# Oxygen Pathways and Allostery in Monomeric Sarcosine Oxidase via Single-Sweep Free-Energy Reconstruction

Anthony Bucci and Cameron F. Abrams*

Dept. Chemical and Biological Engineering, Drexel University, Philadelphia, Pennsylvania 19104, United States

S *Supporting Information*

ABSTRACT: Monomeric sarcosine oxidase (MSOX) is a flavoprotein D-amino acid oxidase with reported sarcosine and oxygen activation sites on the *re* and *si* faces of the flavin ring, respectively. O_2 transport routes to the catalytic interior are not well understood and are difficult to ascertain solely from MSOX crystal structures. A composite free-energy method known as single-sweep is used to map and thermodynamically characterize oxygen sites and routes leading to the catalytically active Lys265 from the protein surface. The result is a network of pathways and free energies within MSOX illustrating that oxygen can access two free-energy minima on the *re* face of the reduced flavin from four separate solvent portals. No such minimum is observed on the *si* face. The pathways are geometrically similar for three major states of the enzyme: (1) apo with a closed flavin cleft, (2) apo with an open flavin cleft, and (3) inhibitor-bound with a closed flavin cleft. Interestingly, free energies along these transport pathways display significantly deeper minima when the substrate-mimicking inhibitor 2-furoic acid is bound at the sarcosine site, even at locations far from this site. This suggests a substrate-dependent allosteric modulation of the kinetics of O_2 transport from the solvent to the active site.
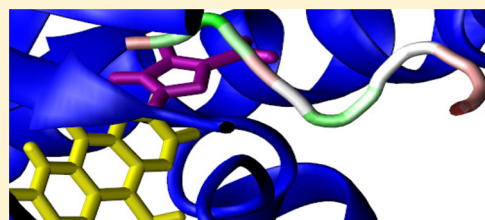
1. INTRODUCTION

Reduction of molecular oxygen is essential to aerobic biology. Flavoprotein D-amino acid oxidases reduce O_2 to H_2O_2 in order to regenerate the substrate-reactive oxidized state of their covalently bound flavin cofactors. Flavin oxidation by O_2 reduction is slow in free solution because of the requirement for an intermediate step in what would otherwise be a spin-forbidden two-electron transition.[1] Flavoprotein oxidases can speed this reaction up several orders of magnitude, although there does not appear to be a consensus mechanism.[2] Nevertheless, detailed structure-based studies suggest the importance of (a) the protein environment near the flavin for possible stabilization of intermediates and (b) the protein structure in channeling O_2 from solution to the flavin site.[3]

Probably the best-characterized flavoprotein oxidase in terms of structure and activity is monomeric sarcosine oxidase (MSOX),[4−11] a bacterial enzyme which catalyzes the oxidation of sarcosine (*N*-methylglycine) to produce formaldehyde, glycine, and hydrogen peroxide:

$$CH_3NHCH_2COOH + O_2 + H_2O$$

$$\rightarrow HCHO + NH_2CH_2COOH + H_2O_2$$
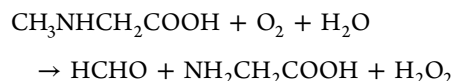
MSOX is a 44 kDa two-domain protein consisting of 385 residues containing a single covalently bound flavin adenine dinucleotide (8α-S-cysteinyl-FAD) whose flavin ring sits at the interface between the so-called catalytic and flavin domains. The flavin domain contains a cleft that admits and shelters the cofactor from solution.

A major question regarding the O_2 reactivity of MSOX (and all flavoprotein oxidases) is the location of O_2 activation relative

to the flavin ring and surrounding side chains. The prevailing theory is that the two-electron reduction of O_2 passes through a superoxide anion intermediate. This hypothesis has gained support because there are positively charged side chains near the flavin isoalloxazine ring.[12] X-ray crystal structures show the Lys265 ε-amino oriented over the *si* face of the flavin ring with a bridging water linking it to the flavin ring N(5).[4] Lys265 is indeed crucial for O_2 reduction: mutations to Ala, Gln, or Met at position 265 result in an 8000-fold decrease in oxygen reactivity, and mutation to Arg results in a 250-fold decrease.[8,10] It has therefore been proposed that sarcosine oxidation and oxygen reduction may occur on the *re* and *si* faces of the flavin ring, respectively,[10,11] which is consistent with the so-called modified ping-pong mechanism wherein O_2 can oxidize FADH⁻ before the product imine is released.[5] However, one potential complication with this interpretation is that it is based on structures containing the oxidized flavin, while it is the reduced anionic form, with a protonated N(5) and a negative charge on N(1), with which O_2 must react. It is not clear that precise crystallographic arrangement of the Lys265 side chain and associated waters is retained with reduced flavin, which leaves unanswered the question as to whether or not this arrangement is a result of flavin oxidation rather than its cause.

A second major question regards how O_2 accesses the flavin cavity from the solvent. In the absence of bound substrate, there is clearly room for a single O_2 to reach the *re* face of the

---

flavin ring based on the crystal structure. However, the modified ping-pong mechanism suggests that O_2 may access the flavin while substrate is bound. The addition of a substrate blocks the channel observed in the static crystal structure, suggesting O_2 accesses the cavity through a separate channel. No such channels are readily apparent in the MSOX crystal structures. Given that small dissolved gases access protein interiors through relatively small channels (e.g., the histidine gate in myoglobin[13,14]), it is plausible that enzymes which utilize gas molecules as reactants may also exhibit such channels. Understanding these channels would provide important new structural insight into the biochemistry of flavoprotein oxidases and may even provide clues to the location of the O_2 activation site.

Both of these questions can be addressed by querying the statistical likelihood of observing an O_2 molecule at any position inside the protein. In similar settings, this has been approached using analysis of static crystal structures and all-atom molecular dynamics (MD) simulations, as well as enhanced sampling variants thereof. In all such efforts, an accurate picture of the accessible volume in the protein interior is limited by inadequate sampling of protein conformations. This is a severe problem in sampling static structures[15,16] and even in brute-force molecular dynamics (MD),[17,18] especially if one needs enough statistics to map out the entire protein interior. Therefore, enhanced-sampling methods remain an imporant tool for this kind of mapping. For instance, Elber applied the time-dependent Hartree approximation to study CO diffusion in myoglobin.[19] Ceccarelli et al. used the metadynamics method[20] to study the same system.[21] Saam et al. applied implicit ligand sampling (ILS) to study dynamic oxygen access channels in 12/15-lipoxygenase.[22] Recently, a composite method combining temperature-accelerated MD,[23] the single-sweep method of free-energy reconstruction,[24] and the string method in collective variables[25] was used to fully map diffusion pathways and adsorption sites for CO[26] and water[27] in the interior of myoglobin, including thermodynamic characterization of the pathways.

