100 bar.nm (-50 bar.nm per surface) in the x- and y-directions and a compressibility of 5×10^{-5} bar^{-1}. The pressure for the z-direction was obtained from the initial equilibration simulation at constant volume and temperature (0.00931365 bar at 283 K, 0.136525 bar at 300 K, and 0.28539 bar at 317 K) with the compressibility in the z-direction set to zero in order to keep the box size in the direction normal to the monolayer fixed.

Simulations at constant volume and temperature were initially 1 μs long. From $A_{MA} = 0.79$ nm^2 and smaller, the constant volume and temperature simulations were increased to be 5 μs long, allowing more time for the molecules to rearrange in the monolayer. The initial surface tension of -50 bar.nm per surface correlates to a lateral pressure of ~2 bar. The relationship between the pressure and surface tension is given by $\gamma_m = (P_N - P_L) \times L_Z/2$ where P_N and P_L are the normal and lateral pressures of the box with $P_L = (P_{xx} + P_{yy})/2$ and L_z is the box size in the z-direction. For simulations at 283 K and 300 K, the monolayer became hard to compress at lower surface areas (even up to 10 μs simulation would not yield the 10 % decrease in surface area) and therefore the surface pressure was increased as required, in order to yield the 10 % lower surface area within 1 μs simulation. The surface tension used for each compression is listed in Supplementary Table

S1. After reaching $A_{MA} = 0.23\ nm^2$ compression was not achieved with surface tension up to -4000 bar.nm (-2000 bar.nm per surface, ~80 bar).

The large monolayer was simulated in the same stepwise approach at 300 K. Constant volume and temperature simulations were performed for 2 μs before the next compression. A surface tension of -100 bar.nm (-50 bar.nm per surface), correlating to ~0.6 bar lateral pressure, was applied during the compression simulations. Once $A_{MA} = 0.51\ nm^2$ was reached the surface tension was increased to -2500 (-1250 bar.nm per surface, ~12.6 bar) in order to achieve the desired 10 % compression within 1 μs simulation. The surface tension used for each compression is listed in Supplementary Table S2.

Surface tension, order parameters, density profiles, average structures and principal component analysis were calculated from the last 500 ns of the simulation at constant volume and temperature at the relevant A_{MA}. Principal component analysis for the large monolayer was performed on every 100th frame (every 2 ns).

Reference values were obtained by modifying the small monolayer setup. To obtain the water-vacuum surface tension, all MAs were removed and a 50 ns simulation was run with NVT. For the lipid chain-vacuum surface tension, the lipid head groups (P2 and P3 beads) were removed and the molecule linked through the C1 beads. The simulation was run at NVT. Beads at the chain termini as well as at each of the three folding points were position restrained with a force of 1000 $kJ.mol^{-1}.nm^{-2}$ in the z-direction, so that the monolayers do not merge. For the lipid chain-water surface tension, the box used in the lipid chain-vacuum simulation was shortened to 15 nm in the z-direction and filled with water. The simulation was run for 50 ns using semi-isotropic coupling with a pressure of 1 bar in the z-direction and zero compressibility for the x and y direction in order to keep the surface area fixed.

Coarse grained simulation parameters

Molecular dynamics simulations were performed using Gromacs 4.5.5[32] and images were rendered using VMD.[33] The standard simulation parameters for the MARTINI force field were used. Lennard-Jones interactions were shifted to zero

between 0.9 and 1.2 nm and the Coulomb potential was shifted to zero between 0 and 1.2 nm. The relative dielectric constant was 15 and a stepsize of 20 fs was used in all simulations.

All-atom simulation parameters

Atomistic simulations of AMA were performed for 100 ns in vacuum, water and hexane. The OPLS all atom (AA) force field[34] (with additional parameters for cyclopropane, obtained using hybrid ensembles for force matching, shown in Supplementary table S7 and Supplementary table S8)[35] was used and each molecule was placed in the centre of a 10 × 10 × 10 nm box and minimized with a steepest descent algorithm with a maximum stepsize of 0.01 nm, a maximum number of 200 000 steps and a tolerance of 10 $kJ.mol^{-1}.nm^{-1}$. A timestep of 1 fs was used and the neighbour list was updated every 10 fs. Van der Waals interactions were modelled by using a shift function between 0.8 and 0.9 nm and electrostatic interactions were modelled by using PME. V-rescale temperature coupling was used at 300 K with a time constant of 0.5 ps and no constraints were applied. The equation of motion was integrated using a leap-frog algorithm. All atomistic simulations were performed with Gromacs version 4.5.4.[32]

Vacuum simulations were performed with a NVT ensemble.
TIP4P water was used to fill the simulation box for simulations in water. Equilibration was performed with position restraints by applying a force of 1000 $kJ.mol^{-1}.nm^{-2}$ in the x-, y- and z-directions on all carbon and oxygen atoms of the MA. The system was equilibrated by performing 100 ps NVT, followed by 50 ps NPT using Berendsen pressure coupling to scale the box in an efficient way at the beginning of the simulations, and lastly 100 ps NPT using Parinello-Rahman pressure coupling in order to yield the correct ensemble. In both of the NPT-ensembles isotropic pressure coupling at a pressure of 1 bar was used with a 1 ps time constant. Production simulations for MAs in water also used the Parinello-Rahman setup.

A hexane solvent box was built and equilibrated before addition to MA simulation boxes. A 3.6 × 3.6 × 3.6 nm box was filled with 216 hexane molecules, minimised (as described before) and equilibrated at 300 K using a NVT ensemble for 5 ns. The density of the hexane box was 584.201 $kg.m^{-3}$, which is approximately 11 % lower than the experimental density for hexane. The simulation box containing AMA was filled with hexane using the equilibrated hexane box. Equilibration was performed with position restraints by applying a force of 1000 $kJ.mol^{-1}.nm^{-2}$ in the x-, y- and z-directions on all carbon and oxygen atoms of the MA. The system was equilibrated by first increasing the timestep from 10^{-7} ps to 10^{-3} ps in 10-fold increments by running respective simulations of 100 000 steps at 0 K. Secondly, using a 10^{-3} ps timestep, the temperature was increased in 50 K increments from 0 K to 250 K with consecutive 100 ps simulations and a 1 ns simulation at 300 K. Simulations were run using Berendsen pressure coupling at a pressure of 1 bar and a time constant of 1 ps. Production simulations were run using a timestep of 2 fs and with all bonds constrained.

Results and Discussion

Optimisation of bonded parameters

Resemblance of CG distributions to vacuum and hexane atomistic distributions were prioritised over those in water, since they more closely approximate the environment in the monolayer simulation where AMA will be at the water-vacuum interface and each molecule will be adjacent to other identical, non-polar molecules. Bond force constants for CG AMA were higher than usual for MARTINI, since distributions were narrow – especially around the cyclopropane and two polar head group beads. In general, the angle force constants were also higher than usual for MARTINI beads.

Generally, bond and angle distributions correlated well with the mapped atomistic distributions, as shown in Figure 3, Supplementary figure 5 and Supplementary figure 6. The performance of the model in mimicking conformations similar to the reference AA simulations can be shown in the

subsequent dihedral distributions in Figure 4 and in the projection of conformers onto principal component (PC) axes as in Figure 5. In the PC plot each point represents a frame from a single MA simulation. The points to the left of the plot represent folded conformations, while those on the right tend towards a stretched out conformation. It is clear that in hexane MA conformations are more extended, in contrast to water and vacuum where the molecules tend to be folded.

The angle OOH-COH-C1D did not have a very good resemblance with reference distributions (Figure 3, E). This compromise had to be made in order to gain better resemblance of the other bonds and angles including beads COH, OOH and C1D. The bonded interactions involving the carbon chain that would determine AMA conformation were favoured over the peripheral OOH-COH-C1D angle because this angle is less influential on the AMA conformation, with the P1 and P2 beads likely to be in contact with the water beads in the monolayer.

The two cyclopropane beads, M1A and M1B, were treated as identical in this model, with identical adjacent beads (including mapping of the beads) and identical bond and angle parameters. The CG model reproduces the atomistic angle distributions very well in hexane. However, the angles that centre on the M1A and M1B beads are represented less accurately in vacuum and in water (Figure 3, C and D) where it can be seen that smaller angles are more likely for the distal cyclopropane (M1A) than the proximal (M1B), and this difference is missed by the CG distribution. It remains to be shown whether this shortcoming will alter the conformational behaviour of the individual molecule, and hence of the monolayer properties. It should be noted that smaller angles are excluded from the CG distributions and hence folding at the cyclopropane beads may be under-represented – more so for the distal group, M1A.

