Biochemical and Biophysical Research Communications xxx (2017) 1–8

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Structural and mechanistic comparison of the Cyclopropane Mycolic Acid Synthases (CMAS) protein family of *Mycobacterium tuberculosis*

Lucas A. Defelipe [a, b], Federico Osman [a], Marcelo A. Marti [a, b, **], Adrián G. Turjanski [a, b, *]

[a] Departamento de Química Biológica, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Intendente Güiraldes 2620, Ciudad Autónoma de Buenos Aires, Argentina
[b] IQUIBICEN-UBA/CONICET, Intendente Güiraldes 2620, Ciudad Autónoma de Buenos Aires, Argentina

ARTICLE INFO

Article history:
Received 15 June 2017
Received in revised form
31 July 2017
Accepted 27 August 2017
Available online xxx

Keywords:
Tuberculosis
Methyltransferase
Bioinformatics
QM/MM
Mycolic acids
CMAS

ABSTRACT

Tuberculosis (TB) is a chronic disease caused by the bacillus *Mycobacterium tuberculosis(Mtb)* and remains a leading cause of mortality worldwide. The bacteria has an external wall which protects it from being killed, and the enzymes involved in the biosynthesis of the cell wall components have been proposed as promising targets for future drug development efforts. Cyclopropane Mycolic Acid Synthases (CMAS) constitute a group of ten homologous enzymes which belong to the mycolic acid biosynthesis pathway. These enzymes have S-adenosyl-L-methionine (SAM) dependent methyltransferase activity with a peculiarity, each one of them has strong substrate selectivity and reaction specificity, being able to produce among other things cyclopropanes or methyl-alcohol groups from the lipid olefin group. How each CMAS processes its substrate and how the specificity and selectivity are encoded in the protein sequence and structure, is still unclear.

In this work, by using a combination of modeling tools, including comparative modeling, docking, all-atom MD and QM/MM methodologies we studied in detail the reaction mechanism of *cmaA2*, *mmaA4*, and *mmaA1* CMAS and described the molecular determinants that lead to different products. We have modeled the protein-substrate complex structure and determined the free energy pathway for the reaction. The combination of modeling tools at different levels of complexity allows having a complete picture of the CMAS structure-activity relationship.

1. Introduction

According to the last WHO report, near 10.4 million people around the world developed tuberculosis (TB), leading to 1.4 million deaths in 2015 [1]. Traditional therapies for TB involve long treatments with first generation drugs such as isoniazid, rifampicine, pyrazinamide and ethambutol [2]. Nevertheless, the emergence of multidrug resistant (MDR) and extremely drug resistant (XDR) *Mycobacterium tuberculosis (Mtb)* and the negative interaction between various drugs to treat TB and the ones for AIDS, demonstrate the urgent need to develop new therapies. The host's immune response to TB infection relies in phagocytosis of the bacilli by the macrophages leading to granuloma formation that stops bacterial replication. Inside the macrophages bacteria face stressful conditions characterized by hypoxia, inducible nitric oxide synthase derived NO and nutrient deprivation [3]. The bacilli in response switches to a non replicative (latent) state, where it can remain hidden and alive for decades [4]. *Mtb* has an external wall which protects the bacilli from the hostile conditions it faces inside the macrophage. This external wall is composed of glycolipids, an ester of trehalose and mycolic acid (MA), a very long chain fatty acid, with 40–60 carbon atoms [5]. In this sense, proteins involved in the biosynthesis of the external wall are attractive targets for drug design [6,7].

Cyclopropane Mycolic acid synthases (CMAS) are a family enzymes with a methyl transferase activity which act by transferring a

* Corresponding author. Departamento de Química Biológica, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Intendente Güiraldes 2620, Ciudad Autónoma de Buenos Aires, Argentina.
** Corresponding author. Departamento de Química Biológica, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Intendente Güiraldes 2620, Ciudad Autónoma de Buenos Aires, Argentina.
E-mail addresses: marti.marcelo@gmail.com (M.A. Marti), adrian@qi.fcen.uba.ar (A.G. Turjanski).

http://dx.doi.org/10.1016/j.bbrc.2017.08.119
0006-291X/

methyl group from S-adenosyl-L-methionine (SAM or AdoMet) to the double bond of immature mycolic acids. These proteins have the typical Rossmann fold subdomain of nucleic acid binding proteins and a MA recognition domain.

Due to the difficulty in working with these proteins *in vitro*, in particular from a biochemical approach, most of the experiments that evaluate their activity have been performed with knockout strains analyzing the resulting MA lipid profile with thin layer chromatography (TLC) or nuclear magnetic resonance. The proposed reaction mechanism can be described in two distinct steps: the first one, supposedly the same in all CMAS, consist of the transfer of the methyl group from SAM to the double bond of the olefin, yielding S-adenosyl-L-homocysteine and a secondary carbocation. The second step, specific to each member of the family, consists in the resolution of the carbocation to either: i) a cyclopropane, in *cis* as proposed for *cmaA1*, *pcaA*, *mmaA2* and *cmaA2*, the latter also proposed to produce the *trans* isomer [8–12], ii) a *trans* double bond with vicinal methyl group (*mmaA1*) [13] or iii) a (S,S)-methyl-alcohol (*mmaA4*) [14]. An exception to this behaviour is *mmaA3* [15], which takes the product of *mmaA4* and transfers a methyl group to the previously formed alcohol forming an methyl-ether. A summary of the different reactions carried out by CMAS can be seen in Fig. 1.

Even though there is some experimental evidence related to the activity of CMAS we still do not know the molecular determinants for each CMAS specificity. Moreover, due to the difficulty in performing biochemical experiments little is known about the reaction mechanisms and enzyme kinetics. The activity of enzymes in a living cell is based on a complex network of interactions among biomolecules, exchanging information and energy through biochemical processes and conformational changes. These events could go from complex macromolecular arrangements in the micro and nano-scale to atomistic detail changes in the active site. Consequently, different modeling techniques, each proper for a particular scale, are commonly used. Moreover, a single process often spans more than a single time or spatial scale. In this sense, the necessity arises for combining modeling techniques in multi-scale approaches.