The composite approach of Maragliano et al. is well-suited to studying biological systems involving small molecule diffusion (O_2, CO, CO_2, etc.). It is applied here on MSOX to map O_2 sites and pathways to the substrate activation site. In particular, we base our calculations on both substrate-free and inhibitor-bound MSOX with reduced flavin. Free energy as a function of the Cartesian position of the O_2 center of mass in a protein-fixed reference frame is determined. The locations of regions of low free energy are then identified and the string method applied on this free-energy surface to define pathways of minimal free energy connecting these regions. The result is a network of putative O_2 transport pathways and sites within MSOX. Our results show that O_2 can access the active site through several channels, some of which are unexpected. We observe deep free energy minima for O_2 on the *re* face of reduced flavin, even with the substrate-mimicking inhibitor 2-furoic acid bound, and none on the *si* face. We also observed that the locations of the minimum-free energy pathways and low free-energy sites are not strongly sensitive to whether or not the substrate mimic is bound. The free energy profiles along the paths, however, do depend sensitively on the presence of the substrate mimic: specifically the occurrence of several deep minima along O_2 pathways correlates with the inhibitor-bound state. This last finding hints at a manifestation of allostery wherein ligand binding alters the transport pathways through a protein, thereby modulating the rate at which a second substrate accesses the active site.

2. METHODS

2.1. System Description. The composite method developed by Maragliano et al. in their study of CO in myoglobin[26] was employed here to study O_2 in MSOX. The primary technique in this method is all-atom molecular dynamics (MD) underlying a particular series of free-energy calculations. Before explaining those, a discussion of how the all-atom systems of inhibitor-bound and apo MSOX were prepared is necessary. Heavy-atom coordinates for MSOX were taken from the 2gf3 PDB entry.[28] In addition to the protein, this structure contains several crystal waters and the inhibitor 2-furoic acid (FOA), the latter of which mimics how the substrate sarcosine binds. The FOA-bound MD system was generated by adding hydrogens to the 2gf3 coordinates where needed, solvating in TIP3P water,[29] neutralizing with Na^+ ions, to generate a box of 33 226 atoms with dimensions $83.8 \times 64.1 \times 67.6$ Å^3. This initial system was subject to 1000 steps of conjugate gradient minimization followed by 130 ns of NPT MD equilibration.

The apo system was created by deletion of the FOA from this minimized system and subjecting the resulting system to 130 ns of NPT equilibration. A molecule of O_2 was added by mutating one of the crystal waters coordinating Lys265. This system consisted of 33 221 atoms in a box $83.7 \times 64.0 \times 67.5$ Å. We did not observe spontaneous opening of the so-called active site loop (residues 55−60) in the apo equilibration, and all calculations reported here are with this loop in the closed position. It may exist in either open or closed configurations for apo MSOX but only closed for ligand-bound MSOX.[10] However, spontaneous opening of the FAD cleft was observed in the apo equilibration, distinguished by the breakage of a contact between Asn41 and Arg282. The distribution of $C\alpha-C\alpha$ distances for this residue pair is centered at 10 Å when the cleft is open and 8 Å when closed. To assess whether or not this transition altered the ability of O_2 to access the flavin cavity, both apo states are considered in subsequent calculations.

Molecular dynamics simulations employed periodic boundary conditions, a nonbonded cutoff of 10 Å, a particle-mesh Ewald spacing of 2 Å, rigid bonds, and a time step of 2 fs. The temperature was held at 310 K using a standard Langevin thermostat and 1 bar using a Langevin−Nosé−Hoover barostat.[30] Finally, because the subsequent free-energy calculations are performed in a protein-fixed coordinate frame, weak positional/rotational restraints were applied. These consisisted of Cartesian harmonic restraints on the C_α's of residues 25, 100, and 370 with a common spring constant of 1 kcal/mol·Å^2. All simulations were conducted with NAMD v. 2.9[31] using the CHARMM force field.[32,33] We mention in particular that we use the standard neutral CHARMM force-field model for oxygen which cannot model electronic stacking interactions with residues such as PHE. Such interactions can have an influence on O_2 diffusion in other proteins,[34] and we expect that using methods such as PELE[35] which supplement nonbonded interactions with periodic QM/MM calculations could be of some benefit in future refinement of the results presented in this work.

Protonation of the flavin was made consistent with its reduced form, FADH$^-$. CHARMM-style parameters for the adenine and sugar portions of FADH$^-$ were adapted from existing parameters for NADH. The flavin ring was para-

metrized using the AMBER antechamber procedure with parameters, including charges determined independently using geometry optimization at the B3LYP 6-311G* level using Gaussian, translated into CHARMM-style units. FOA was similarly parametrized. The parameter sets for FADH$^-$ and FOA were not further optimized. CHARMM-style topology and parameter files for FADH$^-$ and FOA are available in the Supporting Information.

2.2. Sampling the Protein Interior Accessible to O$_2$ via Temperature-Acceleration. Generally, temperature-accelerated MD (TAMD) accelerates the sampling of collective variables (CV's) $\theta(x)$ in an MD simulation by tethering them to fictitiously hot auxiliary variables z with high friction $\overline{\gamma}$, such that the forces these variables experience approximate negative free-energy gradients on the free-energy surface of the CVs:[23]

$$\overline{\gamma}\dot{z}_j = \kappa[\theta_j(x) - z_j] + \sqrt{2\overline{\beta}^{-1}\overline{\gamma}}\,\eta_j^z(t) \tag{1}$$

Here, κ is a spring-constant-like parameter, $\overline{\gamma}$ the artificial friction coefficient, $\overline{\beta}$ the inverse of the artificial temperature ($\beta \equiv 1/k_{\mathrm{B}}T$, where k_{B} is Boltzmann's constant), and $\eta_j^z(t)$ is white noise with unit variance.

In the approach used by Maragliano et al. for CO in myoglobin, the CVs accelerated were the Cartesian coordinates of the CO center of mass.[26] Analogously, here the Cartesian coordinates of the O$_2$ center of mass are accelerated. Each TAMD simulation, or "sweep," used the same conditions as did the MD equilibrations, with the additional TAMD parameters of $\kappa = 200$ kcal/mol·Å and $\overline{\gamma} = 5$ ps^{-1}.

The goal of TAMD is to sample as much of the O$_2$-accessible volume in the protein as possible. Since three systems were studied, the following naming convention was adopted. Inhibitor bound MSOX is referred to as "bound MSOX," while apo MSOX with the Asn41−Arg282 bridge open or closed is "apo-OB MSOX" and "apo-CB MSOX," respectively. Initial sweeps for bound and apo-CB MSOX were run at ficitious temperatures of $2 \leq \overline{\beta}^{-1} \leq 7$ kcal/mol. Three sweeps were run for each discrete fictitious temperature at both the *re* face near Lys265 and below the *si* face of the flavin ring. Productive sampling was observed between 3 and 5 kcal/mol (see Results and Discussion).