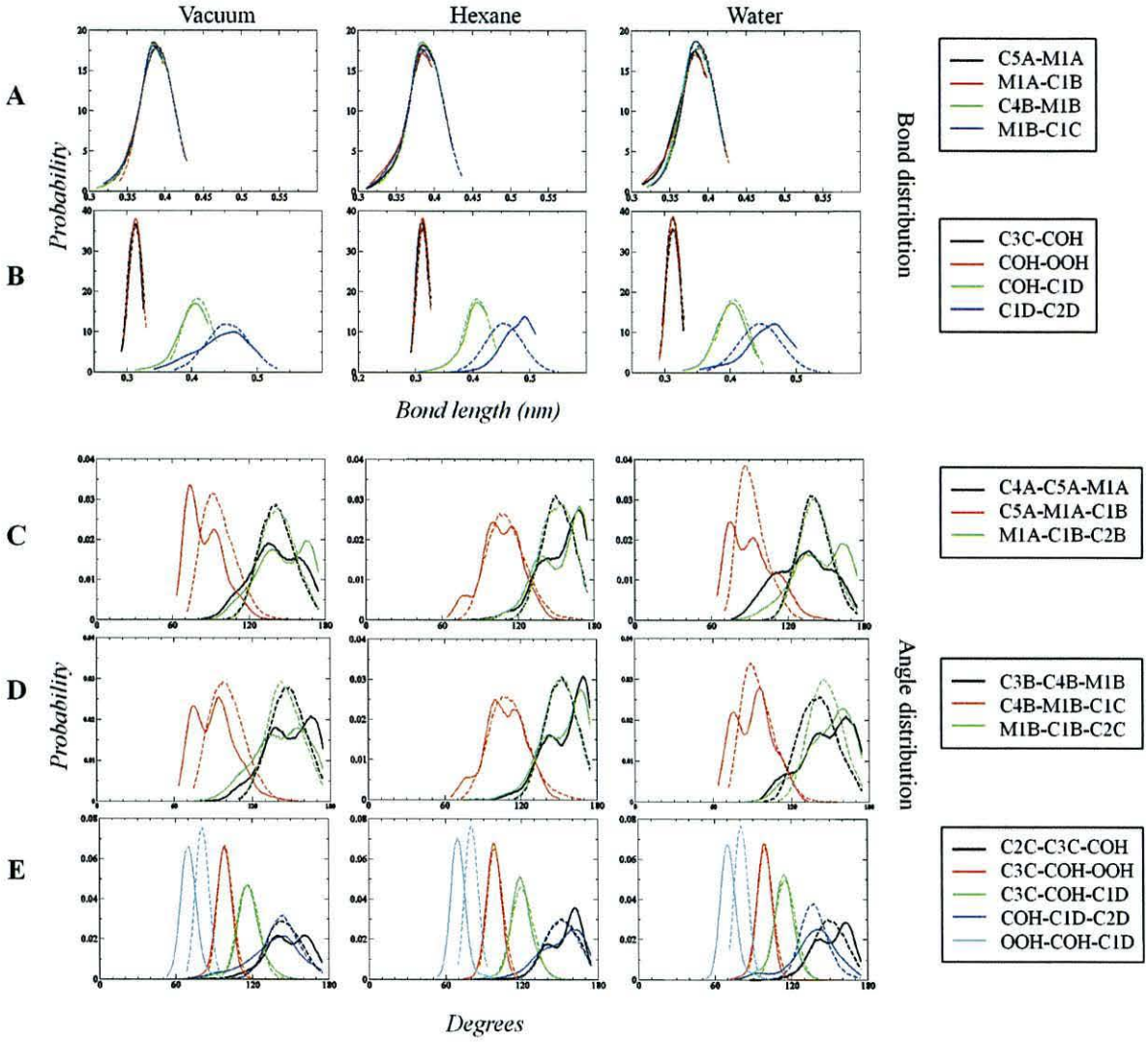


Figure 3: Selected bond and angle distributions around the functional groups that were optimized in vacuum, hexane/butane and water for AMA. Solid lines represent the AA distribution as mapped onto CG beads and dashed lines represent the CG model.

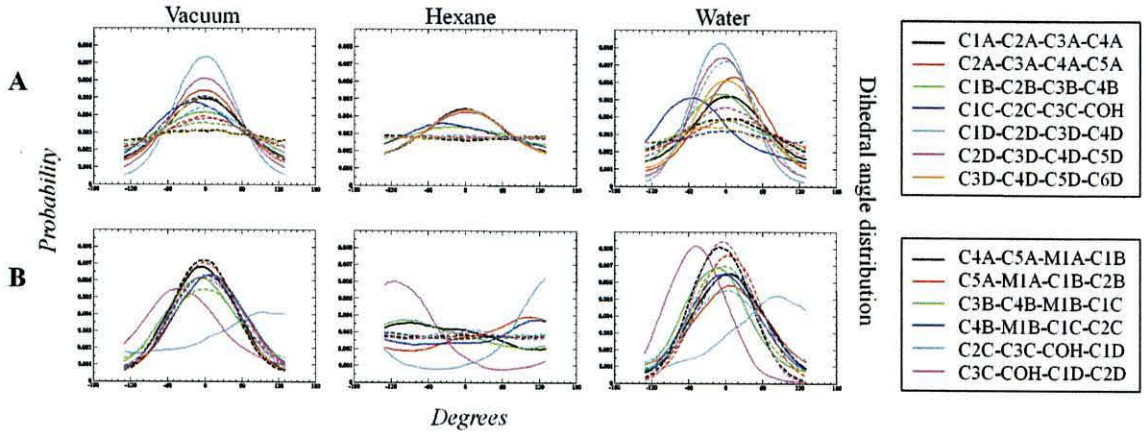


Figure 4: Selected dihedral angle distributions for alkyl chains (A) and dihedral angles involving the functional groups (B) from vacuum, hexane/butane and water for AMA, illustrating the performance of the CG model. Solid lines represent the AA distribution as mapped onto CG beads and dashed lines represent the CG model.

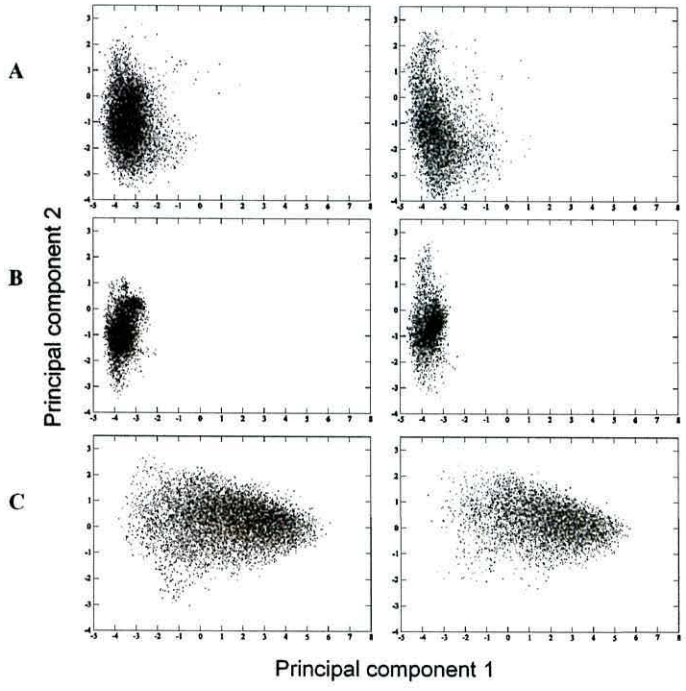


Figure 5: Conformations of AMA shown on the same set of principal component axes in vacuum (A), water (B) and hexane/butane (C). Conformers from the AA reference simulation are shown on the left and the CG model on the right.

Alpha-MA isotherms at different temperatures

Two monolayer systems were simulated: a small system containing 220 MAs (Figure 6) and a large system containing 1972 MAs (Figure 7). The small system was simulated at 283 K, 300 K and 317 K in order to examine the effect of temperature on the monolayer. The large system was simulated at 300 K.

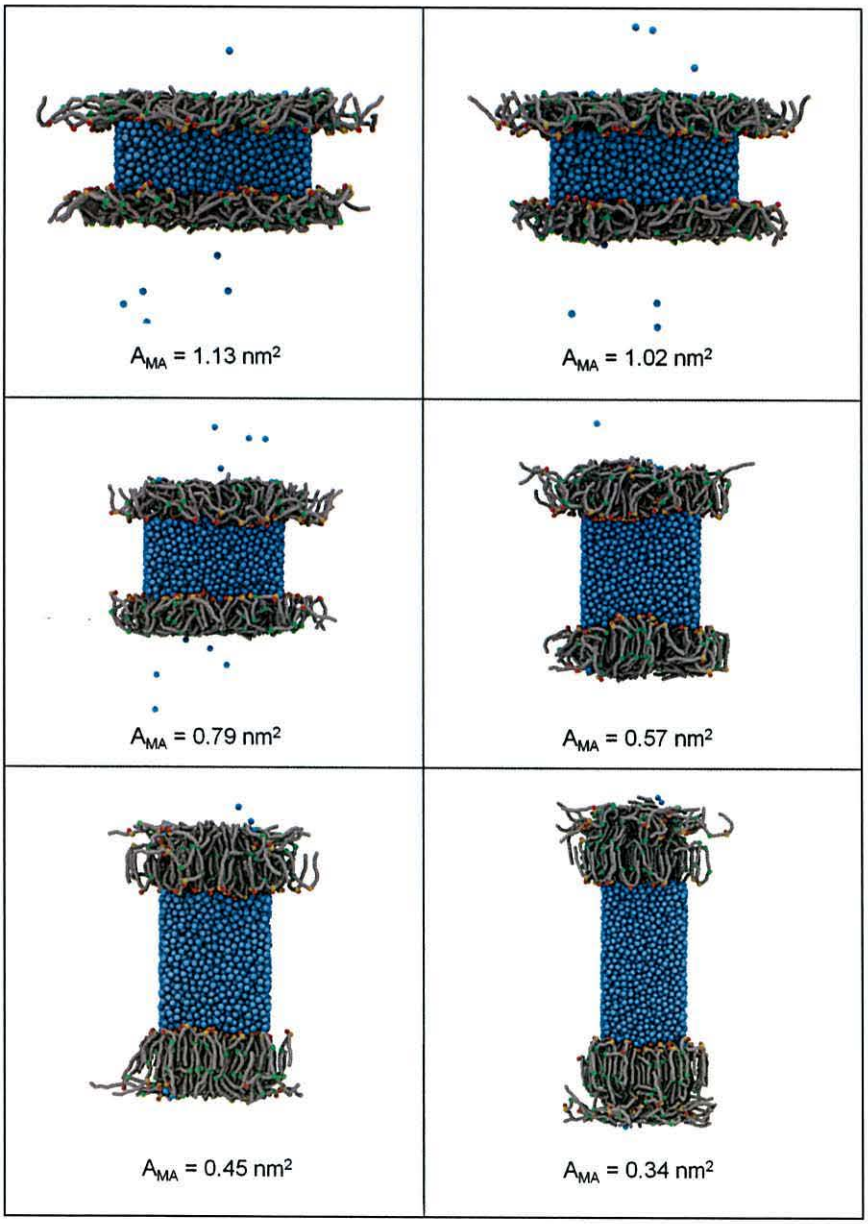


Figure 6: Snapshots of the small monolayer simulated at 300 K, at various A_{MA}, as indicated.

