

The Protein's Role in Substrate Positioning and Reactivity for Biosynthetic Enzyme Complexes: the Case of SyrB2/SyrB1

Rimsha Mehmood^{1,2}, Helena W. Qi^{1,2}, Adam H. Steeves¹, and Heather J. Kulik^{1,*}

¹*Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA*

02139

²*Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139*

ABSTRACT: Biosynthetic enzyme complexes selectively catalyze challenging chemical transformations, including alkane functionalization (e.g., halogenation of threonine, Thr, by non-heme iron SyrB2). However, the role of complex formation in enabling reactivity and guiding selectivity is poorly understood, owing to the challenges associated with obtaining detailed structural information of the dynamically associating protein complexes. Combining over 10 μ s of classical molecular dynamics of SyrB2 and the acyl carrier protein SyrB1 with large-scale QM/MM simulation, we investigate the substrate–protein and protein–protein dynamics that give rise to experimentally observed substrate positioning and reactivity trends. We confirm the presence of a hypothesized substrate-delivery channel in SyrB2 through free energy simulations that show channel opening with a low free energy barrier. We identify stabilizing interactions at the SyrB2/SyrB1 interface that are compatible with phosphopantetheine (PPant) delivery of substrate to SyrB2. By sampling metal–substrate distances observed in experimental spectroscopy of native SyrB2/SyrB1-PPant-*S*-Thr and non-native substrates, we characterize essential protein–substrate interactions that are responsible for substrate positioning, and thus, reactivity. We observe the hydroxyl sidechain and terminal amine of the native Thr substrate to form cooperative hydrogen bonds with a single N123 residue in SyrB2. In comparison, non-native substrates that lack the hydroxyl interact more flexibly with the protein and therefore can orient closer to the Fe center, explaining their preferential hydroxylation and higher turnover frequencies.

1. Introduction

Mononuclear non-heme iron enzymes¹⁻⁶ have attracted great interest for their ability to selectively functionalize otherwise unreactive C-H bonds, playing an essential role in biosynthesis pathways that produce clinically relevant natural products.⁷ As a result, the development of mimetic inorganic complexes of these enzymes has been an area of intense focus.⁸⁻¹² Non-heme iron enzymes differ from their mimetics by operating under a distinct catalytic paradigm in which the enzyme environment protects a deeply buried but reactive metal center and controls specificity in part through substrate delivery.¹⁻⁶ One such enzyme is SyrB2¹³⁻¹⁴, an Fe/ α -ketoglutarate (α KG)-dependent halogenase¹³⁻²⁰ from *Pseudomonas syringae* in which the Fe(II) center is coordinated by two protein His, one Cl, and α KG. Once activated²¹ by substrate delivery via a prosthetic phosphopantetheine (PPant) arm attached to the acyl carrier protein (ACP) SyrB1¹³, SyrB2 chlorinates the C4 position of a native L-Thr substrate. Halogenases like SyrB2 share structural and mechanistic commonalities with well-studied, non-heme iron hydroxylases^{15, 22}. In both hydroxylases and halogenases, an Fe(IV)=O intermediate forms, which carries out H atom abstraction followed either by hydroxylation or halogenation, respectively.^{1, 13-14, 21, 26-31} For hydroxylases a 2His/1(Asp/Glu) facial triad²³⁻²⁴ is essential for hydroxylation, whereas the Asp is replaced with an Ala in halogenases to enable Cl binding to the metal center. Although Ala to Asp mutation in SyrB2 does not lead to hydroxylation¹⁴, replacement of Cl with N donors has catalyzed C-N bond formation²⁵.

Although substrate selectivity is thus governed in part by the ability of SyrB1 to link the substrate to PPant and form a PPant-S-Res moiety, non-native substrates that differ from Thr by lacking its hydroxyl sidechain (i.e., aminobutyric acid or Aba) and/or containing a longer carbon chain (i.e., norvaline or Nva) have been successfully tethered to SyrB1/PPant³². Reactivity

studies³² on these substrates have revealed surprising reactivity patterns: H atom abstraction occurs up to 130x faster for non-native substrates, concomitant with an increase in rebound hydroxylation propensity (up to 100%) over native halogenation. Thus it appears that SyrB2 has a substrate-specific catalytic inefficiency for the native substrate that coincides with halogenation being favored over hydroxylation³². This unexpected behavior has sparked vigorous computational³³⁻³⁸ and spectroscopic³⁹⁻⁴⁰ study of SyrB2 to better understand its unexpected reactivity. One suggestion^{37, 40-41} has been that configurational isomerization of the active site, especially after formation of the oxo intermediate^{38, 41}, reorients Cl to an axial position to make it more accessible for halogenation. However, this mechanism still requires some substrate-specific protein interaction and thus positioning that would enable isomerization for one substrate (e.g., Thr) while disfavoring it for others (e.g., Nva, Aba).

Challenges remain in understanding the role that the greater protein and protein-protein complex must play in SyrB2/SyrB1 reactivity. The crystal structure of SyrB2¹⁴ was obtained with no obvious channel for substrate delivery, and SyrB1 has not been crystallized in isolation or in complex with SyrB2. Advances in structural characterization by crosslinking enzyme/ACP complexes⁴² and forming intermediate analogues⁴³, e.g., vanadyl⁴⁴⁻⁴⁵ intermediates, have provided insight through static snapshots into protein-protein interactions or active site rearrangements during the catalytic cycle of related enzymes. Computational modeling can provide complementary insight into transient interactions and in cases where structural characterization is incomplete. However, most modeling has been carried out on cluster models of SyrB2, with few exceptions^{36, 38}, despite the wider use of whole-protein multi-scale modeling for other non-heme iron enzymes⁴⁶⁻⁵⁰. To date, no SyrB2/SyrB1 complexes have been simulated, with the largest models using truncated^{34, 36-38} PPant-S-Thr models placed in SyrB2 by rotating or

mutating F196.

Shedding insight into structural features of the SyrB2/SyrB1 complex, Silakov and coworkers⁵¹ recently employed hyperfine sublevel correlation (HYSCORE)⁵² spectroscopy to estimate average positions and angles of PPant-loaded substrate methyl and ethyl carbons with respect to the iron center in the active site of the SyrB2/SyrB1 complex, similar to previous efforts that had provided insight into non-heme iron hydroxylases⁵³⁻⁵⁴. This analysis provided a spectroscopic basis for observed reactivity patterns³², previously only inferred³² or suggested from simulation³⁴ to be a consequence of positioning, by identifying a greater distance between Fe(II) and the native Thr substrate than non-native Nva or Aba. Although providing essential quantitative insight into substrate/active-site positioning, the HYSCORE models cannot be used to determine which protein–substrate interactions give rise to these observed positions nor which configurational isomer the active site samples when the distances are measured.

The role of the greater protein environment in conferring differences in substrate-specific reactivity thus remains unknown in SyrB2. In this work, we employ long-time classical molecular dynamics (MD) and large-scale quantum mechanics/molecular mechanics (QM/MM) simulation to reveal the nature of the protein–protein and protein–substrate interactions that confer reactivity and selectivity in the SyrB2/SyrB1 complex. These simulations provide detailed information about how differences between the orientation of the native Thr substrate and those of non-native Nva or Aba result from distinct interactions between Thr and active-site residues; these data represent the first guidance on protein-substrate interactions that govern positioning and reactivity in SyrB2 and reveal the enzyme-specific origins of divergent reactivity.

2. Results and Discussion

2a. Dynamics of Isolated SyrB2 and SyrB1 Proteins.

SyrB1, a member of the ACP family, contains two domains^{13, 55}, the adenylation, or A, domain and a thiolation, or T, domain. The T domain is responsible for delivering substrate to SyrB2 via a PPant arm that is connected by a phosphate group to a Ser residue in SyrB1 (Figure 1). Although SyrB1 has not been crystallized, ACP T domains are highly conserved, and homology modeling of the SyrB1 T domain based on a related ACP⁵⁶ afforded a stable structure as evaluated over 500 ns of dynamics (see Computational Details). In solution, the PPant-S-Thr arm on SyrB1 samples a very broad conformational ensemble, characterized by the radius of gyration, R_g , defined from the phosphate group to the terminal carbon of the Thr substrate, having a peak below 4 Å but a broad overall distribution from 3.5 to over 7.0 Å (Figure 1 and Supporting Information Figure S1). In the smallest R_g configurations, the substrate curls on itself, minimizing solvent exposure, whereas the largest R_g values correspond to an extended, nearly linear configuration of PPant (Figure 1). Analysis of all sampled PPant geometries reveals a few dominant configurations with 4–8 Å P–S separation that are similarly bent, as measured by the angle between vectors connecting P- and S-adjacent C atoms to a central C atom at the vertex, and a second highly sampled configuration with a nearly linear C–C–C angle and a 14–15 Å P–S distance (Supporting Information Figure S2 and Table S1). The observations on the MD ensemble suggest that PPant frequently samples bent, compact configurations but will extend with a small barrier (ca. 1.5 kcal/mol) to form a configuration slightly higher in free energy (ca. 0.5 kcal/mol) that is compatible with docking to SyrB2. This analysis is consistent with the modest fraction (ca. 20%) of the NMR ensemble⁵⁶ of the ACP on which SyrB1 was homology modeled in which the PPant structure is extended (Supporting Information Table S2 and Figure S3).

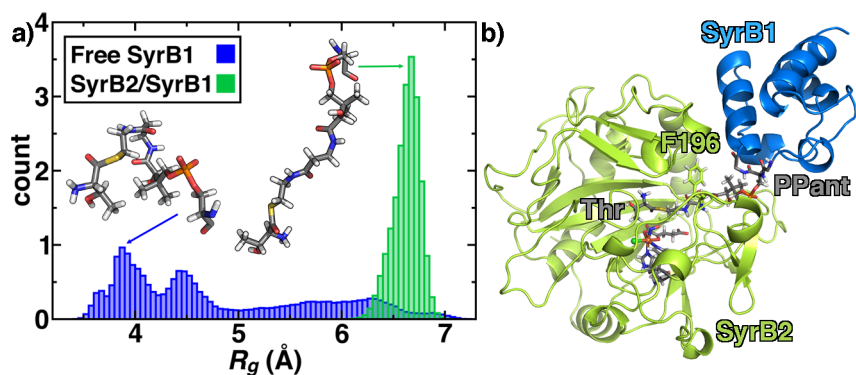


Figure 1. (a) Normalized histograms of R_g (in Å) for PPant evaluated from P to S for free SyrB1 (blue) and the SyrB2/SyrB1 complex (green) with representative snapshots shown in inset, both with Thr attached. (b) Schematic of docked SyrB2/SyrB1 structure with F196 and the active site shown as sticks. Both PPant and Thr are labeled for reference.

In comparison to free SyrB1 PPant, PPant dynamics in a docked SyrB2/SyrB1 complex show a considerably narrower R_g distribution centered around 6.7 Å (Figure 1). The predominant PPant P-S distances and C-C-C angles sampled in the SyrB2/SyrB1 complex are consistent with the elongated local minimum observed in free SyrB1 dynamics, and overall geometries are comparable (Supporting Information Figure S2 and Table S1). Thus, at room temperature, SyrB1 likely samples geometries that enable SyrB2/SyrB1 docking through PPant/substrate chain elongation, although this configuration is not the global minimum for isolated SyrB1 PPant in solution.

The SyrB2 holoenzyme has been crystallized¹⁴ with no apparent opening for SyrB1 to deliver substrate to the SyrB2 active site. However, it has been proposed¹⁴ that rotation of a phenylalanine residue (F196) could allow PPant and substrate to access the non-heme Fe(II) active site. Here, we employ long-time MD and umbrella sampling to quantify structural properties and free energies of SyrB2 protein configurations that enable substrate delivery to the active site. In the crystal structure, F196 in SyrB2 closes the channel entry to the active site with a χ_1, χ_2 dihedral pair of $(-42^\circ, 98^\circ)$, where χ_1 represents the rotation of F196 about the C α -

C β axis, and χ_2 represents the tilt of F196 with respect to the backbone with 180° periodicity due to symmetry in the phenyl ring (Figure 2 inset and Supporting Information Figure S4).

