PCCP

The Infuence of Model Building Schemes and Molecular Dynamics Sampling on QM-cluster Models: The Chorismate Mutase Case Study

Journal:	*Physical Chemistry Chemical Physics*
Manuscript ID	CP-ART-12-2023-006100.R2
Article Type:	Paper
Date Submitted by the Author:	21-Mar-2024
Complete List of Authors:	Agbaglo, Donatus; The University of Memphis,, Chemistry Summers, Thomas; The University of Memphis, Department of Chemistry Cheng, Qianyi; The University of Memphis,, Department of Chemistry DeYonker, Nathan; The University of Memphis,, Chemistry

The Influence of Model Building Schemes and Molecular Dynamics Sampling on QM-cluster Models: The Chorismate Mutase Case Study

Donatus A. Agbaglo,[†] Thomas J. Summers,[†,‡] Qianyi Cheng,[†] and Nathan J. DeYonker[*,†]

†*Department of Chemistry, University of Memphis, Memphis, TN 38152, U.S.A.*

‡*Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, U.S.A.*

E-mail: ndyonker@memphis.edu

Abstract

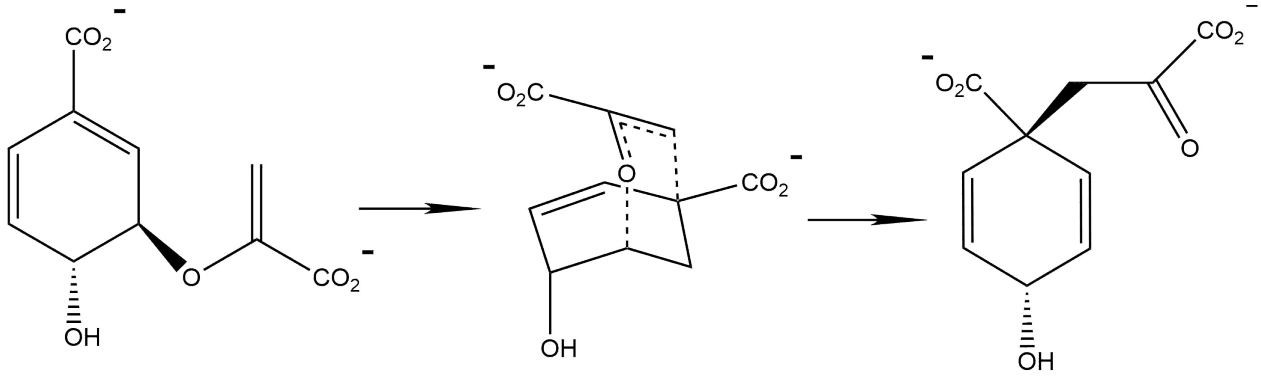
Most QM-cluster models of enzymes are constructed based on X-ray crystal structures, which limits comparison to *in vivo* structure and mechanism. The active site of chorismate mutase from *Bacillus subtilis* and the enzymatic transformation of chorismate to prephenate is used as a case study to guide construction of QM-cluster models built first from the X-ray crystal structure, then from molecular dynamics (MD) simulation snapshots. The Residue Interaction Network-based ResidUe Selector (*RINRUS*) software toolkit, developed by our group to simplify and automate the construction of QM-cluster models, is expanded to handle MD to QM-cluster model workflows. Several options, some employing novel topological clustering from Residue Interaction Network (RIN) information, are evaluated for generating conformational clustering from MD simulation. *RINRUS* then generates a statistical thermodynamic framework for QM-cluster modeling of the chorismate mutase mechanism via refining 250

MD frames with Density Functional Theory (DFT). The 250 QM-cluster models sampled provide a mean $\Delta G^{\ddagger}$ of 10.3 ± 2.6 kcal mol^{-1} compared to the experimental value of 15.4 kcal mol^{-1} at 25 ^{0}C. While the difference between theory and experiment is consequential, the level of theory used is modest and therefore "chemical" accuracy is unexpected. More important are the comparisons made between QM-cluster models designed from the X-ray crystal structure versus those from MD frames. The large variations in kinetic and thermodynamic properties arise from geometric changes in the ensemble of QM-cluster models, rather from the composition of the QM-cluster models or from the active site-solvent interface. The findings open the way for further quantitative and reproducible calibration in the field of computational enzymology using the model construction framework afforded with the *RINRUS* software toolkit.

Introduction

Through multiscale QM/MM or QM-only "cluster model" studies, stationary points along a reaction mechanism can be optimized, which allows a structural probe of the enzyme kinetics that is impossible to directly observe experimentally.[1] As the reliability of computational enzymology and the tractable size of QM-regions increase, a greater focus on cyberinfrastructure is required for building consistent and reproducible atomic-level enzyme models. Our group has developed the Residue Interaction Network ResidUe Selector (*RINRUS*) software toolkit to facilitate studying the reaction mechanisms of enzymes with quantum chemistry.[2–4] Instead of relying on chemical intuition or distance-based criteria to prioritize the critical fragments within the enzyme active site, *RINRUS* algorithmically constructs enzyme models based on several possible qualitative and quantitative criteria. *RINRUS* infrastructure was first developed to build QM-cluster models of enzymes, but adapting the code to also build QM/MM enzyme models is in progress. In this work, we explore the enzyme chorismate mutase in conjunction with a proof-of-concept expansion of *RINRUS* capabilities: interfacing

QM-cluster modeling with Molecular Dynamics (MD) techniques.



Scheme 1: Schematic representation of the Claisen rearrangement of chorismate to prephenate

Chorismate mutase (CM) catalyzes the reaction of chorismate to prephenate, participating in the shikimate pathway that biologically produces phenylalanine and tyrosine amino acids (Scheme 1).[5–14] The shikimate pathway does not occur in the animal kingdom, and thus provides a target for the development of new antibiotics, fungicides, and herbicides.[15] While chorimsate mutase has been widely studied experimentally and computationally, there are still mysteries to be unraveled with respect to the extraordinary kinetic enhancement of its active site. The chorismate mutase enzymatic reaction promotes a 10^6-fold rate acceleration of prephenate production through a Claisen rearrangement in the catalytic elementary step.[16,17] This Claisen rearrangement is one of the few known examples of a naturally-occurring catalyzed pericyclic reaction.[18] In 1993, Lipscomb and coworkers published an X-ray crystal structure of *Bacillus subtilis* chorismate mutase (BsCM, PDB: 2CHT) at 2.2 Å resolution that forms the basis of most theoretical works.[19] This structure contains an endo-oxabicyclic transition state analogue (TSA), 8-hydroxy-2-oxa-bicyclo[3.3.1]non-6-ene-3,5-dicarboxylic acid, which offered structural insight into the enzyme mechanism. Since the pericyclic reaction does not involve covalent substrate-protein bonding or acid-base chemistry, CM makes an intriguing, and in some respects, simplified case study of enzyme catalysis.[20,21]

Mutagenesis, computational enzymology, and biochemical kinetics have been indispens-

able tools to study the mechanism of the CM reaction, especially for exploring the transition state stabilization (TSS) and near-attack conformation (NAC) hypotheses or for describing manifestations of their complementary kinetic and thermodynamic behavior.[9,18,22–25] Mutagenesis experiments of *Escherichia coli* chorismate mutase (EcCM)[26–30] and BsCM revealed the catalytic importance of many charged active site residues for establishing hydrogen bonding with the negatively charged substrate. For example, replacement of Arg90 with a positively-charged lysine still decreases the catalytic efficiency by at least three orders of magnitude in BsCM.[19,31]

Theoretical studies of chorismate mutase with QM/MM-MD first emphasized the importance of a near attack conformation (NAC) as the main catalytic driving power behind the proposed mechanism.[20,32,33] Studies done by Bruice and co-worker showed that NAC rearrangement of chorismate structure is a result of activated carbon and oxygen ligand atoms approaching within the van der Waals contact distance at very small bond angles, creating a favorable orientation of π-orbital overlap.[20,32] The proponents of the NAC hypothesis focus on geometric distortion of the substrate in the active site. However, those who argue for the TSS hypothesis indicate that positively charged residues like Lys39 in EcCM and Arg90 in BsCM stabilize the developing negative charge during bond breaking at the ether oxygen.[25,31,34] Bond-breaking then leads to electrostatic stabilization of active site residues, lowering the activation energy. Subsequent QM/MM and QM-cluster model calculations have provided evidence that catalysis is due to both near attack conformation and transition state stabilization, but with TSS being the main driving force of the proposed mechanism.[21,22,35]

While computational enzymology has advanced rapidly over the last two decades,[6,36–38] one persistent challenge in this research area is designing effective QM-regions that reliably predict catalytic activity, with kinetic and thermodynamic properties that can converge quickly with respect to model size. *Ad hoc* methods of selecting residues for inclusion in the QM regions of QM/MM models or in QM-cluster models are poorly reproducible and

not well calibrated. One technique for QM region selection is to include all residues that are within a specific radial distance from the center of the active site or from the center of mass of the substrate. This construction paradigm is based on the idea that spherical active site models are appropriate. Several studies by our group and others reveal that this is not always the case,[3,4,39–46] though CM active sites are known to be fairly compact and spherical.

To facilitate improved benchmarking in computational enzymology, our group has created the *RINRUS* software toolkit to automate the process of generating QM-cluster models. Our goal is to address various community-wide challenges in computational enzymology, such as standardizing QM-cluster (and eventually QM/MM) model construction, lowering the learning curve for new users, and reducing trial and error caused by *ad hoc* model-building schemes. *RINRUS* uses an automated approach to trim and cap the active site fragments. With a given protein structure and a user-defined "seed", which consists of the substrate and any active site fragments necessary to describe the chemical reaction, *RINRUS* identifies proximal fragments that have important non-covalent interactions with the seed using the graph theory concept of the Residue Interaction Network (RIN).[47,48]

To summarize the *RINRUS* procedure, a protein structure is converted into a RIN graph composed of only a subset of the fragments (referred to as "nodes" in graph theory) that have an identifiable electrostatic and/or steric interaction (referred to as "edges" in graph theory) with the seed nodes.[47,48] The RIN is then processed using one or more user-selected schemes that identify qualitative interaction types (Structural Interaction Fingerprints, SIFs)[49] or quantitative schemes that utilize first-principles interaction energies like symmetry-adapted perturbation theory (SAPT or F/I-SAPT, see below).[50–53] *RINRUS* can also be used to rank fragments via distance-based criteria. Once a ranking scheme is chosen and fragment rank is enumerated, *RINRUS* will algorithmically construct QM-cluster models and provide input files formatted appropriately for several commercial and open-source quantum chemistry software packages.

This work has two major objectives. First, we analyze how specific residues influence the

enzymatic reaction and contribute to the convergence of *RINRUS*-built QM-cluster models of chorismate mutase. Multiple fragment ranking schemes are explored and compared, with models built incrementally, growing by one fragment at a time. Second, we explore QM-cluster modeling in a quasi time-dependent fashion by sampling MD snapshots with refined QM-cluster models to account for conformational averaging. Thermally stable conformational change is one of the most important aspects of regulating protein structure and activity, and conformational sampling of enzymes is typically probed on the micro-second time scale via MD simulations.[54] We have selected 250 snapshots from a 20 ns MD simulation of BsCM and processed each with *RINRUS* to obtain 250 different QM-cluster models. The catalytic transition state for each of the 250 QM-cluster models is optimized, and via computation of the connected reactant and product structures, kinetic and thermodynamic data is obtained.

Methods

All computations were based on the X-ray crystal structure of the *Bacillus subtilis* chorismate mutase taken from PDB entry 2CHT. The 2CHT enzyme is trimeric with three active sites formed at the interface of adjacent monomer chains. The active site of the crystallographic A/C chain was used for QM-model construction in this work. Further justification of using the chain A/C interface is provided in the Supporting Information. Hydrogen atoms were added to the enzyme using the *reduce* program.[55] For all QM-cluster models and MD simulations, the TSA found in the crystallographic active sites was replaced with the native substrate (chorismate).

Incremental QM-cluster model building with *RINRUS*

RINRUS identifies and ranks inter-residue interactions based upon two existing packages that compute the RIN and output node/edge information in machine and human-readable

formats. *Probe*[56] rolls a small sphere over the internal van der Waals surface of a protein structure to identify and classify non-covalent interatomic interactions between fragments of a protein structure; *arpeggio*[57] uses interatomic distance and angle criteria to identify and classify interactions. Throughout this work, "seed", "substrate", and "ligand" synonymously refer to the chorismate molecule shown in Figure S1.

A good fragment ranking scheme is needed to design reliable QM-cluster models, which is a core feature of the open-source *RINRUS* package.[58] There are three different fragment ranking schemes being tested in this work. The *RINRUS-probe* workflow ranks the importance of active site fragments based on the number of contact counts between each fragment and the seed. When incrementally building models, fragments (categorized as residue side chains, residue main chains, or solvent water molecules) are added to the model one at a time in order from the fragment with the highest number of contacts with substrate to the lowest. While *probe* parses interaction types into five simple SIF categories, *arpeggio* classifies fourteen different chemical interaction type, based on the CREDO set of protein-substrate interactions.[59] While *arpeggio* also accounts for typical interaction types like hydrogen bonding and hydrophobic contacts, it can also more flexibly account for weaker inter-residue interactions such as aromatic π-stacking or less common interactions such as halogen bonds. It should be noted that the proximal interactions computed by *arpeggio* are ignored in this study because the focus is on fragments that have recognized intermolecular forces with the chorismate substrate, rather than distance-based metrics.

