# Alma Mater Studiorum – Università di Bologna

## Dottorato di Ricerca in Scienze Farmaceutiche
XXI Ciclo (2006-2008)

## CHIM/08

Coordinatore: Chiar.mo Prof. Maurizio Recanatini

# Computational investigation of the *Plasmodium falciparum* fatty acid biosynthetic pathway toward the discovery of novel antimalarials

Tesi di Dottorato presentata da

## Francesco Colizzi

Supervisor:                                                                 Advisor:

**Chiar.mo Prof.**
**Maurizio Recanatini**                                          **Dr. Andrea Cavalli**

*Doctor Philosophiæ*

**Esame finale anno 2009**

## Abstract

The structural peculiarities of a protein are related to its biological function. In the fatty acid elongation cycle, one small carrier protein shuttles and delivers the acyl intermediates from one enzyme to the other. The carrier has to recognize several enzymatic counterparts, specifically interact with each of them, and finally transiently deliver the carried substrate to the active site. Carry out such a complex game requires the players to be flexible and efficiently adapt their structure to the interacting protein or substrate. In a drug discovery effort, the structure-function relationships of a target system should be taken into account to optimistically interfere with its biological function. In this doctoral work, the essential role of structural plasticity in key steps of fatty acid biosynthesis in *Plasmodium falciparum* is investigated by means of molecular simulations. The key steps considered include the delivery of acyl substrates and the structural rearrangements of catalytic pockets upon ligand binding. The ground-level bases for carrier/enzyme recognition and interaction are also put forward. The structural features of the target have driven the selection of proper drug discovery tools, which captured the dynamics of biological processes and could allow the rational design of novel inhibitors. The model may be perspectively used for the identification of novel pathway-based antimalarial compounds.

*"ma guai a chi cede alla tentazione di scambiare una ipotesi elegante con una certezza"*
Primo Levi. Il Sistema Periodico, 1975.

# Contents

# Prologue

*Bad Air*

*Until the mid-nineteenth century, most scientists thought that noxious swamp gases caused malaria, indeed the word means "bad air" in italian.*[1]

Malaria remains a major and growing threat to the public health and economic development of countries in the tropical and subtropical regions of the world.

This dissertation does not count for explicitly introduce the malaria burden; for this aim, the reader will find better satisfaction looking at the references listed at the end of this session.[1-13] The general aspects of fatty acid biosynthesis in *Plasmodium falciparum* will be briefly discussed in the following pages and some recall can be found along the argumentation of each chapter.

I will focus my dissertation on the advances that the molecular simulations I carried out may have brought in understanding some molecular aspects in the biology of fatty acid biosynthesis in parasites. This biological understanding is integrated with a computational drug discovery effort aimed at the design of small molecules endowed with the capability to interfere with the fatty acid production in *Plasmodium falciparum*. The organization of each chapter will be likely similar to that of a scientific article. The results discussed in each chapter are used as starting point for the following one. First, the substrate delivery issue as intrinsic properties of the carrier is discussed. Then the delivery of substrates is considered from a wider standpoint which takes into account the way the carrier and its counterpart may interact, and recognize each other. Structural plasticity is found to play a critical role in each of the above steps and is therefore taken into account to address the computational medicinal chemistry part of this dissertation.

# 1. Introduction: the type II fatty acids biosynthesis

The synthesis of fatty acids is organized into two distinct biosynthetic pathways based on the enzymes involved. In mammals and fungi, the fatty acid biosynthesis machinery resides on a single multifunctional polypeptide with multiple active sites (Type I).[14] This multifunctional ensemble is supposed to be a product of an evolutionary flow resulting from the fusion of genes encoding individual monofunctional enzymes.[15] These monofunctional enzymes constitute the Type II system.[16-18]

Malaria parasite and other members of the *phylum* apicomplexa harbor a relict plastid, known as apicoplast, homologous to the chloroplast of plants and algae. The apicomplexan, together with plants and most bacteria use a dissociated Type II fatty acids biosynthesis (FAS II) consisting of at least nine separate polypeptides, each of which may have a dimeric, tetrameric or higher order of quaternary structure.[18] A key feature of the FAS II is the presence of a small, acidic and highly conserved acyl carrier protein (ACP) that shuttles all the covalently bound fatty acyl intermediates from one enzyme to the other.

Newly synthesized ACP must be converted from its *apo* into *holo* form, by the attachment of a prosthetic group in order to participate in fatty acid biosynthesis. This posttranslational modification is catalyzed by ACP synthase which transfer a 4'-phosphopantetheine group from coenzyme A (CoA) to a conserved serine residue on the ACP. The acyl intermediates are bound through a thioester linkage to the sulfhydryl of the prosthetic group. Fatty acids biosynthesis begins at the acetyl-CoA carboxilase (ACC). The substrate acetyl-CoA is converted to malonyl-CoA and the malonate group is transferred to ACP by malonyl-CoA:ACP transacylase (FabD) to form malonyl-ACP.
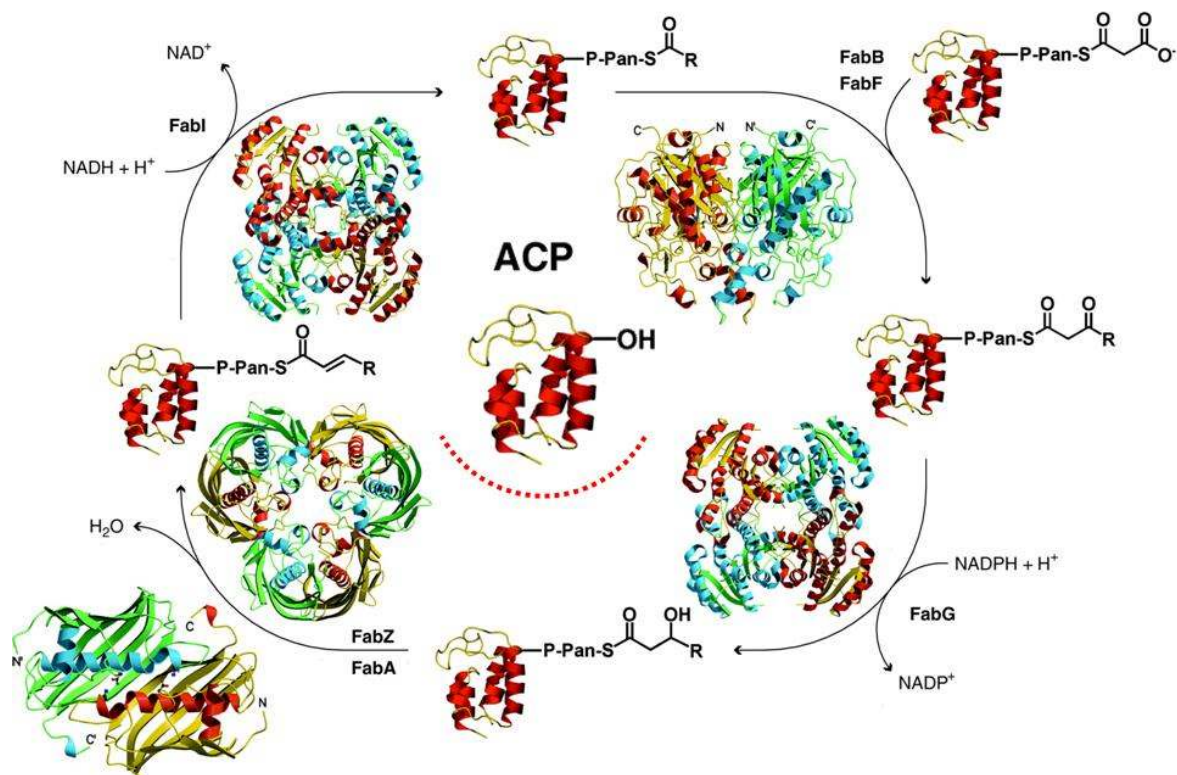
Figure 1. An insightful portrait of Type II fatty acid biosynthesis elongation cycle. Picture adapted from White et al.[18]. Note the central role of the Acyl carrier protein (ACP).

A Claisen condensation assisted by β-ketoacyl-ACP synthase III (FabH) initiates a cycle that elongates the acyl-ACP by two carbons unit for each cycle until a saturated fatty acid of 16 or 18 carbons is made (the elongation cycle has been insightfully depicted by White et al.[18] as shown in Figure 1). The first reaction in this cycle is the NADPH-dependent reduction of β-ketoacyl-ACP to β-hydroxyacyl-ACP by β-ketoacyl-ACP reductase (FabG), and then the β-hydroxyl intermediate is dehydrated to yield trans-2-enoyl-ACP catalyzed by either β-hydroxydecanoyl-ACP dehydratase/isomerase (FabA) or β-hydroxyacyl-ACP dehydratase (FabZ, which is the unique β-hydroxyacyl-ACP dehydratase in *Plasmodium*). The last step of the cycle is the NADH-dependent

9

reduction of the double bond in the trans-2-enoyl-ACP intermediate by an enoyl-ACP reductase I (FabI) or by other analogue reductases. The product of FabI reduction, enoyl-ACP, is the substrate for the subsequent elongation cycle by condensation with malonyl-ACP catalyzed either by β-ketoacyl-ACP synthase I (FabB) or β-ketoacyl-ACP synthase II (FabF). Palmitic acid, for instance, formed by reiteration of this cycle can be either elongated by another set of enzymes or channelized for the formation of phospholipids and other molecules. All Type II systems have this basic set of enzymes to initiate and elongate acyl chains, and the diversity of products is achieved by variation on this theme.[19, 20]

Even if similar, different gene variations and gene expression regulate in various organisms the fatty acids chain length, branches and saturated/unsaturated balance.[20]

**References**

1.      Whitfield, J., Portrait of a serial killer. *Nature (News)* **2002,** 01 Oct.

2.      Bannister, L.; Mitchell, G., The ins, outs and roundabouts of malaria. *Trends Parasitol* **2003,** 19, (5), 209-13.

3.      Breman, J. G., The ears of the hippopotamus: manifestations, determinants, and estimates of the malaria burden. *Am J Trop Med Hyg* **2001,** 64, (1-2 Suppl), 1-11.

4.      Christophides, G. K., Transgenic mosquitoes and malaria transmission. *Cell Microbiol* **2005,** 7, (3), 325-33.

5.      Grabowsky, M., The billion-dollar malaria moment. *Nature* **2008,** 451, (7182), 1051-2.

6.      Guinovart, C.; Navia, M. M.; Tanner, M.; Alonso, P. L., Malaria: burden of disease. *Curr Mol Med* **2006,** 6, (2), 137-40.

7.      Miller, L. H.; Baruch, D. I.; Marsh, K.; Doumbo, O. K., The pathogenic basis of malaria. *Nature* **2002,** 415, (6872), 673-9.

8.      Mills, A.; Lubell, Y.; Hanson, K., Malaria eradication: the economic, financial and institutional challenge. *Malar J* **2008,** 7 Suppl 1, S11.

9.      Targett, G. A.; Greenwood, B. M., Malaria vaccines and their potential role in the elimination of malaria. *Malar J* **2008,** 7 Suppl 1, S10.

10.     White, N. J., The role of anti-malarial drugs in eliminating malaria. *Malar J* **2008,** 7 Suppl 1, S8.

11.     White, N. J., How antimalarial drug resistance affects post-treatment prophylaxis. *Malar J* **2008,** 7, 9.

12.     White, N. J., Qinghaosu (artemisinin): the price of success. *Science* **2008,** 320, (5874), 330-4.

13.     Winzeler, E. A., Malaria research in the post-genomic era. *Nature* **2008,** 455, (7214), 751-6.

14.      Leibundgut, M.; Maier, T.; Jenni, S.; Ban, N., The multienzyme architecture of eukaryotic fatty acid synthases. *Curr Opin Struct Biol* **2008,** 18, (6), 714-25.

15.      Smith, S., The animal fatty acid synthase: one gene, one polypeptide, seven enzymes. *FASEB J* **1994,** 8, (15), 1248-59.

16.      Goodman, C. D.; McFadden, G. I., Fatty acid biosynthesis as a drug target in apicomplexan parasites. *Curr Drug Targets* **2007,** 8, (1), 15-30.

17.      Goodman, C. D.; McFadden, G. I., Fatty acid synthesis in protozoan parasites: unusual pathways and novel drug targets. *Curr Pharm Des* **2008,** 14, (9), 901-16.

18.      White, S. W.; Zheng, J.; Zhang, Y. M.; Rock, The structural biology of type II fatty acid biosynthesis. *Annu Rev Biochem* **2005,** 74, 791-831.

19.      Heath, R. J.; Rock, C. O., The Claisen condensation in biology. *Nat Prod Rep* **2002,** 19, (5), 581-96.

20.      Lu, Y. J.; Zhang, Y. M.; Rock, C. O., Product diversity and regulation of type II fatty acid synthases. *Biochem Cell Biol* **2004,** 82, (1), 145-55.

## 2. Mechanical features of Plasmodium falciparum acyl carrier protein in the delivery of substrates

The Acyl Carrier Protein (ACP) is a key element in the biosynthesis of fatty acids being responsible for the acyl group shuttling and delivery within a series of related enzymes. The molecular mechanism of the delivery process is poorly known, and its characterization is essential for the in-depth understanding of the biosynthetic machinery. A steered molecular dynamics approach has been applied to shed light on the putative delivery pathway, suggesting the small α3-helix to act as gatekeeper for the transfer process. Preventing the delivery mechanism would be an innovative strategy for the development of pathway-based antimalarial compounds.

*Plasmodium falciparum* (*Pf*) infections are the most widespread and lethal form of Malaria. Despite the centenary effort to eradicate or control the disease, more than one third of the human population lives in endemic areas with an estimated half billion of infections annually resulting in about two millions of death.[1-3]

The recently disclosed type II fatty acid synthesis (FAS-II) pathway of *Pf* is offering attractive targets potentially enabling the discovery and development of efficacious and selective antimalarial agents. FAS-II system relies on a dissociative process that exploits a series of individual enzymes that are, indeed, structurally different from the multifunctional type I fatty acid synthase (FAS-I) system of humans.[4,5] A key feature of the FAS-II is the presence of a small (~ 9 kDa), acidic and highly conserved Acyl Carrier Protein (ACP) that shuttles all the acyl intermediates from one enzyme to the other. The acyl substrates are bound to ACP through a flexible arm formed by the serine-bound prosthetic phosphopantetheine (4'-PP) group (Figure 1a).[6] Nevertheless, ACPs are also known to play a fundamental role in numerous other

biosynthetic pathways in which acyl transfer steps are required.[7-9] The geometric properties of several conformationally distinct ACPs have been determined by X-ray crystallography and NMR.[10-15] These experimental data have suggested that structural plasticity is an intrinsic feature of this protein family, providing a possible explanation for ACPs' capability to recognize multiple enzyme partners and transiently deliver the acyl group to the active sites of these enzymes.[16]
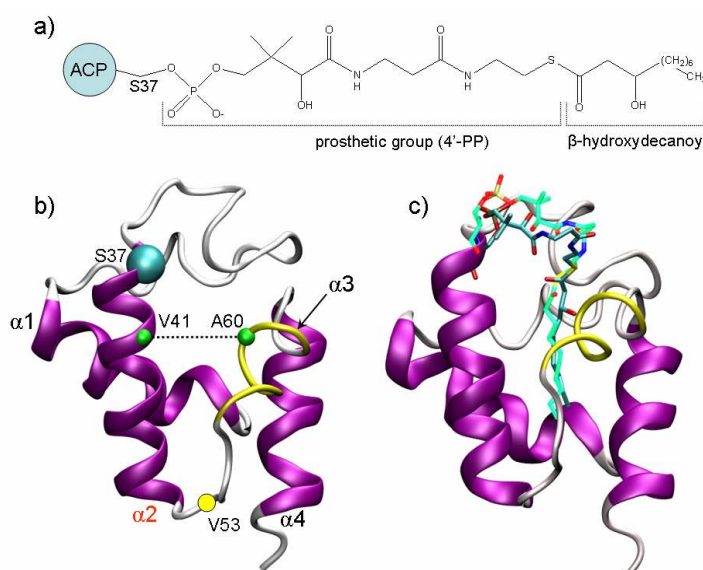


**Figure 1.** a) 2D structure of the acylated 4'-phosphopantetheine (4'-PP) prosthetic group in the β-hydroxydecanoyl-ACP. b) Overall fold of ACPs; the α3-helix (residues 57-62) is depicted in yellow tube. c) Binding mode of the β-hydroxylated substrate (carbon atoms in cyan) after the MD equilibration compared to the decanoyl substrate (carbon atoms in light green) reported in *E. coli* crystal structure (pdb code: 2fae).

ACP structures are composed of four α-helices delimiting a lipophilic core that forms a binding pocket for fatty acids. Helix α2 has highly conserved residues and plays a major role in ACP-protein interactions (Figure 1b)[17] whereas the short α3-helix does not have a conserved folding among the published ACP structures and, accordingly, this protein portion was observed to experience a helix-loop conformational equilibrium in *Pf*ACP.[14] Moreover, the analysis of

experimental and computational studies suggested the protein region between helices α2 and α4 to be part of the putative ACP/enzyme interface.[18-26] The functional implication of the above mentioned structural features are yet unrevealed when observed from the standpoint of the delivery of substrates.

X-ray structures of acylated ACPs have shown the thioester-bound acyl-chain to be embedded into a tunnel-like hydrophobic cavity (Figure 1c), which can harbor acyl substrates of different length.[12,15] Notably, the reactive center of the acyl substrate is buried in the hydrophobic core of the carrier, and thus inaccessible to catalytic activities of FAS-II enzymes. Therefore, delivering the substrate from the inner ACP core to the FAS-II enzyme active site is a mandatory event for each biosynthetic step. The molecular understanding of  ACP's ability to deliver substrates together with the ability of drug designers to interfere with this process might be a new strategy for the development of innovative FAS-II inhibitors that could behave as pathway modulators rather than single-enzyme blockers.

Limited structural information about the complex of ACP with biological counterparts and the mechanism of substrate delivery is currently available.[18,27,28] Recently, Leibundgut et al.,[27] addressed the delivery issue discussing the crystal structure of yeast fatty acid synthase (FAS-I type) system with its ACP stalled at one catalytic domain. They have suggested a general switchblade-like mechanism in which the 4'-PP arm delivers the acyl chain flipping from the ACP core into the catalytic domain active site.[27] However, even though ACPs share the same structural motif in the acyl-substrate binding region, when compared to the FAS-II ACP, the yeast ACP domain (~ 18 kDa) has four additional C-terminal α-helices, which take part in the interaction with the catalytic domain. Therefore, differences in the recognition and delivery process between FAS-I and FAS-II systems are likely to occur. Furthermore, substrate delivery

is a dynamic process, and the detailed but static X-ray picture requires to be complemented with other approaches that are able to directly capture the dynamics of the biomolecule under investigation.

In the present work, the dynamics of the delivery mechanism of the β-hydroxydecanoyl substrate by *Pf*ACP was computationally investigated using a Steered Molecular Dynamics (SMD) approach. Analyzing the trajectories and the force profiles, we were able to identify the lowest resistance pathway for the substrate delivery process. In addition, our simulations pointed out both the role of α3-helix as gatekeeper for the substrate transfer process, and the effect of the substrate β-hydroxylation on the enzyme recognition.
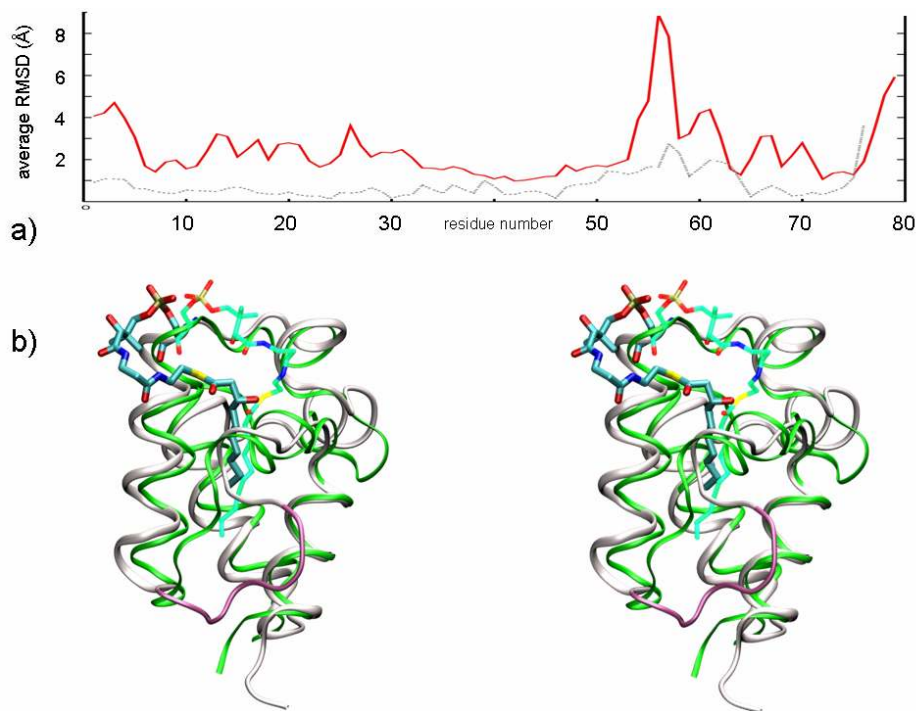


**Figure 2.** a) Root mean square fluctuation (RMSf) of the *Pf*ACP backbone along the whole MD simulation (red line), and RMSf of up-to-date available crystal structures of *E. coli* ACP after Cα superposition (dotted gray line). b) Stereo view of the β-hydroxydecanoyl-ACP (white ribbon) as it appears after 8 ns of MD. The prosthetic group (cyan) lies on the mouth of the fissure between helices α2 and α3. The highly flexible portion going from 53 to 57 is magenta. The decanoyl-ACP crystal structure (in green, pdb code: 2fae) is shown for comparison.

