

RESEARCH ARTICLE | MAY 16 2022

Biotin-painted proteins have thermodynamic stability switched by kinetic folding routes

Special Collection: [2022 JCP Emerging Investigators Special Collection](#)

Frederico Campos Freitas; Michelli Maldonado; Antonio Bento Oliveira Junior; ... et. al



J. Chem. Phys. 156, 195101 (2022)

<https://doi.org/10.1063/5.0083875>



CrossMark

Articles You May Be Interested In

Emergence of allostery through reorganization of protein residue network architecture

J. Chem. Phys. (February 2023)

Intracellular self-assembly of TPE-biotin nanoparticles enables aggregation-induced emission fluorescence for cancer-targeted imaging

Chinese Journal of Chemical Physics (December 2018)



Time to get excited.

Lock-in Amplifiers – from DC to 8.5 GHz



Find out more



Zurich
Instruments

Biotin-painted proteins have thermodynamic stability switched by kinetic folding routes

Cite as: J. Chem. Phys. 156, 195101 (2022); doi: 10.1063/5.0083875

Submitted: 30 December 2021 • Accepted: 27 April 2022 •

Published Online: 16 May 2022



Frederico Campos Freitas,¹ Michelli Maldonado,² Antonio Bento Oliveira Junior,³
José Nelson Onuchic,³ and Ronaldo Junio de Oliveira^{1,a)}

AFFILIATIONS

¹ Laboratório de Biofísica Teórica, Departamento de Física, Instituto de Ciências Exatas, Naturais e Educação, Universidade Federal do Triângulo Mineiro, Uberaba, MG 38064-200, Brazil

² Departamento de Matemática, Instituto de Ciências Exatas, Naturais e Educação, Universidade Federal do Triângulo Mineiro, Uberaba, MG 38064-200, Brazil

³ Center for Theoretical Biological Physics, Rice University, BioScience Research Collaborative, 6566 Main St., Houston, Texas 77030, USA

Note: This paper is part of the 2022 JCP Emerging Investigators Special Collection.

a) Author to whom correspondence should be addressed: ronaldo.oliveira@uftm.edu.br

ABSTRACT

Biotin-labeled proteins are widely used as tools to study protein–protein interactions and proximity in living cells. Proteomic methods broadly employ proximity-labeling technologies based on protein biotinylation in order to investigate the transient encounters of biomolecules in sub-cellular compartments. Biotinylation is a post-translation modification in which the biotin molecule is attached to lysine or tyrosine residues. So far, biotin-based technologies proved to be effective instruments as affinity and proximity tags. However, the influence of biotinylation on aspects such as folding, binding, mobility, thermodynamic stability, and kinetics needs to be investigated. Here, we selected two proteins [biotin carboxyl carrier protein (BCCP) and FKBP3] to test the influence of biotinylation on thermodynamic and kinetic properties. Apo (without biotin) and holo (biotinylated) protein structures were used separately to generate all-atom structure-based model simulations in a wide range of temperatures. Holo BCCP contains one biotinylation site, and FKBP3 was modeled with up to 23 biotinylated lysines. The two proteins had their estimated thermodynamic stability changed by altering their energy landscape. In all cases, after comparison between the apo and holo simulations, differences were observed on the free-energy profiles and folding routes. Energetic barriers were altered with the density of states clearly showing changes in the transition state. This study suggests that analysis of large-scale datasets of biotinylation-based proximity experiments might consider possible alterations in thermostability and folding mechanisms imposed by the attached biotins.

Published under an exclusive license by AIP Publishing. <https://doi.org/10.1063/5.0083875>

I. INTRODUCTION

Biotin was discovered in mammals almost a century ago,¹ and it is one of our important vitamins among the complex B, also known as vitamin B₇ or H.² The effects of biotin deficiency in humans can be quite severe, including skin lesions, immune dysfunction, neurological disorders, and abnormal fetal development. However, biotin deficiency is rare once it can be found in several commonly used food products.³ Biotin serves as a coenzyme for five carboxylases involved in a wide range of key metabolic processes, primarily related to the catalysis of fatty acids, carbohydrates, and amino acids.⁴ After biosynthesis, proteins can undergo

post-translational modifications (PTMs) such as biotinylation, the attachment of a biotin molecule to a protein residue. Vitamin biotin is also involved in regulating gene expression by being covalently attached to lysine (K) protein residues of chromatin histones H2A, H3, and H4.^{5–7} Biotinylated chromatin condensates to a smaller nucleosome structure with less DNA being wrapped around the nucleosome, and it is repeatedly detected in repressed loci.⁸ There are other common PTMs, such as phosphorylation, acetylation, glycosylation, methylation, and ubiquitylation.⁹ PTMs often modulate protein functions by combinatorial formation of different protein complexes that might be derived from the same gene.¹⁰

An increasing list of biotechnological industry applications that isolate biopolymers, such as proteins and enzymes, has found on biotin (and its chemically modified versions) an effective tool as affinity tags.¹¹ *In vivo* biotinylation enzymatically couples the biotin molecule to a protein acceptor domain. One example is the high-affinity, on-rate, and specific binding between biotin and streptavidin/avidin protein molecules (dissociation constant, $K_D \sim 10^{-15} M$), which is used for protein immobilization.^{12,13} The biotinylated proteins are typically captured by (strept)avidin bound to beads, followed by elution of the sample, which leaves only the biotinylated proteins bound to (strept)avidin-beads. Modified proteins are then isolated by standard pull-down methods and identified by mass spectrometry. This biochemical assay is used in the purification and detection of several proteins in order to map the protein–protein interactions of a biological system of interest.¹¹ *In vivo* protein biotinylation technologies have been rapidly developed to study protein–protein interactions and proximity in living cells.^{14–20}

Proximity-labeling technologies based on biotinylation of protein partners are broadly used in proteomic methods that map the pathways of transient encounters of biomolecules at the level of their subcellular compartments.^{21–24} They rely on a concept: biotinylation is increased in protein targets close to the biotin-activating enzyme fused to bait proteins. These technologies are high-throughput assays since one single study can directly detect thousands of biotin sites in hundreds of proteins.^{25–29} The most common market available proximity-labeling technologies are the enzyme-mediated activation of radical sources (EMARS),^{30,31} the engineered ascorbate peroxidase (APEX),^{32–34} and the proximity-dependent biotin identification (BioID).^{35,36} EMARS and APEX demonstrated higher spatial and temporal resolution. However, the requirement of using H_2O_2 limits *in vivo* applications.²⁹

BioID is becoming the industry standard for proximal labeling by using a highly promiscuous mutated form of *Escherichia coli* (*E. coli*) biotin ligase (BirA) and its variants.^{21,23,35,37–44} The estimated effective biotinylation radius of BirA is around 1–1.5 Å.^{29,45} BioID biotinylates the surrounding lysine residues of the fished proteins by forming a highly reactive and short-lived biotinoyl-5'-AMP, which attaches covalently to the nucleophilic ϵ -amino side chain group of lysines. On the other hand, APEX biotinylates nearby tyrosine residues (Y). The *E. coli* enzyme BirA is a biotin protein ligase (BPL) that catalyzes the endogenous biotinylation of a specific lysine residue on acetyl coenzyme A (CoA) carboxylase.^{22,46} Figure 1 shows the biotinylation of the specific residue K122 contained in the biotin carboxyl carrier protein (BCCP), a domain of CoA. One disadvantage of BioID is the experiment time span (>16 h), once its slow kinetics requires high biotin concentration to generate sufficient biotinylated materials for proteomic analysis.⁴¹ This extra biotin environment in the subcellular compartments could act as additional crowding agents, resulting in more than one biotin-labeled site on the interacting protein partner.

Proteins typically coexist within a highly crowded environment⁴⁸ where affinities between individual proteins might be relatively weak. Yet, it is amazing how precise and efficient biotin-labeling uncovers the vast protein–protein interaction network in almost intact living cells.²² Protein biotinylation alongside other PTMs, cellular crowded environment, and many other aspects can influence proteins in several aspects, such as folding, binding,

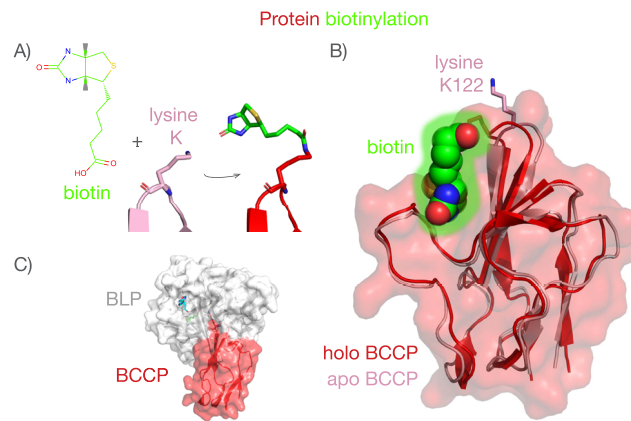


FIG. 1. Biotin Carboxyl Carrier Protein (BCCP) biotinylation. (a) A biotinyl group is attached to the apo biotin carboxyl carrier protein (apo BCCP) at the NH_3^+ side-chain group of the amino acid residue lysine 122 (K122), forming the (b) holo BCCP that contains the biotinyl-lysine residue (biotin in green). (c) The biotinyl group is post-translationally attached to BCCP by a biotinyl protein ligase (BPL), which catalyzes the formation of an amide linkage between the biotin carboxyl group and the N^6 -amino group of a special lysine residue. (a) and (b) present the apo and holo BCCP domains of the acetyl coenzyme A (CoA) carboxylase of *Escherichia coli*.⁴⁷ The *Pyrococcus horikoshii* OT3 BLP-BCCP dimer⁴⁶ is shown in (c). BirA is a promiscuous mutated BPL used in BioID-based proximity biotinylation methods.

protein mobility, thermostability, enzymatic kinetics, and gene expression.^{22,49–53} The biotinylated protein folding to the native state might be destabilized, interfering with other protein/enzyme wild-type characteristics. Under these circumstances, it is pertinent to evaluate the biotinylation influence on the protein folding thermodynamic stability and kinetics at the molecular level.

Recent investigations hypothesized that biotinylation-based proximity methods could be biased by structural features of target proteins.⁵⁴ Their results show the enrichment of cellular biotinylation events in intrinsically disordered regions (IDRs) of proteins in HEK293 cells, pointing to the re-purposing of *in vivo* biotinylation to reach broad conformational proteomics studies in intact cells. Additionally, they concluded that biotin proximity tagging favors local disordered portions in proteins. The authors suggest the application of biotin-painting as a method to solely gain insights into *in vivo* condition-dependent subcellular plasticity of proteins in large-scale datasets.⁵⁴ Yet, it is worth noting that saturated conditions of activated biotins and others covalently bound cofactors may alter the protein folding, structural integrity, binding dynamics, and cellular uptake.^{55–60} The extent of the modifications supplied by biotin-labeling sites covalently added to a typical enriched protein, which is pursued by a BirA fused protein, can be estimated by simulations of simplified computational models.

Theoretical models have been successfully employed in computer simulations of proteins modified by amino acid mutations and PTMs.^{61–65} Nanoscale molecular dynamics with a coarse-grained Martini model, designed to simulate experiments that exploit streptavidin-functionalized Au nanoparticles, have shown how biotinylated particles interfered critically in the diffusion process on lipid membranes. Biotin-labeling had significant effects on the lateral diffusion of the biotinylated lipids.⁶⁴ Two other protein

PTMs (glycosylation and ubiquitination) have been studied using the well-known structure-based model (SBM),⁶⁶ where the change in thermodynamic stability of the native state dynamics for the modified proteins could be successfully identified.^{61,62}

Structure-based models (SBMs, also known as Gō models) are simplified/minimalist models underpinned by the energy landscape theory.^{67–69} In this theory, the underlying protein folding energy landscape is multidimensional in nature and funneled toward the native state, which is defined as the surface global minimum.^{70,71} The stochastic folding process is best portrayed in terms of diffusion along a rough downhill funneled landscape.^{72–78} Since the structural techniques can only resolve low free-energy states, the energetics of SBMs are defined to stabilize the given structures instead of assigning energetic interactions based only on the chemical composition (as usually performed by semi-empirical models).⁷⁹ The SBM properties of the energy landscape are encoded in smooth energy functions through the use of a simplified representation instead of including every type of physical interaction of fully atomistic models.^{80–82} Despite the simplifications, SBM simulations are able to fully characterize the protein density of states and return the thermodynamic and kinetic profiles by converging many long runs at different temperatures^{71,83} and have been largely used over the last three decades to uncover many biological mechanisms related to folding and function, such as domain swapping, binding mechanisms, and functional conformation dynamics.^{84–95}

Subcellular biotinylation patterns in proteins vary with the extent of their local flexibility: intrinsically disordered proteins (IDPs) tend to have a higher number of attached biotins, opposing to completely folded proteins that are prone to have less biotin-labeled sites.⁵⁴ As reported by large-scale datasets of biotinylation-based proximity experiments, some proteins can be labeled by one or a few biotins, covalently added to lysine (K) or tyrosine (T) amino acid side chains.^{25–28} The steric and entropic effects of biotinylation might impact the protein thermodynamics with possible changes in folding routes, the folding temperature, and the free-energy barrier. Thus, cutting-edge experiments of biotin proximity tagging, aiming to investigate the *in vivo* protein plasticity and the protein–protein interaction patterns, cannot neglect these assumptions when designing and analyzing such experiments. In this context, we propose that structure-based models can be used to investigate biotin-labeling effects on proteins' energy landscape and gain information on their thermodynamic and kinetic features.