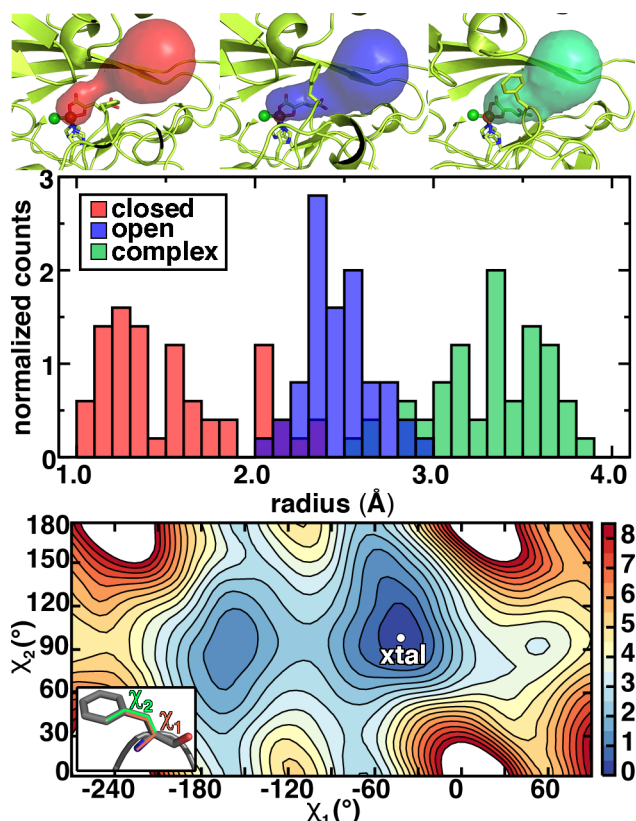


Figure 2. (top) Visualization of median cavities connecting Syrb2 active site to solvent in the closed (crystal) minimum (red, left), open minimum (blue, middle), and from a Syrb2/Syrb1 complex with Syrb1 omitted (green, right). The residues coordinating Fe(II) are shown as sticks and spheres, and the gating F196 is shown as sticks. (middle) Normalized histograms of the radius of the cavity evaluated at the F196 C β along the Syrb2 channel evaluated on 50 snapshots from the closed minimum (red), open minimum (blue), and from a Syrb2/Syrb1 complex with Syrb1 omitted (green). (bottom) Free energy surface (in kcal/mol, color bar at right) for F196 dihedrals χ_1 and χ_2 evaluated as illustrated in inset. The crystal (xtal) configuration is indicated with a white circle and the open minimum is the second minimum to the left of the crystal minimum.

We first carried out over 1.6- μ s of aggregate MD simulations of Syrb2 starting from the experimental structure and nine other rotameric states of F196 (Supporting Information Table S3 and Figure S5). Six of the simulations exclusively sampled in the vicinity of the crystal structure F196 configuration, albeit with broad distributions having widths of 60° in χ_1 and 90° in χ_2

(Supporting Information Figure S6). Without apparent dependence on starting rotamer, four other simulations sampled a number of distinct rotameric states, with multiple ns-timescale transitions observed during each 250 ns simulation (Supporting Information Figure S7).

To identify the relative free energy of metastable F196 rotameric states, we then computed the two-dimensional free energy surface (FES) defined by χ_1 and χ_2 of F196 (see Computational Details). Consistent with the equilibrium MD simulations, the crystal structure minimum is observed to be the FES global minimum, but other local minima are evident (Figure 2). Namely, a second minimum near $(-165^\circ, 95^\circ)$ is about 1 kcal/mol above the global minimum. The path connecting the two minima corresponds to rotation of the F196 phenyl ring around the C α -C β axis without significant change in the tilt of the ring and is characterized by a 2.5 kcal/mol free energy barrier (i.e., χ_1 variation with fixed χ_2 , Figure 2 and Supporting Information Figure S8).

To confirm that this second minimum corresponds to an open, active-site-accessible configuration of SyrB2 in comparison to the closed crystal structure configuration, we carried out cavity analysis⁵⁷ that has previously been fruitfully applied to related non-heme iron dioxygenases⁵⁸ (see Computational Details). In the crystal configuration, the cavity connecting the solvated exterior of SyrB2 and the Fe(II)-centered active site appears pinched at F196, with a large number of MD snapshots having a cavity radius at the C β of F196 of only approximately 1.2 Å (Figure 2). Repeating this analysis on the $(-165^\circ, 95^\circ)$ local minimum reveals a uniformly wider cavity with a radius of approximately 2.5 Å at the C β of F196 (Figure 2). Analysis on the overall bottleneck of the channel, rather than just the width at the F196 C β , shows consistent results (Supporting Information Figure S9). Thus, rotation of F196 about the C α -C β axis, which likely occurs on ns-timescales, will make substrate delivery to the active site possible.

We also evaluated the F196 dihedral in a SyrB2/SyrB1 complex to identify how F196 orientations and the channel might change in the presence of PPant and the SyrB2/SyrB1 interface. Overall, the χ_1 distribution is unchanged, but the average value of χ_2 decreases slightly, indicative of a tilting of the phenyl ring that would have a small free-energy cost in isolated SyrB2 (Figure 2 and Supporting Information Figure S10). Cavity analysis on the SyrB2/SyrB1 complex with SyrB1 removed reveals an enlarged cavity in the complex with a radius of approximately 3.4 Å to accommodate PPant and substrate (Figure 2). The slight change in F196 tilt can only partly explain the enlargement, as cavity analysis of SyrB2-only configurations with the shifted dihedral of the SyrB2/SyrB1 complex only increases the peak of the radius distribution at the C β of F196 by approximately 0.4 Å (Supporting Information Figure S11). Comparing SyrB2 in the SyrB2/SyrB1 complex with isolated SyrB2 in the open or closed configuration indeed reveals additional differences beyond just F196 rotation. Namely, the F196-containing loop moves to open the cavity, a short N62–N65 α -helix is displaced, and an adjacent I57–G59 loop reorients (Supporting Information Figure S12). These motions all contribute to further enlarge the cavity for PPant/substrate delivery upon complexation.

2b. SyrB2/SyrB1 Interface Interactions and Dynamics.

After docking SyrB1-PPant-S-Thr to an open configuration of SyrB2, the overall complex is stable, without significant increases in center-of-mass distances between the two proteins over long-time dynamics (Supporting Information Figure S13). Classical energy decomposition analysis (i.e., generalized Born or GBSA, see Computational Details) reveals that several protein-protein interface interactions stabilize the complex. Favorable interactions are predominantly electrostatic in nature and occur at the interface of flexible loops and small helices on SyrB2 with adjacent residues on SyrB1 that were shown to shift upon complexation

(Supporting Information Figure S12 and Table S4). The strongest such interactions occur between SyrB1 D589 and SyrB2 S58 and in the form of a salt bridge between SyrB1 R575 and SyrB2 D181 (Supporting Information Tables S4-S5). This model of the SyrB2/SyrB1 interface provides targets for experimental mutagenesis to stabilize a long-lived SyrB2/SyrB1 complex (Supporting Information Figure S14).

Within the SyrB2/SyrB1 complex, the PPant arm forms a number of favorable stabilizing interactions with the SyrB2 cavity, as judged through geometric hydrogen bonding (HB) and GBSA analysis. These interactions are distributed along the 14-Å length of the PPant. The PPant phosphate group hydrogen bonds at the SyrB2 surface with N184, which also participates in SyrB2/SyrB1 interface interactions (Figure 3). Favorable van der Waals interactions are observed between the gating F196 and PPant as well as the adjacent F195 (Figure 3 and Supporting Information Table S5). HB analysis reveals a strong interaction between an E111 HB acceptor and a PPant hydroxyl (~2.6-2.7 Å) with high frequency (>70% of all MD snapshots, Supporting Information Table S6). The SyrB2 S185 is a HB donor (40%) to the PPant phosphate and N62 is a HB donor to a carbonyl in a PPant amide bond with a longer bond length (2.8-2.9 Å) that means it is detected less frequently (6%, Supporting Information Table S7).

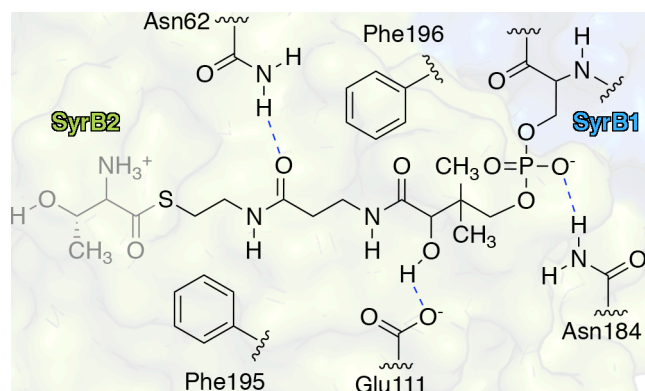


Figure 3. Schematic of key hydrogen bonding interactions (blue dashed lines) and noncovalent (stacking or electrostatic) interactions (no lines shown) between PPant and surrounding SyrB2 residues as obtained from geometric and GBSA analysis with SyrB2/SyrB1 interface indicated by background coloring.

In recent HYSCORE experiments by Silakov and coworkers⁵¹, an Fe-bound NO molecule provided an EPR-active probe, revealing that native Thr substrates are positioned further from the metal center than non-native substrates as judged by the distance of iron to the closest terminal hydrogen atoms (Thr: 4.2 ± 0.3 Å, Nva: 3.4 ± 0.3 Å, Aba 3.7 ± 0.3 Å)⁵¹. Overall, the substrates were observed to be at an acute angle to the Fe-N bond ($85 \pm 10^\circ$ for Thr/Aba and $64 \pm 7^\circ$ for Nva⁵¹), but these distances and angles leave open several possibilities for how the substrate could be oriented in the active site given ongoing uncertainty over the configurational isomer of the active site^{37-38, 41, 51} (Supporting Information Figure S15).

In the absence of available high-resolution crystal structures, MD simulation of the SyrB2/SyrB1 complex is expected to provide valuable insight into substrate delivery, substrate–protein interactions that drive substrate specificity, and dynamics at the SyrB2 active site. However, the noncovalent interactions that govern the substrate-protein positioning in the active site are notoriously difficult to describe correctly with standard, point-charge force fields⁵⁹⁻⁶⁵. Indeed, μ s-length MD of SyrB1-PPant loaded with each of the three substrates in combination with the three most probable active-site configurational isomers seldom sample the HYSCORE distances or angles, instead favoring a broad distribution (ca. 4–11 Å and up to 15 Å for Aba substrates) of terminal methyl H atom (Me) to Fe distances (Fe-Me, Supporting Information Table S8). The experimentally observed, substrate specific Fe-Me distance trend (i.e., Thr > Aba > Nva) can be discerned from the minimum sampled distances in any of the configurational isomers but is not preserved for the average sampled distances (Supporting Information Table S8).

Unlike experiments⁵¹ that show longer ethyl group hydrogen (Et) atom distances to Fe (i.e., Fe-Et) in comparison to Fe-Me (ca. 1.0 Å for Aba and substantial for Nva), MD-sampled

distances are instead comparable (Supporting Information Table S8). Longer Fe-Me and Fe-Et distances lead to sampling of a wide range of Fe-N-Me and Fe-N-Et angles (0-125°) for Fe-N-Me in all configurational isomers. At these longer distances, equatorial(eq)-NO configurations are more likely than axial(ax)-NO to sample the experimental Fe-N-Me and Fe-N-Et angles regardless of substrate (Supporting Information Figures S16-S21). This observation provides preliminary support for an eq-NO active site configuration but is counterbalanced by the fact that shorter Fe-Me distances are sampled in the ax-NO configuration (Supporting Information Table S8). The ax- α KG isomer samples the appropriate Fe-N-Me and Fe-N-Et angles for Thr substrates but, like the other isomers, does not adequately sample the short distances (Supporting Information Figure S22). In cases where free dynamics of ax- α KG samples better distances, such as for Nva and Aba substrates, agreement with experimental angles worsens (Supporting Information Figures S23-S24).

