Atomic-Level Characterization of the Chain-Flipping Mechanism in Fatty-Acids Biosynthesis

Francesco Colizzi,[*,†,¶] Matteo Masetti,[†] Maurizio Recanatini,[†] and Andrea Cavalli[†,‡]

†*Department of Pharmacy and Biotechnology, Alma Mater Studiorum-University of Bologna, via Belmeloro 6, 40126 Bologna, Italy*

‡*CompuNet, Istituto Italiano di Tecnologia, via Morego 30, 16163 Genova, Italy*

¶*Current address: Institute for Research in Biomedicine (IRB Barcelona) Carrer Baldiri Reixac 10, 08028, Barcelona, Spain*

E-mail: cecio.colizzi@gmail.com

SUPPLEMENTARY TEXT

SUPPLEMENTARY INFORMATION PROVIDED:

Supplementary Methods

Molecular Models. The initial 3D coordinates of *Pf*FabZ were retrieved from the Protein Data Bank (*apo* structure; pdb code: 1z6b).[1] Chains A and B were selected to simulate the protein dimer and the disordered portions were rebuilt using standard parameters of the loop modeling routine implemented in MODELLER v7.0.[2]

The first NMR conformer (out of 20) of *holo Pf*ACP (pdb code: 2fq0)[3] was used to model ACP structure. The substrate was covalently bound to Ser 37 via the 4´-phosphopantetheine (4'-PP) prostetic group. 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 pose of the high resolution X-ray structure of E. coli decanoyl-ACP (pdb code: 2fae).[4]

Protein-Protein docking. We used the combination of two docking algorithms to elucidate the 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 surroundings of the top-ranked outcomes generated by the global search and a more accurate scoring function is generally used.

Here we used the PatchDock server[5–7] with default settings for the global search of the configurational space between ACP and FabZ. Given two interacting counterparts, PatchDock finds and ranks docking transformations according to shape-complementarity criteria. The energy landscape around each of the 10 best-scoring PatchDock poses was further explored generating 1000 decoys with the local-refinement protocol of RosettaDock v3.1 server in which a detailed energy function is combined with side chain flexibility.[8,9] RosettaDock's local perturbation includes ± 3 Å in the direction between the two proteins, 8 Å in the directions sliding the proteins relative to each other along their surfaces, 8° of tilt of the proteins,

and a complete 360° spin around the axis between the centers of the two proteins. Out of the 10 best-ranked PatchDock poses, only one pose (the 4th) showed the typical energy funnel after local refinement. The funnel is related to the presence of a stable energy minimum surrounded by a broad region of attraction.[10]

System and Simulations set up. The ACP-FabZ complex was solvated with a 10 Å thick layer of water using the *solvate* package of VMD[4] and one Cl- anion was added to neutralize the system. 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 a further 1000 steps. Periodic boundary conditions were applied and long-range electrostatic interactions were calculated every time step using the particle mesh Ewald method.[4] 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 constrained using the SHAKE algorithm.[11] In the NPT ensemble, the system was heated from 1 up to 300 K by increments of 50 K every 15 ps 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 and the Nosé-Hoover Langevin piston was used to set the target pressure at 1 atm. The system was further equilibrated in the isothermal−isobaric ensemble for 30 ns fully unrestrained. The final snapshot was used as starting configuration for both the definition of the preassigned delivery pathway by targeted MD and the metadynamics production run.

All simulations were performed with the molecular dynamics program NAMD v2.6[12] patched with the free-energy plugin PLUMED v1.2.1,[13] using the CHARMM22 force field[14,15] and the TIP3P water model.[16] CHARMM parameters for the prosthetic group together with the β-hydroxydecanoyl substrate were generated as described previously by Colizzi *et al.*[17]

Definition of the reference path. The mechanism of the enzymatic reaction catalyzed by β-hydroxyacyl dehydratases in fatty acid biosynthesis (FAS-II) was used as a primer for the definition of the reference delivery path. In particular, the analysis of *E.coli* FabA crystal structure covalently bound to 3-decynoyl-N-acetylcysteamine inhibitor (pdb access code: 1mka)[18] via the catalytic His 70 (*E.coli*, numeration) has provided the structural details of the mechanism-based inhibition of the dehydratase. That is, the requirement of the close proximity between the nucleophile Nϵ2 of the catalytic Histidine (133 in *Pf*) and the polarized α-C-H bond of the substrate (as shown in Figure 5b of the manuscript). For the enzymatic reaction to take place, it has been argued that the reactivity of the substrate would be enhanced by the hydrogen bond of the thioester carbonyl group (that becomes positively polarized) to the main chain nitrogen of Gly 142 and by the N-terminal positive dipole moment of helix α2 (also shown in Figure 5b of the manuscript). Moreover, the α-C-H bond is positively polarized by His 133' and the β-C-OH bond is positively polarized by Glu 147, thus lowering the activation energy for dehydration.[1,18,19]