II. RESULTS AND DISCUSSION

A. Selection of the simulated proteins and biotinylation sites

The first chosen protein was the biotinyl domain of acetyl-coenzyme A (CoA), also known as the biotin carboxyl carrier protein (BCCP). BCCP is a typical biotin-dependent enzyme with three functional domains related to a carboxybiotinyl intermediate that catalyzes the first step of fatty acid biosynthesis. The biotinyl BCCP domain transfers its covalently attached biotin from the active site of the biotin carboxylase domain to the carboxyl transferase domain.⁹⁶ As shown in Fig. 1, BCCP is a capped β -barrel sandwich structure with the biotinylated lysine located at the hairpin β -turn that connects the N-terminal and the C-terminal halves. It is a relatively

small 80-residue protein with a common fold. Figure 1(b) shows the *E. coli* BCCP structures extracted from the Protein Data Bank for the holo (PDB code, 1BDO⁹⁶) and apo (PDB code, 3BDO⁴⁷) biotinyl domains. Holo (with biotin) and apo (without biotin) BCCP structures are very similar with a backbone root-mean-square deviation (RMSD) of 1.3 Å when both are compared.⁴⁷ The exposed biotinyl-lysine residue is the major difference between the apo and holo BCCPs, which makes these two natural proteins suitable candidates for a test case of the biotin influence on protein thermodynamic stability.

The not biotinylated FK506-(and rapamycin)-binding protein (FKBP3),²⁶ a *cis-trans* peptidyl-proline isomerase enzyme involved in cellular protein folding, was also selected to investigate the effects of biotin “painting.” FKBP3 contains two globular domains connected by a long disordered and flexible loop (Fig. 3). The N-terminal domain is a structural 5-helical bundle nucleoplasmin-like (NPL) domain. The C-terminal domain packs six β -strands and one α -helix that tightly binds to the immunosuppressive drugs FK506 and rapamycin (FKBD).^{97,98} A novel function for a subclass of FKBP3s was recently discovered, which facilitates chromatin binding by being recruited to link regions of chromatin.⁹⁹ FKBP3 was found biotinylated on four independent, recent, and orthogonal large-scale proximity proteomics studies by labeling biotins in lysine^{25,26} and tyrosine^{27,28} residues. Numerous directly identified biotinylation sites targeting different subcellular niches by independent biotin–peptide enrichment strategies were found.⁵⁴ In this work, the biotin K-labeled occurrences identified by two proximity tagging experiments were chosen to be investigated: DiD-BiT²⁵ and SpotBioID²⁶ (Fig. S4). In total, 23 biotin sites along the FKBP3 residue sequence (Fig. 3) were gathered. These identified sites were “painted” with biotin to understand how the steric and entropic aspects of biotinylation can change the FKBP3 dynamics and thermodynamic properties.

B. Apo and holo biotinyl domains show distinct thermodynamic and kinetic folding profiles

As a biotinyl domain of CoA, the biotin carboxyl carrier protein (BCCP) has its biotin moiety catalyzed by a biotin carboxylase enzyme. There are some carboxylase partners for the BCCP domain with different structures, oligomerizations, and sequences in different organisms^{46,100–102} although they are highly conserved enzymes. Even though *Escherichia coli* has shown to be a model organism to study biotin biosynthesis, BCCP has its structure and sequence conserved among different animals, as shown in Fig. S1. Yet, for most organisms, the significant difference is the lack of the thumb domain when compared with the *E. coli* BCCP.⁴⁷ Figure S1B shows an extremely conserved amino acid residue sequence surrounding the biotinyl K122 residue in both apo⁴⁷ and holo⁹⁶ *E. coli* BCCPs. This homology is also reflected in their respective structures, which can be seen when the BCCPs are aligned (Fig. S1C).

The all-atom structure-based model (SBM) was employed separately on both BCCP protein structures, which consider all non-hydrogen atoms. A batch of very long simulations at different temperatures of apo and holo BCCP was run, commonly exceeding 50 μ s. Figure 2(d) presents the changes in the number of native contacts (*Q*) monitored along the simulated molecular dynamics time

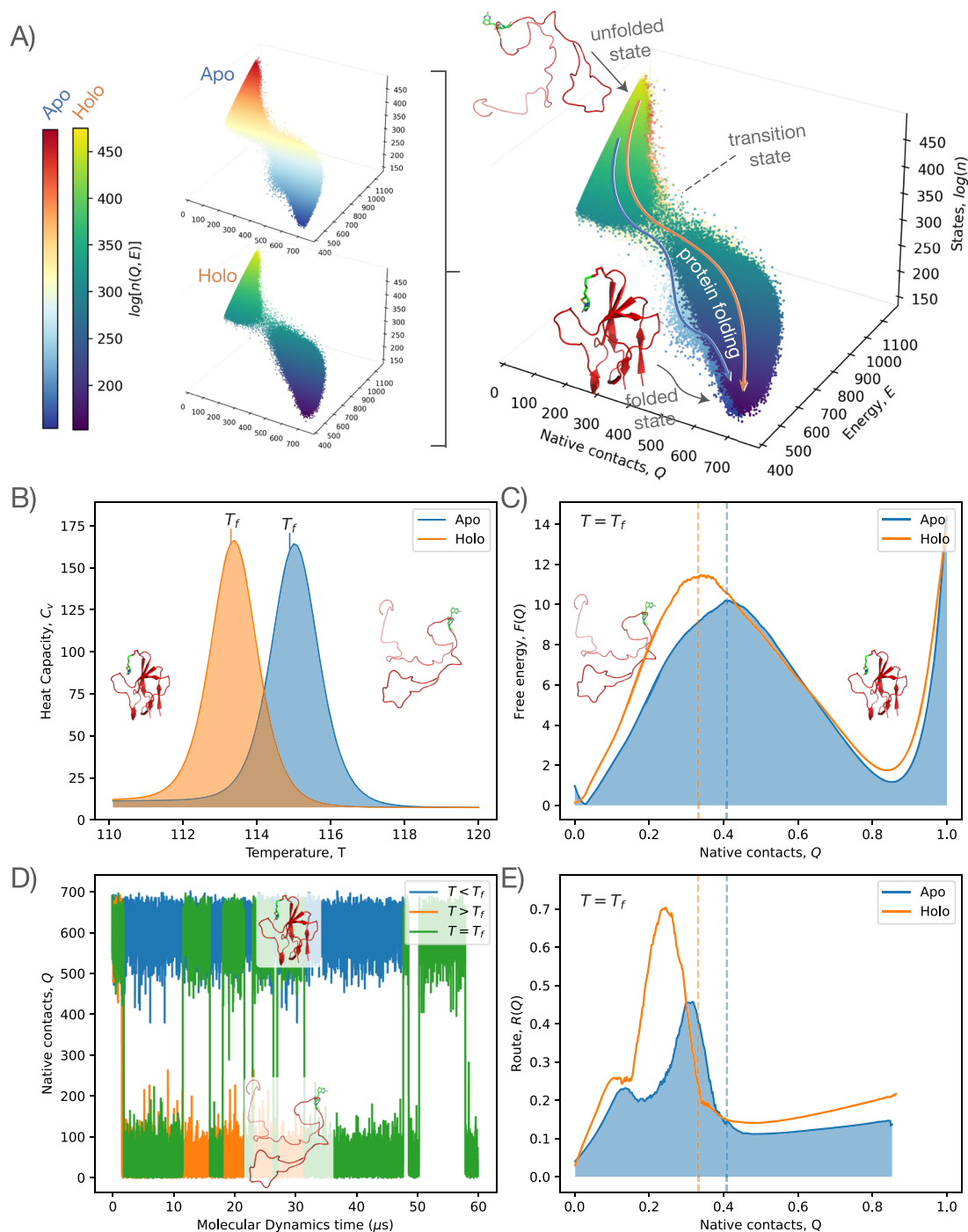


FIG. 2. BCCP folding thermodynamic and kinetic analyses. (a) Folding energy landscape of the two (apo and holo) BCCPs portrayed by the density of states, $\log[n(Q, E)]$. Q is the reaction coordinate (the number of native contacts), and E is the total energy. The two n surfaces on the left are superposed on the right with the colors representing the logarithm of the number of states for a given (Q, E) . The continuous lines over the surfaces illustrate possible protein folding routes. (b) Specific heat (C_v) as a function of the temperature of the simulated apo and holo enzymes. The peak of each curve corresponds to the critical transition temperature (T_f) between folded (low temperatures) and unfolded (high temperatures) states of each BCCP. (c) Free-energy profiles as a function of the normalized Q [$F(Q)$] at the respective protein T_f . (d) Trajectory time traces in Q of the apo BCCP at three different temperatures. (e) Folding routes along Q [$R(Q)$] at the folding T_f of each protein. Vertical dashed lines were placed at the respective protein F peak, representing the transition state (TS) maximum energetic barrier. Holo BCCP has a higher and shifted F barrier with increased specificity in its folding route. Energy is in units of $k_B T_f$ with T_f being the respective system's folding temperature in GROMACS reduced units.

at three chosen temperatures: a “low” temperature in which the protein is folded all the time (blue); a “high” temperature in which it is unfolded (orange); and the transition temperature (T_f) between folded and unfolded ensembles (green). In protein folding simulations, Q is a common choice of reaction coordinates because it evaluates the nativeness, quantifying how similar is a given structure in time compared to the native structure.⁹⁵ Figure 2(d) also shows that some transitions took about 10 μ s to complete a folding/unfolding event at the critical T_f . Regarding or not the biotinyl moiety of BCCP, these are long transition times for proteins around this size, which typically fold one or two orders of magnitude faster on similar SBM simulations.^{71,103–109}

A wide range of temperatures for each group of BCCP runs (apo and holo) were separately analyzed using the weighted histogram analysis method¹¹⁰ through a Python framework (PyWHAM).¹¹¹ By converging all the trajectory histograms,¹¹² WHAM computes the microcanonical density of states, which is the base of all the thermodynamic quantities to be further analyzed. Figure 2(a) shows the density of states (n) for the apo and holo BCCPs in the logarithm scale, projected onto two coordinates: the number of native contacts (Q) and the total energy (E). n of the biotinylated (holo) BCCP is shown over the apo BCCP, revealing the energy landscape for the simulated data projected along Q and E . Ideally, the energy landscape has a bottleneck at the transition state between the folded state (at the funnel bottom) and the unfolded state (at the top of the surface), which is funneled in entropy.^{67,71} In Fig. 2(a), $n(Q, E)$ bears resemblance with the folding landscape hypothesized by the theory. Considering that n is in the log scale, Fig. 2(a) presents n as a funneled surface with numerous accessible states at the top (unfolded states with low Q and high E), which are dramatically reduced as both proteins fold en route to the native state (folded state with high Q and low E). The transition state emerges in n as a bottleneck of an even reduced number of accessible states, which could also be visually spotted by the reduced number of dots between the folded and unfolded states. The holo BCCP density of states at the transition state is compressed with less accessible states compared with the apo BCCP, which also has more diverging states at the folded basin around $Q = 400$. The addition of biotin to BCCP reduces the number of accessible states for the protein folding along the energy landscape surface.

Figure 2(b) presents the specific heat (C_v) as a function of the temperature for the apo and holo BCCPs. The thermodynamic $C_v(T)$ is a direct result of the obtained density of states with each C_v peak defining the critical phase transition temperature (T_f) between folded (low temperature) and unfolded (high temperature) states. In Fig. 2(b), the biotinylated BCCP has shown a decreased folding temperature T_f although the shape and height of the two curves were practically identical.

Figure 2(c) shows F as a function of the normalized number of native contacts (Q) for the apo and holo BCCPs at each respective folding temperature (T_f). Both proteins' all-atom simulations resulted in thermodynamic F profiles characteristic of two-state folders,⁷⁸ i.e., two minima, one representing the unfolded ensemble (the basin located at $Q \sim 0.5$) and the other representing the folded state (at $Q \sim 0.9$). Between the minima, there is an energetic F barrier representing the transition state for folding, which is the same transition state seen in Fig. 2(a). When compared with the apo BCCP

free-energy profile, the holo BCCP displayed a higher free-energy barrier ($\Delta F = 2k_B T$) and shifted (Q decreased from 0.40 to 0.35).

It is often a difficult task for experimentalists to obtain projections of this landscape on one-dimensional free-energy profiles using the number of native contacts as a reaction coordinate due to their intrinsically multidimensional nature.¹¹³ Consequently, experimental works more frequently study F over extension-based coordinates.¹¹⁴ Thus, our simulations were also monitored along with two other reaction coordinates: radius of gyration (R_g) and root-mean-square deviation (RMSD). On average, the most extended/deviated structures obtained in the simulations showed R_g and RMSD above 4 nm (Figs. S2A and S2B). Free-energy landscape projections at T_f for R_g and RMSD showed almost the same energetic barrier of nearly $10k_B T$ for both BCCP proteins (Fig. S2C). However, the basins related to the unfolded state for R_g and RMSD were shifted to higher values for the apo BCCP, suggesting that the non-biotinylated protein has an unfolded ensemble comprising more extended states. Free-energy barriers found for both BCCPs in all reaction coordinate projections (around 10 – $12k_B T$) are considered high when compared with proteins of similar sizes and folds.^{71,103–109} The estimated μ s dynamical time-scale regime in Fig. 2(d) could be explained by these high F energetic barriers in which the proteins could take a long time to overcome a back and forth motion between the folded/unfolded basins. No direct biophysical experiment determining free energy was found. However, experimental assays have shown that BCCPs are generally stable structures in solution.¹¹⁵

Figure S3 shows the amino acid root-mean-square fluctuation (RMSF) for the simulated apo and holo BCCPs alongside the x-ray crystallographic B-factors as a measure of local mobility of the holo BCCP.⁹⁶ The RMSF from the simulations corroborates with the B-factors seen in Fig. S3. Small differences can be observed and could probably be due to the dynamic nature of the all-atom SBM, opposing to the static and packing effects of the crystal. Simulated RMSFs of both proteins are comparable, as shown in Fig. S3. The small range of exhibited RMSF values (between 1.0 and 3.5 Å), which is of the order of the crystal structure and nuclear magnetic resonance (NMR) resolutions,^{47,96} agrees with the experimental results on protein residue fluctuations. Both BCCPs had increased fluctuations in the surroundings of the biotin-binding region, given by the turns around K122, the protruding thumb between β_2 and β_3 , and the residues between strands β_6 and β_7 . Experiment and simulations point to further evidence for small, localized conformational changes upon post-translational modification.