To achieve consistent protein interior volume sampling across the three systems, 1 ns sweeps were initialized from four separate sites within MSOX, two in each domain. Initial production sweeps were run from two sites in the catalytic domain, the *re* face near Lys265 and below the *si* face of the flavin ring. Sites were required in the FAD binding domain because regions were accessible to only apo or bound sweeps. Therefore, additional production sweeps were launched from near Leu212 and C(5)′ on FADH but not in the cleft. For any site, a water was mutated into oxygen from the equilibrated starting structure. A total of 36 sweeps were required for apo-CB MSOX. Fifty-one sweeps were required for apo-OB, and 64 for bound MSOX. A total of 151 ns of production sweeps were required to achieve consistent interior volume sampling among all three systems.

The primary output of production sweeps is a dense set of O$_2$ positions both inside and outside the protein. These sets were culled by beginning with an interior location and including any location not closer than 2.5 Å to any already-included location. Each harvested location is referred to as a "center," and the kth center is indexed as z_k. A total of 262

centers were harvested for apo-CB MSOX, 355 for apo-OB MSOX, and 416 for bound MSOX.

2.3. Mean-Force Calculations and Free-Energy Reconstruction. The main idea of single-sweep is to use mean forces computed on a small number of important locations as the basis for reconstruction of a complete analytical free energy as a function of O$_2$ position in the protein. Mean-force calculations for each center proceeded as follows. The MD system for the center is run under the TAMD protocol with the protein restraints active and the fictitious friction effectively infinite, so the auxiliary variables do not evolve. The mean force vector f at center position z_k is computed as the following time-average on the atomic trajectory $x(t)$ as

$$f(z_k) \equiv f_k = \frac{1}{T}\sum_{j=0}^{T}\kappa[\theta(x(t_j)) - z_k] \tag{2}$$

where T is the number of time increments in the trajectory and θ is the collective variable (instantaneous O$_2$ center of mass). Saturation of mean forces was observed to occur in less than 5 ns of MD integration using $\delta t = 1$ fs time-steps.

The reconstructed free energy $\tilde{A}(z)$ is represented analytically as a radial-basis function expansion:

$$\tilde{A}(z) = \sum_{k=1}^{K} a_k \varphi_\sigma(|z - z_k|) + C \tag{3}$$

where φ_σ is a Gaussian with width σ, a_k is the kth coefficient in the basis-function expansion, and C is an irrelevant constant that adjusts the overall height of $\tilde{A}(z)$. Via standard linear-algebra methods, the fitting parameters a_k and σ are determined by minimizing an error function given by

$$E(a, \sigma) = \sum_{k=1}^{K}\left|\sum_{k'=1}^{K} a_{k'}\nabla_z\varphi_\sigma(|z_k - z_{k'}|) + f_k\right|^2 \tag{4}$$

The optimal basis function width (σ) was 2.5 Å with a relative residual error of approximately 0.66 kcal/mol/Å for all three configurations.

2.4. Mapping the Reconstructed Free Energy Surface. Local minima on each FES are found using multiple-walker steepest-descent minimization. From the final walker locations, a set of local minima is identified. To find pathways of minimum free energy between any two of these minima, the zero-temperature string method is used.[36] Briefly, for two minima A and B, a line segment connecting them is discretized into N sites, each of which is a walker that is allowed to move according to the local gradient in $\tilde{A}$ subject to a reparameterization step that keeps the site−site separation distance along the string uniform. The string of sites thus "falls" into a minimum free-energy path (MFEP) connecting the two minima. String convergence for this investigation was achieved when the change in free energy between successive calculations was approximately 10^{-5} kcal/mol. We typically used $N = 50$ discretization points on our strings.

3. RESULTS AND DISCUSSION

3.1. Equilibration and Sampling. Figure 1 shows root-mean squared deviation (RMSD) traces for each MD equilibration, indicating that both systems equilibrate after about 10 ns. The apo system undergoes slightly larger fluctuations than the inhibitor-bound system, which may stem from the fact that the inhibitor's interactions with the protein
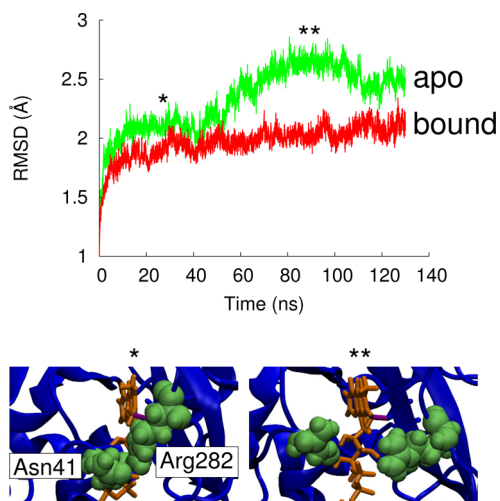
Figure 1. (upper) C_α root-mean squared deviation (RMSD) from MD simulations of apo and FOA-bound MSOX. (lower) Images from the apo simulations showing detail of the interaction between Asn41 and Arg282; here, "*" refers to the "closed-bridge" state and "**" to the "open-bridge" state.

suppress certain fluctuations. As can be seen in the apo RMSD trace, a jump in RMSD occurs at about 50 ns. This corresponds to the spontaneous opening of the Asn41−Arg282 bridge of the flavin cleft, as illustrated by the renderings in the lower panels of Figure 1.

Under normal MD, O_2 sampling of the protein interior is severely limited on accessible computational time scales. Figure 2 depicts sampling of O_2 locations from 1 ns MD and 1 ns
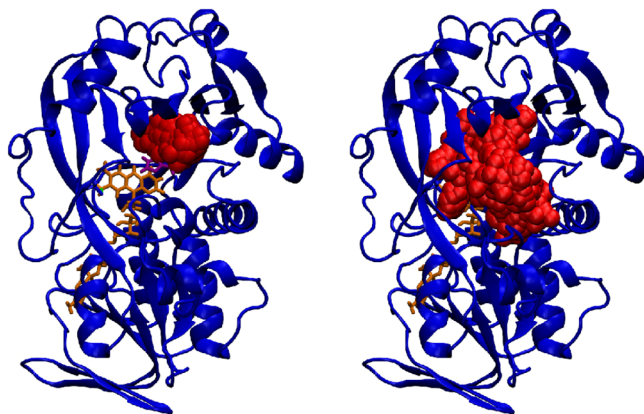
Figure 2. Cartoon structures of bound MSOX showing the location of sampled O_2 positions in red. (Left) 1 ns MD trajectory and (right) 1 ns TAMD trajectory with $\bar{\beta}^{-1}$ = 3 kcal/mol.

