Simulations of the regulatory ACT domain of human phenylalanine hydroxylase (PAH) unveil its mechanism of phenylalanine binding

Received for publication, July 16, 2018, and in revised form, September 17, 2018 Published, Papers in Press, October 4, 2018, DOI 10.1074/jbc.RA118.004909

Yunhui Ge[‡], Elias Borne[§], Shannon Stewart[§], Michael R. Hansen[§], Emilia C. Arturo[§¶], Eileen K. Jaffe[§], and Vincent A. Voelz[‡1]

From the [‡]Department of Chemistry, Temple University, Philadelphia, Pennsylvania 19122, [§]Fox Chase Cancer Center, Temple University Health System, Philadelphia, Pennsylvania 19111, and [¶]Drexel University College of Medicine, Philadelphia, Pennsylvania 19129

Edited by Norma M. Allewell

Phenylalanine hydroxylase (PAH) regulates phenylalanine (Phe) levels in mammals to prevent neurotoxicity resulting from high Phe concentrations as observed in genetic disorders leading to hyperphenylalaninemia and phenylketonuria. PAH senses elevated Phe concentrations by transient allosteric Phe binding to a protein–protein interface between ACT domains of different subunits in a PAH tetramer. This interface is present in an activated PAH (A-PAH) tetramer and absent in a resting-state PAH (RS-PAH) tetramer. To investigate this allosteric sensing mechanism, here we used the GROMACS molecular dynamics simulation suite on the Folding@home computing platform to perform extensive molecular simulations and Markov state model (MSM) analysis of Phe binding to ACT domain dimers. These simulations strongly implicated a conformational selection mechanism for Phe association with ACT domain dimers and revealed protein motions that act as a gating mechanism for Phe binding. The MSMs also illuminate a highly mobile hairpin loop, consistent with experimental findings also presented here that the PAH variant L72W does not shift the PAH structural equilibrium toward the activated state. Finally, simulations of ACT domain monomers are presented, in which spontaneous transitions between resting-state and activated conformations are observed, also consistent with a mechanism of conformational selection. These mechanistic details provide detailed insight into the regulation of PAH activation and provide testable hypotheses for the development of new allosteric effectors to correct structural and functional defects in PAH.

Phenylalanine hydroxylase (PAH;[2] EC 1.14.16.1) functions in humans to control free phenylalanine (Phe), an essential amino acid that is neurotoxic at elevated levels. Failure to control Phe, most often due to defects in PAH, results in hyperphenylalaninemia or phenylketonuria (PKU), which is the most common inborn error of amino acid metabolism. PAH catalyzes the conversion of Phe to tyrosine at the enzyme active site but also binds Phe at an allosteric site that sits at a subunit–subunit interface of a PAH tetramer. This intersubunit interface, which is present in activated PAH (A-PAH) and absent in resting-state PAH (RS-PAH), lies between ACT subdomains located diagonally across the tetramer (Fig. 1, *a* and *b*). ACT domains, which can serve in ligand sensing, were named for the first three proteins in which they were identified (1). Phe stabilization of A-PAH controls the equilibrium between RS-PAH and A-PAH and thus allows Phe to regulate PAH activity (2). The Phe-stabilized conformational change is coupled to exposure of the enzyme active site (3), thus activating the enzyme.

The crystal structure of a Phe-bound ACT domain dimer from truncated human PAH (Protein Data Bank (PDB) code 5FII) (4) shows key differences from a homology model of the putative ligand-free ACT domain dimer (3). The homology model was constructed using monomeric rat PAH (rPAH; see Fig. S1 for sequence comparison) ACT domain (PDB code 1PHZ) as a template with the dimeric form modeled by threading onto the ACT domain dimer from phosphoglycerate dehydrogenase (PDB code 1PSD). Unliganded monomeric ACT domains have identical conformations in most available crystal structures that represent RS-PAH (PDB codes 1PHZ, 2PHM, 5DEN, and 5FGJ) (5–7). This conformation differs from the Phe-bound crystal structure (PDB code 5FII) by having a helical turn at residues 61–64 (*e.g.* PDB code 5DEN), which, when overlaid on the Phe-bound ACT domain dimer structure, fills the cavity of the allosteric Phe-binding site. The homology

This work was supported by National Institutes of Health Grants 1R01GM123296-01 (to V. A. V.), 1S10OD020095-01, 1R01NS100081 (to E. K. J.), and P30CA006927 (to Fox Chase Cancer Center) and in part by the National Science Foundation through Major Research Instrumentation Grant CNS-09-58854. The *in vitro* studies were also supported by the National PKU Alliance and BioMarin Pharmaceuticals. The authors declare that they have no conflicts of interest with the contents of this article. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.
This article was selected as one of our Editors' Picks.
This article contains supporting methods, Figs. S1–S27, Tables S1–S9, and Movie S1.

[1] To whom correspondence should be addressed. Tel.: 215-204-1973; Fax: 215-204-1532; E-mail: voelz@temple.edu.

[2] The abbreviations used are: PAH, phenylalanine hydroxylase; RS-PAH, resting state phenylalanine hydroxylase; A-PAH, activated phenylalanine hydroxylase; tICA, time-structure–based independent component analysis; MSM, Markov state model; PKU, phenylketonuria; rPAH, rat PAH; TPT, transition path theory; IEC, ion-exchange chromatography; tIC, tICA component; SUMO, small ubiquitin-like modifier; Bistris propane, 1,3-bis[tris(hydroxymethyl)methylamino]propane; YT, yeast extract–tryptone.

model and the crystal structure also differ in their β-strand register at the dimer interface.

Molecular simulations of the homology model and crystal structure, in monomeric and dimeric forms and in the presence and absence of Phe, offer valuable insights into the Phe binding mechanism not available by other methods. Whereas previous experimental studies have estimated binding equilibria for Phe to the preformed ACT domain dimer (8), simulations can provide estimates of Phe binding pathways and rates crucial to understanding the molecular mechanism of allosteric activation (9–14). In particular, because the putative unliganded homology model represents an alternative dimer conformation where the monomers are very close to the RS-PAH conformation, *ab initio* binding simulations can distinguish between conformational selection *versus* induced fit ligand binding mechanisms. Moreover, the conformational motions seen in molecular simulations can be used to understand the effects of mutations and to generate testable hypotheses about alternative conformational states that could be targeted by allosteric effectors (15).

Until now, only submicrosecond simulations of the ACT domain monomer in the absence of Phe have been performed (16). Here, ~633 μs of trajectory data obtained from parallel explicit-solvent molecular dynamics simulations were performed and analyzed to elucidate 1) the conformational dynamics of the ACT domain monomer, 2) the conformational dynamics of Phe-binding encounter complexes of the ACT domain dimer, and 3) pathways and rates of Phe binding to ACT domain dimer conformations. As described below, Markov state model analysis of the trajectory data implicates a conformational selection mechanism for Phe association with ACT domain dimers and reveals key conformational motions coupled to Phe binding. One of these motions corresponds to a ligand gating mechanism, whereas another reflects mobility in a hairpin loop region. Experimental measurements on PAH with a bulky amino acid substitution in this region do not show a shift in the RS-PAH to A-PAH equilibrium, corroborating the role of flexibility in the hairpin loop.

Results

Estimates of binding rates of free Phe to the ACT domain dimer

Multiple independent methods were used to estimate Phe binding rates to the preformed regulatory ACT domain dimer poses: 1) the distribution of observed binding times, 2) the numbers of binding events, 3) implied timescales from Markov state models, and 4) binding flux estimates from transition path theory (TPT).

Estimates of binding rates from the distribution of binding times

Bayesian inference was used to estimate the rate of free Phe binding to an ACT domain dimer from the distribution of binding times. Given N observed binding events, each with an observed time to binding t_i, $i = 1, \ldots, N$, each binding time t_i was assumed to come from a Poisson distribution, $P(t|k) = k \exp(-kt)$, for some unknown binding rate, k. The likelihood of observing the data given the rate k is $P(t_1, t_2, \ldots, t_N|k) = \Pi_i P(t_i|k) = k^N \exp(-k\sum_i t_i)$. By Bayes' theorem, the posterior probability of the value of k, given the observed binding times t_i, is $P(k|t_1, t_2, \ldots, t_N) \propto P(t_1, t_2, \ldots, t_N|k)\, P(k)$ where $P(k)$ is a prior distribution. Two common choices were considered for the prior distribution: the uniform distribution ($P(k) \sim 1$) and the noninformative Jeffreys prior ($P(k) \sim 1/k$). Using the $N = 29$ observed binding times (Table S4), the maximum posterior probability yields an estimate of $k = 6.28 \times 10^7$ s^{-1} M^{-1} (95% confidence interval, 4.63–8.50 $\times 10^7$ s^{-1} M^{-1}) for uniform prior and $k = 6.03 \times 10^7$ s^{-1} M^{-1} (95% confidence interval, 4.65–8.54 $\times 10^7$ s^{-1} M^{-1}) using a Jeffreys prior (Fig. S10).