Few structural changes were found on native conformations upon biotinylation. However, this post-translational modification delivers different functionalities for the protein, and the simulations showed dissimilarities in the folding thermodynamics when compared with apo BCCP. The biotin attached in the biotinyl domain of CoA alters the protein density of states by decreasing the number of accessible states during folding. These effects had a direct impact on the number of protein possible folding routes. Figure 2(e) shows the route measurement as a function of Q [$R(Q)$] for the simulated proteins. $R(Q)$ is a kinetic measurement that quantifies the average number of folding routes at a given position in Q .^{48,94} For this reason, the number of native contacts formed at a given time (Q) is a suitable reaction coordinate by monitoring the protein folding process using its “nativeness degree.”¹¹⁶ During folding,

the protein backbone adopts a low native-like topology in the early stages (lower Q) and progressively forms more native-like contacts as secondary structure elements approach the native conformation (higher Q).¹¹⁷ As folding proceeds, native contacts are formed by bringing closely structured elements far in the protein sequence, also forming a small number of non-native contacts. In most cases, folding proceeds by a single dominant route in which native-like elements appear in an order correlated with their propensity to form in the unfolded state.¹¹⁷ According to the theory, the transition state acts as a filter selecting most native-like routes for protein folding

and binding.^{85,118–120} R profiles on Fig. 2(e) reveal these features. For both BCCPs, there is an increase in specificity given by the increase in R before the peak of the transition state [dashed line aligned with the top of the F barrier in Fig. 2(c)]. The holo BCCP $R(Q)$ peak (before the transition state) is higher, suggesting fewer folding routes being accessed when compared with the apo BCCP. There were a few possible routes to cross the transition state that selects most native-like routes for folding. After crossing the transition state, both BCCPs show a downhill motion in F to the native state direction, while the route values abruptly decrease and remain

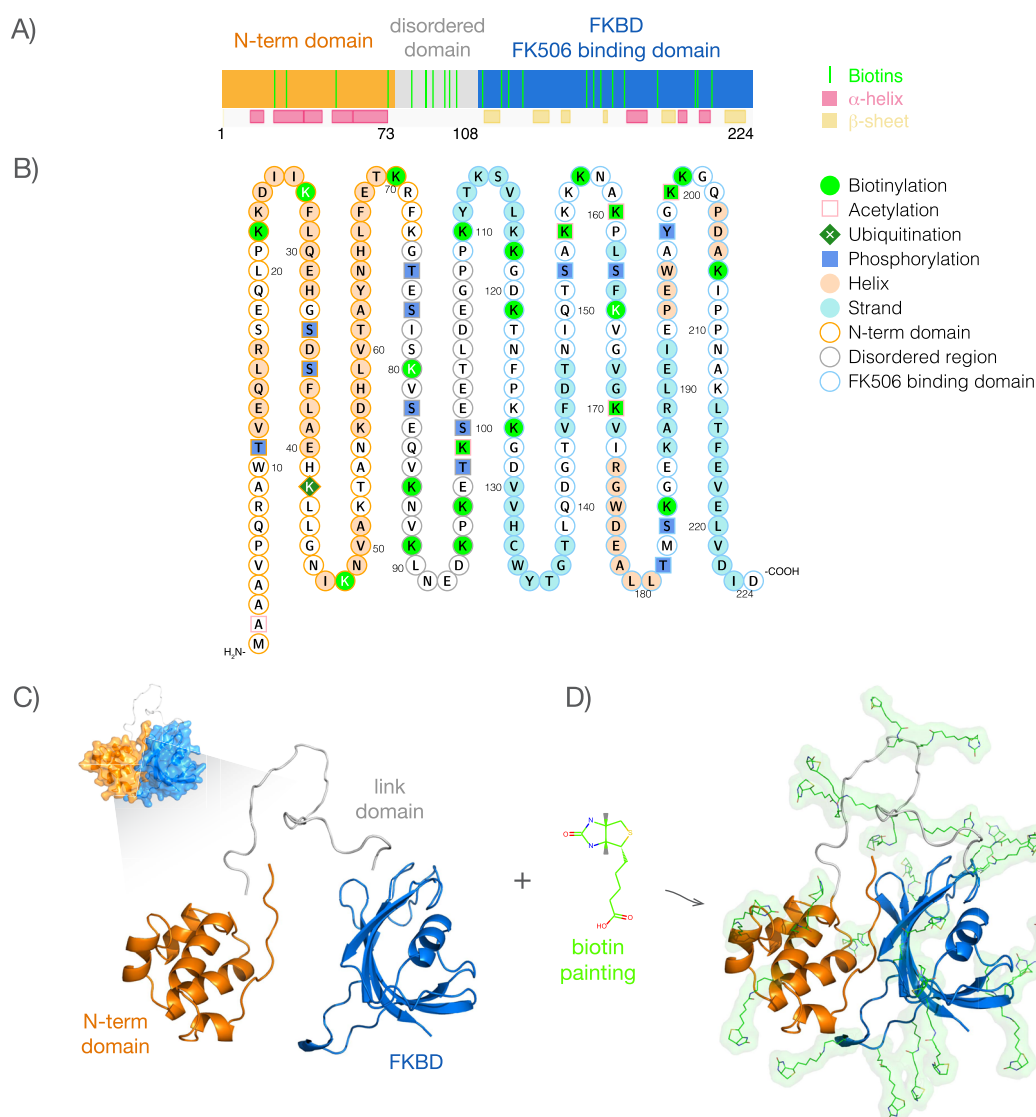


FIG. 3. Biotinylation of the FK506 binding protein 3 (FKBP3). (a) and (b) show the protein secondary structural content and the three FKBP3 domains in the one-dimensional representation: N-terminal domain (orange), disordered link domain (gray), and FK506/rapamycin binding domain (FKBD). 23 biotinylated lysines are marked on the protein sequence.⁵⁴ (b) The primary sequence of FKBP3 with the post-translational modifications (PTMs) found experimentally in the Uniprot data bank (accession code, Q00688); except for biotinylation sites that were extracted from the biotin-labeled occurrences found by the two biotin proximity tagging experiments: DiDBiT²⁵ and SpotBioID.²⁶ (c) Three-dimensional structure of FKBP3 exploded view with the three domains. (d) FKBP3 structure with the biotin-painted residues.

roughly constant (and with approximately same values) up to the folded free-energy minima. Folding thermodynamics and kinetics were altered by biotinylation, specially by changing the number of available states in the unfolded ensemble.

C. FK506 binding protein 3 is biotinylated across four independent biotin proximity tagging experiments

The FK506 binding protein 3 (FKBP3) is highly susceptible to post-translational modifications. The Uniprot data bank lists the common found PTMs acetylation, ubiquitination, and phosphorylation, as shown in Fig. 3(b) where they are marked in the protein sequence. Large-scale proximity proteomics studies (DiD-BiT,²⁵ SpotBioID,²⁶ BioSITE,²⁷ and AB-APEX2²⁸) also found many occurrences of these aforementioned PTMs in FKBP3, including 13 phosphorylation, 5 ubiquitination, and 4 acetylation sites, differing from Uniprot only by one ubiquitinated residue. FKBP3 in Fig. 3 represents a reasonable model for folding studies as it contains two globular domains (N-term and FKBD) with structured α -helix, β -sheet, and coil parts, joined by one IDR. Figures 3(a) and 3(b) show the PTM distribution on these three domains, including 23 biotinylation sites on the protein sequence. We chose to investigate only the K-labeled datasets, DiDBiT²⁵ and Spot-BioID,²⁶ in order to not mix with Y-labeled effects of BioSITE²⁷ and AB-APEX2.²⁸ This choice led to a small difference of biotinylation sites mentioned in the recent biostatistical biotin tagging study.⁵⁴ Lysine and tyrosine are amino acids of opposite types: the former (K) promotes protein disorder, and it is a residue normally exposed on the protein surface. The latter (Y) is not common in IDRs and is commonly found buried in the protein core.^{121–123} Just one biotin tag on the folded structure was found on each of the two Y-labeled datasets, opposing to the many K-labeled occurrences retrieved by the two other biotin proximity tagging experiments shown in Fig. S4. Among the FKBP3 biotin-tagged sites, around 70% are within IDRs, local coil structures, or short and exposed α -helical segments of the protein three-dimensional structure. The remaining are within stable and structured α -helix bundles or β -sheet motifs.

D. FKBP3 biotin painting progressively increases the folding temperature

The full FKBP3 x-ray crystallographic structure (PDB code 2MPH) was selected for the simulations. In order to generate the biotinylated versions, the wild-type (apo) protein was progressively “painted” by attaching one biotin in each lysine residue mapped in Fig. S4. All FKBP3 versions (apo and biotinylated) were extensively simulated by using the all-atom structure-based model (SBM) representation, which considers all non-hydrogen atoms. It was executed as a group of runs covering a wide range of temperatures for each biotin-painted addition. The simulations were quite long, mostly between 2 and 4 μ s, with some runs reaching more than 15 μ s. Normally, it used the same running time for all cases. Nonetheless, there were some reasons to opt for these two ranges of running times. FKBP3 is a long 224-residue multi-domain protein that conventionally would not be chosen for folding/unfolding simulations. Typically, large proteins (with more than one hundred residues) face high energetic barriers for folding transitions, jeopardizing thermodynamics and kinetic analyses. Another layer of difficulty added

by FKBP3 is the difference between the time scales involved to fold/unfold each of the two globular domains. Similar cases frequently use enhanced sampling techniques to overcome the barriers imposed by such limitations.^{91,124–126} Instead, we preferred to run “brute force” simulations, meaning to wait for spontaneous transitions to occur, even if it took a few months to complete. In addition to that, the computational cost to simulate and to store all the data using the longest simulation time for all runs is prohibitive.

FKBP3 simulations started with the wild-type protein structure (apo with no biotin added). After running a batch comprising a wide range of temperatures, one biotin was then added at the lysine biotinylation site previously identified. For each new biotin added, a new set of simulations was run. The sets were nominated according to the number of FKBP3 biotin-painted residues (from 00 to 23), as shown in Fig. 4. Thermodynamic analysis was performed using the WHAM algorithm^{110,111} by following the same protocol employed in the early BCCP section. Figure 4(d) shows the specific heat as a function of the temperature [$C_v(T)$] for all cases, while Fig. 4(e) highlights the set of FKBP3 simulations with ten biotins added. The two C_v peaks (temperatures T_{f1} and T_{f2}) appeared on all simulated sets (apo and holo FKBP3) with both temperatures monotonically increasing, as shown in Figs. 4(d), S5A, S5C, and S5D. Vertical lines in Fig. 4(e) are equivalent to the simulated temperatures shown in Fig. 4(a) using the transition temperature of each set accordingly. There were minimal oscillations on the C_v profiles with small deviations in the curves’ height and width.

Figure 4(a) presents a subset of six temperature runs out of 40 (the total set) of the full FKBP3 folding/unfolding simulation trajectories in which the chosen reaction coordinate is the number of native contacts formed (Q). Figure 4(b) projects a normalized histogram of each time series shown in Fig. 4(a). Each of the two globular FKBP3 domains undergoes folding transitions in different time and temperature scales, resulting in two critical transitions and, consequently, two defined folding temperatures: T_{f1} and T_{f2} . For temperatures $T < T_{f1}$, FKBP3 remains fluctuating around its completely folded structure during the simulation, as illustrated by the trajectory in blue in Fig. 4(a). In FKBP3, Fig. 4(c), its corresponding representative structure is marked by ①. Other representative structures also shown in Fig. 4(c) were marked up to ⑤ with colors matching their respective temperature in Figs. 4(a) and 4(e). At $T = T_{f1}$ (green time series), the enzyme has a dynamic behavior that alternates between its completely folded state and an ensemble where only the N-terminal domain is unfolded. At $T > T_{f1}$ [orange time series in Fig. 4(a)], the simulation shows bound and unbound N-term tail structures. The unbound N-term tail ensemble, representative structure ③, was generated at a temperature that completely unfolds the N-term domain while the protein C-terminal domain remains folded most of the time even though showing a slight increase in its dynamical state.