Over all sampled configurations, no substrate/configurational isomer combination yields a structure with all four distances and angles within the experimental measurement uncertainty. However, configurations close to the experimental values are infrequently observed, especially for Thr, which had the longest experimental target distances (Figure 4 and Supporting Information Figure S15 and Table S9). In these configurations, Me-Cl distances are comparable to or longer than Me-N for Thr except in the ax- α KG isomer, but Cl-proximity is also observed in this isomer for Nva, inconsistent with proximity-directing reactivity³² arguments (Supporting Information Table S10).

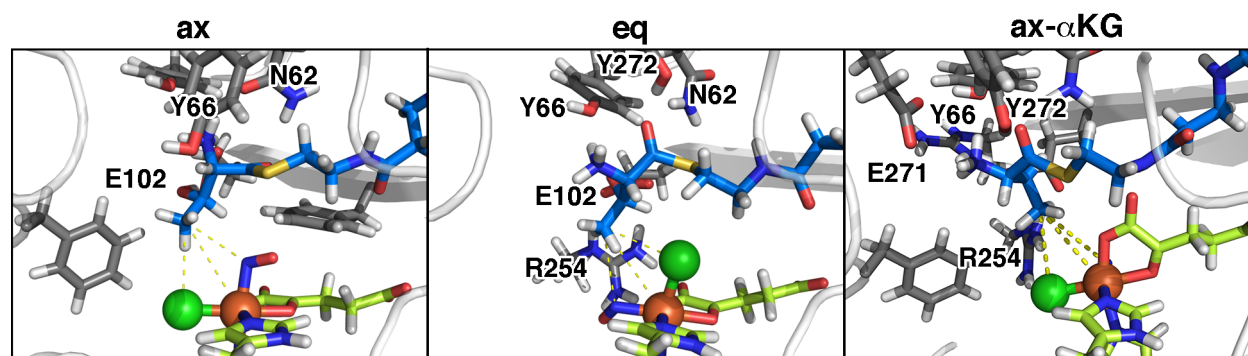


Figure 4. Representative frames from classical MD that correspond most closely to experimentally observed distances for Thr in three configurational isomers, axial (ax), equatorial (eq), and axial- α KG (ax- α KG). The protein cartoon is shown in translucent white, the active site is shown in yellow sticks, PPant-S-Thr is shown in blue, and residues with heavy atom distances within 3.25 Å of the substrate atoms are shown in gray sticks. The vectors connecting the methyl carbon to Fe(II), Cl, and N are indicated with yellow dashed lines. Select relevant residues shown in sticks are labeled.

Clustering each full trajectory reveals a range of Fe–substrate distances with one, typically non-dominant, cluster having Fe–substrate distances that are shorter than the other clusters, but these clusters do not vary strongly in terms of their protein–substrate interactions (Supporting Information Tables S11-S17). Although more differences are observed across configurational isomers than across substrates, similar interactions are observed throughout: hydrogen bonding and electrostatic interactions with N62, Y66, E102, E271, and Y272 stabilize substrate NH_3^+ or OH sidechains (Figure 4 and Supporting Information Figure S15 and Tables S11-S17). Overall, these interactions appear to be the driving force for substrate distances that do not conform to HYSORE distances. To overcome limitations of classical MD that often prevent the sampling of short noncovalent interactions⁶⁴⁻⁶⁶, we therefore turn to enforcing the HYSORE distances and identifying the protein–substrate interactions favored in configurations that satisfy experimental restraints.

2c. Restrained MD Identifies Protein-Substrate Interactions.

Adjusted restraints (see Computational Details) were used to enforce experimentally observed⁵¹ distances and angles. In all cases, reasonable structures were achieved without any

apparent distortion of the bond lengths and angles of the substrate or the active site, supporting the reported experimental values as feasible in the active SyrB2/SyrB1 complex for any of the configurational isomers. In all cases, the Fe-Me distances are somewhat easier to enforce, and all restraints are most easily satisfied in the crystal-structure-like ax-NO isomers (Figure 5 and Supporting Information Figures S25-S33). Our observation is in contrast to earlier motivations⁵¹ for alternate configurational isomers to the ax-NO crystal structure analogue based on the expectation that acute Fe-N-Me/Et angles would be hard to satisfy at such short distances in the ax-NO isomer. Average distances and angles preserve observed HYSCORE trends⁵¹ in all isomers, suggesting we cannot immediately rule out any of the configurational isomers as incompatible with the spectroscopically inferred distances and angles⁵¹ (Figure 5 and Supporting Information Table S18).

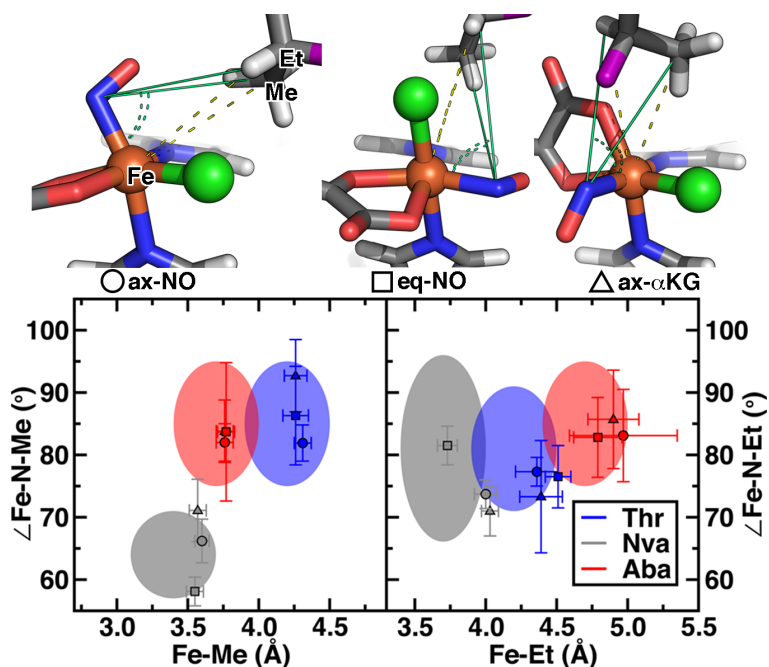


Figure 5. (top) Distance (yellow lines) and angle (green lines) definitions for the crystal-like, axial-NO (ax-NO, circles), equatorial-NO (eq-NO, squares), and axial- α KG (ax- α KG, triangles) configurational isomers. The labeled Me and Et definitions hold for all substrates, but the purple stick in the substrate indicates where the hydroxyl would be for Thr. (bottom) Fe-N-Me angle (in °) vs. Fe-Me distance (in Å) and Fe-N-Et angle (in °) vs. Fe-Et distance (in Å) for Thr (blue), Nva (gray), and Aba (red) substrates in each configurational isomer as indicated in top. The experimental values⁵¹ and uncertainties are indicated by translucent oval shaded regions. The

restrained MD average values are indicated by the symbols, and the ranges represent the standard deviation. The largest overlapping angular Fe-N-Me standard deviation is eq-NO Aba.

An open question is whether target distances may be satisfied in only one substrate orientation. Most distance and angle distributions are approximately normal with a single distinct pose (e.g., for Nva or Thr in ax/eq-NO isomers), except for Aba that samples multiple configurations as indicated by clustering analysis and broad Fe-Et distance distributions (Figure 5 and Supporting Information Figures S27, S30, and S33 and Table S19). This difference in Aba corresponds to its ability to satisfy Fe-Me target distances with both gauche and trans dihedral (i.e., $\text{N}(\text{H}_3^+)-\text{C}-\text{C}(\text{Et})-\text{C}(\text{Me})$) angles (Supporting Information Table S19 and structures provided in the Supporting Information).

To determine which favorable protein-substrate interactions underlie the experimentally observed substrate positioning and, secondarily, to rationalize why these distances were not predominantly sampled in free MD, we analyze and compare interaction (i.e., GBSA) and hydrogen bonding strengths of the trajectories (see Computational Details). In the ax/eq-NO isomers, strong interactions (i.e., from GBSA or HB analysis) are observed between all substrates and polar, neutral residues (i.e., N62, Y66, N123, and Y272). This is in contrast to unrestrained MD, in which interactions with charged residues (i.e., E102, E271) dominate. These interactions are reduced in strength in the restrained MD of Thr/Nva and absent in the case of Aba (Supporting Information Tables S12-S17 and S20-S25). For Thr, HYSCORE-restrained MD shifts $-\text{OH}$ HBs to N123 instead of E102, which instead interacts with the NH_3^+ group (Supporting Information Tables S15 and S22). Poor sampling of experimentally observed⁵¹ distances in free MD can be attributed to limitations of point charge force fields⁵⁹⁻⁶⁵. The critical role of N123 in substrate positioning will be revisited in the following section.

Although all three configurational isomers can support the experimentally observed distances, observations thus far suggest eliminating ax- α KG as a probable isomer. With respect to other configurational isomers, fewer protein–substrate interactions are observed in ax- α KG: the NH_3^+ groups of Nva and Aba form no substantial HBs and, when present, the Thr –OH/E102 HB prevents E102 from cooperatively interacting with NH_3^+ , which occurs readily in other isomers (Supporting Information Tables S14-S17 and S22-S25). Additionally, the larger α KG (vs. NO or Cl) in the axial position orients the substrate, and the NH_3^+ group in particular, away from enzyme active-site residues (Supporting Information Figure S34). These structural differences also explain the difficulty with which Fe-Et distances and angles were satisfied in this isomer (Figure 5). Although ax- α KG has been observed in clavamate synthase⁶⁷, this alternate active-site configuration for SyrB2 was proposed in the context of the expected difficulty of satisfying the restraint distances with crystal structure analogue ax-NO. Given the lack of disproportionate support for eq-NO or ax- α KG in our work, we focus on the crystal-structure analogue ax-NO isomer for further consideration of substrate-specific differences, with alternate isomer data provided in the Supporting Information.

In experimentally guided simulations, Fe-Me and Fe-Et distances appear to be related to the depth at which substrates sit in the active-site cavity. We now define a quantitative measure of the depth of the PPant-substrate delivery to SyrB2 to provide more insight into substrate–active-site interactions. After identifying immobile regions of the protein, we selected two subsets of beta-sheet residues, over which we track the center of mass of all C α atoms to serve as reference points, B11: M249-V256 and B12: Y255-V256 (Supporting Information Figure S35). In addition to following the same beta-sheet strand, these two points run adjacent to the vector along which PPant enters SyrB2, with B12 corresponding to greater active-site depth than B11.

Thus, substrate delivery depth can be judged by relative distance to B11 and B12, $d_{B12}-d_{B11}$, where smaller values correspond to deeper substrate delivery. As expected, smaller $d_{B12}-d_{B11}$ values for the Me carbon of Thr (i.e., 2.7 Å average in ax-NO) than for Nva or Aba (i.e., around 3.5-3.6 Å) indicate deeper substrate delivery for Thr (Figure 6 and Supporting Information Table S26). The $d_{B12}-d_{B11}$ values for the N in NH_3^+ rank the substrates as Thr<Aba<Nva (i.e., 2.6 Å Thr average in ax-NO vs. 3.2 and 4.5 Å for Aba and Nva), indicating the native substrate is distinguished by positioning away from the axial NO species and deeper into the active site (Figure 6 and Supporting Information Table S26). In comparison to these observations on ax-NO, eq-NO and ax- α KG isomers do not preserve these trends, positioning Thr equivalently to either Nva or Aba Me C, respectively, due to shallower and broader overall positioning of substrates in the active sites especially for ax- α KG (Supporting Information Table S26 and Figure S36).