In this work we link each CMAS sequence to its activity and to the production of each specific MA modification. We used comparative modeling and docking to model the proteins structures and ligand binding, molecular dynamics to understand the conformational regulation of the activity and QM/MM to study the underlying reaction mechanism. We propose that each CMAS controls its activity at different levels, by either regulating the solvent accessibility of the active site, the binding small molecules in the active site, or simply changing certain amino acids involved in the reaction mechanisms. The combination of modeling tools at different levels of complexity yields a complete picture of the CMAS structure-activity relationship.

2. Materials and methods

Comparative Modeling. Models of *umaA*, *ufaA*, *Rv3720*, *mmaA1* and *mmaA3* were done with MODELLER [18] manually curating the alignment. A more detailed description is presented in the Supplementary Methods section.

Docking. Docking of the olefin substrate was performed with rDock [19,20] for 100 runs using the standard docking protocol. Visual inspection of the solutions was performed to discard poor conformations and the lowest energy solution was kept in each system.

Classical molecular dynamics. MD simulations were done with AMBER14 [21]. Complex structures were solvated in TIP3P water 10 Å radius truncated octahedron box. Amber14SB [21] was used for the protein while GAFF [22] was used for the substrates (Bicarbonate, Olefin and SAM/SAH). Simulations were performed treating electrostatic interactions with Particle Mesh Ewald [23] with a time step of 2fs. Production simulations were done for 150ns for cmaA2 and mmaA1 and 500ns for mmaA4. Further information on classical simulation parameters is provided in the Supplementary Material.

QM/MM Steered molecular dynamics. The chosen level of theory was DFTB [24,25] as it offered a balance between quality and needed computational time. The simulation strategy is similar to classic simulations. A minimization was performed for 2000 steps with CG in periodic boundary conditions. Heating was performed with the Berendsen thermostat from 10K to 300K in 50 ps with a coupling constant of 1 ps. Finally, the density equilibration was performed with the Langevin thermostat [26] and the Berendsen barostat for another 50 ps. From there a 500 ps NPT simulation with a fixed reaction coordinate was performed to take snapshots every 12.5 ps for steered molecular dynamics (SMD) simulation. Free energy profiles are obtained from the Jarzynski relationship [27,28].

For all cases, the first reaction step of was done with a 1 fs timestep using a Hybrid Differential Relaxation Algorithm (HyDRA)

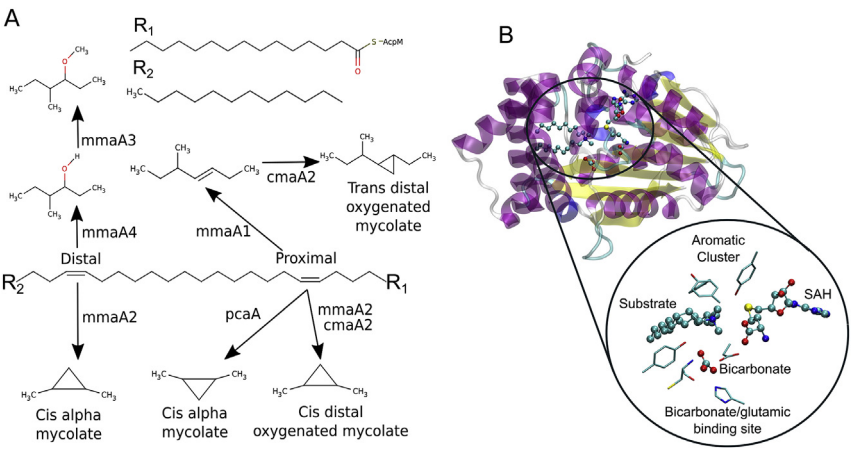

Fig. 1. General information of CMAS protein family. A) Proposed activities of different CMAS. B) General fold and depict of the active site. Depending on the CMAS the active site can have a bicarbonate ion and a glycine or a glutamic acid residue. We only depicted the Gly/bicarbonate active site.

Please cite this article in press as: L.A. Defelipe, et al., Structural and mechanistic comparison of the Cyclopropane Mycolic Acid Synthases (CMAS) protein family of *Mycobacterium tuberculosis*, Biochemical and Biophysical Research Communications (2017), http://dx.doi.org/10.1016/j.bbrc.2017.08.119

previously developed in our group [29]. The reaction was computed along 50,000 simulation steps (12,500 QM/MM steps). The reaction coordinate used is the following:

$$RC = d(C_{SAM\_CH3} - S_{SAM}) - d(C_{SAM\_CH3} - C_{DoubleBond})$$

where C_{SAM_CH3} is the methyl group to be transferred, S_{SAM} is the sulfur atom of SAM y $C_{DoubleBond}$ is one of the atoms from the double bond of the olefin. The coordinate moving velocity was 0.04 Å per ps.

For the second step in *cmaA2*, the chosen reaction coordinate for *cmaA2* was:

$$RC = d(H_{SAM\_CH3} - C_{SAM\_CH3}) - d(H_{SAM\_CH3} - O_{HCO3})$$

where H_{SAM_CH3} is the proton to be transferred and O_{HCO3} is the oxygen acceptor. The guide velocity was 0.08 Å per ps.

The second step in *mmaA4 and mmaA1*, The simulation is done for 50000 steps with a 1fs timestep. In *mmaA4* no reaction coordinate was required. In *mmaA1* we used the following shown below:

$$RC = d(H_{Carbocation} - C_{Carbocation}) - d(H_{Carbocation} - O_{D139})$$

3. Results

Comparative modeling of CMAS family membersProteins of the CMAS family have the typical Rossmann fold of methyltransferases. The particular CMAS fold consists of 7 beta sheets (6 parallel and one antiparallel) and 13 alpha helix of variable length. CMAS fold can be divided in two distinct sub structures: the nucleoside binding motif, mainly made of beta sheets, common to the methyltransferase superfamily; and the substrate binding motif, mostly alpha helical, specific to each family. The active site presents, a series of aromatic residues (Tyrosines and Phenylalanines) which form an aromatic cluster involved in substrate binding [30], the AdoMet/AdoCys binding site is composed of polar amino acids which interacts with the nucleoside and the backbone of the Met/Cys of the cofactor. One remarkable characteristic of some members of the enzyme family is the emergence of a bicarbonate/carbonate binding site (Fig. 1B). This site is formed by a histidine, a cysteine backbone amide, a tyrosine and a glycine. In *mmaA1*, *mmaA3* and *mmaA4* the glycine is replaced by an aspartic or glutamic acid, thus displacing the bicarbonate.