The red curve in Fig. 4(a) shows the FKBP3 simulation close to the second transition temperature ($T < T_{f2}$) and represents an ensemble with an even higher dynamical C-term domain and the N-term domain completely unfolded. The C-terminal (FKBD) domain starts to appear less structured even though FKBD still retains its globular shape, preserving some secondary structures, as seen in the representative structure ④ [Fig. 4(c)]. At $T = T_{f2}$ (brown curve in Fig. 4(a), the globular FKBD sampled folded and unfolded

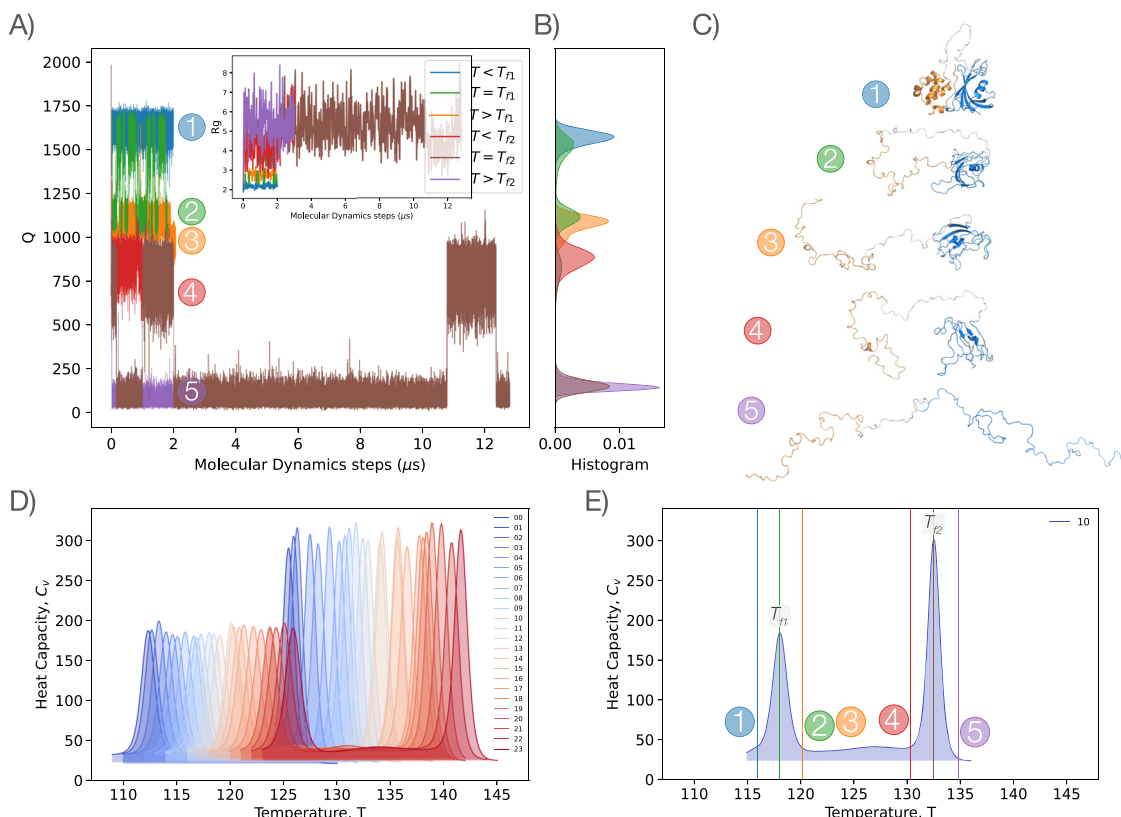


FIG. 4. (a) Trajectory time traces of apo FKBP3 with the number of native contacts $[Q(t)]$ and the radius of gyration $[R_g(t)]$, inset, monitored along the simulated time at six different temperatures. (b) Density of sampled Q in (a). (c) Representative structures of the most common sampled states for each FKBP3 simulation temperature. Numbering is consistent among all (a), (c), and (e). (d) Specific heat as a function of temperature $[C_v(T)]$ of the simulated apo (00, without biotin) and all biotinylated (with 01 up to 23 biotins) FKBP3 enzymes. (e) $C_v(T)$ of FKBP3 for the case of ten biotins added. Vertical lines were placed at the corresponding six simulated temperatures shown in (a). The two peaks of each curve were defined as two transition temperatures (T_{f1} and T_{f2}) in which they correspond to the folding/unfolding transitioning of the N-term and FKBD domains, respectively, of the FKBP3 protein. Energy is in units of $k_B T$, and temperature is in GROMACS reduced units.

states, the latter representing the completely unfolded FKBP3 conformations (conformation ⑤). For temperatures above T_{f2} ($T > T_{f2}$), the enzyme is completely unfolded and unstructured during all the simulated time. Figure 4(b) exhibits histograms reflecting the normalized distributions of Q sampled in each temperature run shown in Fig. 4(a). The histogram distributions are dual-peaked curves at the critical temperatures T_{f1} (green) and T_{f2} (brown), demonstrating the consecutive folding/unfolding transitions of the N-term and C-term domains, respectively. The N-terminal domain has 75 amino acid residues that possess a folding/unfolding transitioning dynamics on the ns regime, whereas the 118-residue FKBD showed much slower transitions (on the μ s timescale).

There was no specific reason for the choice to attach biotin to FKBP3 in the sequential order, as it was performed in this work. In real-world situations, biotinylation might occur on lysine or tyrosine residues of proteins close to the proximal fused biotin ligase enzymes, and probably, it happens in a random order. Proximity-directed labeling experiments retrieve the washed peptides corresponding to a portion of the whole protein. After biostatistical analysis, each peptide can have one or a few biotinylated sites that are

accounted for in the full biotinylated protein. It would be impractical to simulate all the possible combinations of having one up to 23 biotin-painted full FKBP3 proteins as the simulations demonstrated a high usage of computational time and disk space for each run. Nonetheless, our theoretical experiment disclosed a trend in the thermodynamic stability of the globular domains as a function of the biotin-painted systems.

E. FKBP3 biotin painting alters the protein energy landscape

The WHAM density of states can be obtained by converging all trajectory histograms in different temperatures for a given protein. Figure 5(a) presents the density of states (n) as a function of Q and E for the apo (00) and biotinylated (with 23 biotins) FKBP3 enzyme. In the supplementary material, Fig. S8 contains the density of state (n) landscapes of all FKBP3 labeled cases. In Figs. 5(a) and S8, n is presented for each biotin-painted enzyme (colored from yellow to blue) over the apo protein (colored from red to blue). Both monitored coordinates, Q and E , have a wide range of values as FKBP3

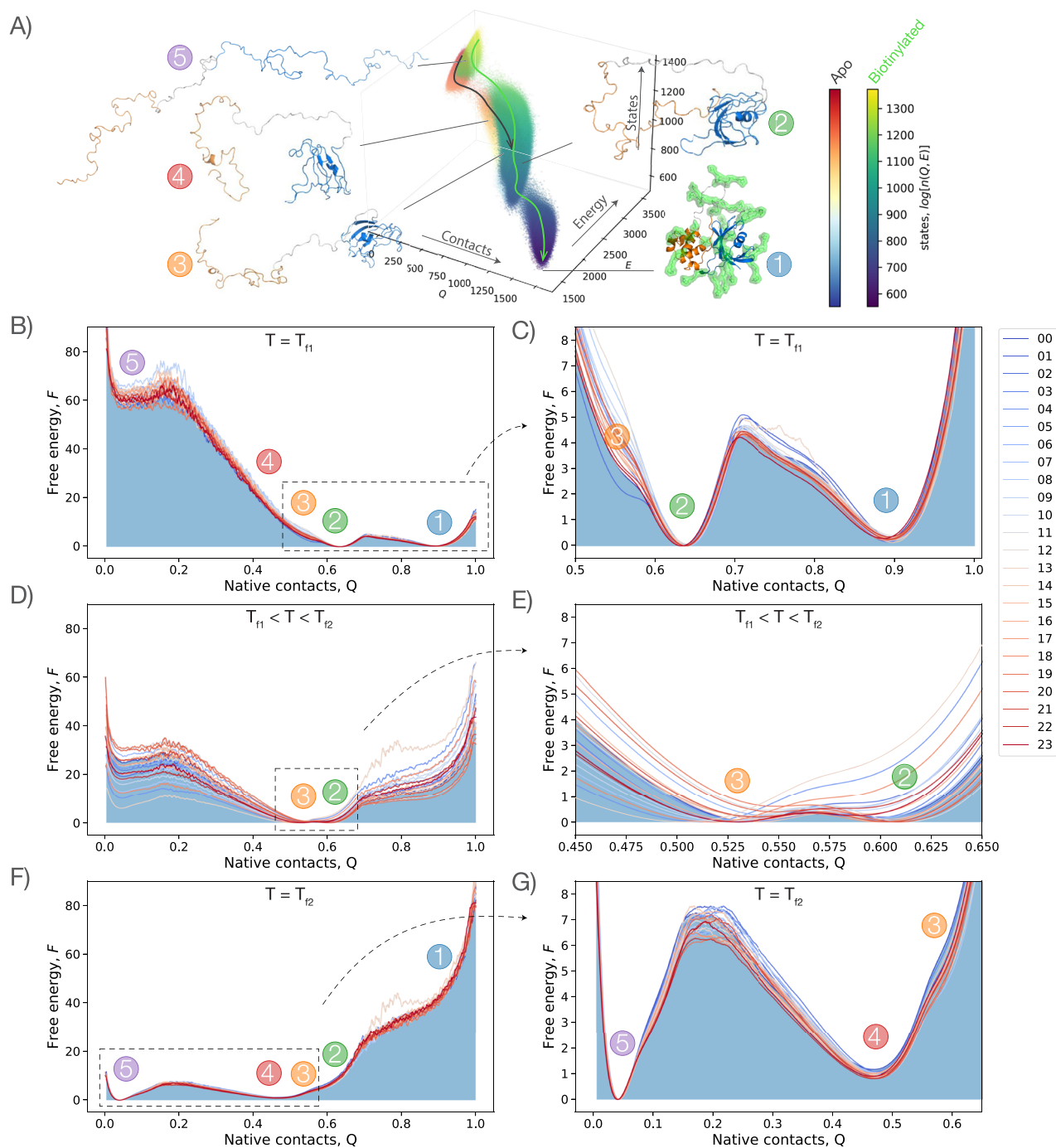


FIG. 5. (a) Folding energy landscape of the apo (without biotin) and biotinylated (with 23 biotins) FKBP3 enzymes portrayed by the density of states, $\log[n(Q, E)]$. Q is the number of formed native contacts, and E is the total energy. (b)–(g) show the free-energy profiles along the normalized native contacts $[F(Q)]$ of the modified full FKBP3 enzymes for three different temperatures: $T = T_{f1}$ (top panels), $T_{f1} < T < T_{f2}$ (middle panels) and $T = T_{f2}$ (bottom panels). Right panels are zooming the two states with near equal probabilities of transition, where sampling was enhanced for each case. Panels (b) and (c) represent the folding/unfolding of the N-term domain; panels (d) and (e) show the transition of unbinding the most extreme part of the N-term segment out of the C-term FKBD interface; bottom panels (f) and (g) display the folding transition of the FKBD domain. Representative structures of each most sampled states of FKBP3 in (a)–(g) are the same as in Fig. 4. F curves are shown for FKBP3 without biotin (00) and with an increasing number of biotinylated sites (01 up to 23). Energy is in units of $k_B T_f$ with T_f being the respective system's folding temperature in GROMACS reduced units.

is a multi-domain enzyme, spanning many native contacts between atoms of the same and different domains. The energy landscapes of the apo FKBP3 and the modeled biotin-painted enzymes resulted in surfaces with three wide domains narrowed by bottlenecks. The first region at the top of the n surface (low Q and high E) represents completely extended FKBP3 conformations. By crossing to the next region at intermediary Q and E , the FKBD domain folds, and thus, this first bottleneck crossing region can be considered as a transition state. The last dense region at high Q and low E is related to completely folded/native conformations of the full enzyme. Crossing from the intermediary region in n to the native FKBP3 folds the N-term domain, and thus, this bottleneck is another transition state. Both transition states are sparse compared to the states with each folded domain, which restrict passage of the chain between the three more dense regions.

It is possible that FKBP3 folding is sequential from the C-term to the N-term domain. Inspection of $n(Q, E)$ shows that the density of states of the apo FKBP3 changes upon each biotinylation. The differences between apo and biotinylated proteins mostly appear at the C-term (FKBD) states related to its folding event. Folded states of the FKBP3 C-terminal domain remain almost the same. Consequently, SBM simulations suggest that the FKBP3 energy landscape is funneled toward folding of the drug binding FKBD domain, followed by folding of the N-term domain. The energy surface contains two transition states, and the surface changes upon biotin painting mostly around the C-term folding states.

Panels in Figs. 5(b)–5(g) exhibit the free-energy profiles as a function of the normalized number of native contacts $[F(Q)]$ for the apo (00) and biotinylated FKBP3 (01–23). F is displayed for three temperatures: T_{f1} [Figs. 5(b) and 5(c)], $T_{f1} < T < T_{f2}$ [Figs. 5(d) and 5(e)], and T_{f2} [Figs. 5(f) and 5(g)]. Each FKBP3 modeled system revealed its own T_{f1} and T_{f2} ; thus, each F is shown at the respective critical temperature of the corresponding panel. F curves in Figs. 5(b)–5(g) show three minima separated by transition state barriers with the barrier minimum magnitudes changing with the respective temperature. Two free-energy profile $[F(Q)]$ minima correspond to the N-term domain folded ($Q \approx 0.90$) and unfolded ($Q \approx 0.65$) with the C-term domain remaining folded. There is one F barrier at the transition state with a peak around $Q \approx 0.70$ between these two ensembles. The other minimum at $Q \approx 0.10$ is related to the completely unfolded FKBP3 in which FKBD is the last domain to unfold. The F barrier adjacent to the completely unfolded state arises at $Q \approx 0.20$ with a high peak.