In SMD simulations, which have been widely and successfully applied to explore the properties of non-equilibrium processes of biomolecules,[29,30] a moving harmonic potential is used to induce a motion along a reaction coordinate. The free end of a virtual spring is moved at constant velocity, while a set of "pulled atoms", attached to the other end of the spring, are subject to steering forces. The applied forces are determined by the extension of the spring, and can be monitored throughout the entire simulation. If the pulled atoms can easily advance along the selected reaction coordinate, the applied force is small and its profile is rather flat. Conversely, if the pulled atoms encounter hindrance along the pathway, the force increases to allow the pulled atoms overcoming energy barriers, thus resulting in quite relevant drops in the force profile. Based on the magnitude of the exerted force, it is possible to determine how easily a pathway can be coursed.[29,30]

The simulations of the covalently bound β-hydroxydecanoyl-ACP complex were performed in explicit water, using the CHARMM[31] force field and the program NAMD[32] [see Appendix (Ap) for details]. To get insights into intrinsic properties of the system, the complex was first simulated for 8 ns of unrestrained MD. The average value of RMSD calculated for the backbone along the whole simulation time was 2.9 Å. Nonetheless, the region including residues 53-62 showed high flexibility (RMSf up to 8 Å, red plot in Figure 2a) reflecting the higher ratio of conformational diversity observed by superimposing ACP crystal structures (dotted grey plot in Figure 2a). During the free MD simulations the β-hydroxyacyl moiety of the bound ligand behaved like a fishing float inside the lipophilic core, and the acyl moiety fluctuated assuming several conformations, according to the existence of multiple low-occupancy conformers observed in the crystal structure of decanoyl-ACP.[15] The substrate β-hydroxyl group was able to interact with several polar residues (mainly the backbone of D59, A60, and I63) lining the

entrance of the ACP binding cavity. As a result, the β-hydroxydecanoyl was less buried in the protein core when compared to the experimentally reported decanoyl analogue. This had the consequence that the prosthetic group (4'-PP) was more relaxed and able to fluctuate among several conformations. Figure 2b shows the prosthetic group protruding towards the solvent, self-docking at the upper fissure between helices α2 and α3. In such a conformation, the FAS-II enzymatic counterparts might still recognize the conserved α2-helix, selectively interact with the prosthetic group, and establish connections with protein segment 53-62. Roujenikova et al.[15] proposed that the fairly flexible prosthetic arm may adopt a set of different conformations, likely related to the carried acyl substrate, to allow optimal interaction of acyl-ACP intermediates to partner enzymes. However, the MD simulations did not provide any mechanistic explanation of the delivery process. This is likely a consequence of high potential energy barriers implicated, which cannot be sampled by ns-time-scale simulations. To overcome this drawback, we applied SMD (Figure 3) to drive the substrate along several putative reaction coordinates (see Figure 3a).
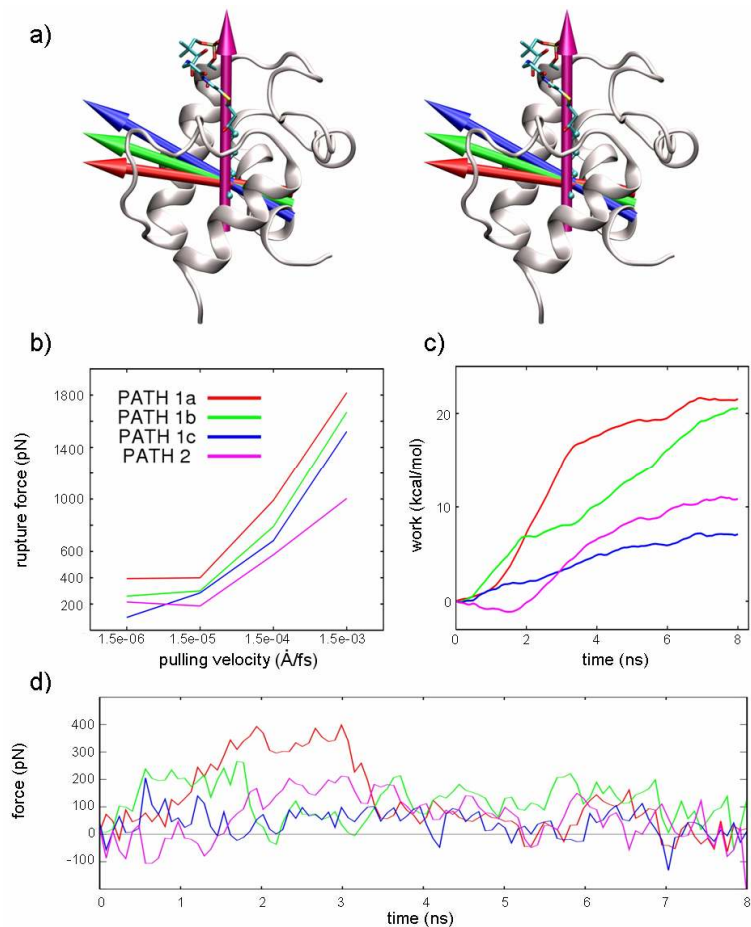
**Figure 3.** a) Acyl-ACP and the delivery pathways investigated. The arrows represent the pulling directions of the applied forces. b) Correlation of the pulling velocity with the rupture force. At 1.5e-06 Å/fs the process is close to equilibrium. c) Mechanical work done on the system calculated by numerical integration. d) Force profiles of the investigated pathways at 1.5e-06 Å/fs. For sake of clarity the curves are obtained averaging the forces every 30 points.

Based on the inspection of the structural features of the reported acyl-ACPs, and on the aforementioned experimental data,[18-26] several delivery pathways were initially considered. However, some of them were discarded because of either they generated unlikely structural distortion or were scarcely in agreement with the available experimental data.[18-26] Two main putative delivery pathways were finally investigated (Figure 3a): the first one accounting for the substrate exposure through the fissure formed by helices α2 and α3 (path 1); the second one

18

accounting for a sword-unsheathed-like mechanism, in which the substrate is unthreaded away in a parallel direction with respect to the major axis of the binding pocket (path 2). Since for path 1 a protein conformational rearrangement was required, and considering critical the role of the pulling direction, three slightly different sub-paths (path 1a, 1b, and 1c) were investigated. For path 2, preliminary investigations revealed that no relevant protein conformational rearrangements were required, and that slight modifications of xyz components of the pulling direction provided very similar force profiles (see Ap). Hence, we assumed that the chosen pulling direction in path 2 was likewise representative of the sword-unsheathed-like mechanism.

In SMD, the pulling velocity ($v$) largely influences both the results of simulations and the profile of the applied forces. As $v$ decreases, the pulled atoms have more time to sample the conformational space and to search for lowest resistance path along the selected direction. At the same time, the non pulled atoms can relax following the movement of the pulled atoms, and thus reducing the friction rate of the process. As shown in Figure 3b, a lowering of the pulling velocity resulted in a decrease of the observed rupture force likely because of the reduction of non-equilibrium effects.[29,33] Remarkably, at the pulling velocity of 1.5e-06 Å/fs, a plateau was observed, pointing to this pulling rate as the more appropriate one to minimize the friction influence on the delivery process. Moreover, at such a slow pulling velocity, some spontaneous and SMD-unrelated conformational changes of ACP could also be sampled.

In Figure 3d and 3c, the force profile and the mechanical work done during each simulated pathway are shown, respectively. For each pathway, the force profile correlated well with the rupture and formation of interaction between the outgoing substrate and the binding site residues; a detailed description of the events is reported in Ap.

19

The system in both path 1a and 1b encountered relatively high hindrance in the early ns of simulations because the substrate moved against the deep portion of the cleft formed by α2- and α3-helix, and lined by the α2-α3 loop. Here, the α3-helix could not spontaneously follow the pulled atoms and as a consequence the exerted force increased.

In path 1c (Figures 3 [blue line] and 4), the force profile was much flatter than in paths 1a and 1b. We observed an initial force peak (~200 pN, Figure 3d) due to the breaking of an H-bond network involving the substrate β-OH and the backbone oxygen of A60 and I63. Then, the whole process evolved with a force less than 100 pN. A strong electrostatic interaction involving the β-OH group and the side chain of D59 occurred between ~1.5-2.5 ns (Figure 4c). Concertedly, a slow "gate opening" was manifested by the increasing distance between V41 and A60 (see also Figure 1b). The flatness of the applied force profile indicated that both the pulled substrate and the constraint point were concertedly moving, and the decreasing of the overall vdW contacts meant that the substrate was leaving the binding cavity quite slowly. The lack of a significant rupture point in the force profile suggested that the substrate did not encounter any relevant resistance despite the remarkable shift of the α3-helix. This was possible because, during the substrate pulling, the ACP α3-helix could spontaneously rearrange, allowing the substrate exposure. This observation was in good agreement with both free MD and experimental data, [18-26] which pointed to α3-helix as one of the most dynamic segments of ACP structure (Figure 2a). Noteworthy, in the same simulation time scale, in the free MD, it was sampled both the helix folding of residues 19-26 and the partial unfolding of the short α3-helix (see Ap).
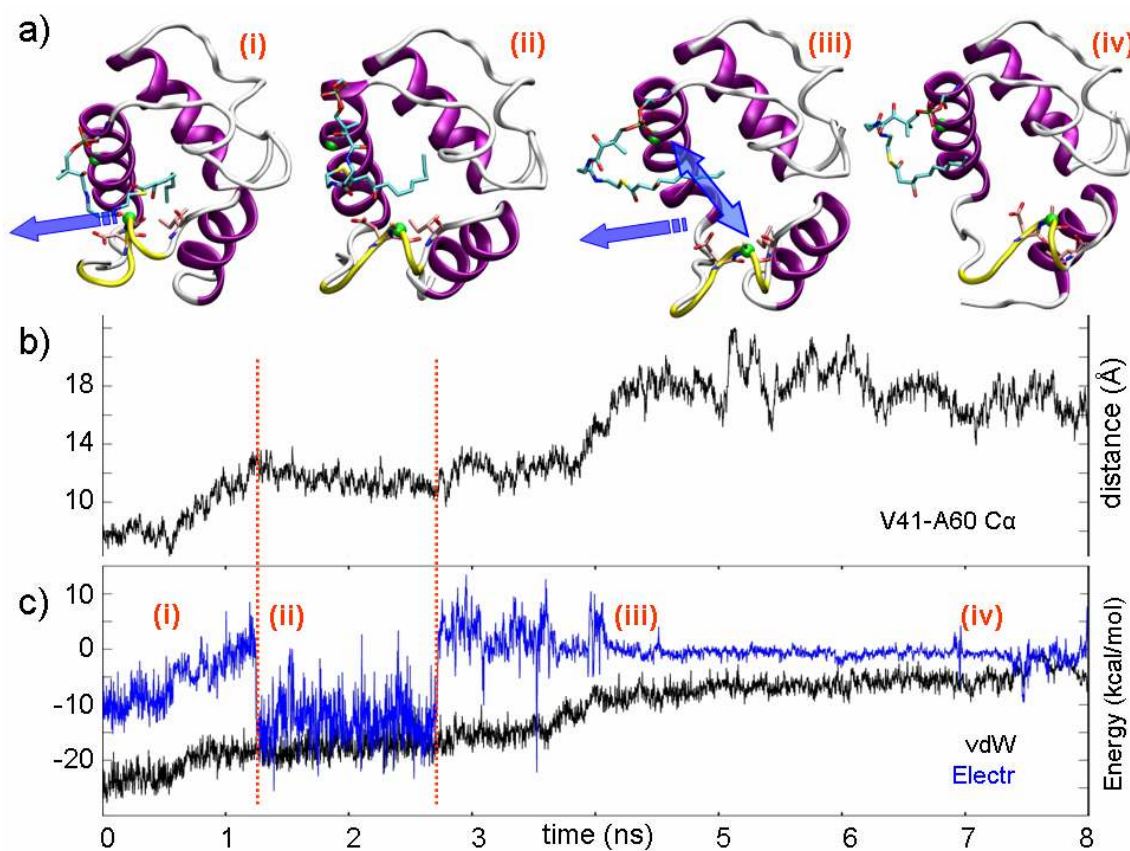
**Figure 4.** a) Snapshots isolated from the SMD trajectory sampled in Path 1c at (i) 300 ps, (ii) 1.6 ns, (iii) 4.2 ns, and (iv) 7 ns. The blue arrow in (i) shows the pulling direction and in (iii), the double end arrow indicates the opening movement of the fissure formed by helices α2 and α3. b) Evolution of the distance between Cα of V41 and A60. c) Electrostatic and vdW interaction energies between the β-hydroxydecanoyl moiety and the protein residues during the SMD.

Path 1 delivery hypothesis was then compared to path 2, to investigate if a sword-unsheathed-like mechanism could provide a flatter force profile. As shown in Figure 3d (magenta line), the unthreading of the substrate had a magnitude of forces similar to that registered for path 1b and higher than that observed in path 1c. Such a mechanism poorly complied with an ACP-enzyme interface comprising both α2-helix and its connection to α4-helix. However, a sword-unsheathed-like mechanism could be proposed for enzymes, whose interactions with ACP more extensively involve the α1-α2 loop connection.[16,20]

In summary, our simulations suggest that the substrate delivery through the cleft between α2- and α3-helix is a feasible pathway that exploits intrinsic conformational plasticity of ACP (path 1c). We also show that ACP portion including the loop connecting helices α2 and α3 and the α3-helix itself can play a critical role in both the delivery and, as previously reported, the recognition processes.[18-26] As hypothesized on the basis of X-ray experiments,[12,15] we also point out the ability of the prosthetic group to function as label for the acyl-intermediates carried by the ACP. Such aspects are related to ACP plasticity and may account for a mechanism in which ACP and its enzymatic counterparts minimize the solvent exposure of the lipophilic substrate moiety. The delivery model would include an ACP/enzyme interface formed by ACP helices α2 and α3; it might be hypothesized that a slight conformational change of the α3-helix could easily allow the substrate to glide into the corresponding enzyme active site.

Can we use the above uncovered features of ACP to address the development of innovative antimalarial compounds? This study sheds light on the functional and plastic features of ACP and attempts to contextualize them in the protein-protein interactions network occurring in the FAS-II system. During the elongation cycle, ACP-enzyme interactions and the delivery of the acyl substrate to the active site are compulsory steps for fatty acid production. These processes are currently subject of drug discovery efforts. In this respect, blockers of substrate delivery would function as pathway-based antimalarial compounds rather than single FAS-II enzyme inhibitors. The design of a delivery-blocker might be achieved through a strategy that considers as template-structure for a lead-candidate that region of the α3-helix usually interacting with α2. The interaction of such a mimetic compounds with *Pf*ACP would interfere with the gate opening function managed by the α2/α3 cleft. Moreover, the low-resistance delivery pathways here investigated might be used as starting point for further studies aiming at identifying local energy

minima along the reaction coordinate. As matter of fact, our simulations suggest that, while the substrate transiently leaves the carrier core moving toward the catalytic partner, the ACP core itself might become a peculiar binding pocket for small molecules able to interfere with the relocation of the substrate into the carrier. Likewise, if the conformational "transient" state of the acyl-ACP is an highly populated energy minimum, the probability to block the delivery process would dramatically increase.

**Methodological details**

Starting from the first NMR conformer (out of 20) of the *holo* structure of *Pf*ACP (pdb code: 2fq0), the prosthetic group carrying a β-hydroxydecanoyl substrate was built using the Sybyl 7.3 molecular modeling suite of program (Tripos Inc., St. Louis, MO) and manually docked at the acyl binding pocket of the carrier using as template the binding solution of the high resolution X-ray structure of E. coli decanoyl-ACP (pdb code: 2fae). The covalently bound β-hydroxydecanoyl-ACP complex was first energy-minimized in gas phase for 1000 steps using the conjugated gradient method keeping fixed the backbone atoms and restraining the side chain with a 10 kcal mol$^{-1}$ Å$^{-2}$ spring constant. The complex was then solvated with a 10 Å thick layer of water using the solvate package of VMD and seven Na$^{+}$ cations were added to neutralize the system. The simulations were performed using periodic boundary conditions and long-range electrostatics calculated by using the particle-mesh Ewald (PME) method with a charge grid spacing < 1Å. A cutoff of 10 Å was used for van der Waals and short-range electrostatic interactions with a switching function started at 8 Å to ensure a smooth cutoff. Time integration step of 2 fs was used and the length of all bonds involving hydrogen atoms was fixed using the SHAKE algorithm. The solvated system was minimized with the conjugate gradient method for 1000 steps restraining the heavy atoms of the protein with a force constant of 10 kcal mol$^{-1}$ Å$^{-2}$, followed by 1000 steps with a force constant of 3 kcal mol$^{-1}$ Å$^{-2}$ and finally fully minimizing the system for further 1000 steps. The insertion of the β-hydroxydecanoyl moiety in the *Pf*ACP core produced only slightly reorientations in the side chains of residues shaping the cavity, with the minimized structure having an average overall rmsd less than 0.5 Å. Six subsequent steps were

used to heat the system from 1 to 300 K. Every 15 ps the system was heated up by 50 K and correspondingly the alpha-carbon were gradually unrestrained (till 2 kcal mol$^{-1}$ Å$^{-2}$) starting from a 6 kcal mol$^{-1}$ Å$^{-2}$ harmonic restraint force constant. In these equilibration steps, constant volume was maintained and the temperature was controlled by Langevin dynamics with a dumping coefficient of 5 ps$^{-1}$. The system was subsequently switched to the isothermal-isobaric (NPT) ensemble once the temperature was stabilized at 300 K. During the switching, a soft harmonic restraint of 2 kcal mol$^{-1}$ Å$^{-2}$ was still applied to the C$\alpha$ atoms and gradually turned off in the next 60 ps. The constant pressure control was applied using the Nosé-Hoover Langevin piston method and 1 atm was set as target pressure. The system evolved for further 100 ps and the final state was used as starting point for both SMD study and free MD production.

All simulations were performed with the molecular dynamics program NAMD 2.6 (Phillips et al., 2005) using the CHARMM22 force field (MacKerell et al., 1998) for proteins and the TIP3P model for all water in the system (Jorgensen et al., 1983). CHARMM compliant parameters for prosthetic group together with the β-hydroxydecanoyl substrate were generated using the paratool plugin implemented in VMD (Humphrey et al., 1996. http://www.ks.uiuc.edu/Research/vmd/). Missing parameters were estimated from similar terms within the force field using an empirical/additive approach. Mulliken charges were calculated at the HF/6-31G* level of theory and refined by analogy in order to preserve consistency with CHARMM style charges (i.e., always $q_H = 0.09$).

SMD simulations were performed at constant pulling velocity using a steering velocity of 1.5e-06 Å/fs with a spring constant of 5 kcal mol$^{-1}$ Å$^{-2}$ (several velocities were further considered as discussed later in the Appendix). Variation of the constant $K$ of the harmonic restrain influences the profile/fluctuation of the applied forces. Here, we used a value of 5 kcal mol$^{-1}$ Å$^{-2}$ that coupled with an appropriate $v$ gave a force profile in which it was still possible to notice a drop of the forces, but at the same time permitted the thermal fluctuation of the pulled atoms to be similar to the perturbation arising from the pulling force.

Using as reference coordinates the structure of the E. coli decanoyl-ACP (pdb code: 2fae subunit A) the $x$, $y$ and $z$- components of the normalized pulling direction were: for path 1a, -0.9388, 0.2181, 0.2665; for path 1b, -0.9198, 0.2927, 0.2611; for path 1c, -0.8512, 0.4658, 0.2416; for path 2, 0.0326, 0.9300, 0.3640. To avoid shifting of the system during pulling, in path 1, the alpha-carbons of protein going from residue 1 to 35 were harmonically restrained to their initial

position using a 1 kcal mol$^{-1}$ Å$^{-2}$ spring constant and, in path 2 the restraints were applied to residue portions 1-16 and 48-52.

The value of the exerted force ($F$) were outputted every ($dt$) 1 ps of simulation and the work $W(t) = \int_0^{t'} F(t')v dt'$ done on the system during the SMD was calculated by numerical integration; the pulling velocity ($v$) was 1.5e-06 Å/fs.

CHARMM force field topology for prosthetic group and substrate can be found at http://pubs.acs.org/doi/suppl/10.1021/ci800297v/suppl_file/ci800297v_si_001.pdf

## Appendix

### Pulling velocity and force profiles

To explore the influence of the pulling velocity on the behavior of the system we tested four decreasing $v$ values starting from 1.5e-03 Å/fs (0.003 Å/timestep) leading to 1.5e-06 Å/fs (0.000003 Å/timestep). For the higher velocity used it was possible to correlate the RMSD of the protein with the magnitude of the rupture force, calculated as mean of three individual simulation for each path (Figure 1Ap-a), required for the substrate exposure along each investigated pathway. As shown, starting from the most "perturbative" pulling direction path 1a, the protein was subjected to lower conformational rearrangement going toward the path 2 direction (Figure 1Ap-b). The reported RMSD of the whole protein actually reflected the deviation whose residues 57 to 62 were subject (Figure 1Ap-c). This means that at higher velocities path 1a pushed the substrate in the deep portion of the cleft formed by α2 and α3 helices and lined by their α2-α3 loop connection. The protein could only suffer this rustling having not the possibility to give into the pushing atoms. As the pulling direction became less drastic as it was in path 1b and 1c, the substrate could find less opposition to its flowing, leading to a decrease of the exerted force as well as of the RMSD. In path 2, the conformational shift of α3-helix was not remarkable and the substrate could be easily threaded away.

As the pulling velocity is lowered one should expect a decreasing trend of the rupture force because of the reduction of non-equilibrium effects. In Figure 1Ap-d this trend is shown, and a *plateau* was reached when $v$ approached to 1.5e-06 Å/fs. Around this velocity the necessary forces to extract the β-hydroxyacyl substrate became more similar among the chosen paths and

an in-depth inspection of the system was required. Remarkably, at the lower pulling velocity used, the substrate took about 6-8 ns to be extracted from the binding cavity. This was the same simulation time scale we used in the free MD, where it was possible to sample both the helix folding of residues 19-26, and the partial unfolding of the short α3-helix (see Figure 9Ap-d). That means that, meanwhile the pulling force was applied to the substrate atoms, the ACP system was able to undergo some intrinsic conformational changes normally occurring without any constraint.
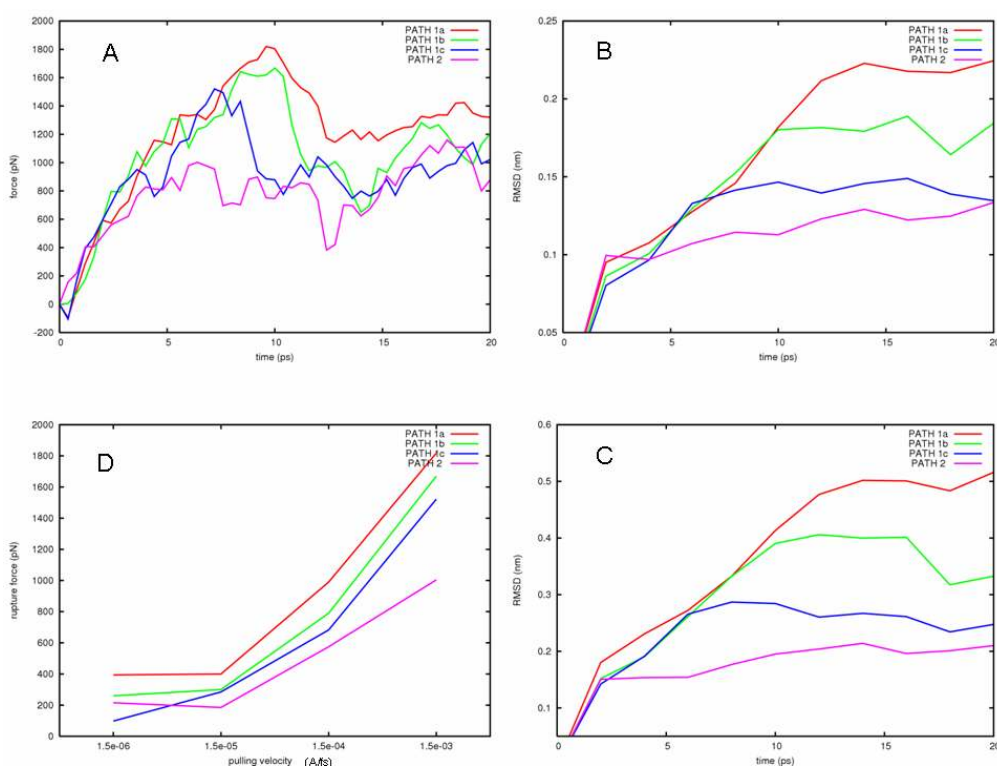


**Figure 1Ap**. Effects of the pulling velocity on the force profiles. A) Force profile obtained for the higher velocity (1.5e-03 Å/fs); calculated as mean of three individual simulation for each path. B) RMSD of the alpha-carbons of the whole protein during the steering process and C) RMSD of the protein portion 57-62. Path 1a induces the wider movement respect to the other paths. D) Computed rupture force as a function of several pulling velocities.
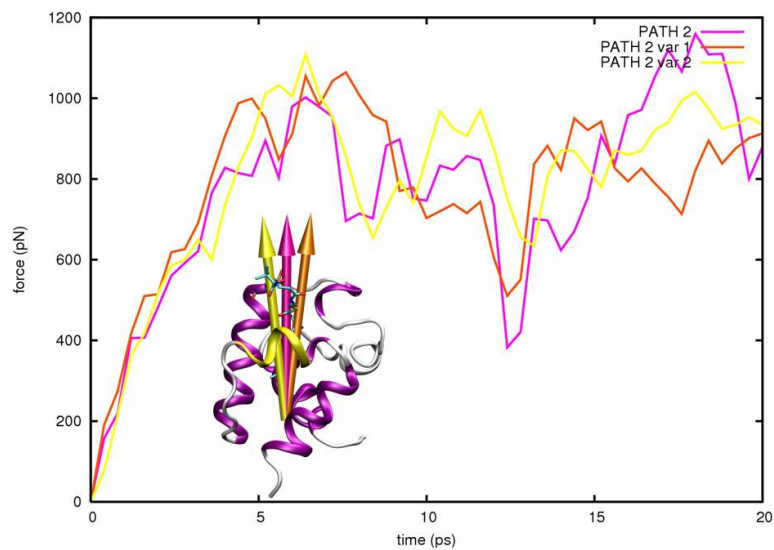
**Figure 2Ap**. Effects of modification of pulling direction on the force profiles in PATH 2. Force profiles obtained using a pulling velocity of 1.5e-03 Å/fs, calculated as mean of three individual simulation run.


