▌**Protein Dynamics** │*Hot Paper*│

The Influence of Chemical Change on Protein Dynamics: A Case Study with Pyruvate Formate-Lyase

Marko Hanževački[+],[a, b] Karmen Čondić-Jurkić[+],[a] Radha Dilip Banhatti,[a] Ana-Sunčana Smith,[a, b] and David M. Smith*[a]

Abstract: Pyruvate formate-lyase (PFL) catalyzes the reversible conversion of pyruvate and coenzyme A (CoA) into formate and acetyl-CoA in two half-reactions. For the second half-reaction to take place, the S—H group of CoA must enter the active site of the enzyme to retrieve a protein-bound acetyl group. However, CoA is bound at the protein surface, whereas the active site is buried in the protein interior, some 20–30 Å away. The PFL system was therefore subjected to a series of extensive molecular dynamics simulations (in the μs range) and a host of advanced analysis procedures. Models representing PFL before and after the first half-reaction were used to examine the possible effect of enzyme acetylation. All simulated structures were found to be relatively stable compared to the initial crystal structure. Although the adenine portion of CoA remained predominantly bound at the protein surface, the binding of the S—H group was significantly more labile. A potential entry channel for CoA, which would allow the S—H group to reach the active site, was identified and characterized. The channel was found to be associated with accentuated fluctuations and a higher probability of being in an open state in acetylated systems. This result suggests that the acetylation of the enzyme assumes a prominent functional role, whereby the formation of the acyl intermediate serves to initiate a subtle signaling cascade that influences the protein dynamics and facilitates the entry of the second substrate.

Introduction

Enzymes are dynamic entities that can adopt a variety of conformations to perform their function. Indeed, they are often found to undergo certain structural changes, which allow for the binding of the substrate and the release of the product, thus enabling the necessary chemical transformations to take place in the active site. Transitions between various states and available conformational ensembles are characterized by both the free energy landscape of the enzyme and the details of its catalytic cycle.[1,2] However, the identification and characterization of specific conformational changes that are coupled with enzyme catalysis, or protein function in general, can be rather challenging. In a recent study addressing the closing and opening of active site loops that regulate enzyme activity in a classic and prototypical system, namely, triosephosphate isomerase,[3] by using advanced simulation techniques with several

hundreds of nanoseconds of sampling, the authors have shown that the conformational changes occurring in the loop dynamics is interesting and complex, contrary to what had been earlier thought.

A potentially even more challenging situation arises when the enzyme itself undergoes a chemical modification, which can in turn modulate the protein conformational space.[4,5] A typical case of chemical modification arises in enzymes in the context of double displacement or ping-pong mechanisms.[6,7] This scenario involves an enzyme-catalyzed reaction with two substrates and a mechanism that includes a temporary intermediate state in which the protein is chemically modified during the reaction with the first substrate. The initial state is then regenerated upon the reaction with the second substrate. A well-studied example is the enzyme *T. cruzi trans*-sialidase in which a classical ping-pong mechanism with acid/base catalysis was detected[8] and confirmed by recent molecular dynamics study.[9]

An intriguing, yet less well-studied, example of ping-pong kinetics can be found in the radical enzyme pyruvate formate-lyase (PFL). PFL plays a key role in the anaerobic glucose metabolism of *E. coli* and other microbes, catalyzing the breakdown of pyruvate and the subsequent acetylation of coenzyme A (CoA) through an acetylated protein intermediate in two half-reactions shown in Scheme 1.[10]

Structurally, PFL is a homodimer comprised of two identical subunits, in which only one subunit per dimer is active (half-site reactivity).[11] The active site of the activated subunit contains a glycyl radical (G734), which serves for radical storage,
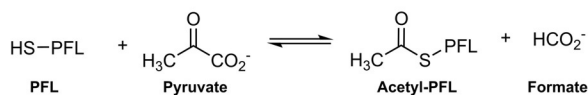
[a] M. Hanževački,[+] Dr. K. Čondić-Jurkić,[+] Dr. R. D. Banhatti,
    Prof. Dr. A.-S. Smith, Dr. D. M. Smith
    Division of Physical Chemistry, Ruđer Bošković Institute
    Bijenička 54, Zagreb (Croatia)
    E-mail: david.smith@irb.hr

[b] M. Hanževački,[+] Prof. Dr. A.-S. Smith
    PULS Group, Institute for Theoretical Physics
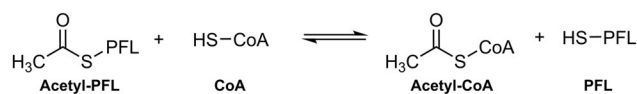   FAU Erlangen-Nürnberg, Staudtstraße 7, Erlangen (Germany)

[+] These authors contributed equally to this work.

Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under:
https://doi.org/10.1002/chem.201900663.

1st half-reaction

HS−PFL	+	H₃C-pyruvate-CO₂⁻	⇌	H₃C-acetyl-S-PFL	+	HCO₂⁻
PFL		**Pyruvate**		**Acetyl-PFL**		**Formate**

2nd half-reaction

Scheme 1. Reaction catalyzed by PFL.

and two neighboring cysteines involved in catalysis, C418 and C419. Both cysteines were demonstrated to play a crucial role in the enzyme function by experiments based on mutagenesis and inhibitor testing.[12] PFL is the first discovered glycyl radical enzyme (GRE), a class of enzymes that all temporarily store radical reactivity on the C_α atom of a glycine residue in the polypeptide chain.[13–15] To become fully active, PFL undergoes posttranslational modification in which the glycyl radical is introduced into the system.[16] This process is catalyzed by the PFL activating enzyme (PFL-AE), which is a member of the radical S-adenosylmethionine (SAM) enzyme superfamily.[17–19] Such radical enzymes are receiving increased interest because of their potential applications in biotechnology and biochemical engineering,[20, 21] and most recently also in the context of enzyme engineering.[22, 23]

From the crystal structure of PFL available in the presence and absence of both substrates, pyruvate and CoA, it emerges that the pyruvate binding site of PFL is buried in the protein interior, potentially serving to reduce the possibility of quenching the radical species in futile side reactions with solvent or other small molecules.[24–26] Further, it is found that CoA binds close to the interface between two subunits in the dimer, with one CoA bound to the surface of each subunit and approximately 30 Å from the active site.[26] Bound CoA adopts the unusual *syn* conformation with respect to the N-glycosidic bond, although the *anti* arrangement is the preferred conformation of free CoA in solution.[27–29] In the *syn* conformation, the thiol group on the pantothenate chain is located in the predominantly hydrophobic pocket formed by the side chains of resi-

dues F200 and H227 of the opposing monomer. Figure 1 shows the above described binding modes of both substrates in more detail.

The currently accepted mechanism for the catalytic cycle which brings forth more details of the two half-reactions outlined in Scheme 1, is shown in Figure 2.[30–34] In the first half-reaction, pyruvate is fragmented into formate and the acetyl group upon the addition of the thiyl radical located at C418. The radical initially stored at G734 (see structure A in Figure 2) is shuttled to C418 via C419 (see B in Figure 2). The addition of radical C418-S• to the central carbon of pyruvate leads to C−C bond dissociation, resulting with formation of $CO_2^{•-}$ and acetyl-C418. The latter species acts as a temporary acetyl carrier and a reactant in the second half-reaction, with the co-substrate CoA, to produce acetyl-CoA (AcCoA). Formation of AcCoA, the final product, closes the catalytic cycle of PFL. This mechanism suggests that $CO_2^{•-}$ is quenched by C419 (see C → E in Figure 2), which then activates CoA for a radical acetyl transfer, although it has been proposed that $CO_2^{•-}$ could alternatively adopt the latter role (see D→E in Figure 2).[33] The sequence of events taking place after the first half-reaction is still somewhat speculative, as most of the available experimental data offer insights into the first half-reaction and the initial generation of the glycyl radical.[35, 36]

Whereas certain radical enzymes, such as those dependent on coenzyme B_{12}, are known to undergo large conformational changes upon substrate binding, activation or product release,[37–44] the situation for PFL is much less clear. In particular, the pathway for CoA from its crystallographic binding position on the protein surface to the buried active site is far from obvious from the available crystal structure.[26] Indeed, apart from the consensus that a conformational change appears to be required for the thiol group of CoA to reach the active site, very little is known about the structural basis for such a change. It has been suggested that the ribose and pantothenate moieties of CoA might rotate around the N-glycosidic bond (Figure 1) and change from the *syn* to the *anti* conformation. This transition is expected to be energetically favorable and it could lead, potentially, to a more favorable positioning of the thiol group for the entry in the active site.[26] Nevertheless, the structural evidence surrounding the conformational change of CoA, in
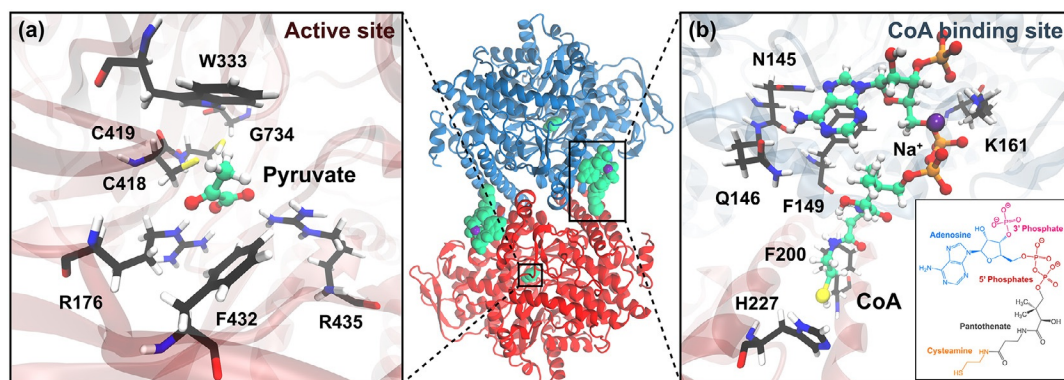