Estimates of binding rates from the number of binding events

As described in Shirts and Pande (17), simulating M parallel trajectories of a Poisson process results in a probability of observing a binding event after time t is given by $P_M(t) = Mk \exp(-Mkt)$. Therefore, for a collection of trajectories of different lengths, the expected number of binding events $\langle n \rangle$ is $\langle n \rangle(k) = \int M(t)k \exp(-M(t)kt)dt$ where $M(t)$ is the number of trajectories that reach a length of t (18). Because the analysis suggests that only crystal-like ACT domain dimer poses are competent for binding (see Fig. 1*g*), a subset of trajectory data corresponding to crystal-like dimer poses was analyzed (starting conformations 0–9; Fig. S11). The 29 binding events that were observed (assuming ±5.2 from binomial finite sampling error) corresponds to rates in the range of 2–6 $\times 10^7$ s^{-1} M^{-1} (Fig. S12).

Estimates of binding rates from a Markov state model of Phe binding

The slowest MSM implied timescale, $\tau_1 = 256 \pm 65$ ns, corresponds to Phe binding (see Fig. 2 and Fig. S6). The observed relaxation rate corresponding to two-state binding is therefore $k_{obs} = k_{on} + k_{off} = 1/c\tau_1$ where $c = 99.52$ mM is the effective concentration of Phe in the simulations. Because $k_{off} \ll k_{on}$, the estimated binding rate is $k_{on} = 1/c\tau_1 = 3.9 \times 10^7$ s^{-1} M^{-1}. A bootstrap estimate of standard error in ($\ln \tau_1$) yields upper and lower estimates of 5.1 and 3.0 $\times 10^7$ s^{-1} M^{-1}, respectively.

Binding rates estimated using multiple independent methods agree well

Of the 480 trajectories generated, 29 independent Phe binding events were observed (Movie S1). Binding was monitored using the average distance between the Phe ligand and three residues that show close contact to the Phe ligands in the Phe-bound ACT domain dimer crystal structure: Leu^{48}, Leu^{62}, and Ile^{65} (Fig. 1*c*). From the distribution of observed binding times (Table S4), a binding rate of 6 $\times 10^7$ s^{-1} M^{-1} to the ACT domain dimer is inferred (Fig. 1*d*). This value is corroborated by estimates from the number of binding events and MSM-based rate estimates (Table 1).

Physiologically, the estimated binding rate is slower than the theoretical diffusion limit (~10^9 s^{-1} M^{-1}) in the range typical for small molecules binding to protein targets (19, 20). For comparison, the experimentally measured binding rate of benzamidine (similar in size to Phe) to trypsin is ~2.9 $\times 10^7$ s^{-1} M^{-1} (21). It should be kept in mind that our rate estimate reflects the binding rate to the *preformed* ACT domain dimer and that ACT domain dimerization would likely limit the rate of PAH activation at high Phe concentrations.

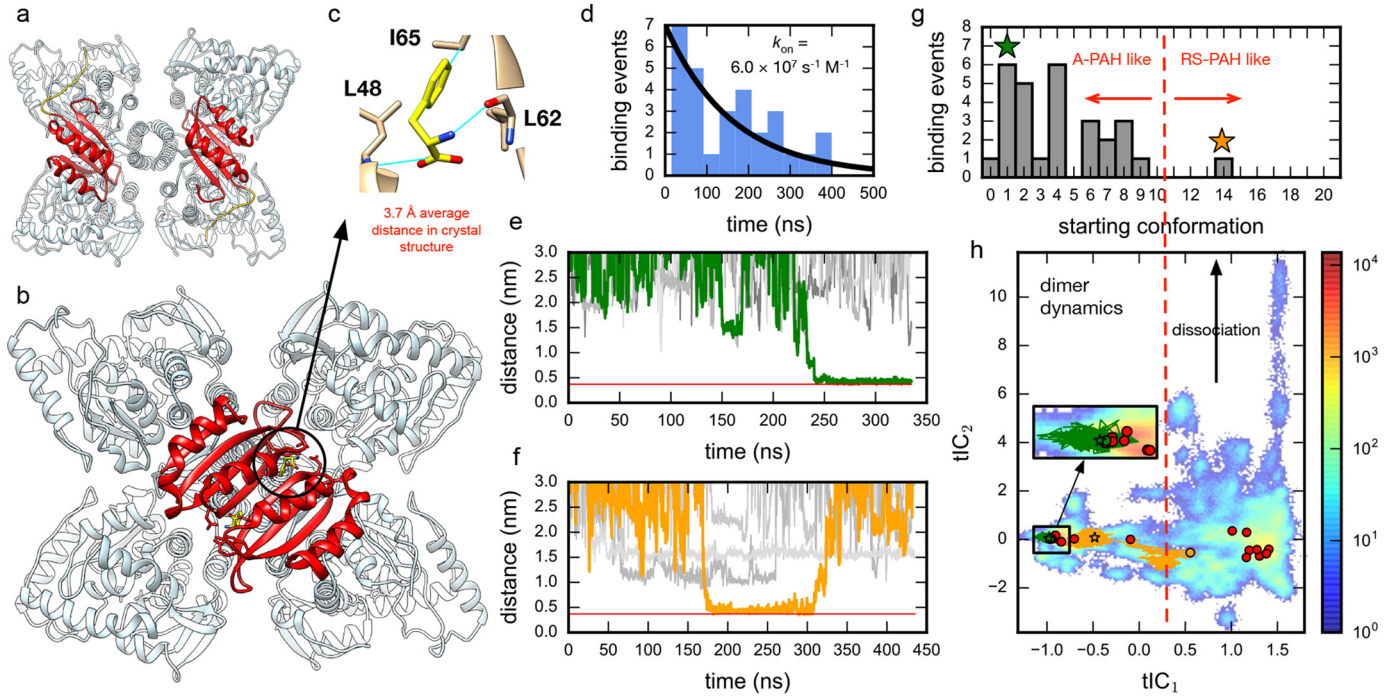

Figure 1. *Ab initio* binding simulations implicate a conformational selection mechanism for Phe association with ACT domain dimers. *a*, the crystal structure of full-length tetrameric mammalian (rat) RS-PAH (PDB code 5DEN) with two ACT domains shown in *red*. The other two ACT domains, *uncolored*, are in the back. *b*, a model of A-PAH (human) prepared using the crystal structure of the Phe-bound ACT domain dimer (PDB code 5FII). Again, two ACT domains are colored *red*, and the two *uncolored* ACT domains are in the back. *c*, average distances (Phe-C)–(Leu^{48}-N), (Phe-C$^{\zeta}$)–(Ile^{65}-C$^{\beta}$), and (Phe-N)–(Leu^{62}-O) were used to monitor productive binding events using a threshold of 0.375 nm. *d*, binding time distribution for 29 observed binding events shown with a Bayesian estimate of the binding rate. *e* and *f*, example traces for Phe binding trajectories (*orange* and *green*) shown with example nonbinding trajectories (*gray*). The *red line* shows the average distance in the crystal structure (PDB code 5FII). *g*, histogram of binding events for initial conformations of the dimer suggests that a crystal-like dimer pose is necessary for productive binding. *Colored stars* mark the starting dimer conformations of the traces shown in *e* and *f*. Only one of the 29 binding trajectories starts from a dimer pose close to the homology model (*orange*); this same trajectory also contains the only unbinding event observed (see *f*). *h*, dimer trajectory data projected to the 2D tICA landscape. *Circles* mark the initial dimer poses with *colored* traces and binding events (*stars*) shown for the trajectories in *e* and *f*.

Sample binding traces are shown in Fig. 1, *e* and *f*. Of the 29 binding events, all but one are found in trajectories started from poses resembling the liganded crystal structure (Fig. 1*g*). The lone exception is found in a trajectory for which an unliganded homology model–like dimer pose first transitions to a more crystal structure–like pose before Phe binds (Fig. 1, *f* and *h*, *orange* trace). This dimer pose may still be suboptimal for binding because unbinding is later observed. No unbinding events are observed in the 28 other binding trajectories or in any of the simulations starting from the Phe-bound crystal structure (PDB code 5FII).

The absence of simulated unbinding events is consistent with estimates of unbinding rates. Zhang *et al.* (8) have used analytic ultracentrifugation to estimate a dissociation constant (K_d) of 8.3 μM for each Phe molecule's association to an ACT domain dimer. Using our estimated k_{on} value, this implies a Phe dissociation rate $k_{off} = k_{on}/K_d$ of about 500 s^{-1} and would suggest that millisecond-timescale trajectories would be required to observe unbinding, whereas our longest trajectories are around 1 μs.

Ab initio binding simulations implicate a conformational selection mechanism for Phe association with ACT domain dimers

To visualize the conformational dynamics of the collection of ACT domain dimer poses seen in the *ab initio* binding trajectories, time-structure– based independent component analysis (tICA) was performed to project trajectory coordinates to a low-dimensional subspace representing the slowest motions of the dimers. The largest tICA component (tIC1) shows slow interconversion between crystal structure–like and homology model–like dimer poses, whereas the next largest component (tIC2) shows motions corresponding to ACT domain dimer dissociation (Fig. 1*h*). Dimer dissociation is not seen for crystal structure–like poses. Despite the numerous interactions between free Phe and the ACT domain dimer poses in the simulations, both ligand binding and dimer association are found to be highly dependent on the dimer pose. This finding suggests a conformational selection mechanism of Phe binding whereby formation of a binding-competent dimer is required before association of Phe; in other words, the formation of the bound dimer is *not* a Phe-induced conformational change.