Crystal structures of *cmaA1-2*, *pcaA, mmaA4* and *mmaA2* have been published so we performed comparative modeling of the remaining CMAS *umaA*, *mmaA1* and *mmaA3* (Figure S2) having some minor differences among them. *umaA* turned out to have a different configuration of the ηX helix and the $\eta 1$ helix, not previously observed. Models of *ufaA* and *Rv3720* were not used for the structural analysis due to low identity in the N-terminal region. To analyze substrate recognition we docked a model substrate of the olefin in the active site of each CMAS by rDock.

Based on the characteristics of its active site we classified the proteins in two distinct families: the ability to bind bicarbonate or instead to have an acidic residue in that position. In the first group we find *cmaA1-2*, *pcaA* and *mmaA2* which have electronic density maps compatible with a bicarbonate ion and are deposited in the PDB as such [31]. The modeled umaA, shares the same structural features as this group of CMAS, His, Cys, Tyr and Gly in the binding site, so we classify it as a bicarbonate binder (Fig. 1).

In the second group we have the proteins which have a Glutamic/Aspartic acid instead of the previously conserved Gly, *mmaA1*, *mmaA3* and *mmaA4*. The acidic side chain replaces the bicarbonate ion. While *mmaA1* has an aspartic acid replacing the bicarbonate, both *mmaA3* and *mmaA4* have a glutamic acid residue. A whole family sequence alignment is shown in Figure S1.

3.1. The reaction mechanism of CMAS

Initially we present the free energy profiles for the methyl transfer and the formation of the carbocation intermediate for three model protein systems, *cmaA2, mmaA4 and mmaA1*, selected as reference cases for each product type: cyclopropane, hydroxyl and methyl olefin. We subsequently show the resolution of the carbocation for the three systems.

3.2. Methyl transfer and carbocation formation

The proposed reaction mechanism for the first reaction step is common to all CMAS family members and in particular for the three ones we studied (*cmaA2*, *mmaA4* and *mmaA1*). We performed the corresponding free energy QM/MM simulations starting from the previous modeled structures. The reaction mechanism, the transition state and the free energy profiles are shown in Fig. 2a,b and c respectively. In all cases the transition state configuration is closer to products than to reactants, with $\Delta G^{\#}$ values of 14.39 kcal/mol in *cmaA2* (Fig. 2c, Figs. S3a and b), 14.03 kcal/mol for *mmaA4* (Fig. 2c, Figure S5a) and 12.23 kcal/mol for *mmaA1* (Fig. 2c, Figure S5c) and a ΔG^{0} of 2.7 kcal/mol for *cmaA2*, 7.40 kcal/mol for *mmaA4* and 2.01 kcal/mol for *mmaA1* (Table S1). The same reaction but in water solvent occurs with a $\Delta G^{\#}$ of 16.67 kcal/mol and a ΔG^{0} of 10.50 kcal/mol(Figure S3d). The stabilization of the TS provided by the protein environment can be estimated to be small, only 2.28 kcal/mol for *cmaA2* while the effect is higher for the products.

As already mentioned, Tyr 22 is an active site conserved residue. In order to study its role in CMAS reaction we mutated it to Alanine using *mmA4* as model. The computed free energy profile shows a higher activation barrier (18.34 kcal/mol) and a lower product stability, 9.23 kcal/mol, comparable to the one observed in water (Fig. 2c and Figure S5c). Certainly, the conserved tyrosine plays a role in the stabilization of the transition state and the product. This is also observed when mutating the equivalent tyrosine (Y30) in *cmaA2*, produces a higher barrier and lower stability product (See Figure S3c). Interestingly, when comparing the reaction mechanism by using the so called More O'Ferrall–Jencks plot, in which is possible to see in a consistent manner if the reaction proceeds in a concerted, associative or dissociative way by plotting bond breaking distance (Y axis) and bond forming distance (X axis). If the reaction is concerted a Y = X straight line is plotted, when the bond breaks before the other bond is formed (a dissociative mechanism) the line goes below the concerted one whereas if the reaction has an associative character the line goes above the concerted line. The reaction in the three proteins proceed in an almost concerted manner whereas Y22A mutant present a rather associative mechanism (See also Figure S7 and Figure S9). Therefore, it is clear that Tyr 22 helps the Methyl to be released from SAM. Mulliken populations (Fig. 4) in *cmaA2* are rather stable during the reaction until the transition state when a charge inversion between SAM (ca 0.7 to 0 charge units) and the olefin (0–0.9 charge units) is observed. When looking at the same property in the mutant (mmaA4 Y30A, Figure S11a and c) and the reaction in water solvent a more abrupt increase in charge at early stages is observed as well as a higher absolute value (close to 1 charge unit) for the formed carbocation. This clearly suggest that the aromatic residue present in the proximity of the active site is playing a role by electrostatic stabilization (possibly by means of cation-Pi interactions) of the transition state and the product. This seems to be a conserved feature of

Please cite this article in press as: L.A. Defelipe, et al., Structural and mechanistic comparison of the Cyclopropane Mycolic Acid Synthases (CMAS) protein family of *Mycobacterium tuberculosis*, Biochemical and Biophysical Research Communications (2017), http://dx.doi.org/10.1016/j.bbrc.2017.08.119