Figure 1. The crystal structure of the homodimeric form of PFL, with a closer view of the pyruvate bound in the active site (a) and the binding site of CoA (b). The building blocks of CoA are also shown as an inset in panel (b).
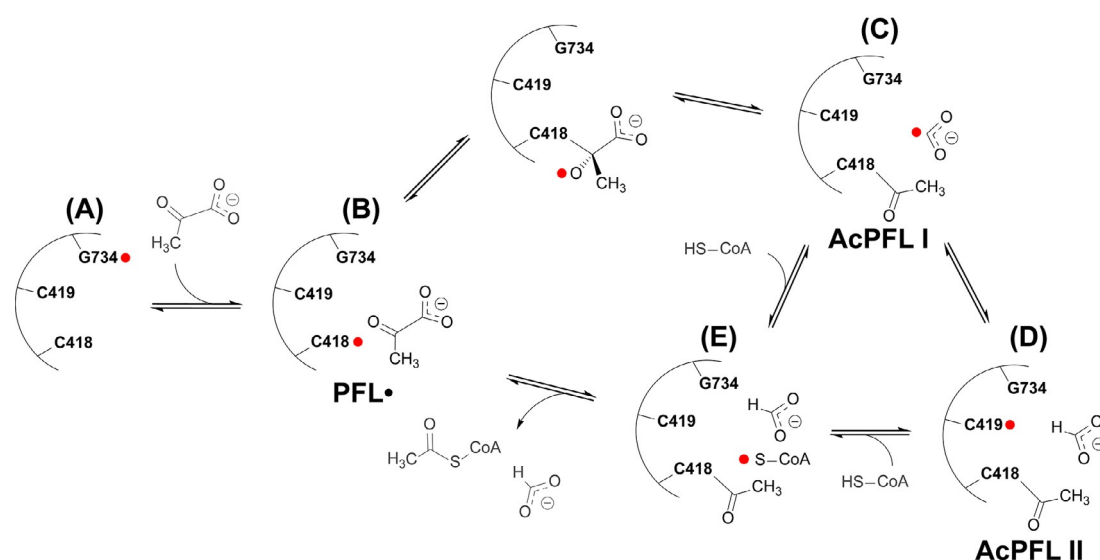
Figure 2. The currently accepted reaction mechanism for PFL catalysis. The structures represent the active site of PFL before (**PFL·**) and after (**AcPFL I** and **AcPFL II**) the first half-reaction (see text for a detailed description). The location of the radical at each step is denoted by the symbol ·.

the context of the PFL mechanism, has been even recently described as elusive.[22]

The experimental data show that the presence of CoA is not mandatory for the first half-reaction, but it can be bound to the enzyme in a spectator mode. It is also known that small thiols, such as mercaptoethanol and dithiothreitol, are efficient deactivators of PFL, whereas larger thiols, such as cysteine or glutathione, show no detectable deactivation.[45] These observations indicate that a conformational change is not only required for the second half-reaction to take place, but that its timing may also need to be regulated. In particular, if the S—H group of CoA or any other thiol molecules were to enter the active site before the completion of the first half-reaction, it could conceivably result in the premature quenching of the radical activity. This could explain why the active site needs to be occluded, as observed in the crystal structure and, potentially, why a conformational change might be a requisite aspect of the overall reaction mechanism.

In the past decades, molecular dynamics (MD) simulations have become an indispensable tool in studying protein structure-function relationships at a molecular level, especially for questions that are difficult to access experimentally.[46,47] In this vein, we have chosen to apply both standard and advanced MD simulation tools in an attempt to clarify the open questions surrounding the mechanism of PFL and, particularly, the elusive role of CoA. In terms of the aforementioned timing of the poorly understood conformational change, we have investigated whether acetylation of the enzyme could act as a trigger, in which the chemical modification of the enzyme can serve to signal the onset of a conformational change. We substantiate our hypothesis that a pathway for CoA could indeed appear as a result of the said conformational changes, by first analyzing the global protein conformation followed by examining the binding properties of CoA in detail. With robust tools we then endeavor to identify a potential entry channel for CoA

into the active site and examine what effects of the system topology would allow it to approach the active site from its experimentally observed, surface-bound position. Before we make a systematic and sequential presentation of these results, we start by defining the model systems considered. The parameters, methods and analytical tools employed in our extensive atomistic MD study of the PFL system are described in the Theoretical Methodology section.

Results and Discussion

System models

For the purpose of this study, we constructed three different PFL models representing the system before (**PFL·**) and after (**AcPFL I** and **AcPFL II**) the first half-reaction. It must be noted from Figure 2 that each of these three models contains a different radical carrier as described in the Introduction. In addition, all models contain CoA initially bound at the protein surface according to the crystal structure data.

As described in the Introduction, PFL is a homodimer with only one active site activated, based on the experimentally determined half-site radical occupancy.[35,48–50] However, the binding site of CoA in the available crystal structure spans both subunits so that the nucleotide moiety interacts with one monomer, whereas the cysteamine group is placed between a histidine and phenylalanine residue of the opposing monomer. To understand the role of the spectating subunit in the catalysis we additionally introduce two sets of models–a monomeric and dimeric set. The monomer-based models were also useful for performing preliminary calculations at the lower computational cost. When presenting the results, the prefix "d-" is used for dimeric models, whereas "m-" is used for monomeric models: for example, **dPFL·**, and **mPFL·**.

Each of the above-mentioned structural models was derived from a single crystal structure containing PFL in a complex with pyruvate and CoA (PDB ID: 1H16).[26] Before further processing, the original PDB file was modified by removing duplicate entries and assigning the protonation states of titratable residues using the H++ server.[51] Assignments made by the H++ server were additionally verified and confirmed by visual inspection of the local environments of the residues. Sodium ions and water molecules were retained, while other co-crystalized species were removed from the PDB file, including the Mg^{2+} ion, 1-treitol and tetraethylene glycol. The final PDB files used to generate the relevant topologies are provided in Supporting Information.

Active site geometries of PFL and AcPFL

Before turning to the long-range effects of chemical change on protein dynamics, it is instructive to view the geometry of the active site obtained within our MD production runs. In Figures 3b–d, we present the active site of PFL in both non-acetylated (**PFL·**) and acetylated (**AcPFL I** and **AcPFL II**) forms.
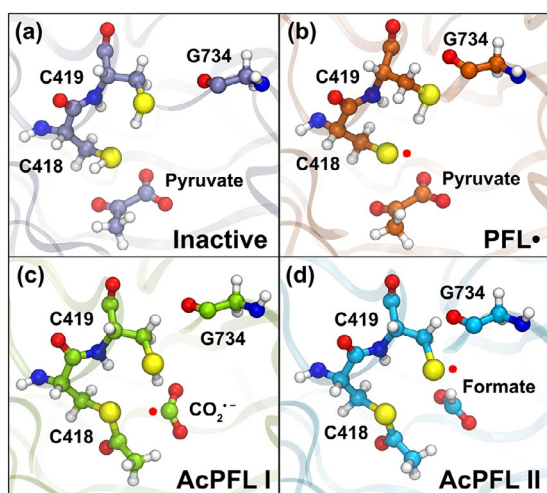

Figure 3. Models representing the snapshots of the PFL active site in the inactive subunit of a dimer (a) and before (b) and after (c and d) the first half-reaction. All models also contain CoA occupying its crystallographic binding site at the protein surface and the location of the radical is denoted by the symbol ·.

For completeness, we also present the active site of the inactive subunit of a dimer in Figure 3a. It must be noted that that Figures 3b–d correspond to structures B, C and D in Figure 2, respectively. We here wish to draw attention to the fact that, in case of **PFL·**, both the active and inactive subunits share similar features of the active-site architecture. In both Figures 3a and 3b, pyruvate is tightly bound in the active site under the cysteine loop, in a very similar position to that observed in the crystal structure, despite the long simulation times. In contrast, we notice distinct changes are introduced in response to changing the connectivity of the atoms and relaxing the system, in the acetylated forms. Thus, in Figures 3c,d, we note

that pyruvate is separated into two parts with the acetyl group bound to the C418, whereas the other fragment (be it $CO_2^{·-}$ or HCO_3^-) stays noncovalently bound in the active site. This visualization should be useful in understanding various results that follow. The next step is to examine carefully the various conformational changes in our model systems that will help in understanding aspects that might facilitate the entry of CoA.