TAMD simulations on bound MSOX. Both were initialized from the same O_2 location. The TAMD simulation was run with $\bar{\beta}^{-1}$ = 3 kcal/mol. This illustrates that TAMD provides much more extensive sampling at a fraction of the computational cost of standard MD. The complete sampling provided by TAMD in both systems is shown in Figure 3, which depicts the locations of all centers harvested from the TAMD sweeps for all three systems. Though the number of centers varies greatly, the vast majority of additional centers occupy the solvent outside the protein but within the 10 Å cutoff distance for interatomic interactions. As will be shown later, these centers do not impact the pathways within MSOX. They are

included to smooth the PMF outside the protein. Note that the volume spanned by the set of centers includes many important residues, including the active site loop (residues 55−60), Lys265, Phe256, Arg49, Cys315, and both the *re* and *si* faces of the flavin ring.[8−10,12,37,38]

3.2. Mean Force Estimation and Free Energy Reconstruction. Figure 4 shows running average forces [eq 2] in the three Cartesian directions for a representative center. Most of the 1033 individual mean-force calculations required about 1 ns to achieve saturation of the forces. The calculations that required more than 1 ns saturated within 5 ns.

The deepest minima observed were on the order of −10 kcal/mol, relative to O_2 in solvent. The deepest minima identified were observed to occur at the same locations for all three systems. Fifty-one unique minima were identified for the apo-CB system, 50 for apo-OB, and 96 for bound MSOX. The bound system was observed to display many more shallow local minima which, as will be shown, had very little effect on the free-energy profiles along the pathways of minimal free energy.

Each of the three systems displayed the same two distinct minima on the *re* face of the flavin ring. One occurs near flavin atom C(6) and the other near N(3) (see Figure 5). These two minima are features of a larger hydrophobic basin bounded by the *re* face. Lysine residues near the flavin ring have been shown to be important catalytically not only in MSOX but in histone demethylases LSD1 and LSD2 as well as other systems.[8,18,39,40] Lys265 is directly above C(6), and a local minimum nearby further supports experimental observations. However, both minima are located on the *re* face of the ring, suggesting that at least the initial interaction of neutral molecular oxygen with the isoalloxazine ring is a metastable encounter complex with the O_2 on the *re* face. It may be that the first electron transfer event creates a superoxide anion and the one-electron reduced flavin radical such that the superoxide transits to the end of the Lys265 side-chain to the *si* face, but this would require relative motion of the flavin and Arg49 to make room. The limitation of only considering molecular oxygen in our simulations means we cannot directly address the hypothesis that activation and sarcosine oxidation may occur on opposite faces of the flavin ring, as suggested by Zhao et al.[8] The minimum observed by N(3) on the flavin ring is close to the inhibitor and not shallow, approximately 6−7 kcal/mol. On the basis of the location of the substrate-mimicking inhibitor and the presence of an oxygen minima there in the apo systems, this location may be the sarcosine activation site, consistent with experimental results.[4]

3.3. O_2 Pathways From the Protein Surface to the Active Site. For each minimum near the surface, the string method located an MFEP to the active site. The active site is defined as the minimum near N(3) on the *re* face of the flavin ring. It is approximately where the substrate-mimicking inhibitor lies. This resulted in four distinct pathways, designated I−IV. With the exception of pathway I, each displays a high degree of geometrical similarity among the three systems. System-specific pathways are referred to using prefixes "apo-CB," "apo-OB," or "bound." The pathways are shown as tubes in Figure 6.

Pathway I enters MSOX between the loop connecting βF2 with αC1 (residues 32−59) and the loop connecting βC7 with αC4 (residues 268−289). These loops form part of the cleft that admits FADH into MSOX and are seen to open and close in a gate-like fashion in the standard MD simulations. Inside MSOX, pathway I passes next to the catalytically active Lys265.
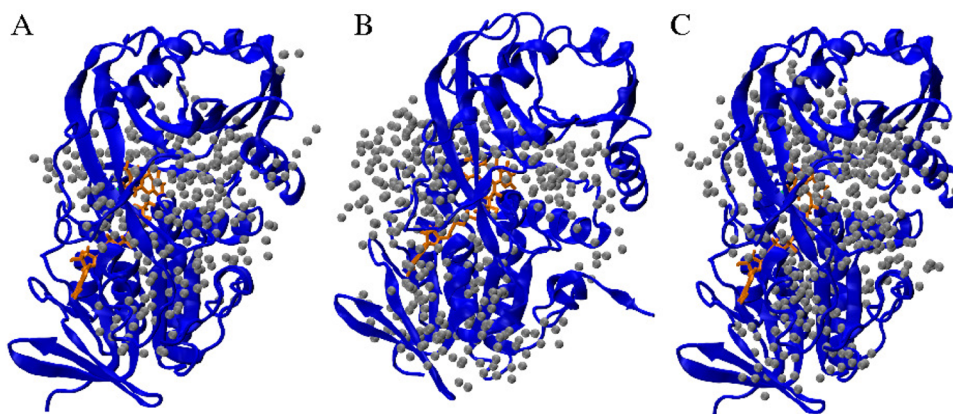
Figure 3. Cartoon structures of MSOX showing the location of centers harvested from (A) apo-CB MSOX, (B) apo-OB MSOX, and (C) bound TAMD sweeps.
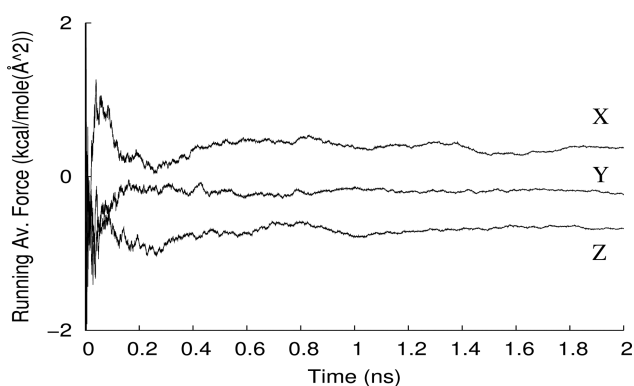

Figure 4. Running average forces for one representative center.

Both apo-CB-I and apo-OB-I enter MSOX this way. Apo-CB-I then passes over the *si* face of the flavin ring and directly under Lys265 when viewing the flavin's *si* face before terminating at the active site. It must be stressed that while apo-CB-I passes through what would be the oxygen activation site reported in the literature,[8] no free energy minimum is observed there. Apo-OB-I goes over the Cys315 linkage before passing through both minima on the *re* face and terminating at the active site. The bound-I string, however, enters MSOX above the entrance loops. It then passes between Arg49 and Lys265, crossing above the *si* face and terminating at the active site. Consequently, bound-I passes the loop connecting βC(6) to βC7 (residues 256−264). While apo-OB-I passes through the two distinct minima on the *re* face of the flavin ring, apo-CB-I and bound-I avoid the minima near C(6).