Simulations of an ACT domain dimer in high concentrations of Phe additionally provide a detailed picture of Phe-binding hot spots on the dimer surface other than the allosteric Phe-binding site. Significant binding propensity between free Phe and Phe^{80} (Fig. S13) is observed, which is intriguing because a somewhat buried Phe^{80} of RS-PAH participates in stabilizing that subunit conformation through cation–π interactions with arginine residues on both the catalytic and multimerization domains on the same subunit (6). Further analysis, however, suggests that the observed binding propensity arises mainly

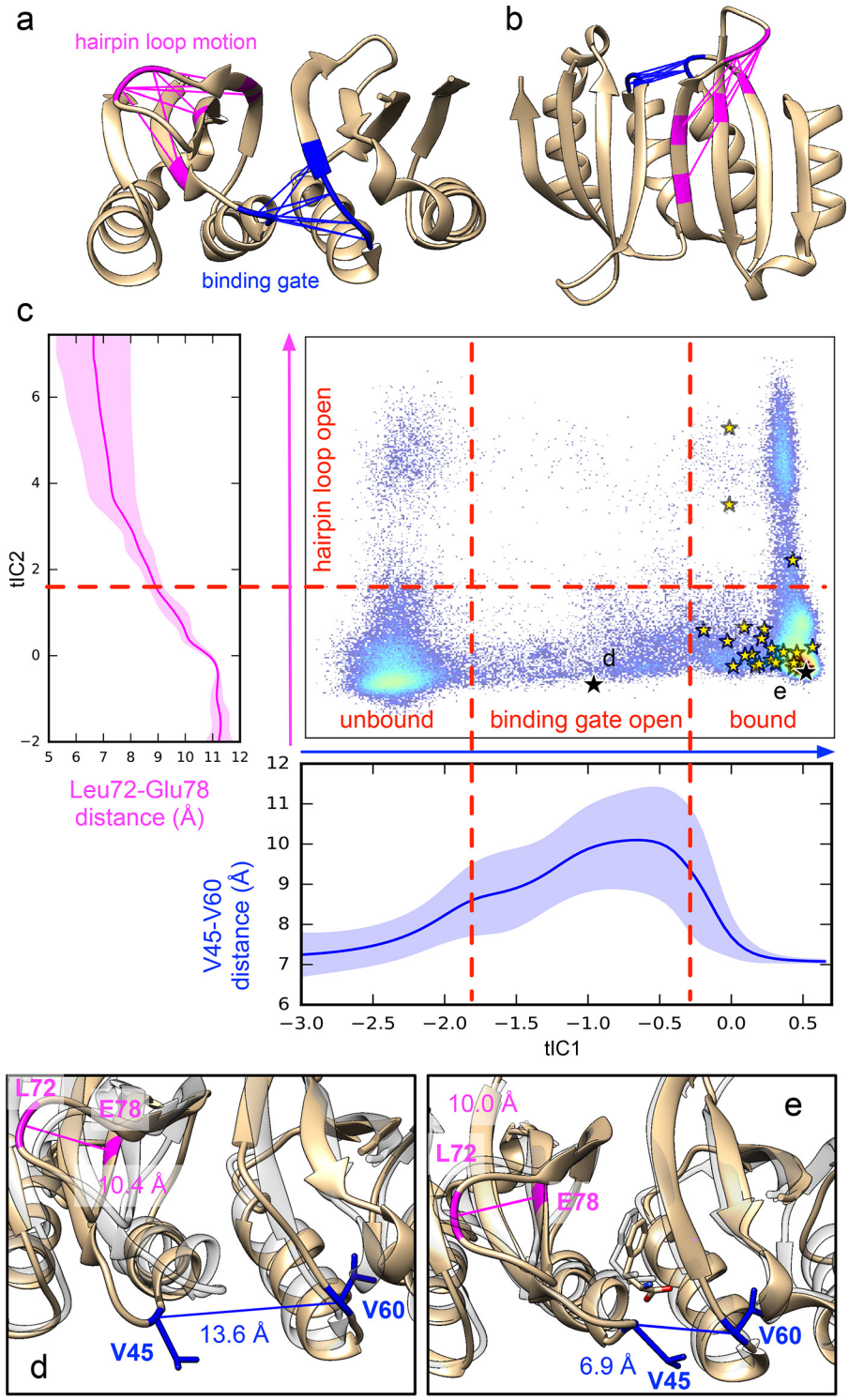

Figure 2. Simulated binding pathways reveal a ligand gating mechanism. *a* and *b*, the slowest dynamics (along tIC1) corresponds to a binding gate motion coupled to Phe association, whereas the next-slowest dynamics corresponds to a hairpin loop motion. The interresidue distances that change greatly during these motions are shown in *blue* and *pink*, respectively. *c*, trajectory data are shown projected to the 2D tICA landscape along with average atomic distances (Val^{45}-C^{α})–(Val^{60}-C^{α}) (*blue*) and (Leu^{72}-C^{β})–(Glu^{78}-C^{α}) (*magenta*) as a function tIC1 and tIC2, respectively. *Yellow stars* denote Phe binding events. *Black stars* show the locations of snapshots shown in *d* and *e*. Representative snapshots (*tan*) are shown of an open binding gate conformation (*b*) and the Phe-bound structure after binding (*e*). For comparison, each are superposed with the crystal structure of Phe-bound ACT domain dimer (transparent *gray*) (PDB code 5FII).

from the high concentration of free Phe used in the simulations; on- and off-rates suggest low overall affinity between free Phe and Phe^{80} of the ACT domain dimer ($K_d \sim 200$ mM; see supporting methods, Fig. S14, and Table S5). Note that Phe^{80}, which is on the eight-stranded β-sheet of the ACT domain dimer, is predicted to be facing inward toward the C-terminal four-helix bundle of the A-PAH tetramer model (refer to Fig. 1*b*).

Simulations of the ACT dimer poses in the absence of Phe show dimer dissociation for homology-like dimer poses but not

Table 1
Estimates of binding rates of free Phe to the preformed ACT domain dimer

Method	Rate[a]	Uncertainty (lower–upper bound)
	($\times 10^7\ s^{-1}\ M^{-1}$)	
Binding time distribution		
Uniform prior	6.28	4.63–8.50
Jeffreys prior	6.03	4.65–8.54
Numbers of binding events	4.0	2.0–6.0
MSM implied timescales	3.9	3.0–5 .1
Transition path theory	6.0	1.0–6 .3

[a] The average of all methods is $5 \pm 3 \times 10^7\ s^{-1}\ M^{-1}$.

for crystal structure–like poses (Fig. 1*h*). This finding is consistent with significant *in vitro* data on the isolated PAH ACT domain that suggest spontaneous dimer formation in the absence of Phe and the stabilization of this dimer in the presence of Phe (8, 22, 23). Other studies show that although full-length RS-PAH in the absence of Phe samples both tetramer and dimer, the addition of Phe favors a stable tetramer (presumably A-PAH) (3).

Simulated binding pathways reveal a binding gate mechanism

To elucidate the mechanism of Phe binding to the ACT domain dimer, MSMs of the conformational dynamics associated with binding using the coordinates of Phe ligands and the protein residues surrounding a single binding site were constructed (see "Experimental procedures" and supporting information). Before doing this, however, the possibility of any Phe-binding cooperativity was probed by building MSMs of both binding sites in the Phe-bound dimer (PDB code 5FII). The results suggest that conformational dynamics in each binding site are independent of each other (data not shown). Binding statistics also support noncooperative Phe binding to the preformed dimer: of the 29 binding events observed in 480 trajectories, two of those 29 trajectories show double Phe binding events (Fig. S15), consistent with a ~6% independent probability of observing binding for each site ($29/480 \approx 2/29$). Consistent with this result are experimental measurements of Phe-dependent ACT domain dimerization in which the model that best fit sedimentation data was the one with identical and independent Phe binding to the preformed ACT domain dimer (8) as well as isothermal titration calorimetry studies of the truncated ACT dimer (7).

We then proceeded to construct an MSM of 75 metastable conformational states for Phe ligands and protein residues surrounding a single binding site. The results reveal structural intermediates along the Phe binding pathway coupled to protein dynamics (Fig. 2). The tICA projections show that the slowest-timescale motions are coupled to binding events with the most significant changes in interresidue distances corresponding to residues that gate the entry of Phe (Glu^{44}/Val^{45} and $Asp^{59}/Val^{60}/Asn^{61}$) (Fig. 2, *blue*). Strikingly, the MSM shows that the protein loop containing Val^{45} must swing open to allow access to the binding site. In all binding trajectories, this gate must open before binding can occur. Once the ligand is bound, the gate closes, helping to stabilize the bound state. More details describing this motion can be found in Table S6 and Figs. S16 and S17. In the absence of free Phe, the slowest two motions of the dimer are also found to occur within the same area that functions as the binding gate for Phe binding (Table S7 and Fig. S18). This suggests that this intrinsic motion is independent of Phe concentration, again consistent with conformational selection.

The next-slowest conformational motion corresponds to bending of the hairpin loop containing Leu^{72} (Fig. 2, *magenta*). Although this motion occurs on a timescale (~100 ns) similar to that of binding, it does not appear coupled to binding; binding occurs regardless of whether the hairpin loop is bent. Using transition path theory, it is estimated that about 8% of binding flux occurs through bent-hairpin pathways (Fig. S8). The mobility of this hairpin loop helps explain a number of experimental observations. First, in the crystal structure of the Phe-bound ACT domain dimer (PDB code 5FII), Leu^{72} is unresolved in one of the four chains, indicative of high mobility. Second, hydrogen/deuterium exchange studies on full-length rPAH (25) showed that a peptide containing residues 67–81, which includes a significant portion of the β-strands on either side of the hairpin loop containing Leu^{72}, displays modest hydrogen/deuterium exchange. There is no difference in hydrogen/deuterium exchange for the peptide in the absence or presence of 5 mM Phe, which is consistent with the prediction from the current molecular dynamics simulations that the hairpin loop's mobility is inherent to the structure and not dependent on Phe binding.