Global protein conformation

We begin our analysis by investigating the global protein conformational changes and behavior of monomeric and dimeric systems in aqueous solution. We calculated the root-mean-square deviation (RMSD) of the backbone atoms (N, C_α and C) from the initial crystal structure positions for monomeric and dimeric systems, which is one of the standard measures of protein structural stability. The boxplots for two longest trajectories (500 ns) for each of our three model systems shown in Figure 4a (monomer and dimer) are constructed by using the RMSD time traces shown in Figures S2a and S3a in the Supporting Information. An inspection of similar RMSD time traces for all the remaining 100 ns trajectories (see Figures S4a and S5a in the Supporting Information) confirms that the boxplots shown in Figure 4a are indeed representative of also the results of the shorter runs. Furthermore, the inactive subunit in the dimer also has RMS deviations of the protein backbone atoms similar to the monomeric and the active subunit of the dimeric systems (see Figures S3a and S5a, Supporting Information). Turning our attention back to Figure 4a, the average measured RMSD was within 1.2–1.6 Å, whereas the maximum values did not exceed 2 Å in any of the simulated systems. From all the above we infer that both monomeric and dimeric sets of PFL models are relatively stable with respect to the reference crystal structure, and do not display major structural changes.

Root-mean-square fluctuation (RMSF) of the backbone atoms was also calculated by measuring the fluctuations of each C_α atom and is a measure of an average atomic mobility of the protein during the MD simulations. RMS fluctuations of the backbone atoms were estimated by averaging fluctuations over 500 ns shown in Figures S2b and S3b (Supporting Information) for both sets of PFL models. Notably, these fluctuations are measured at each of the C_α backbone atoms of the residues constituting the protein. These are plotted in Figure 4b. Once again, the RMSF values from the ten shorter simulations (see Figures S4b and S5b in Supporting Information) show very similar features to the longer simulations (Figures S2b and S3b, Supporting Information), confirming the structural stability of PFL with respect to the starting crystal structure on both simulated timescales, be it 100 or 500 ns. A closer look at Figure 4b shows that the solvent-exposed regions and protein termini show similar fluctuations in both dimeric and monomeric systems. However, the loops found at the interface between two subunits in dimeric models show significantly diminished fluctuations compared to their counterparts in the monomer simulations (highlighted emerald in Figure 4b).
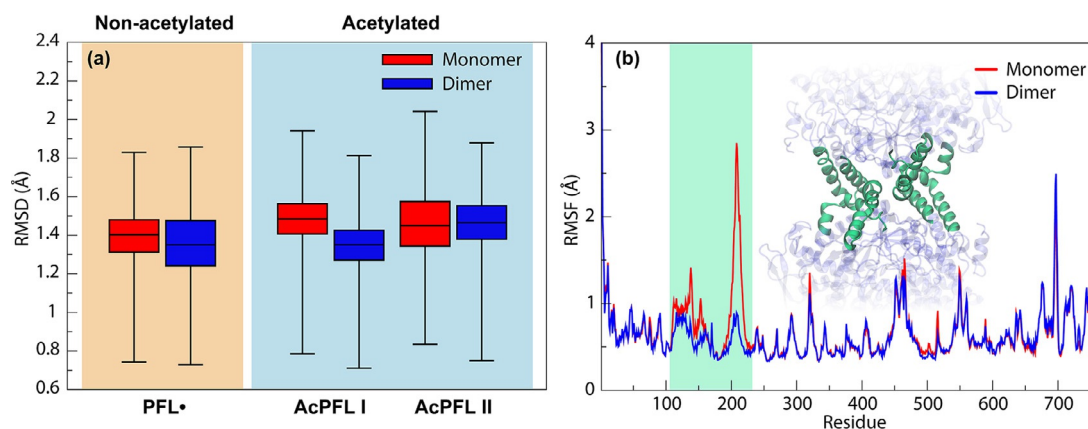
Figure 4. Average RMSD (a) and RMSF (b) values of the protein backbone from the crystal structure in monomeric and dimeric (active subunit only) systems during 500 ns of MD simulation. The MD snapshots from both non-acetylated (**PFL·**) and acetylated (**AcPFL I** and **AcPFL II**) models were taken into account. The RMSD values are shown in a boxplot representation, where the rectangle indicates the interquartile range. The median is shown by the line in the box while the whiskers go from the minimum to the maximum of all data. For RMSD time series of monomers and dimers see Figures S2a and S3a (Supporting Information), respectively. The residues at the interface of two subunits in dimeric models are highlighted in emerald in the panel (b).

Coenzyme A and its binding site

CoA is a linear, flexible molecule with many possible rotamers. To better understand the effect of the binding of CoA to PFL on the conformational space, we also performed two additional 500 ns simulation of CoA in water to obtain the reference state for the unbound molecule. We start by using the previously outlined procedure of constructing boxplots from time traces of RMS deviations, taking into account the non-hydrogen atoms of CoA of 500 ns (see Figures S6a and S7, as well as Figure S8 of the Supporting Information in which short 100 ns time traces are also representative of the longer time traces). We find from Figure 5, RMSD values of up to about 5 Å, which are much larger compared to that in the protein backbone. The more dynamic behavior of CoA is, in some sense, to be completely expected given its flexibility. As outlined in the Introduction, it is known from the crystal structure that CoA bridges two units of the PFL and, thus, prefers the *syn* confor-
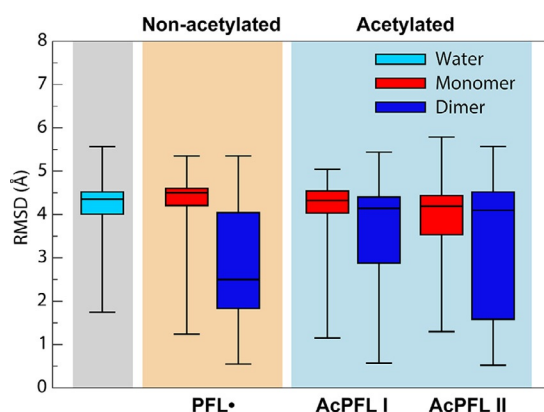

Figure 5. RMSD boxplot of non-hydrogen atoms of CoA from the crystal structure position in water, monomeric and dimeric systems during 500 ns of MD simulation. For RMSD time series of CoA in water and bound to the protein see Figures S6a and Figure S7 (Supporting Information), respectively.

mation, whereas the *anti* conformation is dominant in its unbound state (Figure 6a).[26–29]

What is interesting in this context is to examine if evidence for conformational isomerism of CoA can be obtained which would potentially enable it more easily to reach the buried active site of PFL. We have indeed found such evidence by monitoring two geometrical parameters, namely, the N-glycosidic dihedral angle O4′-C1′-N9-C8, and the head-to-tail distance between the cysteamine moiety and the adenine. These two parameters when plotted against each other, for example by using average values from time traces shown in pairs of Figures S6b,c (Supporting Information) for CoA in solution, yield a heatmap as shown in Figure 6b. Figure 6c is made for CoA in monomeric systems (by using sets from Figures S9 and S11a in the Supporting Information, as well as by using results of short-time runs on more independent production runs shown in Figures S10 and S12a, Supporting Information). Figures 6d,e correspond to heatmaps of CoA in the active and inactive subunit of the dimer, respectively (using sets from Figures S9 and S11b, as well as Figures S10 and S12b in the Supporting Information). From Figure 6b it is evident that, in solution, CoA prefers folded states over extended ones, whereas the *syn* and *anti* conformations are similarly populated. Surprisingly, in the monomeric system (Figure 6c) folded *anti* conformations are preferred. In the two dimeric systems, despite the starting conformation bias (Figures S9b, S10b, S11b, and S12b, Supporting Information) that makes the extended *syn* conformations appear dominant, transitions towards the more compact states and *anti* rotamers were consistently observed during simulations (Figures 6d,e).