Accordingly, we exploited this enzymatic/mechanicistic information to perform a 10 ns targeted-MD (TMD, k=100 kcal mol$-$1 Å$-$2) simulation and drive the α-C-H and the thioester carbonyl group at interacting distances from FabZ. Namely, the α-C-H was driven from its initial position within the ACP core to 3.8 Å from Nϵ2 of His 133(FabZ), and the thioester carbonyl oxygen to 2.6 Å from the N of Gly 142(FabZ). The position of the acyl chain was then manually docked in the fatty acid binding site of *Pf*FabZ using as reference the binding mode of the acyl chain shown in the FabA/3-decynoyl-NAC crystal structure.[18] The resulting complex with the β-hydroxydecanoyl substrate within FabZ was minimized and further equilibrated for 10 ns keeping the restrains on the targeted distances. The position of carbon atoms of the acyl chain were then used as reference and included in the targeted-atom subset and used to run another 10 ns TMD simulation. The trajectory was discretized and a preliminary path produced optimizing the distance between 20 frames as described in reffs: 20, 21. Finally, the discretized path was used to run a Steered MD (SMD)

simulation in the $s(q)$ CV space, where q are the microscopic coordinates of the system. In 20 ns, the CV value of $s(q)$ was steered (k=100 kcal mol−1 units−2) from 1 to 20 while $z(q)$ was restrained to a maximum value of 9 Å^2. The resulting delivery trajectory was further analyzed and the choice of the frames optimized as described in reffs: 20, 21. Eventually, 20 frames were selected which were sufficient to describe the chain-flipping mechanism with the necessary resolution using metadynamics.

Metadynamics simulations. Metadynamics is an established method aimed at accelerating molecular dynamics simulations and subsequently recovering free-energy landscapes.[20,21]

Having generated M=20 frames (or nodal points $\{\tilde{q}(i)\}$) from SMD simulations (see the previous section) representative of the chain-flipping transition, a variable that allows one to map the conformations along the ideal progress along the path can then be built as:[20,22]

$$s(q) = \frac{1}{M-1} \frac{\sum_{i=1}^{M} (i-1) \exp(-\lambda ||q - \tilde{q}(i)||)}{\sum_{i=1}^{M} \exp(-\lambda ||q - \tilde{q}(i)||)}$$

where i is a discrete index ranging from 1 to M, and $||q - \tilde{q}(i)||$ is the distance between the current molecular dynamics configuration and the i$^{\text{th}}$ node. The parameterλ is chosen according to the average inter-frame distance and in our metadynamics simulations we used λ=0.8 Å^{-2}.

In this metric, a complementary variable can also be used to define the distance from the closest node of the path itself:

$$z(q) = -\frac{1}{\lambda} \ln \textstyle\sum_{i=1}^{M} \exp(-\lambda ||q - \tilde{q}(i)||)$$

Taken together, $s(q)$ and $z(q)$ can define the position of any microscopic configuration q in the CVs space and allowed us to map the free-energy landscape of the chain-flipping mechanism also exploring regions far from the originally assigned path. In this metric, $||q - \tilde{q}(i)||$ was calculated as mean square displacement of the carbon atoms of the acyl moiety plus the thioester carbonyl and the β-hydroxy group of the substrate after optimal alignment of the α-carbons of the main α2 helices (residues 142-159) of FabZ homodimer

(Figure S1). Several preliminary metadynamics simulations were performed to define the optimal subset of atoms representative of the substrate transition with the best resolution and computational performance. To optimize the efficiency of sampling while preserving a non-local exploration of the mechanism, $z(q)$ was restrained to a maximum value of 9 Å^{2}by a wall potential with k=300 kcal mol−1 Å−4.

With the above definition of the path CV, we ran a cumulative time of 400 ns of metadynamics simulations. As metadynamics accelerates the sampling of relevant regions of the configurational space, it is worth to mention that this time does not have a direct physical meaning. The height of the Gaussians was set to 0.2 kcal/mol and the addition frequency to 1 ps. The width of the Gaussians was set to 0.1 for both s and z.