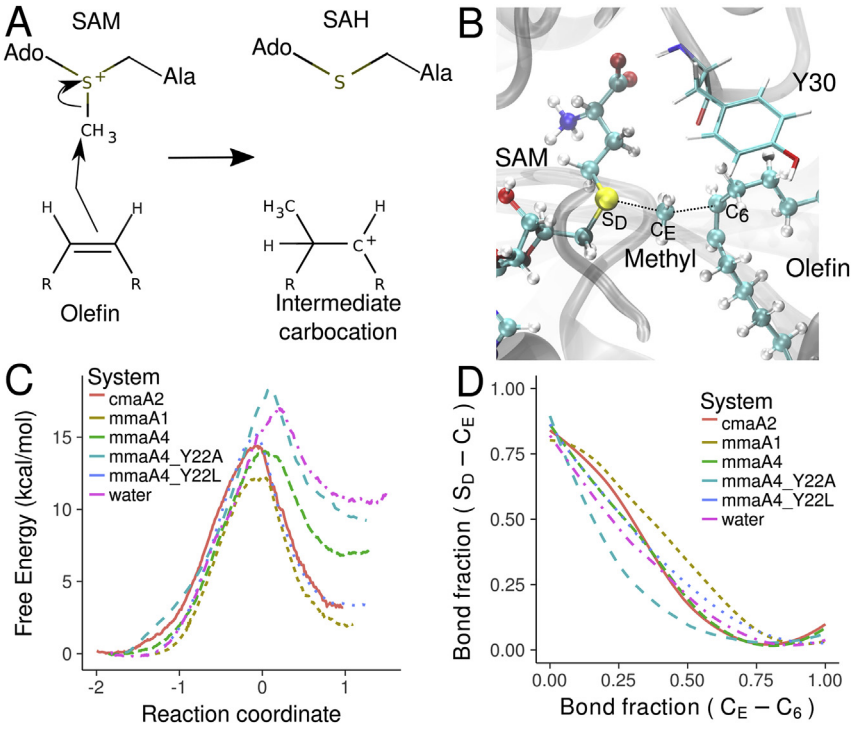

Fig. 2. Reaction mechanism for the carbocation formation in *cmaA2*, *mmaA4*, *mmaA1* and water. A) Reaction mechanism scheme of the first step. B) Transition state snapshot from *cmaA2* showing the planar methyl group. Dashed lines show the sulfur bonds. C) Free energy profile of the first step of the reaction in *cmaA2*, *mmaA4*, *mmaA1* and water. D) More O'Ferrall–Jencks plot.

the three analyzed CMAS. From this data we can conclude that there is no significant difference in the reaction mechanism for the first step in between the three proteins.

3.3. Carbocation resolution

3.3.1. Formation of cyclopropane in cmaA2

The free energy profiles for the carbocation resolution in *cmaA2* were performed in the WT protein and also in the E148A mutant, since E148 is a conserved residue in the cyclopropane subfamily that could be involved in the reaction as it is adjacent to the active site. As mentioned before, *cmaA2* and all members of the cyclopropane subfamily, bind a bicarbonate ion which has been proposed to be responsible for the abstraction of a proton from the methyl group (Fig. 3a), as bicarbonate has a $pK_a = 10.3$. According to our results HCO_3^- abstracts a proton from the methyl group while concomitantly transferring its own proton to E148 during the formation of the new C–C bond. When comparing the $\Delta G^{\#}$ between WT and E148A there is a small but significant difference in the barrier (1.36 vs 2.20 kcal/mol). The reaction in water has a barrier that is significantly higher (24.44 kcal/mol) and the product present a lower stability than in the protein (Fig. 3a and Figure S4). The high $\Delta G^{\#}$ difference could be explained by the presence of the strong bicarbonate base in the right position to subtract the proton from the methyl group. In water, the solvent has to act as a base, a process less favorable. Mulliken populations of the carbocation reach a neutral charge in all simulations. In the case of HCO_3^-, in the WT protein has a value close to −1 while in E148A is closer to 0 (−0.3) (Figure S12). The effect of E148 seems minimal in the reaction mechanism of *cmaA2*. Distance analysis suggests that the three reactions (WT, E148A, and water) show that the reaction mechanism is similar in all systems. (Figure S8). It should be noted that as the QM system is rather small, differences seen between WT and E148A could be lower and non significant. Also the level of theory in use has an RMSE (Root Mean Square Error) higher than 2 kcal/mol which does not allow to discern between WT and E148A reactions.(see Fig. 4)

3.4. Formation of Methyl-alcohol in mmaA4

In *mmaA4* formation of the Methyl-alcohol occurs spontaneously during equilibration time when a rearrangement of carbocation dihedral angles occur. The reaction proceeds with the transfer of a proton from a water molecule to E149 and the concomitant attack of the hydroxyl to the carbocation (Figure S10a and S13). The active site is ready to produce the attack as soon as the dihedral rearrangement happens. Key for the reaction is the presence of E146 and E149 that function as a water trap, keeping a water molecule confined and ready for the attack, once the carbocation is correctly positioned. Results can be seen in Fig. 3c and d, distances and mulliken populations are in Figures S10a and Figure S13a. At first glance *mmaA3* and *mmaA4* have identical structures, but if we take a closer look, there is a significant difference in the configuration of the N-terminal region. The conformation of the ηX helix may determine the solvent accessibility to the active site and therefore regulate the specific enzymatic activity of the CMAS (Figure S16). In *mmaA4* holo and apo structures an "open" conformation is observed. We propose that the open conformation allows water to enter, even when the substrate is bound and therefore *mmaA4* is able to hydroxylate the substrate and not *mmaA3*.

3.5. Formation of Methyl-olefin in mmaA1

The resolution of the carbocation in *mmaA1* also happens with no barrier, similar to the observed results in *cmaA2* (Fig. 3e,f, Figure S6, Figure S10b and Figure S13b). The concomitant transfer of the proton from the carbocation to D139 derives in the formation

Please cite this article in press as: L.A. Defelipe, et al., Structural and mechanistic comparison of the Cyclopropane Mycolic Acid Synthases (CMAS) protein family of *Mycobacterium tuberculosis*, Biochemical and Biophysical Research Communications (2017), http://dx.doi.org/10.1016/j.bbrc.2017.08.119