To understand this better, we look at various features revealed in these two figures when contrasted to Figure 6c. Essentially, the main difference between the monomer and dimer-based simulations originate from the interactions of CoA cysteamine thiol group with the aromatic sidechains of F220 and H227 of the opposing subunit, as observed in the crystal structure model.[25, 26] This interaction appears relatively persis-
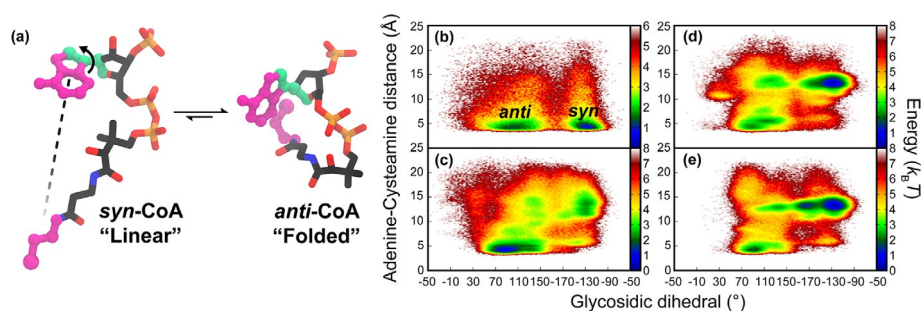
Figure 6. (a) The equilibrium between *syn* and *anti* conformation of CoA. Heatmaps for CoA conformations in (b) water, (c) monomer, (d) active and (e) inactive subunit of a dimer. The snapshots were taken from both non-acetylated (**PFL'**) and acetylated (**AcPFL I** and **AcPFL II**) systems. Two geometrical parameters (descriptors) were monitored, namely glycosidic dihedral O4'-C1'-N9-C8 (cyan) were plotted against distance between adenine and cysteamine (magenta). The heatmap surface is given by $W/k_BT = -\ln(N/N_{tot})$ in which N_{tot} represents the total number of configurations of each individual case (N_{tot} is 100 000 and 540 000 for CoA in water and bound to a protein model, respectively).

tent under the simulation conditions. However, the thiol group of CoA eventually disengages in the majority of the simulations, demonstrated both in the rise of the RMSD (see Figures S7 b and S8 b, Supporting Information) and in the increasing distance between the cysteamine moiety of CoA and two interacting residues F220 and H227 (Figures S13 and S14, Supporting Information). This release of the thiol group then allows sampling of the more folded conformations observed in the heatmaps presented in Figures 6 d,e. The total absence of this interaction in the monomeric systems allows for greater mobility of the cysteamine and panthotenate groups in the solvent from the very beginning of the simulation, resulting with a different conformational distribution (compare Figure 6 c with Figures 6 d,e). The RMSD patterns further illustrate the difference between monomers and dimers (see Figures S7 a and S8 a for monomers and S7 b and S8 b for dimers, Supporting Information).

To appreciate the differences observed in the distribution of conformation of CoA in monomeric and dimeric PFL, it is important to note that the binding mode of the adenosine part of CoA, present in both monomeric and dimeric models, is achieved via interactions of the nucleotide moiety with the (activated) subunit of PFL. The relevant binding site comprises a short α-helix that includes the strictly conserved N145, Q146 and F149. In the crystal structure (see Figure 1), the latter residue engages in stacking interactions with the imidazole ring of adenine moiety, while N145 and Q146 further strengthen binding through the formation of hydrogen bonds with the adenine amino group. Additional interaction between CoA and PFL is provided by a salt bridge between the 3' and 5' phosphates and K161. The distance between the adenine base and its stacking partner in the crystal structure, F149, showed that adenine generally remains tightly or partially bound in all examined cases (see Figures S15 and S16, Supporting Information). What could be observed from snapshots of MD trajectories is that the phosphate groups of CoA seemed to form more of an exclusive interaction with the previously shared K161, but also with a number of other neighboring lysine or arginine residues (K118, K159, R160, K615 and K617). This could potentially affect the adenine-F149 stacking and hydrogen bonding with N145 and Q146. However, it is also seen

that stacking and hydrogen bonding were able to reform during the investigated simulation times, keeping the adenine of CoA predominantly bound to the protein surface. These changes were also reflected in RMSD plots of CoA (see Figure 5 and Figures S7 and S8 in the Supporting Information).

The most important findings from a functional point of view can, therefore, be stated as follows. Whereas the binding mode of the adenosine fragment of CoA remains predominantly as observed in the crystal structure, the binding of S—H group is significantly more labile. In this connection, we observe, both in monomeric and dimeric model systems and within the simulation time of 500 ns, a systematic change of the CoA conformation with respect to the N-glycosidic bond, from *syn* to *anti*, changing the adenosine dihedral angle from around −110° to 80° (Figures 6 and Figure S17 in the Supporting Information). This is in accordance with the previously mentioned hypothesis, namely, that ribose and pantothenate moiety might rotate around N-glycosidic bond, in turn potentially enabling CoA to more easily reach the buried active site.

Potential CoA entry channel

A visual inspection of the free simulations of PFL in both monomeric and dimeric systems did not reveal any obvious or large-scale conformational changes that would allow CoA to enter the active site from the surface of the protein where it is bound. Thus, despite the previous demonstration of a range of conformations readily adopted by CoA when bound to PFL, possible pathways leading from the protein surface toward the active site that are able to accommodate CoA remain a puzzle. An attempt to instead identify dominant channels pointing from the active site towards the surface of the protein has been made by using CAVER–a program using spherical particles to locate cavities in proteins (see Theoretical Methodology for details) in the three monomeric model systems. Irrespective of whether we considered acetylated or non-acetylated systems, these channels were found pointing from the active site to CoA and its binding site. What is more interesting is that in the acetylated systems, the highly ranked channels identified by CAVER when clustered, converged to a single but broad channel starting above two neighboring helices positioned
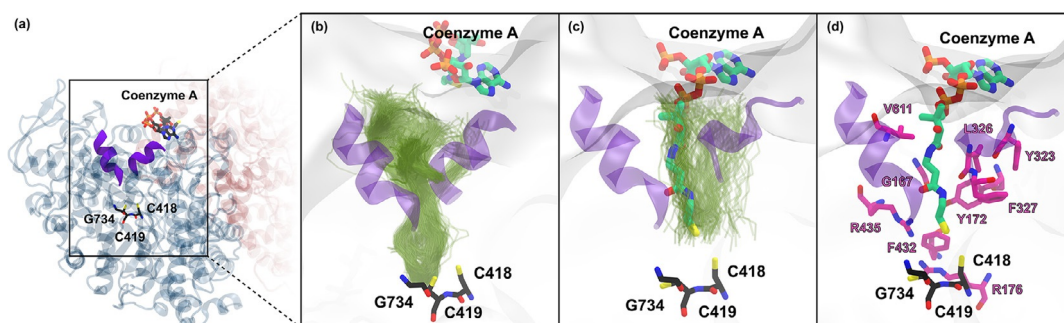
Figure 7. (a) The location of the potential CoA entry channel "gateway" helices in the active subunit of PFL for all three systems. (b) The top ranked collective pathways identified throughout the MD simulations by CAVER. Pathways are shown in one frame as pathway centerlines where only channel with the lowest cost is shown. (c) Ensemble of CoA (pantothenate and cysteamine blocks shown as green lines) geometries derived from SMD simulations in **AcPFL I**, in which CoA was pulled into the active site independently 200 times. (d) The representative CoA binding mode derived from SMD simulations, forming the potential "gateway" in the channel, with residues in direct contact with bound CoA in **AcPFL I**. The representative CoA binding mode is shown in a licorice representation.

above the active site (Figure 7 a,b). Other possible pathways that were mostly pointing away from the CoA site were also identified and these could potentially be relevant in the PFL activation process or product release after the completion of the catalytic cycle (Figure S18, Supporting Information). Figure S18 also contains results obtained on the dimeric form of the enzyme, indicating that the results from the monomer are sufficient.

The size of the spheres used in CAVER for probing cavities in the protein are usually smaller than the size of a CoA molecule. To ascertain that this dominant single but broad channel shown in Figure 7 b can indeed be a possible pathway for entry of a CoA, we performed a series of steered MD (SMD) simulations on the three monomeric model systems, in which the cysteamine group of CoA was driven from the protein surface into the active site (see Theoretical Methodology for details). The distance between the cysteamine moiety and the backbone atoms of C418 and C419 was used as a reaction coordinate. In these pulling experiments, CoA was able to penetrate the protein dynamically and successfully reach the active site, see Figures 7 c,d. Furthermore, when tracing the main-chain atoms of the deeply bound CoA significant overlap with the previously identified channel by CAVER was revealed (see Figure 7 b,c). This surprisingly good agreement between the CAVER modelling (exit) and tracing of the SMD trajectories (entry) supports the plausibility of the suggested binding channel. For **AcPFL I** and **AcPFL II**, we indeed find that the binding pose is stable for 10 ns of unrestrained dynamics following the forward pulls (Figure S19, Supporting Information).

A careful inspection of the pulling simulations allowed us to identify the residues that were most commonly interacting with CoA on its entry pathway. The residues that form the potential channel are: G167, Y172, R176, Y323, L326, F327 F432, R435 and V611 (Figure 7 d). Three of these residues (R176, F432 and R435) also participate in pyruvate binding and only interact with CoA when it is deeply bound (see Figure 7 d). There are three residues with aromatic sidechains (Y323, F327 and Y172), stacked together and forming a spiral lid that further covers the active site. The aliphatic L326 and V611 are sitting

at the top of two opposing helices, which ordinarily keep the channel closed. These helices are a part of two longer sequences (321–327 and 607–615, respectively) that enclose the channel (shown as purple ribbons in Figure 7 a). The focal point of the resulting "gateway" (effectively between L326 and V611) is positioned approximately 15 Å above the catalytic cysteines and this is the point where the multiple CAVER channels converged to a single one.