Symmetry adapted perturbation theory (SAPT) has become an increasingly popular approach for computing non-covalent interaction energies between two molecules or fragments.[50–52,60–63] SAPT calculations are especially useful in that the interaction energies are readily decomposed into electrostatic, exchange-repulsion, induction, and dispersion components. Functional-group SAPT (F-SAPT)[52] is an extension of SAPT that provides an effective secondary two-body partition of the SAPT components. This additional partitioning allows computation of interaction energy between a fragment A (in this case study, the

chorismate ligand) and user-defined sub-fragments of a fragment B (the various side chain and backbone fragments of the active site). F-SAPT is leveraged to decompose the interaction energy between chorismate and individual residue main chains or side chains, without cutting or capping fragments differently from what is used in the parent QM-cluster models. We will use the F-SAPT interaction energies between chorismate and surrounding residue fragments to rank incremental QM-cluster model building. This work uses the zeroth-order formulation of F-SAPT, F-SAPT0, described by the equation:

$$E_{int} = E^{(1)}_{elec} + E^{(1)}_{exh} + [E^{(2)}_{ind} + E^{(2)}_{exch-ind} + \delta E^{(2)}_{HF}]_{ind} + [E_{disp} + E^{(2)}_{exch-disp}]_{disp} \quad (1)$$

F-SAPT0 computations employed the jun-cc-pVDZ basis set[51,52] for all atoms and frozen core electrons via the *PSI4* v1.3 package.[64] The jun-cc-pVDZ basis set has been demonstrated to provide reliable SAPT interaction energies.[65]

In recent work, a poor correlation between number of probe contacts and F-SAPT interaction energies was observed.[66,67] We then hypothesized that F-SAPT interaction energies will be a more quantitatively reliable metric for ranking the importance of active site residues. However, SAPT calculations are computationally expensive (days of CPU time) compared to the near-negligible effort required to compute and parse a RIN from *probe* or *arpeggio* ranking ($<$ 20 seconds of CPU time).

QM-cluster models were generated using the *RINRUS* software.[58] Trimming of residue fragments is performed algorithimcally by *RINRUS* depending on if the backbone NH, backbone CO, and/or side chain of a residue has interatomic contacts with chorismate. Where covalent bonds are broken in the trimming procedure (typically across C_α atoms), *RINRUS* automatically adds hydrogen atoms to satisfy carbon valency. We refer throughout to the QM-cluster model that contains all fragments with a quantifiable interaction with the chorismate ligand as a "maximal model". Trimming details for the maximal model of the X-ray crystal structure active site are shown in Table S1. To maintain the general shape and mimic

the semi-rigid character of the protein tertiary structure, all C_α atoms, along with the C_β atoms of any Arg, Lys, Glu, Met, Trp, and Phe side chains were frozen to their crystallographic positions (if obtained from the X-ray crystal structure) or frozen at their positions in the respective MD frame (if obtained from MD simulation). All chorismate atoms were unconstrained in the QM-cluster model computations.

The QM computations were carried out using the Gaussian16 software package.[68] The geometries of the models were optimized using density functional theory (DFT) with the B3LYP exchange-correlation functional.[69,70] The 6-31G(d′) basis set was used for N, O, and S,[71] and the 6-31G basis set was used for C and H atoms.[60] The Grimme D3 (Becke-Johnson) dispersion correction (GB3BJ) was also included,[72] along with a conductor-like polarizable continuum model (CPCM) using UAKS sets of atomic radii, a non-default electronic scaling factor of 1.2, and default cavity parameters for water but with an attenuated dielectric constant of $\varepsilon = 4$.[73,74] Transition states were located for the elementary step of the proposed mechanism, and the reactants and products were then confirmed by following the intrinsic reaction coordinate (IRC). [1] The zero-point energies (ZPE) and thermal enthalpy/free energy corrections were calculated at 1 atm and 298.15 K.

MD trajectory-based QM-cluster models

For the MD simulations, some pre-processing of the X-ray crystal structure was necessary.[67] Missing residues in the 2CHT X-ray crystal structure were added from the C-terminus using PDB entry 1DBF,[75] a BsCM structure without substrate or TSA in complex with the protein. The two structures were globally aligned and atomic coordinates from 1DBF were added to the 2CHT structure based upon the point where the two structures begin a common structural alignment. Specifically, residues 1 and 116-127 from 1DBF were added to 2CHT for chain A, residues 1 and 115-127 were added for chain B, and residues 1-2 and 115-127

[1] It is important to note that our group employs the "freeze code" scheme in Gaussian16, in which all Hessian elements are zero when two frozen Cartesian coordinates are involved. The phenomenon in which several small magnitude imaginary vibrational frequencies appear in thermochemical analysis does not occur in our treatment of the Hessian matrix.

were added for chain C (with residues 2 and 115-119 of 2CHT chain C being replaced with the corresponding coordinates from 1DBF). Hydrogen atoms were added to this structure via the H++ server using default parameters.[76] The native substrate chorismate in a pre-reactive conformation was used in MD simulations instead of the TSA. The AMBER18 MD package[77] was used to run the MD simulations, and the AMBER force field ff14SB was used with periodic boundary conditions and a cutoff value of 9 Å for non-bonded interactions. The Antechamber package was employed to parameterize the chorismate substrate with the Generalized Amber Force Field (GAFF).[77,78] The protonated structure with chorismate was solvated in a cubic 10 Å box of water with the explicit solvent model TIP3P.[79] The MD model charge was neutralized by adding 9 Na^+ ions.[80]

An energy minimization of the system was first carried out with protein heavy atoms constrained to their crystallographic coordinates using a harmonic positional restraint (k_{pos}) of 200 kcal $mol^{-1}/Å^2$ allowing the solvent bath to be initially relaxed and the hydrogen bonding networks to be established. The protein heavy atom constraints were then iteratively relaxed over five 20 ps simulations using Langevin dynamics under constant-temperature, constant-pressure (NPT) conditions at 300 K and 1 atm; the SHAKE algorithm[81] was used to constrain all bonds involving hydrogen atoms for the initial equilibration simulation. The protein was then allowed to move freely for a 20 ns production-level run. The timescale of each frame was 1 ps, for a total of 20,000 frames. The protein RMSDs of MD trajectories were calculated using the *cpptraj* module of AMBER18.[82]

Schemes for selection of frames for the QM-cluster models from MD trajectories

Designing QM-cluster models from a large number of MD frames will allow consideration of conformational influence on kinetic and thermodynamic quantities. Eight schemes are considered in an attempt to cover a diverse sampling of conformations and non-equilibrium structures. From each scheme, 20 to 40 MD frames are selected and then used to construct

a QM-cluster model of the active site. The first scheme considered ($\mathbf{S}_1$) is perhaps the most common scheme for MD simulation sampling, and involves selecting MD frames at equal intervals over the course of an equilibrated simulation. This approach is effectively random and unbiased. For the next set of schemes ($\mathbf{S}_2$, $\mathbf{S}_3$, and $\mathbf{S}_4$) we chose frames similar to the X-ray crystal structure. Furthermore, it may be better to consider only the structural variations of the active site residues rather than of the whole protein, and this idea is incorporated into $\mathbf{S}_3$, $\mathbf{S}_4$, $\mathbf{S}_6$, $\mathbf{S}_7$, and $\mathbf{S}_8$. For the final set of schemes ($\mathbf{S}_5$, $\mathbf{S}_6$, $\mathbf{S}_7$, and $\mathbf{S}_8$) frames were grouped by a specific metric and then k-means clustering divided the frames into 3 or 4 clusters. These schemes should increase the structural diversity of QM-cluster model refinement. Again, note that the Chain A/C interface was used to construct the QM-cluster models from each selected MD frame. Detailed frame selection criteria are as follows.

$\mathbf{S}_1$ - Twenty frames were selected from the MD simulation at equal intervals of 1,000 ps over the entire 20 ns equilibrated simulation.

$\mathbf{S}_2$ - The RMSD of the backbone atoms (C, O, C_α, N, and H) of the entire protein structure compared to the X-ray crystal structure was measured for each frame. Frames with an RMSD within $\pm$ 1 standard deviation (0.76 Å) of the mean RMSD (2.66 Å) were isolated, and a random number generator was used to select 30 frames from this data set.

$\mathbf{S}_3$ - The RMSD of the backbone atoms of a selection of active site residues compared to the X-ray crystal structure was measured for each frame. The subset of active site residues was defined as all residues present in any of the QM-cluster models obtained from $\mathbf{S}_1$: Arg7, Glu78, Arg90, Tyr108, Leu115, Phe57, Ala59, Lys60, Arg63, Val73, Thr74, and Cys75. Frames with an active site backbone RMSD within $\pm$ 1 standard deviation (0.09 Å) of the mean RMSD (0.84 Å) were isolated, and a random number generator was used to select 30 frames from this data set.

$\mathbf{S}_4$ - This scheme used the RMSD of the side chain atoms of the active site residues (listed in $\mathbf{S}_3$) compared to the X-ray crystal structure. Frames with a side chain backbone RMSD within $\pm$ 1 standard deviation (0.16 Å) of the mean RMSD (1.66 Å) were isolated, and a

random number generator was used to select 30 frames from this data set.

$\mathbf{S}_5$ - The RMSD of all heavy (non-hydrogen) atoms of protein and chorismate compared to the X-ray crystal structure was measured for each frame. K-means clustering was used to group the frames into three distinct clusters based on the gap statistic and elbow plots shown in Figure S2, and a random number generator was used to select 10 frames from each of the three clusters.

$\mathbf{S}_6$ - The RMSD of the backbone atoms of only the active site residues (from $\mathbf{S}_3$) compared to the X-ray crystal structure was measured for each frame. Based on analysis of the RMSD using the gap statistic and elbow plots in Figure S3, it became apparent that there is only one unique k-means cluster. We then subdivided the data into four clusters and randomly selected 10 frames from each of the four clusters.

$\mathbf{S}_7$ - This scheme used the RMSD of the side chain atoms of the active site residues compared to the X-ray crystal structure instead of the backbone atoms. Similar to $\mathbf{S}_6$, k-means clustering with the active site side chain atom RMSD values was not a useful technique (Figure S4). The MD frames were still split into another four arbitrary clusters and randomly selected to provide an unbiased sampling of 40 additional MD frames.

$\mathbf{S}_8$ - The number of *probe* contacts between chorismate and surrounding residues was measured for each frame of the MD trajectory. K-means clustering grouped the frames into distinct clusters. However, the gap statistics and elbow plots shown in Figure S5 indicate our MD frames are not easily clustered into less than 10 sets, so the clustering is truncated at k = 3. A random number generator was used to randomly select 10 frames from each of the three clusters.

From the eight selection schemes, a total of 250 unique MD frames were chosen and then refined into QM-cluster models generated by *RINRUS*. Note that the composition of the QM-cluster models is not uniform. Each QM-cluster model constructed from MD includes all fragments recognized by the *probe* software as having inter-residue interactions with chorismate *for that specific MD frame*. Interestingly, nearly adjacent and even adjacent

frames that were selected by the various schemes showed non-uniform RIN composition [frames 159 ($\mathbf{S}_8$) and 161 ($\mathbf{S}_7$); 1218 ($\mathbf{S}_6$) and 1221 ($\mathbf{S}_7$); 2473 ($\mathbf{S}_8$) and 2475 ($\mathbf{S}_6$); 7603 ($\mathbf{S}_6$) and 7607 ($\mathbf{S}_5$); 9748 ($\mathbf{S}_6$) and 9750 ($\mathbf{S}_2$); 12378 ($\mathbf{S}_6$) and 12379 ($\mathbf{S}_5$), 19719 ($\mathbf{S}_2$) and 19721 ($\mathbf{S}_7$)]. Active site RIN composition of the adjacent frames 12378 and 12379 is shown in Table S2.

Results and Discussion

Building QM-cluster models with different ranking schemes

We began by examining how different schemes to prioritize residue interactions affect the construction of QM-cluster models and the convergence of predicted reaction properties. Full information about model size, model charge, and kinetic and thermodynamic properties are provided for all iterative building schemes in Tables S1 and S3. Using the X-ray crystal structure, 12 QM-cluster models were built by incrementally adding one residue based on their cumulative *probe* contact counts with chorismate within the X-ray crystal structure. The maximal *probe*-derived model, which includes all residues with any *probe* contact with the chorismate, contains 203 atoms and is shown in Figure 1.

The computed $\Delta G^{\ddagger}$ and ΔG_{rxn} values for the Claisen rearrangement reaction are plotted in Figure 2 for the *probe*-based model building scheme. First, the highest *probe*-ranked fragment, Phe57 side chain, is added to the chorismate to constitute the first QM-cluster model, containing 42 atoms and a total model charge of -2. The second QM-cluster model is generated by adding the Arg7 side chain to the first model. The second model now contains 64 atoms and a total model charge of -1. The list of fragments provided in Table S1 gives details about subsequent models, culminating in our "maximal model" of 203 atoms.

For this study, we define models as being converged for a given building scheme if *both* $\Delta G^{\ddagger}$ and ΔG_{rxn} of a model *and all subsequent larger models* are within $\pm$ 1 (tight convergence criteria) or $\pm$ 3 kcal mol^{-1} (loose convergence criteria) of their reference values ($\Delta G^{\ddagger}$ and

ΔG_{rxn} of the maximal model), respectively. The maximal *probe*-based *RINRUS*-designed QM-cluster model has values of $\Delta G^{\ddagger}$ = 9.1 kcal mol^{-1} and ΔG_{rxn} = -16.3 kcal mol^{-1}. As the size of the model increases, the predicted $\Delta G^{\ddagger}$ and ΔG_{rxn} become converged at the 155-atom model within the defined metric of convergence of ± 3 kcal mol^{-1}, after Arg63 was added to the 133-atom model. None of the *probe* models converge both $\Delta G^{\ddagger}$ and ΔG_{rxn} to within 1.0 kcal mol^{-1} of the maximal QM-cluster model. Overall, the ΔG_{rxn} value qualitatively agrees with other QM-cluster model and QM/MM studies of the chorismate mutase catalytic step that exhibited strongly exergonic reaction free energies.[5,12,16,22,35,83,84] Below, we will explain why the computed $\Delta G^{\ddagger}$ of converged and maximal QM-cluster models is significantly lower than the known experimental value.