## Further details on the pathways


In Figure 3Ap, we report on the force profiles of the four pathways (paths 1a-c and path 2).
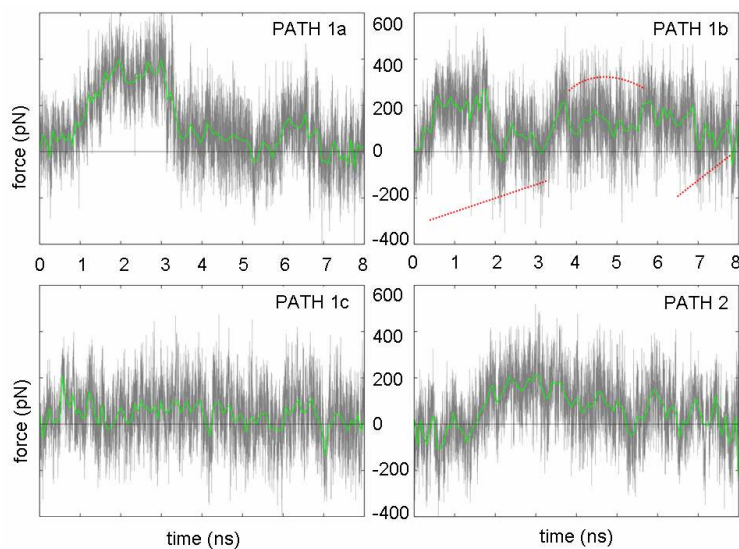


**Figure 3Ap**. Force profiles of pathways investigated at 1.5e-06 Å/fs. The green thick curve is obtained averaging the forces every 30 points.

**PATH 1A**. In Figure 4Ap, the path 1a is summarized in terms of force profile, V41-A60 distance between their alpha-carbons, and electrostatic and van der Waals interaction energy between ACP and the β-hydroxydecanoyl moiety of the substrate. The plateau observed at 1.5-3 ns (Figure 4Ap-a) was due to the breaking of H-bond interaction between the β-hydroxyl group of the substrate and A60 and I63 backbone. Then, the acyl chain followed the unfolding of α3-helix, being it still able to transiently establish an H-bond interaction with A60 and I63 (Figure 4Ap-d).
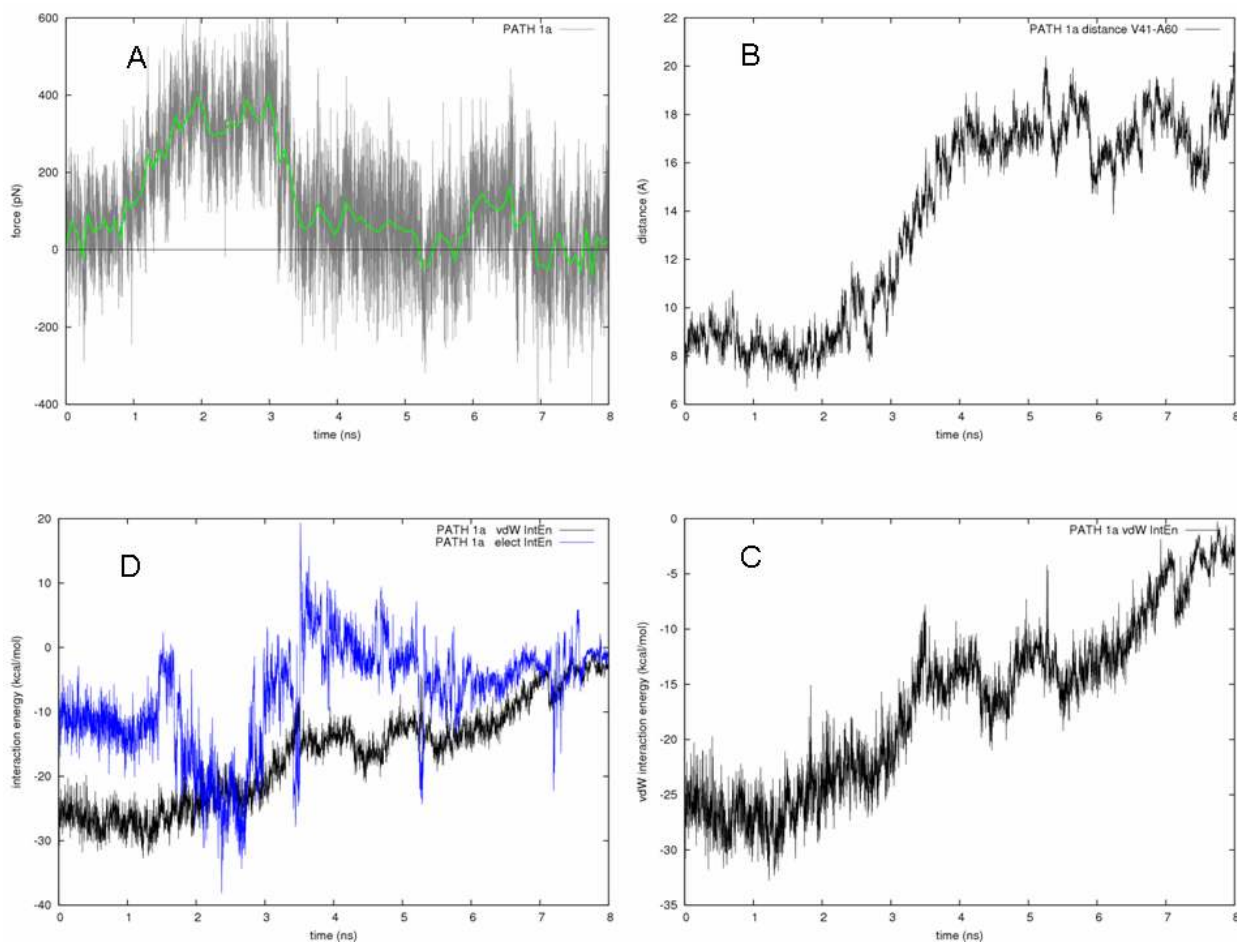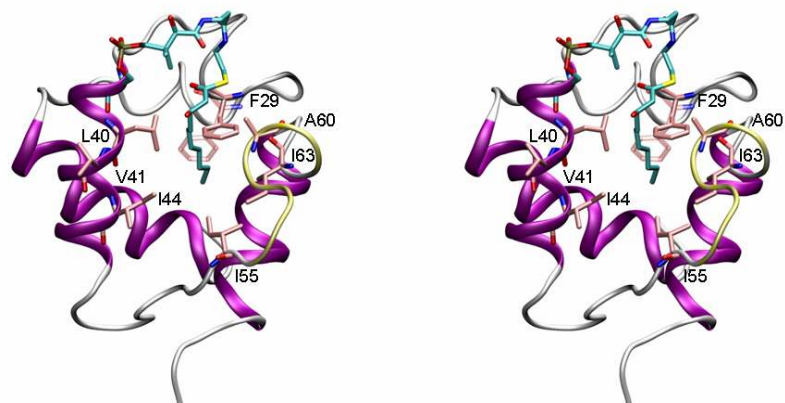


**Figure 4Ap**. PATH 1A. A) Force profile obtained at $v$ = 1.5e-06 Å/fs. B) Evolution of the distance between the alpha-carbons of V41 and A60 during SMD simulations. C) Electrostatic and van der Waals interaction energy between the β-hydroxydecanoyl moiety and the protein during SMD simulations. D) Magnification of the van der Waals interaction energy contribution.

**PATH 1B.** In Figure 5Ap, the path 1b is summarized. As shown in Figure 5Ap-b, an opening/closure of the α2-α3 fissure was observed. The process could be divided in three steps: approach, transit, and escape. During the early stage of the simulation the acyl end was trapped into a sub-pocket toward the α1-helix and formed by F29, L47, L44, and I8; this explained the force increasing despite both electrostatic and vdW interactions remained constant. Around 1.6 ns the pulling force released the acyl group from the sub-pocket and slowly it moved towards the fissure formed by the juxtaposition of α2 and α3 helices. In the simulation time-range from ~3.7-6.5 ns, the substrate protruded into the fissure formed by I41, I44, I55 and A60, and by L40 and I63 (Figure 6Ap). Furthermore, the side chain of F29 underwent a conformational change; F29 acted as lid maintaining the cavity core water-free. Finally, in the third step, the substrate completely left the binding cavity as suggested by both the decreasing of interaction energies and forces magnitude.
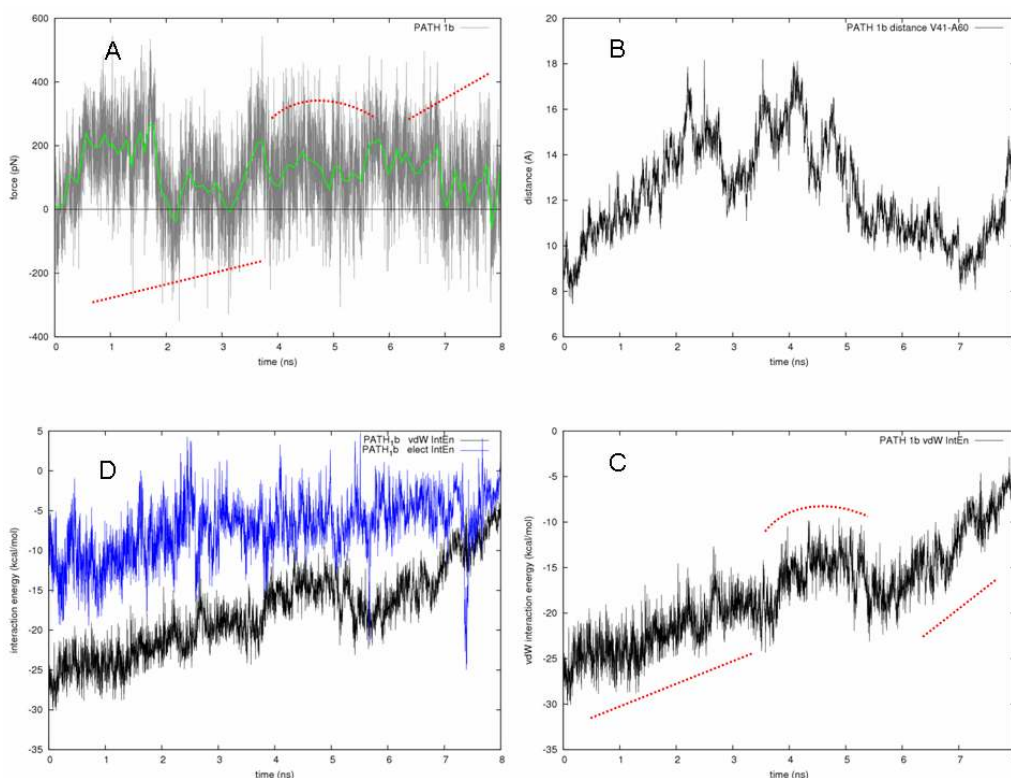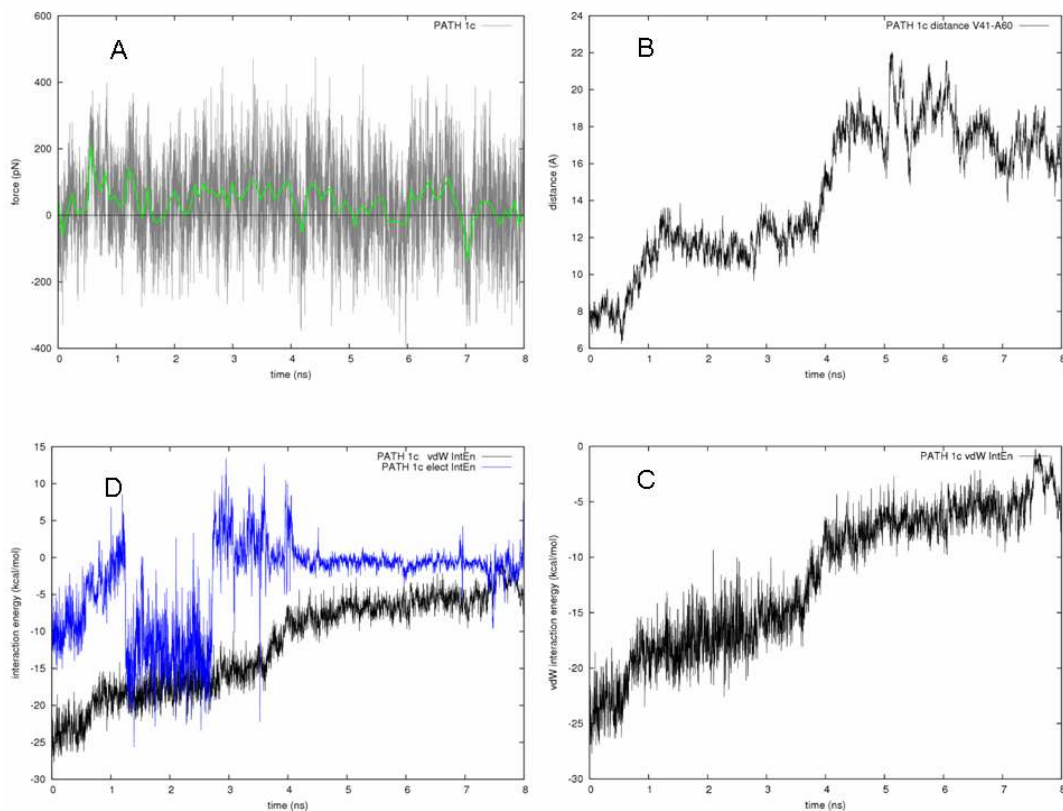


**Figure 5Ap**. PATH 1B. A) Force profile obtained at $v$ = 1.5e-06 Å/fs. B) Evolution of the distance between the alpha-carbons of V41 and A60 during SMD simulations. C) Electrostatic and van der Waals interaction energy between the β-hydroxydecanoyl moiety and the protein during SMD simulations. D) Magnification of the van der Waals interaction energy contribution.

**Figure 6Ap**. Stereo view of the transient state sampled in Path 1b (see text). The conformation of F29 related to the fully embedded substrate is shown in transparent pink stick. Note the β-OH group of the substrate exposed to putative metabolizing enzyme activity.

**PATH 1C.** In Figure 7Ap, the path 1c is summarized. This pathway is in-depth described in the paper.



**Figure 7Ap**. PATH 1C. A) Force profile obtained at $v$ = 1.5e-06 Å/fs. B) Evolution of the distance between the alpha-carbons of V41 and A60 during SMD simulations. C) Electrostatic and van der Waals interaction energy between the β-hydroxydecanoyl moiety and the protein during SMD simulations. D) Magnification of the van der Waals interaction energy contribution.

**PATH 2**. The unthreading of the substrate is summarized in Figure 8Ap. At the beginning of the simulation the force average fluctuated to both positive and negative values reflecting the floating behavior of the substrate. The β-OH of the substrate was found to interact with the backbone carbonyl group of I63 (Figure 8Ap-d). As the pulling evolved, another H-bond was formed with A60 backbone and the exerted forces increased till about 3.8 ns (Figure 8Ap-d). At 4 ns the acyl moiety started to leave the ACP core as revealed by the decreasing vdW contacts. The outgoing of the substrate was related to a decreasing of the V41-A60 distance reflecting the core contraction as the substrate left its location. During the outgoing process the prosthetic group was, finally, harbored in the mouth of α2-α3 helix cleft.
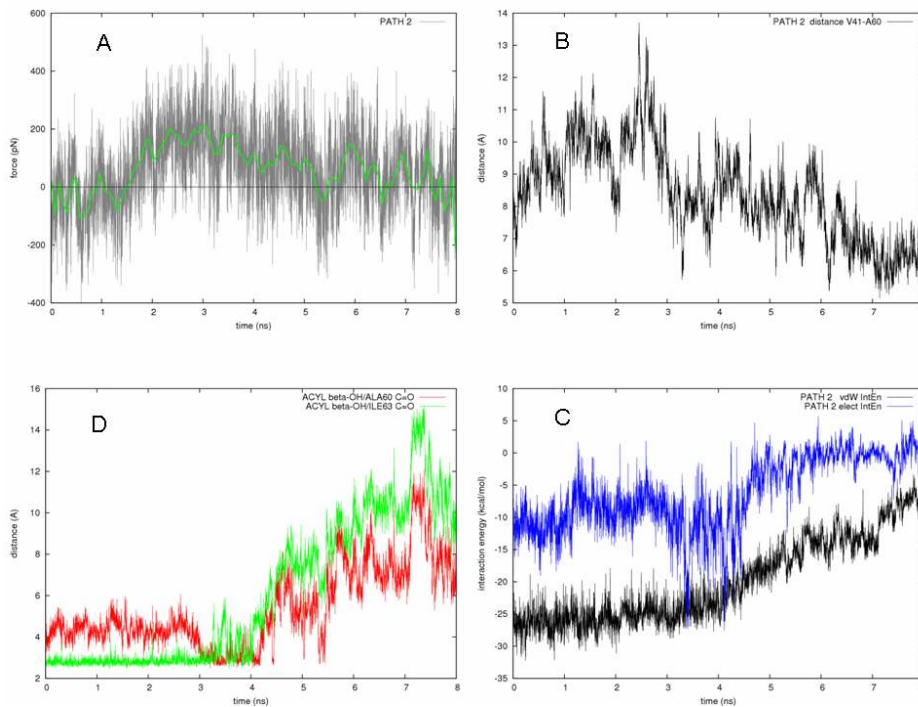


**Figure 8Ap**. Path 2. A) Force profile obtained at $v$ = 1.5e-06 Å/fs. B) Evolution of the distance between the alpha-carbons of V41 and A60 during SMD simulations. C) Electrostatic and van der Waals interaction energy between the β-hydroxydecanoyl moiety and the protein during SMD simulations. D) Monitored distances between H-bonding moieties.

**Monitored features of the free MD simulation**
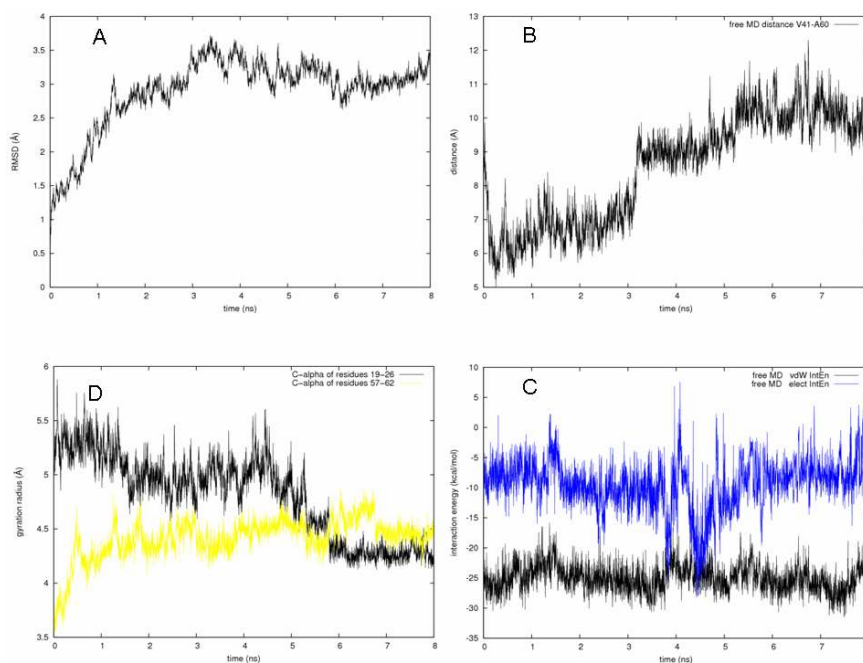


**Figure 9Ap**. Monitored features of the free MD. A) Time dependence of the RMSD calculated for the Cα  B) Evolution of distance between the alpha-carbons of V41 and A60 during the molecular dynamics simulation. C) Electrostatic and van der Waals interaction energy between the β-hydroxydecanoyl moiety and the protein during the simulation. D) Time dependence of the radius of gyration calculated for the loop portion 19-26 that completed the α helix folding around the 6th ns of simulation. Vice versa, the small α3 helix partially unfolded into a loop at the beginning of the simulation.

### References

(1)     Breman, J. G. The ears of the hippopotamus: manifestations, determinants, and estimates of the malaria burden. *Am. J. Trop. Med. Hyg.* **2001**, *64*, 1-11.

(2)     Greenwood, B. M.; Bojang, K.; Whitty, C. J.; Targett, G. A. Malaria. *Lancet* **2005**, *365*, 1487-1498.

(3)     Snow, R. W.; Guerra, C. A.; Noor, A. M.; Myint, H. Y.; Hay, S. I. The global distribution of clinical episodes of Plasmodium falciparum malaria. *Nature* **2005**, *434*, 214-217.

(4)     Surolia, A.; Ramya, T. N.; Ramya, V.; Surolia, N. 'FAS't inhibition of malaria. *Biochem. J.* **2004**, *383*, 401-412.

(5)     Lu, Y. J.; Zhang, Y. M.; Rock, C. O. Product diversity and regulation of type II fatty acid synthases. *Biochem. Cell Biol.* **2004**, *82*, 145-155.

(6)     White, S. W.; Zheng, J.; Zhang, Y. M.; Rock, C. O. The structural biology of type II fatty acid biosynthesis. *Annu. Rev. Biochem.* **2005**, *74*, 791-831.