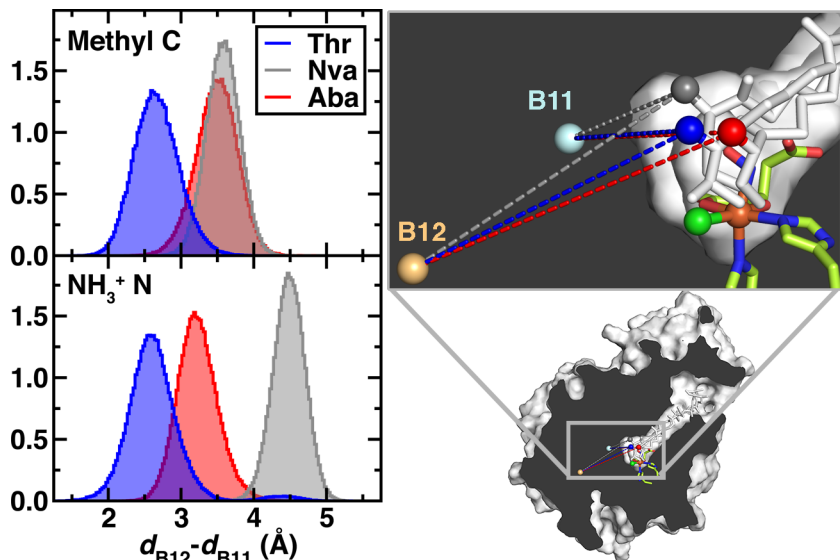


Figure 6. (left) Substrate depth delivery metric ($d_{B12}-d_{B11}$, in Å) for the methyl carbon and NH_3^+ N of Thr (blue), Nva (gray), and Aba (red) substrates. (right) Cut of SyrB2 showing PPant-S-Res delivery to active site along with B11 (light blue) and B12 (light orange) anchor point spheres for the whole protein (bottom) and enlarged zoom (top, as indicated in bottom). The distance to B11 and B12 are indicated by dashed lines for Thr (blue), Nva (gray), and Aba (red) NH_3^+ N. B11 is projected perpendicular to the plane of the page, making the distance appear somewhat shorter in this plane than it actually is (e.g., d_{B11} for Nva NH_3^+ N is 10 Å vs. 14.5 Å for d_{B12}).

It has been argued^{32, 34, 51} that substrate- and atom-specific hydroxylation/halogenation branching ratios result from substrate positioning: substrates proximal to Cl should be readily chlorinated and substrates closer to the formed OH should be hydroxylated. In both ax-NO and eq-NO configurations, distances of the methyl C to Cl span a narrow 3-4 Å range, with all substrates having average distances of around 3.3-3.5 Å (Figure 7 and Supporting Information Table S27 and Figure S37). In contrast, the C-to-N distance clearly distinguishes substrates with Nva<<Aba<Thr for both ax-NO and eq-NO (Figure 7 and Supporting Information Table S27). The enforcement of experimentally observed Fe-Me/Et distances and Fe-N-Me/Et angles provides an indirect enforcement of this order, but C-N or C-Cl distances have not been constrained in any way. Only in the case of Nva methyl (i.e., C5) are the sampled C-N distances ever shorter than C-Cl distances (Figure 7 and Supporting Information Table S27 and Figure S37). Although Aba was previously noted to sample multiple orientations, neither orientation leads to it frequently sampling shorter C-N than C-Cl distances (Supporting Information Figure S38).

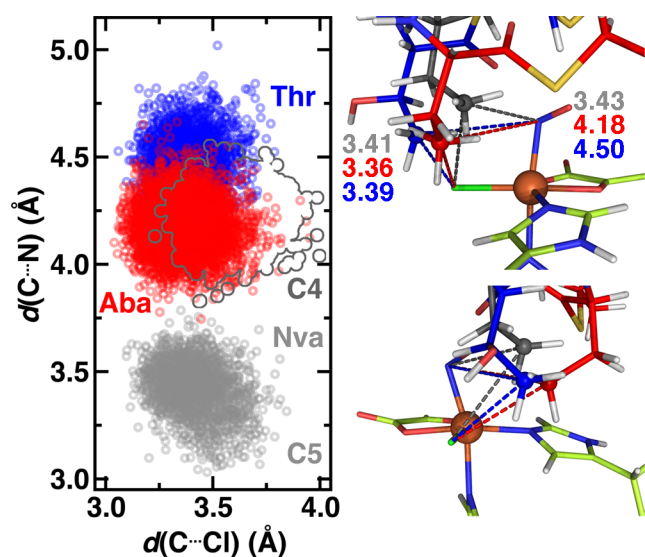


Figure 7. (left) Distribution of C...N and C...Cl distances observed for the ax-NO configurational isomer of each substrate: Thr (blue circles), Aba (red circles), and Nva (C5 carbon in gray circles, outline of C4 carbon shown in dark gray). (right) Representative median snapshots from Thr (blue sticks), Aba (red sticks), and Nva (gray sticks) shown in two views (top and bottom)

with methyl carbons shown as sphere and C \cdots N or C \cdots Cl distances shown as dashed sticks with distances (in Å) indicated as color coded at top.

According to previous experimental studies, the ethyl carbon (C4) of Nva reacts similarly to the methyl carbon of Aba in terms of both rates and branching ratios with both carbon atoms undergoing a mixture of hydroxylation and chlorination³². Similar orientations with respect to the Fe center have also been indicated⁵¹, and thus in our experimentally restrained simulations the Nva C4 has similar C-Cl and C-N distance distributions to Aba (Figure 7 and Supporting Information Figure S39). This correspondence is clearest in the ax-NO isomer (Supporting Information Figure S39). In comparison, the ethyl carbons of Aba and Thr sample much longer C-N distances in both isomers (Supporting Information Figure S40). Overall, our results are consistent with the idea that increased halogenation of Thr with lower turnover rates^{32, 51} is governed by longer Thr (vs. Nva or Aba) distances to the oxo species rather than reduced distances to Cl (Figure 7). Both ax-NO and eq-NO isomers obey this trend, indicating that the active-site configuration does not control relative C-Cl vs. C-N distance (Figure 7 and Supporting Information Figure S39).

2d. Analysis of Substrate-Dependent Hydrogen Bonding.

Although we have shown that experimentally observed distances could be enforced in dynamics, longer distances were preferred in free dynamics. Owing to the limitations of point charge force fields, we therefore turn to QM/MM geometry optimizations to further characterize active site interactions and to determine if the experimentally observed substrate/active-site positions are favorable at the QM/MM level of theory. Over nearly 50 QM/MM geometry optimizations, the distances and angles remained close to the experimental distances, with up to 0.15 Å elongation in the Fe-Me distance, especially for Nva or Aba in the ax-NO and eq-NO

isomers (Supporting Information Table S28). This elongation corresponds to a small change (typically, a decrease) in the Fe-N-Me angle (ca. 5-10°) with similar or smaller changes to the Et distances and angles (Supporting Information Table S28). In comparison, Thr Fe-Me/Et distances elongate only slightly (ca. 0.01 Å on average), and angle changes are also smaller (Supporting Information Table S28). The distinction between Thr and Nva/Aba highlights that the shortest distances proposed in the HYSCORE experiments may be a slight underestimate, consistent with recent observations in oxygenases where both HYSCORE and X-ray crystallographic distances were available for comparison⁶⁸.

Having confirmed the feasibility of the experimentally observed distances, we now turn to analyzing substrate/active-site interactions through QM/MM simulation. We particularly focus on electronic structure properties by using bond critical points (BCPs⁶⁹) from quantum theory of atoms in molecules (QTAIM, see Computational Details) to identify and quantify the strength of HB interactions. QM analysis of HBs between the substrates and the greater protein active site largely confirms observations based on classical geometric HB analysis but provides additional insight (Supporting Information Tables S15-S17 and S22-S24). Notably, Thr forms additional HBs in the ax-NO configuration, as judged by QM analysis after QM/MM optimization, that were not observed in geometric analysis alone (Supporting Information Figure S41 and Table S29). The HB between the sidechain amide H of N123 and Thr –OH that was observed in geometric analysis is confirmed here with an estimated -7.2 kcal/mol stabilizing interaction with an O...H distance of 1.9 Å (Supporting Information Table S29). In addition, a second strong interaction of -5.2 kcal/mol is simultaneously observed between the carbonyl O of the N123 sidechain and the NH₃⁺ of Thr residue with approximately 2.2 Å O...H separation (Figure 8 and Supporting Information Table S29). Aba also interacts with N123 in the ax-NO configurational

isomer, but the single HB ($d = 1.9 \text{ \AA}$) between the carbonyl O of N123 and NH_3^+ of -6.9 kcal/mol is much weaker than the sum of the two cooperative interactions for Thr (Supporting Information Table S29). The single interaction with the NH_3^+ functional group also allows Aba to position closer to the Fe center than the simultaneous interactions do in the case of Thr, as we will discuss next (Figure 8 and Supporting Information Table S29). The same closer positioning of Nva due to a single interaction with NH_3^+ is also observed, but the HB interactions are formed instead with Y66, E102, and E271 (Supporting Information Table S29). The QM HB analysis also reinforces the lack of likely validity of the ax- α KG isomer, as no HBs are detected for the Nva substrate and no OH-specific HBs are detected for Thr, in comparison to those present in the other two isomers (see Supporting Information Table S29).

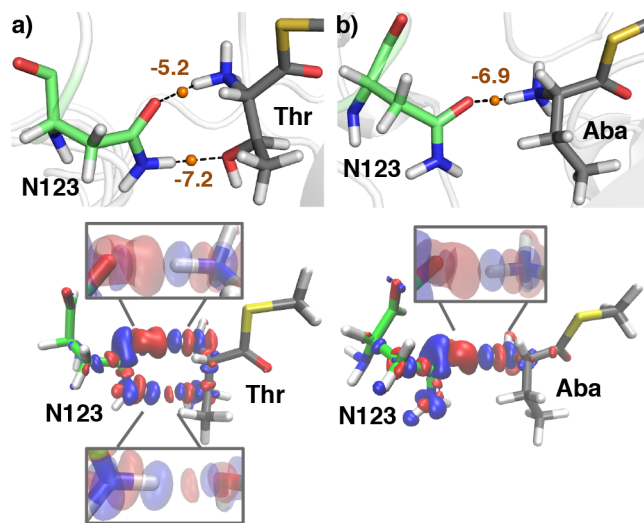


Figure 8. Comparison of quantum mechanical properties of hydrogen bonding interactions for a) Thr in a double HB configuration with N123 and b) Aba in a single HB with N123 at distances sampled during classical MD. The visualization of the QTAIM bond critical points and their hydrogen bond energies (in kcal/mol) is shown at top, and the density difference on a minimal model of the interaction between the complexed residues and isolated species is shown at bottom with an isosurface of 0.0012 e and density gain shown in red and density loss shown in blue.

To further quantify the unique HB configuration observed in Thr, we extracted, for QM/MM optimization, MD snapshots at varying Thr NH_3^+ /N123 carbonyl distances that were identified less frequently as HBs by classical analysis (Supporting Information Table S30). For

MD snapshots with two initially short HBs, QM/MM optimization preserves the double HB configuration and exhibits total HB energies as large as -21 kcal/mol for the shortest QM/MM-stabilized 1.7-1.9 Å HBs (Supporting Information Table S30). These bond lengths are similar to those in a spectroscopically characterized cooperative hydrogen bonding interaction between formamide and a water chain⁷⁰. Optimization of structures with a single HB to the Thr –OH does not result in reorientation to recapture the second HB, as expected from a double well potential (Supporting Information Table S30). In fact, in intermediate orientations, the single HB ($d = 1.8$ Å) that remains between Thr –OH and N123 amide H (where $d = 4.0$ Å for the other HB) can be as strong as -10 kcal/mol, stronger than the same interaction in double HB configurations (Supporting Information Table S30).

Although the double HB configuration was not frequently observed in classical geometric HB analysis, such analysis employs a tight heavy atom cutoff ($d(\text{H}\cdots\text{X}) = 3.0$ Å) and short noncovalent interactions are disfavored by standard force field parameters. Thus, we directly computed the distances of the Thr –OH and –NH_3^+ to the N123 amide H and carbonyl O throughout MD (Supporting Information Figures S42 and S43). With extensive sampling, we indeed observed distances consistent with a double HB configuration, though it occurred less frequently than a single HB (Supporting Information Figure S42). Over 1 μs of dynamics, we estimate the relative free energy of the configurations from their probabilities and find the double HB to be within 1 kcal/mol of the single HB global minimum and to be separated from it by a small free energy barrier (Figure 9). Single HB configurations are strongly preferred in the eq-NO configuration over comparable sampling times (Supporting Information Figure S43). Neither HB to N123 is observed over 500 ns of MD for the ax- αKG case, with the non-Thr-specific NH_3^+ HB being the only one that forms appreciably (Supporting Information Figure S43). Analyzing

the ax-NO double HB approximate free energies, we conclude that the relative stability of the two configurations is likely sensitive to the force field parameters, and both ax-NO and eq-NO isomers have the potential to favor a double HB configuration with N123 (Figure 9).