The mutation L72W does not shift the PAH structural equilibrium toward A-PAH

To investigate the mobility of the hairpin loop region, the L72W variant of the rat protein was prepared and tested for the robustness of the RS-PAH ⇔ A-PAH structural equilibrium. In the full-length crystal structure of rat RS-PAH (6), the hairpin loop containing Leu^{72} is located between the multimerization domain and a neighboring catalytic subunit. This is in contrast to an early composite homology model of RS-PAH, made by combining two two-domain PAH structures, both containing the catalytic domain, which suggested a potential clash between Leu^{72} of the ACT domain and Ile^{432} in the multimerization domain (6, 26).

The RS-PAH ⇔ A-PAH structural equilibrium was characterized via measurements of the enzyme's kinetics, intrinsic Trp fluorescence, and affinity for an ion-exchange resin as described below. Taken together, these data are uniformly consistent with the notion that position 72 can accommodate a large bulky amino acid without significantly shifting the RS-PAH ⇔ A-PAH structural equilibrium.

The first, most straightforward measure of the position of the RS-PAH ⇔ A-PAH equilibrium is the assessment of PAH catalytic activity with and without preincubation with Phe. It is well established that the initial velocity of a PAH activity assay is dramatically dependent upon the order of addition of assay components (3). This occurs, in part, because the substrate Phe is also an allosteric activator (2). Thus, one measure of the position of the RS-PAH ⇔ A-PAH is the -fold activity activation seen when PAH is preincubated with Phe prior to the assay. Table 2 includes the kinetic parameters for rPAH and L72W.

Table 2

Kinetic parameters for the PAH-catalyzed production of tyrosine from Phe at pH 7.25 and 25 °C.

The reported fold activation is the ratio of V_{max} values. The value in parentheses is the ratio of values when the assay is done at 1 mM Phe. The latter is included because it is the "standard" approach to reporting activation ratios in the PKU literature.

Variant	Preincubation with Phe	Apparent V_{max}	Apparent K_m for Phe	Hill coefficient	-Fold activation V_{max} (1 mM Phe)
		nmol tyrosine min^{-1} mg^{-1}	*μM*		
rPAH	−	969 ± 163	273 ± 78	1.8 ± 0.7	5.5 (6.9)
rPAH	+	5383 ± 139	148 ± 8	2.9 ± 0.4	
L72W	−	1189 ± 33	678 ± 39	1.4 ± 0.1	3.9 (4.9)
L72W	+	4584 ± 76	532 ± 16	2.1 ± 0.1	

V_{max} values for L72W are comparable with the rPAH protein, whereas the K_m values for Phe for L72W are elevated ~ 2–3-fold. For rPAH and L72W, the -fold activation values by preincubation with 1 mM Phe are comparable and within the range of published values, consistent with L72W existing predominantly as RS-PAH in the absence of Phe.

The intrinsic fluorescence of mammalian PAH, which also changes dramatically upon allosteric activation, provides a second measure of the position of the RS-PAH ⇔ A-PAH equilibrium. A red shift is observed upon addition of Phe to PAH as was first reported by Kaufman and co-workers (27). The fluorescence change was later determined to arise predominantly from a change in the environment of Trp^{120} (28). Trp^{120} is near the hinge between the regulatory and catalytic domains of PAH that is predicted to rotate ~90° during the RS-PAH to A-PAH transition, thus changing the Trp^{120} environment. Fig. 3*a* shows the intrinsic fluorescence for rPAH and L72W before and after the addition of 1 mM Phe. It is not surprising that the intrinsic fluorescence of L72W is not identical to rPAH due to the addition of a tryptophan at position 72. Nevertheless, the similar extent of fluorescence red shift between rPAH and L72W observed upon addition of 1 mM Phe is consistent with the variant existing predominantly as RS-PAH in the absence of added Phe and transitioning to A-PAH upon addition of Phe. No additional red shift was observed when Phe was increased to 5 mM (not shown).

The third measure of the position of the RS-PAH ⇔ A-PAH equilibrium is the behavior of the protein during ion-exchange chromatography (IEC) with and without inclusion of Phe in the running buffer. It was previously shown that inclusion of 1 mM Phe in the column buffer increases the retention of rPAH to a Q column (3). Fig. 3, *b* and *c*, illustrate this phenomenon for rPAH and L72W at 0 and 1 mM Phe. As before, 1 mM Phe causes a dramatic increase in protein retention for rPAH (Fig. 3*b*). This shift is interpreted as indicative of the RS-PAH to A-PAH transition. For L72W, whose intrinsic affinity for the IEC column is greater than rPAH, 1 mM Phe causes a similar increase in retention on the IEC column (Fig. 3*c*).

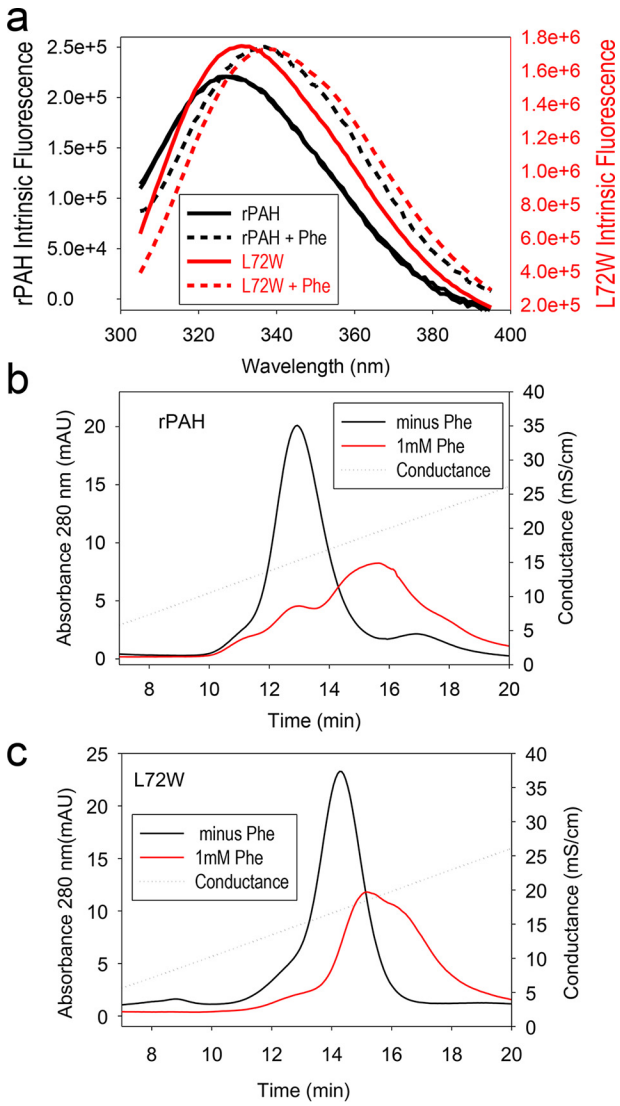

Figure 3. Evaluation of the RS-PAH ⇔ A-PAH equilibrium for rPAH and L72W. *a*, intrinsic protein fluorescence for rPAH (*black*) and L72W (*red*) in the absence (*solid lines*) and presence (*dashed lines*) of 1 mM Phe. *b*, ion exchange behavior of rPAH at 0 (*black*) and 1 mM (*red*) Phe. *c*, ion exchange behavior of L72W at 0 (*black*) and 1 mM (*red*) Phe. *mAU*, milli-absorbance units; *mS*, millisiemens.

Simulations of the ACT domain monomer reveal a spontaneous transition between A-PAH–like and RS-PAH–like conformations

Although crystal structures of monomeric ACT domains (*e.g.* PDB code 5DEN) show a helical turn at residues 61–64, the recent structure of a Phe-bound ACT domain dimer (PDB code 5FII) shows these residues in a β-strand conformation, which likely enables Phe to bind at the dimer interface. To examine whether these RS-PAH–like and A-PAH–like conformations can interconvert via intrinsic conformational dynamics of the monomer, over 185 μs of trajectory data were obtained using simulations initiated from the 21 interpolated poses, with and without Phe, according to a similar protocol as was used for the dimer simulations.