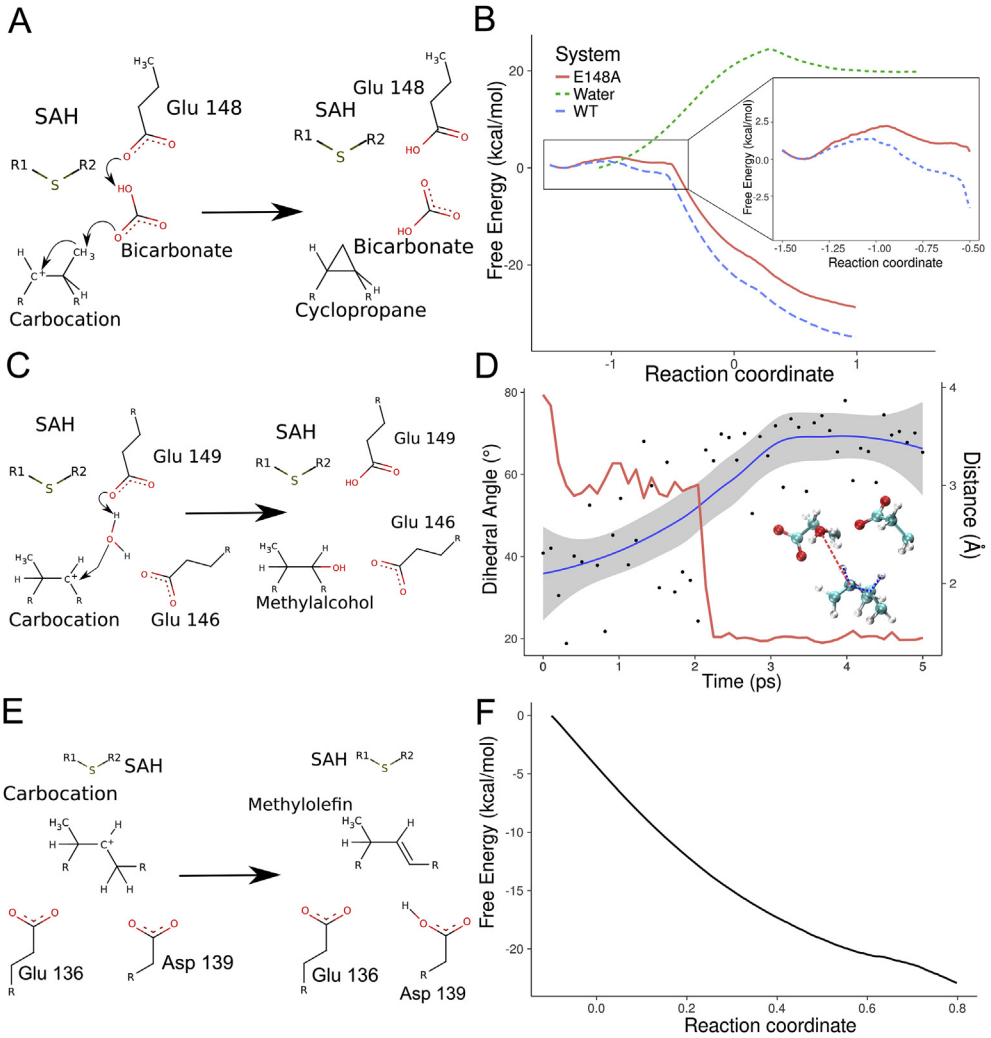

Fig. 3. Reaction mechanism for the resolution of the carbocation in cmaA2, mmaA4 and mmaA1. A) Reaction mechanism scheme for cmaA2. B) Free energy profile for cmaA2. C) Reaction mechanism scheme for mmaA4. D) Dihedral (black point and a trend line in blue, 95% confidence interval in grey) and distance (red line) for barrierless reaction of mmaA4. E) Reaction mechanism scheme for mmaA1. F) Free energy profile for mmaA1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of a new double bond. The reaction is highly exothermic (−22,91 kcal/mol, Table S2).

4. Discussion

4.1. The reaction mechanisms in CMAS

In this work we present the QM/MM reactions for three CMAS (*cmaA2*, mmaA4 and *mmaA1*). We show that the rate limiting step in all three the proteins is the methyl transfer to the olefin, and the formation of the carbocation intermediate which has a relatively high free energy barrier (ca 15 kcal/mol). This result is in agreement with previous reports for *cmaA1* performed with restraint optimizations [32] and experimental kinetic measurements determined in an *E. coli* cyclopropane synthase variant [33,34]. In those works, the authors performed the activity measurements in the presence of the chalcogen series for SAM with Selenium and Tellurium Adenosyl-L-methionine, showing that the reaction rate depends strongly on the identity of the atom, being higher with Selenium and drastically lower with Tellurium. Interestingly, the authors show that, in water solvent the second reaction step becomes the rate limiting step. So CMAS catalyze the reaction, first by bringing reactants and product together, second by lowering the barrier of the first step even though is not a dramatic stabilization, and third by clearly lowering the barrier of the second steps.

When comparing the reactions between *cmaA2*, *mmaA4* and *mmaA1* with respect to the reaction in water, a difference in approximately 4 kcal/mol is observed for the activation energy and of 8 kcal/mol for the products. The stabilization of the carbocation intermediate can be due to the presence of a conserved Tyr because the reaction with an alanine mutant gives a similar barrier to the one in water solvent. This observation is in agreement with a recent study [35] done with a model system for terpene synthase. The study shows that weak interactions such as cation-π can change tendencies for competing products deriving in a strong product selectivity, as the one observed in the case of CMAS. An aromatic residue (Phe, Tyr and Trp) is capable of stabilizing the positive charge with the quadrupolar moment generated by its π electrons. Again, our calculation show that Tyr 30 in *cmaA2* or Tyr 22 in *mmaA4* play a critical role in the stabilization of the transition state (2.4 kcal/mol for $\Delta G^{\#}$ and 4.6 kcal/mol for ΔG^{0} in *cmaA2* while in *mmaA4* the differences are higher, 3 kcal/mol and almost 8 kcal/mol).

As expected, the second step is different for each enzyme since

Please cite this article in press as: L.A. Defelipe, et al., Structural and mechanistic comparison of the Cyclopropane Mycolic Acid Synthases (CMAS) protein family of *Mycobacterium tuberculosis*, Biochemical and Biophysical Research Communications (2017), http://dx.doi.org/10.1016/j.bbrc.2017.08.119