Pathways apo-OB-II, apo-CB-II, and bound-II pass through the middle of the gate in pathway I between Asn42 and Arg282. Oxygen then travels along the sugar backbone of FADH, entering the binding pocket after passing between the *re* face of the flavin ring and the loop connecting βF9 with αF4 (residues 340−350). All three paths are geometrically nearly identical and pass through the two distinct minima identified on the *re* face.

Pathways apo-OB-III, apo-CB-III, and bound-III traverse the entire FAD binding domain and most of the catalytic domain to arrive at the active site. The path begins at an entry portal far from the site, bounded by two loops: the first loop connects
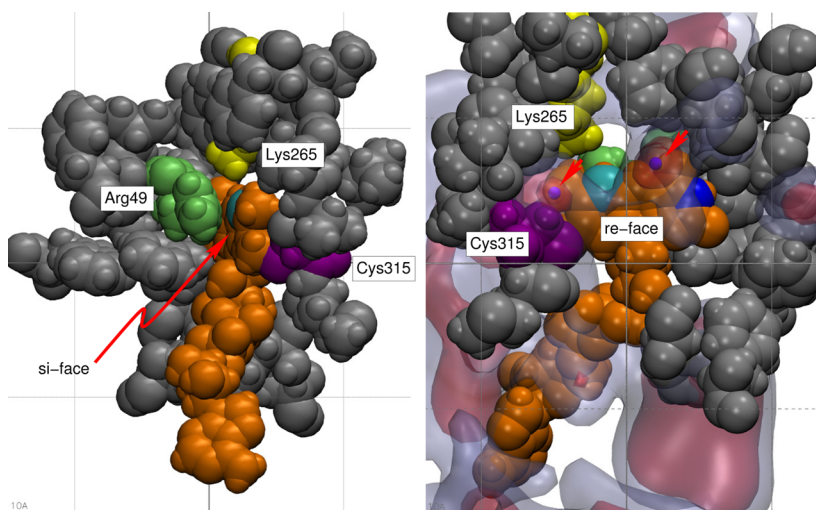
Figure 5. (left) Space-filling view of the flavin cofactor (orange) and all atoms in hydrophobic residues within 10 Å of the flavin atoms C(6) and N(3). The crowded *si* face has little to no room for molecular oxygen to interact with the isoalloxazine ring. (right) Rotated view of the same set of atoms as in the left panel, showing the *re* face of the isoalloxazine ring. Superimposed on this view are isosurfaces of the free energy at −5 kcal/mol (red) and −3 kcal/mol (light blue). Two local free-energy minima on the *re* face are indicated with red arrows. Flavin atoms C(6) is shown in cyan and N(3) in blue. A 10 Å grid is overlaid on both panels.
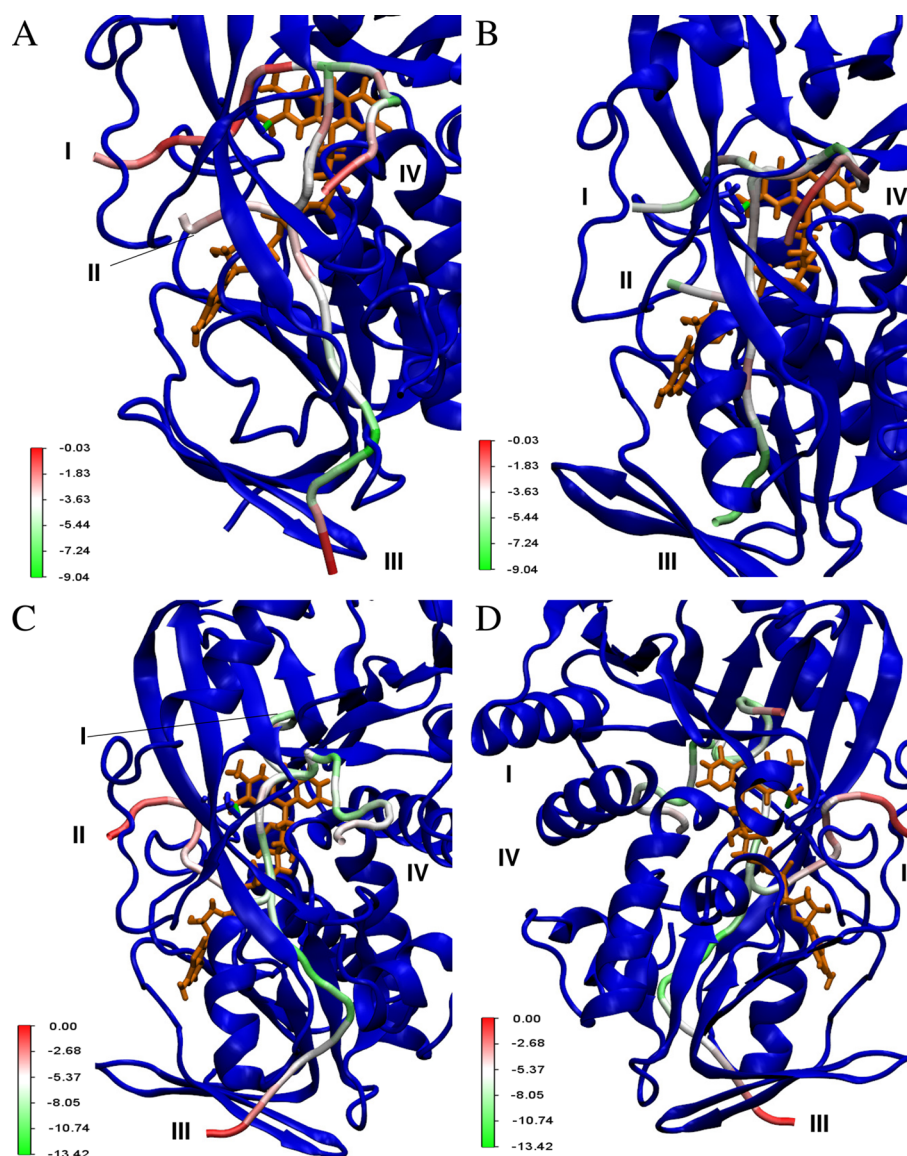
Figure 6. (A) Pathways apo-CB-I, -II, -III, and -IV overlaid on the MSOX structure. (B) Pathways apo-OB-I, -II, -III, and -IV (C) Pathways bound-I, -II, -III, and -IV. (D) Rotated view (180° along of bound pathways to provide a better view of pathways bound-II and -IV). Free energy along each pathway is indicated by the pathway color.