In the metadynamics simulations with the restrains on the α3 helix of ACP, a wall potential with k=300 kcal mol−1 Å−4 was used to restrain the RMSD of the α-carbons of residues 56-63 to a maximum value of 2 Å. The restrained metadynamics was performed for 96 ns and the free-energy profile was compared to the one obtained from equivalent time of unrestrained simulations.

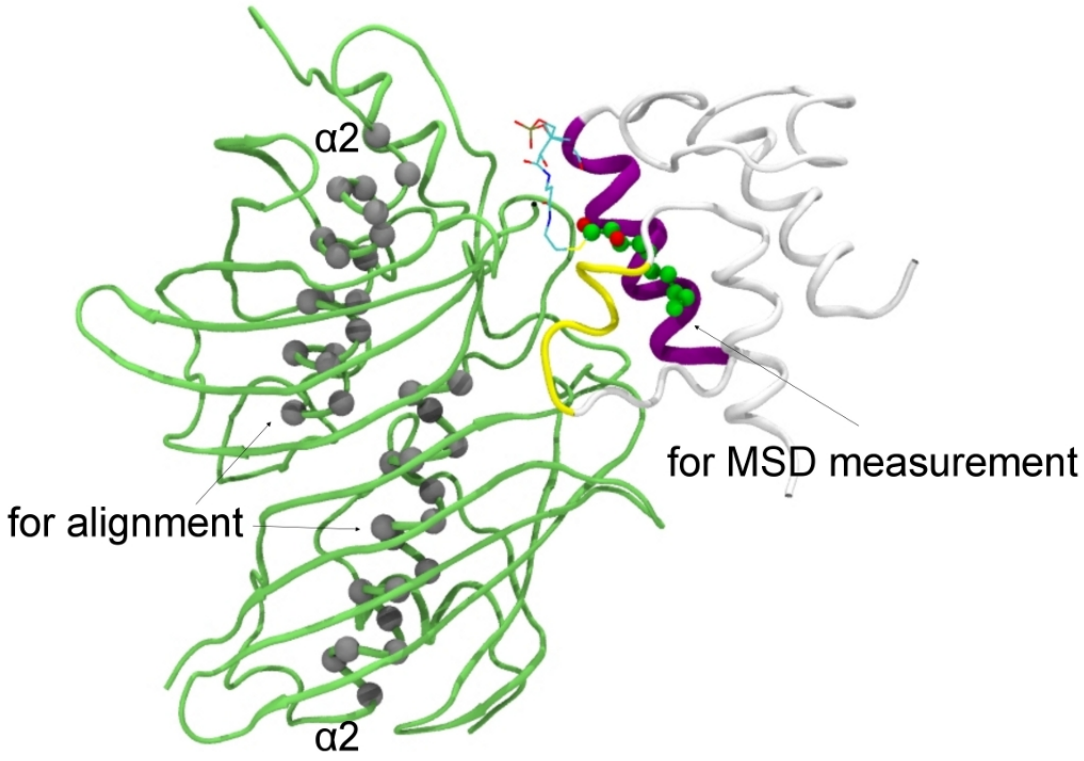

Figure S1: Path collective variable. Graphical representation of the subset of atoms used for the structural alignment (gray spheres) and for the measurement (green and red spheres in the substrate) of mean square displacement along the path (see the supporting text for details).

Supplementary Discussion

Comparison of *Pf* ACP/FabZ docking complex with *Ec* ACP/FabA cocrystal structure. The third step in chain elongation during fatty acid biosynthesis in *Plasmodium falciparum* (*Pf*) is carried out by the unique β-hydroxyacyl-ACP dehydratase (*Pf*FabZ) which corresponds to the primary dehydratase (*Ec*FabZ) participating in fatty acid biosynthesis of *E. coli*. A second dehydratase in *E. coli* (*Ec*FabA) is involved in the dehydration and subsequent isomerization of β-hydroxydecanoyl-ACP to cis-3-decenoyl-ACP.[1,18,23] The residues of the active site of *Ec*FabA and *Ec*FabZ are essentially identical.[18]