At $T = T_{f1}$, in Figs. 5(b) and 5(c), biotin painting simulations of FKBP3 resulted in F barriers for N-term folding/unfolding dynamics ranging between 4 and $5k_B T$. In addition, unfolding of the C-term drug-binding domain exhibited barriers between 55 and $70k_B T$. At T_{f1} , N-term folding/unfolding event sampling is enhanced, opposing to the low sampling for FKBD transitions due to the nearly prohibitive high F barrier at this temperature. Figures 5(d) and 5(e) attempt to find a temperature in which sampling of the unbinding transition of the N-term tail segment out of the C-term interface is intensified. In Figs. 5(d) and 5(e), this temperature is between $T_{f1} < T < T_{f2}$ where small energetic barriers were found (below $1k_B T$). Thermodynamic stability of this transition is marginal compared with the others, probably leading to biotin-painted cases where two equally leveled minima are now seen. At $T = T_{f2}$, the full

enzyme native state is destabilized, presenting folding and unfolding ensembles of the FKBD domain equally populated [as seen on the F profiles of Figs. 5(f) and 5(g)]. Figures 5(f) and 5(g) present energetic F barriers between 6 and $7.5k_B T$ for the labeled proteins. At $T = T_{f2}$, the native FKBP3 minima at $Q \approx 0.90$ are completely destabilized assuming high energies around $60k_B T$.

Biotinylation of FKBP3 resulted in increasing transition temperatures, as shown in Fig. 4. Free-energy profiles were altered in the vicinity of the transition states (Fig. 5) and their minima (Fig. S10), i.e., thermodynamics changed upon the addition of biotins to the FKBP3 enzyme. However, structural dynamics of apo and biotinylated FKBP3 showed small variations at their native state. Figure S6 shows the root-mean-square fluctuation (RMSF) for the FKBP3 models at temperatures slightly below their two transition temperatures T_{f1} (left panels) and T_{f2} (right panels). The IDR domain that connects the N-term and C-term domains displayed enhanced dynamics when compared with the two other domains. Biotinylation did not abruptly change the protein dynamic fluctuations as it is shown by the Δ RMSF (Figs. S6E and S6F) and their observed distributions (Fig. S6G and S6H). RMSF variations were centered around zero with an amplitude of up to 2–3 Å, representing atomic fluctuations of the order of experimental structure resolution. Thus, we proceeded to evaluate the biotin painting effects on the enzyme dynamics along the entire folding event.

The computed free-energy barriers required to fold the C-term domain are higher when compared with the N-term folding dynamics. This is a consequence of the density of states characterized for this multi-domain enzyme. FKBP3 folds the FKBD domain by crossing a transition state that has a tenuous and narrow density of states, probably one of the reasons to have high F barrier to fold this domain. As folding of the full enzyme proceeds, the N-term domain goes through a transition state with more connected states although it is still a bottleneck. The F barrier to fold the N-term domain is lower than for FKBD; however, $5k_B T$ is considered a moderate-to-high barrier.⁷¹ Predictions by this simplified model might sound reasonable, once the retrieved high energy barriers at transition states could be related to this long and multi-domain enzyme. Folding is related to function; thus, high energetic free-energy barriers for folding of the drug binding FKBD might retain its catalytic role inside the cell nucleus or cytoplasm.

It is also interesting to notice that the positions in Q of the F barriers and minima are conserved while increasing biotin painting. This does not happen for the distance-dependent F profiles, as shown in Figs. S10 and S11. F as a function of R_g (in Fig. S10) and along RMSD (in Fig. S11) resulted in lower barriers if compared to $F(Q)$ as well as shifted barrier positions upon biotin addition. The protein energy landscape is multidimensional, and the protein might experience different mechanisms along projections over different reaction coordinates.^{127,128}

F. Biotin painting alters the folding routes of the C-terminal drug binding domain FKBD

SBM all-atom simulations of the entire FKBP3 enzyme showed slow dynamics, particularly for the C-terminal domain. The N-terminal domain has folding/unfolding transitions on the ns timescale, while the C-terminal FKBD is on the μ s regime. The

two globular domains transitioned in different timescales and T_f s. The previous Secs. II D and II E successfully characterized the enzyme thermodynamics. However, the kinetic analysis could be improved with more sampled transitions, specially for the FKBD folding events. To address this issue, the C-terminal (FKBD) domain containing 118 residues was extracted from the full FKBP3 enzyme. FKBD and its respective biotin-painted constructions were subjected to all-atom SBM simulations with a batch of wide range of temperature runs for each case. This choice of simulating the domain individually decreased the simulation time and hard disk usage, allowing to find more transitions without using exogenous forces to enhance sampling.

Thermodynamic analysis was performed over the temperature runs, as shown in Fig. 6. Biotinylation of the FKBD domain was created preserving the same order and position used in the full FKBP3 biotin-painted enzyme of the previous Secs. II C, II D, and II E. The FK506/rapamycin domain (FKBD) had up to 13 biotinylated lysine residues [Fig. 6(a)]. Numbering of biotin painting in Fig. 6 agrees with that adopted for the full FKBP3, i.e., the domain has up to 13 biotins and its numbering was kept from 11 to 23. In Fig. 6(b), the FKBD specific heat curves as a function of the

temperature [$C_v(T)$] showed increasing folding temperature (T_f) as the biotins were sequentially attached.

The free-energy profiles of FKBD simulations [$F(Q)$, $F(R_g)$, and $F(RMSD)$ in Fig. 6] also resulted in two-state folders as seen in the full FKBP3 dynamics after identification of the domain folding at temperature T_{f2} . In Figure 6(d), F decreased $1-2k_B T$ on average as the biotins were added to FKBD. Energetic barriers are about $2k_B T$ higher for $F(Q)$ and around $6-8k_B T$ higher for $F(R_g)$ and in $F(RMSD)$ if compared with analysis of the full FKBP3 at T_{f2} . This difference for the thermodynamic F barriers could be a result of the enhanced sampling of FKBD when simulated apart from the full FKBP3.

The thermodynamic quantities for FKBD were obtained by computing the microcanonical density of states using the WHAM algorithm, the same procedure of previous Secs. II B, II D, and II E. Figure 7(a) presents the density of states (n) in the logarithm scale projected onto the two coordinates: the number of native contacts and the total energy ($\log[n(Q, E)]$) for FKBD. In Fig. S9, n is presented for all biotin painting constructions for the sake of completion. Figure 7(a) shows the funneled energy landscape surface toward the native state at higher Q and lower E . The transition state

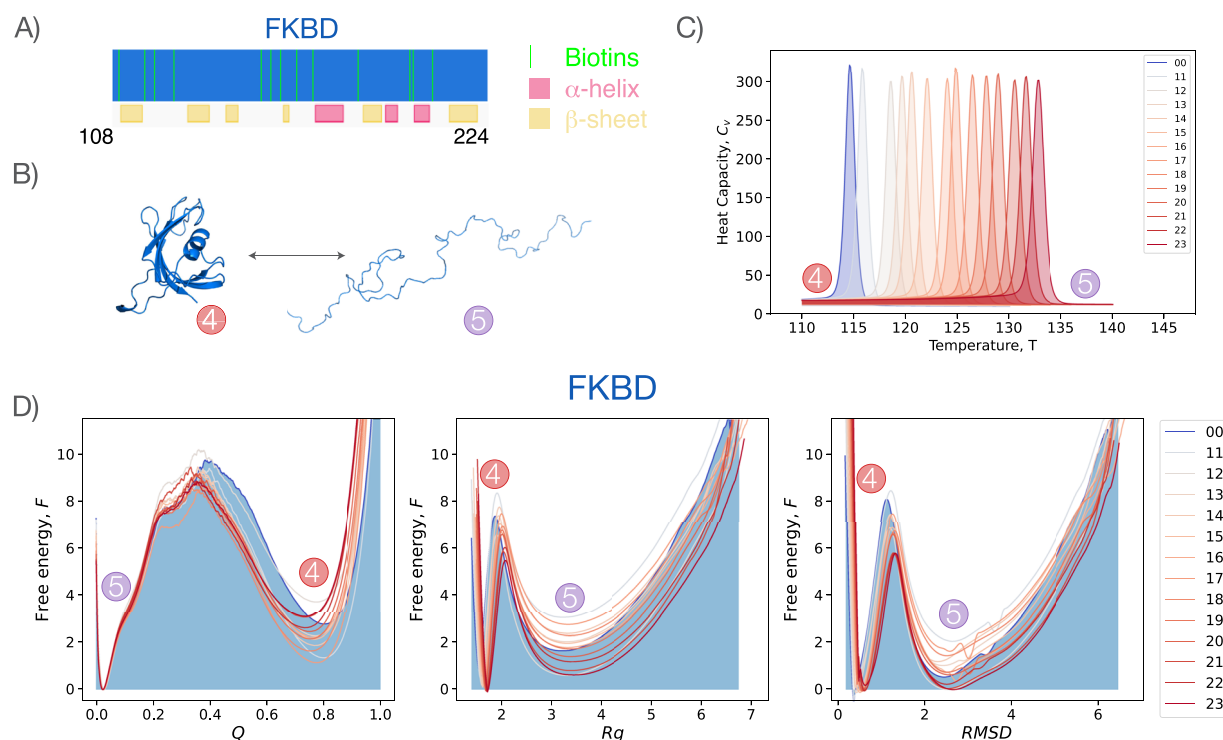


FIG. 6. Thermodynamics of the C-term FKBD domain simulated separately. (a) Protein secondary structural content of the FK506/rapamycin binding domain (FKBD) in the one-dimensional representation. The biotinylated lysines are marked on the protein sequence.⁵⁴ (b) FKBD folding schematics. (c) Specific heat as a function of temperature [$C_v(T)$] for the simulated FKBD with apo (00) and holo (11–23) FKBD curves shown. The numbering of FKBD runs is consistent with the FKBP3 simulations. (d) Free energy (F) profiles at the transition temperature (T_f). The monitored reaction coordinates were the normalized number of native contacts (Q , left panel), radius of gyration (R_g , middle panel), and root-mean-square deviation ($RMSD$, right panel). Thermodynamic profiles are shown for the C-term domain analyzed without biotin (00) and with an increasing number of biotinylated sites. Each construction retrieved an increased T_f following the growing number of biotins attached. Energy is in units of $k_B T_f$ with T_f being the respective system's folding temperature in GROMACS reduced units, and distances (R_g and $RMSD$) are given in nm.

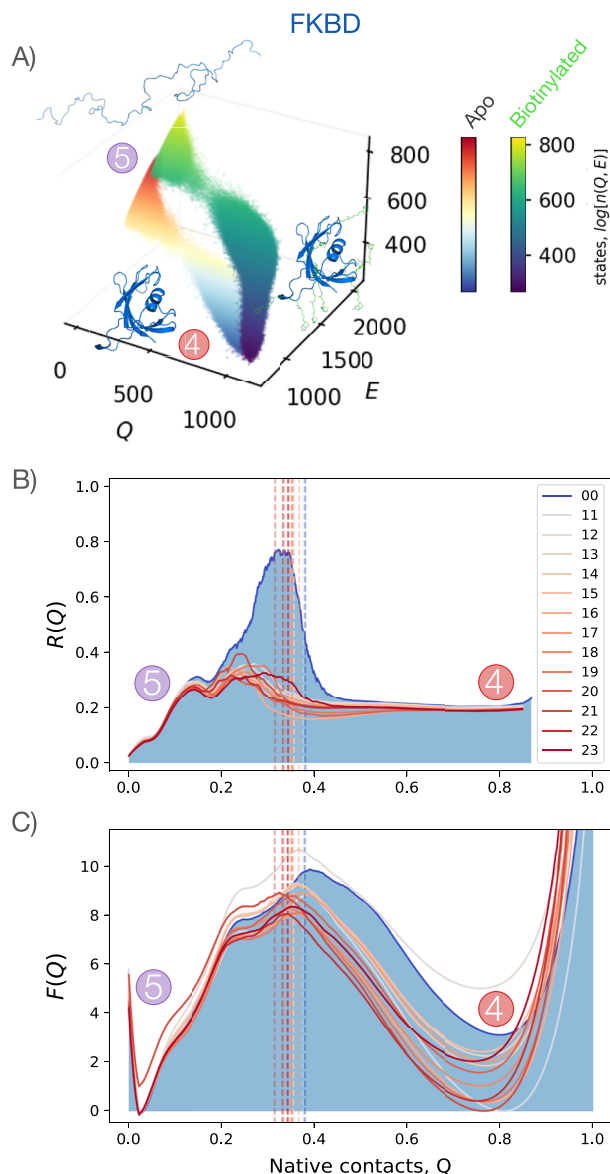


FIG. 7. Kinetics and thermodynamics of FKBD simulated separately. (a) Folding energy landscape of the apo (without biotin) and biotinylated (with 13 biotins) FKBD portrayed by the density of states, $\log[n(Q, E)]$. Q is the number of native contacts, and E is the total energy. (b) Folding routes (R) and (c) free energy (F) curves as a function of Q near the folding temperature (T_f) for each FKBD construction. Vertical dashed lines represent the maximum energetic barrier at the transition state (TS) of each biotin-painted protein. Biotinylation altered the energy landscape, the folding route, and the free-energy profile of the apo FKBD. Energy is in units of $k_B T_f$ with T_f being the respective system's folding temperature in GROMACS reduced units.

is positioned in n at the intermediary values of Q and E , where the number of accessible states is reduced for the C-term domain. The surface is in resemblance with the funneled energy landscape theory for folding of globular two-state proteins.^{67,71} The full FKBP3

density of states had three major states [Fig. 5(a)], and the density of states found for the FKBD domain [Fig. 7(a)] resembles its top part.

