Title

Predictive methods for computational metalloenzyme redesign – a test case with carboxypeptidase A

Journal

Physical Chemistry Chemical Physics, 18(46)

ISSN

0956-5000

Authors

Valdez, Crystal E
Morgenstern, Amanda
Eberhart, Mark E
et al.

Publication Date

2016-11-23

DOI

10.1039/c6cp02247b

Peer reviewed

PCCP

PAPER

View Article Online
View Journal

Cite this: DOI: 10.1039/c6cp02247b

Predictive methods for computational metalloenzyme redesign – a test case with carboxypeptidase A†

Crystal E. Valdez,[a] Amanda Morgenstern,[b] Mark E. Eberhart*[b] and Anastassia N. Alexandrova*[ac]

Computational metalloenzyme design is a multi-scale problem. It requires treating the metal coordination quantum mechanically, extensive sampling of the protein backbone, and additionally accounting for the polarization of the active site by both the metal cation and the surrounding protein (a phenomenon called electrostatic preorganization). We bring together a combination of theoretical methods that jointly offer these desired qualities: QM/DMD for mixed quantum-classical dynamic sampling, quantum theory of atoms in molecules (QTAIM) for the assessment of electrostatic preorganization, and Density Functional Theory (DFT) for mechanistic studies. Within this suite of principally different methods, there are both complementarity of capabilities and cross-validation. Using these methods, predictions can be made regarding the relative activities of related enzymes, as we show on the native Zn^{2+}-dependent carboxypeptidase A (CPA), and its mutant proteins, which are hypothesized to hydrolyze modified substrates. For the native CPA, we replicated the catalytic mechanism and the rate in close agreement with the experiment, giving validity to the QM/DMD predicted structure, the DFT mechanism, and the QTAIM assessment of catalytic activity. For most sequences of the modified substrate and tried CPA mutants, substantially worsened activity is predicted. However, for the substrate mutant that contains Asp instead of Phe at the C-terminus, one CPA mutant exhibits a reasonable activity, as predicted across the theoretical methods. CPA is a well-studied system, and here it serves as a testing ground for the offered methods.

Received 5th April 2016,
Accepted 22nd September 2016

DOI: 10.1039/c6cp02247b

www.rsc.org/pccp

I. Introduction

Nature provides well-folded protein scaffolds that catalyze some of the most difficult reactions known to humankind. Harnessing this catalytic power of enzymes is a goal within the enzyme design community, in particular to provide the opportunity to create "greener" ways to produce molecules of interest that encompass pharmaceutical and industrial scale production.[1,2] Introducing slight modifications to naturally occurring proteins allows one to shift functionality by tuning properties such as substrate specificity, inducing stereospecific reactions,[3] altering protein stability, and even catalyzing non-native reactions.[4] In particular, computational redesign of natural enzymes is a powerful technique. Having the computational resources and capabilities to redesign natural enzymes through small adjustments in the active site or second coordination sphere residues, altering the function and/or catalytic ability without designing the protein from the ground up and worrying about the ability of the protein to fold and/or remain stable proves to be a prevailing strategy to protein design. Given the complexity of the problem, it is remarkable that modern computational design grows to become competitive with the more traditional protein design techniques such as directed evolution (not discussed or reviewed in this paper, since it is not the focus). The computational design and redesign of enzymes without catalytic metals is vast.[5–12] Many studies demonstrate that the substitution of a single to just a few amino acids can alter catalytic activity and specificity. For example, the substrate preference of a serine peptidase was altered to favor a new peptide substrate with a proline–glutamine (PQ) dipeptide motif over the native proline–arginine or proline–lysine (PR or PK) substrates.[13]

Partly due to the methodological limitations, efforts to design metal-containing biocatalysts are rare, especially on the computational front.[14,15] In a broader context of creating

[a] Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, CA, 90095, USA. E-mail: ana@chem.ucla.edu
[b] Molecular Theory Group, Colorado School of Mines, Golden, Colorado 80401, USA. E-mail: meberhar@mines.edu
[c] California NanoSystems Institute, Los Angeles, CA, 90095, USA

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c6cp02247b

metal-containing biocatalysts *via* inserting homogenous catalysts in macromolecular scaffolds, there is a thriving community.[16–19] It would be of even greater interest to have metals bound to the amino acid ligands, like in the natural enzymes. It is also of interest to strategically modify natural metalloenzymes with the native metals for operating on different substrates. In a purely experimental effort, Edge *et al.* performed mutagenesis on a zinc-containing peptidase, specifically targeting one residue in the binding pocket of the protein and changing the polarity of the pocket.[20] There are some successful efforts in *de novo* metalloprotein design and repurposing.[21–24] The function of a family of *de novo* proteins, di-iron oxo-proteins or "Due Ferri" proteins,[25] was altered through modifications of a loop far from the active site, mutations chosen to redirect the enzyme's intended function.[26] The addition of an extra metal-binding residue and adjusting the substrate cavity switch a *due ferri* protein originally intended to oxidize hydroquinones to catalyze the selective *N*-hydroxylation of arylamines.[27]