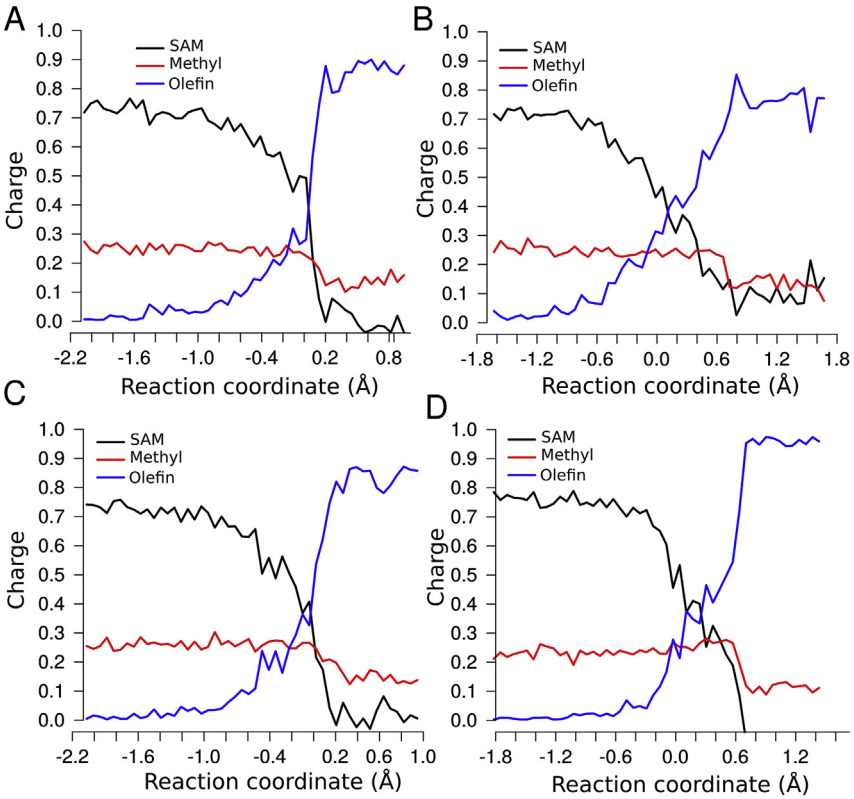

Fig. 4. Mulliken populations for carbocation formation in cmaA2 and water. A) cmaA2@C10, B) cmaA2@C11, C) cmaA2-Y30A@C10 and d) water.

their product is different. In the case of *cmaA2* the cyclization reaction occurs with a low barrier (ca. 2 kcal/mol) while in solution, if it happens at all, it must overcome a very high activation barrier (>24 kcal/mol). The catalytic effect produced by *cmaA2* can be explained in terms of base catalysis; water is a very poor base to subtract a proton from the methyl group while in the protein the bicarbonate ion acts as a base thus providing the catalytic power needed for the proton transfer to happen. As mentioned above differences between systems in cmaA2, due to the small size of the QM region and the intrinsic methodology error could not be significant and the reaction be a barrierless process.

In a previous report [34], the fundamental role of bicarbonate is shown by means of a carbon dioxide scrubbing experiment, allowing to control the exact concentration of bicarbonate in solution. Completely removing CO_2 from solution results in a total activity of 3% compared to the normal system. Also when performing the experiments with amino acid mutants at the bicarbonate binding site (Y317F and H266A) a significant reduction in activity was observed.

In *mmaA4* the reaction directly occurs without any barrier. A simple reorientation of the substrate during simulation time (5 ps) is sufficient for the reaction to happen. The protein scaffold favours the positioning of the water molecule between two glutamic acids which act as a base for the hydroxyl transfer to the carbocation in a selective fashion. Water molecules have a higher residence time near two glutamic acids present in the active site (Glu 146 and 149) than in bulk. Long MD simulations (500ns) of the apo structure and analysis with the WATCLUST tool [36] helped determine this water site strength. Water is 10 times more prone to be on this site than in bulk water, underscoring the role of the protein matrix in trapping the water. Waters can also enter the active site of *cmaA2* but the strength of the site is much lower, about 3 times that of bulk. The observed closed active site in apo *cmaA2* and the lower binding of water may help explain why while *mmaA4* produces the hydroxyl product, *cmaA2* produces cyclopropane (Figures S14 and S15). Looking at the aminoacidic composition we observe that V21 is replaced in mmaA4 to Isoleucine, not allowing for specific hydrogen bonds to form between the helix and the rest of the protein due to the bigger sidechain. In particular, this region of mmaA4 is not able to form two critical interactions (His24-Tyr42 and Tyr25-His150) which stabilizes the helix in the closed conformation in other CMAS. We hypothesize that these destabilize the alpha-helix leading to the formation of a 3−10 helix in an open conformation.

Finally, regarding *mmaA1*, the computed free energy profile predicts no barrier for the reaction. Most likely, the difference in activity between *cmaA2* (and other bicarbonate binding CMAS) and *mmaA1* is the presence of an aspartic acid instead of a glycine and bicarbonate ion. In this sense, the length of the aspartic acid chain also influences the possibility of an attack to the methyl group, being able to attack only the sp^3 carbon next to the carbocation constrained by the protein matrix.

4.2. A predictive model for the activity of CMAS

A combination of the reaction mechanism results from *cmaA2*, *mmaA4* and *mmaA1* (Free Energy QM/MM calculations), dynamical configuration of the proteins and their interaction with solvent(all atom classical MD) and the structural/sequence alignment (Comparative Modeling and Docking tools) helped us to propose a model that explains the differential activity in the whole CMAS family as shown in Fig. 5 and Figure S16. Starting with the presence of a glycine and a bicarbonate ion, we can conclude that these proteins (*cmaA1-2*, *pcaA* and *mmaA2*) will produce a cyclopropane as their main product. If in that position an Aspartic acid is found, the common product will be a methyl-olefin as produced by

Please cite this article in press as: L.A. Defelipe, et al., Structural and mechanistic comparison of the Cyclopropane Mycolic Acid Synthases (CMAS) protein family of *Mycobacterium tuberculosis*, Biochemical and Biophysical Research Communications (2017), http://dx.doi.org/10.1016/j.bbrc.2017.08.119