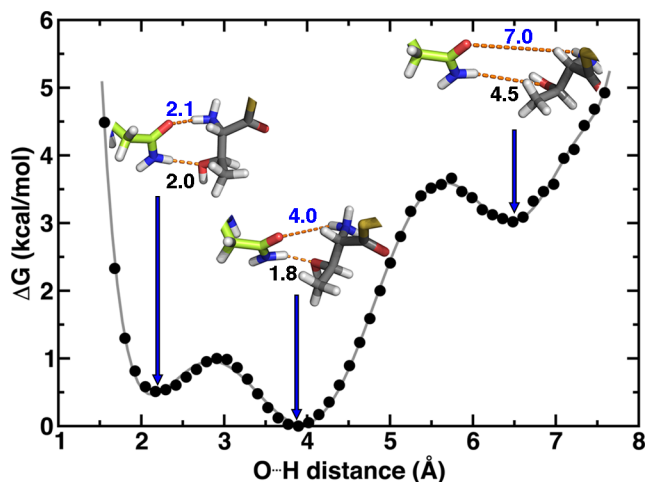


Figure 9. Approximate free energy (in kcal/mol) from $-kT \ln\langle P \rangle$ of a 1 μ s simulation of SyrB2/SyrB1 in the axial-NO configurational isomer over 0.1 Å windows (black dots) and with a 10th order fit (gray line) also shown. The reaction coordinate (O...H distance in Å) is the N123 carbonyl O to closest Thr NH₃⁺ distance. Representative structures with measured distances (orange dashed lines) are shown in inset between N123 in lime green and Thr in gray. The reaction coordinate distance in Å is annotated in blue, and the N123 amide H to Thr -O(H) distance is shown in Å in black.

Given that QM HB analysis revealed interactions and distances shorter than had been observed in MD, we investigated the interaction of minimal QM-only models of not just N123 but also the active-site N62, Y66, E102, and E271 residues with truncated PPant-*S*-Res (Res=Thr, Nva, or Aba) models (Supporting Information Table S31). When the substrates and protein residues are no longer restrained in any way by the protein, Nva and Aba form multiple, strong HBs ($d = 1.7$ - 2.0 Å) with N123, including to the backbone carbonyl O (Supporting Information Table S31 and Figure S44). When we freeze backbone atoms of the PPant-*S*-Res fragment and the protein residues, interactions comparable to the QM/MM model are instead recovered: Nva forms no HBs with N123 and Aba forms one only to the sidechain carbonyl O

atom (Supporting Information Table S32 and Figure S44). Throughout MD, the distance of the amine N of any substrate to the backbone carbonyl O atom is never less than 4.5 Å (this shortest value is observed only for Aba) for any of the configurational isomers (Supporting Information Figure S45). Thus, it appears that the greater protein environment is essential in positioning N123 to form a double HB with the Thr amine and –OH sidechain that is not observed for the non-native substrates (Figure 10 and Supporting Information Figure S44).

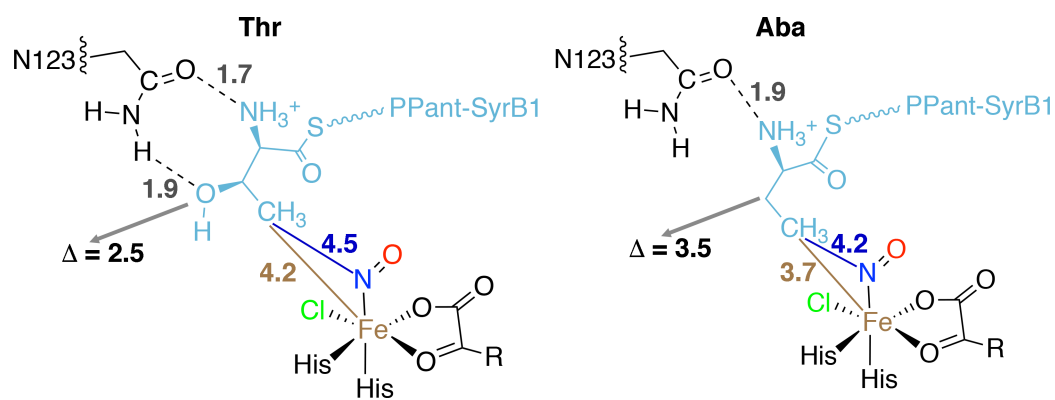


Figure 10. Substrate and PPant orientation in the SyrB2/SyrB1 complex with key distances annotated: dB12-dB11 of the methyl carbon (labeled with gray arrow and Δ , in Å), distance of methyl carbon to N (dark blue, in Å), average distance of methyl H atoms to Fe (brown, in Å), and hydrogen bond distances between the substrate (light blue) and N123 (shown in dark gray and dashed lines, in Å) for Thr (left) and Aba (right).

Based on the distinct HB interactions observed in the native and non-native substrates, it is now possible to rationalize the difference in reactivity based on different substrate positions. Previous experimental and theoretical studies^{32, 34} have emphasized the importance of substrate positioning in the diverging reactivity of SyrB2 for native and non-native substrates and suggested that residues (e.g., R254) could interact with the Fe-OH moiety to prevent rebound hydroxylation for Thr³⁸. However, substrate-specific hydrogen bonding to the protein has not been identified to play a role in substrate positioning and thus reaction selectivity in SyrB2/SyrB1. The protein environment provides two strong HBs between the Thr and N123 that are stabilizing by up to 21 kcal/mol (Figure 10). Alternative interactions observed in non-native

substrates through only the NH_3^+ group are smaller (around 15 kcal/mol) and less directional for Aba or involve interacting with multiple residues, e.g., as in Nva with Y66, E102, and E271 (Figure 8). These more flexible interactions with only one functional group on the substrate (i.e., for Nva or Aba) allow the substrate to remain near Fe(II) (Figure 10). The double HB configuration, on the other hand, pulls Thr further away from the Fe(II) in the active site, and consequently from the reactive Fe(IV)-oxo intermediate in the catalytic cycle, but at an optimum distance from Cl thus allowing SyrB2 to prefer chlorination for Thr instead of hydroxylation as in Nva (Figure 10).

Based on these observations, we would expect mutation of N123 to significantly affect the chlorination of native substrate Thr by the multi-protein assembly SyrB2-SyrB1. Review of the literature allowed us to identify a prior experimental study that showed that the mutation of N123 causes a decrease in production of syringomycin E by SyrB2 as shown by a decrease in antifungal activity. The residual activity after mutation was shown to be 26–30% for N123A in SyrB2³⁰, potentially due to loss of the critical substrate positioning interaction. This relationship could be explored further through HYSCORE of N123A or N123Q mutants that should influence substrate-active-site distances.

3. Conclusions

From well over 10 μs of MD of both free and complexed SyrB2 and SyrB1 along with large-scale QM/MM simulation, we have investigated the substrate–protein and protein–protein dynamics that give rise to experimentally observed substrate positioning and reactivity trends. First, we confirmed low 2-3 kcal/mol free energy barriers for F196 rotation in SyrB2 to open a channel for substrate delivery. We identified concomitant rearrangement of surface residues on SyrB1 (i.e., N62-N65 and I57-G59) that further widened the substrate delivery cavity when a

SyrB1 model was docked to SyrB2, indicating concerted motions that enable substrate delivery. Although in solution the SyrB1 PPant-*S*-Res arm prefers a closed conformation, the elongated orientation compatible with substrate delivery is also transiently observed in solution. In the context of the SyrB2 channel, PPant is restricted in its motion both due to the shape of the cavity and the formation of interactions with essential residues that stabilize PPant (e.g., N184, F195, F196) and the SyrB2/SyrB1 (e.g., D181/R575 and S58/D589) interface.

Although free dynamics of the SyrB2/SyrB1 complex did not sample experimentally observed substrate–Fe(II) active site distances, HYSCORE-derived⁵¹ restraints easily satisfied target distances to enable investigation of protein–substrate interactions. We observed a unique, cooperative pattern of hydrogen bonding that occurred only for the native Thr substrate: both hydroxyl and amine functional groups formed hydrogen bonds with N123. As judged through QM/MM analysis, this interaction was stronger than any alternative amine-only hydrogen bonding configurations. Conversely, Nva and Aba are only able to form hydrogen bonds to the protein with their amine functional groups, enabling them the flexibility to position closer to the Fe center. This analysis was consistent in both *ax*-NO and *eq*-NO configurational isomers, but a lack of protein–substrate interactions in the *ax*- α KG configuration suggested it was unlikely in SyrB2/SyrB1. Overall, the unique orientation of native Thr corresponded to quantitatively deeper substrate delivery than non-native substrates. At odds with some previous expectations, all substrates had similar proximity to Cl, but Thr was most distant from the anticipated location of the oxo species. This work demonstrates how spectroscopically guided simulation provides essential insight into protein–substrate and protein–protein interactions in biosynthetic enzyme complexes, and paves the way for equivalent interactions to be incorporated into biomimetic catalysts for challenging reactions.

4. Computational Details

Protein structure and preparation. Chain A of the crystal structure of a SyrB2 dimer (PDB ID: 2FCT¹⁴) was used for simulation with crystallizing agents removed. The structure was missing two residues at the N-terminus, which were omitted (i.e., the third residue was treated as the N-terminal residue), but a second loop of 4 missing residues (I57, S58, G59, G60) was added back by hand in PyMOL⁷¹. The thiolation (T) domain of SyrB1 (residues 538-609) was homology modeled against an ACP (PDB ID: 2KR5⁵⁶) using Phyre⁷² and equilibrated to remove steric clashes (see equilibration protocol next). The charge state of SyrB2 and SyrB1 apoenzymes were assigned using the H++ webserver⁷³⁻⁷⁵ assuming a pH of 7.0 with all other defaults applied (Supporting Information Tables S33-S34). After manual charge assignment of residues adjacent to cofactors/substrates, the SyrB2 holoenzyme (apoenzyme) has a net charge of -12 (-11), and SyrB1 holo/apo forms both have a net charge of -2 (Supporting Information Tables S33-S34). The SyrB2/SyrB1 complex was obtained by aligning S570 of equilibrated SyrB1 to the Ser residue of a previously docked³⁴ Ser-PPant-S-Thr structure in SyrB2. This procedure was repeated for Nva and Aba substrates and the three active site configurational isomers: equatorial-NO (eq-NO), axial-NO (ax-NO), and axial- α KG (ax- α KG) and prepared for simulation with AMBER⁷⁶ using the tleap utility (Supporting Information).

For non-standard residues (α KG, NO, and PPant), we use the generalized AMBER force field (GAFF)⁷⁷ with restrained electrostatic potential (RESP) charges⁷⁸ obtained at the Hartree-Fock/6-31G*⁷⁹ level using GAMESS-US⁸⁰, as implemented by the R.E.D.S. web server⁸¹⁻⁸³. Cl⁸⁴ and Fe(II) parameters⁸⁵ were obtained from the literature, and standard parameters were used for Fe(II)-coordinating H116 and H235 residues. In eq-NO, Cl⁻ and Fe(II) partial charges were rescaled to -0.5 e and +1.5 e, respectively, to improve sampling dynamics, consistent with

recommendations in MM modeling of ions⁸⁶. The complex was solvated in a periodic rectangular prism box with at least a 10 Å buffer of TIP3P⁸⁷ water and neutralized with 14 Na⁺ counterions for a total simulation of 48,954-48,957 atoms (5,998-6,001 protein/substrate atoms).