Identification of the putative entry channel and its gateway by SMD simulations, as well as the agreement with results from CAVER has opened new directions for analysis as it has narrowed down the protein region that might be relevant for the approach of the second substrate. A potentially useful approach in this respect is principal component analysis (PCA), which we have performed on all six unrestrained model systems, both monomeric and dimeric, as well as acetylated and non-acetylated. We conducted a close inspection of the first ten principal components (PC1-10) thus produced and present the displacement vectors of the first five (PC1-5) in Figure S20 (Supporting Information). The highest variance was typically associated with the movement of the termini and loops, as expected. However, important fluctuations comprising PC5 are found to be localized in the "gateway" region. Specifically, these fluctuations correspond to the "breathing" movement of the two opposing helices that ordinarily keep the channel closed (Figure 8 and Movie S1).

Furthermore, when we consider the histograms constructed by projecting all MD snapshots onto PC5 (Figure S21, Supporting Information), it is evident that in non-acetylated model systems, the distribution is shifted towards the closed configuration (fewer positive values) in comparison to acetylated model systems, for which the probability of the channel being in an open state (more positive values) is larger. This result indicates that the topology could have a potentially important relationship with the identified channel, although more conclusive information in this direction is difficult to obtain from PCA, due to the inherent effects of noise.
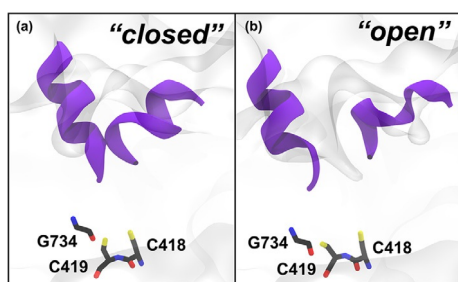
Figure 8. Snapshots from fifth principal component showing a view down the possible "gateway": (a) "closed" and (b) "open" state.

Topology effect

The evidence from PC analysis supports the hypothesis that acetylation might be related to conformational changes in the protein. This in turn evokes the notion that chemical changes at the active site might be involved in transferring a signal, which causes conformational changes in the distal regions of the protein and an associated change in preference from a closed state to an open state of the identified "gateway". We therefore report in Figure 9 the cross-correlation of α-carbon positions to further investigate this possibility by comparing the correlations of non-acetylated and acetylated PFL systems.

The plots in Figure 9 are represented by different colors with highly positive regions ranging from white to cyan and highly negative regions ranging from white to magenta. These are associated with strongly correlated and anticorrelated movements of specific residues, respectively. The correlation maps for non-acetylated and acetylated states exhibit rather similar overall structures, with certain additional correlations present in the acetyla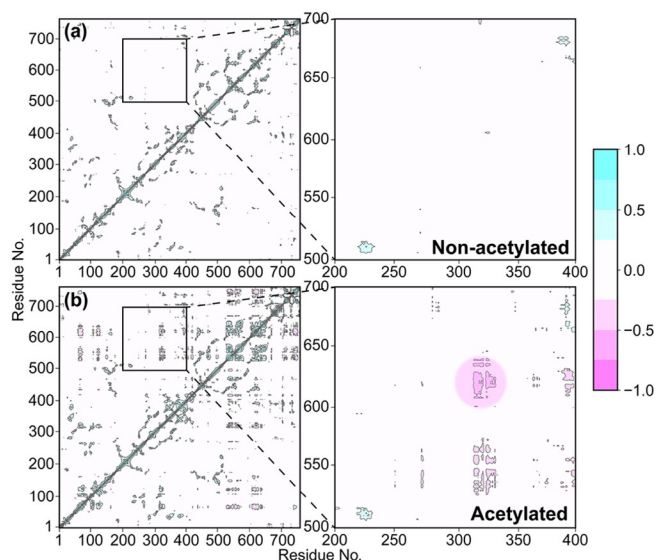ted systems. The additional features are indicative of higher flexibility overall and consist of some enhancement of positive correlations along the diagonal regions, coupled with the appearance of several anticorrelated regions in the off-diagonal regions. The negative signals between C-terminal residues in the range 75–125 and N-terminal residues in the range 550–650 (Figure 9b) correspond to relatively large-scale motions whose appearance, while interesting, is difficult to associate with a specific functional aspect. On the other hand, the anticorrelations of residues 320–330 with the same N-terminal residues and, particularly those between 600 and 620, appear to be highly relevant in the present context.

Namely, the regions from 320–330 and 600–620 (highlighted in Figure 9b), were previously characterized as a channel "gateway" region (Figures 7 and 8), related to the potential entry of CoA to the active site. The fact that these residues experience practically no cross-correlation before acetylation but become negatively cross-correlated after the acetylation event is quite telling. Overall, Figure 9 shows that acetylation of C418 is capable of exerting a subtle effect on the global dynamics of PFL. Among the numerous incarnations of this effect, it is particularly interesting that a mode corresponding to the breathing of an identified channel gateway (Figure 8) is significantly enhanced in this way.

In light of these results, it is instructive to examine the RMSDs of the channel residues listed above (Figure 7d) with respect to the crystal structure for all simulated systems (Figures S22 and S23, Supporting Information). The corresponding values showed that the channel stays close to its reference conformation for the non-acetylated **PFL·** system, while all four acetylated simulations (two **AcPFL I** and two **AcPFL II**) display periods associated with notable deviations. Interestingly, very similar trends were observed in measurements of the single distance between L326 and V611 (Figure 10, see also Figures S25 and S26 in the Supporting Information), which are two proximate residues located on each of the two opposing helices (Figure 7d).

Thus, starting from the general idea of finding dominant channels using CAVER, followed by SMD simulations, PCA procedures, cross-correlation maps and RMSD values, we have been led systematically to deeper insights. A careful analysis of their implications has led to the discovery of the L326–V611 (or L–V) distance as a simple, but informative descriptor for the channel movement. The measurement of the L–V distance (Figure 10) in the acetylated monomeric systems, **mAcPFL I** and **mAcPFL II**, revealed the presence of two distinct states, which were not found in the non-acetylated **mPFL·** system (see Figures S25a and S26a, Supporting Information). The same observations were made for the dimeric system (Figures S25b and S26b). The two states observed in the L–V coordinate can be directly linked to open and closed states of the channel described above. Namely, the channel remained in the closed state (L–V ≈ 7 Å, Figure 10a) in the non-acetylated systems, corresponding to the moderate RMSDs of the channel-forming residues (Figures S22 and S23, Supporting Information). The "open" state (L–V ≈ 9 Å, Figure 10b) found in the acetylated systems also has higher RMSDs associated with the channel residues.

Figure 9. Residue dynamical cross-correlation map for MD simulations of (a) non-acetylated and (b) acetylated PFL systems. Cyan regions (0 to 1) represent positive correlation, whereas magenta regions (0 to −1) represent negative or anticorrelation. Negative values typically indicate regions that move synchronously in opposite directions.
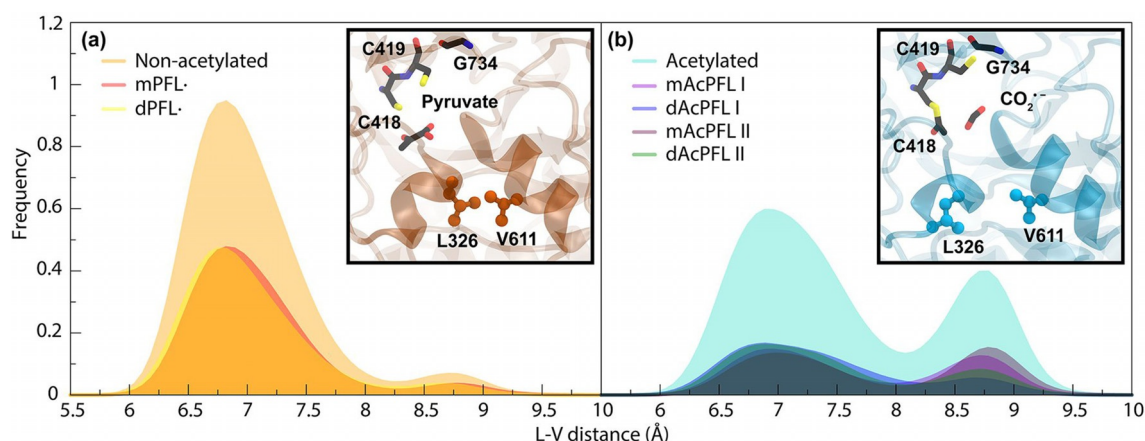
Figure 10. Histograms constructed from L326–V611 distances taking into account all MD simulations of (a) non-acetylated and (b) acetylated monomeric and dimeric systems. The "open" state of the channel, in which the distance between L326 and V611 residues is approximately 8.8 Å, is significantly more pronounced in acetylated forms. A snapshot from free MD simulations showing a view down the possible entrance channel: (a) "closed" state in **mPFL·** model; (b) "open" state in **mAcPFL I** model.