βF8 to βF9 (residues 330−335) and the second αF3 to βC4 (residues 212−219). The pathways continue past a loop connecting βF7 to αF3 (residues 199−205). Subsequently, oxygen travels along the sugar backbone of the FADH, converging with pathway II near the flavin ring. As with II, pathways apo-OB-III, apo-CB-III, and bound-III are geometrically nearly identical and pass through the two distinct minima identified on the *re* face.

Pathways apo-OB-IV, apo-CB-IV, and bound-IV enter MSOX via passage through the large opening identified as the sarcosine entryway.[4] All three pass between the active site loop (residues 55−60) and residues 268−272, part of the loop connecting βC7 with αC4, en route to the sarcosine activation site. Again, all type-IV paths are geometrically nearly identical across the three systems. However, since they enter through the substrate access channel and impinge directly on the minimum near N(6), pathway IV does not pass through the minimum near C(6).

None of the pathways pass through a minimum on the *si* face of the flavin ring before terminating at the sarcosine activation site. Lack of a minimum on the *si* face of the flavin ring is a concern because as we mentioned previously, it has been suggested that it is the site of oxygen activation.[8] The TAMD trajectories from which centers were harvested sample this region, and several centers at which mean forces were calculated were located on the *si* face, so if a minimum did exist it would be detected. To double-check whether such a minimum would occur, an additional center was added on the *si* face under Lys265, and its mean force was evaluated. The PMF was reconstructed again for the apo-CB system, and no additional minimum was observed. The bound system was also observed to contain no additional minimum. However, the minimum near C(6) on the *re* face of the flavin ring is conserved in all three cases. This suggests that the site may be important to catalysis and possibly be the O$_2$ activation site.

The area within 10 Å (the electrostatic cutoff distance) of C(6) encompasses both *re* face minima and exhibits interesting
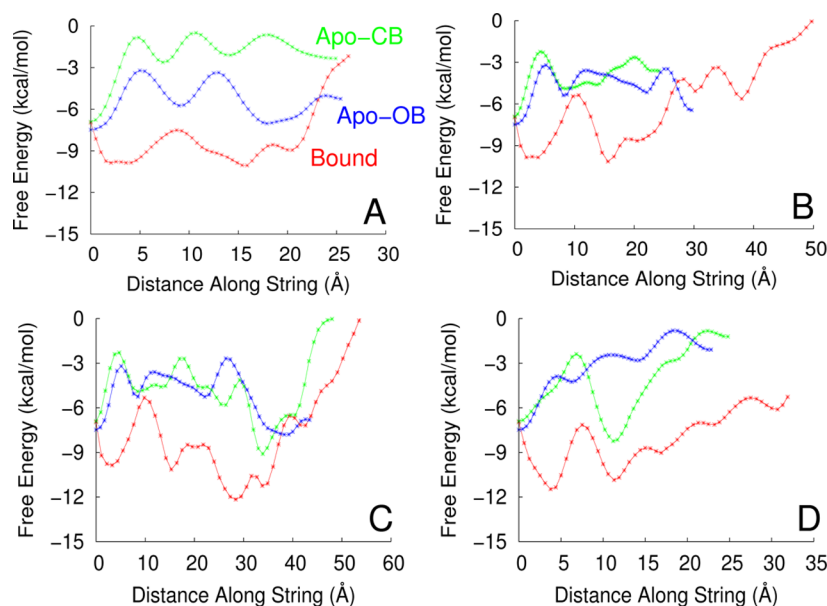
Figure 7. Free energy as a function of distance along each pathway from the activation site for pathways (A) apo-CB-I, apo-OB-I, and bound-I; (B) apo-CB-II, apo-OB-II, and bound-II; (C) apo-CB-III, apo-OB-III, and bound-III; and (D) apo-CB-IV, apo-OB-IV, and bound-IV. Apo-CB is shown in green, while apo-OB is shown in blue and bound in red. All graphs begin with oxygen in the active site and end at the nearest minimum to the protein surface where interactions with the solvent occur. Oxygen in solution has a reference value of 0 kcal/mol.

characteristics. Although the region around the flavin ring is reported as basic, basic residues primarily occur on the *si* face. The *re* face exhibits mostly nonpolar residues. The presence of multiple nonpolar residues is significant. Recent work has suggested that oxygen activation sites in flavoenzymes require a nonpolar residue nearby to aid in desolvation, optimize site geometry, and maximize electrostatic effects on molecular oxygen. It was therefore suggested that Phe256 would be the essential nonpolar residue owing to its proximity to Lys265.[12] Phe256 is well within the cutoff for interactions with oxygen located at the $C(6)$ minimum. The minima near $C(6)$ therefore exhibits nearly ideal conditions for oxygen activation: proximity to the catalytic Lys265, multiple nearby nonpolar residues to stabilize the site, and easily accessible by oxygen in all simulations.

3.4. Free Energy Profiles Along the O_2 Pathways and Allosterically Cryptic O_2 Sites. The free energy as a function of distance along each MFEP can be read directly from the reconstruction. Figure 7 shows the free energies as a function of distance from the sarcosine activation site along each pathway for all 12 pathways identified. Generally, pathways in the ligand-bound system are lower in free energy than their counterparts in the apo-CB or apo-OB systems. This generally inhibits transport to the activation site in the bound system as oxygen will tend to stay in the deepest minima. In contrast, with the ligand unbound, the deep local minima along the pathways attenuate somewhat, which would likely promote transport. It is not surprising that overall geometrical similarities in pathways exist due to structural similarities among the three systems. However, the large differences in free energies along these pathways between apo and ligand-bound MSOX are unexpected. Evidently, ligand binding and unbinding must cause subtle conformational changes which affect oxygen transport. A deeper analysis of the residues along each pathway reveals the bases for many of these differences.

Apo-CB-I and apo-OB-I appear to be the only MFEPs affected by cleft opening and closing. When closed, the flavin

ring shifts approximately 2 Å in the positive direction normal to the *re* face, helping to expose the *si* face. This is a feature conserved in bound-I. Additionally, residues 42−44 are oriented under the *si* face. Apo-OB-I in contrast has a *si* face that is partially blocked by residues 42−44. This slight shift of the flavin ring caused by cleft motion is likely the cause of both the free energy and geometric differences between apo-OB-I and apo-CB-I. It is interesting to note that when examining bound-I, there were MFEPs which caused it to be identical to bound-II. Since the bound system contains a closed bridge, a MFEP similar to apo-CB-I was sought. To have bound-I cross the *si* face required usage of a different starting point. However, it still passed Lys265 approximately where apo-CB-I did. Incidentally, bound-I then failed to pass through the minima near $C(6)$ on the *re* face. Ligand binding may therefore cause the route previously taken by apo-CB-I and apo-OB-I to be adversely affected due to crowding on the *re* face.