The *Arpeggio* interaction classification may be more robust than *probe* in that it is not inherently limited to only the local interatomic contacts. Indeed, all residues identified in the *probe* ranking scheme are included in the *arpeggio* ranking scheme in addition to Glu78, as well as the side chains of Val73, where only backbone atoms had been included in the *probe*-based models as shown in Table S4. Due to differences in which specific residue atoms interact with chorismate, some fragments in the *arpeggio*-based models are trimmed and capped differently. Additional details of the *arpeggio* trimming scheme are shown in Table S1. The maximal *arpeggio*-based model has 245 atoms, which makes it somewhat larger than the maximal *probe*-based model (203 atoms). Figure 3 shows the computed values of $\Delta G^{\ddagger}$ and ΔG_{rxn} when employing the *arpeggio*-based RIN to construct the QM-cluster models. The maximal *arpeggio*-based model (used as the reference for convergence tests) has $\Delta G^{\ddagger}$ = 10.2 kcal mol^{-1} and ΔG_{rxn} = -16.1 kcal mol^{-1}.

Arpeggio-based models predict satisfactory convergence for $\Delta G^{\ddagger}$ and ΔG_{rxn} (Figure 3, magenta plot) once QM-cluster models are larger than 200 atoms. However, we see a dramatic disruption of convergence in the reaction free energy when Thr74 is added to form the 136-atom model. The computed ΔG_{rxn} of -36.3 kcal mol^{-1} is artificially too negative because the chorismate translates far out of the active site in the optimized product struc-

ture. Once Arg90 is added to form the 158-atom model, the chorismate is properly posed; all substrate-arginine hydrogen bonds seen in the maximal model are accounted for. While no *arpeggio*-based models have both ΔG_{rxn} and $\Delta G^{\ddagger}$ converged within ± 1 kcal mol^{-1} of the maximal model, convergence to the looser ± 3 kcal mol^{-1} threshold appears once the 177-atom model is constructed.

One limitation with the *arpeggio* scheme is the more frequent occurrence of tie scores for the number of interaction counts. While the number of *probe* contacts can vary over 3–4 orders of magnitude as it is linked to the continuous inter-residue surface area, *arpeggio* interaction count scores will be much smaller as values arise from summing the categorical presence/absence of interaction types. The *RINRUS* code does not yet preferentially discern between fragments with tied rankings, so there is no chemical significance to the output ordering for those residues. However, depending on which residues are selected in a tie situation, the convergence of $\Delta G^{\ddagger}$ and ΔG_{rxn} values can be affected.

In situations where there is a tie in the number of *arpeggio* contact counts, we have manually reordered the *RINRUS* ranking list. First, the number of *arpeggio* contact types are used to break the tie. However, if there are fragments where the number of contact types is also tied, the following convention was used to manually prioritize ranking: charged residues > polar > non-polar residues. In situations where there is still a tie between residues of the same category, the *probe*-based contact count ranking was used to break the tie, as in the case of Thr74 being added before Tyr108. Improvements to the *RINRUS* code to automatically account for tie-breaking in either *probe* or *arpeggio* rankings are currently in development. Further details about *probe*, *arpeggio*, and tie-broken *arpeggio* QM-cluster models is given in Table S1.

The tie-broken *arpeggio*-based models (Figure 3, brown plot) show quicker convergence to the $\Delta G^{\ddagger}$ value of the maximal model (10.2 kcal mol^{-1}) than in the original *arpeggio* building scheme. Adding Arg90 before Thr74 via the tie-breaking scheme also eliminates the odd disruption of ΔG_{rxn} convergence. However, there is still no model where both the

ΔG_{rxn} and $\Delta G^{\ddagger}$ are converged to within ± 1 kcal mol^{-1} of the maximal model. Tie-broken *arpeggio*-based models have kinetic and thermodynamic values converging to the loose ± 3 kcal mol^{-1} threshold starting with the 191-atom model. The tie-broken models do not have a significant effect on kinetic or thermodynamic convergence beyond fixing the spurious $\Delta G^{\ddagger}$ value. However, avoiding random ordering of fragments that have tied contact count or contact type values seems prudent until an automated approach is available.

A third scheme using quantitative chorismate-residue interaction energies as a ranking method was evaluated. As observed in previous work,[66,67] the F-SAPT interaction energies prioritize important charged residues which play a key role in transition state stabilization. Our analysis of several proteins (including chorismate mutase) indicated no apparent correlation between number of *probe* contact counts and E_{int} between noncovalently interacting biochemical fragments, raising concern that *probe* may de-emphasize residues that have a strong, but directional electrostatic interaction with seed fragments. The substrate-residue interaction energies were computed using F-SAPT0, and a series of 11 QM-cluster models were first constructed by adding fragments ranked from largest negative E_{int} with the chorismate substrate to the largest positive E_{int} value (Table S5). It must be recognized that a negative total F-SAPT interaction energy signifies a favorable interaction between a residue fragment and chorismate, while a positive total F-SAPT interaction energy describes a repulsive interaction. Given a dianionic chorismate substrate, it was expected that positively charged residues will be ranked first, then polar residues, then nonpolar residues, then negatively charged residues. The initial F-SAPT scheme ranked the four positively charged residues highest; Arg7 is first (E_{int} = -140.5 kcal mol^{-1}), followed by Arg63 (E_{int} = -133.2 kcal mol^{-1}), then Arg90 (E_{int} = -113.0 kcal mol^{-1}), and Lys60 (E_{int} = -78.1 kcal mol^{-1}). For comparison to a few polar residues, the E_{int} of Tyr108 and Thr74 are -15.5 kcal mol^{-1} and +10.9 kcal mol^{-1}, respectively.

Matching literature precedence, the *probe* and *arpeggio* schemes for constructing QM-cluster models frequently de-prioritize charged residues compared to F-SAPT.[67] While Arg7

is ranked first or second in all three schemes, Arg90 is ranked 3rd by F-SAPT, 5th by *probe*, and 6th by *arpeggio*, as illustrated in Tables S4 and S3. Arg63 (1st by F-SAPT) was ranked 8th by *probe* and 3rd by *arpeggio*. A visual relationship between *probe* contacts and the orientation of important charged active site arginine residues can be seen in Figure S6. Phe57 is ranked first in the *probe* ranking scheme with a total of 288 contacts with chorismate (highlighted in grey in Figure S6), but only has an F-SAPT E_{int} of -2.0 kcal mol^{-1} and is ranked 10th. Arg63 has only 97 *probe* (highlighted in yellow) interaction counts, making it the 8th ranked fragment, but again has the second largest negative F-SAPT interaction energy. Charged active site amino acid residues are crucial for both NAC and TSS of the chorismate substrate. Yet Arg7 is the only one of four positively charged residues in the BsCM active site that is ranked consistently high in the *probe*, *arpeggio*, and F-SAPT schemes. Our F-SAPT results strongly suggest large residue side chains can be oriented in such a way that they provide strong hydrogen bonds within an active site, but have low RIN contact count values.

In a recent analysis of glycine-N-methyltransferase,[85] we recognized that residues with strongly unfavorable (positive) interaction energies should be ranked higher than residues with near-zero F-SAPT interaction energies. Ranking fragments by $|E_{int}|$ will thus prioritize negatively charged active site fragments that have a large, but unfavorable interaction with the dianionic substrate before fragments that have a small or negligible interaction with the substrate. Semantically, the difference between F-SAPT schemes is subtle, but the quality of QM-cluster models could be substantially affected by this choice. The $\Delta G^{\ddagger}$ and ΔG_{rxn} values for the two F-SAPT ranking schemes (signed in magenta and unsigned in brown) are overlaid in Figure 4. Both schemes overlap until the 139-atom model, where Ala59 is next added in the signed scheme and Thr74 is added in the unsigned scheme.

As the F-SAPT calculations were derived from the maximal *probe* model, the *probe*, signed and unsigned F-SAPT schemes will all have an equivalent maximal model ($\Delta G^{\ddagger}$ = 9.1 kcal mol^{-1} and ΔG_{rxn} = -16.3 kcal mol^{-1}) that does not need to be recomputed. The

unsigned F-SAPT ranking scheme exhibits slightly improved convergence over the signed scheme, as the last three unsigned models were within ± 3 kcal mol^{-1} of the maximal model for both $\Delta G^{\ddagger}$ and ΔG_{rxn} values (Table S3). Despite the expectation that QM-cluster models derived from F-SAPT rankings would be optimal, none of the truncated F-SAPT models are within ± 1 kcal mol^{-1} of both the $\Delta G^{\ddagger}$ and ΔG_{rxn} values. Thus, there is little qualitative difference between the largest QM-cluster models built with the F-SAPT, *probe*, or *arpeggio* ranking schemes. The F-SAPT scheme is also quite computationally expensive on the front end compared to *probe* and *arpeggio* schemes. Generally, only QM-cluster models of chorismate mutase that closely resemble the maximal models are reliable. To ascertain how more liberally truncated models can appropriately reproduce NAC or TSS phenomena, a brute force or combinatorial approach (like the RINRUS-based investigation of Catechol-O-Methyltransferase)[4] would need to be carried out on the chorismate mutase active site.

Previous eznymology studies done by our group have shown that B3LYP generally underestimates free energies of activation compared to experiment.[2–4,85] Accordingly, all ranking schemes had maximal QM-cluster models of the chorismate mutase active site that exhibited $\Delta G^{\ddagger}$ values significantly lower than the experimental value[86] of 15.4 ± 0.5 kcal mol^{-1}. The maximal F-SAPT / *probe*-based model predicted an activation free energy of 9.1 kcal mol^{-1}, while the maximal *arpeggio*-based model predicted 10.2 kcal mol^{-1}. QM-cluster models reported by Burschowsky and coauthors at the B3LYP/6-31G(d)//B3LYP/6-311+G(d,p) level of theory arrived at an even lower $\Delta G^{\ddagger}$ value for the chorismate mutase catalysis (8.6 kcal mol^{-1}).[35] It is important to stress that this work is not concerned with accuracy of the QM-cluster models, but focused on understanding how kinetics and thermodynamics are influenced by the decisions involved in QM-cluster model construction.

Our lab (and others) are exploring much-needed benchmarks of one-electron basis set and density functional on enzyme models.[45,46,87–92] To avoid model construction contributing to kinetic and thermodynamic errors, the current study demonstrates that QM-cluster models require, at minimum, over ∼150 atoms. This lower bound to model size unfor-

tunately guarantees that employing large basis sets and double-hybrid density functionals will be intractable for most production-level exploration of enzyme chemical mechanisms. Ideally, the community will arrive at a consensus on methodological best practices in QM-cluster modeling to accurately and efficiently compare to experimental observation. Until then, dispersion-corrected B3LYP with small Pople-style basis sets is an efficient and mostly reliable level of theory for calibrating the error arising from QM-cluster model composition.

Building QM-cluster models from MD frames

Next, we explore the impact that fluctuations of residue and substrate positioning can have on both the design of QM-cluster modeling and the resulting kinetic and thermodynamic properties. First, 250 frames from a 20 ns MD simulation of solvated chorismate mutase were sampled to construct maximal QM-cluster models of the active site using *probe* contacts. Structures from MD simulations can be advantageous over crystallographic structures in their unambiguous hydration shells and energy relaxation of the active site structure based on *in vivo* substrates rather than inhibitors or transition state analogues. However, building QM-models from MD simulations will incorporate statistical uncertainty, as sampling many MD frames are required to represent the diversity of structural conformations.[93–95] In particular, we examine three features particularly relevant for QM-cluster modeling that are expected to cause variation in the predicted reaction properties: 1) the number and identity of residues included in the model, 2) the number of waters included in the model, and 3) the statistical ensemble of sampled frames.

In plotting the activation and reaction free energies for all 250 MD-derived QM-cluster models (Figure 5, Figure S7, and Table 1), there is a wide range of values wherein the mean activation free energy is 10.3 ± 2.6 kcal mol^{-1} and the mean free energy of reaction is -15.4 ± 3.4 kcal mol^{-1}. These ranges encompass the converged values observed for QM-cluster models built from the X-ray crystal structure, though this is unsurprising given the large standard deviation observed in the ensemble of refined MD frames. The size of the

maximal QM-cluster models ranges from 158 to 240 atoms, with the five smallest models containing only 8 residues and 5 or 6 waters and the largest model containing 13 residues and 3 waters.

Using *probe* to identify active site fragments, a total of 22 residues were identified as having at least one contact interaction with the substrate in at least one frame over the course of the entire MD simulation. Table S6 shows the mean interaction counts of each identified residue with chorismate. There is precedence that crystal packing leads to an increase in protein-substrate contact counts.[67,96] However, replacement of the TSA with chorismate in the X-ray crystal structure without a subsequent geometry relaxation does not create steric clashes with the protein, which might have nonphysically amplified the contact counts. As expected, the Arg90 and Arg7 residues have the highest mean contact counts, 116.3 and 78.3, respectively. Several residues appear in RINs during the entire MD run with very low mean interaction counts (< 0.02) such as Ala9, Pro117, and residues 242-245. None of these residues have inter-residue contacts with the TSA in the X-ray crystal structure. Pro117 is the only "rare" residue from the entire MD simulation that also appears in the 250 selected frames that were refined to QM-cluster models. The mean interaction counts of residues modeled in the 250 QM-cluster models is similar to those observed in the 20000 RINs of the MD simulation (Table S6). This similarity affirms that the selection schemes used to refine MD frames into QM-cluster models are representative of the entire MD simulation. From Tables S6 and S7, we find that consistently high-ranking active site residues common to *probe*, *arpeggio*, and F-SAPT schemes can occasionally be missing entirely from specific MD frames.