We have also investigated the opening events on a molecular level by closely monitoring the behavior of the channel residues as well as the residues comprising the active site, with a focus on the two longest trajectories (500 ns each) for each of the two acetylated monomeric systems. This is compared to the corresponding long trajectories of the non-acetylated systems, despite there being no opening events. Firstly, we recall that the geometry of the active sites after the acetylation (Figure 3 c,d) share similar features, while being distinct from that of the (active) non-acetylated system (Figure 3 b). In particular, we note that, although the pyruvate in Figure 3 b can be more tightly bound to the arginine residues (not shown in Figure 3), in the acetylated forms, either the $CO_2^{·-}$ in Figure 3 c or formate in Figure 3 d, tend to display an increased flexibility in their binding to the arginine residues owing to their smaller size and altered charge distributions. This is indeed revealed to be so from the dynamics. In Figure S24 (Supporting Information), we provide illustrative snapshots showing the changes introduced going from **PFL·** to the acetylated forms following the C—C cleavage.

We are able to observe a series of changes. R435, which appears to be more tightly bound to pyruvate carboxylate group in **PFL·**, is shifted away from the acetyl group in **AcPFL I** and **AcPFL II**. There is a small change in the distance between R435 and F327, which appears to be related to a more marked change in the orientation of F327 and a corresponding rotation of the whole helix containing F327, L326 and Y323, significantly increasing the width of the L–V gate (see Movie S2). Although the described changes are present during numerous opening events examined, the process is more aptly described as a subtle dynamic interplay as opposed to a deterministic sequence of specific events. Given that $CO_2^{·-}$ is smaller than the formate, we observe a larger change for **AcPFL I,** in tune with the results shown in Figure 10.

Notably, the indicated distances in Figure S24 (Supporting Information) are representative for a snapshot and mean values are shown in Figure 10. It must also be noted that the opening and closing events (Figures S25 and S26, Supporting Information) are always associated with rather abrupt jumps in the instantaneous value of the L–V descriptor, which strongly supports the existence of distinct states, despite the fact that the difference in the mean descriptor values is in the same range as the thermal fluctuations.

Furthermore, when considering the contributions to the histograms of the L–V distances shown in Figure 10 for the acetylated and non-acetylated dimeric and monomeric systems, it becomes evident that, although the contributions are about the same for monomeric and dimeric systems for the "closed" state, the "open" state is somewhat more populated by the monomeric systems. For now, we are not able to discern whether this population difference arises from slower sampling of the states in the dimeric models or from a potential mechanistic role of the inactive subunit, possibly to reduce the number of undesired molecules reaching the active site (see Figures S25 b and S26 b, Supporting Information). The putative channel of the inactive subunit in dimers was found to strongly prefer the "closed" state in all performed MD simulations, for both acetylated and non-acetylated cases (Figure S27, Supporting Information). Referring back to Figures 3 a,b, we see that the architecture of the active site of **PFL·** is very similar irrespective of whether we look at the active or the inactive subunit. Our observations from the Movie S2 and from Figure S24 (Supporting Information) support our conjecture that the pyruvate is more tightly bound to the arginine preventing the kind of cascade effects observed in acetylated forms.

Conclusions

Through this systematic study, we have sought to obtain a better understanding of how the temporary acetylation of PFL at C418 could potentially serve as a trigger for necessary conformational changes in the protein. Through the application of a set of standard and advanced MD techniques and analysis procedures, we have obtained results that strongly indicate

that the acetylation event has a marked influence on initiating the opening of a channel of approach for the second substrate, CoA, from its binding position on the protein surface to the active site.

To summarize the results in brief, a series of unrestrained molecular dynamics simulations of 100–500 ns was carried out to investigate conformational space of PFL using topologies that correspond to the protein states before (**PFL·**) and after the first half-reaction (**AcPFL I** and **AcPFL II**) by using monomeric and dimeric representations of the protein. Analysis of the simulated systems revealed that the global protein conformation was stable during the simulations and the measured deviations from the crystal structure were minor in both monomeric and dimeric model systems. On the other hand, the observed conformations of CoA ranged from the extended (*syn*) to the more compact (*anti*) states. The S—H group of CoA frequently unbound from the inactive subunit in the dimer, thus allowing for a counter clockwise rotation around N-glycosidic bond, which was also observed in the monomeric and isolated systems. In all simulations, the adenosine part of CoA remained in close proximity to the binding site identified in the crystal structure,[26] despite certain rearrangements in binding interactions with surrounding residues.

Further analysis of the simulations in monomeric model systems with the CAVER algorithm revealed three major channels connecting the active site with the protein surface. The channel with the highest rank invariably led towards the CoA surface binding site. The other two pointed in different directions and could be relevant for the activation of PFL or for final product release.

Interestingly, SMD simulations on the monomeric model systems, in which the thiol group of CoA was pulled towards cysteines 418 and 419, showed a propensity for the second substrate to be accommodated in a channel reaching to the pyruvate binding site, where CoA was found to stably reside for a further 10 ns of MD simulations. Furthermore, the resulting positions of the CoA tail atoms exhibited a significant overlap with the channel identified by CAVER. This putative channel also represents the shortest and the least crowded pathway when the system is driven along the chosen reaction coordinate. Careful inspection of the SMD trajectories revealed residues capable of interacting with a channel-bound CoA.

Employing the procedure of PCA in analyzing the trajectories of unrestrained dynamics in both monomeric and dimeric model systems, we focused on the extent of fluctuations for the first five principal components. In particular, it was demonstrated that significant fluctuations associated with PC5 are localized in a "gateway" region, which corresponded to a set of residues identified in the context of the CAVER and SMD analyses to encompass the putative channel. A comparative analysis of these residues during unrestrained dynamics on the monomeric models revealed that the channel was mostly found to be closed in the non-acetylated system (**mPFL·**). However, in the monomeric acetylated systems (**mAcPFL I** and **mAcPFL II**), this putative entry channel was frequently found to be in a more open state. In particular, we found that the distance between L326 and V611 (L–V) can be used as an excellent de-

scriptor of the channel opening fluctuations, which also correlates well with the fluctuations associated with PC5. The analysis of the channel residues and the L–V descriptor in the dimeric models (**dAcPFL I** and **dAcPFL II**) also revealed a more frequent visitation of the open state, only after the acetylation event (not for **dPFL·**). On a closer examination, we found that the acetylation loosens the binding of R435 to either the $CO_2^{·-}$ or formate, which is closely related to the change in the orientation of the helix containing L326 and the widening the L–V distance providing an open state of the gateway. This molecular level view provides insight into how the chemical change at the active site transfers the signal to a distal region.

We have thus obtained considerably novel insights into the details of the potential sequence of the PFL mechanism. Namely, the acetylation of the enzyme at the end of the first half-reaction appears to trigger more dynamic fluctuations, which causes the channel for possible CoA entry to appear in an open state more frequently. It would be remiss not to mention that we have not been able to find any signature of spontaneous entry of CoA into the active site, although SMD shows this to be feasible. Using classical MD simulations, which are hampered by the usual issues, such as insufficient sampling time and an inability to deal with the effects of electronic rearrangements on the fly, we have only been able to consider distinct states. Nevertheless we have uncovered a fascinating and remarkable result by which a minor chemical change on a length scale of a couple of ångstroms causes a change in the fluctuation spectrum of the entire enzymatic system. The effect of this change is to shift the dynamic equilibrium of the residues of the "gateway" to favor an "open" state. The resulting channel happens to correspond to the most probable path of the S—H group of CoA from its surface-bound, crystal-structure position to the buried active site it needs to reach to complete the second half-reaction.

Importantly, this putative CoA entry channel becomes available only after the completion of the first half-reaction. This is particularly intriguing in the context of the aforementioned inactivation effects of exclusively small thiol compounds[45] that can presumably access the active site in the absence of a specific channel. Thus, the apparent prevention of the access of CoA to the active site before the completion of the first half-reaction seems to assume a functional significance, which is related to protecting the radical intermediates from the potentially harmful, premature interference of the S—H group of CoA. Overall, our results and interpretations are fully consistent with the catalytic ping-pong mechanism in PFL being triggered by the coupling of a minor chemical change in the active site of the enzyme with its overall dynamics.

Apart from constituting a convenient descriptor for the channel opening, the positions and roles of L326 and V611 could also be utilized as a convenient means to verify our suggested mechanism, experimentally. The targeted mutation of both of these residues to cysteines, coupled with conditions conducive to forming a disulfide bridge, could result in the covalent closure of the CoA entry channel. Our calculations indicate that such a modification should prevent the approach of CoA to the active site. In the best case scenario, such a system

could even correspond to an engineered variant of PFL, the function of which can be conveniently switched at will.