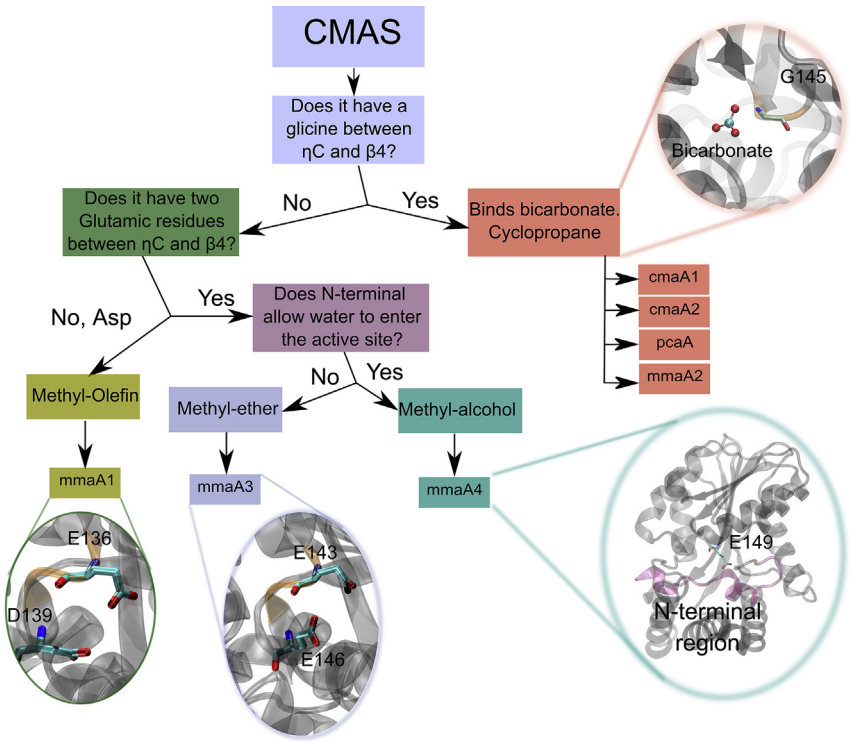

Fig. 5. Flowchart of structural determinants for CMAS differential activity. Each group has a close up view of the important features of its active site or in the overall structure.

mmaA1. The other option is to have a Glutamic acid residue in that position as found in *mmaA3* and *mmaA4*. In this case the product formation is controlled by the entrance of water to the active site by the N-Terminal region. In *mmaA3*, as shown in the model, the N-terminal acts as a lid preventing water from entering the active site whereas in *mmaA4* the position of this "lid" allows the entrance of water molecules. We propose that the lack of water leads to the formation of methyl-ethers in *mmaA3*, while their presence leads to methyl-alcohols, as seen in *mmaA4*. Moreover, this open-closed conformational change produces also a differential positioning of a conserved Tyrosine residue, in *mmaA3* the residue points towards the active site while in *mmaA4* is solvent exposed. This Tyrosine position allows a strong hydrogen bond to be formed with the alcohol group of the olefin, favouring the binding of the hydroxylated product in *mmaA3* and therefore the attack by SAM to yield methyl ether. The highly hydrophobic site could, in turn, shift the pKa of the two glutamic acids to higher values, favouring the proton transfer from alcohol to the glutamic acids during the methyl ether formation *mmaA3*, this result can be seen in propKa pKa computations done from the crystal or modeled structures shown in Table S3. Although propKa is not a precise methodology, it is able to reproduce the tendency of the pKa change. cmaA2, the one that generates a cyclopropane, has a bicarbonate and a glutamic that will finally get the proton, has a pKa of 7.1. On the other hand mmaA4, the one that generates a methyl alcohol, has two glutamic acids which according to propKa are coupled and have a pKa of 5.7 and 8.3, indicating that one of them can easily subtract a proton from a water molecule generating the OH− that will form the alcohol. While mmaA3, the one that generates the methyl ether, has two glutamic acids and have a pka of 6.1 and 11.6, this strong basic glutamic is able to subtract a proton from the alcohol produced by mmaA4.The combination of modeling tools at different levels of complexity helped us to shed light into the determinants of product selectivity in this important protein family of *Mtb*.

Acknowledgements

This work was supported by grants PICT 2010−2805, PICTO GSK 2012-0057 and PIP 11220130100469CO awarded to AGT and MAM. The authors would like to thank Centro de Cómputos de Alto Rendimiento (CeCAR), FCEN-UBA and TUPAC−CSC−CONICET for granting use of computational resources which allowed us to perform the experiments included in this work. LAD is a CONICET postdoctoral fellow. AGT and MAM are members of the CONICET.

Transparency document

Transparency document related to this article can be found online at http://dx.doi.org/10.1016/j.bbrc.2017.08.119.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.bbrc.2017.08.119.

References

[1] W.H. Organization, Others, Global tuberculosis report 2016, 2016. http://apps.who.int/iris/bitstream/10665/250441/1/9789241565394-eng.pdf.
[2] J.A. Caminero, G. Sotgiu, A. Zumla, G.B. Migliori, Best drug treatment for multidrug-resistant and extensively drug-resistant tuberculosis, Lancet Infect. Dis. 10 (2010) 621–629.
[3] M.I. Voskuil, I.L. Bartek, K. Visconti, G.K. Schoolnik, The response of Mycobacterium tuberculosis to reactive oxygen and nitrogen species, Front. Microbiol. 2 (2011).
[4] A. Demissie, E.M.S. Leyten, M. Abebe, L. Wassie, A. Aseffa, G. Abate, H. Fletcher, P. Owiafe, P.C. Hill, R. Brookes, G. Rook, A. Zumla, S.M. Arend, M. Klein, T.H.M. Ottenhoff, P. Andersen, T.M. Doherty, VACSEL Study Group, Recognition of stage-specific mycobacterial antigens differentiates between acute and latent infections with Mycobacterium tuberculosis, Clin. Vaccine Immunol. 13 (2006) 179–186.
[5] A.M. Abdallah, N.C.G. van Pittius, P.A.D. Champion, J. Cox, J. Luirink, C.M. Vandenbroucke-Grauls, B.J. Appelmelk, W. Bitter, Type VII