Figure 7(b) shows the folding route quantity as a function of the FKBD protein native contacts [$R(Q)$]. The FKBD free-energy profile as a function of Q [$F(Q)$] is displayed in Fig. 7(c) with the vertical dashed lines that cross F and R representing the maximum of each respective transition state F barrier. $R(Q)$ and $F(Q)$ curves are presented near each respective T_f cases.

For the apo (00) C-term domain in Figure 7(b), R increases until reaching 0.8 right before the transition state (maximum F barrier), which represents a more specific folding pathway for this apo domain. After reaching its maximum R in the early folding reaction, the apo FKBD route quantity decreases while crossing the transition state until it approaches 0.2. In Fig. 7(b), biotinylation of FKBD had the consequence of decreasing this maximum specificity in R to 0.3–0.4. Biotin painting of FKBD rendered much less specific folding pathways to this domain. This steep change in R might be related to the sudden change in the n surface of the biotin-painted FKBD in Fig. 7(a). Folding temperatures increased and energetic barriers decreased on average as biotins were sequentially attached to FKBD. Consequently, the kinetic folding routes were altered by decreasing the specificity of the biotinylated proteins. Folding is frequently related to function; thus, the PTM introduced by biotinylation of one or a few residues could have a cascade effect of tampering with the cell behavior. Nevertheless, this is not an attempt to find a universal trend as only two proteins were analyzed. In addition, other features involved in protein folding mechanisms might not be captured by the simple model employed in this study.

III. CONCLUSIONS

Biotin-labeled proteins are key components of proximity-labeling experiments.¹¹ Such technologies employ biotin as a high affinity small-molecular tag that is easily detected by specific binding of bait proteins.¹¹ Proximity-labeling is a high-throughput assay that directly detects thousands of biotin sites in hundreds of proteins, being widely applied to uncover *in vivo* protein–protein interaction networks in almost intact living cells.^{21–24} In addition, biotin proximity tagging experiments can be repurposed to investigate conformational proteomics and protein plasticity inside the cell of large-scale datasets.⁵⁴ A higher number of biotin labels are found in proteins with extensive local structural disordered regions. In addition, the lower the number of labeled sites, the more folded the protein native conformation.⁵⁴ Despite the fact that protein biotinylation can give useful information on protein–protein interaction and on the protein structure content, the small-molecules attached on the protein surface might influence some of its biophysical features.

This work studied the thermodynamic and kinetic properties of biotin painting of two proteins—CoA with one biotin attached (BCCP) and FKBP3 with up to 23 biotins added—by analyzing all-atom structure-based model simulations. This is a simplified computational model with its foundation on the energy landscape theory for proteins.¹¹² Modeling of biotin-painted lysine residues of the proteins was performed considering all non-hydrogen atoms in simulations that are many microseconds long. The results showed that thermodynamics and kinetics were altered for both studied

proteins due to biotinylation. The thermodynamic stability temperature of the labeled proteins, reflected by T_f , decreased for CoA and increased for FKBP3. In addition, the thermodynamic free-energy barrier at the transition state for folding changed for the biotinylated proteins: it was increased for CoA and decreased for FKBP3. Kinetics, given by the folding routes of both proteins, were altered by the attached biotins: the folding routes specificity increased for CoA and decreased for the FKBP3 C-term domain (FKBD) upon biotinylation. CoA had the biotinylated side chain pointed to its thumb domain, which conferred a tighter native state, opposing to FKBP3 that had loose biotin-painted lysine residues placed at the protein surface, and this could be one of the reasons for these opposing thermodynamic and kinetic trends found for the two simulated proteins.

A direct cause of the altered folding routes of both biotin-painted proteins was related to the folding energy landscape that changed after the proteins were biotinylated. The theory states that the energy landscape contains all the protein features, including kinetics and thermodynamics of the biomolecules.^{67,71} The multidimensional energy landscape was obtained by projecting the density of states over two dimensions (native contacts and total energy) of each constructed protein model. The increased folding route specificity of the holo CoA resulted from the narrowed transition state found on its density of states. The opposing effect of decreasing the folding route specificity for the biotin-painted FKBDs was related to the projected density of states that changed its shape and width after biotinylation.

Regarding the model, the all-atom SBM employed on the protein simulations could have potential limitations to the acceptance of the conclusions of this investigation. The coarse-grained protein simulations could limit our understanding of the obtained results due to its accuracy with relation to the real world protein experiments. One fact could be attributed to the effective time scale of the simulations. The structure-based model of this class allows for simulations with longer effective time scales while using reduced units. For instance, the mass of each bead is set to be one, implying an atom mass scale of ~ 12 amu, and the length scale is kept at 1 nm. As discussed elsewhere,¹²⁹ a comparison between the all-atom SBM and experimental measurements of base-stacking interactions suggests a modeled energy scale of ~ 1 kcal/mol, indicating a reduced time unit of ~ 3 ps.^{129,130} Such assumptions cannot provide a precise real time scale; yet, they allow for a reasonable approximation for the theoretical estimates. A more precise correction could be performed, for instance, with the comparison between mean-first passage times of a specific model and experiments if one have those in hand. However, this approach can mask any potential disagreement between experimental and theoretical barrier heights. In addition, the correction factor would depend on other parameters, such as the drag coefficient chosen for the Langevin dynamic simulations. Structure-based models are not expected to produce quantitatively accurate results as they use arbitrary energy and temperature units that are not that straightforward to convert to physical units. It has been demonstrated, however, that often they can reproduce the qualitative features of protein folding quite accurately.

In summary, simulations of the structure-based model in the all-atom representation are characterized the thermodynamics and kinetics of the apo BCCP and FKBP3 enzymes. The two proteins had their estimated thermodynamic folding stability changed by

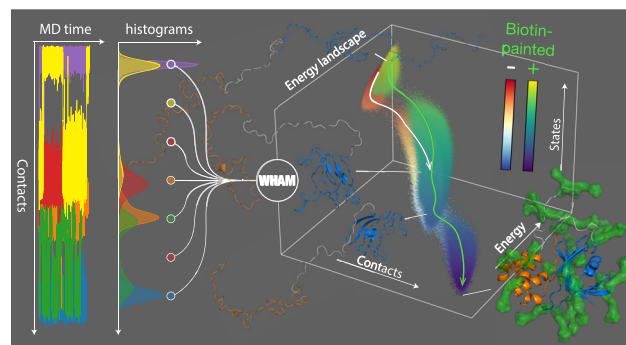


FIG. 8. Scheme of the thermodynamic and kinetic characterization of apo and biotinylated FKBP3 enzyme. The protein energy landscape (on the right) is constructed by converging all trajectory histograms (on the left) with the WHAM algorithm.¹¹⁰ WHAM computes the microcanonical density of states, which is the base of all the thermodynamic quantities. Folding kinetics down the energy funnel is illustrated by the continuous lines over the surfaces and quantified by the route analysis $R(Q)$.

altering their energy landscape upon biotinylation as schemed in Fig. 8. The two biotin-painted proteins had their kinetic folding routes down the energy funnel altered in comparison with their respective apo versions although biotinylation did not modify the proteins' coordinate fluctuations at their native/folded basin. Thus, biotinylation is another post-translational modification that might reshape the protein folding mechanism with a direct impact into the protein thermodynamic and kinetic attributes. As a consequence, repurposing of large-scale datasets of biotinylation-based proximity experiments aiming to study the conformation and plasticity of globular proteins should take into account the potential change in thermodynamic and folding mechanisms by attached biotins. In addition, intrinsic disordered proteins are favored by biotinylation; this class of proteins was not contemplated by the computational model of this work. IDPs have the increasing attention of many scientific studies, and it would be significant to investigate the impact of biotinylation to the IDP dynamics in the future theoretical essays.

IV. METHODS

A. Computational model

Protein simulations were made using all heavy (non-hydrogen) atoms by employing the structure-based model (SBM),¹⁰⁶ also known as Go-model.^{89,131} SBMs have stabilized energetic interactions based on knowledge of an experimentally obtained conformation.^{106,112} SMOG2 software (version 2.3-beta) was used to generate the force field by using standard options for the all-atom model.⁸⁰ The simulation protocol was based on the previous work⁸³ using GROMACS suite 5.1.4.¹³² The biotin topology was added to the standard topology file, observing the following protocol: biotins were attached to the respective lysine NH_3^+ side-chain group of the proteins using the UCSF Chimera¹³³ software. Biotinylated proteins had their energy and topology minimized and refined by the Rosetta software^{134,135} (rosettacommons.org).

The SMOG2 class of structure-based models has a simple potential energy, given by¹⁰⁶

$$V = \sum_{\text{bonds}} \frac{\epsilon_r}{2} (r_{ij} - r_{ij,0})^2 + \sum_{\text{angles}} \frac{\epsilon_\theta}{2} (\theta_{ijk} - \theta_{ijk,0})^2 + \sum_{\text{improper}} \frac{\epsilon_{\chi_{\text{imp}}}}{2} (\chi_{ijkl} - \chi_{ijkl,0})^2 + \sum_{\text{planar}} \frac{\epsilon_{\chi_{\text{planar}}}}{2} F_P(\varphi_{ijkl}) + \sum_{\text{backbone}} \epsilon_{bb} F_D(\phi_{ijkl} - \phi_{ijkl,0}) + \sum_{\text{sidechains}} \epsilon_{sc} F_D(\phi_{ijkl} - \phi_{ijkl,0}) + \sum_{\text{contacts}} \epsilon_C \times \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \sum_{\text{non-contacts}} \epsilon_{nc} \left(\frac{\sigma_{nc}}{r_{ij}} \right)^{12}, \quad (1)$$

where $F_D(\phi) = [1 - \cos(\phi)] + \frac{1}{2}[1 - \cos(3\phi)]$ and $F_P(\varphi_i) = [1 - \cos(2\varphi_i)]$. The parameters related to the minima of each term were obtained from each native structure ($r_{ij,0}$, $\theta_{ijk,0}$, $\chi_{ijkl,0}$, and $\phi_{ijkl,0}$). The planar ring dihedrals were maintained by cosine potentials of periodicity two. Dihedrals were grouped by their middle bond, and each group was given a summed weight of ϵ_{bb} (backbone dihedrals) or ϵ_{sc} (side chain dihedrals). The ratio $R_{bb/sc} = \frac{\epsilon_{bb}}{\epsilon_{sc}}$ was set to one (nucleic acids) or two (proteins). ϵ_{bb} was defined to be equal for protein and nucleic acids. Contact and dihedral strengths were scaled such that

$$R_{C/D} = \frac{\sum \epsilon_C}{\sum \epsilon_{bb} + \sum \epsilon_{sc}} = 2$$

and

$$\sum \epsilon_C + \sum \epsilon_{bb} + \sum \epsilon_{sc} = N\epsilon,$$

where N is the number of atoms in the system and ϵ is the reduced energy unit, which is equal to $2k_B T$. Contact pairs were defined using the shadow contact map algorithm with default values.¹³⁶ σ_{ij} was set to be the distance pair found in the native structure.

B. Analysis details

Thermodynamic analysis was carried out for each protein system simulated over a batch of different temperature runs from lower temperatures (fully folded/native states) to higher temperatures (unfolded/denatured states), including the transition folding temperature between the extreme states (T_f). Temperature is shown in units of the GROMACS reduced units. The analyses were accomplished by the Weighted Histogram Analysis Method (WHAM)¹¹⁰ with a Python script, the PyWham package.¹¹¹ WHAM computes the microcanonical density of states that generate the free-energy profiles.⁷¹ T_f was defined as the temperature related to the specific heat peak between folding/unfolding states for each protein. The number/fraction of native contacts within a given structure (Q) was used to evaluate nativeness as the reaction coordinate, which describes how similar is a given structure with respect to the initial native structure.⁹⁵ Route $[R(Q)]$ profiles were calculated by an in-house Perl script based on our previous publication.⁹⁴ Root-mean-square fluctuation (RMSF), radius of gyration (R_g), and root-mean-square deviation (RMSD) were computed with the analysis package contained in the GROMACS suite. The simulation time

ranged from 1 to 25 μ s with integration steps of 2 fs and snapshots of 4 ps. The all-atom SBM simulations were executed using periodic boundary conditions with no constraint or enhanced sampling techniques. Enough sampling for the statistical calculations of folding/unfolding states was achieved at T_f and temperatures in its vicinity.

Secondary structure elements and sequence alignments were identified by the ENDscript 2 web server.¹³⁷ Post-translation modifications of FKBP3 were drawn by using the Protter web server¹³⁸ (wlab.ethz.ch/protter). Protein structures were visualized, and their imaging was performed with PyMOL (pymol.org) Visual Molecular Dynamics (VMD),¹³⁹ UCSF Chimera,¹³³ and Protein Imaging¹⁴⁰ (3dproteinimaging.com) softwares. Two and three dimensional curves were plotted with Grace (plasma-gate.weizmann.ac.il/Grace), Gnuplot (gnuplot.info), and Matplotlib¹⁴¹ packages.