Surprisingly, QM-cluster models with atypical composition do not necessarily create kinetic or thermodynamic outliers. Frame 394 is the only member of the 250-frame subset to not have any *probe* contacts with the Arg90 side chain. It also does not contain an Arg63 fragment, making it the only QM-cluster model with net -2 charge. The missing fragments result in a spuriously high free energy of activation (see below). The QM-cluster models

made from frames 9464, 14007, 16450 are the only three of the 250 that have no *probe* contacts between substrate and Leu115, yet all three have kinetic/thermodynamic properties within the uncertainty range of the total set. Frames with rare residues have a small impact on the overall kinetic and thermodynamic values. For example, the five QM-cluster models that contain Pro117 have mean $\Delta G^{\ddagger}$ and ΔG_{rxn} values of 11.2 ± 2.9 kcal mol^{-1} and -15.1 ± 4.3 kcal mol^{-1}, respectively.

Mean *probe* contact counts of the 250 QM-cluster models arising from MD sampling emphasize charged residues more than the X-ray crystal structure, but interestingly, Figure S8 still shows a lack of correlation with F-SAPT $|E_{int}|$ values computed at the X-ray crystal structure. MD-averaged *probe* counts rank the first five residues as Arg90, Arg7, Leu115, Ala59, and Arg63. The Lys60 residue has a mean contact count of only 2.9, but as demonstrated earlier, has the 4th-largest $|E_{int}|$ with the substrate. The mean *probe* contact counts for Leu115 are large (72.3), but it has the smallest absolute F-SAPT interaction energy. Of the uncharged side chain fragments, Tyr108 has the smallest mean *probe* count (28.8) and the largest $|E_{int}|$ value. These conflicting results demonstrate how various schemes rank residue importance differently. Great challenges remain in quantifying the impact of specific amino acid fragments on protein-substrate reactivity.

The catalytic activity of chorismate mutase is particularly driven by charge stabilization interactions, which might be susceptible to differences in net model charge. Thus, it is of interest to examine whether differences in model charge of QM-cluster models refined from individual MD frames can account for the broad range of activation and reaction free energies observed. Figure 6 shows the distribution of the net model charges for the 250 QM-cluster models compared to the range of $\Delta G^{\ddagger}$ and ΔG_{rxn} values for each model. The net charge of our 250 QM-cluster models varies from -2 to +2, with the majority (200 models) having an overall neutral charge. QM-cluster models with a neutral model charge had mean $\Delta G^{\ddagger}$ and ΔG_{rxn} values of 10.1 ± 2.4 and -15.7 ± 3.3 kcal mol^{-1}, respectively. Only one MD-based QM-cluster model (frame 394) has a -2 net charge model and it provides anomalously high

values of $\Delta G^{\ddagger}$ and ΔG_{rxn}, 20.0 and -7.7 kcal mol^{-1}, respectively. The outlying energetics of frame 394 are likely due to missing Arg90 and Arg63 fragments, which have proven to be critical for the enzyme catalysis.[19,31] The 33 QM-cluster models with net +1 charge show the largest range of $\Delta G^{\ddagger}$ values, encompassing the highest (19.1 kcal mol^{-1}, frame 4114) and lowest (4.1 kcal mol^{-1}, frame 8310) values. However, the mean energetic values are in reasonable agreement with the complete set, 11.1 $\pm$ 3.6 kcal mol^{-1} for $\Delta G^{\ddagger}$ and -14.0 $\pm$ 3.3 kcal mol^{-1} for ΔG_{rxn}. The net charge of the QM-cluster models do not systematically influence the $\Delta G^{\ddagger}$ and ΔG_{rxn} values.

We have shown the maximal QM-cluster models based on the X-ray crystal structure, from any of our building schemes, are expected to provide kinetics and reaction thermodynamics that are reliably converged at a given level of theory (Figure 2). The 250 maximal QM-cluster models derived from MD will have significant variations in the residues that are included in each RIN. This heterogeneity opens the question: when comparing QM-cluster models with the same fragment composition but with different active site conformation and/or relative frozen atom positions, will the computed reaction kinetics and thermodynamics show consistent values or large variance? To disentangle model composition from model structure, the dataset is trimmed to only include MD-derived QM-cluster models that have an identical composition. This data filtering ignores distinguishing models with different water molecule positioning. The subset contained 144 total models in 37 different bins (Figure S9). Among the groups of models with identical designs but taken from different snapshots, the groups still show a wide distribution of $\Delta G^{\ddagger}$ and ΔG_{rxn} values, with ranges from 4.1 to 16.4 kcal mol^{-1} for $\Delta G^{\ddagger}$ and -28.8 to -6.7 kcal mol^{-1} for ΔG_{rxn}. No patterns seem to emerge from this data. If the bins in Figure S9 showed a narrow distribution of kinetics and thermodynamics, we would conclude that the observed wide distribution of values in the 250 QM-cluster models manifested from differences in active site fragment composition. However, data in Figure S9 match the large variation of the total set of QM-cluster models refined from the MD simulation. The variation must be due to conformational fluctuation

of active site residues and water molecules during the course of the MD trajectory.

The active site RIN from the X-ray crystal structure contains only a single crystallographically resolved water molecule shown to have interactions with the substrate captured by *probe*. The chorismate mutase active site is small and quite solvent-exposed, but the lack of crystallographically resolved water molecules is unsurprising (though rarely quantified in the literature). The 3D protein structure is typically of greater interest than the poorly resolved oxygen nuclei of the bulk solvent. In contrast, the QM-cluster models generated from the MD simulation encompass a comprehensive hydration shell. In the 250 MD frames selected for QM-cluster model refinement, 2 to 10 water molecules are identified by *probe* as having an interaction with chorismate (Figure 7). Intriguingly, the *RINRUS*-built QM-cluster models of chorismate mutase derived from MD frames have on average 5.6 water molecules interacting with the substrate. Frame 6981 is the only QM-cluster model with 2 waters in the active site, and $\Delta G^{\ddagger}$ is predicted to be 10.8 kcal mol^{-1}. At the other extreme, the two QM-cluster models with 10 waters have a mean $\Delta G^{\ddagger}$ value of 11.3 kcal mol^{-1}. Only 29 models total have 2, 3, 8, 9, or 10 water molecules in the RIN. Despite low occurrence in the sampled MD frames, these models have mean predicted $\Delta G^{\ddagger}$ and ΔG_{rxn} values of 10.9 $\pm$ 3.0 kcal mol^{-1} and -14.8 $\pm$ 2.9 kcal mol^{-1}, respectively; kinetics and thermodynamics are within uncertainties of the total set of 250 models. The 221 QM-cluster models with 4 to 7 water molecules are qualitatively similar, 10.2 $\pm$ 2.6 kcal mol^{-1} for $\Delta G^{\ddagger}$ and -15.4 $\pm$ 3.4 kcal mol^{-1} for ΔG_{rxn}. Clearly, the number of waters in the BsCM active site has minimal influence on the kinetic and thermodynamic properties of QM-cluster models. However, the inclusion of *any type of water network* at the active site-solvent boundary in our MD-derived QM-cluster models may be a factor in the $\sim$2 kcal mol^{-1} higher free energies of activation observed compared to models constructed from the X-ray crystal structure.

Finally, we analyze groupings of the statistical ensemble of QM-cluster models (Table 1), which showed minimal statistical difference with the overall mean kinetic and thermodynamic values ($\Delta G^{\ddagger} = 10.3 \pm 2.6$ kcal mol^{-1} and $\Delta G_{rxn} = -15.4 \pm 3.4$ kcal mol^{-1}). Schemes labeled

$\mathbf{XS}_2$ to $\mathbf{XS}_8$, are expanded versions of $\mathbf{S}_2$ to $\mathbf{S}_8$, and include all frames from the 250 QM-cluster models that fit the criteria of each Scheme. For example, $\mathbf{XS}_2$ includes the 30 frames from $\mathbf{S}_2$ and the additional 118 frames from the 250 frame set that have an RMSD within 0.76 Å of the mean backbone atom RMSD. Kinetic and thermodynamic results for the expanded schemes are given in Table 2.

The first scheme, $\mathbf{S}_1$, contains 20 frames and should be representative of a random and unbiased distribution of activation and reaction free energies over the course of the entire MD simulation. Mean $\Delta G^{\ddagger}$ and ΔG_{rxn} values of the 20 frames used in $\mathbf{S}_1$ are lower than the total set, but in reasonable agreement. Establishing that k-means clustering of $\mathbf{S}_6$ and $\mathbf{S}_7$ was invalid, these two schemes also represent a random selection of frames. We combined the frames of $\mathbf{S}_1$, $\mathbf{S}_6$, and $\mathbf{S}_7$ (100 total) into an expanded Scheme ($\mathbf{S}_1 + \mathbf{S}_6 + \mathbf{S}_7$) in Table 2. Interestingly, the kinetic and thermodynamic values of $\mathbf{S}_1 + \mathbf{S}_6 + \mathbf{S}_7$ are within 0.10 kcal mol^{-1} of the entire data set. This improved agreement suggests 20 randomly selected frames (8% of the total data set) may not be a robust amount. Since most of the expanded schemes have mean kinetic and thermodynamic values very similar to the total set of 250 MD frames, then a sample of 100 frames (40% of the data points in total set) may be an upper bound needed to emulate the total set.

The next sets of schemes ($\mathbf{S}_2$, $\mathbf{S}_3$, and $\mathbf{S}_4$), take into account the fluctuation of the active site residues and discard MD frames geometrically dissimilar to the X-ray crystal structure. All three schemes predict mean $\Delta G^{\ddagger}$ values slightly lower than the entire dataset. $\mathbf{S}_2$ and $\mathbf{S}_4$ mean ΔG_{rxn} values are lower than the total mean, while the $\mathbf{S}_4$ mean is slightly higher than $\mathbf{S}_2$ and $\mathbf{S}_3$. The extended $\mathbf{XS}_3$ and $\mathbf{XS}_4$ schemes (Table 2) are closer to the total mean statistics than $\mathbf{XS}_2$.

The $\mathbf{S}_5$ scheme used k-means clustering of the RMSDs (ranging from 1.46 to 4.22 Å shown in Table S8) of the active site residues to group similar frames into clusters. The three clusters for $\mathbf{S}_5$ are ordered from largest centroid RMSD value ($\mathbf{S}_5$-C_1) to the lowest ($\mathbf{S}_5$-C_3). The ($\mathbf{S}_5$-C_1) and ($\mathbf{S}_5$-C_3) clusters have nearly the same mean $\Delta G^{\ddagger}$ value, below

the mean $\Delta G^{\ddagger}$ value of the total data set. The ($\mathbf{S}_5$-C_2) cluster in contrast, is higher (11.0 kcal mol^{-1}) than the total data set. Values of ΔG_{rxn} become less negative as the centroid RMSD value decreases from C_3 to C_1, and the extended Scheme $\mathbf{XS}_5$-C_3 to $\mathbf{XS}_5$-C_1 follows the same pattern. Scheme $\mathbf{S}_5$ gives a mean $\Delta G^{\ddagger}$ value closest to that of the total data set. The mean ΔG_{rxn} value for $\mathbf{S}_5$ is also quite close, but effectively random sampling in $\mathbf{S}_6$ and $\mathbf{S}_7$ give a slightly better match to the total set.

The last scheme ($\mathbf{S}_8$) classified MD frames with k-means clustering according to *probe* interatomic contacts between the chorismate ligand and surrounding residues. All three clusters of $\mathbf{S}_8$ predicted the mean $\Delta G^{\ddagger}$ value to be 0.46 - 1.15 kcal mol^{-1} more negative than the mean of the total dataset. The statistics of the expanded clusters of $\mathbf{XS}_8$ are much closer to the total dataset. Notwithstanding, the largest magnitude differences between any frame selection scheme and mean values of the 250 QM-cluster models are 0.62 kcal mol^{-1} for $\Delta G^{\ddagger}$ and 1.56 kcal mol^{-1} for ΔG_{rxn}. For the expanded schemes, the largest absolute differences decrease to 0.17 kcal mol^{-1} for mean $\Delta G^{\ddagger}$ and 0.78 kcal mol^{-1} for mean ΔG_{rxn}.

In summary, efforts to find a subset of MD frame selection schemes that best reflect the kinetic and thermodynamic values of a large statistical ensemble were inconclusive, yet promising. All eight schemes shown in Table 1, with 20 - 40 MD frames in each refined to QM-cluster models, give reasonable approximations to the larger set of 250 MD frames. Expanded schemes with 69 - 186 selected MD frames give mean values even closer to the larger data set. Schemes employing k-means clustering to partition frames via structural metrics did not perform better than schemes with completely random selected MD frames. However, the QM-cluster models were built from one of three trimeric BsCM active sites (the Chain A/C interface) that exhibited the least conformational fluctuation during the course of the 20 ns MD simulation. Machine-learned selection procedures like k-means clustering may be more beneficial for enzymes with more disordered regions or that undergo substantial conformational changes during the simulation time.

Conclusions

Over 50 QM-cluster models of *Bacillus subtilis* chorismate mutase based on the X-ray crystal structure, and an additional 250 QM-cluster models obtained from sampling MD frames were extensively tested with the *RINRUS* software package being developed by our group. *RINRUS* automatically identifies and trims fragments that interact with a substrate, allowing quantitative and reproducible analysis of how the active site fragments affect enzyme catalysis.