(7)     Shen, B.; Summers, R. G.; Gramajo, H.; Bibb, M. J.; Hutchinson, C. R. Purification and characterization of the acyl carrier protein of the Streptomyces glaucescens tetracenomycin C polyketide synthase. *J. Bacteriol.* **1992**, *174*, 3818-3821.

(8)     Brozek, K. A.; Carlson, R. W.; Raetz, C. R. A special acyl carrier protein for transferring long hydroxylated fatty acids to lipid A in Rhizobium. *J. Biol. Chem.* **1996**, *271*, 32126-32136.

(9)     Sweet, C. R.; Williams, A. H.; Karbarz, M. J.; Werts, C.; Kalb, S. R. et al. Enzymatic synthesis of lipid A molecules with four amide-linked acyl chains. LpxA acyltransferases selective for an analog of UDP-N-acetylglucosamine in which an amine replaces the 3"-hydroxyl group. *J. Biol. Chem.* **2004**, *279*, 25411-25419.

(10)    Kim, Y.; Prestegard, J. H. Refinement of the NMR structures for acyl carrier protein with scalar coupling data. *Proteins* **1990**, *8*, 377-385.

(11)    Xu, G. Y.; Tam, A.; Lin, L.; Hixon, J.; Fritz, C. C. et al. Solution structure of B. subtilis acyl carrier protein. *Structure* **2001**, *9*, 277-287.

(12)    Roujeinikova, A.; Baldock, C.; Simon, W. J.; Gilroy, J.; Baker, P. J. et al. X-ray crystallographic studies on butyryl-ACP reveal flexibility of the structure around a putative acyl chain binding site. *Structure* **2002**, *10*, 825-835.

(13)    Wong, H. C.; Liu, G.; Zhang, Y. M.; Rock, C. O.; Zheng, J. The solution structure of acyl carrier protein from Mycobacterium tuberculosis. *J. Biol. Chem.* **2002**, *277*, 15874-15880.

(14)    Sharma, A. K.; Sharma, S. K.; Surolia, A.; Surolia, N.; Sarma, S. P. Solution structures of conformationally equilibrium forms of holo-acyl carrier protein (PfACP) from Plasmodium falciparum provides insight into the mechanism of activation of ACPs. *Biochemistry* **2006**, *45*, 6904-6916.

(15)    Roujeinikova, A.; Simon, W. J.; Gilroy, J.; Rice, D. W.; Rafferty, J. B. et al. Structural studies of fatty acyl-(acyl carrier protein) thioesters reveal a hydrophobic binding cavity that can expand to fit longer substrates. *J. Mol. Biol.* **2007**, *365*, 135-145.

(16)    Kim, Y.; Kovrigin, E. L.; Eletr, Z. NMR studies of Escherichia coli acyl carrier protein: dynamic and structural differences of the apo- and holo-forms. *Biochem. Biophys. Res. Commun.* **2006**, *341*, 776-783.

(17)    Zhang, Y. M.; Marrakchi, H.; White, S. W.; Rock, C. O. The application of computational methods to explore the diversity and structure of bacterial fatty acid synthase. *J. Lipid Res.* **2003**, *44*, 1-10.

(18)    Parris, K. D.; Lin, L.; Tam, A.; Mathew, R.; Hixon, J. et al. Crystal structures of substrate binding to Bacillus subtilis holo-(acyl carrier protein) synthase reveal a novel trimeric arrangement of molecules resulting in three active sites. *Structure* **2000**, *8*, 883-895.

(19)     Flaman, A. S.; Chen, J. M.; Van Iderstine, S. C.; Byers, D. M. Site-directed mutagenesis of acyl carrier protein (ACP) reveals amino acid residues involved in ACP structure and acyl-ACP synthetase activity. *J. Biol. Chem.* **2001**, *276*, 35934-35939.

(20)     Zhang, Y. M.; Rao, M. S.; Heath, R. J.; Price, A. C.; Olson, A. J. et al. Identification and analysis of the acyl carrier protein (ACP) docking site on beta-ketoacyl-ACP synthase III. *J. Biol. Chem.* **2001**, *276*, 8231-8238.

(21)     Worsham, L. M.; Earls, L.; Jolly, C.; Langston, K. G.; Trent, M. S. et al. Amino acid residues of Escherichia coli acyl carrier protein involved in heterologous protein interactions. *Biochemistry* **2003**, *42*, 167-176.

(22)     Gong, H.; Byers, D. M. Glutamate-41 of Vibrio harveyi acyl carrier protein is essential for fatty acid synthase but not acyl-ACP synthetase activity. *Biochem. Biophys. Res. Commun.* **2003**, *302*, 35-40.

(23)     Keatinge-Clay, A. T.; Shelat, A. A.; Savage, D. F.; Tsai, S. C.; Miercke, L. J. et al. Catalysis, specificity, and ACP docking site of Streptomyces coelicolor malonyl-CoA:ACP transacylase. *Structure* **2003**, *11*, 147-154.

(24)     Zhang, Y. M.; Wu, B.; Zheng, J.; Rock, C. O. Key residues responsible for acyl carrier protein and beta-ketoacyl-acyl carrier protein reductase (FabG) interaction. *J. Biol. Chem.* **2003**, *278*, 52935-52943.

(25)     Zhang, L.; Liu, W.; Xiao, J.; Hu, T.; Chen, J. et al. Malonyl-CoA: acyl carrier protein transacylase from Helicobacter pylori: Crystal structure and its interaction with acyl carrier protein. *Protein Sci.* **2007**, *16*, 1184-1192.

(26)     Liu, W.; Du, L.; Zhang, L.; Chen, J.; Shen, X. et al. Helicobacter pylori acyl carrier protein: expression, purification, and its interaction with beta-hydroxyacyl-ACP dehydratase. *Protein Expr. Purif.* **2007**, *52*, 74-81.

(27)     Leibundgut, M.; Jenni, S.; Frick, C.; Ban, N. Structural basis for substrate delivery by acyl carrier protein in the yeast fatty acid synthase. *Science* **2007**, *316*, 288-290.

(28)     Rafi, S.; Novichenok, P.; Kolappan, S.; Zhang, X.; Stratton, C. F. et al. Structure of acyl carrier protein bound to FabI, the FASII enoyl reductase from Escherichia coli. *J. Biol. Chem.* **2006**, *281*, 39285-39293.

(29)     Izrailev, S.; Stepaniants, B. I.; Isralewitz, B.; Kosztin, D.; Lu, H. et al. Steered molecular dynamics. In *Computational Molecular Dynamics: Challenges, Methods, Ideas*; Springer Ed., 1999; pp 39-64.

(30)     Sotomayor, M.; Schulten, K. Single-molecule experiments in vitro and in silico. *Science* **2007**, *316*, 1144-1148.

(31)     MacKerell Jr., A. D., Bashford, D., Bellott, M., Dunbrack Jr., R.L., Evanseck, J., et al. All-atom empirical potential for molecular modeling and dynamics studies of protein. *J. Phys. Chem. B* **1998**, *102*, 3586 -3616.

(32)     Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E. et al. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781-1802.

(33)     Niu, C.; Xu, Y.; Luo, X.; Duan, W.; Silman, I. et al. Dynamic mechanism of E2020 binding to acetylcholinesterase: a steered molecular dynamics simulation. *J. Phys. Chem. B* **2005**, *109*, 23730-23738.

# 3. ACP/FabZ interaction

The role of ACP's α3-helix in the delivery of substrate is discussed in the context of the protein-protein interaction between *Pf*ACP and *Pf*FabZ. Several experimental observations reporting on the ACP residues involved in interaction with enzymatic counterparts were taken into account and the frame of the recognition and delivery process is also outlined.

Whereas the α2-helix of ACP is thought to play the major role in ACP-enzyme interactions,[1-4] other ACP portions are likely involved in the protein-protein binding process. Residues going from 57 to 62 (*Pf*ACP numeration) and forming the short α3-helix do not have a conserved secondary structure among the available ACPs. NMR investigations on *Pf*ACP have reported the α3-helix to experience a helix-loop equilibrium which in turn formed a longer loop connection between α2 and α4 helices. The high mobility of the ACP portion including residues 53 to 62 was sampled in simulations[5] and is consistent with structural data discussed in the previous chapter. The putative physiological role of the portion 53-62 was hypothesized in virtue of its proximity to the recognition α2-helix and considering its contribution in forming the boundary of the cavity pocket hosting the acyl substrate. Moreover, several experimental and computational studies aimed at the identification of the residues involved in the interaction of ACP with enzymatic counterparts identify this portion, together with α2-helix, as being part of the interacting ACP interface. Chemical shift perturbation protein NMR studies of the complex ACP/FabA in *Escherichia coli* have indicated residues E53 and E60 as mediator of the interaction.[6] In the X-ray crystal structure of *Bacillus subtilis* ACP-AcpS complex,[2] residues I54, D56 and E60 of ACP are within an interacting distance from the AcpS. Zhang et al.,[4] described the "outsider" mediation of I54 in the ACP-FabG interaction and hypothesized a role for the α2-

α3 loop in the conformational change associated with ACP binding. Using an environmentally sensitive fluoroprobe, the interacting residues of selectively labeled myristoil-ACP with HlyC were found to be located in the α1-α2 loop, α2 helix and α2-α4 loop I54, D56 and K61.[7] Figure 1 resume the above experimental observation.
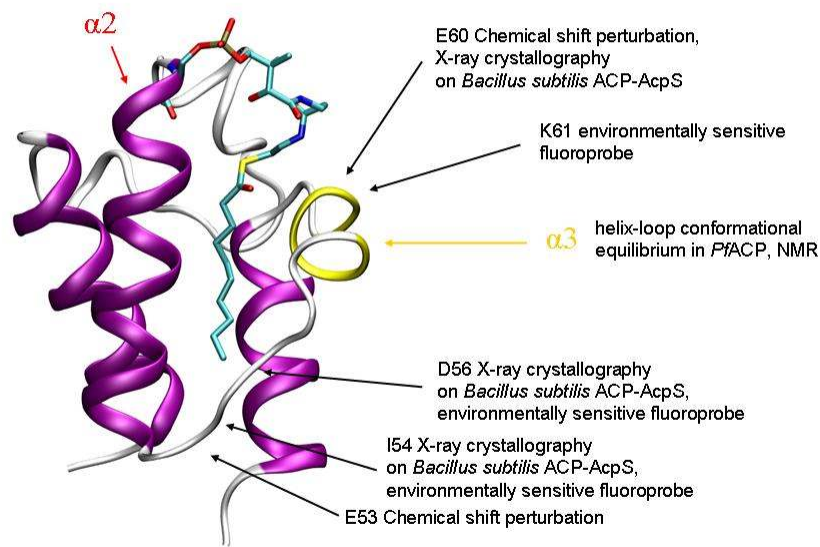


Figure 1. Crystal structure of E.coli decanoyl-ACP (pdb code: 2fae). The localizations of residues (black arrows) involved in the interactions with enzymatic counterparts are shown.

Typically consisting of 70-100 residues, ACP belongs to a broad family of conserved carrier protein. The acyl substrates are bound to ACP through a flexible arm formed by the serine-bound prosthetic phosphopantetheine (4′-PP) group. ACP multiple sequence alignment (Figure 2a) revealed a high degree of identity centered on the serine linking the 4′-PP group. Conserved residues highlight their evolutionary functional importance: Asp36, Ser37, Leu38, Glu42, and Glu48, I55, D57 and position 58 which is either occupied by Glu or Asp. The lipophilic residues

are mainly located in core region of the carrier while charged residues tend to be at the protein surface. Importantly, lipophilic amino acids line the core pocket of ACP and allow optimal hosting of the substrate acyl chain. However, hydrophobic patches are usually present at protein interfaces[8, 9] and in the specific case of FAS-II they may account for selectivity among different ACP-interacting enzymes. The negatively charged residues strongly influence the electrostatic properties of ACP surface. The negative electrostatic potential distribution is focused on both α2 and α3 helices, therefore increasing the capability of ACP to interact *via* its α2-α3 face with the positively charged residues lining the active site entrance of FAS-II enzymes.[3]

The functional role of the ACP portion including helices α2 and α3 is also suggested by an evolutionary conservation analysis of ACPs sequence (Figure 2d).[10-16] Enzymatic activity as well as protein-protein interactions are mediated by clusters of evolutionarily conserved residues which are spatially related to each other. Accounting for the phylogenetic relations between aligned protein and for the stochastic nature of evolution,[17] it is possible to assign a conservation score for each residue of a protein sequence/structure. Figure 3d shows how the most conserved regions (deep purple) of ACPs included the α2 and α3 helices suggesting their connection with the biological function that the carrier protein has to absolve.

Concordantly, enzymatic and *in vivo* experiments have illustrated how acyl-ACPs from different species can be interchangeably used as substrate by *E. coli* FAS-II system. Since charged and lipophilic residues are conserved on ACPs surface, then ACP/protein interactions are likely achieved through such a conserved set of electrostatic and/or hydrophobic contacts.
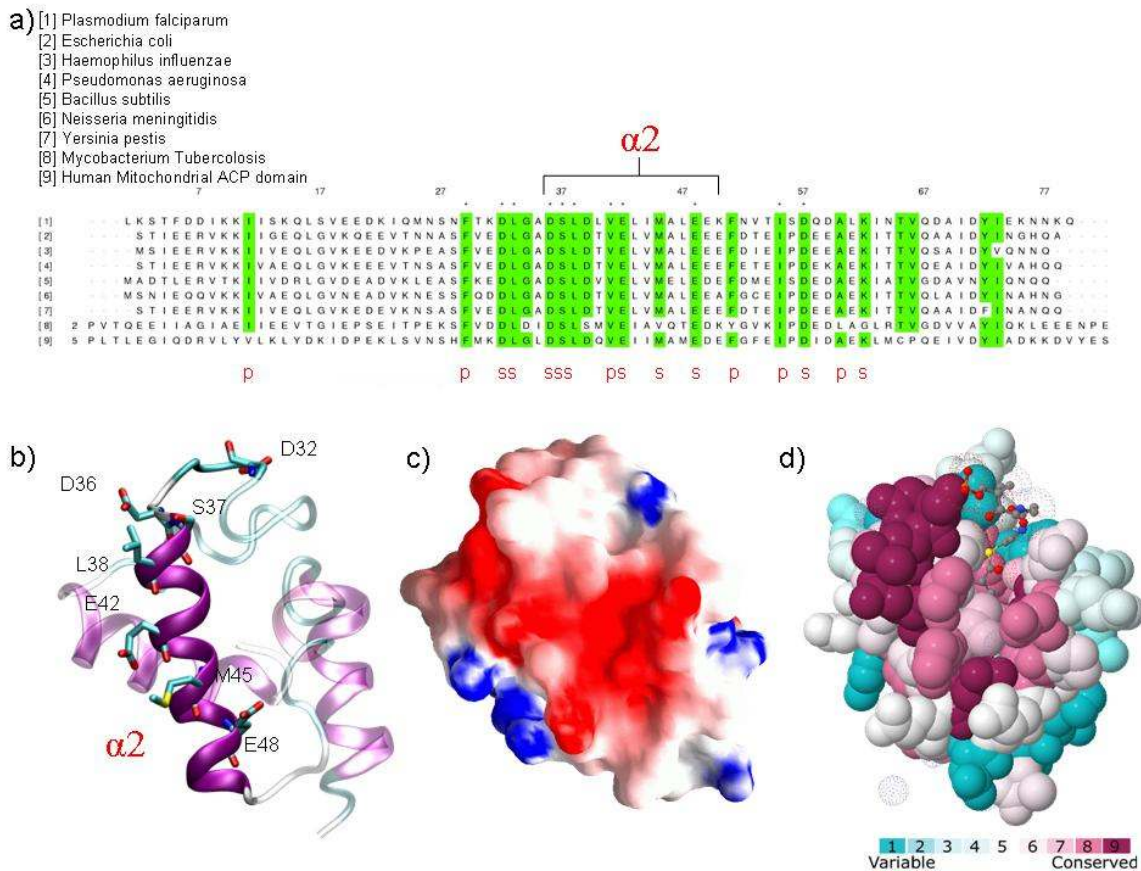
Figure 2. Acyl carrier protein from an evolutionary point of view. a) Sequence alignment of ACP homologues using the structural matrix implemented in BODIL.[18] b) 3D structure and c) electrostatic potential surface for the same view.[19] The range was from -8 (red) to 8 (blue) $k_bT/e$. d) amino acids are colored by their conservation level, with turquoise and purple indicating the most variable and conserved residues respectively.

Whereas the α2-helix has been widely postulated to represent a "recognition helix,[1, 3] the functional role of α3-helix in the delivery[5] as well as in the recognition and interaction of the enzymatic counterparts is here outlined for the first time.

In the ACP/FabZ interaction, the relevance of a negative charge distribution on ACP surface is confirmed by the observation that the binding tunnel of *Pf*FabZ is surrounded by basic residues such as R178, K180, K181 and K199 (Figure 3). Favorable long range Coulombic electrostatic

forces increase the rate of association rate of a protein complex,[20, 21] and the affinity of the ACP/enzyme complexes is thought to be strongly affected by electrostatic contributions.
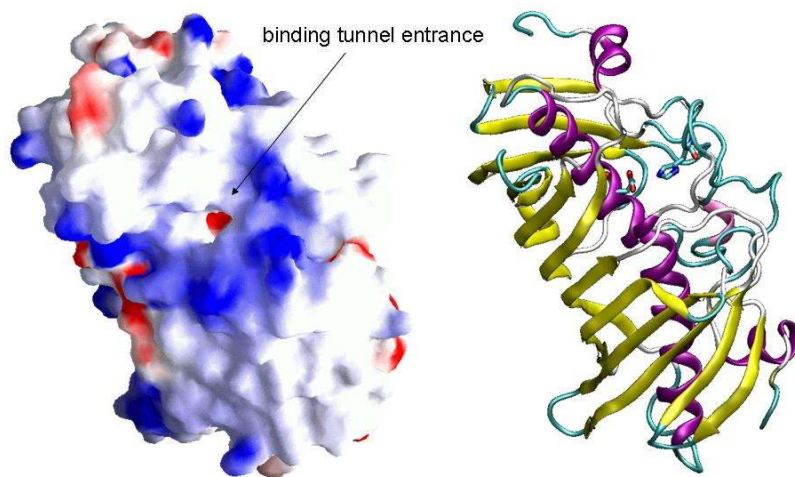


Figure 3. Electrostatic potential surface of PfFabZ. The contours were drawn at 8 $k_bT/e$ (with red negative and blue positive charge). In the cartoon model the catalytic residues Glu147 and His133 are shown.

We used a combination of docking algorithms to elucidate the putative ACP/FabZ mode of interaction. Typically, a protein-protein docking protocol is composed by two subsequent phases. In the first, called global search, the configurational space is sampled thoroughly using efficient posing algorithms combined with low resolution energy functions. In the second phase, or local search, the configurational sampling is limited to the surrounding of the top ranked outcomes generated by the global search and a more accurate scoring function is generally be used.

In figure 4 is shown a converged outcome of a global search docking of *Pf*ACP against *Pf*FabZ. Given two interacting counterparts, the algorithms aimed at finding docking transformations that yielded good molecular shape complementarity between the macromolecules.[22, 23] Several top ranked docking poses were further refined using a more detailed energy function and accounting for side chain flexibility.[24] However, only for the pose shown in Figure 4a it was observed the

typical energy funnel (Figure 4b) which is related to the presence of a stable energy minimum surrounded by a broad region of attraction.
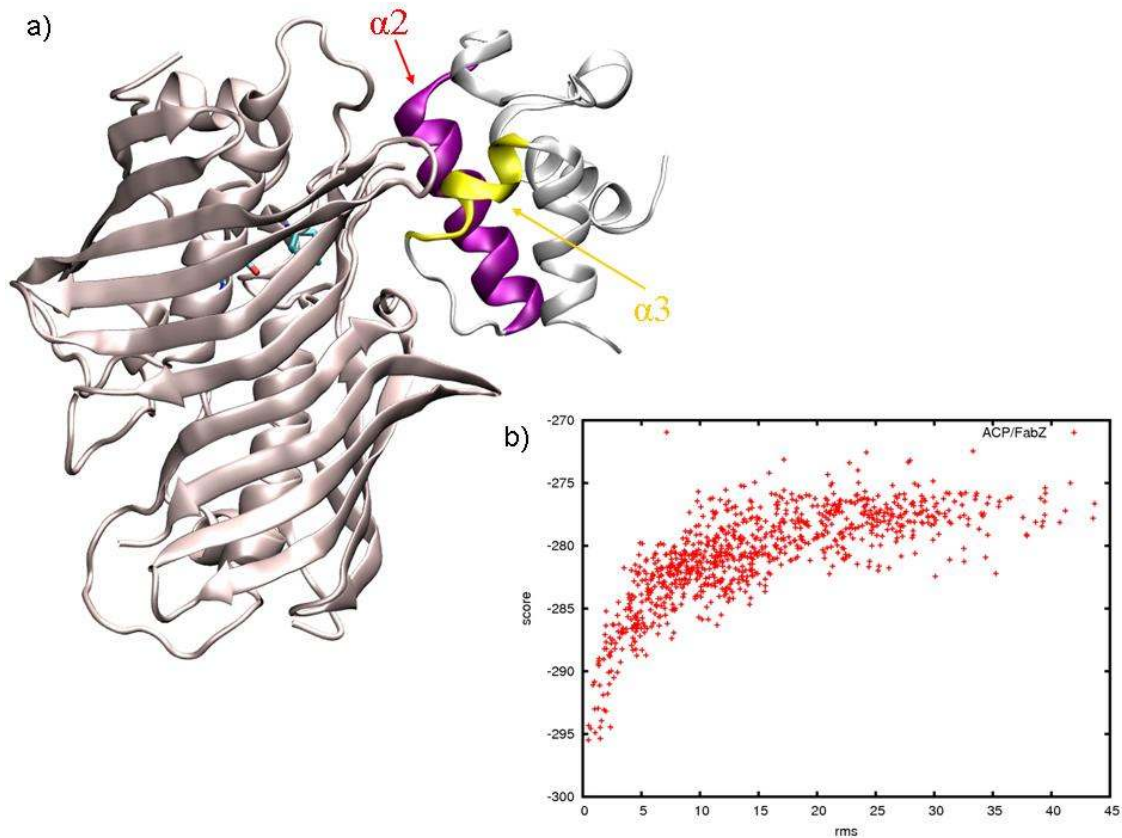


Figure 4. Interaction model for the PfACP/PfFabZ complex. a) Is shown the protein complex which generated b) the energy funnel. Protein-protein docking calculations were performed using the ZDOCK,[22] PatchDock[23] and *RosettaDock*[24] servers.

The binding mode is consistent with a delivery mechanism in which the α3-helix is the gatekeeper of the process. The ACP/FabZ interface is formed by ACP helices α2 and α3 and it might be hypothesized that a slight conformational change of the α3-helix could easily allow the substrate to move into the corresponding enzyme active site.

To gain further insight in the interaction process between the two macromolecules, preliminary molecular dynamics simulations were performed. Trajectories of the binding event were

generated in vacuum to reduce the simulation time and to enhance the attractive electrostatic contribution to the association of ACP and *Pf*FabZ. Starting from the refined binding model, ACP molecule was translated away from PfFabZ in the direction connecting their center of mass (c.o.m). In turn, ACP's c.o.m resulted about 15 Å farther its original docked position. Using a dielectric constant (ε) of 1 and 80, five independent short time (100ps) MD simulations were performed for each ε value with a cut off 20Å. Few alpha-carbons of FabZ were restrained to their original position so that only ACP was freely movable. One of the best ranked docking solutions, which not generated any energy funnel, was used as comparison term.