A 2D tICA projection of the trajectory data shows that the slowest motion corresponds to transitions between RS-PAH–like and A-PAH–like conformations (Fig. 4), both in the

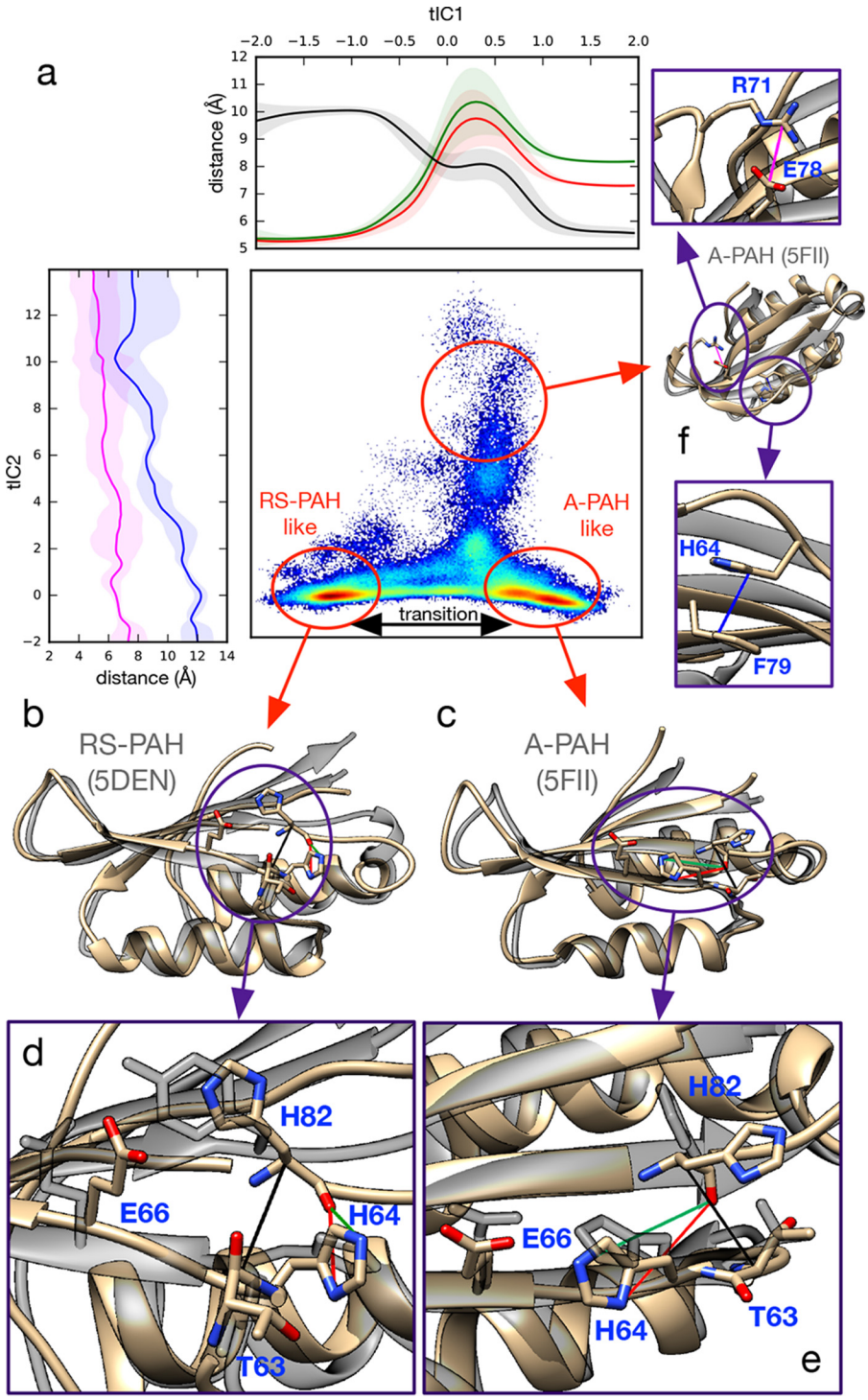

Figure 4. Simulations of the ACT domain monomer reveal spontaneous transitions between RS-PAH–like and A-PAH–like conformations in the absence of Phe. *a*, monomer trajectory data projected to the 2D tICA landscape shows that RS-PAH–like (*b*) and A-PAH–like (*c*) conformations are significantly populated metastable states. Atomic distances are shown to highlight structural changes: along tIC1, (His^{64}-$N^{\delta1}$)–(His^{82}-O) (*red*), (His^{64}-$N^{\epsilon2}$)–(His^{82}-O) (*green*), and (Thr^{63}-C^{α})–(His^{82}-C^{α}) (*black*); along tIC2, (Arg^{78}-C^{ζ})–(Glu^{78}-C^{δ}) (*pink*) and (His^{64}-C^{γ})–(Phe^{79}-C^{γ}) (*blue*). Ribbon structures (*tan*) show selected conformations from the simulation trajectory data superposed with ACT domain crystal structures (*transparent gray*) representative of the RS-PAH state (PDB code 5DEN) and A-PAH state (PDB code 5FII). Key structural details for RS-PAH–like (*d*) and A-PAH–like (*e*) states are shown. *f*, an off-pathway conformational state sampled by the simulations.

absence and presence of Phe (Figs. S19–S21). In the RS-PAH–like conformation, Leu^{62}, Thr^{63}, and His^{64} have an α-helical backbone conformation, stabilized by a hydrogen bond between Asn^{61}(O^{δ}) and Thr^{63}(NH). The His^{64} side chain is near the backbone carbonyl oxygen of His^{82}, preventing cross-strand H-bonding. Transition to an A-PAH–like conformation occurs via Leu^{62} and His^{64} transitioning to a β-sheet backbone conformation, allowing Thr^{63} to form a backbone hydrogen bond with His^{82} with rotation of the His^{64} side chain to form a hydrogen bond with Glu^{66}. In over 91 μs of trajectory data for

the monomer in the absence of free Phe, 35 and 80 independent transitions were observed from RS- to A-PAH–like conformations and from A- to RS-PAH–like conformations, respectively, with 10 trajectories containing two interconversions. In the presence of free Phe (over 95 μs of trajectory data), 20 and six transitions were observed, respectively. To estimate the timescales associated with the RS-PAH–like to A-PAH–like transition, a 40-state MSM was constructed from the trajectory data of the monomer in the absence of free Phe. Interconversion rates, estimated using transition path theory and MSM-implied timescales, suggest roughly equal interconversion rates on timescales ranging from 3 to 14 μs (supporting methods and Table S8).

The second-slowest motion identified in the monomer trajectory data corresponds to a transition to an off-pathway state in which the His^{64} side chain favorably stacks with Phe^{74}, disrupting β-strand pairing (see Fig. 4). This conformation is partially stabilized by a salt bridge between Arg^{71} and Glu^{78} whose formation is facilitated by the same hairpin loop motions seen in the dimer simulations. Rates and binding affinities were also estimated for free Phe binding to both the allosteric binding site and Phe^{80} on the monomer using methods similar to the presented analysis of the ACT dimer; in all cases, similar association and dissociation rates, which are close to the diffusion limit, were found, and there was negligible binding affinity (Table S9).

Discussion

Our combined simulation and experimental study gives unprecedented insights into the conformational dynamics of the regulatory ACT domain of PAH and the molecular mechanism by which free Phe binds these domains. In both the absence and presence of free Phe, simulations of monomeric and dimeric ACT domains suggest that spontaneous transitions occur between RS-PAH–like and A-PAH–like conformations. MSM analysis of *ab initio* binding simulations suggests that Phe binding is only possible when the ACT domains are dimerized in an A-PAH–like pose, thus following a conformational selection mechanism.

Our simulations of the PAH regulatory domain are the most extensive to date and the first to incorporate information about the ACT domain dimer. Prior molecular dynamics studies of allosteric Phe binding to PAH considered a putative allosteric Phe site in RS-PAH (16, 29). Those studies were initiated prior to the original suggestion of an allosteric Phe site on A-PAH ACT domain dimer (3) and completed coincident with the publication of the Phe-bound ACT domain dimer structure (4). An allosteric Phe-binding site on RS-PAH is not supported by recent crystal structures of RS-PAH obtained at high concentrations of Phe (7).

A unique aspect of the current simulation work is the opportunity to observe the statistics of Phe binding events across large data sets of parallel simulations and compare rate estimates from direct observations (*i.e.* distributions of binding times and numbers of events) with rate estimates derived from MSM methods (*i.e.* implied timescales and TPT). The estimates agree well with each other, validating the accuracy of the MSMs. The MSMs, however, provide a wealth of information about how conformational motions of the ACT domain dimer are coupled to binding, which is difficult to obtain by other means.

An additional strength of the current study is the comparison of ACT domain simulations in both the absence and presence of Phe (~100 mM; to observe sufficient binding statistics). Although some superficial differences in simulations performed with and without Phe are observed, there is good reason to believe that the conformational dynamics of both monomeric and dimeric ACT domain are mostly independent of Phe concentration. Root mean square fluctuations of dimer and monomer residues are similar in the presence and absence of Phe (Fig. S22). The ACT domain has similar fluctuations in both monomeric and dimeric forms except for the binding gate region, which has smaller fluctuations in the dimer (Fig. S23).

The simulations identify two motions important to PAH function. One is a "binding gate" motion in which the helices containing Val^{45} and Val^{60} on opposite subunits must open to allow entry and association of Phe. Remarkably, it appears that opening of the binding gate is required for binding, which in turn may have important functional implications. The disease-associated variant G46S dramatically reduces allosteric Phe binding and has been recently found to promote insoluble aggregates *in vitro* (4, 30), consistent with the importance of conformational dynamics in this region. Furthermore, the open states of the binding gate predicted from the simulations may serve as potential cryptic binding sites that can be targeted by small-molecule drugs (15, 31). For instance, it may be possible to discover allosteric effectors that bind the dimer interface much like Phe but are better optimized to bind the "open gate" dimer pose.