While design efforts focused on metal-free enzymes have yielded many useful *in silico* tools, these tools are insufficient for metal-containing enzymes.[28–30] When designing metalloenzymes, it is critical to properly treat the electronic structure of the metal. It is also desirable to efficiently sample the protein backbone, since multiply charged cations create strong polarization in the binding site and subsequent restructuring might be significant. Additionally, electrostatic preorganization, which is critical for the function of all enzymes,[31,32] is particularly sensitive in metalloenzymes, as the mutual polarization between the protein and the active site is significant. Our focus here is on the computational methodology that would be free of the aforementioned limitations, and enable predictions concerning altered and artificial metalloenzymes.

Using a suite of principally different and complementary computational tools, we redesign the natural Zn^{2+}-dependent hydrolase, Carboxypeptidase A (CPA) (E.C. 3.4.17.1),[33–35] to hydrolyze a non-native peptide substrate. CPA is an exopeptidase found in the gut that preferentially cleaves C-terminal aliphatic and aromatic amino acids from dietary proteins.[36] Because CPA was crystallized early on, compared to other members of the zinc hydrolase superfamily, it served as a prototype for understanding catalysis and substrate binding of such enzymes.[37,38] Therefore, much is known about the structure and the function of CPA, and it provides a well-studied model system. Our goal is to modify the substrate and then predict mutations to the protein that lead to proper positioning of the substrate with respect to the catalytic Zn and important residues in the active site, and create