The surface tension-area isotherm for the small monolayer system is shown in Figure 8A. Each point on the plot above $A_{MA} = 0.79\ nm^2$ is obtained from the last 500 ns of a 1 μs simulation. From $0.79\ nm^2$ and lower each point is obtained from the last 500 ns of a 5 μs simulation. At around $0.79\ nm^2$ AMA molecules started to move away from the water, forming protrusions into the membrane (Supplementary figure S1). From this point on to smaller A_{MA}, it is seen that the longer the simulation is carried out at fixed surface area, the higher the surface tension obtained. This correlates to the MA molecules from protrusions breaking away from the water surface and moving toward the vacuum interface causing the molecules at the water interface to be more spread out again, obtaining a higher surface tension. Therefore longer equilibration times are allowed for $A_{MA} = 0.79\ nm^2$ and smaller, although no noticeable difference is seen in a monolayer which is compressed without intermittent equilibration at larger A_{MA} (Supplementary figure S4).

MAs leaving the water interface early on result in a difference between A_{MA} plotted on isotherms and the true A_{MA}. This difference becomes larger at lower A_{MA} as more MAs move away from water. For the small monolayer at 300 K it is estimated from density distributions of the head groups (Supplementary figure S2) that the error is $0.0075\ nm^2$ per MA at $A_{MA} = 0.68\ nm^2$, increasing to $0.1\ nm^2$ per MA at $A_{MA} = 0.23\ nm^2$.

It is unexpected for MAs to move away from the water interface at large A_{MA} since the impermeability of the mycobacterial cell wall implies a closely packed layer of MAs, as is supported by monolayer experiments in which stable monolayers are formed. From monolayer experiments of AMA it has been shown that the molecules will undergo conformational change upon compression before monolayer collapse. This raises the question whether the monolayer of AMA modelled here is perhaps less stable than expected due to its homogeneity. It is possible that the heterogeneity within a single MA-type (ie variation in chain lengths and stereochemistry, as in monolayer experiments) or a monolayer with different types of MAs (as in the cell wall) adds stability. Alternatively it may imply that a bilayer is preferred, as in the cell wall where

free lipids effectively form a bilayer with the MA monolayer and may help to stabilise the MA monolayer. This observation could also be an artefact of our model or of limited sampling, which we are unable to determine.

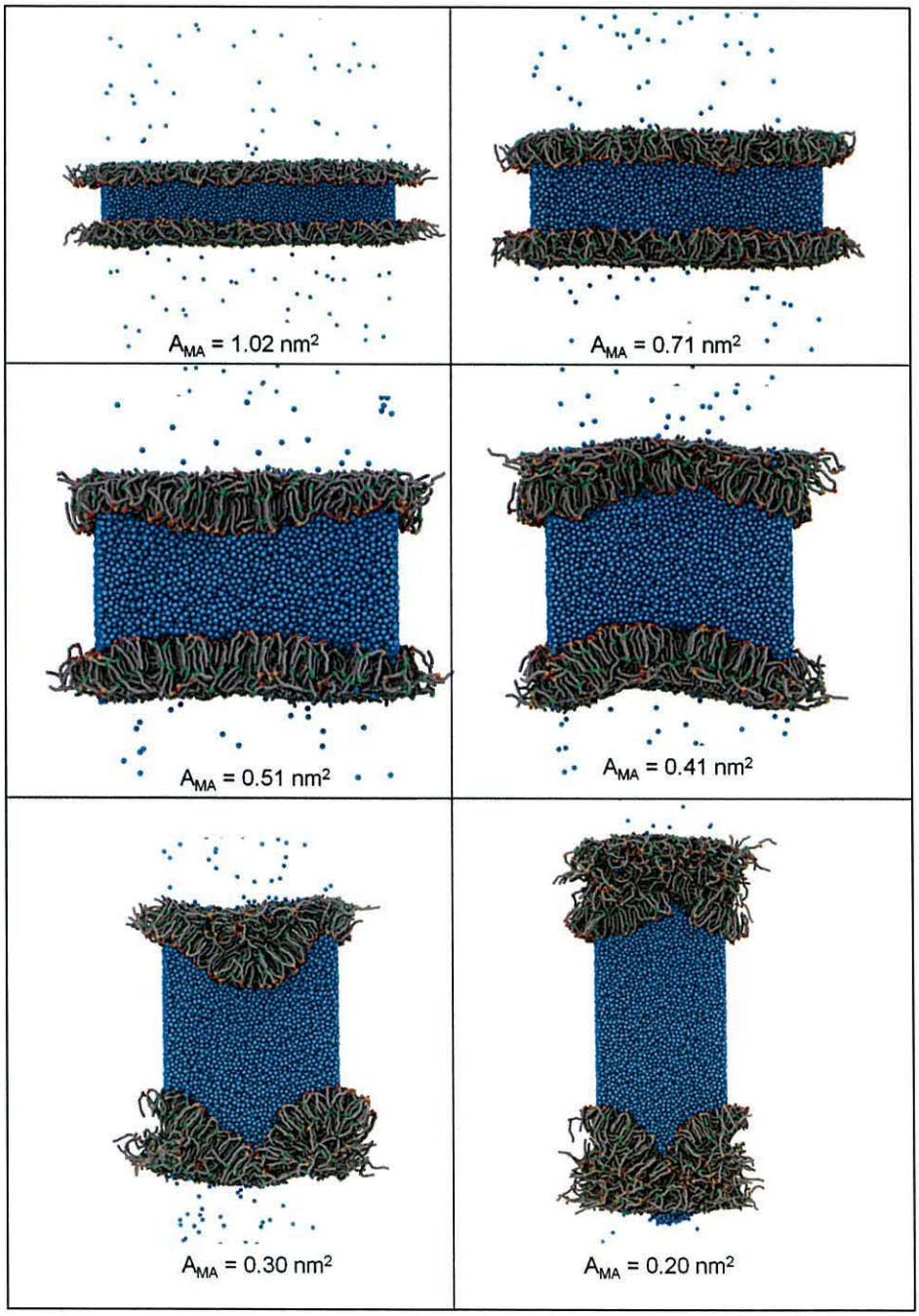


Figure 7: Snapshots of the large monolayer simulated at 300 K, at various A_{MA}, as indicated.

No collapse plateau is seen for the small monolayer system at 300 K, likely because the monolayer is too small to bend significantly, and hence molecules are not displaced out of the monolayer plane. A drop to negative surface tension is seen at $A_{MA} = 0.34\ nm^2$, indicating that the monolayer is unstable. The subsequent increase in surface tension would be due to a new monolayer forming at the water surface after numerous MAs moved away to the vacuum.

Simulation of the small monolayer system at 283 K and 317 K show good correlation with experimental data with lower surface tension as temperature increases at high A_{MA}. At 283 K a drop to negative surface tension is seen at $0.45\ nm^2$, but none at 317 K.

The surface tension-area plot for the large monolayer system is shown in Figure 8, B. Each point on the plot is obtained from the last 500 ns of a 2 μs simulation. MAs are also seen to leave the water interface in the large system, starting with very few molecules at $0.91\ nm^2$ and more at smaller A_{MA}. The large monolayer system showed a collapse point at $0.41\ nm^2$ followed by a plateau. The monolayer was large enough to accumulate collective out-of-plane displacements. The monolayer starts to bend slightly at $0.71\ nm^2$ and has very large undulations at smaller values for A_{MA} such as $0.30\ nm^2$ (Figure 7). Monolayer bending in addition to MAs moving away from the water interface results in larger differences between A_{MA} plotted on isotherms and true A_{MA} in the large monolayer. The approximate error in A_{MA} due to MAs moving away from water, as estimated from density distributions of MA head groups (Supplementary figure S3), is $0.001\ nm^2$ per MA at $A_{MA} = 0.91\ nm^2$, increasing up to $0.14\ nm^2$ per MA at $A_{MA} = 0.14\ nm^2$. After the collapse point errors are estimated to be as large as $0.5 - 0.8\ nm^2$ per MA, which is in line with numerous MAs being expelled from the monolayer at collapse.

Experimental isotherms are reported as surface pressure (Π)-area plots. In order to convert the surface tension (γ_m)-area plots so that they are comparable to experimental data, the surface tension for the lipid chain-water, lipid chain-vacuum and vacuum-water interfaces were obtained for the CG model. The

surface tension obtained for lipid chain-water (43 $mN.m^{-1}$) and lipid chain-vacuum (14 $mN.m^{-1}$) compare reasonably well with experimental values for alkanes (52 and 22 $mN.m^{-1}$, respectively)[36] and values reported for MARTINI.[31] However, the surface tension of 32 $mN.m^{-1}$ obtained for the water-vacuum interface, γ_{aw}, is significantly lower than the experimental value of 72 $mN.m^{-1}$.[36] Due to this difference, the usual conversion from surface tension to surface pressure, $\Pi(A_{MA}) = \gamma_{aw} - \gamma_m(A_{MA})$, cannot be applied as it stands. A simple modification in the form of $\Pi^{exp}(A_{MA}) = \gamma_{aw}^* - \gamma_m(A_{MA})$ is applied to correct for the difference in surface tensions, with γ_{aw}^* a constant.[31]