Because of the short simulation time used, the association of ACP and FabZ was sampled only for the lowest value of ε. The distance between the c.o.m was monitored along the simulations (Figure 5) and only the funneled binding mode yielded to associating trajectories. Interestingly, the trajectories converged into the original docking pose.
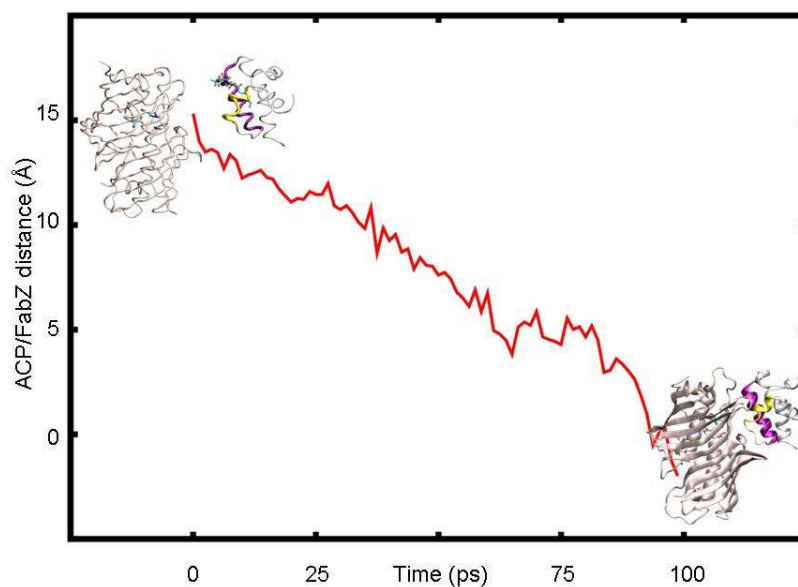


Figure 5. Lowering distance in the association between *Pf*ACP and *Pf*FabZ.

What can be argued from such simulations? α3-helix of ACP is a small flexible protein portion containing an high density of charged residues (D57, E59 and K61). The following scenario might be hypothesized: the approaching ACP points its α2 and α3 helices towards the active site of FabZ. The active site of FabZ is surrounded by positively charged residues that would have a long-rage attractive effect on the acid residues present on ACP surface. The dynamics of α3-helix might be perturbed by approaching the surface of FabZ. Namely, ACP residues D57, and E59 might tend toward FabZ even before the interaction of the two surfaces. The perturbation of the α3-helix might trigger a conformational rearrangement in ACP structure[5] leading to the exposure of the carried substrate.

**References**

1. Byers, D. M.; Gong, H., Acyl carrier protein: structure-function relationships in a conserved multifunctional protein family. Biochem Cell Biol 2007, 85, (6), 649-62.
2. Parris, K. D.; Lin, L.; Tam, A.; Mathew, R.; Hixon, J.; Stahl, M.; Fritz, C. C.; Seehra, J.; Somers, W. S., Crystal structures of substrate binding to Bacillus subtilis holo-(acyl carrier protein) synthase reveal a novel trimeric arrangement of molecules resulting in three active sites. Structure 2000, 8, (8), 883-95.
3. Zhang, Y. M.; Marrakchi, H.; White, S. W.; Rock, C. O., The application of computational methods to explore the diversity and structure of bacterial fatty acid synthase. J Lipid Res 2003, 44, (1), 1-10.
4. Zhang, Y. M.; Wu, B.; Zheng, J.; Rock, C. O., Key residues responsible for acyl carrier protein and beta-ketoacyl-acyl carrier protein reductase (FabG) interaction. J Biol Chem 2003, 278, (52), 52935-43.
5. Colizzi, F.; Recanatini, M.; Cavalli, A., Mechanical features of Plasmodium falciparum acyl carrier protein in the delivery of substrates. J Chem Inf Model 2008, 48, (12), 2289-93.
6. Hill, R. B. A NMR Approach to the Study of Protein-Protein Interactions: The Interaction of 4-fluorodecanoyl-Acyl Carrier Protein with Bloch Dehydrase. Yale University, New Haven, CT, 1995.

7.      Worsham, L. M.; Earls, L.; Jolly, C.; Langston, K. G.; Trent, M. S.; Ernst-Fonberg, M. L., Amino acid residues of Escherichia coli acyl carrier protein involved in heterologous protein interactions. Biochemistry 2003, 42, (1), 167-76.

8.      Bogan, A. A.; Thorn, K. S., Anatomy of hot spots in protein interfaces. J Mol Biol 1998, 280, (1), 1-9.

9.      Keskin, O.; Gursoy, A.; Ma, B.; Nussinov, R., Principles of protein-protein interactions: what are the preferred ways for proteins to interact? Chem Rev 2008, 108, (4), 1225-44.

10.     Lichtarge, O., Getting past appearances: the many-fold consequences of remote homology. Nat Struct Biol 2001, 8, (11), 918-20.

11.     Lichtarge, O.; Sowa, M. E., Evolutionary predictions of binding surfaces and interactions. Curr Opin Struct Biol 2002, 12, (1), 21-7.

12.     Madabushi, S.; Yao, H.; Marsh, M.; Kristensen, D. M.; Philippi, A.; Sowa, M. E.; Lichtarge, O., Structural clusters of evolutionary trace residues are statistically significant and common in proteins. J Mol Biol 2002, 316, (1), 139-54.

13.     Armon, A.; Graur, D.; Ben-Tal, N., ConSurf: an algorithmic tool for the identification of functional regions in proteins by surface mapping of phylogenetic information. J Mol Biol 2001, 307, (1), 447-63.

14.     Glaser, F.; Pupko, T.; Paz, I.; Bell, R. E.; Bechor-Shental, D.; Martz, E.; Ben-Tal, N., ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. Bioinformatics 2003, 19, (1), 163-4.

15.     Goldenberg, O.; Erez, E.; Nimrod, G.; Ben-Tal, N., The ConSurf-DB: pre-calculated evolutionary conservation profiles of protein structures. Nucleic Acids Res 2009, 37, (Database issue), D323-7.

16.     Landau, M.; Mayrose, I.; Rosenberg, Y.; Glaser, F.; Martz, E.; Pupko, T.; Ben-Tal, N., ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. Nucleic Acids Res 2005, 33, (Web Server issue), W299-302.

17.     Pupko, T.; Bell, R. E.; Mayrose, I.; Glaser, F.; Ben-Tal, N., Rate4Site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. Bioinformatics 2002, 18 Suppl 1, S71-7.

18.     Lehtonen, J. V.; Still, D. J.; Rantanen, V. V.; Ekholm, J.; Bjorklund, D.; Iftikhar, Z.; Huhtala, M.; Repo, S.; Jussila, A.; Jaakkola, J.; Pentikainen, O.; Nyronen, T.; Salminen, T.; Gyllenberg, M.; Johnson, M. S., BODIL: a molecular modeling environment for structure-function analysis and drug design. J Comput Aided Mol Des 2004, 18, (6), 401-19.

19.    Nicholls, A.; Sharp, K. A.; Honig, B., Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins 1991, 11, (4), 281-96.

20.    Schreiber, G.; Fersht, A. R., Rapid, electrostatically assisted association of proteins. Nat Struct Biol 1996, 3, (5), 427-31.

21.    Selzer, T.; Albeck, S.; Schreiber, G., Rational design of faster associating and tighter binding protein complexes. Nat Struct Biol 2000, 7, (7), 537-41.

22.    Chen, R.; Li, L.; Weng, Z., ZDOCK: an initial-stage protein-docking algorithm. Proteins 2003, 52, (1), 80-7.

23.    Schneidman-Duhovny, D.; Inbar, Y.; Nussinov, R.; Wolfson, H. J., PatchDock and SymmDock: servers for rigid and symmetric docking. Nucleic Acids Res 2005, 33, (Web Server issue), W363-7.

24.    Lyskov, S.; Gray, J. J., The RosettaDock server for local protein-protein docking. Nucleic Acids Res 2008, 36, (Web Server issue), W233-8.

# 4. Conformational plasticity in *Pf* FabZ

This little chapter resumes and augments some conformationals studies which I started during my master thesis and completed at the early stage of my PhD. It might be considered a *trait d'union* between the previous and of the following chapter.

Crystal structure ascertainment of ß-hydroxyacyl-ACP dehydratase (FabZ) from diverse organisms suggested that a certain extent of structural plasticity may play a critical role influencing the capabilities of the enzyme both to interact with ACP but also to harbor acyl-substrates with different chain length. Hence, the hypothesis is investigated using Langevin dynamics simulations.

*Pf*FabZ is the β-hydroxyacyl-ACP dehydratase that catalyzes the third step in chain elongation during fatty acid biosynthesis. The biological relevant form of *Pf*FabZ is a dimer, with two active sites symmetrically formed at the interface. The catalytic residues are His133 and Glu147' (where ' indicates the residue of the symmetric chain). Together with His98' they form the only hydrophilic site in the otherwise completely hydrophobic active site.

Previous studies[1] have shown (Figure 1) that the effects of conformational changes of few residues lining the active site pocket of *Pf*FabZ may have large effect on the shape of the binding tunnel. In particular, the conformation of Phe169 highly influenced the extension of the binding tunnel and accounts for the capability of *Pf*FabZ to metabolize substrate of different chain length. Such conformational change was thought to be likely induced upon binding of hindered ligands. The inhibitory activity of the bioflavonoid molecule, amentofalvone,[2] was interpreted as proof of the concept that *Pf*FabZ might undergo various conformational change upon ligand binding (Figure 1).
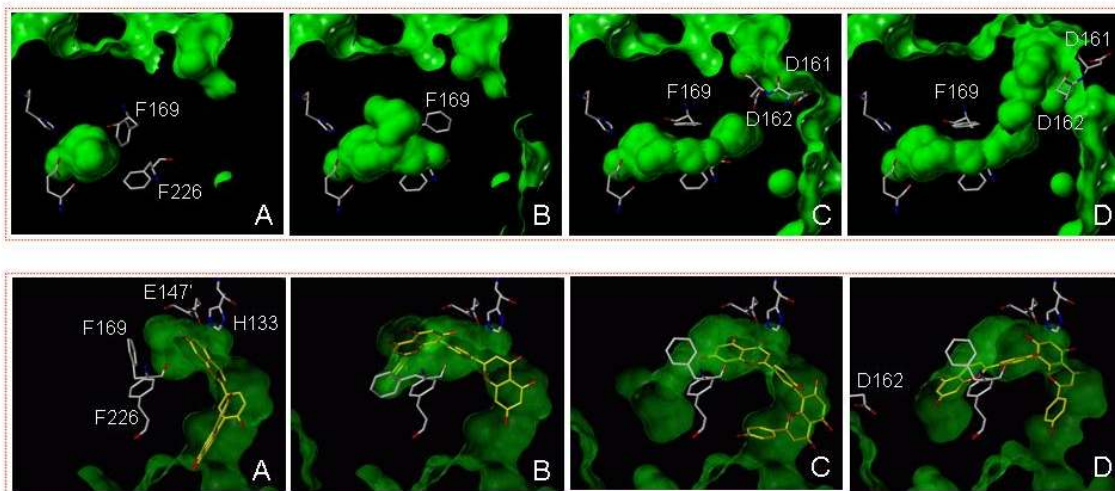
Figure 1. PfFabZ active site. Connolly surfaces according to small conformational rearrangement of Phe169 (upper row). Structural rearrangement allowed proper docking of the large bioflavonoid inhibitor (lower row).

Molecular dynamics simulations were employed to shed light on conformational properties of *Pf*FabZ structure. Several conformational ensembles of the enzyme were used as starting point for the atomistic simulations. Few issues are reported and shortly discussed below:

1) A large conformational change of the α2-β3 loop was sampled. The entity of such rearrangement is shown in Figure 2. Fluctuations in the distance between L168 and L96 were also sampled and were interpreted as the natural and functional "breath" of the protein structure.

2) Three phenylalanine residues (F169, F171 and F226 referred as the triad of phenylalanine) showed high mobility fluctuating around a "stable" conformation. The stable conformation corresponded to a native-similar rearrangement of the triad in which, Phe171 and Phe226 stack with a parallel-displaced geometry and Phe169 and Phe226 via an edge-to-face interaction.
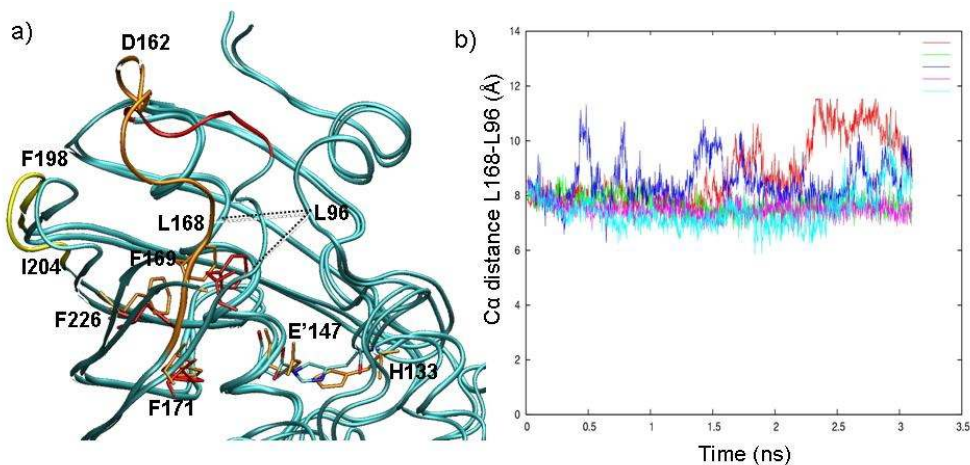
Figure 2. PfFabZ α2-β3 loop conformational switching. a) Difference between conformations at 0ns (red) and 3ns (orange). b) Monitored Cα distance between L168-L96 for five independent simulations with different starting configuration.
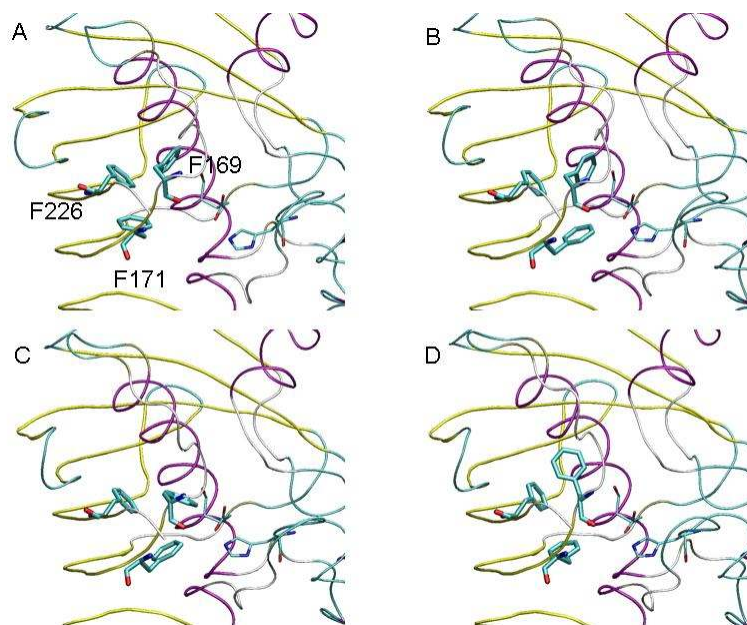


Figure 3. Snapshot along a 3ns MD trajectory. (A) Initial configuration, (B) Phe171 slips away from its position and occupies the catalytic environment nearby His133 and Glu147'. Phe171 conformational switch is stabilized by π-π interaction with Trp179 (not shown). (C) Phe226 lost the interaction with Phe169. (D) Phe169 is less packed and populates a different conformation energy minimum. Phe171 returns in its original position swaying with Phe226. After ~2.5 ns of simulation a native like conformational ensemble is restored.

48

Concluding, using molecular dynamics simulations we confirmed the hypothesis arising from crystal structure ascertainment. *Pf*FabZ is conformationally versatile and its adaptability may account for functionality. The conformational plasticity was associated mainly with a wide loop breathing movement and with conformational fluctuations of the active site and in particular of the phenylalanine triad.

**Methodological details**

MD simulations were performed using NAMD with the CHARMM22 force field for protein and the TIP3P model for water molecules. The protein was immerged into a 10 Å thick water box leading to a system of about 32,000 atoms. Six chlorine ions were added to neutralize the system and PME was used for full electrostatic. The cutoff was of 10 Å with a switching distance of 8 Å. The time integration step was 2fs and the SHAKE algorithm was therefore used. The system was minimized for 3000 steps gradually releasing the harmonic restraints on protein heavy atoms. The Langevin equation was used to generate the Boltzmann distribution for canonical (NVT) ensemble simulations. A damping coefficient of $5ps^{-1}$ was used for the whole simulation length. After the temperature was stabilized at 300K, the constant pressure control was applied with the default Nos´e-Hoover Langevin piston method and 1 Atm was set as target pressure. The NPT simulation ran for 3ns.

1.      Colizzi, F., Modeling conformational changes in PfFabZ. In 2005.
2.      Perozzo, R., Personal Communication IC50=0.4uM. In.

# 5. Atomistic simulations discern active from inactive ligands of the β-hydroxyacyl-ACP dehydratase of *Plasmodium falciparum.*

Understanding ligand-protein recognition and interaction processes is of primary importance for any structure-based drug design project. Several approaches combining molecular docking and molecular dynamics simulations have been exploited to investigate the physicochemical properties of putative complexes of pharmaceutical interest. Molecular docking is often able to rationalize the structure-activity relationships of known inhibitors. However, it is well known that, even if the geometric properties of a modeled protein-ligand complex can be well-predicted by computational methods, scoring functions are prone to rank series of analogue ligands not in a consistent fashion with available biological data.

In the unique β-hydroxyacyl-ACP dehydratase of Plasmodium falciparum (PfFabZ), the application of standard molecular docking procedures was partially sufficient to rationalize the activity of previously discovered inhibitors. Complementing docking results with atomistic simulations in the steered molecular dynamics (SMD) framework, we could hypothesize a binding model able both to fully explain the biological activity of known ligands, and to provide sufficient insight to address the design of novel enzyme inhibitors. To our knowledge, this is the first time that SMD-derived force profiles have been proven to be useful to clearly discern active from inactive compounds.

## *Introduction*

Malaria kills more than one million people each year and is one of the major causes of death in children in the developing world. Plasmodium falciparum infections are estimated to be about half billion annually and continuously threaten lives as well as the economic-development opportunity of countries in the tropical and subtropical zones of the world.

Plasmodium species have two distinct replicating life cycle forms in the mammalian host. The first occurs in the liver following the inoculation of the parasite by the bite of an infected mosquito. Afterwards, an intense hepatic replication occurs and the parasite invades the blood stream and here

infects red blood cell and initiate the second cyclic replication. Blood stage infection is responsible for most of the malaria symptoms.

Development of antimalarial compounds is a critical and hanging emergence due to the high incidence of drug-resistant parasite strains and since an effective vaccine is still lacking.[1] Existing alternative medications or prophylaxis are virtually unaffordable in the countries affected most seriously (ref). Malaria burden has stimulated investigators to seek out novel inhibitors and drug targets. In the last years, the biosynthetic machinery involved in fatty acid production has been looked at as a promising antimalarial target because of the different structural organization of the enzymatic moieties between plasmodium and human. Together with its structural peculiarities, the parasitic fatty acid biosynthesis (FAS II) is essential for the membrane biogenesis necessary during invasive stage. Recently, independent studies have shown that FAS II plays a vital role in liver-stages development and is not essential for the blood stages.[2, 3] Moreover, using knockout parasites Vaughan et al. have demonstrated that the lack of FAS II renders the pre-erythrocytic parasite unable to successfully infect the mammalian host. In this scenario, FAS II inhibitors would be endowed with the capability to interfere with the transition stage in which the parasite moves from the liver into the blood stream, and initiate red blood cell infection. Ideally, such a prophylactic strategy might significantly contribute to malaria eradication.

The elongation of the acyl chain in FAS II is catalyzed by four key enzymes, FabB/F, FabG, FabZ, and FabI. The growing acyl substrate is covalently bound to the acyl carrier protein (ACP) which shuttles and delivers the substrate from one enzyme to the other. The unique β-hydroxyacyl-ACP dehydratase of Plasmodium falciparum (*Pf*FabZ) catalyzes the dehydration of β-hydroxyacyl-ACP to form trans-2-enoyl-ACP, and its three dimensional structure has been well characterized by X-ray crystallography.[4, 5] Crystal structures ascertainment and comparison reveal a certain extent of

51

structural plasticity in the substrate-binding region, reflecting the functional role of metabolizing acyl-substrates with different chain-length. *Pf*FabZ structure has been described as a homodimer in which each monomer adopts the typical "hot dog" fold, where six anti-parallel β-sheets wrap around a long central α-helix. The dimer forms two independent substrate-binding tunnels with the catalytic sites at the interface. The whole substrate-binding site is prevalently hydrophobic with the catalytic amino acids His133 and Glu147' (from the other subunit), together with His98', representing the only hydrophilic spots. Recently, Zhang et al. have also described the X-ray crystal structure of Helicobacter pylori FabZ (*Hp*FabZ).[6] Whereas the overall three dimensional structure of *Hp*FabZ is similar to other FabZs,[5, 7, 8] it contains an additional two-turns helix, α4, which has not been observed in similar structures, and which plays a peculiar role in shaping and rigidifying the substrate-binding tunnel. Moreover, the short helix α4 of *Hp*FabZ is formed by extra residues, where there is usually a flexible loop in other FabZ structures.[6] Three different flavonoids, quercetin, apigenin, and sakuranetin, have been co-crystallized and solved in the binding site of *Hp*FabZ.[9] They have been reported as competitive inhibitors against *Hp*FabZ by either binding at the entrance of substrate tunnel or nearby the catalytic site. However, since the surrounding of the binding site of *Hp*FabZ is significantly different from those previously reported and particularly from *Pf*FabZ, both a different biochemical and structural behavior of the same inhibitor is likely to occur. As a matter of fact, apigenin has been reported to be a micromolar inhibitor of *Hp*FabZ, while it is inactive against *Pf*FabZ.[10] Therefore, the binding modes observed for flavonoids at *Hp*FabZ active site are unlikely exportable to other FabZ structures.

Flavonoids are a common component in human diet since they are ubiquitously present in fruits and vegetables as well as in different types of beverages; flavonoids are also reported as the major bioactive component of several herbal preparations for medical use. The understanding and

characterization at the molecular level of their interaction with biologically relevant targets is of primary importance in view of designing more potent derivatives and optimizing their medical application.

Despite the three dimensional characterization of *Pf*FabZ and structural information available for other FabZ enzymes, no structure-based drug design attempt of novel anti-malarial lead candidates has been reported. Nevertheless, there are several recent studies reporting on both synthetic and natural compounds able to inhibit *Pf*FabZ enzymatic activity.[10, 11] In particular, Tasdemir et al. have screened a large flavonoid library against *Pf*FabZ, finding competitive inhibitors showing mid- to low-micromolar activity. Despite the large number of tested compounds, structure-activity relationships (SARs) interpretation for such a series was found to be far from trivial, and therefore a systematic SARs study is still missing.