The smallest QM-cluster models built with *probe*, *arpeggio* and F-SAPT schemes showed critical differences in how the kinetic and thermodynamics were altered by subsequent addition of residues. Once model building schemes approach the size of the "maximal" model, all three iterative schemes behaved similarly. We have seen some methodological issues with the *arpeggio* ranking scheme where ties can occur in the number of contact counts or contact types. The tie issue in *arpeggio* was resolved manually, and fixed an outlying reaction free energy that was observed in one of the smaller QM-cluster models. The solution to tie interaction counts or types will require more automation to be incorporated into *RINRUS* functionality.

The F-SAPT-based interaction energies highlight the importance of active site charged residues. We recommend always using absolute values of F-SAPT interaction energies to rank active site fragments in QM-cluster model construction. Rankings via signed interaction energies may de-prioritize important active site fragments that exhibit electrostatic repulsion with a substrate. The unsigned F-SAPT ranking scheme showed slight improvement of convergence compared to *probe* and *arpeggio* schemes, but no truncated models in any of the schemes converged to within 1 kcal mol^{-1} of the respective maximal models. We again validate that there is no correlation between the number of *probe* contact counts and E_{int} obtained from F-SAPT computations. More case studies are required to determine if the small performance differences between schemes is related to the compact size of the BsCM active site. Nevertheless, *probe*-based models, *arpeggio* and F-SAPT maximal models

are similar, providing evidence that the largest *RINRUS*-generated QM-cluster models are robust and reliable.

As is widely known in the community and seen in our previous studies, B3LYP-GD3BJ with small Pople-style basis sets and implicit solvation with CPCM systematically underestimates the free energies of activation of enzyme mechanisms compared to the experimental kinetic value. A focus on the quality of the quantum chemical level of theory is purposefully avoided in this work, to instead efficiently provide insight about QM-cluster model building approaches.

The crystallographic protein structure was then solvated within an explicit water bath and, over a 20 ns equilibrated MD simulation, 250 frames were selected to construct 250 QM-cluster models of the active site. The proposed catalyzed Claisen rearrangement mechanism was computed for all QM-cluster 250 models, and the reaction thermodynamics are observed to fluctuate, with the activation free energy spanning 10.34 ± 2.62 kcal mol^{-1} and the reaction free energy spanning -15.38 ± 3.40 kcal mol^{-1}. The variation is shown to be primarily due to the changes in residue/solvent/ligand positioning and conformation that occur over the MD simulation, rather than differences in residue composition among the models. For example, we noted that some active site residues highly ranked in the *probe*, *arpeggio*, and F-SAPT schemes can be absent from specific MD frames when the residues shift to different placements, but the computed kinetic and thermodynamic properties of those complexes can still be reasonable given the QM-cluster model is suitably constructed. Furthermore, while the catalytic mechanism is largely derived from charge stabilization interactions, and we thus might expect the QM-cluster models to be very sensitive to changes in net model charge. The results show most of the variation in $\Delta G^{\ddagger}$ and ΔG_{rxn} values is largely among models with neutral net charge and a general insensitivity in predicted values with net charge between ± 1 was observed. The active site interface with bulk solvent is shown to influence kinetics and thermodynamics of the QM-cluster models. However, the number of explicit water molecules included in the models appear to be inconsequential.

Collectively, results from the MD to QM-cluster model refinement point to the changing molecular positioning rather than model composition as the main source for changing reaction thermodynamics over the sampled times. We attempted to trace the thermodynamic differences to simple, easily quantifiable structural differences among the models, specifically by grouping models based upon RMSDs in backbone or side chain atoms. Ultimately, none of the metrics were better than random selection for acceptably sampling a statistical ensemble of structures. A more multifaceted technique will be required to efficiently cluster MD frames for QM-cluster model refinement, especially if the enzyme undergoes major conformational changes during the MD simulation.

This study exemplifies diverse features of the *RINRUS* toolkit by comparing the structural variation between X-ray crystal structure-based models and MD-based models of bacterial chorismate mutase. Composition of QM-cluster models, or the QM region of a QM/MM model is an essential part of reliability and accuracy in computational enzymology. For far too long, a lack of automated model building techniques and software has hampered advancement of the field as well as the reproducibility of seminal work. Here, QM-cluster modeling provided insight into the enzymatic activity of chorismate mutase by connecting the model composition, the contribution of charged residues, the influence of explicit solvent water molecules, and positioning and orientation of active site residues to the computed kinetic and thermodynamic values. Accompanying data can be easily used to perform further cheminformatic analysis or to calibrate accuracy with more reliable quantum chemistry methodologies; *RINRUS* was designed with reproducibility as a core feature.

Acknowledgments

This work is supported by the National Science Foundation (NSF) CAREER BIO-1846408 (for NJD and DAA), the Department of Energy (DOE) BES SBIR DE-SC0021568 (for NJD), the National Institute of General Medical Sciences of the National Institutes of Health under

award number 1R35GM145206-01 (for QC), and the NSF Graduate Research Fellowship Program under Grant No. 1451514 (for TJS). The authors are grateful for the support provided by the University of Memphis High Performance Computing Center.

Author Contributions

Donatus A. Agbaglo: Conceptualization (equal); Formal analysis (lead); Investigation (lead); Methodology (equal); Visualization (lead); Writing – original draft (lead); Writing – review & editing (equal). **Thomas J. Summers**: Conceptualization (equal); Formal analysis (supporting); Methodology (equal); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (equal). **Qianyi Cheng**: Writing – review & editing (equal); Investigation (supporting). **Nathan J. DeYonker**: Conceptualization (equal); Resources (lead); Writing – original draft (supporting); Writing – review & editing (equal).

References

(1) Kraut, D. A.; Carroll, K. S.; Herschlag, D. Challenges in enzyme mechanism and energetics. *Annu. Rev. Biochem.* **2003**, *72*, 517–571.

(2) Cheng, Q.; DeYonker, N. J. Acylation and deacylation mechanism and kinetics of penicillin G reaction with Streptomyces R61 DD-peptidase. *J. Comput. Chem.* **2020**, *41*, 1685–1697.

(3) Cheng, Q.; DeYonker, N. J. QM-Cluster Model Study of the Guaiacol Hydrogen Atom Transfer and Oxygen Rebound with Cytochrome P450 Enzyme GcoA. *J. Phys. Chem. B* **2021**, *125*, 3296–3306.

(4) Summers, T. J.; Cheng, Q.; Palma, M. A.; Pham, D.-T.; Kelso III, D. K.; Web-

ster, C. E.; DeYonker, N. J. Cheminformatic quantum mechanical enzyme model design: A catechol-O-methyltransferase case study. *Biophys. J.* **2021**, *120*, 3577–3587.

(5) Woodcock, H. L.; Hodošček, M.; Sherwood, P.; Lee, Y. S.; Schaefer III, H. F.; Brooks, B. R. Exploring the quantum mechanical/molecular mechanical replica path method: a pathway optimization of the chorismate to prephenate Claisen rearrangement catalyzed by chorismate mutase. *Theor. Chem. Acc.* **2003**, *109*, 140–148.

(6) Lee, Y. S.; Worthington, S. E.; Krauss, M.; Brooks, B. R. Reaction mechanism of chorismate mutase studied by the combined potentials of quantum mechanics and molecular mechanics. *J. Phys. Chem. B* **2002**, *106*, 12059–12065.

(7) Wiest, O.; Houk, K. On the transition state of the chorismate-prephenate rearrangement. *J. Org. Chem.* **1994**, *59*, 7582–7584.

(8) Wiest, O.; Houk, K. Stabilization of the transition state of the chorismate-prephenate rearrangement: An ab initio study of enzyme and antibody catalysis. *J. Am. Chem. Soc.* **1995**, *117*, 11628–11639.

(9) Lyne, P. D.; Mulholland, A. J.; Richards, W. G. Insights into chorismate mutase catalysis from a combined QM/MM simulation of the enzyme reaction. *J. Am. Chem. Soc.* **1995**, *117*, 11345–11350.

(10) Davidson, M.; Guest, J.; Simon Craw, J.; Hillier, I.; Vincent, M. Conformational and solvation aspects of the chorismate–prephenate rearrangement studied by *ab initio* electronic structure and simulation methods. *J. Chem. Soc., Perkin Trans. 2* **1997**, 1395–1400.

(11) Hall, R. J.; Hindle, S. A.; Burton, N. A.; Hillier, I. H. Aspects of hybrid QM/MM calculations: the treatment of the QM/MM interface region and geometry optimization with an application to chorismate mutase. *J. Comput. Chem.* **2000**, *21*, 1433–1441.

(12) Khanjin, N. A.; Snyder, J. P.; Menger, F. Mechanism of chorismate mutase: Contribution of conformational restriction to catalysis in the Claisen rearrangement. *J. Am. Chem. Soc.* **1999**, *121*, 11831–11846.

(13) Madurga, S.; Vilaseca, E. SCRF study of the conformational equilibrium of chorismate in water. *Phys. Chem. Chem. Phys.* **2001**, *3*, 3548–3554.

(14) Crespo, A.; Scherlis, D. A.; Martí, M. A.; Ordejón, P.; Roitberg, A. E.; Estrin, D. A. A DFT-based QM-MM approach designed for the treatment of large molecular systems: Application to chorismate mutase. *J. Phys. Chem. B* **2003**, *107*, 13728–13736.

(15) Dewick, P. M. The biosynthesis of shikimate metabolites. *Nat. Prod. Rep.* **1995**, *12*, 101–133.

(16) Andrews, P.; Smith, G. D.; Young, I. Transition-state stabilization and enzymic catalysis. Kinetic and molecular orbital studies of the rearrangement of chorismate to prephenate. *Biochem.* **1973**, *12*, 3492–3498.

(17) Gorisch, H. Mechanism of chorismate mutase reaction. *Biochem.* **1978**, *17*, 3700–3705.

(18) Freindorf, M.; Tao, Y.; Sethio, D.; Cremer, D.; Kraka, E. New mechanistic insights into the Claisen rearrangement of chorismate–a Unified Reaction Valley Approach study. *Mol. Phys.* **2019**, *117*, 1172–1192.

(19) Chook, Y. M.; Ke, H.; Lipscomb, W. N. Crystal structures of the monofunctional chorismate mutase from Bacillus subtilis and its complex with a transition state analog. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 8600–8603.

(20) Bruice, T. C. A view at the millennium: the efficiency of enzymatic catalysis. *Acc. Chem. Res.* **2002**, *35*, 139–148.

(21) Claeyssens, F.; Ranaghan, K. E.; Lawan, N.; Macrae, S. J.; Manby, F. R.; Harvey, J. N.; Mulholland, A. J. Analysis of chorismate mutase catalysis by QM/MM modelling of

enzyme-catalysed and uncatalysed reactions. *Org. & Biomol. Chem.* **2011**, *9*, 1578–1590.

(22) Burschowsky, D.; van Eerde, A.; Ökvist, M.; Kienhöfer, A.; Kast, P.; Hilvert, D.; Krengel, U. Electrostatic transition state stabilization rather than reactant destabilization provides the chemical basis for efficient chorismate mutase catalysis. *Proc. Natl. Acad. Sci.* **2014**, *111*, 17516–17521.

(23) Zhang, X.; Zhang, X.; Bruice, T. C. A definitive mechanism for chorismate mutase. *Biochem.* **2005**, *44*, 10443–10448.

(24) Lamb, A. L. Pericyclic reactions catalyzed by chorismate-utilizing enzymes. *Biochem.* **2011**, *50*, 7476–7483.

(25) Štrajbl, M.; Shurki, A.; Kato, M.; Warshel, A. Apparent NAC effect in chorismate mutase reflects electrostatic transition state stabilization. *J. Am. Chem. Soc.* **2003**, *125*, 10228–10237.

(26) Galopin, C. C.; Zhang, S.; Wilson, D. B.; Ganem, B. On the mechanism of chorismate mutases: clues from wild-type E. coli enzyme and a site-directed mutant related to yeast chorismate mutase. *Tetrahedron Lett.* **1996**, *37*, 8675–8678.

(27) Liu, D. R.; Cload, S. T.; Pastor, R. M.; Schultz, P. G. Analysis of active site residues in Escherichia coli chorismate mutase by site-directed mutagenesis. *J. Am. Chem. Soc.* **1996**, *118*, 1789–1790.

(28) Zhang, S.; Kongsaeree, P.; Clardy, J.; Wilson, D. B.; Ganem, B. Site-directed mutagenesis of monofunctional chorismate mutase engineered from the E. coli P-protein. *Bioorg. Med. Chem.* **1996**, *4*, 1015–1020.

(29) Schnappauf, G.; Sträter, N.; Lipscomb, W. N.; Braus, G. H. A glutamate residue in

the catalytic center of the yeast chorismate mutase restricts enzyme activity to acidic conditions. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 8491–8496.

(30) Lassila, J. K.; Keeffe, J. R.; Kast, P.; Mayo, S. L. Exhaustive mutagenesis of six secondary active-site residues in Escherichia coli chorismate mutase shows the importance of hydrophobic side chains and a helix N-capping position for stability and catalysis. *Biochem.* **2007**, *46*, 6883–6891.

(31) Kienhöfer, A.; Kast, P.; Hilvert, D. Selective stabilization of the chorismate mutase transition state by a positively charged hydrogen bond donor. *J. Am. Chem. Soc.* **2003**, *125*, 3206–3207.

(32) Hur, S.; Bruice, T. C. Enzymes do what is expected (chalcone isomerase versus chorismate mutase). *J. Am. Chem. Soc.* **2003**, *125*, 1472–1473.

(33) Zhang, X.; Bruice, T. C. The proficiency of a thermophilic chorismate mutase enzyme is solely through an entropic advantage in the enzyme reaction. *Proc. Natl. Acad. Sci.* **2005**, *102*, 18356–18360.