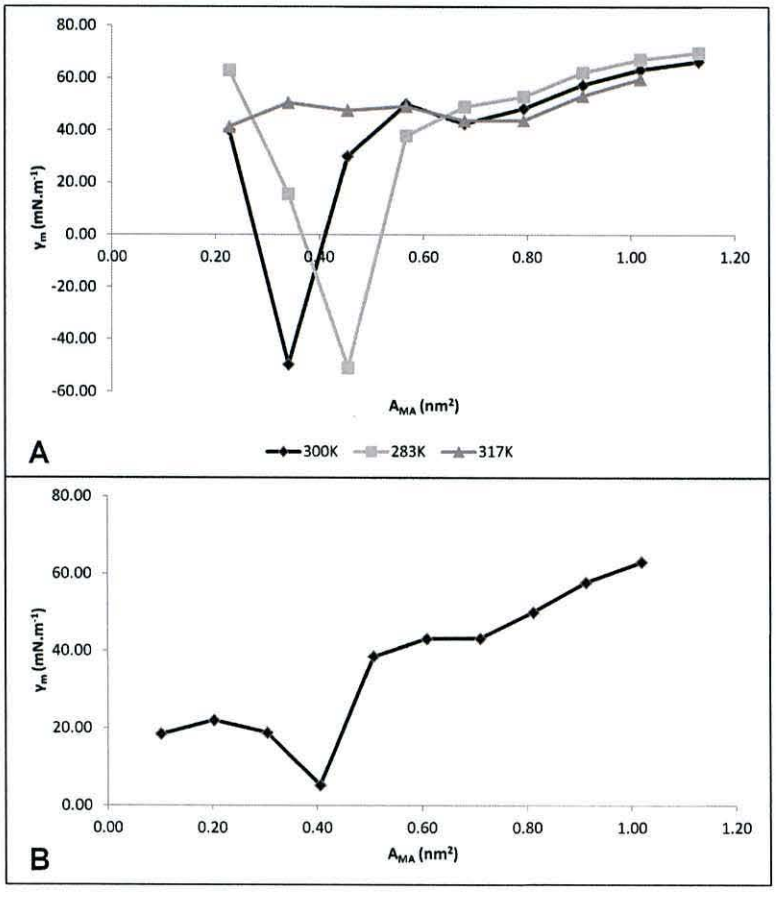


Figure 8: Surface tension - area isotherms for the small monolayer system containing 220 MAs at 300 K, 283 K and 317 K (A) and for the large monolayer system containing 1972 MAs at 300 K.

When fitting the curve from simulation to experimental phase transition surface pressures or collapse surface pressures, it was found that the surface pressure at larger A_{MA} is negative. This is an unrealistic scenario as it implies that the surface tension with the MA monolayer is larger than the water-vacuum surface tension, while it is usually lowered. Therefore $\gamma_{aw}^* = 63.12\ mN.m^{-1}$ was applied to the large monolayer system (Figure 9B) and $\gamma_{aw}^* = 69.71\ mN.m^{-1}$ to the small monolayer system (Figure 9A) in order to produce positive surface pressure values.

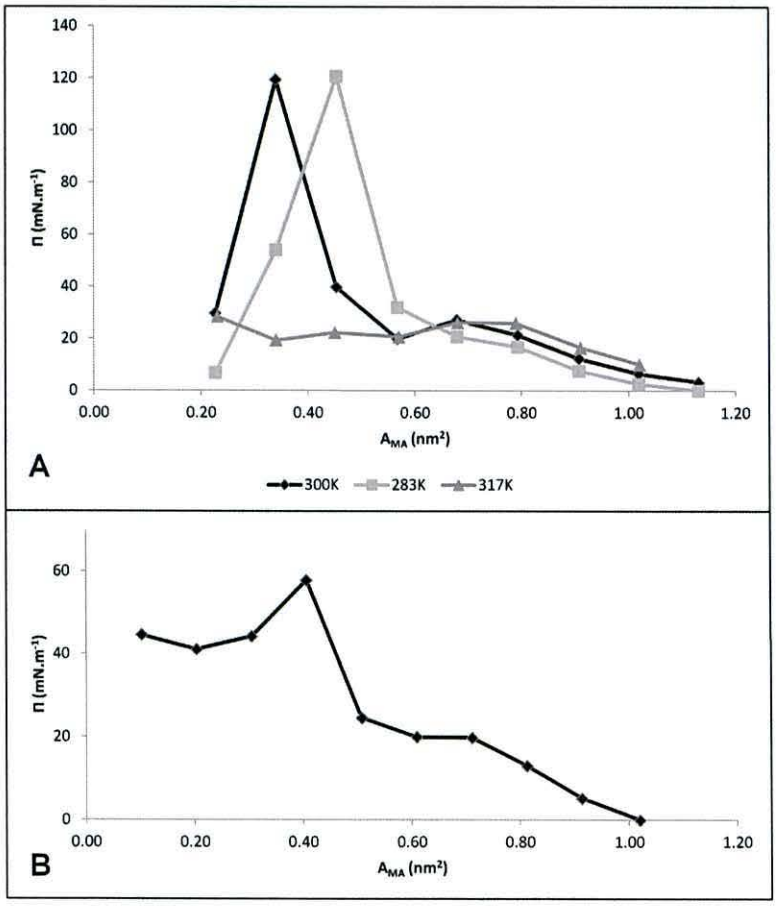


Figure 9: Surface pressure – area isotherms for the small monolayer system containing 220 MAs at 300 K, 283 K and 317 K (A) and for the large monolayer system containing 1972 MAs at 300 K.

With this adjustment, the collapse surface pressure of 57.5 $mN.m^{-1}$ for the large system correlated well with the experimentally reported value of about 60 $mN.m^{-1}$ as reported by one group.[24,26] In addition, the A_{MA} from the simulation, 0.41 nm^2, corresponds well to the experimental range of 0.35 – 0.40 nm^2 for collapse of the monolayer. For the small monolayer system at 283 K the phase transition at 31.85 $mN.m^{-1}$ is slightly higher than the experimental surface pressure of 10-15 $mN.m^{-1}$ while the $A_{MA} = 0.57\ nm^2$ for the phase transition correlates well with the experimental A_{MA} ~0.6 nm^2.[24,26] Due to the flat isotherm curve at 317 K it is hard to determine where the phase change occurs at this temperature, and none is reported experimentally at a temperature of 317 K.

Phases of simulated alpha-MA monolayers

In experimental monolayer studies on MAs, the compressibility modulus, calculated as $K_A = A_{MA} \times \delta\gamma_m / \delta A_{MA}$, indicates the liquid condensed (LC) phase with a value between 80-140 $mN.m^{-1}$, while a value between 20 – 80 $mN.m^{-1}$ indicates the liquid expanded (LE) phase.[24] The compressibility modulus for the systems simulated is shown in Figure 10. Accordingly, the large MA monolayer simulated here is in the LE phase at A_{MA} = 0.91 - 0.51 nm^2. At $A_{MA} = 0.41\ nm^2$ the is in the LC phase, indicating an ordered phase before the monolayer collapses (Figure 7). Similar elasticity values have also been seen for MAs experimentally.[24] The small monolayer at 300 K is in the LE phase at A_{MA} = 1.02 - 0.68 nm^2 and then $K_A > 140$ at $A_{MA} = 0.34\ nm^2$when the monolayer is highly compressed (Figure 6). At 283 K the phases followed the same trends as in the simulation at 300 K, but reaching the most highly compressed state with $K_A >$ 300, representing a very ordered monolayer, but not indicative of the solid condensed state ($K_A > 1000$). At 317 K the monolayer is in the LE phase at A_{MA} = 0.91 – 0.79 nm^2. No values of $K_A > 140\ mN.m^{-1}$ are seen at 317 K, indicating the more fluid nature of the monolayer at higher temperature.

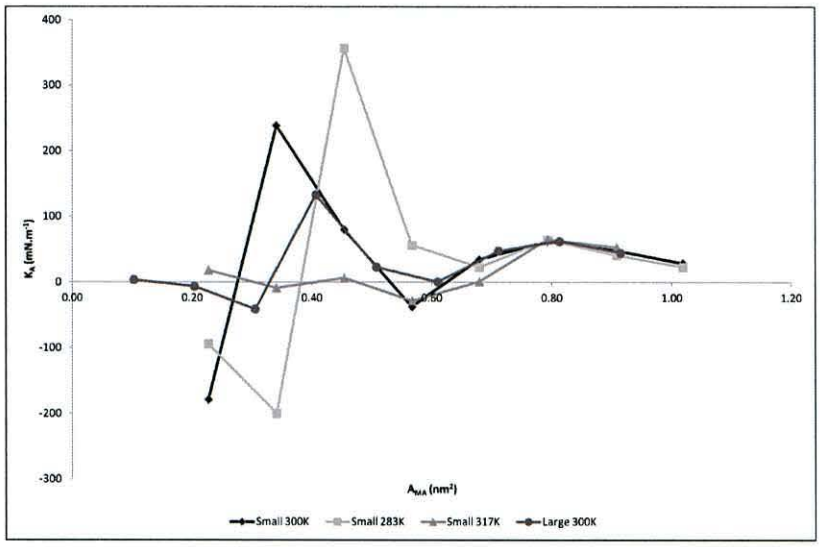


Figure 10: Compressibility modulus, K_A, of the small and large monolayer systems simulated.

Mycolic acid conformation in monolayer compression

It has been proposed in experimental monolayer studies with MAs that at large A_{MA} the molecules are folded at each functional group with four alkyl chains parallel to each other. As the monolayer is compressed, the molecule is proposed to extend its mero-chain, thereby decreasing its surface area to 3 or 2 parallel chains. A model for the unfolding of MA is shown in Figure 2.