In this paper, taking advantage of the recent work on flavonoid derivatives of Tasdemir et al., we report on a structure-based computational approach to investigate putative modes of interaction of such a flavonoid series with *Pf*FabZ. In particular, combining molecular docking and steered molecular dynamics (SMD) simulations, we propose a dualistic binding mode able to shed light on the hazy SARs of this series. The model of binding was then perspectively exploited to identify novel flavonoid inhibitors. Moreover, analyzing the SMD-trajectories and force profiles we were able to discern active from inactive compounds. To the best of our knowledge, the present is the first report on a systematic exploitation of SMD-force profiles in structure-based drug design.

## Results and Discussion

**The hydroxy-flavone series**

Among the flavonoids derivatives shown in Table1, luteolin (**12**) has been originally considered the lead compound, and therefore its structure is here taken as reference. The minimal structural requirements needed to confer inhibitory activity to the flavonoid scaffold is represented by the hydroxy-flavone compounds (cmpds 2-5) shown in Figure1a and in Table1. Biological data reported in Table 1 suggested the favorable effects on PfFabZ inhibition by 6- and 7-hydroxy substitutions (**4** and **5**), while positioning a hydroxyl group in 5- and 3- (**2** and **3**) led to completely loose the inhibitory activity. To provide a possible explanation of such a behavior, we used standard molecular docking calculations and the obtained binding mode is reported in Figure 1.
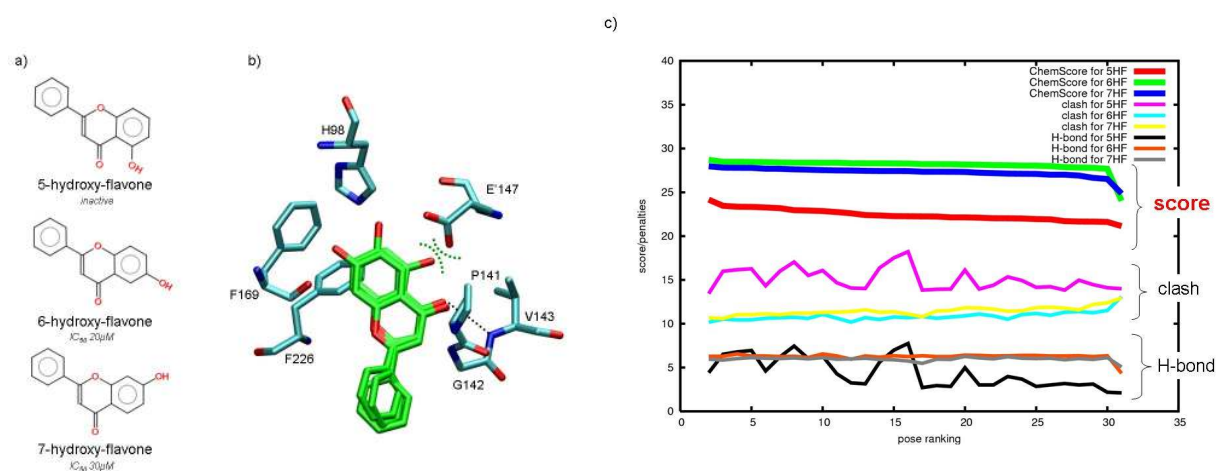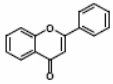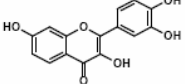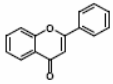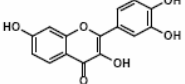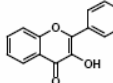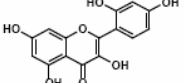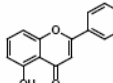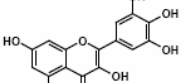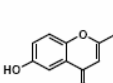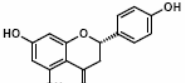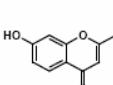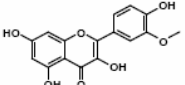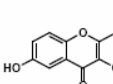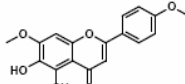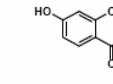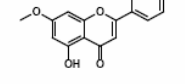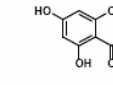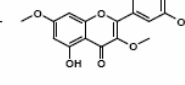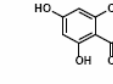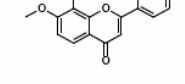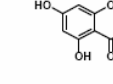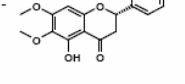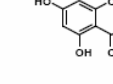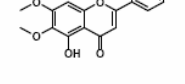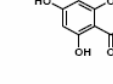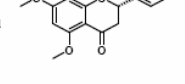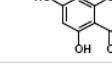


Figure 1. The hydroxy-flavone series. a) Chemical structures of b) 5-, 6-, and 7-hydroxyflavone (HF), carbon atoms in green, in the catalytic pocket of PfFabZ. The green dotted curves highlight the steric clashes between E147' and the 5-hydroxyl group of 5HF. For each ligand the representative pose of the unique significantly populated cluster is reported. c) Representation of the scoring function trend over 30 docked poses. The concerted behavior of the black and the pink lines shows that the more the 5-HF try to find optimal H-bond interactions the more clash penalties are generated as a consequence of positioning the 5-hydroxyl group too close to the E147' side chain.

The docked compounds were embedded into the tunnel leading to the catalytic dyad and shared a similar interaction pattern, since the carbonyl group of flavones formed a pivotal H-bond with the backbone of Val143. The chromone ring interacted with Phe226 side chain, and the whole molecule was surrounded by hydrophobic residues. The 5-, 6-, or 7-hydroxy substituent were involved in interactions with Glu147 and His98 of *Pf*FabZ binding site. We could detect that the 6- and 7-hydroxy derivatives established a network of H-bond interactions with side chains of Glu147 and His98 or with the backbone of Phe169, while the 5-hydroxy derivative although compatible for interacting with Glu147, suffered from relevant steric clashes. Such repulsion was accounted for by the scoring function and its nature was well described by the H-bond and steric-clash scoring (Figure 1c). Hence, the more the favorable H-bond interaction is sampled (ascending black line), the more the steric clash term (violet line) increase. As a consequence, when compared to 6- and 7-hydroxy analogues, 5-hydroxy flavone obtained the lowest score (red line) because the H-bond interaction was abolished by the steric clash penalties. Similarly, the 3-hydroxyl group of compound **2** clashed against the backbone of G142 generating only low scored docking solutions (not shown). In agreement with the experimental data, these results point to a destabilizing role for the 3- and 5-hydroxyl groups of the flavone core, whereas at the 6- and 7- position the same substituent is able to efficiently interact with *Pf*FabZ binding site.

Table1. Inhibition of *Pf*FabZ by Flavonoids. (Adapted from Tasdemir et al.[10])

| compd | name | | IC$_{50}$ (μM) FabZ | compd | name | | IC$_{50}$ (μM) FabZ |
|---|---|---|---|---|---|---|---|
| 1 | flavone |  | n.a. | 14 | fisetin |  | 2 |
| 2 | 3-hydroxy-flavone |  | n.a. | 15 | morin |  | 8 |
| 3 | 5-hydroxy-flavone |  | n.a. | 16 | myricetin |  | 2 |
| 4 | 6-hydroxy-flavone |  | 30 | 17 | naringenin[d] |  | n.a. |
| 5 | 7-hydroxy-flavone |  | 20 | 18 | isorhamnetin |  | n.a. |
| 6 | 3,6-dihydroxy-flavone |  | n.a. | 19 | ladanein |  | n.a. |
| 7 | 3,7-dihydroxy-flavone |  | n.a. | 20 | genkwanin |  | n.a. |
| 8 | chrysin |  | n.a. | 21 | quercetin-3,7,3',4'-tetramethylether |  | n.a. |
| 9 | galangin |  | n.a. | 22 | 7,8-dimethoxy-flavon |  | n.a. |
| 10 | apigenin |  | n.a. | 23 | 5,4'-dihydroxy-6,7-dimethoxy-flavanone[d] |  | n.a. |
| 11 | kaempferol |  | n.a. | 24 | cirsimaritin |  | n.a. |
| 12 | luteolin |  | 5 | 25 | 5,7-dimethoxy-8-methyl-flavanone[d] |  | 40 |
| 13 | quercetin |  | 1.5 | | | | |

**Luteolin as reference compound: investigation of its interaction with PfFabZ**.

The good agreement between scoring function outcomes and biological data was not any longer observed when dealing with more complex flavonoid derivatives (compounds 8-16 in Table1). Moreover, the geometric properties of ligand-protein complexes were rather different as a consequence of multiple binding modes with similar energy scores. This could be due both to the ability of polyhydroxylated flavonoids to interact with the biological counterpart in different and degenerate way,[9] but also to the difficulties of the scoring function to discriminate amongst equivalent binding modes.

Fast and inexpensive docking protocols have often been used in combination with accurate but more computational costly molecular dynamics (MD) techniques to predict reliable ligand-protein complexes.[12] The synergistic combination of the two techniques can overcome the weaknesses of each method taken alone.[13] Molecular docking is usually used to rapidly explore the configurational space of ligands providing MD simulations with more reliable starting configurations, which are then evolved along ns-trajectories. This approach can significantly allow reshaping of the receptor conformation, therefore allowing induced-fit effects and postprocessing analyses which lead to likely discriminate among configurations with different thermal stability.[14] Notably, MD simulations of ligand-protein complexes are often used to obtain more accurate energetic evaluation of binding.[12, 15, 16] Seeking out a general binding paradigm useful to lay down rational SARs for the present series of flavonoid derivatives, we therefore used MD simulations to assess the relevance of different binding modes of the representative compound, **12** (Table 1).

Docking poses of **12** at the binding site of *Pf*FabZ were generated using a genetic algorithm, which provides a population of ligand configurations at the receptor binding site (Jones, 1997). The

outcomes were then processed by a hierarchical-agglomerative clustering procedure, which has been proven to efficiently reduce the dimensionality of the original dataset and to highlight the most relevant docking poses. We obtained two clusters: one significantly populated according to the Chauvenet criterion, and the other less populated but containing the best ranked docking pose. Together, the two clusters covered more than 90% of the sampled configurational space.

The two modes of binding (hereafter referred to as A and B, respectively) of **12**, corresponding to the best ranked docking pose and the most populated one are shown in Figure 2.
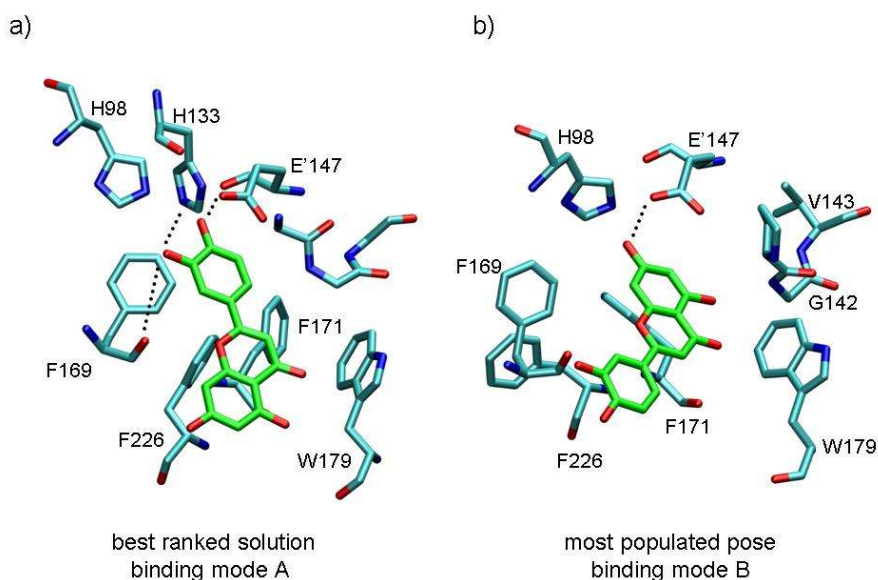


Figure 2. Molecular details of the binding model corresponding to the a) best ranked docking solution and b) to the most populated docking solution (His133 was not displayed for sake of clarity). Both binding models are here shown as they appear at the end of a 3ns MD trajectory.

In both models, **12** is embedded in the same region of the binding site occupying the tunnel entrance in juxtaposition to the catalytically relevant residues Glu147', His133, and His98'. In A, the catechol ring is buried, while in B, it is more solvent exposed. In Figure 3a, the catechol moiety (the B ring) pointed deep into the binding pocket with both the hydroxyl groups in 3' and 4' establishing H-bond

58

interactions with side chain of residues Glu147', His133, His98, and backbone of Phe169. Several lipophilic residues, such as the triad of phenylalanine (169, 171, and 226) shaped the cavity and allowed the favorable docking of the flavonoid core. Other polar groups of the ligand were poorly involved in interactions with the protein and, in particular, the 5- and 7-hydroxyl groups tended to be partially exposed to the solvent bulk. The di-hydroxylated B ring of the molecule played a pivotal role in docking **12** at the enzyme. Vice versa, in the binding mode shown in Figure 3b, the B ring scarcely interacted with the protein, whereas the 5- and 7-hydroxyl groups were embedded in the binding pocket. As shown in Figure 3b, only the 7-hydroxyl group was actually able to find H-bond interactions with Glu147' and His133, while the 5-hydroxyl and the carbonyl group were shifted away from putative interacting residues (like Glu147' and the backbone of Val143 or Gly142). This was likely due to the lack of flavones bearing the 5-hydroxyl groups to satisfy the sterical and electronic features required for interacting with *Pf*FabZ using the interaction pattern suggested by Figure 1b for 6- and 7-hydroxy-flavone ligands.

In Figure 3a, the root mean square deviation (rmsd) of **12** in A and B over 3 ns of MD simulations is shown. "A" corresponded to the most stable but less populated pose, while B corresponded to the pose representative of the most statistically populated cluster. The green plot (B), showed a less stable evolution of the associated binding mode, likely reflecting the presence of a wider energy well in which several local minima were separated by energy barriers lower than the thermal energy $k_b T$, and therefore accessible in the ns-time-scale of simulations. Such a behavior is consistent with the most-populated-cluster-pose origin of the starting configuration.[17] The most populated cluster could be a consequence of a region of the configurational space in which both sampling and scoring converged more frequently when compared to other binding modes. This is likely to occur if, for instance, the sampled region corresponds to a wide, rather than narrow, energy minimum (Figure

3b). A narrow energy minimum, instead, is likely to correspond to the best ranked docking pose (red plot in Figure 3), which exhibited a more stable time dependent evolution along the sampled trajectory. Whereas we could assume a greater enthalpic stabilization for the narrow energy minimum, the entropic contribution could be more relevant for a wide basin. However, this is generally not accounted for when the energy of a single ligand-protein configuration is calculated. Therefore, finding the narrower energy minimum of a ligand-protein complex might be meaningless when its entropy accounts for a sizeable fraction of its free energy. In this regard, clustering may help in detecting near-native states by determining whether or not a minimum is located in broad energy basins which will be favored for entropic contribution.[17-19]
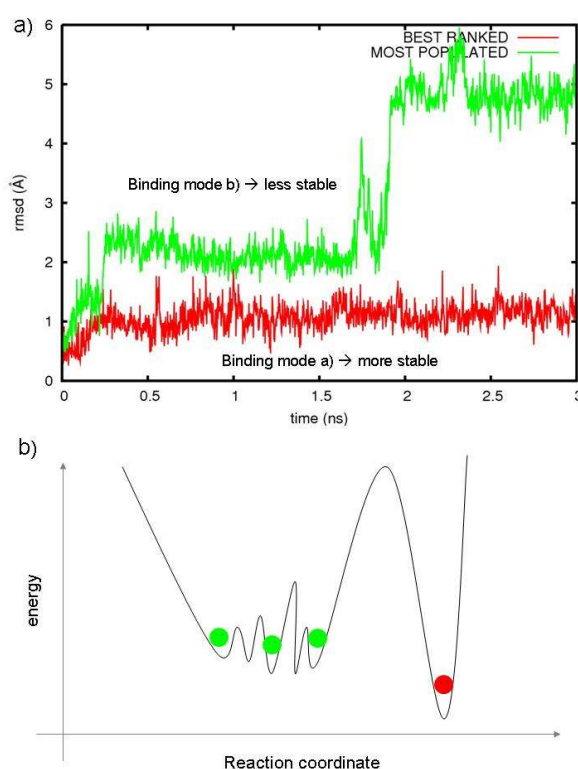


Figure 3. Qualitative differences between most populated (in green) and best ranked (in red) docking pose. a) Root mean square deviation (rmsd) of the thermalized docking poses along 3ns of unrestrained MD trajectories. b) Schematic representation of the energy landscape along and arbitrary association coordinate. The most populated pose likely belongs to a wider energy minimum in which similar configurations can easily interchange to each others. The best ranked pose fall into a narrow minimum and poorly interchange.

60

The analysis so far conducted well described the thermal stability of each binding mode, but no quantitative information about the depth of the free energy wells was provided. In particular, we were concerned whether the cluster population could point to a remarkable increase of entropic contribution for B when compared to A.

Various physic-based methods have been developed and successfully applied to quantitatively characterize ligand-protein interactions, and to reconstruct the free energy profile of the recognition and binding processes.[12, 16] The binding/unbinding of a ligand to a receptor, as well as large conformational rearrangements of a macromolecule, is an event whose energy barrier is often too high to allow its spontaneous sampling by ns-time-scale MD simulations. Among several possible approaches to this issue, SMD allows one to apply a time dependent external bias to encourage the ligand to overcome energy barriers, and to sample a predefined unbinding pathway. In the SMD framework, the ligand is harmonically restraint to a constant velocity moving point thus making it possible to obtain the mechanical irreversible work necessary for the transition by integrating the exerted force on the system along the unbinding reaction coordinate. Furthermore, if the harmonic restrain is large enough so that the ligand does not drift far from the biasing point, the Jarzynski nonequilibrium work theorem[20] can be exploited to discount the dissipated work, and to reconstruct the free energy profile along the selected reaction coordinate (potential of mean force, PMF).[21-23] Notably, SMD is emerging as a promising tool in studying protein-ligand recognition and interaction.[24]

**The unbinding reaction coordinate.**

Here, the unbinding reaction coordinate was investigated using the Random Expulsion Molecular Dynamics approach introduced by Luedemann et al.,[25] which shows the following unique features in the SMD framework: i) low computational cost; ii) objective pathway search. Hence, an unbiased search for ligand escape pathways is pursued by reiterating individual SMD simulations in which a randomly oriented force is applied to the ligand in addition to the standard force field. By collecting and clustering all the successful expulsion trajectories of **12** from the binding site of *Pf*FabZ it was possible to unambiguously identify the preferred ligand escape pathway (Figure 4).
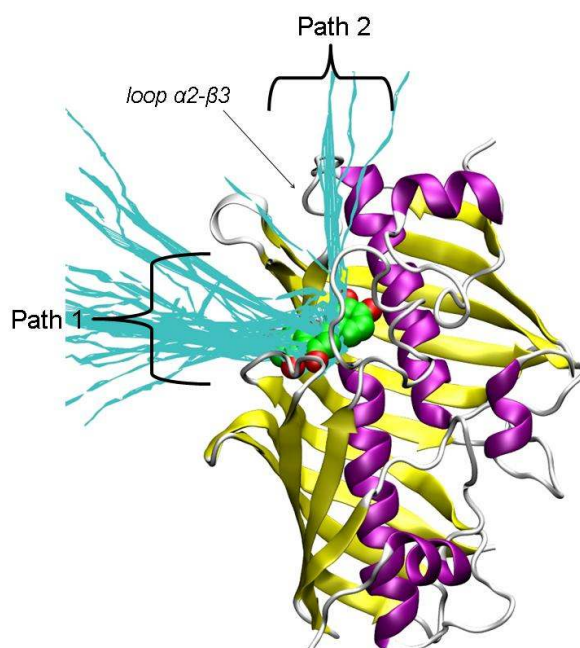


Figure 4. Unbinding trajectories generated by Random Expulsion Molecular Dynamics. The cyan segments represent the trail of c.o.m. of luteolin (carbon atoms in green) escaping from the PfZabZ (purple, yellow and white cartoons) binding site. The unbinding trajectories, path1, and path2, correspond to the more perturbing force applied within a REMD session.

The egress path was then used as reaction coordinate (Figures 5) to reconstruct the PMF profile, and therefore quantify the energy well-depth, of binding modes A and B obtained from molecular docking. As shown in Figure 4, two ligand unbinding pathways were identified without any *a priori* knowledge. Path1, accounting for ligand escape through the most direct connection between the ligand and the protein surface, and Path2 exploiting the flexibility of the loop connecting the central helix α2 to β3, an intrinsic plastic feature of PfFabZ structure. Importantly, path2 was sampled only few times using the highest values of perturbing force and its occurrence tended to zero for smaller forces (see Methods). Therefore, we selected the most occurring pathway, path1, as the representative one for the unbinding event (Figure 5). Although path2 was discarded from further investigation in the current study, it is worth to mention how the "breathing" of the flexible α2-β3 loop connection may play an important role in facilitating the interaction with the acyl carrier protein and therefore the delivery of the substrate from the carrier core to the enzyme active site.



Figure 5. Luteolin docked into the binding site of PfFabZ. The red thick arrow represents the direction of the unbinding reaction coordinate derived from averaging the trajectories of path1. The loop connecting α2 to β3 is labeled.

**PMF calculation**.

Besides the unbinding reaction coordinate definition, constant-velocity SMD was used to provide a detailed energy profile of the unbinding process. Although employing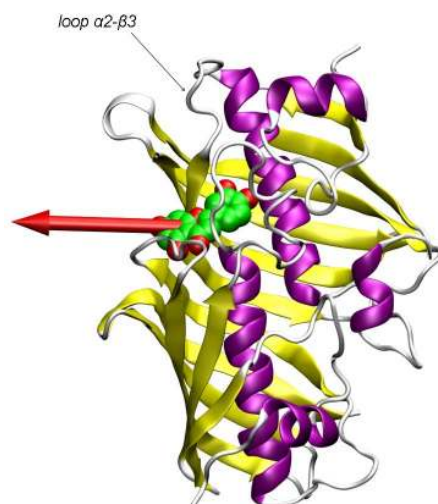 the Jarzynski's equality (JE) eq. 1 allows in theory the irreversible work to be discounted for any arbitrarily irreversible process,

$$e^{-\beta \Delta F} = \lim_{N \to \infty} \left\langle e^{-\beta W} \right\rangle_N \qquad\qquad \beta = 1/k_b T \quad \text{Thermal factor} \qquad (1)$$

in practice, its direct application is limited by the number of collectable trajectories together with the complexity of the biological system which often leads to a standard deviation ($\sigma$) of the work several times higher than $k_b$T.[26] Estimation of the exponential average $\left\langle e^{-\beta W} \right\rangle$ crucially depends on rarely sampled trajectories corresponding to the left tail of Gaussian work distribution.[20, 23, 26] On top of that, SMD pulling paths often sample the region around the peak rather than the tails of the Gaussian work distribution. As the spread of $\sigma$ increases, the probability to sample a region far from the peak likely decreases and the accuracy in reconstructing the potential of mean force is strongly biased. Such a systematic statistical uncertainness has been variously treated in the last years leading to more effective applications of the Jarzynski nonequilibrium work theorem.[21, 23] The cumulant expansion approach has been widely applied[21, 23, 27] to reconstruct the free energy profile of biomolecular processes and here exploited to compare the energy well depth of two putative binding modes of **12**.