Another key motion identified in the simulations is in the "hairpin loop" region of the ACT domain that is flexible and populates multiple conformational states in simulations of both dimer and monomer poses. The experimental studies presented here indicate that a bulky mutation at the hairpin loop does not shift the PAH structural equilibrium toward the A-PAH state; rather, it is likely accommodated by the loop's inherent mobility.

Missing from our current mechanistic picture of PAH activation are information about the rates and pathways by which regulatory ACT domains dimerize and how this dimerization is coupled to conformational changes in the PAH tetramer that accompany enzyme activation. In the future, we hope to gain further insight into this process by using new MSM methodologies along with biased conformational sampling to estimate rates, pathways, and affinities of ACT domain dimerization from simulations. Several disease-associated PAH variants with mutations in the regulatory ACT domain have been characterized (32), providing an exciting opportunity to use simulation studies to determine the extent to which mutations perturb the dimerization process as opposed to ligand binding.

Conclusions

In this work, we used over 633 μs of explicit-solvent trajectory data to assemble a detailed mechanistic picture of how free Phe associates with the regulatory ACT domain dimer of phenylalanine hydroxylase. An MSM analysis of *ab initio* binding

pathways and rates strongly suggests a conformational selection mechanism in which binding can only occur to a preformed dimer in an activated pose. The MSM analysis identifies two important motions in the regulatory ACT domain dimer: 1) a ligand gating mechanism in which the opening of a binding gate is required to allow ligand entry and 2) a hairpin loop motion. We give evidence for the importance of mobility in this hairpin through experimental studies showing that the L72W variant does not shift the structural equilibrium of the enzyme toward the activated state. ACT domain monomer simulations reveal spontaneous transitions between A-PAH–like and RS-PAH–like conformations. Structural intermediates identified from the simulations may be helpful in future studies of disease-associated mutations and provide new directions toward the development of phenylketonuria therapeutics.

Computational methods

Simulation setup

To test whether Phe could induce tightly bound dimer conformations from states other than the crystal structure of the Phe-bound ACT domain dimer (PDB code 5FII), parallel simulations were initiated from 21 minimized and equilibrated structures. These structures were generated by interpolation from the crystal structure (with Phe removed) to the unliganded ACT domain dimer (*i.e.* the homology model) via a rigid-body morphing algorithm (33) implemented in UCSF Chimera (34) (Fig. S2). Simulation systems were prepared for the following structures: 1) the crystallographic dimer bound by two Phe ligands, 2) the 21 morphed dimer poses in the absence of free Phe ligand, and 3) the 21 morphed dimer poses in the presence of 19 free (unbound) Phe at an effective concentration of 99.5 mM. A similar strategy was used in a series of monomer simulations including only one chain of the 21 morphed dimer poses (Fig. S3). A molecular topology for the zwitterionic Phe ligand was constructed using the AmberTools software package (35). The antechamber program was used to derive partial charges using the AM1-BCC method (36) (Fig. S4), and force field parameters were provided by the general Amber force field (GAFF) (37). The ACPYPE program (38) was then used to convert the topology file format for use with the GROMACS molecular dynamics package (39).

Parallel molecular dynamics simulations were performed using GROMACS 4.5.4 on the Folding@home distributed computing platform (40). The AMBER ff99sb-ildn NMR force field (41) was used in combination with the TIP3P explicit-solvent model. Cubic periodic boxes of $(\sim 7\ \text{nm})^3$ were filled with solvated protein and counterions ($\sim$100 mM NaCl) to neutralize the system. A full list of particle numbers and box sizes can be found in Tables S1 and S2. Simulations were minimized and then pressure-equilibrated at 1 atm for 200 ps using constant-pressure molecular dynamics coupled to a Berendsen barostat with a time constant of 1 ps and compressibility of 4.5×10^{-5} bar^{-1}. Trajectory data were generated using constant-volume molecular dynamics at 300 K, a stochastic (Langevin) integration with a 2-fs time step, and friction constant of 1 ps^{-1} coupled to a Berendsen thermostat. Hydrogen bonds were constrained using the LINCS algorithm (42), and particle mesh Ewald electrostatics was used with nonbonded cutoffs of 9 Å. Snapshots of protein atoms were recorded every 100 ps, and all atoms were recorded every 1 ns. The average trajectory length was over 200 ns, and the longest trajectory was over 1.3 μs (Table S3).

Time-lagged independent component analysis

tICA was applied to interatomic distance observables (Fig. S5) to determine the conformational degrees of freedom along which the most time-correlated (*i.e.* slowest) motions occurred in the simulations (43–45). The tICA components (tICs) are found by maximizing the objective function $\langle \alpha_i | \mathbf{C}^{(\Delta t)} | \alpha_i \rangle$ subjected to certain constraints where $|\alpha_i\rangle$ are the tICs, and $\mathbf{C}^{(\Delta t)}$ is a time-lagged correlation matrix of elements $C_{ij} = \langle x_i(t) x_j(t + \Delta t) \rangle$ for (zero-mean) structural observables x_i and x_j. The tICA calculations were performed with MSMBuilder (46) using a tICA lag time of $\Delta t = 5$ ns. Additional details can be found in the supporting methods.

Markov state model construction

MSMBuilder (46) was used to construct MSMs from the simulation trajectory data. To adequately capture the important conformational motions associated with binding, MSMs were constructed from 28 trajectories of Phe binding events ($\sim$11.3 μs in total) and 29 trajectories starting from the Phe-bound crystal structure ($\sim$12.3 μs in total) where the Phe remained bound. The final MSM consisted of 75 metastable conformational states obtained by *k*-centers clustering of trajectory data projected to the four largest tICA components. The matrix of transition probabilities $\mathbf{T}^{(\tau)}$ between metastable states for a given lag time τ were computed from the observed number of transitions using a maximum-likelihood estimator that enforces detailed balance. The slowest relaxation timescales of the dynamics, *i.e.* the so-called *implied timescales*, are given by $\tau_n = -\tau/(\ln \mu_n)$ where μ_n are the largest eigenvalues of $\mathbf{T}^{(\tau)}$. Implied timescales were calculated as a function of lag time plateau beyond a lag time of 20 ns, indicative of Markovian dynamics (Fig. S6). Therefore, a lag time of $\tau = 20$ ns was chosen to construct the final MSM. The optimal number of metastable states (75) for MSM construction was determined using the generalized matrix Rayleigh quotient (GMRQ) variational cross-validation method (47) (Fig. S7).

Transition path theory analysis

TPT (48, 49) was used to estimate binding pathways, fluxes, and rates using the 75-state MSM of Phe association constructed as described above. TPT uses the transition probability matrix $\mathbf{T}^{(\tau)}$ to solve a set of self-consistent equations to obtain commitor values q_i^+ for every state *i*. The commitor value q_i^+ is the probability that a trajectory started from state *i* will reach a set of *sink* states (*B*) before reaching a set of *source* states (*A*). Once the commitor values are determined, the reaction rate k_{AB} can be computed as follows.

$$k_{AB} = \frac{F}{\tau \sum_i \pi_i (1 - q_i^+)} \quad \text{(Eq. 1)}$$

$$F = \sum_{i \in A} \sum_{i \notin A} \pi_i T_{ij}^{(\tau)} q_i^+ \quad \text{(Eq. 2)}$$

where F is the total flux and π_i is the equilibrium population of state i. The rate k_{AB} is the expected number of observed $A \rightarrow B$ transitions per time unit τ, which can be converted to a predicted rate constant by $k'_{AB} = k_{AB}/c$ where c is the simulated molar concentration. See supporting methods and Figs. S8 and S9 for more details.

Experimental procedures

Plasmid construction

The rPAH gene was amplified from pERPH5 (50) using primers 5′-AGATTGGTGGCGCAGCTGTTGTCCTGGAGAATGG-3′ and 5′-GAGGAGAGTTTAGACTTACGACTTTATCTTCTGCAG-3′ and inserted into pETHSUL (a gift from P. J. Loll) using the ligation-independent cloning method (51). The sequence-confirmed construct is named pMRH162. The L72W mutation was introduced by GeneWiz, resulting in a plasmid named pMRH162-L72W. The entire PAH gene sequences of both plasmids were verified (GeneWiz).

Protein expression and purification

Work with PAH variants designed to have a shifted quaternary structure equilibrium required abandoning the traditional purification of mammalian PAH, which takes advantage of the high affinity of A-PAH for phenyl-Sepharose resin and the low affinity of RS-PAH for the same resin (52). In this method, inclusion/exclusion of Phe from the chromatography buffer drives effective protein purification. For PAH variants designed to shift the RS-PAH ⇔ A-PAH equilibrium, a cleavable N-terminal His_6-SUMO purification tag, wherein the cleaved PAH starts at Ala^2, was chosen. In addition, L72W expression was carried out in the presence of the GroEL/GroES chaperone (24).

For L72W, the His_6-SUMO fusion construct was expressed and purified as follows. *Escherichia coli* BL21(DE3), containing plasmids pGro7 (24) and pMRH162-L72W, was grown overnight at 37 °C in 2× YT medium supplemented with 100 μg/ml ampicillin, 25 μg/ml chloramphenicol, and 1% glucose. The overnight culture was used to inoculate 6 liters of 2× YT medium supplemented with 100 μg/ml ampicillin, 25 μg/ml chloramphenicol, 3.33 mM arabinose, 10 μM $FeCl_3$, and 4 mM $MgSO_4$ and grown at 25 °C. Protein expression was induced with 0.5 mM isopropyl 1-thio-β-D-galactopyranoside once the A_{600} reached 0.5 and grown overnight at 18 °C. Cells were harvested after 29 h, flash frozen in liquid N_2, and stored at −80 °C. Expression of rPAH from the His_6-SUMO fusion was carried out similarly in the absence of the pGro7 plasmid.