In a recent work, Nguyen *et al*[19] have elegantly solved the *Ec* ACP/FabA cocrystal structure by successfully crosslinking ACP with the FabA dehydratase using a sulfonyl 3-alkynyl pantetheinamide probe.[19] Based on the insight from the *Ec* Acp/FabA complex structure and multiple-sequence alignment, **Nguyen *et al* have also observed that the arginine/lysine positive patch of FabZ enzymes is shifted when compared to FabA**. **The authors have argued that differences in the positive patch between FabZ and FabA may result in a different docking interface of FabZ relative to ACP**, and that the different binding mode may prevent the isomerization activity of FabZ—activity that in turn is needed to generate the reactive crosslinking intermediate. **Notably, the sulfonyl 3-alkynyl pantetheinamide probe crosslinking to *Ec*FabA was indeed not active with the *Ec*FabZ.**[19]

Accordingly, we found that the binding mode of *Pf* ACP/FabZ is slightly different from the one observed for the *Ec* Acp/FabA complex. Notably, the orientation of the α2-α3 ACP region toward the active site of the enzyme is conserved in both ACP/enzyme complexes. However, whereas the interacting residues on ACP are virtually identical to those observed in the *Ec* Acp/FabA complex, the *Pf*FabZ interacting interface changes (Figure S2). Positively charged residues at the interacting interface of *Pf*FabZ included Lys 94, Lys 137, Lys 181, Lys 216, and Arg 99 (highlighted by green rectangles in Figure S2c, left side). While at the

Ec Acp/FabA interface the reported positive patch included Arg 132, Arg 136, Arg 137 and Lys 161 (highlighted by green rectangles in Figure S2c, right side), which align with *Pf*FabZ's Ser 197, Ser 200, Ser 201, and Thr 225, respectively (highlighted by gray rectangles in Figure S2c, left side).[19]