Theoretical Methodology

Force field parameters

Parameters assigned to standard amino-acid residues in PFL were taken from the ff14SB[52] force field available within the AMBER16 software package.[53] For non-standard residues, the missing parameters were derived using the R.E.D. Server[54] and AmberTools17 suite.[53] These non-standard residues include radical cysteine (Cys$^{•}$), acetylated cysteine (AcCys), and the substrates and intermediates: CoA, pyruvate, formate (HCO$_2^-$), and carbon dioxide radical anion (CO$_2^{•-}$). CoA parameters were obtained by combining molecular fragments of usual cofactors in biochemistry from the R.E.D. Data Base under the project F-91 which was uploaded by Dupradeau.[55] All phosphate groups of CoA were fully charged. For rest of the non-standard residues, bonding and non-bonding parameters were taken from General Amber Force Field (GAFF)[56,57] and are available in SI. The missing charges for these residues were obtained by following the standard RESP procedure.[58] The charges for substrates and modified cysteines were derived from quantum mechanical (QM) calculations at B3LYP/cc-pVTZ//HF/6-31G(d,p) level of theory combined with an IEFPCM ($\varepsilon = 4.335$)[59] continuum dielectric model mimicking the polarization of the medium. The anionic forms of pyruvate, CO$_2^{•-}$ and the formate were used in parameterization. All QM calculations were performed using the Gaussian09 software package.[60]

The topology files describing the PFL system before and after pyruvate cleavage were built using the *leap* module of AmberTools17 with all the necessary parameters taken from the standard ff14SB force field and custom-built libraries. Furthermore, each system was solvated with TIP3P[61] waters in a truncated octahedron box. The edge length of the resulting box of solvent was about 140 Å for dimeric systems and approximately 100 Å for systems containing monomers (see Figure S1 a,b, Supporting Information). As mentioned earlier, all crystal water molecules and seven sodium ions per monomer present in the PDB file were retained. The Mg^{2+} ion was removed and replaced by a Na$^+$ ion (see Comment S1 in the Supporting Information). In addition, eight sodium ions per monomer were required to neutralize the system. The number of waters added to monomeric systems was approximately 24 000 molecules, whereas dimers were surrounded by about 55 600 solvent molecules. In order to compare bound CoA structures in protein models to the reference case of free CoA in water, we built an additional system where the CoA molecule was placed in a truncated octahedron (≈ 60 Å) of approximately 5500 TIP3P water molecules and neutralized by adding four sodium cations (Figure S1 c, Supporting Information).

Simulation conditions

All systems (Table S1, Supporting Information) were treated using periodic boundary conditions (PBC). Long range electrostatic interactions were calculated with Particle-Mesh Ewald (PME) technique with a default non-bonded cut-off of 8 Å to limit the direct space sum. The temperature in all simulations was controlled by coupling the system with the Langevin thermostat with a collision frequency of 2 ps^{-1}.[62] An integration time step of 2 fs was used and the SHAKE algorithm was employed to constrain bonds involving hydrogen atoms during the MD simulation.[63]

Relaxation of the systems was carried out in several steps: 1) steepest descent minimization was applied to the aqueous solution of protein-substrate-cofactors complexes (solute) with harmonic positional restraints on solute molecules (2 kcal mol^{-1} Å^2). 2) Heating dynamics was performed with continued solute restraints at constant volume (NVT). Thereby, the temperature was increased from 0 K to 300 K over 60 ps and kept at that value for another 40 ps. 3) Subsequently, 400 ps of constant pressure (NPT) dynamics at 300 K was performed, with isotropic position scaling at a target pressure of 1 bar and a pressure relaxation time of 0.2 ps by using the Berendsen barostat.[64] 4) Finally, an unrestrained NPT simulation at 300 K and 1 bar was performed for a duration of 500 ps.

The equilibrated systems were subjected to ten independent, unrestrained MD production runs for minimally 100 ns each. We also extended two simulations (from these ten) to a total of 500 ns each, giving rise to an overall simulation time of 1.8 μs per investigated model. All simulations were propagated at constant volume and temperature (300 K), saving the snapshots every 10 ps. Simulations of the monomeric and dimeric systems were carried out using the GPU accelerated code of *pmemd* algorithm,[65–67] whereas the temperature was controlled by the Langevin thermostat in all performed production simulations. The unrestrained MD simulation of free CoA in aqueous solution was propagated for 1 μs by using an analogous procedure.

Steered MD

With unrestrained MD simulations or any other experimental technique, it has not been possible to observe or examine the spontaneous entry of CoA into the active site. Steered molecular dynamics (SMD) provides an enhanced technique to drive and investigate such a process.[68–71] Within the monomeric set of models, trajectories that correspond to the process of CoA entering the active site are generated with this technique at constant volume and temperature (300 K). The pulling was unidirectional for all systems and no reverse trajectories were computed. A suitable reaction coordinate was defined as the distance between the centers of mass of the cysteamine group of CoA and the backbone of atoms C418 and C419. This initial distance between the two groups was approximately 30 Å, while the final distance was set to 5 Å in all performed pulls. The force constant of the harmonic potential used to drive the system along the points of the chosen coordinate was set to 5 kcal mol^{-1} Å^2. The duration of each pull was 10 ns.

The starting ensemble of conformations for the pulling was generated by running restrained MD during which the system was restrained to the initial value of the reaction coordinate by a harmonic potential (1 kcal mol^{-1} Å^2). These restrained simulations were started from an adequate snapshot taken from free dynamics. We performed ten independent distance restrained simulations for each monomeric system. The duration of the restrained MD simulations was 20 ns and starting snapshots for pulls were collected every nanosecond, which gave rise to a total of 200 SMD trajectories per investigated model. Note that this procedure was implemented to avoid any bias of non-independent starting points for the 200 forward pulls, since they start from 200 different initial structures. The final ensembles of CoA structures, obtained with SMD, were collected and shown as a geometrical overlap of CoA (cysteamine and pantothenate moieties) in a putative entry channel.

Analysis

All the data obtained from the molecular dynamics simulations were subsequently processed and analyzed using the *cpptraj*

module of the AMBER16 program package. Principal component analysis (PCA) was performed as a standard exercise to track the conformational dynamics and enzyme motions. In addition, CAVER 3.0[72] was used for identification and analysis of possible CoA entry pathways in MD trajectories of PFL systems processing multiple snapshots from a 500 ns classical MD simulation of monomeric PFL systems. The molecules in the active site (pyruvate, CO_2^{-} or formate) were chosen as starting points for dynamic pathway calculation. All standard and non-standard amino acids were included in the channel calculation. The second substrate, CoA, was not included in the channel calculation, along with all waters and sodium ions present in our MD trajectories. In each snapshot, all possible pathways with the bottleneck radius equal to or larger than 0.5 Å were identified. The pathways were clustered by the average-link algorithm based on the pairwise distances of the pathways using the clustering threshold of 3.5 Å. The dynamical cross-correlation (DCC) matrices were constructed to represent cross correlated displacements of backbone C_α fluctuations across MD trajectories in non-acetylated and acetylated systems. DCC between the residue-based fluctuations during 500 ns of simulation was calculated using the Bio3D package.[73] All structures were visualized using VMD 1.9.3.[74]

Acknowledgements

Conflict of interest

The authors declare no conflict of interest.