MM Equilibration and Dynamics. The complexes were equilibrated with MM MD using the GPU-accelerated PMEMD code in AMBER⁷⁶. Equilibration steps were: i) restrained (1000 steps) and unrestrained (2000 steps) minimizations, ii) 10-ps NVT heating to 300 K with a Langevin thermostat with collision frequency of 5.0 ps⁻¹ and a random seed, and iii) 1-ns NpT equilibration using the Berendsen barostat with a pressure relaxation time of 2 ps. The core active site was held fixed throughout the simulation with 200 kcal/mol·Å² bond distance force constants around experimental crystal structure¹⁴ values (Supporting Information Table S35). Unrestrained production dynamics were collected for at least 250 ns for each active site configurational isomer and substrate combination (Supporting Information Table S36). The SHAKE algorithm⁸⁸ was applied with a 2-fs timestep for all MD, and the particle mesh Ewald method was used for long range electrostatics with a 10-Å electrostatic cutoff.

Restrained MD Dynamics. Snapshots in which a weighted average of the core-substrate distances and angles sampled were closest to target values from free production MM dynamics were chosen for subsequent restrained dynamics (see files in Supporting Information and Supporting Information Tables S37-S38). Starting from step ii) of the equilibration procedure, harmonic restraints were added to enforce the target distances and angles between the active site Fe and N and the methyl and ethyl carbon atoms of the substrate, as obtained from experimental HYSCORE data⁵¹ with the flat portion of the restraint corresponding to reported values and their uncertainties⁵¹ (Supporting Information Tables S37-S38). These restraints were then adjusted iteratively in 10° or 0.1 Å increments to enforce sampling of target distances (Supporting

Information Table S39).

Umbrella Sampling and Analysis. MD free energy surfaces were obtained for the SyrB2 F196 channel-gating mechanism by applying the weighted-histogram analysis method (WHAM)⁸⁹⁻⁹⁰ using the Grossfield software package⁹¹ to unbiased distributions obtained with umbrella sampling.⁹² In total, 400 ps NpT equilibration and 200 ps NpT production was obtained for each equally spaced window. Over the $\chi_1=-180-180^\circ$ and $\chi_2=0-180^\circ$ ranges in F196 dihedrals, 5° windows were sampled with targets enforced by 200 kcal/(mol·rad²) force constants. The WHAM free energy weights were converged to within 1×10^{-8} . For each relevant configuration, 50 representative snapshots were extracted and analyzed with CAVER⁵⁷ to quantify cavities, with a probe radius of 1.0 Å, shell radius of 5.5 Å, shell depth of 7.0 Å, and a starting point in the center of the protein active site. We verified visually that the cavity was between the protein exterior and the active site via F196, which was also the largest detected cavity in most cases. The overall cavity bottleneck was extracted directly from CAVER output, and the sphere radius in the path closest to the C β of F196 was used to determine the cavity radius at F196.

Analysis of MD Trajectories. Snapshots from trajectories spaced 16 ps apart were clustered by RMSD on heavy atoms in the PPant/substrate using a bottom-up approach with minimum distance between clusters of 4 Å and the average linkage algorithm and a target of five clusters based on guidelines in Ref. 93. Interaction analysis employed the AMBER MMPBSA.py⁹⁴ utility, which follows protocols outlined in Ref. 95. This analysis used the Generalized Born (GB)⁹⁶ approximation on snapshots obtained from MD both with a noncovalent ligand present and rigidly removed for up to 625 snapshots from a cluster spaced 2 ps frames apart. Pairwise residue interactions and residue electrostatic and van der Waals'

contributions to binding were computed using the "OBC1" model⁹⁷, as suggested by benchmarks⁹⁸. Geometric hydrogen bonding strength and other computed distances and angles in the active site were obtained with the ptraj utility in AMBER⁷⁶.

QM/MM or QM-only Simulation and Analysis. Snapshots from MD production runs were extracted for QM/MM geometry optimizations. The periodic box was post-processed using PyMOL⁷¹ to generate the largest spherical droplet centered around each protein and prepared with tleap to enforce spherical cap boundary conditions with a restraining potential of 1.5 kcal/mol Å². All QM/MM simulations were carried out using TeraChem⁹⁹⁻¹⁰⁰ for the QM portion as was previously motivated¹⁰¹. The 382-atom (410 with link atoms) QM region (net charge: -1, spin multiplicity: 6) was selected with systematic analysis¹⁰²⁻¹⁰⁴ of PPant/protein interactions (Supporting Information Table S40). The QM region is modeled with unrestricted density functional theory (DFT) using the range-separated hybrid ω PBEh¹⁰⁵ ($\omega=0.2$ bohr⁻¹) with an LANL2DZ effective core potential¹⁰⁶ on Fe and 6-31G** for the remaining atoms. The quantum theory of atoms in molecules (QTAIM) bond critical points (BCPs)⁶⁹ were identified with Multiwfn¹⁰⁷ on QM/MM snapshots, and HB energies were estimated from the potential energy density of the closest BCP.¹⁰⁸ QM-only optimizations were carried out at the same level of theory on minimal models of the PPant-bound substrate (i.e., to the thioester linkage, which was capped with CH₃) and protein residues (H-capped N123, Y66, E102, and E271). Structures were extracted from protein snapshots for optimizations in which the backbone atoms (i.e., C α , N, C, O in the residue, methyl C for the substrate) were either free or constrained, and the protein environment was mimicked with the COSMO¹⁰⁹⁻¹¹⁰ implicit solvent model ($\epsilon=4$).

ASSOCIATED CONTENT

Supporting Information. Residue protonation states; active site restraints; production run time; initial target restraints and snapshots; final restraints; QM region definition; histogram of SyrB1 PPant distances; distance and radii of gyration relationships; SyrB1 homologue NMR ensemble characteristics; free energy of P-S coordinate for SyrB1; dihedral analysis for SyrB2: MD sampled dihedrals, time series, and 1D cut of the free energy surface; bottleneck analysis of open and closed SyrB2 configurations; dihedral and radius analysis of F196 in SyrB2/SyrB1 complex; complexed and free SyrB2 structural analysis; SyrB1-SyrB2 dynamics analysis; SyrB2/SyrB1 interface GBSA and surface interaction visualization; PPant-SyrB2 hydrogen bonding; methyl and ethyl distance definitions and distances in free MD for three substrates in three configurational isomers; properties of target snapshots in free MD; clustering, GBSA, and hydrogen bonding analysis in free MD; distance analysis, clustering, GBSA, and hydrogen bonding analysis in restrained MD; anchor point movements along the trajectory; methyl/ethyl-Cl and methyl/ethyl-NO distances; QM/MM geometry optimization results; QTAIM analysis of QM/MM and minimal model QM geometry optimizations; hydrogen bond N123 dynamics analysis and time series; analysis of backbone N123-substrate distances. (PDF)

Centroids of free and restrained dynamics in PDB format; initial configurations for free and restrained MD for all proteins; raw data of several dynamics indicators for SyrB1 and SyrB2/SyrB1 complex; adjusted restraint files. (ZIP)

This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

*email: hjkulik@mit.edu, phone: 617-253-4584

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

Support for this research was provided by a core center grant P30-ES002109 from the National Institute of Environmental Health Sciences, National Institutes of Health and an NEC Corporation Grant from the MIT Research Support Committee. H.J.K. holds a Career Award at the Scientific Interface from the Burroughs Wellcome Fund, which supported the work. H.W.Q. was supported in part by a Department of Energy Computational Science Graduate Fellowship (DOE-CSGF). This work was carried out in part using computational resources from the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562. This work used the XStream computational resource, supported by the National Science Foundation Major Research Instrumentation program (ACI-1429830).

References

1. Krebs, C.; Galonic Fujimori, D.; Walsh, C. T.; Bollinger Jr, J. M., Non-Heme Fe (IV)–Oxo Intermediates. *Acc. Chem. Res.* **2007**, *40*, 484-492.
2. Solomon, E. I.; Brunold, T. C.; Davis, M. I.; Kemsley, J. N.; Lee, S.-K.; Lehnert, N.; Neese, F.; Skulan, A. J.; Yang, Y.-S.; Zhou, J., Geometric and Electronic Structure/Function Correlations in Non-Heme Iron Enzymes. *Chem. Rev.* **2000**, *100*, 235-350.
3. Hausinger, R. P., Fe (II)/A-Ketoglutarate-Dependent Hydroxylases and Related Enzymes. *Crit. Rev. Biochem. Mol. Biol.* **2004**, *39*, 21-68.
4. Costas, M.; Mehn, M. P.; Jensen, M. P.; Que, L., Dioxygen Activation at Mononuclear Nonheme Iron Active Sites: Enzymes, Models, and Intermediates. *Chem. Rev.* **2004**, *104*, 939-986.
5. Solomon, E. I.; Goudarzi, S.; Sutherlin, K. D., O₂ Activation by Non-Heme Iron Enzymes. *Biochemistry* **2016**, *55*, 6363-6374.
6. Timmins, A.; de Visser, S., A Comparative Review on the Catalytic Mechanism of Nonheme Iron Hydroxylases and Halogenases. *Catalysts* **2018**, *8*, 314.
7. Walsh, C. T., The Chemical Versatility of Natural-Product Assembly Lines. *Acc. Chem. Res.* **2007**, *41*, 4-10.
8. Oloo, W. N.; Que Jr, L., Bioinspired Nonheme Iron Catalysts for C–H and C–C Bond Oxidation: Insights into the Nature of the Metal-Based Oxidants. *Acc. Chem. Res.* **2015**, *48*, 2612-2621.
9. Nam, W., High-Valent Iron (IV)–Oxo Complexes of Heme and Non-Heme Ligands in Oxygenation Reactions. *Acc. Chem. Res.* **2007**, *40*, 522-531.
10. Pangia, T. M.; Davies, C. G.; Prendergast, J. R.; Gordon, J. B.; Siegler, M. A.; Jameson, G. N.; Goldberg, D. P., Observation of Radical Rebound in a Mononuclear Nonheme Iron Model Complex. *J. Am. Chem. Soc.* **2018**, *140*, 4191-4194.
11. Hill, E. A.; Weitz, A. C.; Onderko, E.; Romero-Rivera, A.; Guo, Y.; Swart, M.; Bominaar, E. L.; Green, M. T.; Hendrich, M. P.; Lacy, D. C., et al., Reactivity of an Fe^{IV}-Oxo Complex with Protons and Oxidants. *J. Am. Chem. Soc.* **2016**, *138*, 13143-13146.
12. Rana, S.; Biswas, J. P.; Sen, A.; Clémancey, M.; Blondin, G.; Latour, J.-M.; Rajaraman, G.; Maiti, D., Selective C–H Halogenation over Hydroxylation by Non-Heme Iron (IV)–Oxo. *Chem. Sci.* **2018**.
13. Vaillancourt, F. H.; Yin, J.; Walsh, C. T., SyrB2 in Syringomycin E Biosynthesis Is a Nonheme Fe^{II} A-Ketoglutarate- and O₂-Dependent Halogenase. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 10111-10116.
14. Blasiak, L. C.; Vaillancourt, F. H.; Walsh, C. T.; Drennan, C. L., Crystal Structure of the Non-Haem Iron Halogenase SyrB2 in Syringomycin Biosynthesis. *Nature* **2006**, *440*, 368-371.
15. Galonić, D. P.; Barr, E. W.; Walsh, C. T.; Bollinger Jr, J. M.; Krebs, C., Two Interconverting Fe (IV) Intermediates in Aliphatic Chlorination by the Halogenase CytC3. *Nat. Chem. Biol.* **2007**, *3*, 113.
16. Wong, C.; Fujimori, D. G.; Walsh, C. T.; Drennan, C. L., Structural Analysis of an Open Active Site Conformation of Nonheme Iron Halogenase CytC3. *J. Am. Chem. Soc.* **2009**, *131*, 4872-4879.
17. Vaillancourt, F. H.; Yeh, E.; Vosburg, D. A.; O'connor, S. E.; Walsh, C. T., Cryptic Chlorination by a Non-Haem Iron Enzyme During Cyclopropyl Amino Acid Biosynthesis. *Nature* **2005**, *436*, 1191.
18. Mitchell, A. J.; Zhu, Q.; Maggiolo, A. O.; Ananth, N. R.; Hillwig, M. L.; Liu, X.; Boal, A. K., Structural Basis for Halogenation by Iron-and 2-Oxo-Glutamate-Dependent Enzyme Welo5. *Nat. Chem. Biol.* **2016**, *12*, 636.
19. Khare, D.; Wang, B.; Gu, L.; Razelun, J.; Sherman, D. H.; Gerwick, W. H.; Håkansson, K.; Smith, J. L., Conformational Switch Triggered by A-Ketoglutarate in a Halogenase of Curacin a Biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 14099-14104.