Access to and from pathways apo-II and bound-II is largely controlled by loops 268−289 and 32−45. Initial hypotheses suggested that opening and closing of the cleft would strongly influence the free energy near the end of the MFEP. However, examination of the free energy profiles indicates that opening and closing of the cleft does not influence the energetics significantly. The largest free energy differences are within the cutoff distance (10 Å) of the inhibitor. There are also no other major structural differences between the apo and bound states near the MFEPs. Thus, inhibitor binding plays a larger role in influencing energetics along pathway II than disposition of the cleft.

Access to apo-CB-III, apo-OB-III, and bound-III is controlled by loops 212−219, 330−335, and βF4. Apo-CB-III and apo-OB-III share similar energetics and structurally are nearly identical within the vicinity of the MFEP. When compared to bound-III, several subtle differences become apparent. One of the controlling loops, 212−219, rotates inward, apparently trapping oxygen. This may be the cause of the deep minima seen approximately 20 Å along the string. Ligand unbinding is

correlated with relaxation of the loop, opening the cavity allowing oxygen to escape. Rotation of this loop may not be caused by ligand binding, however, as it is within the cutoff for interactions with residues 375−385. Residues 375−385 include the C-terminus and are completely solvent exposed. They fluctuate greatly over the course of any simulation and interact with loop 212−219. It is more likely that these interactions are the primary cause of loop shifts resulting in deeper minima. Path III for all three systems is the longest of the four identified.

Path IV, which traverses the substrate entryway, bears three MFEPs which are nearly identical. For both apo and ligand-bound MSOX, the active site loop (residues 55−60) remains in the closed configuration for the duration of all simulations. When comparing apo-CB-IV and apo-OB-IV, the energetics are similar except when the MFEP is near Glu57. Apo-CB-IV features Glu57 pointed inward toward the flavin ring. This places it nearly in van der Waals contact with the MFEP at the deepest minima near 10 Å. Apo-OB-IV contains Glu57 pointed out toward the solvent. As such, it is further from the MFEP. This may not be the sole cause for the large discrepancy in free energy between the MFEPs, but it is the only major structural difference between the two in the region of the MFEP. When compared to bound-IV, there are several differences which must be noted. Tyr55 points away from the channel when bound, yet toward the channel in the apo systems. Glu57 is pointed toward the flavin ring similar to apo-CB-IV. This makes sense because the free energy near Glu57 for apo-CB-IV and bound-IV are similar in that region. It also reinforces the observation that orientation of Glu57 affects free energy along the path. The final major difference is binding of the inhibitor. The largest free energy differences occur within the cutoff distance (10 Å) for interactions with the inhibitor. Therefore, interactions with the inhibitor, as well as shifts in Tyr55 and Glu57, result in a path of significantly lower free energy.

The sensitivity observed in the O_2 pathway thermodynamics to ligand binding suggests a possible link to the observation that high product concentrations enhance flavoenzyme kinetics.[12] One interpretation is that positively charged products play the role of stabilizing the superoxide anion for certain flavoenzymes that, unlike MSOX, lack positively charged side-chains in the vicinity of the flavin ring. Our results hint at another possibility, at least for MSOX. When MSOX is substrate-bound, many deep local minima for O_2 along channels connecting solvent to the re face cavity help it absorb O_2 from solution. If they were static features of the protein structure, however, these minima would presumably be detrimental for the processing of O_2 at the flavin ring. We see that they in fact are not static but attenuate significantly in apo-MSOX, meaning that the channels should easily provide access for O_2 to the substrate/product-free flavin ring. A higher substrate concentration would lengthen the time MSOX spends in a substrate-bound state, sponging O_2 from the surrouding solution so that, once product release is initiated, a nearby O_2 can readily access the flavin ring and begin the next catalytic cycle. We are able to identify the structural shifts in backbone segments and side chains that accompany ligand binding to explain how the depths of these minima are allosterically modulated. It is worth pointing out that the ILS simulations of O_2 in 12/15-lipoxygenase by Saam et al.[22] show that binding the ligand arachidonic acid closes one O_2 diffusion channel but opens another. Our work on MSOX shows that ligand binding can act more subtly by altering pathway thermodynamics without completely shutting them off.

It is not clear from this work whether or not MSOX is special in displaying ligand-dependent "cryptic" allosteric sites for O_2. But the ability of proteins to display otherwise hidden secondary binding sites upon binding of a primary partner molecule is certainly gaining attention. For example, recent analysis of long all-atom MD simulations by Bowman and Geissler clearly illustrate the existence of cryptic sites for small molecules in a variety of proteins.[41] Our work suggests that the single-sweep approach is also a viable choice of methods for discovering such sites.

3.5. Concluding Remarks. The composite method of Maragliano et al.[26] has been used to map pathways and sites relevant for O_2 transport in MSOX as well as to gain a better understanding of the impact of substrate binding on these pathways. We observe two O_2 binding sites near the reduced flavin ring on the re face, suggesting that O_2 reduction is at least initiated on the same side of the flavin ring on which sarcosine oxidation occurs. Generally, we observed that multiple plausible pathways exist by which O_2 can access the sites in the re face cavity, regardless of whether or not the substrate-mimicking inhibitor 2-furoic acid is bound. These pathways are geometrically similar in the three different systems examined (apo, bound, and apo with an open flavin cleft) but show marked differences in their free energies between ligand-bound and unbound states. These differences manifest in the appearance of free-energy minima separate from the flavin cavity and the protein surface. We show that fluctuations in relative loop placements and side-chain orientations among the three systems explain the appearance of these cryptic minima. This may point to a mechanism of allostery by which transport of small molecule reactants is dependent upon large-molecule ligand binding similar to that seen in 12/15-lipoxygenase.[22] The presence of deep minima along each pathway that attenuate with ligand unbinding also suggests a sort of organic "pump" that ensures O_2 is able to reach the substrate activation site in MSOX rapidly to begin a new catalytic cycle. Nevertheless, the identification of important residues and loops in the models presented should serve to guide experiments aimed at elucidating small-molecule transport pathways in MSOX. A detailed investigation into transport rates along each pathway is the next step and can be determined using techniques such as milestoning.[42] We have shown that the composite method of Maragliano et al.[26] has been easily adapted to study oxygen diffusion pathways in the flavoenzyme MSOX. It has also been used to study myoglobin, characterizing multiple pathways for CO diffusion. This composite method may be adapted to study many biological systems, including other flavoenzyme oxidases (GOX, TSOX, HSOX, etc.) as well as larger systems such as hemoglobin.