Protein purification was as follows. Specifics are provided for L72W. A cell pellet (20 g) was thawed and resuspended in 92 ml 50 mM Tris-HCl (pH 7.4), 0.5 M NaCl, 50 μM EDTA, 10% glycerol, 20 mM imidazole, and 0.1% Triton X-100 (Buffer Ni-20). Cells were lysed by passing through a cell press eight times. Following centrifugation, the supernatant was passed through 0.45-μm Millex filter (Millipore) and applied onto a prepacked HisTrap HP (GE Healthcare) 5-ml column pre-equilibrated with Buffer Ni-20. After extensively washing with Buffer Ni-20, the fusion protein was eluted in 50 mM Tris-HCl (pH 7.4), 0.5 M NaCl, 50 μM EDTA, 10% glycerol, 0.1% Triton X-100, and 500 mM imidazole (Buffer Ni-500). Fractions (containing ~120 mg of fusion protein) were pooled, 120 μg of the catalytic domain of the *Saccharomyces cerevisiae* SUMO hydrolase (51) was added, and the mixture was dialyzed overnight at 4 °C against 2 × 2 liter of buffer without imidazole (Buffer Ni-0).

The digested dialyzed protein (22 ml) was loaded onto the HisTrap column that was pre-equilibrated with Buffer Ni-20. Detagged protein came out in the flow-through and Buffer Ni-20 wash. The pooled protein (~88 mg) was dialyzed overnight against 30 mM Tris-HCl (pH 7.4) and 15% glycerol.

The dialyzed sample (30 ml; 4.38 mg/ml) was passed through 0.45-μm Millex filter. Further purification was done in two batches on a 1-ml HiTrap Q HP column (GE Healthcare) pre-equilibrated with 95% 30 mM Tris-HCl (pH 7.4) and 15% equilibrated with 30 mM Tris-HCl (pH 7.4), 15% glycerol, 20 mM KCl. Once the protein was loaded onto the column in this buffer, it was eluted in a 30-column volume linear gradient of the same buffer varying the KCl from 20 mM to 400 mM. Fractions (2 ml) containing PAH protein were flash frozen and stored at −80 °C.

PAH activity assays

PAH activity was measured at 25.0 °C by following tyrosine fluorescence intensity at an excitation wavelength of 275 nm and an emission wavelength of 305 nm using a PTI Fluorescence System spectrophotometer equipped with a USHIO xenon short arc lamp and Felix for Windows software. The standard assay mixture (2-ml final volume) contained 20 mM Bistris propane (pH 7.3), L-phenylalanine (at 10 μM, 50 μM, 100 μM, 200 μM, 300 μM, 600 μM, 800 μM, and 1 mM), 10 μM ammonium Fe(II) sulfate hexahydrate, 40 μg/ml catalase (Sigma; bovine liver, aqueous suspension, CAS 9001-05-2, catalogue number C30-100 mg), 75 μM tetrahydrobiopterin (Sigma), and 75 μM dithiothreitol (Fisher). Approximately 2–8 μg of PAH protein was preincubated for 60 s in a total of 22 μl of assay mixture without Phe (for the −Phe condition) or with Phe in a concentration equal to the assay mixture (for the +Phe condition). That preincubation mixture (20 μl) was then added to 1980 μl of reaction mixture, and data collection was started after 20 s. Although hysteresis can be observed, tyrosine production during the first 50 s is nearly linear and is used to calculate initial velocity. Tyrosine concentration was determined from a calibration curve using 10–40 μM L-tyrosine in assay buffer. Data were plotted using SigmaPlot and fit to a sigmoidal Hill equation yielding apparent V_{max}, K_m, and Hill coefficient values.

PAH intrinsic fluorescence determination

PAH intrinsic fluorescence was measured at 25.0 °C with an excitation wavelength of 295 nm and an emission spectrum of 305–400 nm using a PTI Fluorescence System spectrophotometer equipped with a USHIO xenon short arc lamp and Felix for Windows software. PAH, which had been eluted from a preparative HiTrap Q run, was diluted to a concentration of 0.5 μM subunit in 2 ml of 30 mM Tris-HCl (pH 7.4) and 150 mM KCl. Data were formatted for display using SigmaPlot 10.0.

IEC behavior of PAH

Analytical IEC was performed as reported previously (3) with the exception that all proteins, which had been purified on the

HiTrap Q HP column, were first buffer-exchanged into 30 mM Tris-HCl (pH 7.4), 20 mM KCl, and 15% glycerol. As before, 100 μg of protein at ~1 mg/ml was injected onto a 1-ml HiTrap Q HP column and eluted in 30 mM Tris-HCl (pH 7.4) and 15% glycerol using a 20–400 mM KCl gradient containing either 0 or 1 mM Phe in the running buffer. The column was run at 0.5 ml/min using a 10-column-volume gradient. Blank gradients were run between samples.

Author contributions—Y. G. data curation; Y. G. and V. A. V. software; Y. G. and E. K. J. formal analysis; Y. G., E. K. J., and V. A. V. validation; Y. G., E. B., S. S., M. R. H., E. C. A., E. K. J., and V. A. V. investigation; Y. G., E. K. J., and V. A. V. visualization; Y. G., E. B., M. R. H., E. C. A., E. K. J., and V. A. V. methodology; Y. G., M. R. H., E. K. J., and V. A. V. writing-original draft; Y. G., E. B., E. C. A., E. K. J., and V. A. V. writing-review and editing; E. K. J. and V. A. V. conceptualization; E. K. J. and V. A. V. resources; E. K. J. and V. A. V. supervision; E. K. J. and V. A. V. funding acquisition; E. K. J. and V. A. V. project administration.

Acknowledgments—We thank the participants of Folding@home, without whom this work would not be possible.