20. Pratter, S. M.; Light, K. M.; Solomon, E. I.; Straganz, G. D., The Role of Chloride in the Mechanism of O₂ Activation at the Mononuclear Nonheme Fe (II) Center of the Halogenase Hctb. *J. Am. Chem. Soc.* **2014**, *136*, 9385-9395.
21. Matthews, M. L.; Krest, C. M.; Barr, E. W.; Vaillancourt, F. H.; Walsh, C. T.; Green, M. T.; Krebs, C.; Bollinger Jr, J. M., Substrate-Triggered Formation and Remarkable Stability of the C–H Bond-Cleaving Chloroferryl Intermediate in the Aliphatic Halogenase, SyrB2. *Biochemistry* **2009**, *48*, 4331-4343.
22. Mitchell, A. J.; Dunham, N. P.; Bergman, J. A.; Wang, B.; Zhu, Q.; Chang, W.-c.; Liu, X.; Boal, A. K., Structure-Guided Reprogramming of a Hydroxylase to Halogenate Its Small Molecule Substrate. *Biochemistry* **2017**, *56*, 441-444.
23. Koehntop, K. D.; Emerson, J. P.; Que, L., The 2-His-1-Carboxylate Facial Triad: A Versatile Platform for Dioxygen Activation by Mononuclear Non-Heme Iron (II) Enzymes. *JBIC, J. Biol. Inorg. Chem.* **2005**, *10*, 87-93.
24. Que Jr, L., One Motif—Many Different Reactions. *Nat. Struct. Mol. Biol.* **2000**, *7*, 182.
25. Matthews, M. L.; Chang, W.-c.; Layne, A. P.; Miles, L. A.; Krebs, C.; Bollinger Jr, J. M., Direct Nitration and Azidation of Aliphatic Carbons by an Iron-Dependent Halogenase. *Nat. Chem. Biol.* **2014**, *10*, 209.
26. Groves, J. T., Key Elements of the Chemistry of Cytochrome P-450: The Oxygen Rebound Mechanism. ACS Publications: 1985.
27. Price, J. C.; Barr, E. W.; Glass, T. E.; Krebs, C.; Bollinger, J. M., Evidence for Hydrogen Abstraction from C1 of Taurine by the High-Spin Fe (IV) Intermediate Detected During Oxygen Activation by Taurine: A-Ketoglutarate Dioxygenase (Taud). *J. Am. Chem. Soc.* **2003**, *125*, 13008-13009.
28. Price, J. C.; Barr, E. W.; Tirupati, B.; Bollinger, J. M.; Krebs, C., The First Direct Characterization of a High-Valent Iron Intermediate in the Reaction of an A-Ketoglutarate-Dependent Dioxygenase: A High-Spin Fe (IV) Complex in Taurine/A-Ketoglutarate Dioxygenase (Taud) from Escherichia Coli. *Biochemistry* **2003**, *42*, 7497-7508.
29. Blasiak, L. C.; Drennan, C. L., Structural Perspective on Enzymatic Halogenation. *Acc. Chem. Res.* **2008**, *42*, 147-155.
30. Fullone, M. R.; Paiardini, A.; Miele, R.; Marsango, S.; Gross, D. C.; Omura, S.; Ros-Herrera, E.; di Patti, M. C. B.; Lagana, A.; Pascarella, S., Insight into the Structure–Function Relationship of the Nonheme Iron Halogenases Involved in the Biosynthesis of 4-Chlorothreonine–Thr3 from Streptomyces Sp. Oh-5093 and SyrB2 from Pseudomonas Syringae Pv. Syringae B301dr. *FEBS J.* **2012**, *279*, 4269-4282.
31. Solomon, E. I.; Light, K. M.; Liu, L. V.; Srnc, M.; Wong, S. D., Geometric and Electronic Structure Contributions to Function in Non-Heme Iron Enzymes. *Acc. Chem. Res.* **2013**, *46*, 2725-2739.
32. Matthews, M. L.; Neumann, C. S.; Miles, L. A.; Grove, T. L.; Booker, S. J.; Krebs, C.; Walsh, C. T.; Bollinger, M. J., Substrate Positioning Controls the Partition between Halogenation and Hydroxylation in the Aliphatic Halogenase, SyrB2. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 17723-17728.
33. de Visser, S. P.; Latifi, R., Carbon Dioxide: A Waste Product in the Catalytic Cycle of A-Ketoglutarate Dependent Halogenases Prevents the Formation of Hydroxylated by-Products. *J. Phys. Chem. B* **2009**, *113*, 12-14.
34. Kulik, H. J.; Drennan, C. L., Substrate Placement Influences Reactivity in Non-Heme Fe (II) Halogenases and Hydroxylases. *J. Biol. Chem.* **2013**, *288*, 11233–11241.
35. Kulik, H. J.; Blasiak, L. C.; Marzari, N.; Drennan, C. L., First-Principles Study of Non-Heme Fe(II) Halogenase SyrB2 Reactivity. *J. Am. Chem. Soc.* **2009**, *131*, 14426-14433.
36. Rugg, G.; Senn, H. M., Formation and Structure of the Ferryl [Fe=O] Intermediate in the Non-Haem Iron Halogenase SyrB2: Classical and QM/MM Modelling Agree. *Phys. Chem. Chem. Phys.* **2017**, *19*, 30107-30119.
37. Borowski, T.; Noack, H.; Radon, M.; Zych, K.; Siegbahn, P. E., Mechanism of Selective Halogenation by SyrB2: A Computational Study. *J. Am. Chem. Soc.* **2010**, *132*, 12887-12898.

38. Huang, J.; Li, C.; Wang, B.; Sharon, D. A.; Wu, W.; Shaik, S., Selective Chlorination of Substrates by the Halogenase SyrB2 Is Controlled by the Protein According to a Combined Quantum Mechanics/Molecular Mechanics and Molecular Dynamics Study. *ACS Catal.* **2016**, *6*, 2694-2704.
39. Srnc, M.; Wong, S. D.; Matthews, M. L.; Krebs, C.; Bollinger Jr, J. M.; Solomon, E. I., Electronic Structure of the Ferryl Intermediate in the A-Ketoglutarate Dependent Non-Heme Iron Halogenase SyrB2: Contributions to H Atom Abstraction Reactivity. *J. Am. Chem. Soc.* **2016**, *138*, 5110-5122.
40. Wong, S. D.; Srnc, M.; Matthews, M. L.; Liu, L. V.; Kwak, Y.; Park, K.; Bell III, C. B.; Alp, E. E.; Zhao, J.; Yoda, Y., Elucidation of the Fe (IV)= O Intermediate in the Catalytic Cycle of the Halogenase SyrB2. *Nature* **2013**, *499*, 320.
41. Srnc, M.; Solomon, E. I., Frontier Molecular Orbital Contributions to Chlorination Versus Hydroxylation Selectivity in the Non-Heme Iron Halogenase SyrB2. *J. Am. Chem. Soc.* **2017**, *139*, 2396-2407.
42. Nguyen, C.; Haushalter, R. W.; Lee, D. J.; Markwick, P. R.; Bruegger, J.; Caldara-Festin, G.; Finzel, K.; Jackson, D. R.; Ishikawa, F.; O'dowd, B., Trapping the Dynamic Acyl Carrier Protein in Fatty Acid Biosynthesis. *Nature* **2014**, *505*, 427.
43. Martinez, S.; Fellner, M.; Herr, C. Q.; Ritchie, A.; Hu, J.; Hausinger, R. P., Structures and Mechanisms of the Non-Heme Fe (II)-and 2-Oxoglutarate-Dependent Ethylene-Forming Enzyme: Substrate Binding Creates a Twist. *J. Am. Chem. Soc.* **2017**, *139*, 11980-11988.
44. Martinie, R. J.; Pollock, C. J.; Matthews, M. L.; Bollinger Jr, J. M.; Krebs, C.; Silakov, A., Vanadyl as a Stable Structural Mimic of Reactive Ferryl Intermediates in Mononuclear Nonheme-Iron Enzymes. *Inorg. Chem.* **2017**, *56*, 13382-13389.
45. Mitchell, A. J.; Dunham, N. P.; Martinie, R. J.; Bergman, J. A.; Pollock, C. J.; Hu, K.; Allen, B. D.; Chang, W.-c.; Silakov, A.; Bollinger Jr, J. M., Visualizing the Reaction Cycle in an Iron (II)-and 2-(Oxo)-Glutarate-Dependent Hydroxylase. *J. Am. Chem. Soc.* **2017**, *139*, 13830-13836.
46. Godfrey, E.; Porro, C. S.; de Visser, S. P., Comparative Quantum Mechanics/Molecular Mechanics (QM/MM) and Density Functional Theory Calculations on the Oxo- Iron Species of Taurine/A-Ketoglutarate Dioxygenase. *J. Phys. Chem. A* **2008**, *112*, 2464-2468.
47. Wang, B.; Usharani, D.; Li, C.; Shaik, S., Theory Uncovers an Unusual Mechanism of DNA Repair of a Lesioned Adenine by Alkb Enzymes. *J. Am. Chem. Soc.* **2014**, *136*, 13895-13901.
48. Du, L.; Gao, J.; Liu, Y.; Zhang, D.; Liu, C., The Reaction Mechanism of Hydroxyethylphosphonate Dioxygenase: A QM/MM Study. *Org. Biomol. Chem.* **2012**, *10*, 1014-1024.
49. Quesne, M. G.; Latifi, R.; Gonzalez-Ovalle, L. E.; Kumar, D.; de Visser, S. P., Quantum Mechanics/Molecular Mechanics Study on the Oxygen Binding and Substrate Hydroxylation Step in Alkb Repair Enzymes. *Chem. Eur. J.* **2014**, *20*, 435-446.
50. Proos Vedin, N.; Lundberg, M., Protein Effects in Non-Heme Iron Enzyme Catalysis: Insights from Multiscale Models. *JBIC, J. Biol. Inorg. Chem.* **2016**, *21*, 645-657.
51. Martinie, R. J.; Livada, J.; Chang, W.-c.; Green, M. T.; Krebs, C.; Bollinger Jr, J. M.; Silakov, A., Experimental Correlation of Substrate Position with Reaction Outcome in the Aliphatic Halogenase, SyrB2. *J. Am. Chem. Soc.* **2015**, *137*, 6912-6919.
52. Shane, J.; Höfer, P.; Reijerse, E.; De Boer, E., Hyperfine Sublevel Correlation Spectroscopy (Hyscore) of Disordered Solids. *J. Magn. Reson.* **1992**, *99*, 596-604.
53. Casey, T. M.; Grzyska, P. K.; Hausinger, R. P.; McCracken, J., Measuring the Orientation of Taurine in the Active Site of the Non-Heme Fe (II)/A-Ketoglutarate-Dependent Taurine Hydroxylase (Taud) Using Electron Spin Echo Envelope Modulation (Eseem) Spectroscopy. *J. Phys. Chem. B* **2013**, *117*, 10384-10394.
54. Muthukumar, R. B.; Grzyska, P. K.; Hausinger, R. P.; McCracken, J., Probing the Iron- Substrate Orientation for Taurine/A-Ketoglutarate Dioxygenase Using Deuterium Electron Spin Echo Envelope Modulation Spectroscopy. *Biochemistry* **2007**, *46*, 5951-5959.