(34) Shurki, A.; Štrajbl, M.; Villa, J.; Warshel, A. How much do enzymes really gain by restraining their reacting fragments? *J. Am. Chem. Soc.* **2002**, *124*, 4097–4107.

(35) Burschowsky, D.; Krengel, U.; Uggerud, E.; Balcells, D. Quantum chemical modeling of the reaction path of chorismate mutase based on the experimental substrate/product complex. *FEBS. Open Bio.* **2017**, *7*, 789–797.

(36) Mulholland, A. J. Computational enzymology: modelling the mechanisms of biological catalysts. 2008.

(37) Ahmadi, S.; Barrios Herrera, L.; Chehelamirani, M.; Hostaš, J.; Jalife, S.; Salahub, D. R. Multiscale modeling of enzymes: QM-cluster, QM/MM, and QM/MM/MD: A tutorial review. *Intl. J. Quant. Chem.* **2018**, *118*, e25558.

(38) Guo, H.; Rao, N. Chorismate-Mutase-Catalyzed Claisen Rearrangement. *The Claisen Rearrangement: Methods and Applications* **2007**, 1–23.

(39) Kulik, H. J.; Zhang, J.; Klinman, J. P.; Martínez, 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.

(40) 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.

(41) Karelina, M.; Kulik, H. J. Systematic quantum mechanical region determination in QM/MM simulation. *J. Chem. Theory Comput.* **2017**, *13*, 563–576.

(42) Sumner, S.; Soderhjelm, P.; Ryde, U. Effect of geometry optimizations on QM-cluster and QM/MM studies of reaction energies in proteins. *J. Chem. Theory Comput.* **2013**, *9*, 4205–4214.

(43) Hu, L.; Eliasson, J.; Heimdal, J.; Ryde, U. Do quantum mechanical energies calculated for small models of protein-active sites converge? *J. Phys. Chem. A* **2009**, *113*, 11793–11800.

(44) Delcey, M. G.; Pierloot, K.; Phung, Q. M.; Vancoillie, S.; Lindh, R.; Ryde, U. Accurate calculations of geometries and singlet–triplet energy differences for active-site models of [NiFe] hydrogenase. *Phys. Chem. Chem. Phys.* **2014**, *16*, 7927–7938.

(45) Wappett, D. A.; Goerigk, L. A guide to benchmarking enzymatically catalysed reactions: the importance of accurate reference energies and the chemical environment. *Theor. Chem. Acc.* **2021**, *140*, 1–15.

(46) Kromann, J. C.; Christensen, A. S.; Cui, Q.; Jensen, J. H. Towards a barrier height benchmark set for biologically relevant systems. *PeerJ* **2016**, *4*, e1994.

(47) Di Paola, L.; De Ruvo, M.; Paci, P.; Santoni, D.; Giuliani, A. Protein contact networks: an emerging paradigm in chemistry. *Chem. Rev.* **2013**, *113*, 1598–1613.

(48) Doncheva, N. T.; Klein, K.; Domingues, F. S.; Albrecht, M. Analyzing and visualizing residue networks of protein structures. *Trends Biochem. Sci.* **2011**, *36*, 179–182.

(49) Deng, Z.; Chuaqui, C.; Singh, J. Structural interaction fingerprint (SIFt): a novel method for analyzing three-dimensional protein-ligand binding interactions. *J. Med. Chem.* **2004**, *47*, 337–344.

(50) Szalewicz, K. Symmetry-adapted perturbation theory of intermolecular forces. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2*, 254–272.

(51) Parker, T. M.; Burns, L. A.; Parrish, R. M.; Ryno, A. G.; Sherrill, C. D. Levels of symmetry adapted perturbation theory (SAPT). I. Efficiency and performance for interaction energies. *J. Chem. Phys.* **2014**, *140*, 094106.

(52) Parrish, R. M.; Parker, T. M.; Sherrill, C. D. Chemical assignment of symmetry-adapted perturbation theory interaction energy components: the functional-group SAPT partition. *J. Chem. Theory Comput.* **2014**, *10*, 4417–4431.

(53) Parrish, R. M.; Thompson, K. C.; Martinez, T. J. Large-scale functional group symmetry-adapted perturbation theory on graphical processing units. *J. Chem. Theory Comput.* **2018**, *14*, 1737–1753.

(54) Heilmann, N.; Wolf, M.; Kozlowska, M.; Sedghamiz, E.; Setzler, J.; Brieg, M.; Wenzel, W. Sampling of the conformational landscape of small proteins with Monte Carlo methods. *Sci. Rep.* **2020**, *10*, 1–13.

(55) Word, J. M.; Lovell, S. C.; Richardson, J. S.; Richardson, D. C. Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J. Mol. Biol.* **1999**, *285*, 1735–1747.

(56) Word, J. M.; Lovell, S. C.; LaBean, T. H.; Taylor, H. C.; Zalis, M. E.; Presley, B. K.; Richardson, J. S.; Richardson, D. C. Visualizing and quantifying molecular goodness-of-fit: small-probe contact dots with explicit hydrogen atoms. *J. Mol. Biol.* **1999**, *285*, 1711–1733.

(57) Jubb, H. C.; Higueruelo, A. P.; Ochoa-Montaño, B.; Pitt, W. R.; Ascher, D. B.; Blundell, T. L. Arpeggio: a web server for calculating and visualising interatomic interactions in protein structures. *J. Mol. Biol.* **2017**, *429*, 365–371.

(58) Cheng, Q.; DeYonker, N. J.; Summers, T. J.; Agbaglo, D. A.; Suhagia, T.; Palma, M. A. GitHub - natedey/RINRUS: Residue Interaction Network ResidUe Selector (RINRUS) public release. https://github.com/natedey/RINRUS (accessed 2022-09-14). 2022.

(59) Schreyer, A.; Blundell, T. CREDO: a protein–ligand interaction database for drug discovery. *Chem. Biol. Drug Des.* **2009**, *73*, 157–167.

(60) Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self—consistent molecular orbital methods. XII. Further extensions of Gaussian—type basis sets for use in molecular orbital studies of organic molecules. *J. Chem. Phys.* **1972**, *56*, 2257–2261.

(61) Misquitta, A. J.; Podeszwa, R.; Jeziorski, B.; Szalewicz, K. Intermolecular potentials based on symmetry-adapted perturbation theory with dispersion energies from time-dependent density-functional calculations. *J. Chem. Phys.* **2005**, *123*, 214103.

(62) Bukowski, R.; Cencek, W.; Jankowski, P.; Jeziorska, M.; Jeziorski, B.; Kucharski, S.; Lotrich, V.; Misquitta, A.; Moszynski, R.; Patkowski, K., et al. SAPT2008: An *ab initio* program for many-body symmetry-adapted perturbation theory calculations of intermolecular interaction energies. *University of Delaware and University of Warsaw* **2008**,

(63) Spronk, S. A.; Glick, Z. L.; Metcalf, D. P.; Sherrill, C. D.; Cheney, D. L. A quantum

chemical interaction energy dataset for accurately modeling protein-ligand interactions. *Sci. Data* **2023**, *10*, 619.

(64) Parrish, R. M.; Burns, L. A.; Smith, D. G.; Simmonett, A. C.; DePrince III, A. E.; Hohenstein, E. G.; Bozkaya, U.; Sokolov, A. Y.; Di Remigio, R.; Richard, R. M., et al. Psi4 1.1: An open-source electronic structure program emphasizing automation, advanced libraries, and interoperability. *J. Chem. Theory Comput.* **2017**, *13*, 3185–3197.

(65) Burns, L. A.; Faver, J. C.; Zheng, Z.; Marshall, M. S.; Smith, D. G.; Vanommeslaeghe, K.; MacKerell Jr, A. D.; Merz Jr, K. M.; Sherrill, C. D. The BioFragment Database (BFDb): An open-data platform for computational chemistry analysis of noncovalent interactions. *J. Chem. Phys.* **2017**, *147*, 161727.

(66) Summers, T. J.; Daniel, B. P.; Cheng, Q.; DeYonker, N. J. Quantifying Inter-Residue Contacts through Interaction Energies. *J. Chem. Inf. Model* **2019**, *59*, 5034–5044.

(67) Summers, T. J.; Hemmati, R.; Miller, J. E.; Agbaglo, D. A.; Cheng, Q.; DeYonker, N. J. Evaluating the Active Site-Substrate Interplay Between X-ray Crystal Structure and Molecular Dynamics in Chorismate Mutase. *J. Chem. Phys.* **2023**, *158*, 065101.

(68) Frisch, M.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Scalmani, G.; Barone, V.; Petersson, G.; Nakatsuji, H., et al. Gaussian 16. 2016.

(69) Beck, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648–6.

(70) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785.

(71) Petersson, G.; Al-Laham, M. A. A complete basis set model chemistry. II. Open-shell

systems and the total energies of the first-row atoms. *J. Chem. Phys.* **1991**, *94*, 6081–6090.

(72) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32*, 1456–1465.

(73) Barone, V.; Cossi, M. Quantum calculation of molecular energies and energy gradients in solution by a conductor solvent model. *J. Phys. Chem. A* **1998**, *102*, 1995–2001.

(74) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model. *J. Comput. Chem.* **2003**, *24*, 669–681.

(75) Worthington, S. E.; Roitberg, A. E.; Krauss, M. An MD/QM study of the chorismate mutase-catalyzed Claisen rearrangement reaction. *J. Phys. Chem. B* **2001**, *105*, 7087–7095.

(76) Gordon, J. C.; Myers, J. B.; Folta, T.; Shoja, V.; Heath, L. S.; Onufriev, A. H++: a server for estimating p K as and adding missing hydrogens to macromolecules. *Nucleic Acids Res.* **2005**, *33*, W368–W371.

(77) Case, D.; Cerutti, D.; Cheatham, T.; Darden, T.; Duke, R.; Giese, T.; Gohlke, H.; Goetz, A.; Greene, D.; Homeyer, N., et al. Amber18 (University of San Francisco). **2017**,

(78) 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.

(79) 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.

(80) Joung, I. S.; Cheatham III, T. E. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B* **2008**, *112*, 9020–9041.

(81) Andersen, H. C. Rattle: A "velocity" version of the shake algorithm for molecular dynamics calculations. *J. Comput. Phys.* **1983**, *52*, 24–34.

(82) Roe, D. R.; Cheatham III, T. E. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* **2013**, *9*, 3084–3095.

(83) Ishida, T. Probing protein environment in an enzymatic process: All-electron quantum chemical analysis combined with *ab initio* quantum mechanical/molecular mechanical modeling of chorismate mutase. *J. Chem. Phys.* **2008**, *129*, 09B618.

(84) Ishida, T. Effects of point mutation on enzymatic activity: correlation between protein electronic structure and motion in chorismate mutase reaction. *J. Am. Chem. Soc.* **2010**, *132*, 7104–7118.

(85) Cheng, Q.; DeYonker, N. J. The Glycine N-Methyltransferase Case Study: Another Challenge for QM-Cluster Models? *J. Phys. Chem. B* **2023**,

(86) Kast, P.; Asif-Ullah, M.; Hilvert, D. Is chorismate mutase a prototypic entropy trap? - Activation parameters for the Bacillus subtilis enzyme. *Tetrahedron lett.* **1996**, *37*, 2691–2694.

(87) Wappett, D. A.; Goerigk, L. Toward a quantum-chemical benchmark set for enzymatically catalyzed reactions: important steps and insights. *J. Phys. Chem. A* **2019**, *123*, 7057–7074.

(88) Goerigk, L.; Grimme, S. A thorough benchmark of density functional methods for gen-

eral main group thermochemistry, kinetics, and noncovalent interactions. *Phys. Chem. Chem. Phys.* **2011**, *13*, 6670–6688.

(89) Jurečka, P.; Šponer, J.; Černỳ, J.; Hobza, P. Benchmark database of accurate (MP2 and CCSD (T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985–1993.

(90) Antony, J.; Grimme, S. Density functional theory including dispersion corrections for intermolecular interactions in a large benchmark set of biologically relevant molecules. *Phys. Chem. Chem. Phys.* **2006**, *8*, 5287–5293.

(91) Kesharwani, M. K.; Karton, A.; Martin, J. M. Benchmark *ab initio* conformational energies for the proteinogenic amino acids through explicitly correlated methods. Assessment of density functional methods. *J. Chem. Theory Comput.* **2016**, *12*, 444–454.

(92) Paiva, P.; Ramos, M. J.; Fernandes, P. A. Assessing the validity of DLPNO-CCSD (T) in the calculation of activation and reaction energies of ubiquitous enzymatic reactions. *J. Comput. Chem.* **2020**, *41*, 2459–2468.

(93) Ribeiro, A. J.; Santos-Martins, D.; Russo, N.; Ramos, M. J.; Fernandes, P. A. Enzymatic flexibility and reaction rate: a QM/MM study of HIV-1 protease. *ACS Catal.* **2015**, *5*, 5617–5626.

(94) Ryde, U. How many conformations need to be sampled to obtain converged QM/MM energies? The curse of exponential averaging. *J. Chem. Theory Comput.* **2017**, *13*, 5745–5752.

(95) Dokainish, H.; Gauld, J. Computational Approach Choice in Modeling Flexible Enzyme Active Sites. **2019**,

(96) Mei, Z.; Treado, J. D.; Grigas, A. T.; Levine, Z. A.; Regan, L.; O'hern, C. S. Analyses

of protein cores reveal fundamental differences between solution and crystal structures. *Proteins* **2020**, *88*, 1154–1161.