It is possible to see the distribution of conformers relative to the structures in Figure 2 by creating a principal component (PC) map of the structures in Figure 2 and then projecting the MAs from simulations onto the same space (Figure 11). On the PC plot, principal component 1 provides resolution on MA conformers in terms of the number of parallel chains. Lower to higher values on the axis move from the W-fold with four parallel chains to the Z-folds with three parallel chains, and finally the U-fold with two parallel chains prior to the fully extended single chain conformation. The sZ conformer has large overlap between the chain termini, and therefore has three parallel chains for most of its length, in contrast to eZ having a single chain section. Therefore sZ is shifted more to the

left relative to eZ. The relative energy of the conformers, shown in Figure 11, indicates that the close proximity of parallel chains leads to more stable conformers with lower energies. The W-fold with four parallel chains is lowest in energy, Z-conformers ranged from 10.52 – 36.30 kJ.mol^{-1}, eU significantly higher at 89.48 kJ.mol^{-1} and the completely extended conformer has the highest relative energy of 94.13 kJ.mol^{-1}. On the second principal component axis the conformers are separated further, with an approximate zero value on this axis representing a conformer with no folding at any functional group, while negative values characterise conformers folded at the proximal cyclopropane (*c*) in contrast to conformers that do not fold at *c*, located at positive values on this axis.

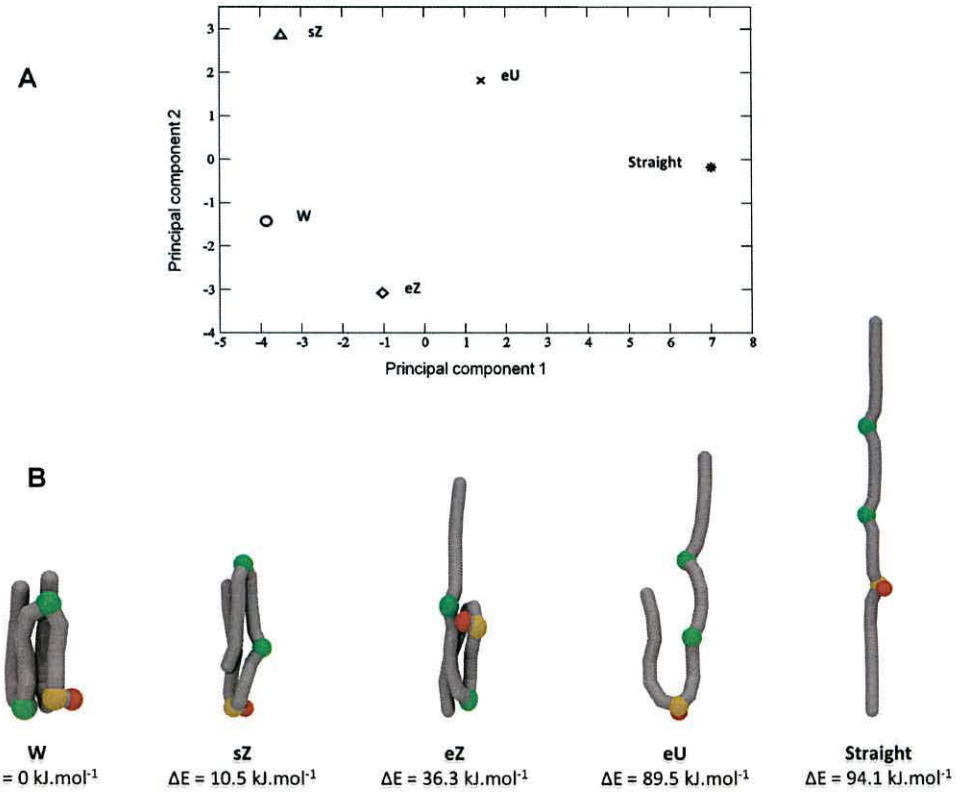


Figure 11: Principal component plot showing the positions of W, sZ, eZ, eU and straight conformers with symbols (A), and the relative energies of these structures (B).

In Figure 12 each MA is represented by a point on the PC plot from several frames at the relevant A_{MA} of the large monolayer system. Each MA was taken from a frame taken every 2 ns from the last 500 ns at each A_{MA}. The PC plot shows that even at large surface areas the molecules did not stay in a four-chain folded conformation, which was the starting conformation. This finding contradicts current popular belief that MAs assume a folded conformation, such as a W-fold at large surface areas. In fact, when taking the average structure of all the molecules in each surface area, it is seen that the molecules are stretched out from the start, only bending at the acid group, which is in contact with the water surface. The finding that MAs fold more rigidly into W-conformations when an α-methyl *trans*-cylcopropane group is present than *cis*-cylopropane,[27] supports this finding that the AMA mero-chain is largely unfolded. At large areas of $A_{MA} = 1.02 - 0.71$ nm^2 the two chains are separated from each other, with most MAs having a conformation intermediate between eU and the extended straight conformer. The average structures here show a distance between *a* and *d* between 2.38 - 3.64 nm. As the surface area is decreased, the chains come closer to each other. From $A_{MA} = 0.71$ - 0.61 nm^2, the distribution of MA conformers on the PC plot starts to narrow down to the eU conformer, and at $A_{MA} = 0.51 - 0.41$ nm^2 there is a narrow distribution around the eU- conformer. This shows that at $A_{MA} = 0.51 - 0.41$ nm^2 the MAs are tightly packed in the monolayer.

The distance between *a* and *d* for the average structure here is down to 1.76 nm, indicating a compact conformation. At smaller $A_{MA} = 0.30 - 0.10$ nm^2 the range of MA conformers on the PC plot expands again toward the extended straight conformer, indicating the collapse of the monolayer and a release of MA molecules from the tightly packed monolayer. Taking the distance between *b* and *e* of the average structures as an indication of the monolayer thickness at various A_{MA}, it can be seen that the MAs stretch out, increasing the *be* distance as A_{MA} decreases. The *be* distance ranges from 4.2 - 4.8 nm, with the maximum at $A_{MA} = 0.30$ nm^2. The same trend was observed in the small monolayer, with the compact double chain fold with a narrow distribution around eU occurring at $A_{MA} = 0.45$ nm^2 at both 283 K and 300 K. At 317 K the conformations remained more disperse throughout, but there is some focus around eU at $A_{MA} = 0.57$ nm^2.

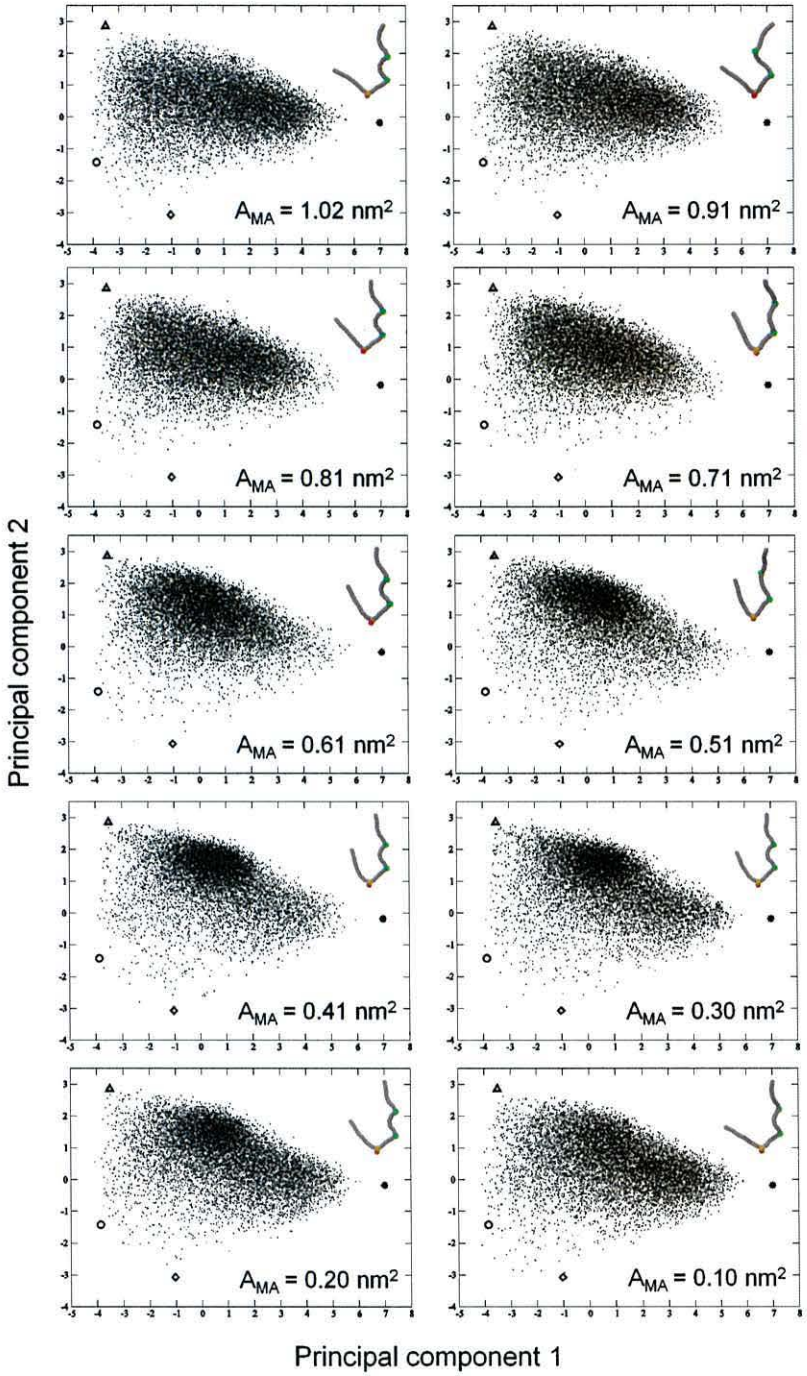


Figure 12: Projection of MAs in the large monolayer at various A_{MA} onto the principal component space with reference structures indicated with symbols as in Figure 11. MAs were taken from a frame every 2 ns from the last 500 ns of each simulation. The average structure is inserted in the top right-hand corner of each plot.

The uniaxial order parameter, S, was calculated for all the bonds between the coarse grained beads with the formula: $S = 0.5*(3*\langle\cos^2\theta\rangle - 1)$, where θ is the angle between each bond and the monolayer normal and averaging is done for all the MAs over the last 500 ns of the simulation. When $S = 1$, the bond is aligned parallel to the bilayer normal, while $S = -0.5$ signifies perpendicular alignment and $S = 0$ random orientation.