SUPPLEMENTARY MATERIAL

The [supplementary material](#) contains extra figures with the sequence alignments of BCCPs, the trajectory traces and free-energy profiles, root-mean-square fluctuations (RMSF), the density of states, and specific heat of the simulated proteins.

ACKNOWLEDGMENTS

The financial support for R.J.d.O. was provided, in part, by the Brazilian agencies Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG, Grant No. APQ-02303-21) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Grant Nos. 438316/2018-5 and 312328/2019-2). J.N.O. was supported by the National Science Foundation (NSF), Grant Nos. CHE-1614101 and PHY-2019745, and The Welch Foundation (Grant No. C-1792). J.N.O. is a Cancer Prevention and Research Institute of Texas (CPRIT) scholar in cancer research. A.B.O.J. acknowledges the Robert A. Welch Postdoctoral Fellow program. Computational resources were provided by the Núcleo de Computação Científica da Universidade do Estado de São Paulo (NCC/GridUnesp) and Centro Nacional de Processamento de Alto Desempenho em São Paulo (CENAPAD-SP).

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

- M. A. Boas, *Biochem. J.* **21**, 712 (1927).
- H. M. Said, in *Water Soluble Vitamins*, Subcellular Biochemistry, edited by O. Stanger (Springer, Dordrecht, The Netherlands, 2012), Vol. 56, pp. 1–19.
- K. Dakshinamurti and J. Chauhan, *Vitam. Horm.* **45**, 337–384 (1989).

- ⁴W. T. Penberthy, M. Sadri, and J. Zemleni, *Present Knowledge in Nutrition* (Elsevier, 2020), pp. 289–303.
- ⁵G. Camporeale, E. E. Shubert, G. Sarath, R. Cerny, and J. Zemleni, *Eur. J. Biochem.* **271**, 2257 (2004).
- ⁶K. Kobza, G. Camporeale, B. Rueckert, A. Kueh, J. B. Griffin, G. Sarath, and J. Zemleni, *FEBS J.* **272**, 4249 (2005).
- ⁷Y. C. Chew, G. Camporeale, N. Kothapalli, G. Sarath, and J. Zemleni, *J. Nutr. Biochem.* **17**, 225 (2006).
- ⁸V. Pestinger, S. S. K. Wijeratne, R. Rodriguez-Melendez, and J. Zemleni, *J. Nutr. Biochem.* **22**, 328 (2011).
- ⁹G. A. Khoury, R. C. Baliban, and C. A. Floudas, *Sci. Rep.* **1**, 90 (2011).
- ¹⁰I. Bludau and R. Aebersold, *Nat. Rev. Mol. Cell Biol.* **21**, 327 (2020).
- ¹¹J. J. Falke and J. A. Corbin, *Encyclopedia of Biological Chemistry* (Elsevier, 2013), pp. 61–65.
- ¹²N. Green, *Biochem. J.* **89**, 599 (1963).
- ¹³U. Piran and W. J. Riordan, *J. Immunol. Methods* **133**, 141 (1990).
- ¹⁴M. Fernández-Suárez, T. S. Chen, and A. Y. Ting, *J. Am. Chem. Soc.* **130**, 9251 (2008).
- ¹⁵A. Kulyassov, M. Shoaib, A. Pichugin, P. Kannouche, E. Ramanculov, M. Lipinski, and V. Ogryzko, *J. Proteome Res.* **10**, 4416 (2011).
- ¹⁶P. Li, J. Li, L. Wang, and L.-J. Di, *Proteomics* **17**, 1700002 (2017).
- ¹⁷D. Z. Bar, K. Atkash, U. Tavares, M. R. Erdos, Y. Gruenbaum, and F. S. Collins, *Nat. Methods* **15**, 127 (2018).
- ¹⁸H. Ummethum and S. Hamperl, *Front. Genet.* **11**, 450 (2020).
- ¹⁹P. Samavarchi-Tehrani, R. Samson, and A.-C. Gingras, *Mol. Cell. Proteomics* **19**, 757 (2020).
- ²⁰J. Zheng, X. Chen, Y. Yang, C. S. H. Tan, and R. Tian, *Anal. Chem.* **93**, 598 (2021).
- ²¹C. M. Tron, I. W. McNae, M. Nutley, D. J. Clarke, A. Cooper, M. D. Walkinshaw, R. L. Baxter, and D. J. Campopiano, *J. Mol. Biol.* **387**, 129 (2009).
- ²²J. S. Rees, X.-W. Li, S. Perrett, K. S. Lilley, and A. P. Jackson, *Mol. Cell. Proteomics* **14**, 2848 (2015).
- ²³D. I. Kim, S. C. Jensen, K. A. Noble, B. Kc, K. H. Roux, K. Motamedchaboki, and K. J. Roux, *Mol. Biol. Cell* **27**, 1188 (2016).
- ²⁴S. Han, N. D. Udeshi, T. J. Deerinck, T. Svinkina, M. H. Ellisman, S. A. Carr, and A. Y. Ting, *Cell Chem. Biol.* **24**, 404 (2017).
- ²⁵L. M. Schiapparelli, D. B. McClatchy, H.-H. Liu, P. Sharma, J. R. Yates, and H. T. Cline, *J. Proteome Res.* **13**, 3966 (2014).
- ²⁶S.-Y. Lee, H. Lee, H.-K. Lee, S.-W. Lee, S. C. Ha, T. Kwon, J. K. Seo, C. Lee, and H.-W. Rhee, *ACS Cent. Sci.* **2**, 506 (2016).
- ²⁷D. I. Kim, J. A. Cutler, C. H. Na, S. Reckel, S. Renuse, A. K. Madugundu, R. Tahir, H. L. Goldschmidt, K. L. Reddy, R. L. Haganir, X. Wu, N. E. Zachara, O. Hantschel, and A. Pandey, *J. Proteome Res.* **17**, 759 (2018).
- ²⁸N. D. Udeshi, K. Pedram, T. Svinkina, S. Fereshetian, S. A. Myers, O. Aygun, K. Krug, K. Clauser, D. Ryan, T. Ast, V. K. Mootha, A. Y. Ting, and S. A. Carr, *Nat. Methods* **14**, 1167 (2017).
- ²⁹Y. Xu, X. Fan, and Y. Hu, *Cell Biosci.* **11**, 27 (2021).
- ³⁰N. Kotani, J. Gu, T. Isaji, K. Uda, N. Taniguchi, and K. Honke, *Proc. Natl. Acad. Sci. U. S. A.* **105**, 7405 (2008).
- ³¹K. Honke and N. Kotani, *Sensors* **12**, 16037 (2012).
- ³²J. D. Martell, T. J. Deerinck, Y. Sancak, T. L. Poulos, V. K. Mootha, G. E. Sosinsky, M. H. Ellisman, and A. Y. Ting, *Nat. Biotechnol.* **30**, 1143 (2012).
- ³³S. S. Lam, J. D. Martell, K. J. Kamer, T. J. Deerinck, M. H. Ellisman, V. K. Mootha, and A. Y. Ting, *Nat. Methods* **12**, 51 (2015).
- ³⁴C. James, M. Müller, M. W. Goldberg, C. Lenz, H. Urlaub, and R. H. Kehlenbach, *J. Biol. Chem.* **294**, 16241 (2019).
- ³⁵E. Choi-Rhee, H. Schulman, and J. E. Cronan, *Protein Sci.* **13**, 3043 (2008).
- ³⁶K. J. Roux, D. I. Kim, M. Raida, and B. Burke, *J. Cell Biol.* **196**, 801 (2012).
- ³⁷K. J. Roux, D. I. Kim, and B. Burke, *Curr. Protoc. Protein Sci.* **74**, 19.23.1 (2013).
- ³⁸M. Ramanathan, K. Majzoub, D. S. Rao, P. H. Neela, B. J. Zarnegar, S. Mondal, J. G. Roth, H. Gai, J. R. Kovalski, Z. Siprashvili, T. D. Palmer, J. E. Cavette, and P. A. Khavari, *Nat. Methods* **15**, 207 (2018).
- ³⁹T. C. Branon, J. A. Bosch, A. D. Sanchez, N. D. Udeshi, T. Svinkina, S. A. Carr, J. L. Feldman, N. Perrimon, and A. Y. Ting, *Nat. Biotechnol.* **36**, 880 (2018).
- ⁴⁰Y. Zhang, Y. Li, X. Yang, Z. Wen, U. Nagalakshmi, and S. P. Dinesh-Kumar, *J. Visualized Exp.* **159**, 60728 (2020).
- ⁴¹K. Kido, S. Yamanaka, S. Nakano, K. Motani, S. Shinohara, A. Nozawa, H. Kosako, S. Ito, and T. Sawasaki, *eLife* **9**, e54983 (2020).
- ⁴²I. M. Schopp, C. C. Amaya Ramirez, J. Debeljak, E. Kreibich, M. Skribbe, K. Wild, and J. Béthune, *Nat. Commun.* **8**, 15690 (2017).
- ⁴³C. Kwak, S. Shin, J.-S. Park, M. Jung, T. T. M. Nhung, M.-G. Kang, C. Lee, T.-H. Kwon, S. K. Park, J. Y. Mun, J.-S. Kim, and H.-W. Rhee, *Proc. Natl. Acad. Sci. U. S. A.* **117**, 12109 (2020).
- ⁴⁴K. F. Cho, T. C. Branon, S. Rajeev, T. Svinkina, N. D. Udeshi, T. Thoudam, C. Kwak, H.-W. Rhee, I.-K. Lee, S. A. Carr, and A. Y. Ting, *Proc. Natl. Acad. Sci. U. S. A.* **117**, 12143 (2020).
- ⁴⁵F. Rudolph, C. Fink, J. Hüttemeister, M. Kirchner, M. H. Radke, J. Lopez Carballo, E. Wagner, T. Kohl, S. E. Lehnart, P. Mertins, and M. Gotthardt, *Nat. Commun.* **11**, 3133 (2020).
- ⁴⁶B. Bagautdinov, Y. Matsuura, S. Bagautdinova, and N. Kunishima, *J. Biol. Chem.* **283**, 14739 (2008).
- ⁴⁷E. L. Roberts, N. Shu, M. J. Howard, R. W. Broadhurst, A. Chapman-Smith, J. C. Wallace, T. Morris, J. E. Cronan, and R. N. Perham, *Biochemistry* **38**, 5045 (1999).
- ⁴⁸V. G. Contessoto, P. H. B. Ferreira, J. Chahine, V. B. P. Leite, and R. J. Oliveira, *J. Phys. Chem. B* **125**, 11673 (2021).
- ⁴⁹D. F. Coelho, M. V. F. Ferraz, E. T. A. Marques, R. D. Lins, and I. F. T. Viana, *J. Mol. Graphics Modell.* **93**, 107442 (2019).
- ⁵⁰J. X. Huang, G. Lee, K. E. Cavanaugh, J. W. Chang, M. L. Gardel, and R. E. Moellering, *Nat. Methods* **16**, 894 (2019).
- ⁵¹Y. Wang, M. Sarkar, A. E. Smith, A. S. Krois, and G. J. Pielak, *J. Am. Chem. Soc.* **134**, 16614 (2012).
- ⁵²M. Tabaka, T. Kalwarczyk, and R. Holyst, *Nucleic Acids Res.* **42**, 727 (2014).
- ⁵³C. M. Davis, M. Gruebele, and S. Sukenik, *Curr. Opin. Struct. Biol.* **48**, 23 (2018).
- ⁵⁴D.-P. Minde, M. Ramakrishna, and K. S. Lilley, *Commun. Biol.* **3**, 38 (2020).
- ⁵⁵J. Song, C. Midson, E. Blachly-Dyson, M. Forte, and M. Colombini, *J. Biol. Chem.* **273**, 24406 (1998).
- ⁵⁶J. Groll, E. V. Amigoulova, T. Ameringer, C. D. Heyes, C. Röcker, G. U. Nienhaus, and M. Möller, *J. Am. Chem. Soc.* **126**, 4234 (2004).
- ⁵⁷K. E. Taylor and C. W. van den Berg, *Immunology* **120**, 404 (2007).
- ⁵⁸P. Curnow and P. J. Booth, *J. Mol. Biol.* **403**, 630 (2010).
- ⁵⁹Y.-C. Chang and J. U. Bowie, *Proc. Natl. Acad. Sci. U. S. A.* **111**, 219 (2014).
- ⁶⁰C.-X. Wang, S.-L. Ai, B. Wu, S.-W. Huang, and Z. Liu, *J. Mater. Chem. B* **8**, 3557 (2020).
- ⁶¹Y. Gavrilov, D. Shental-Bechor, H. M. Greenblatt, and Y. Levy, *J. Phys. Chem. Lett.* **6**, 3572 (2015).
- ⁶²Y. Gavrilov, T. Hagai, and Y. Levy, *Protein Sci.* **24**, 1580 (2015).
- ⁶³V. U. Antunes, E. E. Llontop, F. N. d. C. Vasconcelos, Y. López de los Santos, R. J. Oliveira, N. Lincopan, C. S. Farah, N. Doucet, A. Mittermaier, and D. C. Favaro, *Biochemistry* **58**, 3604 (2019).
- ⁶⁴A. A. Gurtovenko, M. Javanainen, F. Lolicato, and I. Vattulainen, *J. Phys. Chem. Lett.* **10**, 1005 (2019).
- ⁶⁵R. Aguayo-Ortiz and L. Dominguez, *J. Chem. Inf. Model.* **60**, 777 (2020).
- ⁶⁶V. de Godoi Contessoto, A. B. de Oliveira Junior, J. Chahine, R. J. de Oliveira, and V. B. Pereira Leite, *Rev. Bras. Ensino Fis.* **40**, e4307 (2018).
- ⁶⁷P. G. Wolynes, J. N. Onuchic, and D. Thirumalai, *Science* **267**, 1619 (1995).
- ⁶⁸R. Baldwin, *J. Biomol. NMR* **5**, 103 (1995).
- ⁶⁹K. A. Dill, S. Bromberg, K. Yue, H. S. Chan, K. M. Ftebig, D. P. Yee, and P. D. Thomas, *Protein Sci.* **4**, 561 (1995).
- ⁷⁰J. N. Onuchic, H. Nymeyer, A. E. García, J. Chahine, and N. D. Socci, *Adv. Protein Chem.* **53**, 87 (2000).
- ⁷¹J. Wang, R. J. Oliveira, X. Chu, P. C. Whitford, J. Chahine, W. Han, E. Wang, J. N. Onuchic, and V. B. P. Leite, *Proc. Natl. Acad. Sci. U. S. A.* **109**, 15763 (2012).
- ⁷²R. J. Oliveira, P. C. Whitford, J. Chahine, V. B. P. Leite, and J. Wang, *Methods* **52**, 91 (2010).
- ⁷³R. J. Oliveira, P. C. Whitford, J. Chahine, J. Wang, J. N. Onuchic, and V. B. P. Leite, *Biophys. J.* **99**, 600 (2010).
- ⁷⁴R. J. de Oliveira, *J. Chem. Phys.* **149**, 234107 (2018).
- ⁷⁵R. Zwanzig, *Proc. Natl. Acad. Sci. U. S. A.* **85**, 2029 (1988).