The cumulants up to the second order are shown below (eq. 2). Third and higher cumulants are identically zero since within the stiff-spring approximation the distribution of work W is Gaussian.

$$\Delta F = \langle\, W\, \rangle - \frac{\sigma_W{}^2}{2kT} \qquad\qquad \sigma_W{}^2 = \langle\, W^2\, \rangle - \langle\, W\, \rangle^2 \qquad\qquad (2)$$

Eq. 2 directly suggests that free energy differences between states can be obtained averaging the irreversible work $\langle\, W\, \rangle$ and discounting its variance $\sigma_W{}^2$. Here, follows that higher irreversible work trajectories are expected to have higher variance so that discounting its value one can obtain the same $\Delta F$. Whereas with limited SMD trajectories one can determine fairly accurately the average work $\langle\, W\, \rangle$ the variance is generally underestimated due to the above mentioned sampling limitations. The application of irreversible work, obtained from forward and reverse SMD trajectories, has shown to be a valuable approach to overcome the above shortcomings.[28, 29] However, the bidirectional approach is not straightforwardly feasible in processes such as the undocking of a ligand from an enzyme in explicit solvent. For the reconstruction of the PMF, we preliminarily tested several unbinding pulling rate balancing the needs for a proper sampling with those of limited simulation time. We finally used the extremely slow pulling velocity value of 0.5Å/ns.[27, 30] Such a slow pulling rate was chosen considering the peculiar features of the binding modes here investigated. In Figure 2 is shown how, regardless the binding mode selected, the anchoring interaction between luteolin and PfFabZ was represented by hydroxyl groups of the ligand H-bonding the hydrophilic catalytic residues and particularly Glu147' with whom the interaction was enforced by the net charge. The catalytic residues are located at the end of the tunnel entrance and thus not easily accessible to solvent molecules especially if luteolin physically occupies the entrance itself. Nonetheless, the binding site exhibits a wide range of structural flexibility (e.g. the "breathing" of the α2-β3 loop connection), and beside the tunnel entrance, several water accessible pathways might be available to create a catalytically competent environment around the active site. Using a relatively high pulling velocity would generate more unlikely vacuum around the breaking

interactions respect to a slower rate in which water molecules would be able to properly solvate the interacting counterparts and assist the breaking of the interactions at a lower energetic cost. Since the irreversible work reflects the energy penalty for inducing a system transformation at a faster velocity than its slowest relaxation rate, employing a slower pulling velocity would not only reduce the irreversible work done but would also capture intrinsic system peculiarities which would improve the likeliness of the simulated event and thereof the meaning of its free energy profile.

(The breaking of the anchoring interactions between luteolin and PfFabZ was manifested by a drop in the force profile and generally this coincided with the highest rupture force along the whole steered simulation. Notably, for binding mode A it was univocally possible to identify the rupture point from the force profile, while for binding mode B a rupture point was generally more spread or not directly deductible from the force profile graphs.)
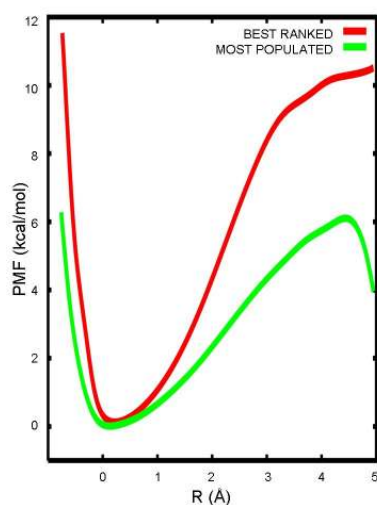


Figure 6. Potential of mean force (PMF) along the unbinding reaction coordinate of luteolin. The free energy profile of the best ranked pose (binding mode A) and the most populated (binding mode B) are compared.

In Figure 6 is plotted the potential of mean force relative to the unbinding of luteolin as to the interacting model A and B. The most populated pose (model B) would belong to a wider free energy basin whereas the best scored pose (model A) to a narrower one. As shown in Figure 6 the width of

the free energy basins was appreciably different and, accordingly the green plot corresponding to docking model B described a wider energy well reflecting a broader region of ligand attraction when compared to the narrower red plot which is related to the binding model A. Notably, the free energy well-depth associated to each of the binding modes investigated can be also deduced from Figure 6. The unbinding of luteolin as it interacts with PfFabZ according to binding mode A required to overcome an energy barrier of about 9kcal/mol whereas the interaction pattern of binding mode B led to a flatter energy barrier of about 5kcal/mol. Whereas the magnitude of the free energy barriers here involved cannot be directly compared[31] to the about -7 kcal/mol of standard binding free energy (estimated from the Ki=11μM[10]) which a low micromolar inhibitor such as luteolin would exhibit, the free energy computations could be used as a quantitative tool to assess different binding models. Hence, the higher energy barrier observed for the binding model A pointed out its interaction pattern as the more stable and therefore its structural features were taken as reference to rationalize the SARs of the flavonoids series shown in Table1. It is worth to stress how the binding mode A privileged the catechol portion of luteolin as the most important moiety in the interaction with PfFabZ. Relating the free energy barriers to the pivotal interactions herein involved, we can deduce that at least two hydroxyl groups in the surrounding of the catalytic residues are needed for an effective H-bond interaction network.

**Rationalizing the SARs of luteolin derivatives.**

The importance of bearing at least two hydroxyl groups in the B ring of the flavonoid scaffold is dramatically in agreement with the experimental biological data reported in Table1. In

fact, all the flavonoids compounds not bearing such a feature are inactive. Exceptions to this rule are represented by **4** and **5**, for which compounds molecular docking clearly suggested only one plausible binding pose. In Figure 1b, it is possible to appreciate how **4** and **5** were able to establish interactions with PfFabZ resembling the binding mode B of luteolin. The inactivity of **3** together with its structural rationalization shown in Figure 1b, suggest that the scaffold alone has no intrinsic capabilities to properly interact with PfFabZ. This is for instance the case of chrisyn (**8**), which has no inhibitory activity, likely as a consequence that a binding mode B-like is discouraged by the 5-hydroxyl group and, at the same time, a binding mode A-like is not feasible since the B ring bears no substituent. Apparently, adding one hydroxyl group is still not enough to reach a medium micromolar activity. This is suggested by the inactive apigenin (**10**) in which the hydroxyl group in the B ring is likely insufficient to establish a stable H-bond network in the catalytic environment. The 5-hydroxyl group does not allow a stable B-like binding mode and compounds such as luteolin (**12**) or quercetin (**13**) are actives since they are able to properly compensate the instability arising from the presence of the 5-hydroxyl group with the presence of the di-hydroxylated B ring able to satisfy the energetic requisites for an effective A-like binding mode. Likely for quercetin (**13**), the 3-hydrohyl group in the chromone scaffold can further stabilize the A-like interaction being it able to point toward the backbone of residue Gly142 (Figure 2a).

In this light, we hypothesize the role of the (3-) 5-hyroxyl group as switcher or trigger between binding mode A and B. One the one hand, if the flavonoid scaffold does not contain the (3-) 5-hyroxyl group then a binding mode B-like is accessible while a binding mode A-like is discouraged if the B ring does not contain at least the di-hydroxylation. On the other hand, if the (3-) 5-hyroxyl group is preserved, then the molecule cannot bind *via* the B-like model and it will be a binder only bearing a hydroxyl rich B ring.

A similar argumentation might be put forward for the role of the 3-hydrohyl group in the chromone scaffold. If the B ring is di-hydroxylated, as it is in fisetin (**14**), then the ligand would better adopt the A-like binding model with the 3-hydoxyl group positively influencing the activity since no clashes arise from this substituent introduction. Given the above binding "rules", then also the inactivity of the various O-methyl derivatives can be rationalized (cmpds **18**-**24**). This is, for instance, the case of isorhamnetin (**18**), in which the O-methylation of the 3'-OH deprives the molecule of its capability to form the proper H-bond network typically borne by the catechol ring. The hints arising from the computation were then complemented with further experimental data. We were confident in understanding why, as suggested by the experimental data shown in Table1, it has been possible to convert the inactive kaempferol (**11**) into the active quercetin (**13**, as well as **15** or **16**) just adding one more hydroxyl group at the B ring. However, the observed activity might also be due to several factors that we may have fully neglected in this study, and therefore the observed activity may not be directly connected to the binding model proposed in Figure 2a. To reduce the range of speculation we further considered the opportunity to get hints from the hypothesized binding model to suggest a micromolar flavonoid inhibitor of PfFabZ not previously reported in the literature (Figure 7).
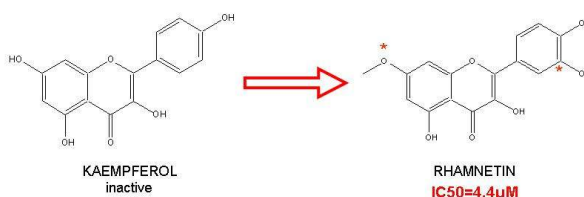


Figure 7. Transforming the inactive kaempferol into the newly disclosed active rhamnetin. Note the presence of at least two hydroxyl groups in the B ring and the 7-O-methyl substitution.

The newly disclosed inhibitor rhamnetin, $IC_{50}=4.4\mu M$ (Personal Communication of Dr. Perozzo at University of Geneva), bears a peculiar structural feature that advances our understanding in linking biological activity to model of interaction. Indeed, the binding model shown in Figure 2a not only suggested the primary role of the catechol ring but also pointed out the non critical role of substituent at position 7 in the chromone core. The 7-hydroxyl group is not directly involved in any interaction with the enzyme but it is also pointing toward the entrance of the tunnel regions suggesting a more sized substitution to be compatible with activity. In contrast, in the binding model B-like, the 7-hydroxyl group is instead the primary interacting moiety of the flavonoid scaffold and substitution at this position would likely largely interfere with biological activity. In this respect, the 7-O-methilated flavonoid inhibitor, rhamnetin, fulfills the gap in the structure-activity relationships of luteolin derivatives reducing the range of speculation about the general validity of binding mode A. Concurrently, as suggested by free energy calculations, the activity of rhamnetin allowed to firmly discard the hypothesis of binding mode B (Figure 2b) for the flavonoid inhibitors bearing the 5-hydroxyl group together with a multi-hydroxylated B ring.

**Pulling away active from inactive compounds**

Finally, we applied an extended SMD approach to further complement the computational as well as experimental studies discussed so far. Particularly, we wanted to explicitly test the general validity of the binding mode here proposed and at the same time evaluate the capabilities of the SMD-approach as a drug discovery tool. SMD simulations have been widely and successfully applied to explore the properties of nonequilibrium processes of biomolecules,[32] and whereas this

methodology has been variously used to investigate and compare the feasibility of putative unbinding pathways, this is the first time, to our knowledge, that such approach is rather used to compare the behavior of different ligands along the same unbinding path. SMD simulations were performed on six representative flavonoid molecules listed in Figure 8a. Three actives, luteolin, rhamnetin and myricetin and three inactive compounds kaempferol, galangin and apigenin were collected together and, in turn, their structure was superimposed to that of PfFabZ-bound luteolin so that their binding mode strictly resembled the one shown Figure 2a. After further equilibrating each ligand-protein complex, an external force was applied in order to steer the ligand along the already defined unbinding reaction coordinate (shown in Figure 5). The applied forces could be monitored throughout the entire simulation length and if the pulled ligand could easily advance along the selected reaction coordinate, then the applied force was small and its profile rather flat. Conversely, if the ligand was more tightly bound to its receptor and it encountered more resistance along the pathway, the applied force increased to overcome the energy barriers, thus resulting in quite relevant drops in the force profile. We compared the magnitude of the exerted force in the computational attempt to keep apart likely binders from non binders. In this particular case, since all the activities of known flavonoids were quite similar we could generalize simply describing this approach as an attempt to discern active from inactive compounds. Should a trend in the force profile be manifested, then the validity of the binding model proposed for luteolin might be likely extended to the whole flavonoid series herein discussed.

The need to carry out a high number of simulations led us to use a pulling velocity which allowed the complete unbinding of the ligand in less than 500 ps. However despite the high velocity, a good resolution was still maintained as shown in Figure 8b. As shown in the figure, to induce the unbinding of each of the inactive compounds it was needed to apply a force not higher that ~400 pN.

71

The force profile corresponding to all the inactives clearly stacked together along the unbinding trajectories. Relatively higher variability was observed for the active compounds. Each force profile likely reflected a slightly different interaction pattern which the compound established along the unbinding process. The maximum exerted force to unbind each of the active flavonoids was at least double (~800pN) the force needed for the inactive ligands.
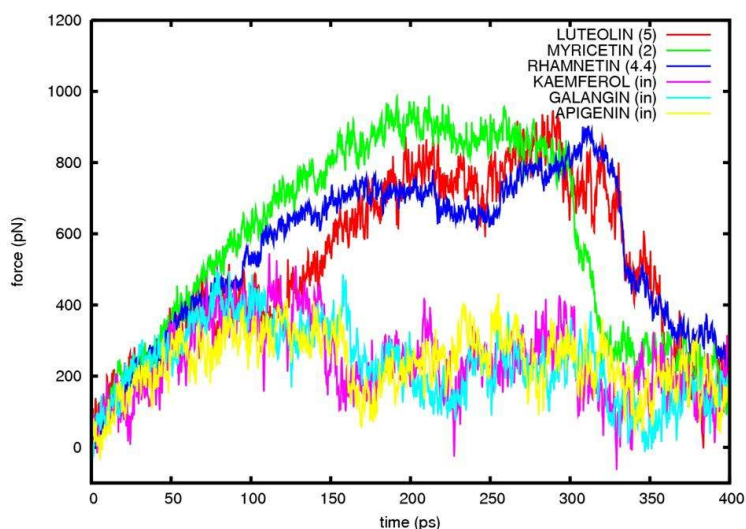
a)
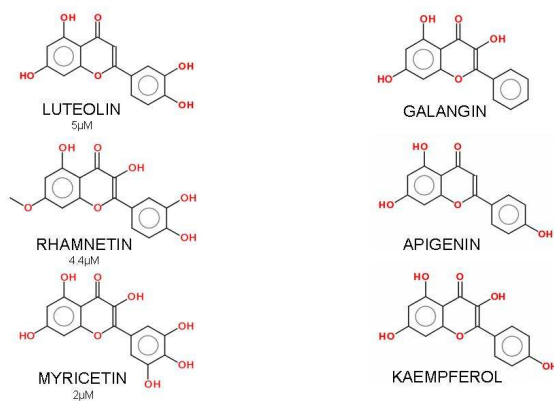


b)



Figure 8. Comparison of force profile of different flavonoids ligands. a) Chemical structures of ligands analyzed. b) Force profiles deriving from pulling the ligands along the unbinding reaction coordinate. For each ligand the plots show the resulting mean values from averaging the force profile from five individual SMD runs.

Whereas it is unlikely to relate the slightly different biological activities of luteolin, myricetin and rhamnetin, to the slightly different magnitude of their force profiles, it is interesting to note how all the inactive compounds were similarly described by the same force profile shape. The inactive kaempferol or apigenin do have one hydroxyl group in the B ring which is pointing in the surrounding of catalytic environment, deep into the entrance tunnel. One would expect for them to show an intermediate magnitude of the force profile respect to the inactive galangin which is totally lacking of substituent on the B ring. However since they are equally not able to establish a productive interaction pattern with PfFabZ, they do equally exhibit a similar force profile. Furthermore, one might deduct that passing from a mono- to a di-hydroxylated B ring could definitely make the difference in terms of fruitful H-bond network interactions.


## Conclusions


Structure activity relationships of a large series of related flavonoids were rationalized using a combined docking and (steered) molecular dynamics approach. MD trajectories of a ligand-protein complex were not only able to suggest the relative stability among different binding modes but allowed to take into account receptor flexibility which was pointed to be critical in FAS-II enzymes. The application of SMD to a ligand-protein system allowed to quantitatively discern between binding modes of luteolin as well as to reconstruct the whole binding event and its free-energy profile. Using SMD-derived forces we were able to pick up active from inactive compounds. In this context, it is worth to mention how the particular features of both the binding site, not extremely buried, and the binding mode, few pivotal interacting moieties, gave us a good case study for testing

our protocol. Further optimization might permit to extend the applicability of SMD-derived forces to discern actives from inactives for several biological relevant targets.

In order to further validate the proposed binding modes and to advance our understanding in ligand-protein interaction, it would be extremely useful keep testing a higher number of luteolin derivatives. Below is reported a short table of molecule whose activity against PfFabZ might strength the validity of the binding model proposed.
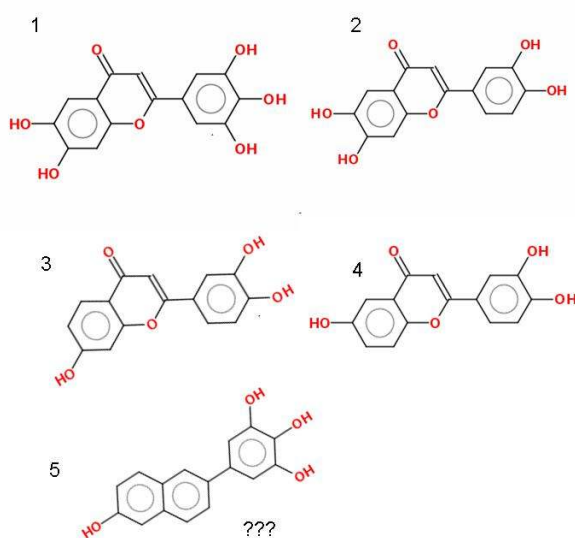


Whereas compound 5 might sound as a naïve provocation, it would be interesting to understand how much farther from the flavonoid scaffold we could go. Compounds 1-4 would be endowed with the capability to satisfy both binding model (A and B). Compounds 3, for instance might allow further quantifying the role of the 3-hydroxyl-group since its structure resembles that of fisetin. Whereas the mono-hydroxylation of the B ring have show to be not enough to confer activity, it would be interesting to test if the inactivity is position-dependent or not.

There might be many other molecules which could be proposed for testing according to the results discussed in this manuscript.

74

**Methodological Details**

The 3D initial coordinates of PfFabZ were taken from the apo x-ray crystal structure stored in the Protein Data Bank (pdb code: 1z6b). Chain A and B were selected to simulate the protein dimer and the two disordered portions, with gaps not exceeding 8 residues, of loops connecting α2 and β3 and between β4 a β5 were rebuilt using standard parameters of the loop modeling routine implemented in MODELLER (ref). The structure of flavonoid ligands was built and geometry optimized using the Sybyl 7.3 molecular modeling suite of program (Tripos Inc., St. Louis, MO). Molecular Docking was carried out using the *default settings* parameters of Gold 3.0.1. The binding site definition included the γC of chain A Phe169 and every residue within 15 Å. For the 5-, 6-, and 7- hydroxy-flavone series ChemScore was used to drive and rank the genetic algorithm search. The preliminary docking calculations on the whole flavonoid series reported Table1 were performed using both GoldScore and ChemScore. Both, geometrical and scoring trends were not observed (see text). Docking poses for luteolin were obtained using both GoldScore and ChemScore and one hundred poses were generated for each scoring function. Docking outcomes were then aclapsterized (clustered by means of the AClAP program) and only one significantly populated cluster was obtained (cardinality of 153). According to the Chauvenet criterion implemented in AClAP, a cluster is significantly populated only if its cardinality is more than twice the standard deviation apart from the average population value for that level of clustering. The second most populated cluster (cardinality of 33) was the one containing the best ranked pose of luteolin.

*MD simulations*: the best ranked docking solutions of top-two populated clusters (poses A and B) were selected and each ligand-protein complex was further investigated by means of molecular dynamics as described below. The complex was solvated with a 8 Å thick layer of water using the solvatebox command of the LEaP program and the electro-neutrality was imposed by equally distributing the excess total charge (+6) over the 4656 atoms of the protein. Periodic boundary conditions were applied and long-range electrostatics were calculated every time step by using the Particle Mesh Ewald (PME) method. A cutoff of 10 Å was used for van der Waals and short-range electrostatic interactions with a smoothing switching function starting at 8 Å. Time integration step of 2 fs was used and the length of all bonds involving hydrogen atoms was fixed using the SHAKE algorithm. The solvated system was minimized for 1000 steps restraining the heavy atoms, except

waters, with a force constant of 5 kcal mol$^{-1}$ Å$^{-2}$, followed by 1000 steps with a force constant of 3 kcal mol$^{-1}$ Å$^{-2}$ applied to the backbone and finally minimizing the unrestrained system for further 1000 steps. At constant volume, the system was heated from 1 to 300K by increments of 50 K every 15ps and correspondingly the α-carbons were gradually unrestrained by lowering the spring constant from 6 to 2 kcal mol$^{-1}$ Å$^{-2}$. The temperature was controlled by the Langevin thermostat with a dumping coefficient of 5 ps$^{-1}$. During the subsequent switching to the isothermal-isobaric (NPT) ensemble, a soft harmonic restraint of 1 kcal mol$^{-1}$ Å$^{-2}$ was still applied to the Cα atoms and gradually turned off in the next 60ps. The Langevin piston method was used to set the target pressure at 1 atm. The system evolved for further 3ns and from the last 300ps were randomly sampled the starting configurations for both unbinding pathway investigation and SMD simulations. All simulations were performed with NAMD 2.6 (Phillips et al., 2005) using the ff99SB AMBER force field for protein and the TIP3P model for all water in the system (Jorgensen et al., 1983). The flavonoid ligands were optimized using the Gussian03 software (Gaussian, Inc. Wallingford, CT) at the B3LYP/6-31G* level of theory and partial atomic charges were assigned using the restricted electrostatic potential fit (RESP) method. The general AMBER force field (GAFF) was used for the ligands and the corresponding topology end parameters files were prepared with the antechamber tool of the AMBER suite of programs.

*Random Expulsion Molecular Dynamic* (REMD) method has been successfully used to disclose putative ligand unbinding pathways in biologically relevant targets.[27, 33] The protocol allows a strongly unbiased search at a relatively low computational cost. A randomly oriented force is applied to the ligand atoms for a defined short amount of time steps $N$. The force has constant magnitude $f$ and accelerates the ligand in the context of the binding pocket. If the ligand encounters hindrance during the route its average velocity will fall below a pre-set threshold, or, in other words, it will not cover in the $N$ time steps the expected distance $r_{min}$. If this is the case, a new direction is chosen randomly and maintained for further $N$ steps, as long as the ligand find a path which allows the coverage of the $r_{min}$ distance. The probability to sample the unbinding event likely depends on the system structural peculiarities, on the type of interactions involved in the ligand-protein complex, and finally on the combination of the adjustable parameters $f$, $N$ and $r_{min}$. For the luteolin/PfFabZ complex we tested three combinations of parameters finding an expulsion rate ranging from 3% to

41% (table below). For each combination, 100 independent trajectories were generated. If no expulsion was observed in the first 50ps, then the run was skipped and considered unsuccessful.

| $f$ (kcal/mol·Å) | $N$ (timesteps) | $r_{min}$ (Å) | expulsion rate (%) | pw1/pw2 (%) |
|---|---|---|---|---|
| 20 | 10 | 0.006 | 41 | 37/4 |
| 15 | 15 | 0.008 | 13 | 13/n.s |
| 10 | 20 | 0.008 | 3 | 3/n.s. |

Vashisth and Abrams have recently implemented Random expulsion MD in NAMD via a *tcl* script interface[27] which we properly modified to fit the needs of our system. The force was applied to all the carbon atoms of the molecule as well as to the intracyclic oxygen. The successful expulsion trajectories were clustered and the components of each cluster were then geometrically averaged to obtain the representative path. Two ligand egress pathways were identified (see text). Preliminary investigations suggested the trend in the expulsion rate ratio between path1 and path2 to be maintained regardless the binding mode of luteolin used as starting configuration.