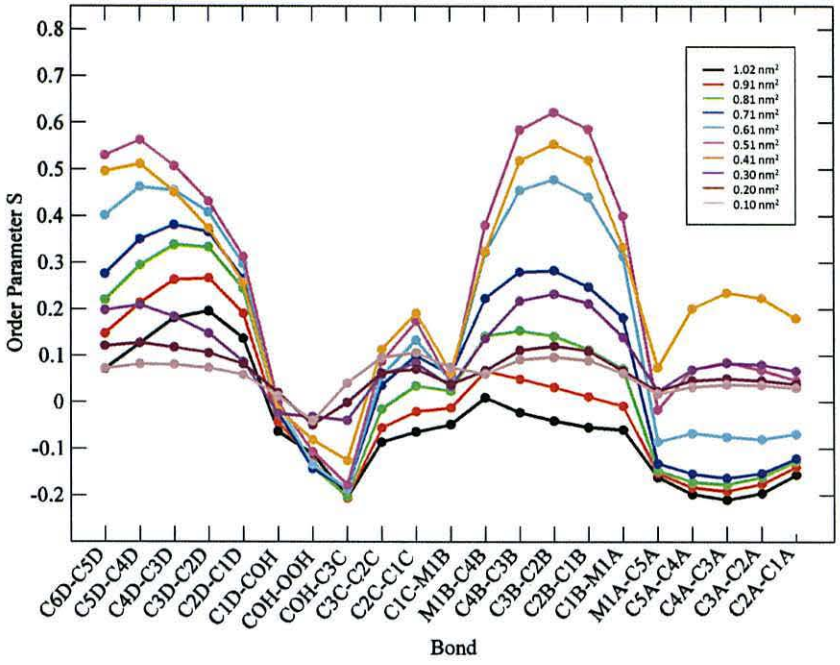


Figure 13: Order parameters for MA bonds with respect to the monolayer normal at various A_{MA} for the large monolayer system. The order parameter at each A_{MA} was calculated from the last 500 ns of each simulation.

As shown in Figure 13 for the large monolayer system, the order parameter for bonds in the alpha-chain region, C6D until C1D and the mero-chain region between the two cyclopropane beads M1B to M1A increases steadily from $A_{MA} = 1.02$ nm^2 where $S \approx 0$ to $A_{MA} = 0.51$ nm^2 where $S \approx 0.6$. This indicates that the orientation of the bonds changes toward a parallel orientation with the monolayer normal as A_{MA} decreases. From $A_{MA} = 0.41$ nm^2 and smaller, the order parameter for these bonds decreases again until reaching a low of about 0.1, indicating random orientations at the end of the monolayer compressions. The

point where S starts to decrease correlates with the collapse point. The order parameters for bonds in the short chain between the acid head group and the proximal cyclopropane (C3C-C1C) remain quite low with a maximum around $S \approx 0.2$.

The bonds at the end of the mero-chain, from C5A to C1A increase very little in order between $A_{MA} = 1.02 - 0.71$ nm^2, with $S \approx -0.2$, indicating that the bonds are tending toward a perpendicular orientation, which means the end of the mero-chain is not stretched out vertically at these surface areas. Only from $A_{MA} = 0.61$ nm^2 is a more significant increase in the order parameter seen. Again a maximum is reached at $A_{MA} = 0.51$ nm^2 before a rapid decrease. However, for this part of the chain the highest value reached is $S \approx 0.24$, which is much lower than the order parameter for the alpha-chain and the more proximal part of the mero-chain. Therefore the parts of the alpha- and mero-chains that are able to match each other in length seem to aid in increasing the order of those regions by supporting each other, while the longer terminus of the mero-chain is not matched and therefore does not pack very rigidly, even when the monolayer is highly compressed. This observation confirms previous interpretations from experimental monolayer studies that the length of the alpha-chain is important in how MAs fold,[22] as well as the relative length of the chain segments in relation to each other.[25,26]

The order parameters for bonds at the functional groups in the molecule are lower than those for the C1-bead chains, with the exception of the short chain between the acid head group and the proximal cyclopropane (C3C-C1C). S is very similar for M1B-C4B and C1B-M1A bonds, but C1C-M1B is generally higher than M1A-C5A. Around the acid head group S stays between -0.2 and 0, indicating perpendicular or random orientations of the bonds in this segment.

The trends above are also seen in the small monolayer system. At 300 K, S for the alpha and mero-chain segment between M1A and M1B increases up to about 0.6 before decreasing again from $A_{MA} = 0.34$ nm^2, similar to the large system. At 283 K, $S \approx 0.8$ at $A_{MA} = 0.57$ nm^2, indicating bonds that are almost parallel in orientation and hence a very ordered monolayer. In contrast, the maximum

reached at 317 K is $S \approx 0.3$, which shows that at this temperature the MAs are not packed in an orderly fashion at any surface area investigated here.

Conclusion

We present here a coarse grained model for an AMA from *M. tb.* Parameterisation of bonded interactions was performed on a single AMA in different environments, using atomistic MD as reference. The coarse grained model replicated the atomistic model well. The peripheral OOH-COH-C1D angle distribution was less precise, in favour of the angles that are more central in determining MA conformation. The two cyclopropane beads were treated as identical, although their atomistic angle distributions varied slightly. This approximation resulted in the smaller angles for the distal cyclopropane bead, M1A being neglected. Hence, more folding may be expected at the distal cyclopropane group than is observed in this model.

Isotherms for the simulated homogeneous AMA monolayers were obtained for a small monolayer system containing 220 MAs at 283 K, 300 K and 317 K as well as for a large monolayer containing 1972 MAs at 300 K. Despite the obvious difference with experimental monolayers that are heterogeneous in composition, even for single MA classes, the isotherms obtained from simulations were able to capture the key features of experimental monolayers such as phase transitions, collapse and temperature-dependent trends with reasonable accuracy. According to the calculated compressibility modulus, the monolayers were generally in the LE phase at higher A_{MA} and in the LC phase at smaller A_{MA}. A high compressibility modulus was obtained just before the monolayer collapse point, indicating a highly compressed monolayer, although a solid-condensed phase was not observed, in accordance with experimental data for AMA.

It was unexpected that MAs moved away from the water surface at larger A_{MA}. This observation may be due to an artefact in the model, or due to stabilising effects from the heterogeneity of experimental monolayers and the cell membrane. MAs moving away from the water surface resulted in errors in the

A_{MA} reported. These errors were small at large A_{MA}, but increased as the monolayer was compressed and more MAs moved away from the water interface. MAs were never seen to enter the water phase, unlike other lipids.[31,37] Investigation into MA conformations in the monolayer revealed that the postulated W-shaped folding at large A_{MA} did not occur. Instead MAs were mostly folded only at the head group, with the two parallel chains becoming more tightly packed as the surface area decreased. It has been shown that the *cis*-cyclopropane group does not promote folding to the extent that the α-methyl *trans*-cyclopropane group does.[27] However, our model may be showing less folding at the cyclopropane beads than is occurring in reality, due to smaller angles being excluded in the CG model's angle distributions. Our model confirms the extension of the mero-chain as the surface area decreases and the molecules become more tightly packed, in accordance with experimental hypotheses.[20-26] The order parameter showed that tight packing with bonds orientated more perpendicular to the membrane normal was observed, as the monolayer was compressed, mostly for the alpha chain and the portion of the mero-chain adjacent and parallel to the alpha-chain. The mero-chain terminus, however, did not pack tightly, even at small A_{MA}. This observation supports the theory that the alpha-chain length plays an important role in determining monolayer properties, as suggested by Hasegawa and co-workers.[22]

To our knowledge this is the first coarse grained model of a mycolic acid. The model cannot incorporate the explicit stereochemistry of a MA, which may be crucial in determining its structure and function.[5,6] Despite this, the model could replicate the key aspects of existing experimental data on MA monolayers. Therefore this approach can serve as a basis for further work on other MAs and their conformations, as well as models of larger components of the mycobacterial cell wall.

Funding Sources

This work was funded by Bangor University through the award of a 125th Anniversary scholarship to WG and was carried out under the HPC-EUROPA2

project (project number: 70100) with the support of the European Commission Capacities Area - Research Infrastructures Initiative.

Acknowledgments

This work was carried out in part with the use of HPC Wales facilities. Motoko Watanabe, Marion Turner and professor David E. Minnikin are acknowledged for their contribution to the development of the WUZ folding concept. Professor David E. Minnikin, professor Mark S. Baird and professor Jan A. Verschoor are acknowledged for discussions on the modeling of the WUZ folding of MAs.

References

(1) WHO *Global tuberculosis report*, 2013.

(2) Getahun, H.; Harrington, M.; O'Brien, R.; Nunn, P. *Lancet* **2007**, *369*, 2042.

(3) *Toman's Tuberculosis case detection, treatment, and monitoring - questions and answers*; 2nd ed.; World Health Organization: Geneva, Switzerland, 2004.

(4) Watanabe, M.; Aoyagi, Y.; Ridell, M.; Minnikin, D. E. *Microbiology* **2001**, *147*, 1825.

(5) Vander Beken, S.; Al Dulayymi, J. R.; Naessens, T.; Koza, G.; Maza-Iglesias, M.; Rowles, R.; Theunissen, C.; De Medts, J.; Lanckacker, E.; Baird, M. S.; Grooten, J. *Eur. J. Immunol.* **2011**, *41*, 450.

(6) Beukes, M.; Lemmer, Y.; Deysel, M.; Al Dulayymi, J. R.; Baird, M. S.; Koza, G.; Iglesias, M. M.; Rowles, R. R.; Theunissen, C.; Grooten, J.; Toschi, G.; Roberts, V. V.; Pilcher, L.; Van Wyngaardt, S.; Mathebula, N.; Balogun, M.; Stoltz, A. C.; Verschoor, J. A. *Chem. Phys. Lipids* **2010**, *163*, 800.