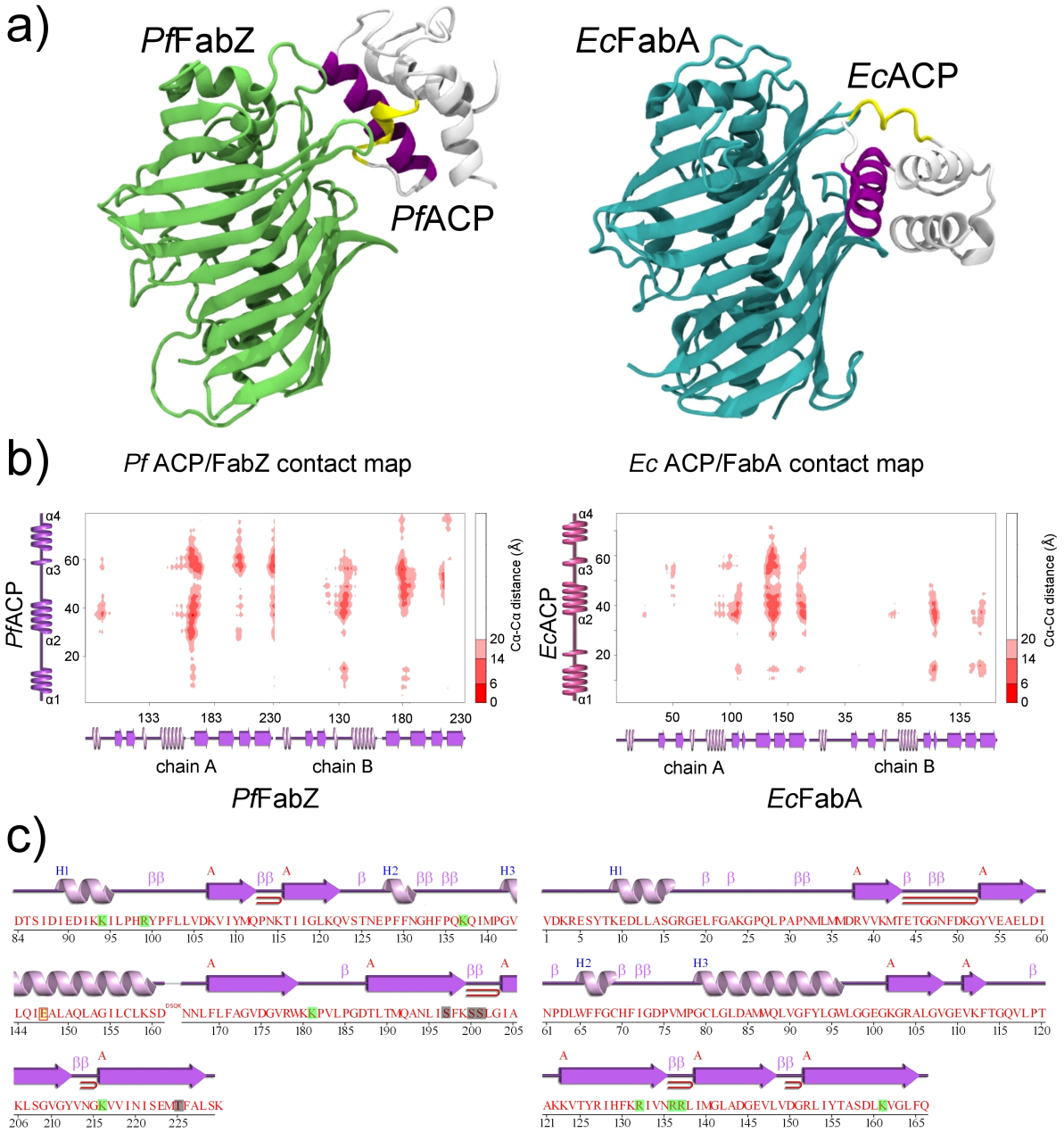

Figure S 2: Comparison of binding modes between *Pf* ACP/FabZ and *Ec* ACP/FabA. a) Structural model of *Pf* ACP/FabZ interaction (left) and the crystal structure of *Ec* ACP/FabA (right, PDB code: 4keh). The ACP's $\alpha 2$ (in violet) and $\alpha 3$ (in yellow) helices at the protein-protein interface are highlighted. b) Contact map for the interactions in *Pf* ACP/FabZ (left) and *Ec* ACP/FabA (right). The secondary structure of the proteins is schematized on the axes; *Pf*FabZ and *Ec*FabA are homodimers. c) Sequence and secondary structure of *Pf*FabZ (left) and *Ec*FabA (right); the arginine/lysine positive patches are highlighted by green rectangles; in gray rectangles are the residues of *Pf*FabZ corresponding to the positive patch in *Ec*FabA according to the multiple sequence alignment of Nguyen *et al.*[19] The images were obtained from the PDBsum server.[24–26]

Details on the Transition State (TS). We identified the movement of the β-hydroxy group—together with the 4'-PP arm—through the α2-β3 loop of FabZ as the molecular determinant of the TS observed on the free-energy landscape (Figure S3). The transition state, together with the neighboring free-energy basins, describe the event in which the reactive center of the substrate actually flips from ACP to FabZ. The α2-β3 loop is a flexible region that shapes the entrance of FabZ active site.[1] At the interface with ACP, residues Leu168, Phe169 and Leu170 form a preferential hydrophobic route for the sleeving of the substrate into FabZ active site (Figure S3). An open question in ACP-dependent enzymatic reactions is how the product can return to the ACP cavity to permit further modifications of the acyl chain along the metabolic pathway.[27] Given the transition state scenario described above, the FabZ-dependent dehydration of the β-hydroxyacyl substrate might facilitate the return of the reaction product (the more rigid and lipophilic *trans*-2-acyl) to the ACP core by exploiting the aforementioned hydrophobic route with a lower energy barrier. Mutations in the aminoacids of the α2-β3 loop of FabZ might regulate the enzyme kinetics by modifying the natural breathing of the protein and modulating the access of β-hydroxylated substrates to the active site.