Comparison of the force profiles among different active and inactive flavonoids were performed using *constant-velocity SMD* with a pulling rate of $5 \times 10^{-5}$ Å/timestep and with a spring constant of 7 kcal/mol Å$^{-2}$. Several pulling velocities were preliminarily tested and the one we chose gave the better balance between resolution among different ligands and simulation time length. The time length of the simulations was 400ps which was sufficient to observe the complete ligand unbinding. The mean force profile for each ligand was obtained by averaging the outcomes of five independent runs. To avoid shifting of the system during pulling, the α-carbons of residues from G186 to I196 and from L207 to N214 on both chains A and B were restrained to their initial position using a spring constant of 1 kcal mol$^{-1}$ Å$^{-2}$. The starting configurations for the luteolin derivatives were obtained by superimposing the flavonoid scaffolds and evolving the resulting ligand-protein complex for further 3ns. The number of water molecules was kept constant for both luteolin binding mode investigation and flavonoids force profile comparison study.

The second order cumulant expansion of the Jarzynski's equality was employed to compute the potential of mean force (PMF) along the unbinding reaction coordinate. Six independent trajectories, each with length of 10ns, were generated for each luteolin binding mode. The value of the exerted force ($F$) was outputted every ($dt$) 1 ps of simulation and the work $W(t) = \int_0^{t'} F(t')v dt'$ done on the

system during the SMD was calculated by numerical integration; the pulling velocity ($v$) was 0.0005 Å/ps ($1 \times 10^{-6}$ Å/timestep). The stiff spring approximation was satisfied by a spring constant of 7 kcal/mol Å$^{-2}$.

## References

1.      Ting, L. M.; Gissot, M.; Coppi, A.; Sinnis, P.; Kim, K., Attenuated Plasmodium yoelii lacking purine nucleoside phosphorylase confer protective immunity. Nat Med 2008, 14, (9), 954-8.

2.      Yu, M.; Kumar, T. R.; Nkrumah, L. J.; Coppi, A.; Retzlaff, S.; Li, C. D.; Kelly, B. J.; Moura, P. A.; Lakshmanan, V.; Freundlich, J. S.; Valderramos, J. C.; Vilcheze, C.; Siedner, M.; Tsai, J. H.; Falkard, B.; Sidhu, A. B.; Purcell, L. A.; Gratraud, P.; Kremer, L.; Waters, A. P.; Schiehser, G.; Jacobus, D. P.; Janse, C. J.; Ager, A.; Jacobs, W. R., Jr.; Sacchettini, J. C.; Heussler, V.; Sinnis, P.; Fidock, D. A., The fatty acid biosynthesis enzyme FabI plays a key role in the development of liver-stage malarial parasites. Cell Host Microbe 2008, 4, (6), 567-78.

3.      Vaughan, A. M.; O'Neill, M. T.; Tarun, A. S.; Camargo, N.; Phuong, T. M.; Aly, A. S.; Cowman, A. F.; Kappe, S. H., Type II fatty acid synthesis is essential only for malaria parasite late liver stage development. Cell Microbiol 2008.

4.      Swarnamukhi, P. L.; Sharma, S. K.; Bajaj, P.; Surolia, N.; Surolia, A.; Suguna, K., Crystal structure of dimeric FabZ of Plasmodium falciparum reveals conformational switching to active hexamers by peptide flips. FEBS Lett 2006, 580, (11), 2653-60.

5.      Kostrewa, D.; Winkler, F. K.; Folkers, G.; Scapozza, L.; Perozzo, R., The crystal structure of PfFabZ, the unique beta-hydroxyacyl-ACP dehydratase involved in fatty acid biosynthesis of Plasmodium falciparum. Protein Sci 2005, 14, (6), 1570-80.

6.      Zhang, L.; Liu, W.; Hu, T.; Du, L.; Luo, C.; Chen, K.; Shen, X.; Jiang, H., Structural basis for catalytic and inhibitory mechanisms of beta-hydroxyacyl-acyl carrier protein dehydratase (FabZ). J Biol Chem 2008, 283, (9), 5370-9.

7.      Kimber, M. S.; Martin, F.; Lu, Y.; Houston, S.; Vedadi, M.; Dharamsi, A.; Fiebig, K. M.; Schmid, M.; Rock, C. O., The structure of (3R)-hydroxyacyl-acyl carrier protein dehydratase (FabZ) from Pseudomonas aeruginosa. J Biol Chem 2004, 279, (50), 52593-602.

8.      Leesong, M.; Henderson, B. S.; Gillig, J. R.; Schwab, J. M.; Smith, J. L., Structure of a dehydratase-isomerase from the bacterial pathway for biosynthesis of unsaturated fatty acids: two catalytic activities in one active site. Structure 1996, 4, (3), 253-64.

9.    Zhang, L.; Kong, Y.; Wu, D.; Zhang, H.; Wu, J.; Chen, J.; Ding, J.; Hu, L.; Jiang, H.; Shen, X., Three flavonoids targeting the beta-hydroxyacyl-acyl carrier protein dehydratase from Helicobacter pylori: crystal structure characterization with enzymatic inhibition assay. Protein Sci 2008, 17, (11), 1971-8.

10.   Tasdemir, D.; Lack, G.; Brun, R.; Ruedi, P.; Scapozza, L.; Perozzo, R., Inhibition of Plasmodium falciparum fatty acid biosynthesis: evaluation of FabG, FabZ, and FabI as drug targets for flavonoids. J Med Chem 2006, 49, (11), 3345-53.

11.   Sharma, S. K.; Kapoor, M.; Ramya, T. N.; Kumar, S.; Kumar, G.; Modak, R.; Sharma, S.; Surolia, N.; Surolia, A., Identification, characterization, and inhibition of Plasmodium falciparum beta-hydroxyacyl-acyl carrier protein dehydratase (FabZ). J Biol Chem 2003, 278, (46), 45661-71.

12.   Huang, N.; Jacobson, M. P., Physics-based methods for studying protein-ligand interactions. Curr Opin Drug Discov Devel 2007, 10, (3), 325-31.

13.   Alonso, H.; Bliznyuk, A. A.; Gready, J. E., Combining docking and molecular dynamic simulations in drug design. Med Res Rev 2006, 26, (5), 531-68.

14.   Cavalli, A.; Bottegoni, G.; Raco, C.; De Vivo, M.; Recanatini, M., A computational study of the binding of propidium to the peripheral anionic site of human acetylcholinesterase. J Med Chem 2004, 47, (16), 3991-9.

15.   Chang, C. E.; Chen, W.; Gilson, M. K., Ligand configurational entropy and protein binding. Proc Natl Acad Sci U S A 2007, 104, (5), 1534-9.

16.   Deng, Y.; Roux, B., Computations of Standard Binding Free Energies with Molecular Dynamics Simulations. J Phys Chem B 2009.

17.   Kozakov, D.; Schueler-Furman, O.; Vajda, S., Discrimination of near-native structures in protein-protein docking by testing the stability of local minima. Proteins 2008, 72, (3), 993-1004.

18.   Xiang, Z.; Soto, C. S.; Honig, B., Evaluating conformational free energies: the colony energy and its application to the problem of loop prediction. Proc Natl Acad Sci U S A 2002, 99, (11), 7432-7.

19.   Lee, J.; Seok, C., A statistical rescoring scheme for protein-ligand docking: Consideration of entropic effect. Proteins 2008, 70, (3), 1074-83.

20.   Jarzynski, C., Rare events and the convergence of exponentially averaged work values. Phys Rev E Stat Nonlin Soft Matter Phys 2006, 73, (4 Pt 2), 046105.

21.   Jensen, M. O.; Park, S.; Tajkhorshid, E.; Schulten, K., Energetics of glycerol conduction through aquaglyceroporin GlpF. Proc Natl Acad Sci U S A 2002, 99, (10), 6731-6.

22.   Minh, D. D.; McCammon, J. A., Springs and speeds in free energy reconstruction from irreversible single-molecule pulling experiments. J Phys Chem B 2008, 112, (19), 5892-7.

23.   Park, S.; Schulten, K., Calculating potentials of mean force from steered molecular dynamics simulations. J Chem Phys 2004, 120, (13), 5946-61.

24. Sotomayor, M.; Schulten, K., Single-molecule experiments in vitro and in silico. Science 2007, 316, (5828), 1144-8.

25. Ludemann, S. K.; Lounnas, V.; Wade, R. C., How do substrates enter and products exit the buried active site of cytochrome P450cam? 1. Random expulsion molecular dynamics investigation of ligand access channels and mechanisms. J Mol Biol 2000, 303, (5), 797-811.

26. Liphardt, J.; Dumont, S.; Smith, S. B.; Tinoco, I., Jr.; Bustamante, C., Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski's equality. Science 2002, 296, (5574), 1832-5.

27. Vashisth, H.; Abrams, C. F., Ligand escape pathways and (un)binding free energy calculations for the hexameric insulin-phenol complex. Biophys J 2008, 95, (9), 4193-204.

28. Chelli, R.; Procacci, P., A potential of mean force estimator based on nonequilibrium work exponential averages. Phys Chem Chem Phys 2009, 11, (8), 1152-8.

29. Forney, M. W.; Janosi, L.; Kosztin, I., Calculating free-energy profiles in biomolecular systems from fast nonequilibrium processes. Phys Rev E Stat Nonlin Soft Matter Phys 2008, 78, (5 Pt 1), 051913.

30. Zhang, D.; Gullingsrud, J.; McCammon, J. A., Potentials of mean force for acetylcholine unbinding from the alpha7 nicotinic acetylcholine receptor ligand-binding domain. J Am Chem Soc 2006, 128, (9), 3019-26.

31. Woo, H. J.; Roux, B., Calculation of absolute protein-ligand binding free energy from computer simulations. Proc Natl Acad Sci U S A 2005, 102, (19), 6825-30.

32. Isralewitz, B.; Baudry, J.; Gullingsrud, J.; Kosztin, D.; Schulten, K., Steered molecular dynamics investigations of protein function. J Mol Graph Model 2001, 19, (1), 13-25.

33. Winn, P. J.; Ludemann, S. K.; Gauges, R.; Lounnas, V.; Wade, R. C., Comparison of the dynamics of substrate access channels in three cytochrome P450s reveals different opening mechanisms and a novel functional role for a buried arginine. Proc Natl Acad Sci U S A 2002, 99, (8), 5361-6.

## Concluding remarks

The elongation of the acyl chain in FAS II is catalyzed by four key enzymes, FabB/F, FabG, FabZ and FabI and for each elongation cycle the chain length is increased by two carbon units. Structural studies have shown that the binding core of ACP is able to expand its central cavity in order to accommodate the growing acyl chain.[1] Comparison between *apo* (no ligand bound) and *holo* (ligand bound) structures have shown that a highly conserved phenylalanine residue might assist the harbouring of the acyl chain by switching with different conformers.[2, 3] Steered molecular dynamics simulations discussed in this dissertation have induced the exposure (delivery) of the β-hydroxyacyl substrate. During the trajectory, (along Path 1b) the egress of the acyl chain was supported by Phe29. As shown the Figure 1, the side chain of Phe29, which is highly conserved in ACPs, underwent a conformational change in order to assist the outgoing substrate; with this movement, Phe29 acted as lid maintaining the cavity core water-free and therefore permitting the rapid relocation of the substrate after the catalytic cycle. The avoided solvation of the core would also increase the stability of the ACP folding during the delivery process.
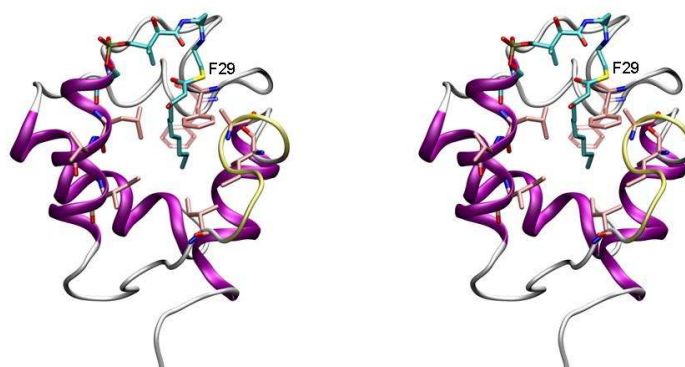


Figure 1. The role of F29 in assisting the outgoing substrate. The starting configuration, before the conformational switch, of F29 is in transparent pink.

Assuming *Pf*FabZ as the enzymatic counterpart receiving the delivered β-hydroxyacyl substrate, then a structural modification is likely to occur after the substrate delivery by ACP. Phe169 has shown high mobility both in experimental and theoretical studies. Its conformational switches reshaped the tunnel of the active site of FabZ in a remarkable matter (Figure 2).
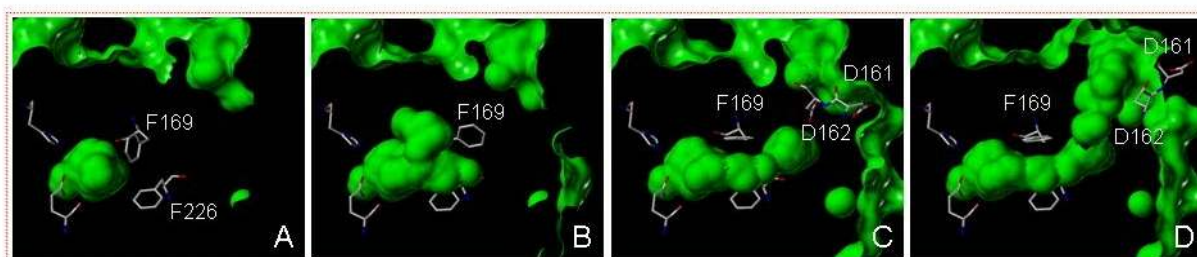


Figure 2. Role of F169 in shaping the binding pocket of Pf FabZ. The green surfaces represent the accessible volume of the binding tunnel.

While it is still under debate whether or not the acyl substrate has to be completely delivered to the enzymatic counterparts, we suppose that the role of Phe169 might be equivalent to that of Phe29 in ACP. That is, allow substrates of different chain length to be optimally accommodated in the binding tunnel. Phe169 was also observed to influence the packing of the active site region. The mutation of any residue of the so-called "phenylalanine triad" into Glycine reduced the thermal stability of the mutant respect to the wild type.[4] Different conformations of Phe69 may affect the packing of the active site and modulate the "breath" (i.e. fluctuations of the α2-α3 loop) of the structure. Fluctuation of the α2-α3 loop may also affect the size of the ACP binding grow surrounding FabZ active site. In particular, a wider cavity has a bigger attraction on ACP.[5] In virtue of the ACP/FabZ binding model I discussed earlier in this dissertation; one might also

hypothesize a role of the loop in facilitating the transfer (delivery) of the substrate from the core of the carrier into the enzyme active site.

The critical role associated with Phenylalanine residues in ACP and FabZ enzymes was our prompt to further investigate the structure of enzymes involved in FAS-II elongation cycle. *Pf*FabI, a NADH-dependent reductase, is probably the more investigated enzyme among the biosynthetic pathway. Several crystal structure are available both in binary (with cofactor) and in ternary (with cofactor and inhibitor) complexes. The ascertainment of *Pf*FabI structures revealed that, while the binding mode of the ligands is conserved, the protein structure differs for one major component in the surrounding of the active site: two different populations of F368 side chain were observed (Figure 3). The consequence is that *Pf*FabI active site can assume a pocket-like as well as a tunnel-like shape. The switch in the active site shape is fully controlled by Phe368 side chain.
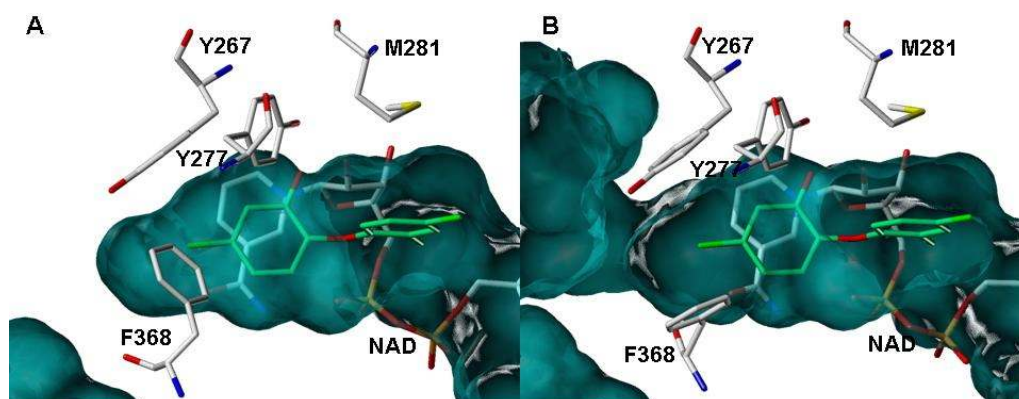


Figure 3. Active site of PfFabI. A) F368 closes the pocket (pdb code: 1nhg). B) F368 is turned up and the active site pocket is in the extended form (pdb code 1zsn). In both pictures triclosan (from 1nhg) is shown with carbon atoms coloured in green.

Whereas a large scale investigation should be carried out before lay down any general conclusion, this preliminary observation shows that phenylalanine residues laudly claim their key

role in active site dynamics. This might be interpreted as follow: Nature provided cells with a plenty of fatty acid bearing different saturation level and especially with different chain length. Phenylalanine residues might have been privileged by Nature in the accomplishment of modulating the shape of fatty acid binging pockets. Phenylalanine has the intrinsic capabilities to act as "doorkeeper" as well as "tailor" of binding pockets. Its fluctuation among few conformers can generate sized tunnel from apparently small cavities. Protein entrances can be generated by switching with only two conformers. Apparently, Nature exploited Phenylalanine to optimize the biosynthesis of fatty acids. We might thereof image a universal role of Phe at fatty acid binding site in order to adapt the shape of the pocket to the size of the fatty acid which is going to be metabolized.

What is the relevance for drug discovery? The intrinsic plasticity borne by phenylalanine might be exploited to identify new lead compounds in the context of fatty acid binding proteins. Below is shown how the increased pocket size of *Pf*FabI might be employed to address the design of "extended" ligands which might be able to cover the receptor space generated by the conformational switches of phenylalanine residues.
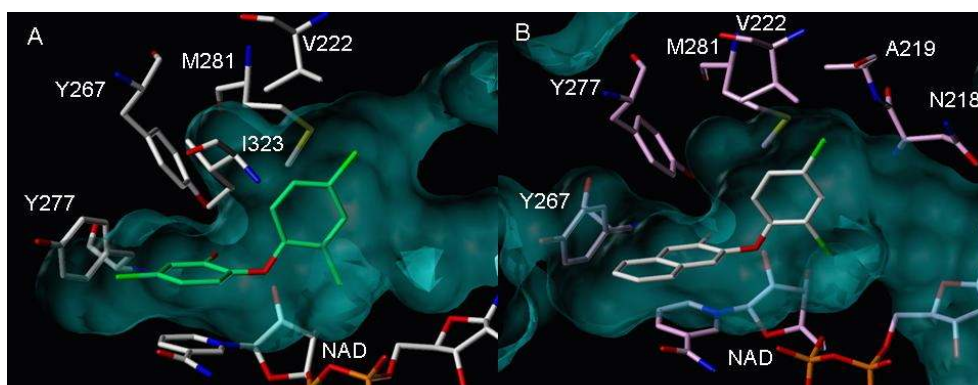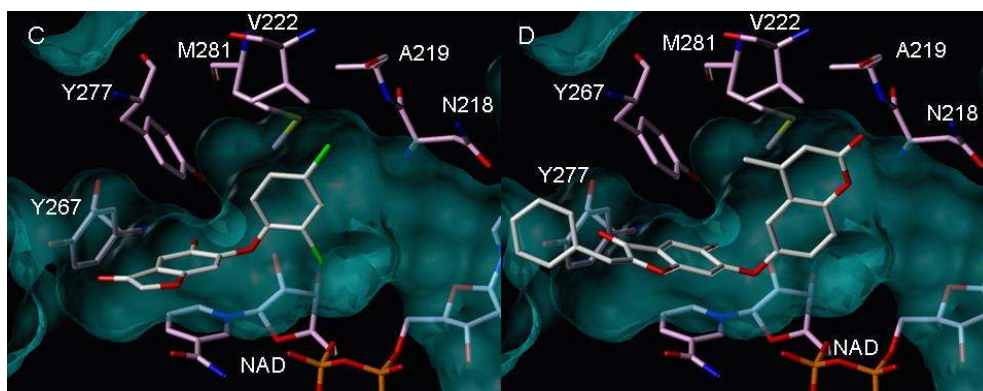
A) Triclosan into the active site of *Pf*FabI. B, C, D) Putative ligands able to take advances of the conformational switch of F368 (not shown for sake of clarity).

I conclude this dissertation highlighting the value of staring at the same target from different standpoints. The first would be the ambitious perspective of the "evolution", the one that Mother Nature has. It might allow identifying peculiar target properties, interpreting structure-function relationships. The second is the more affordable, yet difficult, medicinal chemist standpoint which would be able to take advances of the intrinsic properties of a target (or of a biological system) to rationally design more effective ligands.

### References

1.    Roujeinikova, A.; Simon, W. J.; Gilroy, J.; Rice, D. W.; Rafferty, J. B.; Slabas, A. R., Structural studies of fatty acyl-(acyl carrier protein) thioesters reveal a hydrophobic binding cavity that can expand to fit longer substrates. *J Mol Biol* **2007,** 365, (1), 135-45.

2.    Roujeinikova, A.; Baldock, C.; Simon, W. J.; Gilroy, J.; Baker, P. J.; Stuitje, A. R.; Rice, D. W.; Rafferty, J. B.; Slabas, A. R., Crystallization and preliminary X-ray crystallographic studies on acyl-(acyl carrier protein) from Escherichia coli. *Acta Crystallogr D Biol Crystallogr* **2002,** 58, (Pt 2), 330-2.

3.    Roujeinikova, A.; Baldock, C.; Simon, W. J.; Gilroy, J.; Baker, P. J.; Stuitje, A. R.; Rice, D. W.; Slabas, A. R.; Rafferty, J. B., X-ray crystallographic studies on butyryl-ACP reveal flexibility of the structure around a putative acyl chain binding site. *Structure* **2002,** 10, (6), 825-35.

4.    Rocha, A. P. Travail de Diplome. University of Geneva, 2007.

5.    Zhang, L.; Liu, W.; Hu, T.; Du, L.; Luo, C.; Chen, K.; Shen, X.; Jiang, H., Structural basis for catalytic and inhibitory mechanisms of beta-hydroxyacyl-acyl carrier protein dehydratase (FabZ). *J Biol Chem* **2008,** 283, (9), 5370-9.