(7) Peyron, P.; Vaubourgeix, J.; Poquet, Y.; Levillain, F.; Botanch, C.; Bardou, F.; Daffè, M.; Emile, J.-F.; Marcou, B.; Cardona, P.-J.; de Chastellier, C.; Altare, F. *PLoS Pathog.* **2008**, *4*, e1000204.

(8) Korf, J.; Stoltz, A.; Verschoor, J.; De Baetselier, P.; Grooten, J. *Eur. J. Immunol.* **2005**, *35*, 890.

(9) Korf, J.; Stoltz, A.; Verschoor, J.; De Baetselier, P.; Grooten, J. *Eur. J. Immunol.* **2005**, *35*, 890.

(10) Dao, D. N.; Sweeney, K.; Hsu, T.; Gurcha, S. S.; Nascimento, I. P.; Roshevsky, D.; Besra, G. S.; Chan, J.; Porcelli, S. A.; Jacobs, W. R. *PLoS Pathog.* **2008**, *4*.

(11) Glickman, M. S.; Cox, J. S.; Jacobs, W. R. *Mol. Cell* **2000**, *5*, 717.

(12) Glickman, M. S.; Cahill, S. M.; Jacobs, W. R. *J. Biol. Chem.* **2001**, *276*, 2228.

(13) Rao, V.; Fujiwara, N.; Porcelli, S. A.; Glickman, M. S. *J. Exp. Med.* **2005**, *201*, 535.

(14) Rao, V.; Gao, F.; Chen, B.; Jacobs, W. R.; Glickman, M. S. *J. Clin. Invest.* **2006**, *116*, 1660.

(15) Barkan, D.; Hedhli, D.; Yan, H.-G.; Huygen, K.; Glickman, M. S. *Infect. Immun.* **2012**, *80*, 1958.

(16) Dubnau, E.; Chan, J.; Raynaud, C.; Mohan, V. P.; Lanèelle, M.-A.; Yu, K.; Quèmard, A.; Smith, I.; Daffè, M. *Mol. Microbiol.* **2000**, *36*, 630.

(17) Yuan, Y.; Zhu, Y.; Crane, D. D.; Barry, C. E. *Mol. Microbiol.* **1998**, *29*, 1449.

(18) Zuber, B.; Chami, M.; Houssin, C.; Dubochet, J.; Griffiths, G.; Daffè, M. *J. Bacteriol.* **2008**, *190*, 5672.

(19) Groenewald, W.; Baird, M. S.; Verschoor, J. A.; Minnikin, D. E.; Croft, A. K. *Chem. Phys. Lipids* **2014**, *180*, *15*.

(20) Hasegawa, T. *Biopolymers* **2004**, *73*, 457.

(21) Hasegawa, T.; Amino, S.; Kitamura, S.; Matsumoto, L.; Katada, S.; Nishijow, J. *Langmuir* **2003**, *19*, 105.

(22) Hasegawa, T.; Leblanc, R. M. *Biochim. Biophys. Acta* **2003**, *1617*, 89.

(23) Hasegawa, T.; Nishijo, J.; Watanabe, M.; Funayama, K.; Imae, T. *Langmuir* **2000**, *16*, 7325.

(24) Villeneuve, M.; Kawai, M.; Kanashima, H.; Watanabe, M.; Minnikin, D. E.; Nakahara, H. *Biochim. Biophys. Acta, Biomembr.* **2005**, *1715*, 71.

(25) Villeneuve, M.; Kawai, M.; Watanabe, M.; Aoyagi, Y.; Hitotsuyanagi, Y.; Takeya, K.; Gouda, H.; Hirono, S.; Minnikin, D. E.; Nakahara, H. *Biochim. Biophys. Acta, Biomembr.* **2007**, *1768*, 1717.

(26) Villeneuve, M.; Kawai, M.; Watanabe, M.; Aoyagi, Y.; Hitotsuyanagi, Y.; Takeya, K.; Gouda, H.; Hirono, S.; Minnikin, D. E.; Nakahara, H. *Chem. Phys. Lipids* **2010**, *163*, 569.

(27) Villeneuve, M.; Kawai, M.; Horiuchi, K.; Watanabe, M.; Aoyagi, Y.; Hitotsuyanagi, Y.; Takeya, K.; Gouda, H.; Hirono, S.; Minnikin, D. E. *Microbiology* **2013**, *159*, 2405.

(28) Hong, X.; Hopfinger, A. J. *Biomacromolecules* **2004**, *5*, 1052.

(29) Marrink, S.; Risselada, H.; Yefimov, S.; Tieleman, D.; de Vries, A. *J. Phys. Chem. B* **2007**, *111*, 7812.

(30) Marrink, S. J.; Tieleman, D. P. *Chem. Soc. Rev.* **2013**, *42*, 6801.

(31) Baoukina, S.; Monticelli, L.; Marrink, S. J.; Tieleman, D. P. *Langmuir* **2007**, *23*, 12617.

(32) Pronk, S.; Pall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. *Bioinformatics* **2013**, *29*, 845.

(33) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics Modell.* **1996**, *14*, 33.

(34) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.

(35) Wang, L. P.; Van Voorhis, T. *J. Chem. Phys.* **2010**, *133*.

(36) Aveyard, R.; Haydon, D. A. *An introduction to the principles of surface chemistry*; London Cambridge University Press New York, 1973.

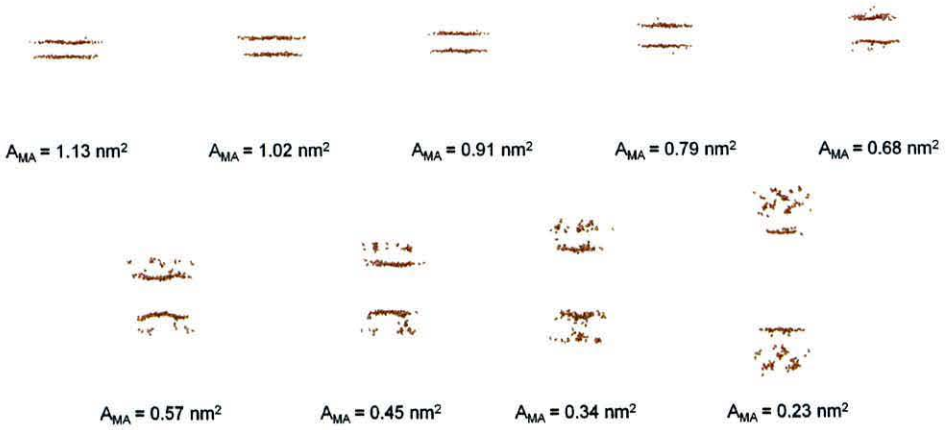
(37) Baoukina, S.; Monticelli, L.; Risselada, H. J.; Marrink, S. J.; Tieleman, D. P. *Proc. Natl. Acad. Sci. U. S. A* **2008**, *105*, 10803.

Supporting Information

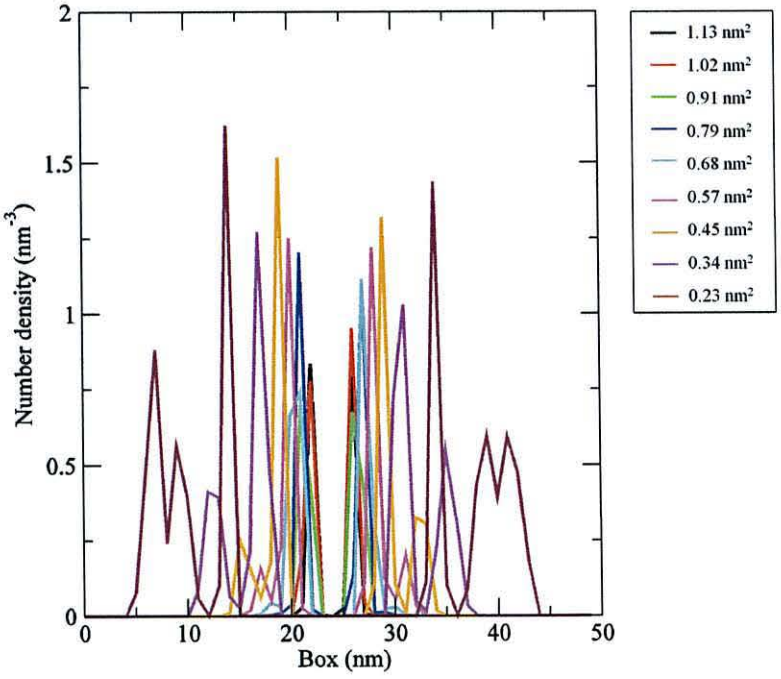
Monolayer Simulation of Coarse Grained Alpha-Mycolic Acid