55. Guenzi, E.; Galli, G.; Grgurina, I.; Gross, D. C.; Grandi, G., Characterization of the Syringomycin Synthetase Gene Cluster: A Link between Prokaryotic and Eukaryotic Peptide Synthetases. *J. Biol. Chem.* **1998**, *273*, 32857-32863.
56. Wattana-amorn, P.; Williams, C.; Płoskoń, E.; Cox, R. J.; Simpson, T. J.; Crosby, J.; Crump, M. P., Solution Structure of an Acyl Carrier Protein Domain from a Fungal Type I Polyketide Synthase. *Biochemistry* **2010**, *49*, 2186-2193.
57. Petřek, M.; Otyepka, M.; Banáš, P.; Košinová, P.; Koča, J.; Damborský, J., Caver: A New Tool to Explore Routes from Protein Clefts, Pockets and Cavities. *BMC Bioinf.* **2006**, *7*, 316.
58. Torabifard, H.; Cisneros, G. A., Computational Investigation of O₂ Diffusion through an Intra-Molecular Tunnel in Alkb; Influence of Polarization on O₂ Transport. *Chem. Sci.* **2017**, *8*, 6230-6238.
59. Bartlett, G. J.; Newberry, R. W.; VanVeller, B.; Raines, R. T.; Woolfson, D. N., Interplay of Hydrogen Bonds and N→ π^* Interactions in Proteins. *J. Am. Chem. Soc.* **2013**, *135*, 18682-18688.
60. Dill, K. A., Dominant Forces in Protein Folding. *Biochemistry* **1990**, *29*, 7133-7155.
61. Hinderaker, M. P.; Raines, R. T., An Electronic Effect on Protein Structure. *Protein Sci.* **2003**, *12*, 1188-1194.
62. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C., Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712-725.
63. Halgren, T. A.; Damm, W., Polarizable Force Fields. *Curr. Opin. Struct. Biol.* **2001**, *11*, 236-242.
64. Riniker, S., Fixed-Charge Atomistic Force Fields for Molecular Dynamics Simulations in the Condensed Phase: An Overview. *J. Chem. Inf. Model.* **2018**, *58*, 565-578.
65. Patra, N.; Ioannidis, E. I.; Kulik, H. J., Computational Investigation of the Interplay of Substrate Positioning and Reactivity in Catechol O-Methyltransferase. *PLoS ONE* **2016**, *11*, e0161868.
66. Zhang, J.; Kulik, H. J.; Martinez, T. J.; Klinman, J. P., Mediation of Donor–Acceptor Distance in an Enzymatic Methyl Transfer Reaction. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 7954-7959.
67. Zhang, Z.; Ren, J.-s.; Harlos, K.; McKinnon, C. H.; Clifton, I. J.; Schofield, C. J., Crystal Structure of a Clavaminate Synthase–Fe(II)–2-Oxoglutarate–Substrate–NO Complex: Evidence for Metal Centred Rearrangements. *FEBS Lett.* **2002**, *517*, 7-12.
68. Dunham, N. P.; Chang, W.-c.; Mitchell, A. J.; Martinie, R. J.; Zhang, B.; Bergman, J. A.; Rajakovich, L. J.; Wang, B.; Silakov, A.; Krebs, C., et al., Two Distinct Mechanisms for C–C Desaturation by Iron(II)- and 2-(Oxo)Glutarate-Dependent Oxygenases: Importance of A-Heteroatom Assistance. *J. Am. Chem. Soc.* **2018**, *140*, 7116-7126.
69. Bader, R. F., A Quantum Theory of Molecular Structure and Its Applications. *Chem. Rev.* **1991**, *91*, 893-928.
70. Blanco, S.; Pinacho, P.; López, J. C., Structure and Dynamics in Formamide–(H₂O)₃: A Water Pentamer Analogue. *J. Phys. Chem. Lett.* **2017**, *8*, 6060-6066.
71. Schrodinger, L. L. C., The PyMOL Molecular Graphics System, Version 1.7.4.3. 2010.
72. Kelley, L. A.; Mezulis, S.; Yates, C. M.; Wass, M. N.; Sternberg, M. J. E., The Phyre2 Web Portal for Protein Modeling, Prediction and Analysis. *Nat. Protoc.* **2015**, *10*, 845.
73. Anandakrishnan, R.; Aguilar, B.; Onufriev, A. V., H++ 3.0: Automating pK Prediction and the Preparation of Biomolecular Structures for Atomistic Molecular Modeling and Simulations. *Nucleic Acids Res.* **2012**, *40*, W537-W541.
74. Gordon, J. C.; Myers, J. B.; Foltz, T.; Shojha, V.; Heath, L. S.; Onufriev, A., H++: A Server for Estimating pK_as and Adding Missing Hydrogens to Macromolecules. *Nucleic Acids Res.* **2005**, *33*, W368-W371.

75. Myers, J.; Grothaus, G.; Narayanan, S.; Onufriev, A., A Simple Clustering Algorithm Can Be Accurate Enough for Use in Calculations of pKs in Macromolecules. *Proteins: Struct., Funct., Bioinf.* **2006**, *63*, 928-938.
76. D.A. Case, J. T. B., R.M. Betz, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, T. Luchko, R. Luo, B. Madej, K.M. Merz, G. Monard, P. Needham, H. Nguyen, H.T. Nguyen, I. Omelyan, A. Onufriev, D.R. Roe, A. Roitberg, R. Salomon-Ferrer, C.L. Simmerling, W. Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, D.M. York and P.A. Kollman Amber 2015, University of California, San Francisco. 2015.
77. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157-1174.
78. Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A., A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem.* **1993**, *97*, 10269-10280.
79. Harihara, P. C.; Pople, J. A., Influence of Polarization Functions on Molecular-Orbital Hydrogenation Energies. *Theor. Chim. Acta* **1973**, *28*, 213-222.
80. Gordon, M. S.; Schmidt, M. W., Advances in Electronic Structure Theory: GAMESS a Decade Later. *Theory Appl. Comput. Chem.: First Forty Years* **2005**, 1167-1189.
81. F. Wang, J.-P. B., P. Cieplak and F.-Y. Dupradeau R.E.D. Python: Object Oriented Programming for Amber Force Fields, Université De Picardie - Jules Verne, SanfordlBurnham Medical Research Institute, Nov. 2013. <http://q4md-forcefieldtools.org/REDServer-Development/> (accessed 10/15/18).
82. Vanquelef, E.; Simon, S.; Marquant, G.; Garcia, E.; Klimerak, G.; Delepine, J. C.; Cieplak, P.; Dupradeau, F.-Y., R.E.D. Server: A Web Service for Deriving RESP and ESP Charges and Building Force Field Libraries for New Molecules and Molecular Fragments. *Nucleic Acids Res.* **2011**, *39*, W511-W517.
83. Dupradeau, F.-Y.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak, P., The R.E.D. Tools: Advances in RESP and ESP Charge Derivation and Force Field Library Building. *Phys. Chem. Chem. Phys.* **2010**, *12*, 7821-7839.
84. Joung, I. S.; Cheatham, T. E., 3rd, Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations.
85. Li, P.; Song, L. F.; Merz, K. M., Parameterization of Highly Charged Metal Ions Using the 12-6-4 Lj-Type Nonbonded Model in Explicit Water. *J. Phys. Chem. B* **2015**, *119*, 883-895.
86. Vazdar, M.; Pluhařová, E.; Mason, P. E.; Vácha, R.; Jungwirth, P., Ions at Hydrophobic Aqueous Interfaces: Molecular Dynamics with Effective Polarization. *J. Phys. Chem. Lett.* **2012**, *3*, 2087-2091.
87. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926-935.
88. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C., Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of N-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327-341.
89. Souaille, M.; Roux, B. t., Extension to the Weighted Histogram Analysis Method: Combining Umbrella Sampling with Free Energy Calculations. *Comput. Phys. Commun.* **2001**, *135*, 40-57.
90. Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A., The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* **1992**, *13*, 1011-1021.
91. Grossfield, A. "Wham: The Weighted Histogram Analysis Method", Version 2.0.9 <http://membrane.urmc.rochester.edu/content/wham> (accessed 10/15/18).
92. Torrie, G. M.; Valleau, J. P., Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977**, *23*, 187-199.

93. Shao, J.; Tanner, S. W.; Thompson, N.; Cheatham, T. E., Clustering Molecular Dynamics Trajectories: 1. Characterizing the Performance of Different Clustering Algorithms. *J. Chem. Theory Comput.* **2007**, *3*, 2312-2334.
94. Miller III, B. R.; McGee Jr, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E., Mmpbsa. Py: An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.* **2012**, *8*, 3314-3321.
95. Massova, I.; Kollman, P. A., Combined Molecular Mechanical and Continuum Solvent Approach (MM-PBSA/GBSA) to Predict Ligand Binding. *Perspect. Drug Discovery Des.* **2000**, *18*, 113-135.
96. Tsui, V.; Case, D. A., Theory and Applications of the Generalized Born Solvation Model in Macromolecular Simulations. *Biopolymers* **2000**, *56*, 275-291.
97. Onufriev, A.; Bashford, D.; Case, D. A., Exploring Protein Native States and Large-Scale Conformational Changes with a Modified Generalized Born Model. *Proteins: Struct., Funct., Bioinf.* **2004**, *55*, 383-394.
98. Hou, T.; Wang, J.; Li, Y.; Wang, W., Assessing the Performance of the MM/Pbsa and MM/Gbsa Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2011**, *51*, 69-82.
99. Ufimtsev, I. S.; Martínez, T. J., Quantum Chemistry on Graphical Processing Units. 3. Analytical Energy Gradients, Geometry Optimization, and First Principles Molecular Dynamics. *J. Chem. Theory Comput.* **2009**, *5*, 2619-2628.
100. Petachem. <http://www.petachem.com>. (accessed 10/15/18).
101. Kulik, H. J.; Luehr, N.; Ufimtsev, I. S.; Martinez, T. J., Ab Initio Quantum Chemistry for Protein Structures. *J. Phys. Chem. B* **2012**, *116*, 12501-12509.
102. Karelina, M.; Kulik, H. J., Systematic Quantum Mechanical Region Determination in QM/MM Simulation. *J. Chem. Theory Comput.* **2017**, *13*, 563-576.
103. Qi, H. W.; Karelina, M.; Kulik, H. J., Quantifying Electronic Effects in QM and QM/MM Biomolecular Modeling with the Fukui Function. *Acta Phys. -Chim. Sin.* **2018**, *34*, 81-91.
104. Kulik, H. J.; Zhang, J.; Klinman, J. P.; Martinez, T. J., How Large Should the QM Region Be in QM/MM Calculations? The Case of Catechol O-Methyltransferase. *J. Phys. Chem. B* **2016**, *120*, 11381-11394.
105. Rohrdanz, M. A.; Martins, K. M.; Herbert, J. M., A Long-Range-Corrected Density Functional That Performs Well for Both Ground-State Properties and Time-Dependent Density Functional Theory Excitation Energies, Including Charge-Transfer Excited States. *J. Chem. Phys.* **2009**, *130*, 054112.
106. Hay, P. J.; Wadt, W. R., Ab Initio Effective Core Potentials for Molecular Calculations. Potentials for the Transition Metal Atoms Sc to Hg. *J. Chem. Phys.* **1985**, *82*, 270-283.
107. Lu, T.; Chen, F., Multiwfn: A Multifunctional Wavefunction Analyzer. *J. Comput. Chem.* **2012**, *33*, 580-592.
108. Espinosa, E.; Molins, E.; Lecomte, C., Hydrogen Bond Strengths Revealed by Topological Analyses of Experimentally Observed Electron Densities. *Chem. Phys. Lett.* **1998**, *285*, 170-173.
109. Klamt, A.; Schuurmann, G., Cosmo: A New Approach to Dielectric Screening in Solvents with Explicit Expressions for the Screening Energy and Its Gradient. *J. Chem. Soc., Perkin Trans. 2* **1993**, *2*, 799-805.
110. Liu, F.; Luehr, N.; Kulik, H. J.; Martínez, T. J., Quantum Chemistry for Solvated Molecules on Graphical Processing Units Using Polarizable Continuum Models. *J. Chem. Theory Comput.* **2015**, *11*, 3131-3144.

TOC GRAPHIC