[1] L.-Q. Yang, P. Sang, Y. Tao, Y.-X. Fu, K.-Q. Zhang, Y.-H. Xie, S.-Q. Liu, *J. Biomol. Struct. Dyn.* **2014**, *32*, 372–393.
[2] S. Hertig, N. R. Latorraca, R. O. Dror, *PLoS Comput. Biol.* **2016**, *12*, e1004746.
[3] Q. Liao, Y. Kulkarni, U. Sengupta, D. Petrović, A. J. Mulholland, M. W. van der Kamp, B. Strodel, S. C. L. Kamerlin, *J. Am. Chem. Soc.* **2018**, *140*, 15889–15903.
[4] U. Doshi, L. C. McGowan, S. T. Ladani, D. Hamelberg, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 5699–5704.
[5] D. D. Boehr, R. N. D'Amico, K. F. O'Rourke, *Protein Sci.* **2018**, *27*, 825–838.
[6] W. W. Cleland, *Biochim. Biophys. Acta Spec. Sect. Enzymol. Subj.* **1963**, *67*, 104–137.
[7] N. N. Ulusu, *J. Mol. Evol.* **2015**, *80*, 251–257.
[8] I. Damager, S. Buchini, M. F. Amaya, A. Buschiazzo, P. Alzari, A. C. Frasch, A. Watts, S. G. Withers, *Biochemistry* **2008**, *47*, 3507–3512.
[9] I. A. Oliveira, A. S. Gonçalves, J. L. Neves, M. von Itzstein, A. R. Todeschini, *J. Biol. Chem.* **2014**, *289*, 423–436.
[10] J. Knappe, H. P. Blaschkowski, P. Grobner, T. Schmitt, *Eur. J. Biochem.* **1974**, *50*, 253–263.
[11] V. Unkrig, F. A. Neugebauer, J. Knappe, *Eur. J. Biochem.* **1989**, *184*, 723–728.
[12] J. Knappe, S. Elbert, M. Frey, A. F. V. Wagner, *Biochem. Soc. Trans.* **1993**, *21*, 731–734.
[13] A. F. Wagner, M. Frey, F. A. Neugebauer, W. Schäfer, J. Knappe, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 996–1000.
[14] J. Yang, S. G. Naik, D. O. Ortillo, R. García-Serres, M. Li, W. E. Broderick, B. H. Huynh, J. B. Broderick, *Biochemistry* **2009**, *48*, 9234–9241.
[15] G. Sawers, B. Suppmann, *J. Bacteriol.* **1992**, *174*, 3474–3478.
[16] J. Buis, J. Broderick, *Arch. Biochem. Biophys.* **2005**, *433*, 288–296.
[17] M. Frey, M. Rothe, A. F. Wagner, J. Knappe, *J. Biol. Chem.* **1994**, *269*, 12432–12437.
[18] R. Külzer, T. Pils, R. Kappl, J. Hüttermann, J. Knappe, *J. Biol. Chem.* **1998**, *273*, 4897–4903.
[19] W. E. Broderick, B. M. Hoffman, J. B. Broderick, *Acc. Chem. Res.* **2018**, *51*, 2611–2619.
[20] C. R. Melchiorsen, K. V. Jokumsen, J. Villadsen, M. G. Johnsen, H. Israelsen, J. Arnau, *J. Bacteriol.* **2000**, *182*, 4783–4788.
[21] L. Zelcbuch, S. N. Lindner, Y. Zegman, I. Vainberg Slutskin, N. Antonovsky, S. Gleizer, R. Milo, A. Bar-Even, *Biochemistry* **2016**, *55*, 2423–2426.
[22] L. R. F. Backman, M. A. Funk, C. D. Dawson, C. L. Drennan, *Crit. Rev. Biochem. Mol. Biol.* **2017**, *52*, 674–695.
[23] C. M. Jäger, A. K. Croft, *ChemBioEng Rev.* **2018**, *5*, 143–162.
[24] H. Conradt, M. Hohmann-Berger, H. P. Blaschkowski, *Arc. Biochem. Biophys.* **1984**, *228*, 133–142.
[25] A. Becker, K. Fritz-Wolf, W. Kabsch, J. Knappe, S. Schultz, A. F. Volker Wagner, *Nat. Struct. Biol.* **1999**, *6*, 969–975.
[26] A. Becker, W. Kabsch, *J. Biol. Chem.* **2002**, *277*, 40036–40042.
[27] C.-H. Lee, R. H. Sarma, *J. Am. Chem. Soc.* **1975**, *97*, 1225–1236.
[28] R. L. Dordine, P. Paneth, V. E. Anderson, *Bioorg. Chem.* **1995**, *23*, 169–181.
[29] W.-J. Wu, P. J. Tonge, D. P. Raleigh, *J. Am. Chem. Soc.* **1998**, *120*, 9988–9994.
[30] C. V. Parast, K. K. Wong, S. A. Lewisch, J. W. Kozarich, J. Peisach, R. S. Magliozzo, *Biochemistry* **1995**, *34*, 2393–2399.
[31] F. Himo, L. A. Eriksson, *J. Am. Chem. Soc.* **1998**, *120*, 11449–11455.
[32] M. de F. Lucas, P. A. Fernandes, L. A. Eriksson, M. J. Ramos, *J. Phys. Chem. B* **2003**, *107*, 5751–5757.
[33] J.-D. Guo, F. Himo, *J. Phys. Chem. B* **2004**, *108*, 15347–15354.
[34] K. Čondić-Jurkić, V. T. Perchyonok, H. Zipse, D. M. Smith, *J. Comput. Chem.* **2008**, *29*, 2425–2433.
[35] Y. Peng, S. E. Veneziano, G. D. Gillispie, J. B. Broderick, *J. Biol. Chem.* **2010**, *285*, 27224–27231.
[36] A. V. Crain, J. B. Broderick, *J. Biol. Chem.* **2014**, *289*, 5723–5729.
[37] F. Mancia, P. R. Evans, *Structure* **1998**, *6*, 711–720.
[38] J. Pang, X. Li, K. Morokuma, N. S. Scrutton, M. J. Sutcliffe, *J. Am. Chem. Soc.* **2012**, *134*, 2367–2377.
[39] Y.-H. Chen, A. N. Maity, P. A. Frey, S.-C. Ke, *J. Am. Chem. Soc.* **2013**, *135*, 788–794.
[40] H.-H. Lo, H.-H. Lin, A. N. Maity, S.-C. Ke, *Chem. Commun.* **2016**, *52*, 6399–6402.
[41] K. R. Wolthers, S. E. J. Rigby, N. S. Scrutton, *J. Biol. Chem.* **2008**, *283*, 34615–34625.
[42] K. R. Wolthers, C. Levy, N. S. Scrutton, D. Leys, *J. Biol. Chem.* **2010**, *285*, 13942–13950.
[43] B. R. K. Menon, K. Fisher, S. E. J. Rigby, N. S. Scrutton, D. Leys, *J. Biol. Chem.* **2014**, *289*, 34161–34174.
[44] V. S. Bharadwaj, A. M. Dean, C. M. Maupin, *J. Am. Chem. Soc.* **2013**, *135*, 12279–12288.
[45] M. R. Nnyepi, Y. Peng, J. B. Broderick, *Arch. Biochem. Biophys.* **2007**, *459*, 1–9.
[46] A. Hospital, J. R. Goñi, M. Orozco, J. L. Gelpí, *Adv. Appl. Bioinform. Chem.* **2015**, *8*, 37–47.
[47] M. C. Childers, V. Daggett, *Mol. Syst. Des. Eng.* **2017**, *2*, 9–33.
[48] J. Knappe, G. Sawers, *FEMS Microbiol. Lett.* **1990**, *75*, 383–398.
[49] T. F. Henshaw, J. Cheek, J. B. Broderick, *J. Am. Chem. Soc.* **2000**, *122*, 8331–8332.
[50] W. Plaga, R. Frank, J. Knappe, *Eur. J. Biochem.* **1988**, *178*, 445–450.

[51] R. Anandakrishnan, B. Aguilar, A. V. Onufriev, *Nucleic Acids Res.* **2012**, *40*, W537–541.

[52] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, C. Simmerling, *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.

[53] Case, D. A, D. S. Cerutti, T. E. Cheatham III, T. A. Darden, R. E. Duke, T. J. Giese, H. Gohlke, A. W. Goetz, D. Greene, N. Homeyer, S. Izadi, A. Kovalenko, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. Mermelstein, K. M. Merz, G. Monard, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D. R. Roe, A. Roitberg, C. Sagui, C. L. Simmerling, W. M. Botello-Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu, L. Xiao, D. M. York, P. A. Kollman AMBER 2017, University of California: San Francisco, CA, **2017**.

[54] E. Vanquelef, S. Simon, G. Marquant, E. Garcia, G. Klimerak, J. C. Delepine, P. Cieplak, F.-Y. Dupradeau, *Nucleic Acids Res.* **2011**, *39*, W511–517.

[55] RED server http://upjv.q4md-forcefieldtools.org/REDDB/projects/F-91/, accessed Feb 15 **2017**.

[56] J. Wang, R. M. Wolf, J. W. Caldwell, P. Kollman, D. Case, *J. Comput. Chem.* **2004**, *25*, 1157–1174.

[57] R. M. Betz, R. C. Walker, *J. Comput. Chem.* **2015**, *36*, 79–87.

[58] C. I. Bayly, P. Cieplak, W. Cornell, P. A. Kollman, *J. Phys. Chem.* **1993**, *97*, 10269–10280.

[59] B. Mennucci, J. Tomasi, *J. Chem. Phys.* **1997**, *106*, 5151–5158.

[60] Gaussian 09, Revision D.01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox, Gaussian, Inc., Wallingford CT, **2009**.

[61] M. W. Mahoney, W. L. Jorgensen, *J. Chem. Phys.* **2000**, *112*, 8910–8922.

[62] B. P. Uberuaga, M. Anghel, A. F. Voter, *J. Chem. Phys.* **2004**, *120*, 6363–6374.

[63] J.-P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, *J. Comput. Phys.* **1977**, *23*, 327–341.

[64] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak, *J. Chem. Phys.* **1984**, *81*, 3684–3690.

[65] A. W. Götz, M. J. Williamson, D. Xu, D. Poole, S. Le Grand, R. C. Walker, *J. Chem. Theory Comput.* **2012**, *8*, 1542–1555.

[66] R. Salomon-Ferrer, A. W. Götz, D. Poole, S. Le Grand, R. C. Walker, *J. Chem. Theory Comput.* **2013**, *9*, 3878–3888.

[67] S. Le Grand, A. W. Götz, R. C. Walker, *Comput. Phys. Commun.* **2013**, *184*, 374–380.

[68] M. Ø. Jensen, S. Park, E. Tajkhorshid, K. Schulten, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 6731–6736.

[69] H. Xiong, A. Crespo, M. Marti, D. Estrin, A. E. Roitberg, *Theor. Chem. Acc.* **2006**, *116*, 338–346.

[70] S. Park, F. Khalili-Araghi, E. Tajkhorshid, K. Schulten, *J. Chem. Phys.* **2003**, *119*, 3559.

[71] G. Hummer, A. Szabo, *Biophys. J.* **2003**, *85*, 5–15.

[72] E. Chovancova, A. Pavelka, P. Benes, O. Strnad, J. Brezovsky, B. Kozlikova, A. Gora, V. Sustr, M. Klvana, P. Medek, L. Biedermannova, J. Sochor, J. Damborsky, *PLoS Comput. Biol.* **2012**, *8*, e1002708.

[73] B. J. Grant, A. P. C. Rodrigues, K. M. ElSawy, J. A. McCammon, L. S. D. Caves, *Bioinformatics* **2006**, *22*, 2695–2696.

[74] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graph.* **1996**, *14*, 33–38.