Wilma Groenewald, Anna K. Croft and Siewert-Jan Marrink

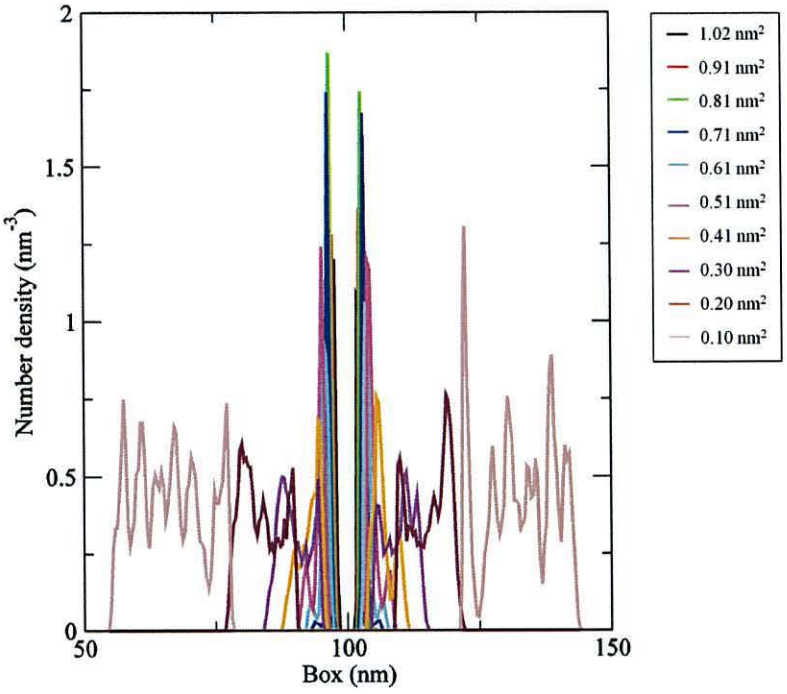
Contents



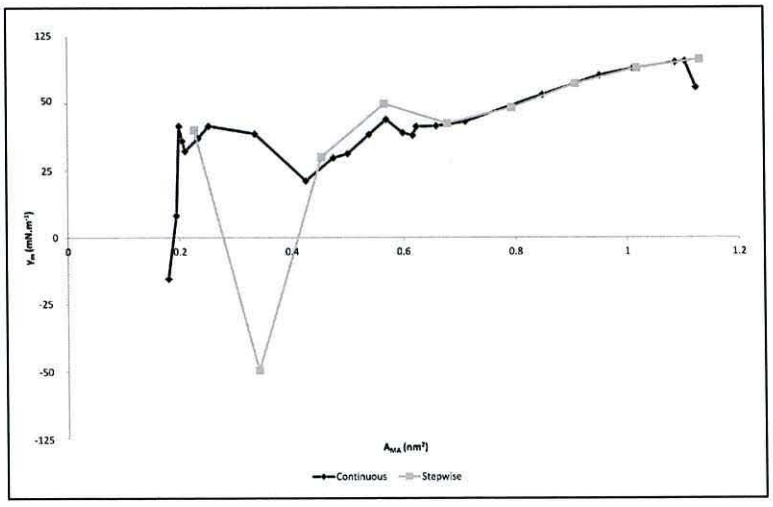
Supplementary figure S1: Head group beads (COH and OOH) at various A_{MA} as indicated for the small monolayer at 300 K, showing head groups moving away from the water interface at A_{MA} = 0.79 nm².



Supplementary figure S2: Number densities for OOH bead for the small monolayer at 300 K at various A_{MA}, as indicated.



Supplementary figure S3: Number densities for OOH bead for the large monolayer at 300 K at various A_{MA}, as indicated.



Supplementary figure S4: Surface tension plots for the small monolayer at 300 K with continuous and stepwise compressions. Each point was obtained from the last 500 ns of the simulation. For the continuous compression each point was a 1 μs simulation and for the stepwise compression the points at 0.91 nm^2 and higher were 1 μs while those at smaller A_{MA} were 5 μs.

Supplementary table S1: Surface tension used for the small monolayer system compressions.

A_{MA}	Surface tension used in compression (bar.nm)		
	283K	300K	317K
1.13 - 1.02	-100	-100	NA
1.02 - 0.91	-100	-100	-100
0.91 - 0.79	-100	-100	-100
0.79 - 0.68	-100	-100	-100
0.68 - 0.57	-100	-100	-100
0.57 - 0.45	-2500	-100	-100
0.45 - 0.34	-3000	-2500	-100
0.34 - 0.23	-2500	-3000	-100

Supplementary table S2: Surface tension used for the large monolayer system compressions.

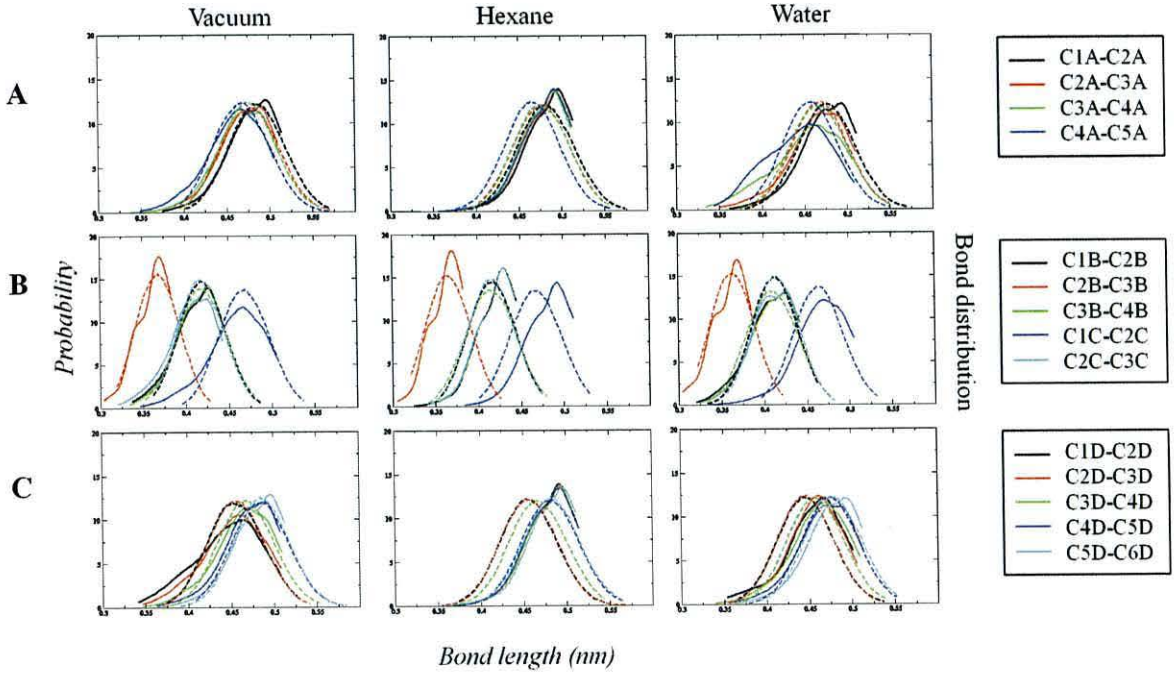
A_{MA}	Surface tension used in compression (bar.nm)
1.02 - 0.91	-100
0.91 - 0.81	-100
0.81 - 0.71	-100
0.71 - 0.61	-100
0.61 - 0.51	-100
0.51 - 0.41	-2500
0.41 - 0.30	-2500
0.30 - 0.20	-2500
0.20 - 0.10	-2500

Supplementary table S3: Additional cyclopropane bonded parameters. Angle (degrees) and force constant ($kJ.mol^{-1}.rad^{-2}$) for the cyclopropane angle parameter.

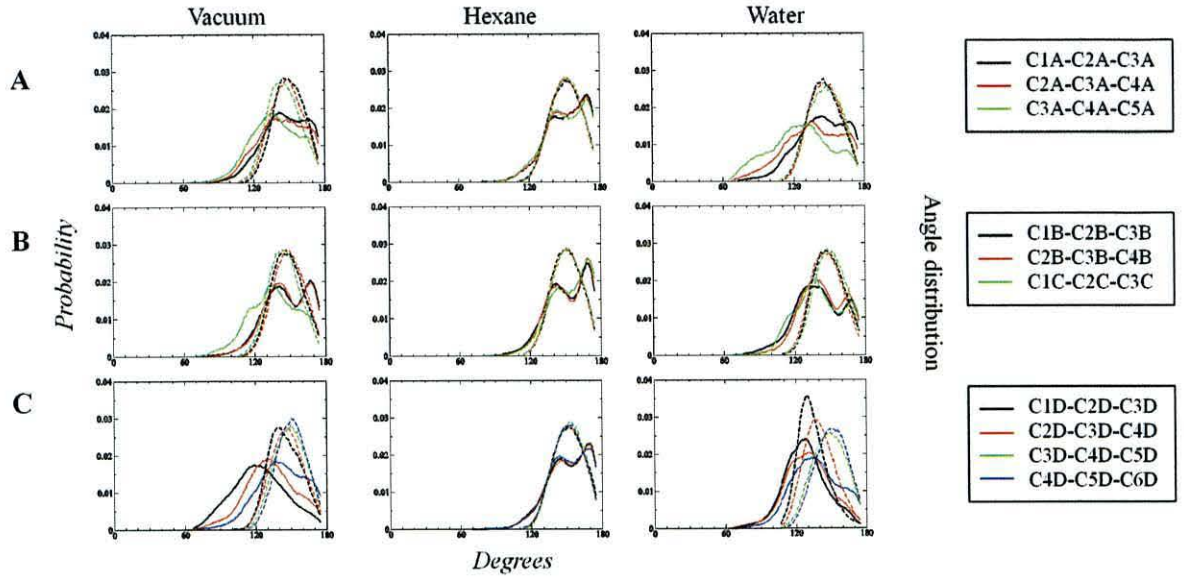
Angle	th0	cth
CT-CT-CY	115.090	612.010

Supplementary table S4: Additional cyclopropane bonded parameters. Torsional Fourier coefficients ($kJ.mol^{-1}$) obtained for cyclopropane.

Dihedral	V_1	V_2	V_3	V_4
CT-CT-CT-CY	1.00000	-22.00200	24.05800	-20.89300
CT-CT-CY-CY	1.00000	14.17700	-20.14000	-9.78690
CY-CT-CT-HC	1.00000	-32.91200	25.29700	2.47730
CT-CT-CY-HC	1.00000	1.43810	21.88000	12.80200
HC-CT-CY-HC	1.00000	9.99550	11.63500	-14.48900



Supplementary figure S5: Bond distributions that were optimised in vacuum, hexane/butane and water for AMA for chain A (A), chains B and C (B) and chain D (C). Solid lines represent the AA distribution as mapped onto CG beads and dashed lines represent the CG model.



Supplementary figure S6: Angle distributions that were optimised in vacuum, hexane/butane and water for AMA for chain A (A), chains B and C (B) and chain D (C). Solid lines represent the AA distribution as mapped onto CG beads and dashed lines represent the CG model.