secretion—mycobacteria show the way, Nat. Rev. Microbiol. 5 (2007) 883–891.
[6] H. Marrakchi, M.-A. Lanéelle, M. Daffé, Mycolic acids: structures, biosynthesis, and beyond, Chem. Biol. 21 (2014) 67–85.
[7] L.A. Defelipe, D.F. Do Porto, P.I. Pereira Ramos, M.F. Nicolás, E. Sosa, L. Radusky, E. Lanzarotti, A.G. Turjanski, M.A. Marti, A whole genome bioinformatic approach to determine potential latent phase specific targets in Mycobacterium tuberculosis, Tuberculosis 97 (2016) 181–192.
[8] Y. Yuan, R.E. Lee, G.S. Besra, J.T. Belisle, C.E. Barry, Identification of a gene involved in the biosynthesis of cyclopropanated mycolic acids in Mycobacterium tuberculosis, Proc. Natl. Acad. Sci. 92 (1995) 6630–6634.
[9] M.S. Glickman, S.M. Cahill, W.R. Jacobs, The Mycobacterium tuberculosis cmaA2 gene encodes a mycolic acid trans-cyclopropane synthetase, J. Biol. Chem. 276 (2001) 2228–2233.
[10] M.S. Glickman, The mmaA2 gene of Mycobacterium tuberculosis encodes the distal cyclopropane synthase of the α-mycolic acid, J. Biol. Chem. 278 (2003) 7844–7849.
[11] D. Barkan, V. Rao, G.D. Sukenick, M.S. Glickman, Redundant function of cmaA2 and mmaA2 in Mycobacterium tuberculosis cis cyclopropanation of oxygenated mycolates, J. Bacteriol. 192 (2010) 3661–3668.
[12] M.S. Glickman, J.S. Cox, W.R. Jacobs Jr., A novel mycolic acid cyclopropane synthetase is required for cording, persistence, and virulence of Mycobacterium tuberculosis, Mol. Cell. 5 (2000) 717–727.
[13] Y. Yuan, D.C. Crane, J.M. Musser, S. Sreevatsan, C.E. Barry, MMAS-1, the branch point between cis-and trans-cyclopropane-containing oxygenated mycolates in Mycobacterium tuberculosis, J. Biol. Chem. 272 (1997) 10041–10049.
[14] Y. Yuan, C.E. Barry, A common mechanism for the biosynthesis of methoxy and cyclopropyl mycolic acids in Mycobacterium tuberculosis, Proc. Natl. Acad. Sci. 93 (1996) 12828–12833.
[15] M.A. Behr, B.G. Schroeder, J.N. Brinkman, R.A. Slayden, C.E. Barry, A point mutation in the mma3 gene is responsible for impaired methoxymycolic acid production in Mycobacterium bovis BCG strains obtained after 1927, J. Bacteriol. 182 (2000) 3394–3399.
[18] N. Eswar, D. Eramian, B. Webb, M.-Y. Shen, A. Sali, Protein structure modeling with MODELLER, in: Structural Proteomics, Springer, 2008, pp. 145–159.
[19] S. Ruiz-Carmona, D. Alvarez-Garcia, N. Foloppe, A.B. Garmendia-Doval, S. Juhos, P. Schmidtke, X. Barril, R.E. Hubbard, S.D. Morley, rDock: a fast, versatile and open source program for docking ligands to proteins and nucleic acids, PLoS Comput. Biol. 10 (2014) e1003571.
[20] S. David Morley, M. Afshar, Validation of an empirical RNA-ligand scoring function for fast flexible docking using RiboDock®, J. Comput. Aided Mol. Des. 18 (2004) 189–208.
[21] D.A. Case, V. Babin, J. Berryman, R.M. Betz, Q. Cai, D.S. Cerutti, T.E. Cheatham Iii, T.A. Darden, R.E. Duke, H. Gohlke, Others, amber 14, (2014).
[22] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Development and testing of a general amber force field, J. Comput. Chem. 25 (2004) 1157–1174.
[23] R. Salomon-Ferrer, A.W. Götz, D. Poole, S. Le Grand, R.C. Walker, Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. explicit solvent particle mesh Ewald, J. Chem. Theory Comput. 9 (2013) 3878–3888.
[24] M. Elstner, The SCC-DFTB method and its application to biological systems, Theor. Chem. Acc. 116 (2006) 316–325.
[25] G. de M Seabra, R.C. Walker, M. Elstner, D.A. Case, A.E. Roitberg, Implementation of the SCC-DFTB method for hybrid QM/MM simulations within the amber molecular dynamics package, J. Phys. Chem. A 111 (2007) 5655–5664.
[26] R.J. Loncharich, B.R. Brooks, R.W. Pastor, Langevin dynamics of peptides: the frictional dependence of isomerization rates of N-acetylalanyl-N′-methylamide, Biopolymers 32 (1992) 523–535.
[27] C. Jarzynski, Nonequilibrium equality for free energy differences, Phys. Rev. Lett. 78 (1997) 2690.
[28] L.A. Defelipe, E. Lanzarotti, D. Gauto, M.A. Marti, A.G. Turjanski, Protein topology determines cysteine oxidation fate: the case of sulfenyl amide formation among protein families, PLoS Comput. Biol. 11 (2015) e1004051–e1004051.
[29] C.L. Ramirez, A. Zeida, G.E. Jara, A.E. Roitberg, M.A. Marti, Improving efficiency in SMD simulations through a hybrid differential relaxation algorithm, J. Chem. Theory Comput. 10 (2014) 4609–4617.
[30] E. Lanzarotti, R.R. Biekofsky, D.A. Estrin, M.A. Marti, A.G. Turjanski, Aromatic–aromatic interactions in proteins: beyond the dimer, J. Chem. Inf. Model 51 (2011) 1623–1633.
[31] C.-C. Huang, C.V. Smith, M.S. Glickman, W.R. Jacobs, J.C. Sacchettini, Crystal structures of mycolic acid cyclopropane synthases from Mycobacterium tuberculosis, J. Biol. Chem. 277 (2002) 11559–11569.
[32] R.-Z. Liao, P. Georgieva, J.-G. Yu, F. Himo, Mechanism of mycolic acid cyclopropane synthase: a theoretical study, Biochemistry 50 (2011) 1505–1513.
[33] D.F. Iwig, A. Uchida, J.A. Stromberg, S.J. Booker, The activity of Escherichia coli cyclopropane fatty acid synthase depends on the presence of bicarbonate, J. Am. Chem. Soc. 127 (2005) 11612–11613.
[34] D.F. Iwig, A.T. Grippe, T.A. McIntyre, S.J. Booker, Isotope and elemental effects indicate a rate-limiting methyl transfer as the initial step in the reaction catalyzed by Escherichia coli cyclopropane fatty acid synthase, Biochemistry 43 (2004) 13510–13524.
[35] S.R. Hare, R.P. Pemberton, D.J. Tantillo, Navigating past a fork in the road: carbocation-π interactions can manipulate dynamic behavior of reactions facing post-transition-state bifurcations, J. Am. Chem. Soc. (2017), http://dx.doi.org/10.1021/jacs.7b01042.
[36] E.D. López, J.P. Arcon, D.F. Gauto, A.A. Petruk, C.P. Modenutti, V.G. Dumas, M.A. Marti, A.G. Turjanski, WATCLUST: a tool for improving the design of drugs based on protein-water interactions, Bioinformatics 31 (2015) 3697–3699.

Please cite this article in press as: L.A. Defelipe, et al., Structural and mechanistic comparison of the Cyclopropane Mycolic Acid Synthases (CMAS) protein family of *Mycobacterium tuberculosis*, Biochemical and Biophysical Research Communications (2017), http://dx.doi.org/10.1016/j.bbrc.2017.08.119