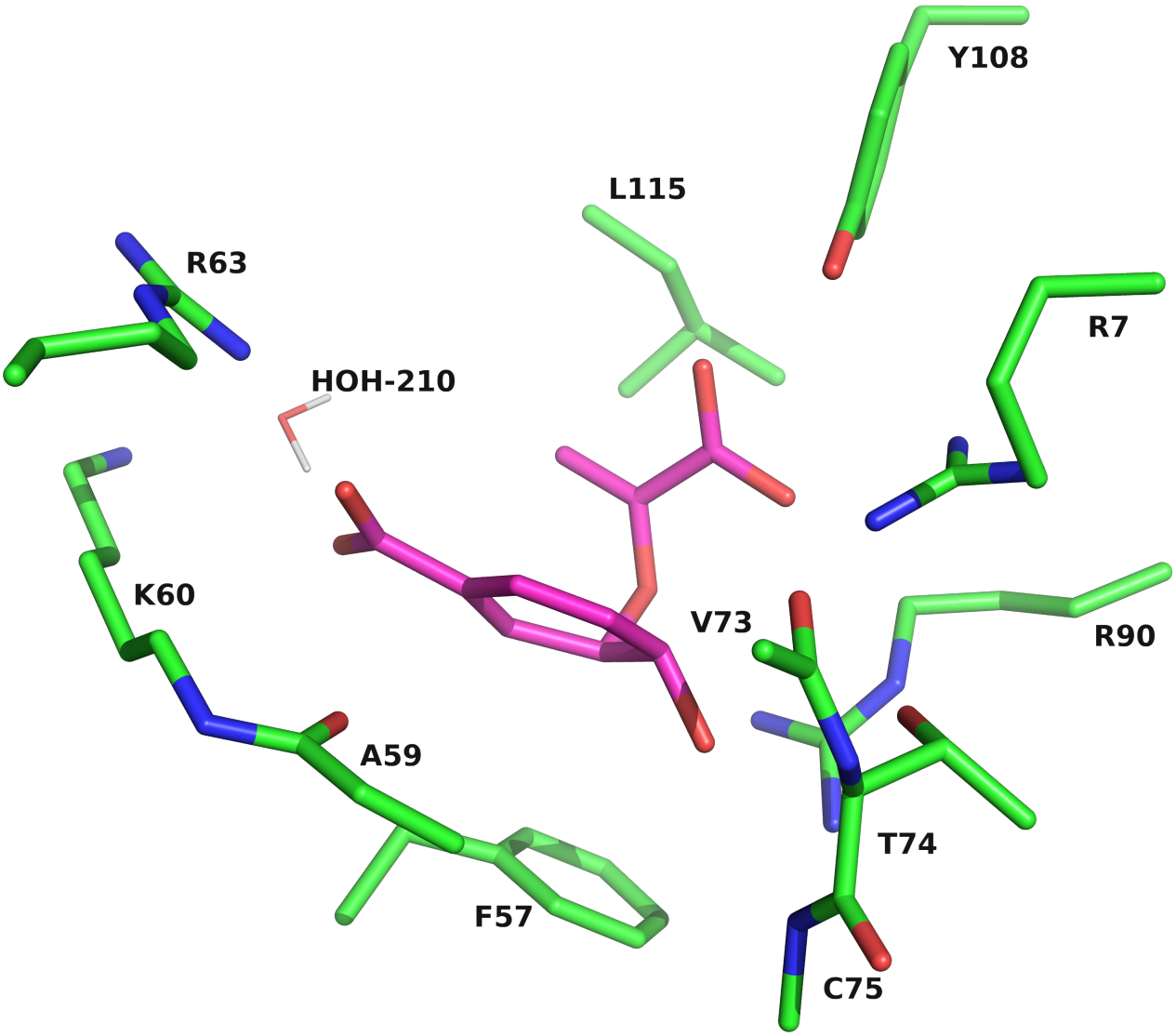


Figure 1: 3D representation of the *RINRUS* maximal model, from the X-ray crystal structure of *Bacillus subtilis* chorismate mutase, using the *probe* ranking scheme. Substrate carbon atoms are colored in magenta. Except for those of the crystallographically resolved water molecule, hydrogen atoms are omitted for clarity.

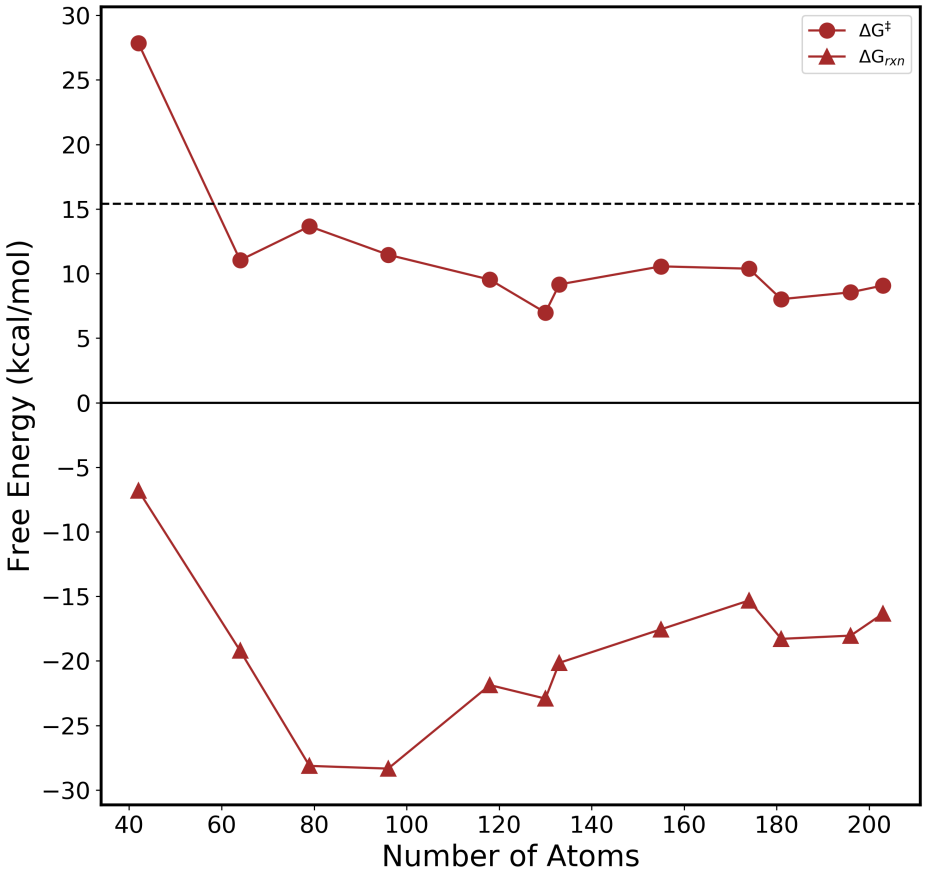


Figure 2: Kinetics and thermodynamics of the iteratively grown QM-cluster models using the *probe* ranking scheme. Computed $\Delta G^{\ddagger}$ values are represented by circles and ΔG_{rxn} values by triangles. The black dashed line shows the experimental $\Delta G^{\ddagger}$ value from reference.[86]

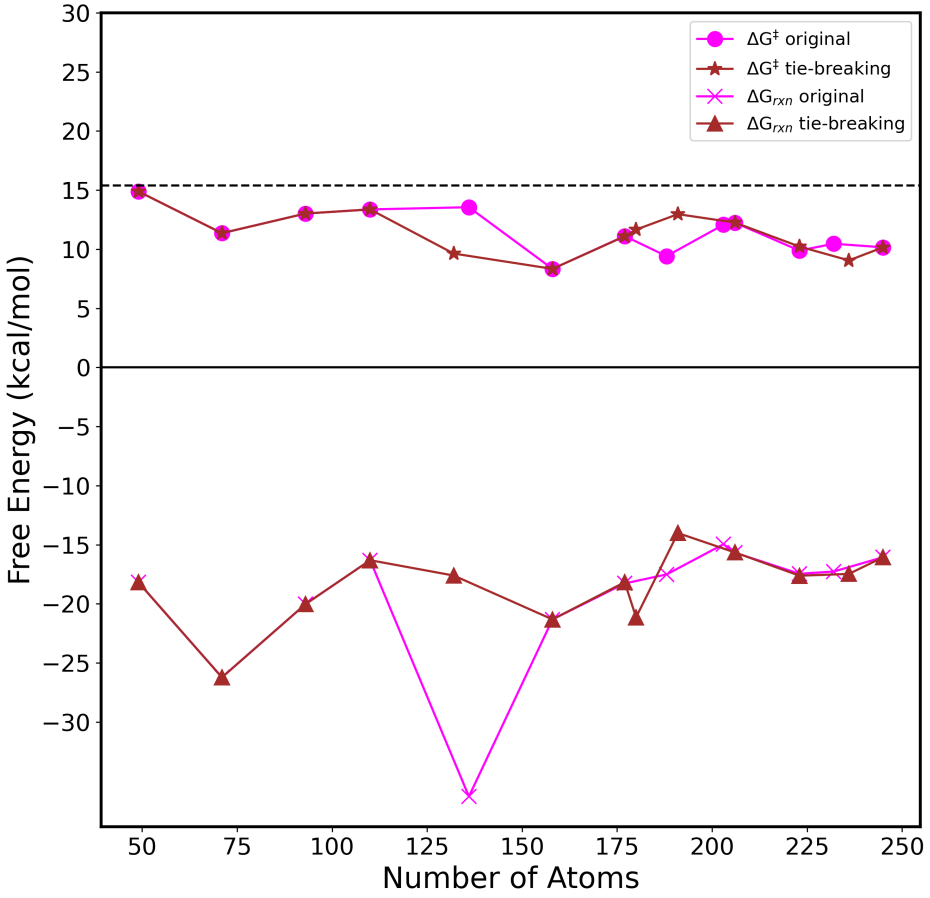


Figure 3: Kinetics and thermodynamics of the iteratively grown QM-cluster models using the *arpeggio* ranking scheme. Computed $\Delta G^{\ddagger}$ values are represented by circles and stars, and ΔG_{rxn} values by crosses and triangles. The original ranking is given in magenta, while values from the tie-breaking scheme are given in brown. The black dashed line shows the experimental $\Delta G^{\ddagger}$ value from reference.[86]

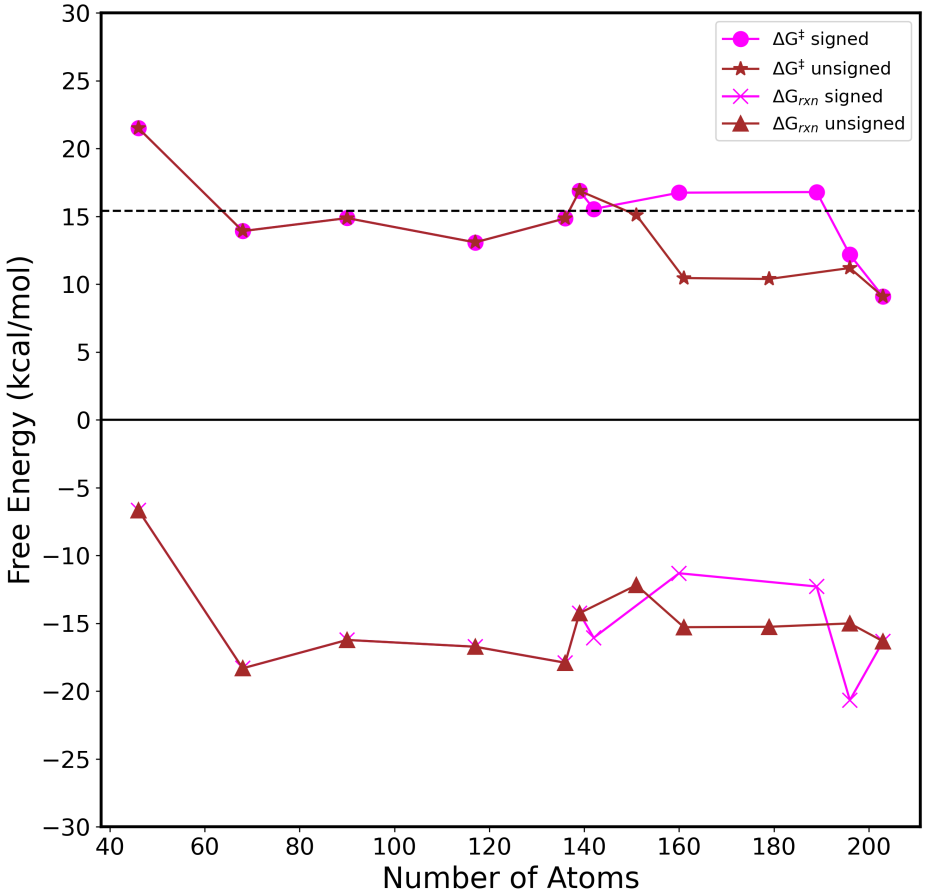


Figure 4: Kinetics and thermodynamics of the iteratively grown QM-cluster models using the F-SAPT ranking scheme. Computed $\Delta G^{\ddagger}$ values are represented by circles and stars, and ΔG_{rxn} values by crosses and triangles. The signed ranking order is given in magenta, while the unsigned ranking order is given in brown. The black dashed line shows the experimental $\Delta G^{\ddagger}$ value from reference.[86]