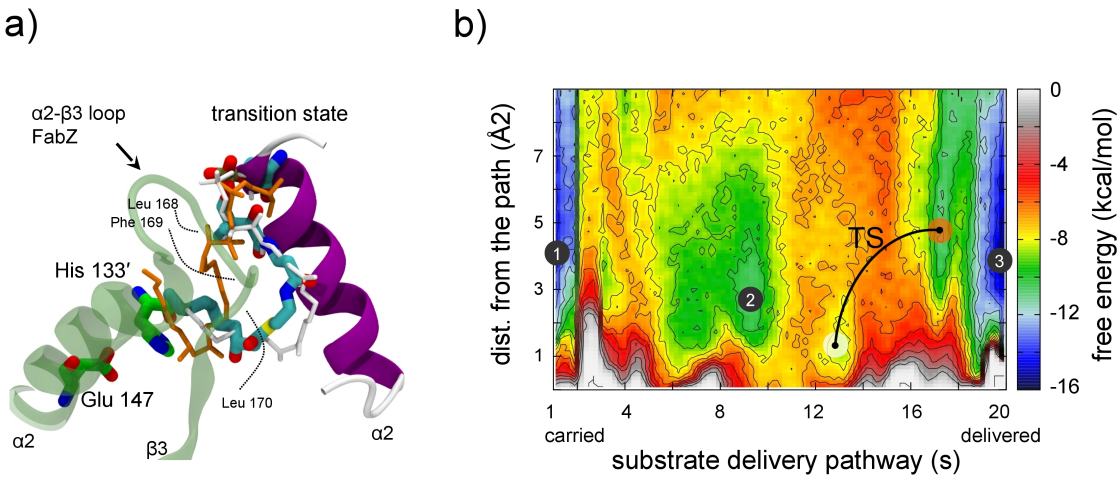

Figure S3: a) Structural details of the transition state (TS) during the chain-flipping mechanism. The carbon atoms of the substrate are in cyan; the white and orange representations of the substrate correspond to the energy basin before and after the TS highlighted in panel *b*.

References

(1) Kostrewa, D.; Winkler, F. K.; Folkers, G.; Scapozza, L.; Perozzo, R. The crystal structure of PfFabZ, the unique β-hydroxyacyl-ACP dehydratase involved in fatty acid biosynthesis of Plasmodium falciparum. *Protein Sci.* **2005**, *14*, 1570–1580.

(2) Fiser, A.; Do, R. K. G.; Šali, A. Modeling of loops in protein structures. *Protein Sci.* **2000**, *9*, 1753–1773.

(3) 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.

(4) 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*, 135–145.

(5) Duhovny, D.; Nussinov, R.; Wolfson, H. J. *Algorithms in bioinformatics*; Springer, 2002; pp 185–200.

(6) Schneidman-Duhovny, D.; Inbar, Y.; Nussinov, R.; Wolfson, H. J. PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res.* **2005**, *33*, W363–W367.

(7) Mashiach, E.; Schneidman-Duhovny, D.; Peri, A.; Shavit, Y.; Nussinov, R.; Wolfson, H. J. An integrated suite of fast docking algorithms. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 3197–3204.

(8) Lyskov, S.; Gray, J. J. The RosettaDock server for local protein-protein docking. *Nucleic Acids Res.* **2008**, *36*, W233–W238.

(9) Gray, J. J.; Moughon, S.; Wang, C.; Schueler-Furman, O.; Kuhlman, B.; Rohl, C. A.; Baker, D. Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *J. Mol. Biol.* **2003**, *331*, 281–299.

(10) 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*, 993–1004.

(11) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.

(12) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802.

(13) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. et al. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comput. Phys. Commun.* **2009**, *180*, 1961–1972.

(14) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S. et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(15) MacKerell, A. D.; Feig, M.; Brooks, C. L. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **2004**, *25*, 1400–1415.

(16) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. F.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926.

(17) 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*, 2289–2293.

(18) 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*, 253–264.

(19) Nguyen, C.; Haushalter, R. W.; Lee, D. J.; Markwick, P. R. L.; Bruegger, J.; Caldara-Festin, G.; Finzel, K.; Jackson, D. R.; Ishikawa, F.; O'Dowd, B. et al. Trapping the dynamic acyl carrier protein in fatty acid biosynthesis. *Nature* **2014**, *505*, 427–431.

(20) Bussi, G.; Branduardi, D. Free-Energy Calculations with Metadynamics: Theory and Practice. *Reviews in Computational Chemistry Volume 28* **2015**, 1–49.

(21) Cavalli, A.; Spitaleri, A.; Saladino, G.; Gervasio, F. L. Investigating Drug–Target Association and Dissociation Mechanisms Using Metadynamics-Based Algorithms. *Acc. Chem. Res.* **2014**, *48*, 277–285.

(22) Branduardi, D.; Gervasio, F. L.; Parrinello, M. From A to B in free energy space. *J. Chem. Phys.* **2007**, *126*, 054103.

(23) 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.

(24) Laskowski, R. A. PDBsum: summaries and analyses of PDB structures. *Nucleic Acids Res.* **2001**, *29*, 221–222.

(25) Laskowski, R. A.; Chistyakov, V. V.; Thornton, J. M. PDBsum more: new summaries and analyses of the known 3D structures of proteins and nucleic acids. *Nucleic Acids Res.* **2005**, *33*, D266–D268.

(26) de Beer, T. A.; Berka, K.; Thornton, J. M.; Laskowski, R. A. PDBsum additions. *Nucleic Acids Res.* **2014**, *42*, D292–D296.

(27) Cronan, J. E. The chain-flipping mechanism of ACP (acyl carrier protein)-dependent enzymes appears universal. *Biochem. J.* **2014**, *460*, 157–163.