- ⁷⁶M. Jacob and F. X. Schmid, *Biochemistry* **38**, 13773 (1999).
- ⁷⁷C. E. Colosqui, *J. Chem. Phys.* **150**, 181102 (2019).
- ⁷⁸W. Xu, Z. Lai, R. J. Oliveira, V. B. P. Leite, and J. Wang, *J. Phys. Chem. B* **116**, 5152 (2012).
- ⁷⁹A. B. Oliveira, V. G. Contessoto, A. Hassan, S. Byju, A. Wang, Y. Wang, E. Dodero-Rojas, U. Mohanty, J. K. Noel, J. N. Onuchic, and P. C. Whitford, *Protein Sci.* **31**, 158 (2022).
- ⁸⁰J. K. Noel, M. Levi, M. Raghunathan, H. Lammert, R. L. Hayes, J. N. Onuchic, and P. C. Whitford, *PLoS Comput. Biol.* **12**, e1004794 (2016).
- ⁸¹J. Hu, T. Chen, M. Wang, H. S. Chan, and Z. Zhang, *Phys. Chem. Chem. Phys.* **19**, 13629 (2017).
- ⁸²C. Nguyen, J. T. Young, G. G. Slade, R. J. Oliveira, and M. E. McCully, *Biophys. J.* **116**, 621 (2019).
- ⁸³F. C. Freitas, P. H. B. Ferreira, D. C. Favaro, and R. J. de Oliveira, *J. Chem. Inf. Model.* **61**, 1226 (2021).
- ⁸⁴D. U. Ferreira, I. E. Sánchez, and G. de Prat Gay, *Proc. Natl. Acad. Sci. U. S. A.* **105**, 10797 (2008).
- ⁸⁵E. Trizac, Y. Levy, and P. G. Wolynes, *Proc. Natl. Acad. Sci. U. S. A.* **107**, 2746 (2010).
- ⁸⁶W.-T. Chu, X. Chu, and J. Wang, *Proc. Natl. Acad. Sci. U. S. A.* **114**, E7959 (2017).
- ⁸⁷R. Sharma, D. De Sancho, and V. Muñoz, *Phys. Chem. Chem. Phys.* **19**, 28512 (2017).
- ⁸⁸D. Bonetti, F. Troilo, M. Brunori, S. Longhi, and S. Gianni, *Biophys. J.* **114**, 1889 (2018).
- ⁸⁹S. Takada, *Biophys. Physicobiol.* **16**, 248 (2019).
- ⁹⁰S. Neelamraju, D. J. Wales, and S. Gosavi, *Curr. Opin. Struct. Biol.* **64**, 145 (2020).
- ⁹¹A. N. Lima, R. J. de Oliveira, A. S. K. Braz, M. G. de Souza Costa, D. Perahia, and L. P. B. Scott, *Eur. Biophys. J.* **47**, 583 (2018).
- ⁹²F. Polotto, E. Drigo Filho, J. Chahine, and R. J. de Oliveira, *Physica A* **493**, 286 (2018).
- ⁹³F. C. Freitas, A. N. Lima, V. d. G. Contessoto, P. C. Whitford, and R. J. d. Oliveira, *J. Chem. Phys.* **151**, 114106 (2019).
- ⁹⁴P. H. B. Ferreira, F. C. Freitas, M. E. McCully, G. G. Slade, and R. J. de Oliveira, *J. Chem. Inf. Model.* **60**, 546 (2020).
- ⁹⁵F. C. Freitas, G. Fuchs, R. J. de Oliveira, and P. C. Whitford, *Biophysica* **1**, 204 (2021).
- ⁹⁶F. K. Athappilly and W. A. Hendrickson, *Structure* **3**, 1407 (1995).
- ⁹⁷A. Prakash, S. Rajan, and H. S. Yoon, *Protein Sci.* **25**, 905 (2016).
- ⁹⁸J. Liang, D. T. Hung, S. L. Schreiber, and J. Clardy, *J. Am. Chem. Soc.* **118**, 1231 (1996).
- ⁹⁹A. Leung, F.-P. Jardim, N. Savic, Y. R. Monneau, R. González-Romero, G. Gudavicius, J. M. Eirin-Lopez, T. Bartke, C. D. Mackereth, J. Ausió, and C. J. Nelson, *Sci. Rep.* **7**, 3795 (2017).
- ¹⁰⁰T. C. Broussard, M. J. Kobe, S. Pakhomova, D. B. Neau, A. E. Price, T. S. Champion, and G. L. Waldrop, *Structure* **21**, 650 (2013).
- ¹⁰¹Y. Liu, M. M. Budelier, K. Stine, and M. St. Maurice, *Nat. Commun.* **9**, 1384 (2018).
- ¹⁰²P. H. Choi, J. Jo, Y.-C. Lin, M.-H. Lin, C.-Y. Chou, L. E. P. Dietrich, and L. Tong, *Nat. Commun.* **7**, 12713 (2016).
- ¹⁰³P. R. Mouro, V. G. Contessoto, J. Chahine, R. J. de Oliveira, and V. B. P. Leite, *Biophys. J.* **111**, 287 (2016).
- ¹⁰⁴V. G. Contessoto, D. T. Lima, R. J. Oliveira, A. T. Bruni, J. Chahine, and V. B. P. Leite, *Proteins: Struct., Funct., Bioinf.* **81**, 1727 (2013).
- ¹⁰⁵C. Clementi, A. E. García, and J. N. Onuchic, *J. Mol. Biol.* **326**, 933 (2003).
- ¹⁰⁶P. C. Whitford, J. K. Noel, S. Gosavi, A. Schug, K. Y. Sanbonmatsu, and J. N. Onuchic, *Proteins: Struct., Funct., Bioinf.* **75**, 430 (2009).
- ¹⁰⁷J. Shimada, E. L. Kussell, and E. I. Shakhnovich, *J. Mol. Biol.* **308**, 79 (2001).
- ¹⁰⁸E. Haglund, J. I. Sulkowska, Z. He, G.-S. Feng, P. A. Jennings, and J. N. Onuchic, *PLoS One* **7**, e45654 (2012).
- ¹⁰⁹L. L. Chavez, J. N. Onuchic, and C. Clementi, *J. Am. Chem. Soc.* **126**, 8426 (2004).
- ¹¹⁰A. M. Ferrenberg and R. H. Swendsen, *Phys. Rev. Lett.* **61**, 2635 (1988).
- ¹¹¹L. Sun, J. K. Noel, J. I. Sulkowska, H. Levine, and J. N. Onuchic, *Biophys. J.* **107**, 2950 (2014).
- ¹¹²C. Clementi, H. Nymeyer, and J. N. Onuchic, *J. Mol. Biol.* **298**, 937 (2000).
- ¹¹³F. O. Tzul, K. L. Schweiker, and G. I. Makhatadze, *Proc. Natl. Acad. Sci. U. S. A.* **112**, E259 (2015).
- ¹¹⁴J. Schönfelder, D. De Sancho, R. Berkovich, R. B. Best, V. Muñoz, and R. Perez-Jimenez, *Commun. Chem.* **1**, 59 (2018).
- ¹¹⁵X. Yao, C. Soden, M. F. Summers, and D. Beckett, *Protein Sci.* **8**, 307 (2008).
- ¹¹⁶S. S. Cho, Y. Levy, and P. G. Wolynes, *Proc. Natl. Acad. Sci. U. S. A.* **103**, 586 (2006).
- ¹¹⁷K. Lindorff-Larsen, S. Piana, R. O. Dror, and D. E. Shaw, *Science* **334**, 517 (2011).
- ¹¹⁸J. O. Z. Luthey-Schulten and P. Wolynes, *Annu. Rev. Phys. Chem.* **48**, 545 (1997).
- ¹¹⁹H. Nymeyer, A. E. García, and J. N. Onuchic, *Proc. Natl. Acad. Sci. U. S. A.* **95**, 5921 (1998).
- ¹²⁰Y. Levy, P. G. Wolynes, and J. N. Onuchic, *Proc. Natl. Acad. Sci. U. S. A.* **101**, 511 (2004).
- ¹²¹L. Lins, A. Thomas, and R. Brasseur, *Protein Sci.* **12**, 1406 (2003).
- ¹²²A. Campen, R. Williams, C. Brown, J. Meng, V. Uversky, and A. Dunker, *Protein Pept. Lett.* **15**, 956 (2008).
- ¹²³J. Bush and G. I. Makhatadze, *Proteins: Struct., Funct., Bioinf.* **79**, 2027 (2011).
- ¹²⁴N. Floquet, M. G. S. Costa, P. R. Batista, P. Renault, P. M. Bisch, F. Raussin, J. Martinez, M. C. Morris, and D. Perahia, *Biophys. J.* **109**, 1179 (2015).
- ¹²⁵R. C. Bernardi, M. C. R. Melo, and K. Schulten, *Biochim. Biophys. Acta, Gen. Subj.* **1850**, 872 (2015).
- ¹²⁶X. Peng, Y. Zhang, H. Chu, and G. Li, *J. Comput. Chem.* **37**, 614 (2016).
- ¹²⁷F. Liu, D. Du, A. A. Fuller, J. E. Davoren, P. Wipf, J. W. Kelly, and M. Gruebele, *Proc. Natl. Acad. Sci. U. S. A.* **105**, 2369 (2008).
- ¹²⁸F. Liu, C. Dumont, Y. Zhu, W. F. DeGrado, F. Gai, and M. Gruebele, *J. Chem. Phys.* **130**, 061101 (2009).
- ¹²⁹H. Yang, P. Bandarkar, R. Horne, V. B. P. Leite, J. Chahine, and P. C. Whitford, *J. Chem. Phys.* **151**, 085102 (2019).
- ¹³⁰M. Kouza, M. S. Li, E. P. O'Brien, C.-K. Hu, and D. Thirumalai, *J. Phys. Chem. A* **110**, 671 (2005).
- ¹³¹H. Taketomi, Y. Ueda, and N. Go, *Int. J. Pept. Protein Res.* **7**, 445 (1975).
- ¹³²D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, *J. Comput. Chem.* **26**, 1701 (2005).
- ¹³³E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin, *J. Comput. Chem.* **25**, 1605 (2004).
- ¹³⁴B. Kuhlman, G. Dantas, G. C. Ireton, G. Varani, B. L. Stoddard, and D. Baker, *Science* **302**, 1364 (2003).
- ¹³⁵A. Leaver-Fay, M. Tyka, S. M. Lewis, O. F. Lange, J. Thompson, R. Jacak, K. W. Kaufman, P. D. Renfrew, C. A. Smith, W. Sheffler, I. W. Davis, S. Cooper, A. Treuille, D. J. Mandell, F. Richter, Y.-E. A. Ban, S. J. Fleishman, J. E. Corn, D. E. Kim, S. Lyskov, M. Berrondo, S. Mentzer, Z. Popović, J. J. Havranek, J. Karanicolas, R. Das, J. Meiler, T. Kortemme, J. J. Gray, B. Kuhlman, D. Baker, and P. Bradley, *Methods Enzymol.* **487**, 545–574 (2011).
- ¹³⁶J. K. Noel, P. C. Whitford, and J. N. Onuchic, *J. Phys. Chem. B* **116**, 8692 (2012).
- ¹³⁷X. Robert and P. Gouet, *Nucleic Acids Res.* **42**, W320 (2014).
- ¹³⁸U. Omasits, C. H. Ahrens, S. Müller, and B. Wollscheid, *Bioinformatics* **30**, 884 (2014).
- ¹³⁹W. Humphrey, A. Dalke, and K. Schulten, *J. Mol. Graphics* **14**, 33 (1996).
- ¹⁴⁰G. Tomasello, I. Armenia, and G. Molla, *Bioinformatics* **36**, 2909 (2020).
- ¹⁴¹J. D. Hunter, *Comput. Sci. Eng.* **9**, 90 (2007).