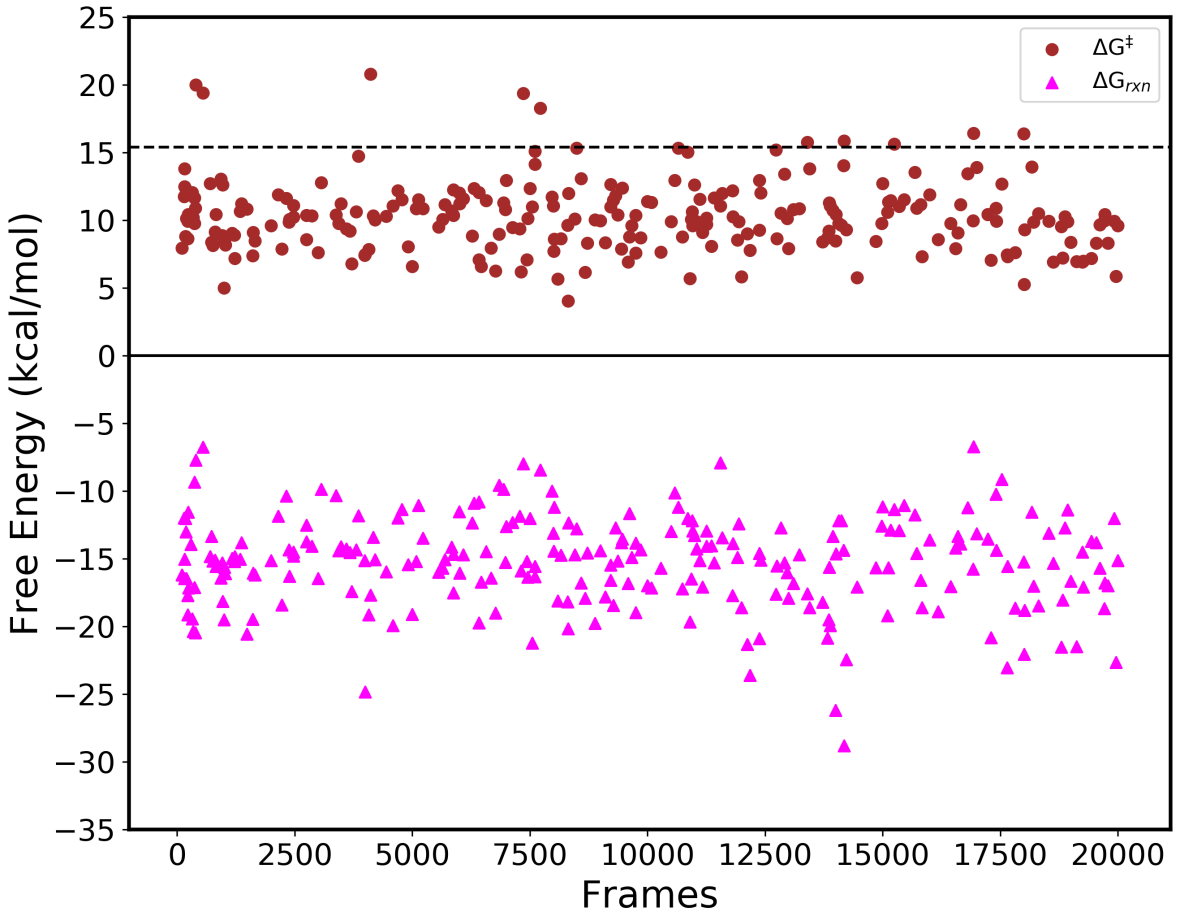


Figure 5: Computed values of $\Delta G^{\ddagger}$ (circle) and ΔG_{rxn} (triangle) for the 250 maximal QM-cluster models plotted against the select frame number (each representing a time scale of 1 ps). The black dashed line at the top is the experimental value from reference.[86]

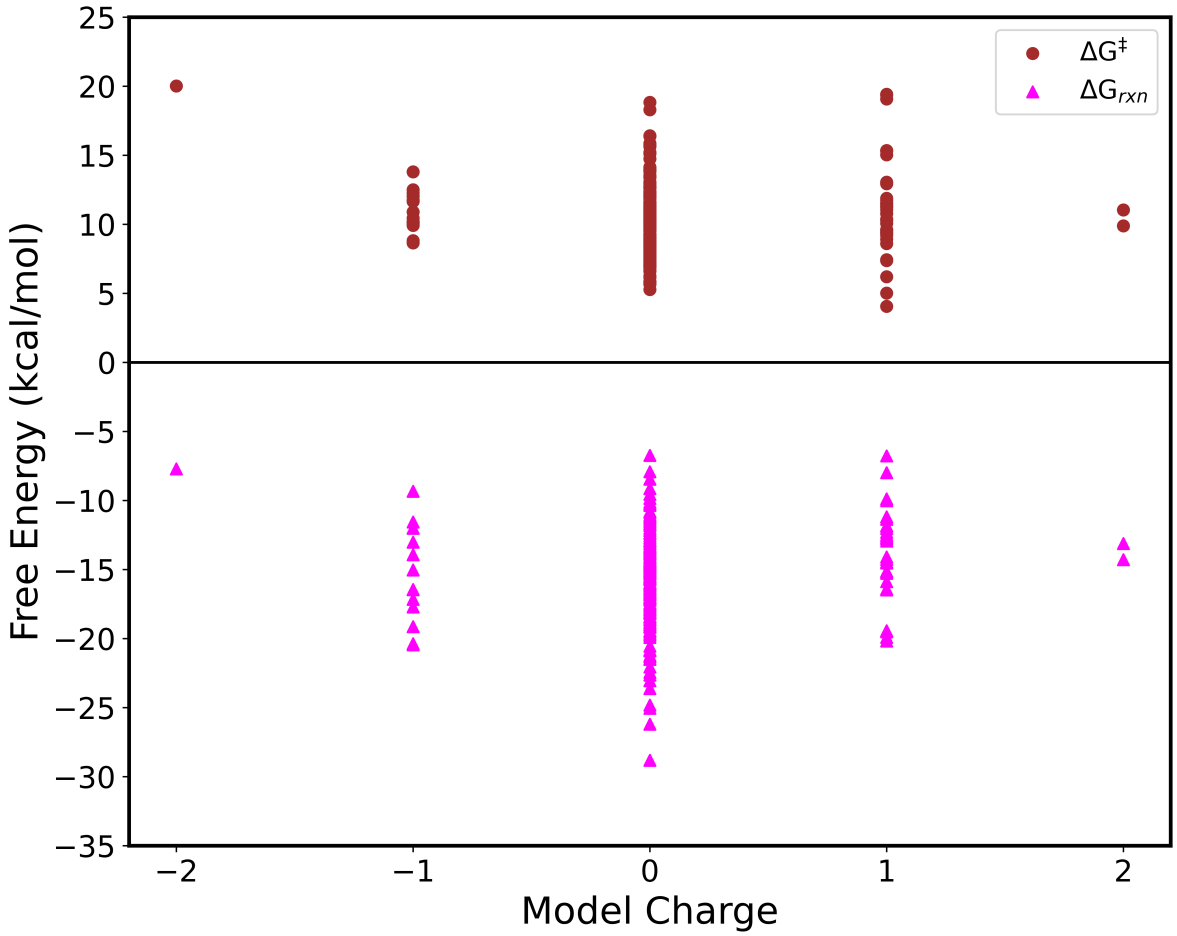


Figure 6: Charge distribution for the 250 QM-cluster models refined from MD frames. The corresponding number of QM-cluster models for each net model charge is: charge -2 = 1 QM-cluster model, charge -1 = 14 QM-cluster models, charge 0 = 200 QM-cluster models, charge +1 = 33 QM cluster models, and charge +2 = 2 QM-cluster models.

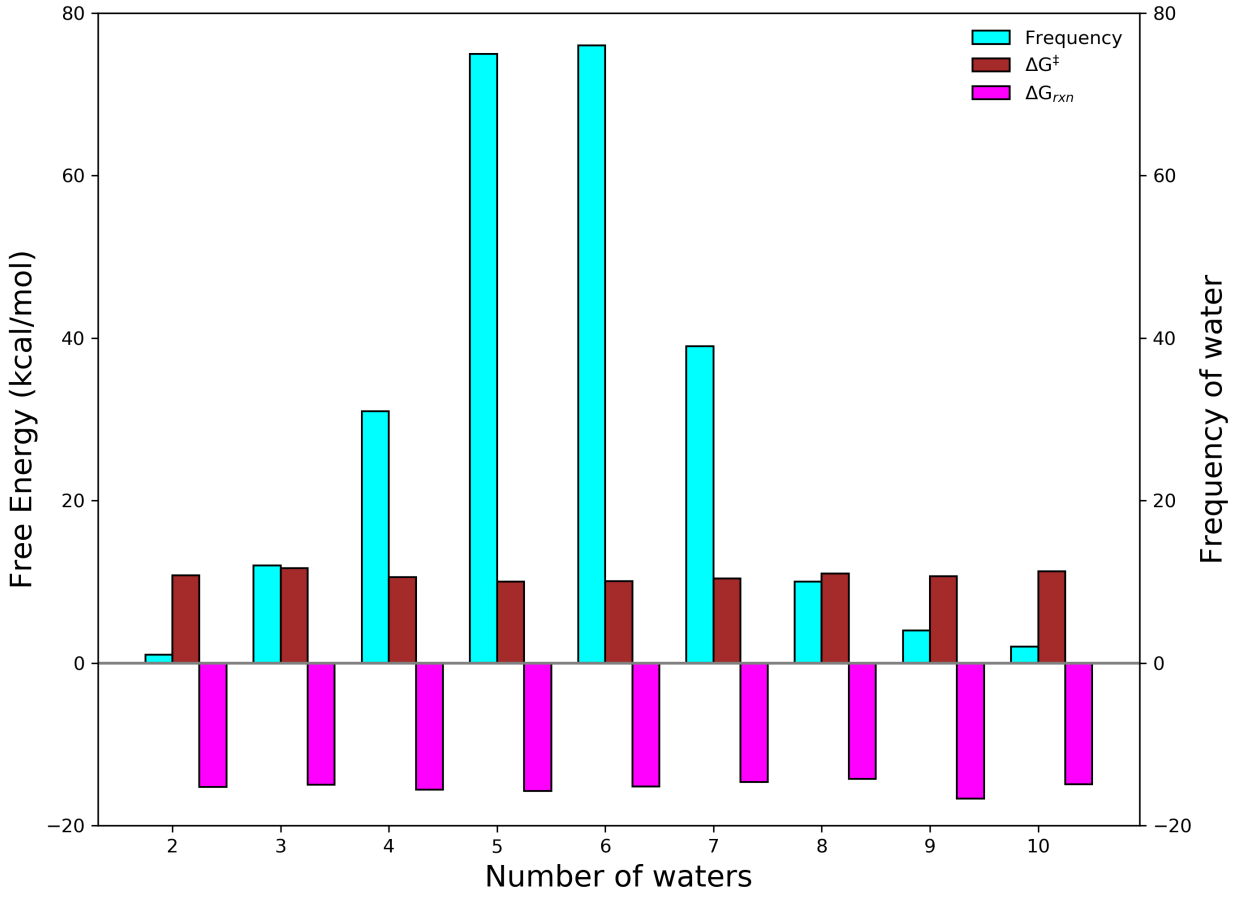


Figure 7: Mean activation free energy (brown), reaction free energy (magenta), and number of QM-cluster models with a given number of explicit water molecules (cyan) identified as having interatomic contacts with the chorismate for the 250 QM-cluster models built from selected MD frames.

Table 1: Mean free energies of activation and reaction for the various MD frame selection schemes. K-means clusters are labelled with a C (all in kcal mol^{-1}).

Scheme	Cluster	# of frames	$\Delta G^{\ddagger}$	σ	ΔG_{rxn}	σ
$\mathbf{S}_1$		20	10.07	± 2.87	-16.23	± 3.90
$\mathbf{S}_2$		30	10.12	± 2.39	-16.28	± 3.82
$\mathbf{S}_3$		30	10.03	± 2.83	-14.99	± 3.06
$\mathbf{S}_4$		30	10.06	± 1.88	-15.92	± 2.86
$\mathbf{S}_5$		30	10.29	± 3.05	-15.57	± 3.09
	C_1	10	9.90	± 2.74	-16.78	± 3.12
	C_2	10	11.03	± 3.98	-15.54	± 3.48
	C_3	10	9.95	± 1.92	-14.40	± 2.02
$\mathbf{S}_6$		40	10.23	± 2.38	-15.24	± 3.04
$\mathbf{S}_7$		40	10.74	± 2.69	-15.35	± 3.52
$\mathbf{S}_8$		30	10.96	± 2.69	-13.82	± 3.36
	C_1	10	10.83	± 1.95	-13.85	± 2.30
	C_2	10	10.80	± 3.94	-13.57	± 4.15
	C_3	10	11.49	± 1.41	-14.06	± 3.34
Combined		250	10.34	± 2.62	-15.38	± 3.40

Table 2: Mean free energies of activation and reaction for the expanded schemes. The individual k-means clusters are labelled XC (all in kcal mol^{-1}).

Scheme	Cluster	# of frames	$\Delta G^{\ddagger}$	σ	ΔG_{rxn}	σ
$\mathbf{S}_1$+$\mathbf{S}_6$+$\mathbf{S}_7$		100	10.40	± 2.63	-15.48	± 3.44
$\mathbf{XS}_2$		148	10.17	± 2.75	-15.64	± 3.44
$\mathbf{XS}_3$		173	10.35	± 2.70	-15.46	± 3.50
$\mathbf{XS}_4$		186	10.39	± 2.64	-15.36	± 3.37
$\mathbf{XS}_5$						
	XC_1	92	10.25	± 2.63	-16.16	± 3.92
	XC_2	89	10.30	± 2.85	-15.00	± 3.05
	XC_3	69	10.50	± 2.28	-14.85	± 2.83
$\mathbf{XS}_8$						
	XC_1	77	10.42	± 2.52	-15.38	± 3.54
	XC_2	81	10.27	± 2.86	-15.05	± 3.39
	XC_3	92	10.32	± 2.49	-15.69	± 3.24
Combined		250	10.34	± 2.62	-15.38	± 3.40