References

1. Aravind, L., and Koonin, E. V. (1999) Gleaning non-trivial structural, functional and evolutionary information about proteins by iterative database searches. *J. Mol. Biol.* **287,** 1023–1040 CrossRef Medline
2. Jaffe, E. K. (2017) New protein structures provide an updated understanding of phenylketonuria. *Mol. Genet. Metab.* **121,** 289–296 CrossRef Medline
3. Jaffe, E. K., Stith, L., Lawrence, S. H., Andrake, M., and Dunbrack, R. L., Jr. (2013) A new model for allosteric regulation of phenylalanine hydroxylase: implications for disease and therapeutics. *Arch. Biochem. Biophys.* **530,** 73–82 CrossRef Medline
4. Patel, D., Kopec, J., Fitzpatrick, F., McCorvie, T. J., and Yue, W. W. (2016) Structural basis for ligand-dependent dimerization of phenylalanine hydroxylase regulatory domain. *Sci. Rep.* **6,** 23748 CrossRef Medline
5. Kobe, B., Jennings, I. G., House, C. M., Michell, B. J., Goodwill, K. E., Santarsiero, B. D., Stevens, R. C., Cotton, R. G., and Kemp, B. E. (1999) Structural basis of autoregulation of phenylalanine hydroxylase. *Nat. Struct. Biol.* **6,** 442–448 CrossRef Medline
6. Arturo, E. C., Gupta, K., Héroux, A., Stith, L., Cross, P. J., Parker, E. J., Loll, P. J., and Jaffe, E. K. (2016) First structure of full-length mammalian phenylalanine hydroxylase reveals the architecture of an autoinhibited tetramer. *Proc. Natl. Acad. Sci. U.S.A.* **113,** 2394–2399 CrossRef Medline
7. Meisburger, S. P., Taylor, A. B., Khan, C. A., Zhang, S., Fitzpatrick, P. F., and Ando, N. (2016) Domain movements upon activation of phenylalanine hydroxylase characterized by crystallography and chromatography-coupled small-angle X-ray scattering. *J. Am. Chem. Soc.* **138,** 6506–6516 CrossRef Medline
8. Zhang, S., Roberts, K. M., and Fitzpatrick, P. F. (2014) Phenylalanine binding is linked to dimerization of the regulatory domain of phenylalanine hydroxylase. *Biochemistry* **53,** 6625–6627 CrossRef Medline
9. Gu, S., Silva, D.-A., Meng, L., Yue, A., and Huang, X. (2014) Quantitatively characterizing the ligand binding mechanisms of choline binding protein using Markov state model analysis. *PLoS Comput. Biol.* **10,** e1003767 CrossRef Medline
10. Huang, X., and De Fabritiis, G. (2014) Understanding molecular recognition by kinetic network models constructed from molecular dynamics simulations. *Adv. Exp. Med. Biol.* **797,** 107–114 CrossRef Medline
11. Stanley, N., Pardo, L., and Fabritiis, G. D. (2016) The pathway of ligand entry from the membrane bilayer to a lipid G protein-coupled receptor. *Sci. Rep.* **6,** 22639 CrossRef Medline
12. Paul, F., Noé, F., and Weikl, T. R. (2018) Identifying conformational-selection and induced-fit aspects in the binding-induced folding of PMI from Markov state modeling of atomistic simulations. *J. Phys. Chem. B* **122,** 5649–5656 CrossRef Medline
13. Plattner, N., and Noé, F. (2015) Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and Markov models. *Nat. Commun.* **6,** 7653 CrossRef Medline
14. Malmstrom, R. D., Kornev, A. P., Taylor, S. S., and Amaro, R. E. (2015) Allostery through the computational microscope: cAMP activation of a canonical signalling domain. *Nat. Commun.* **6,** 7588 CrossRef Medline
15. Bowman, G. R., Bolin, E. R., Hart, K. M., Maguire, B. C., and Marqusee, S. (2015) Discovery of multiple hidden allosteric sites by combining Markov state models and experiments. *Proc. Natl. Acad. Sci. U.S.A.* 201417811
16. Carluccio, C., Fraternali, F., Salvatore, F., Fornili, A., and Zagari, A. (2013) Structural features of the regulatory ACT domain of phenylalanine hydroxylase. *PLoS One* **8,** e79482-e79413 CrossRef Medline
17. Shirts, M. R., and Pande, V. S. (2001) Mathematical analysis of coupled parallel simulations. *Phys. Rev. Lett.* **86,** 4983–4987 CrossRef Medline
18. Voelz, V. A., Bowman, G. R., Beauchamp, K., and Pande, V. S. (2010) Molecular simulation of *ab initio* protein folding for a millisecond folder NTL9(1–39). *J. Am. Chem. Soc.* **132,** 1526–1528 CrossRef Medline
19. Pang, X., and Zhou, H.-X. (2017) Rate constants and mechanisms of protein-ligand binding. *Annu. Rev. Biophys.* **46,** 105–130 CrossRef Medline
20. Kokh, D. B., Amaral, M., Bomke, J., Grädler, U., Musil, D., Buchstaller, H.-P., Dreyer, M. K., Frech, M., Lowinski, M., Vallee, F., Bianciotto, M., Rak, A., and Wade, R. C. (2018) Estimation of drug-target residence times by τ-random acceleration molecular dynamics simulations. *J. Chem. Theory Comput.* **14,** 3859–3869 CrossRef Medline
21. Guillain, F., and Thusius, D. (1970) Use of proflavine as an indicator in temperature-jump studies of the binding of a competitive inhibitor to trypsin. *J. Am. Chem. Soc.* **92,** 5534–5536 CrossRef Medline
22. Zhang, S., and Fitzpatrick, P. F. (2016) Identification of the allosteric site for phenylalanine in rat phenylalanine hydroxylase. *J. Biol. Chem.* **291,** 7418–7425 CrossRef Medline
23. Li, J., Ilangovan, U., Daubner, S. C., Hinck, A. P., and Fitzpatrick, P. F. (2011) Direct evidence for a phenylalanine site in the regulatory domain of phenylalanine hydroxylase. *Arch. Biochem. Biophys.* **505,** 250–255 CrossRef Medline
24. Nishihara, K., Kanemori, M., Kitagawa, M., Yanagi, H., and Yura, T. (1998) Chaperone coexpression plasmids: differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese cedar pollen, Cryj2, in *Escherichia coli*. *Appl. Environ. Microbiol.* **64,** 1694–1699 Medline
25. Li, J., Dangott, L. J., and Fitzpatrick, P. F. (2010) Regulation of phenylalanine hydroxylase: conformational changes upon phenylalanine binding detected by hydrogen/deuterium exchange and mass spectrometry. *Biochemistry* **49,** 3327–3335 CrossRef Medline
26. Erlandsen, H., and Stevens, R. C. (1999) The structural basis of phenylketonuria. *Mol. Genet. Metab.* **68,** 103–125 CrossRef Medline
27. Phillips, R. S., Parniak, M. A., and Kaufman, S. (1984) Spectroscopic investigation of ligand interaction with hepatic phenylalanine hydroxylase: evidence for a conformational change associated with activation. *Biochemistry* **23,** 3836–3842 CrossRef Medline
28. Knappskog, P. M., and Haavik, J. (1995) Tryptophan fluorescence of human phenylalanine hydroxylase produced in *Escherichia coli*. *Biochemistry* **34,** 11790–11799 CrossRef Medline
29. Carluccio, C., Fraternali, F., Salvatore, F., Fornili, A., and Zagari, A. (2016) Towards the identification of the allosteric Phe-binding site in phenylalanine hydroxylase. *J. Biomol. Struct. Dyn.* **34,** 497–507 CrossRef Medline
30. Leandro, J., Saraste, J., Leandro, P., and Flatmark, T. (2017) PKU mutation p.G46S prevents the stereospecific binding of l-phenylalanine to the dimer of human phenylalanine hydroxylase regulatory domain. *FEBS Open Bio* **7,** 195–203 CrossRef Medline
31. Oleinikovas, V., Saladino, G., Cossins, B. P., and Gervasio, F. L. (2016) Understanding cryptic pocket formation in protein targets by enhanced sampling simulations. *J. Am. Chem. Soc.* **138,** 14257–14263 CrossRef Medline

32. Gjetting, T., Petersen, M., Guldberg, P., and Güttler, F. (2001) Missense mutations in the N-terminal domain of human phenylalanine hydroxylase interfere with binding of regulatory phenylalanine. *Am. J. Hum. Genet.* **68,** 1353–1360 CrossRef Medline
33. Krebs, W. G., and Gerstein, M. (2000) The morph server: a standardized system for analyzing and visualizing macromolecular motions in a database framework. *Nucleic Acids Res.* **28,** 1665–1675 CrossRef Medline
34. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 CrossRef Medline
35. Case, D. A., Cerutti, D. S., Cheatham, T. E., III, Darden, T. A., Duke, R. E., Giese, T. J., Gohlke, H., Goetz, A. W., Greene, D., Homeyer, N., Izadi, S., Kovalenko, A., Lee, T. S., LeGrand, S., Li, P., *et al.* (2017) *AMBER 2017*, University of California, San Francisco
36. Jakalian, A., Jack, D. B., and Bayly, C. I. (2002) Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. parameterization and validation. *J. Comput. Chem.* **23,** 1623–1641 CrossRef Medline
37. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004) Development and testing of a general Amber force field. *J. Comput. Chem.* **25,** 1157–1174 CrossRef Medline
38. Sousa da Silva, A. W., and Vranken, W. F. (2012) ACPYPE—AnteChamber PYthon Parser interfacE. *BMC Res. Notes* **5,** 367 CrossRef Medline
39. Abraham, M. J., Murtola, T., Schulz, R., Pall, S., Smith, J. C., Hess, B., and Lindahl, E. (2015) GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 CrossRef
40. Shirts, M., and Pande, V. S. (2000) Screen savers of the world unite! *Science* **290,** 1903–1904 CrossRef Medline
41. Li, D.-W., and Brüschweiler, R. (2010) NMR-based protein potentials. *Angew. Chem. Int. Ed. Engl.* **49,** 6778–6780 CrossRef Medline
42. Hess, B. (2008) P-LINCS: a parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput.* **4,** 116–122 CrossRef Medline
43. Schwantes, C. R., and Pande, V. S. (2013) Improvements in Markov state model construction reveal many non-native interactions in the folding of NTL9. *J. Chem. Theory Comput.* **9,** 2000–2009 CrossRef Medline
44. Pérez-Hernández, G., Paul, F., Giorgino, T., De Fabritiis, G., and Noé, F. (2013) Identification of slow molecular order parameters for Markov model construction. *J. Chem. Phys.* **139,** 015102 CrossRef Medline
45. Naritomi, Y., and Fuchigami, S. (2013) Slow dynamics of a protein backbone in molecular dynamics simulation revealed by time-structure based independent component analysis. *J. Chem. Phys.* **139,** 215102–215111 CrossRef Medline
46. Harrigan, M. P., Sultan, M. M., Hernández, C. X., Husic, B. E., Eastman, P., Schwantes, C. R., Beauchamp, K. A., McGibbon, R. T., and Pande, V. S. (2017) MSMBuilder: statistical models for biomolecular dynamics. *Biophys. J.* **112,** 10–15 CrossRef Medline
47. McGibbon, R. T., and Pande, V. S. (2015) Variational cross-validation of slow dynamical modes in molecular kinetics. *J. Chem. Phys.* **142,** 124105 CrossRef Medline
48. Noé, F., Schütte, C., Vanden-Eijnden, E., Reich, L., and Weikl, T. R. (2009) Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 19011–19016 CrossRef Medline
49. Metzner, P., Schütte, C., and Vanden-Eijnden, E. (2009) Transition path theory for Markov jump processes. *Multiscale Model. Simul.* **7,** 1192–1219 CrossRef
50. Daubner, S. C., Hillas, P. J., and Fitzpatrick, P. F. (1997) Characterization of chimeric pterin-dependent hydroxylases: contributions of the regulatory domains of tyrosine and phenylalanine hydroxylase to substrate specificity. *Biochemistry* **36,** 11574–11582 CrossRef Medline
51. Weeks, S. D., Drinker, M., and Loll, P. J. (2007) Ligation independent cloning vectors for expression of SUMO fusions. *Protein Expr. Purif.* **53,** 40–50 CrossRef Medline
52. Shiman, R., Gray, D. W., and Pater, A. (1979) A simple purification of phenylalanine hydroxylase by substrate-induced hydrophobic chromatography. *J. Biol. Chem.* **254,** 11300–11306 Medline