■ **ASSOCIATED CONTENT**

Ⓢ Supporting Information

The file fadh_foa_charmm.zip is a zipped folder containing CHARMM-style topology and parameter files for furoic acid (FOA) and the reduced flavin adenine dinucleotide anion (FADH$^-$).. This material is available free of charge via the Internet at http://pubs.acs.org.

■ **AUTHOR INFORMATION**

Corresponding Author
*E-mail: cfa22@drexel.edu.

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Massey, V. *J. Biol. Chem.* **1994**, *269*, 22459−22462.
(2) Fitzpatrick, P. F. *Arch. Biochem. Biophys.* **2010**, *493*, 13−25.
(3) Mattevi, A. *TRENDS Biochem. Sci.* **2006**, *31*, 276−283.
(4) Trickey, P.; Wagner, M. A.; Jorns, M. S.; Mathews, F. S. *Structure* **1999**, *7*, 331−345.
(5) Wagner, M. A.; Trickey, P.; Chen, Z.-W.; Mathews, F. S.; Jorns, M. S. *Biochemistry* **2000**, *39*, 8813−8824.
(6) Wagner, M. A.; Jorns, M. S. *Biochemistry* **2000**, *39*, 8825−8829.
(7) Zhao, G.; Jorns, M. S. *Biochemistry* **2002**, *41*, 9747−9750.
(8) Zhao, G.; Bruckner, R. C.; Jorns, M. S. *Biochemistry* **2008**, *47*, 9124−9135.
(9) Hassan-Abdallah, A.; Zhao, G.; Chen, Z.-w.; Mathews, F. S.; Schuman Jorns, M. *Biochemistry* **2008**, *47*, 2913−2922.
(10) Jorns, M. S.; Chen, Z.-w.; Mathews, F. S. *Biochemistry* **2010**, *49*, 3631−3639.
(11) Kommoju, P.-R.; Chen, Z.-w.; Bruckner, R. C.; Mathews, F. S.; Jorns, M. S. *Biochemistry* **2011**, *50*, 5521−5534.
(12) Gadda, G. *Biochemistry* **2012**, *51*, 2662−2669.
(13) Cohen, J.; Arkhipov, A.; Braun, R.; Schulten, K. *Biophys. J.* **2006**, *91*, 1844−1857.
(14) Ayana, T.; Tokushi, S.; Kouhei, I.; Shunsuke, N.; Hirohiko, I.; Matthieu, C.; Fumihiro, K.; Sam-Yong, P.; Takayuki, T.; Takahisa, Y.; Shin-ya, K.; Shin-ichi, A. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 2612−2616.
(15) Moriguchi, T.; Ida, K.; Hikima, T.; Ueno, G.; Yamamoto, M.; Suzuki, H. *J. Biochem.* **2010**, *148*, 491−505.
(16) Chovancova, E.; Pavelka, A.; Benes, P.; Strnad, O.; Brezovsky, J.; Kozlikova, B.; Gora, A.; Sustr, V.; Klvana, M.; Medek, P.; Biedermannova, L.; Sochor, J.; Damborsky, J. *PLoS Comput. Biol.* **2012**, *8*, No. e1002708.
(17) Ruscio, J. Z.; Kumar, D.; Shukla, M.; Prisant, M. G.; Murali, T. M.; Onufriev, A. V. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 9204−9209.
(18) Baron, R.; Binda, C.; Tortorici, M.; McCammon, J. A.; Mattevi, A. *Structure* **2011**, *19*, 212−220.
(19) Elber, R.; Karplus, M. *J. Am. Chem. Soc.* **1990**, *112*, 9161−9175.
(20) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562−12566.
(21) Ceccarelli, M.; Anedda, R.; Casu, M.; Ruggerone, P. *Proteins: Struct., Funct., Bioinf.* **2008**, *71*, 1231−1236.
(22) Saam, J.; Ivanov, I.; Walther, M.; Holzhütter, H.-G.; Kuhn, H. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 13319−13324.
(23) Maragliano, L.; Vanden-Eijnden, E. *Chem. Phys. Lett.* **2006**, *426*, 168−175.
(24) Maragliano, L.; Vanden-Eijnden, E. *J. Chem. Phys.* **2008**, *128*, No. 184110 .
(25) Maragliano, L.; Fischer, A.; Vanden-Eijnden, E.; Ciccotti, G. *J. Chem. Phys.* **2006**, *125*, No. 024106 .
(26) Maragliano, L.; Cottone, G.; Ciccotti, G.; Vanden-Eijnden, E. *J. Am. Chem. Soc.* **2010**, *132*, 1010−1017.
(27) Lapelosa, M.; Abrams, C. F. *J. Chem. Theory Comput.* **2013**, *9*, 1265−1271.
(28) Chen, Z.; Trickey, P.; Jorns, M.; Mathews, F. Structure of the complex of monomeric sarcosine with its substrate analogue inhibitor 2-furoic acid at 1.3 A resolution, NCBI Database.
(29) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1987**, *79*, 10.

(30) Simone, M.; Giovanni, C.; Holian; Lee, B. *Mol. Phys.* **1993**, *78*, 533−544.
(31) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781−1802.
(32) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Du nbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Guo, J. G.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D.; Thom, B. P.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, 3586−3616.
(33) MacKerell, A. D., Jr.; Banavali, N. K. *J. Comput. Chem.* **2000**, *21*, 105−120.
(34) Aitor, H.-O.; Fatima, L.; Patricia, F.; Milagros, M.; Victor, G. *J. Biol. Chem.* **2011**, *286*, 41105−41114.
(35) Borrelli, K. W.; Vitalis, A.; Alcantara, R.; Guallar, V. *J. Chem. Theory Comput.* **2005**, *1*, 1304−1311.
(36) Weinan, E.; Weiqing, R.; Vanden-Eijnden, E. *Phys. Rev. B* **2002**, *66*, 052301.
(37) Hassan-Abdallah, A.; Zhao, G.; Jorns, M. S. *Biochemistry* **2008**, *47*, 1136−1143.
(38) Hassan-Abdallah, A.; Zhao, G.; Jorns, M. S. *Biochemistry* **2006**, *45*, 9454−9462.
(39) Baron, R.; McCammon, J. A.; Mattevi, A. *Curr. Opin. Struct. Biol.* **2009**, *19*, 672−679.
(40) Baron, R.; Riley, C.; Chenprakhon, P.; Thotsaporn, K.; Winter, R. T.; Alfieri, A.; Forneris, F.; van Berkel, W. J. H.; Chaiyen, P.; Fraaije, M. W.; Mattevi, A.; McCammon, J. A. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 10603−10608.
(41) Bowman, G. R.; Geissler, P. L. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 11681−11686.
(42) Vanden-Eijnden, E.; Venturoli, M. *J. Chem. Phys.* **2009**, *130*, No. 194101.