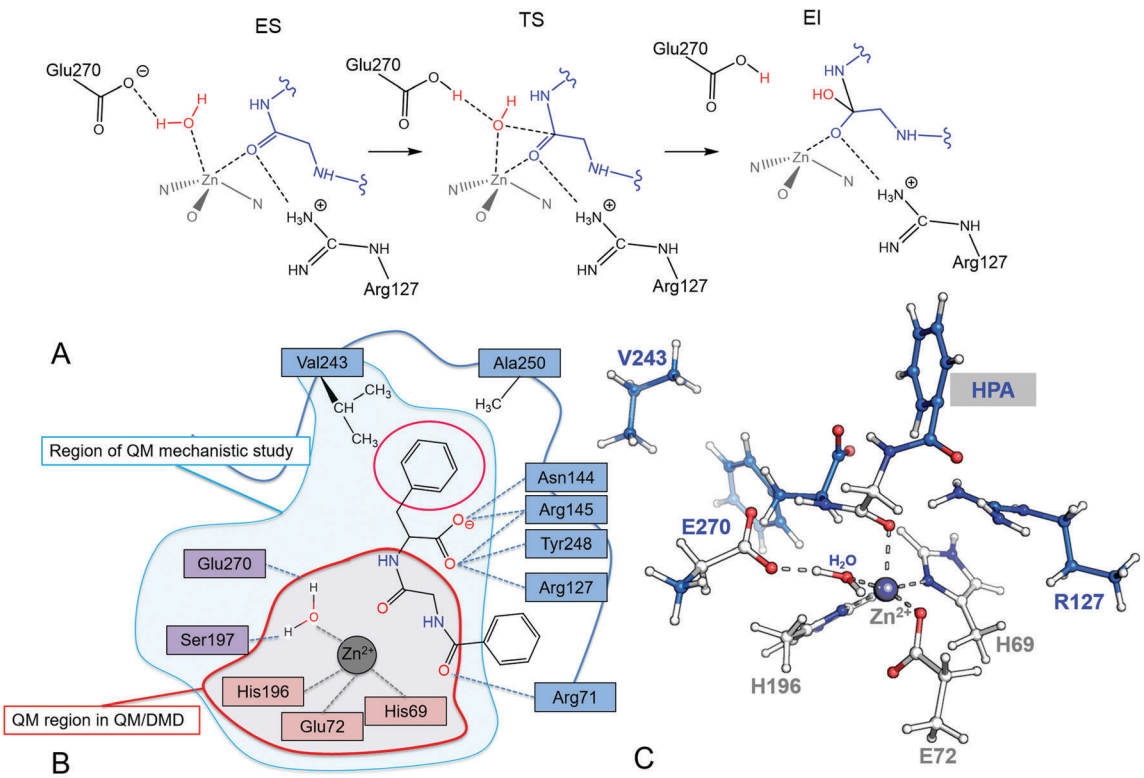

Fig. 1 (A) A schema of the water-promoted mechanism of peptide hydrolysis in CPA (the peptide substrate is truncated for clarity) with enzyme-substrate (ES), transition state (TS) and enzyme-intermediate (EI) scheme structures. Both Arg127 and Glu270 are thought to play vital roles in the reaction mechanism; Glu270 provides a hydrogen bond acceptor for the zinc-bound water and Arg127 stabilizes the negative charge that develops in the tetrahedral intermediate (the four membered ring formed between the substrate carbonyl, water oxygen, and zinc ion). (B) Active site of CPA with the region included in the QM mechanistic study and the truncated region used in QM/DMD simulation. The red boxed residues refer to residues coordinating the metal, the purple boxed ones refer to residues stabilizing the Zn-bound water and the blue boxed ones represent the S′ subsite residues. The pink circle indicates the Phe side chain that binds in the hydrophobic S′ subsite. (C) 3-D view of the active site and regions used in computations: grey – QM/DMD-region and blue – fragment used in the QM mechanistic study.

a charge density distribution in the active site conducive to proper reactivity, resulting in favorable energetics of the catalyzed reaction.

The active site of CPA consists of one Zn^{2+} ion coordinated by two histidines (His69 and His196), one glutamate residue (Glu72), and one water molecule, a common motif for catalytic zinc sites.[38,39] Zn^{2+} acts as a Lewis acid, both lowering the pK_a of the bound water molecule to facilitate the formation of a Zn-bound hydroxide, a more potent nucleophile than a water, and polarizing the carbonyl oxygen of the bound peptide substrate. A higher resolution crystal structure of CPA suggests that the zinc-bound water is in its neutral state.[40] Several computational works, based on small cluster models and sometimes semi-empirical methods, have shed light on the mechanism of peptide hydrolysis in CPA.[41–44] The water-promoted mechanism, schematically depicted in Fig. 1A, is believed to be the most energetically favorable.[43,44] An alternative mechanism proposed for CPA, called the anhydride mechanism, is thought to compete with the water-promoted mechanism for hydrolysis of non-natural substrates, such as esters.[45]

The C-terminal peptide specificity of CPA comes from the highly organized hydrogen bond network within the binding pocket that involves interactions between the terminal carboxylic acid, and residues Asn144, Arg145, Tyr248 and Arg127 (Fig. 1B).[46] Furthermore, the residue specificity comes from the hydrophobic pocket, also known as the S' subsite, that is responsible for the preference for aromatic and aliphatic side chains.[46] Additionally, Arg127 interacts not only with the terminal carboxylic acid, but also with the carbonyl oxygen of the peptide backbone coordinated to zinc, an interaction believed to contribute to catalysis and substrate binding.[46,61] During the course of the reaction, Arg127 stabilizes the oxyanion developed on the tetrahedral intermediate. Thus, in the design, the goal is to preserve these valuable interactions. In this study, based on the predicted structures, reaction barriers, and geometry of the electron density, a particular mutant of CPA (V243R) was predicted to be active toward hydrolysis of hippuryl-L-aspartate.

II. Methods

This study employs several different methods that collectively aim to tackle the challenges described in the Introduction in metalloenzyme design and modeling. We first outline their role in this work, and ways of cross-validating the results, and then present more technical details.

One part of the metalloprotein design problem is structural: predictions of metal coordination, protein backbone structures, and substrate binding require considering the induced fit effect. For the structural equilibration this study employs our QM/DMD[48] method, where DMD is short for discrete molecular dynamics.[48–51] QM/DMD is a hybrid method, a variant of QM/MM, that efficiently captures metalloprotein dynamics on the order of tens of nanoseconds, in conjunction with the quantum mechanical description of the active site. The intricate details of the QM/DMD method are rigorously described elsewhere[48] but the necessary theoretical framework for this work is provided in the Technical details section. QM/DMD has been shown to perform exceptionally well for recapitulating and recovery of native protein structures down to the subtle structural details at the active sites,[48] electronic properties of the bound metals,[49] protein conformational responses to substrate binding and metal replacement,[48,53] and mechanistic studies of enzymatic reactions.[48,54] The main attractive feature of QM/DMD is its affordability as compared to that of other capable QM/MM methods. The strength of this method is in part due to DMD, a remarkably successful classical sampling method for biomolecules.[50,51,55–60] Additionally, the speed of QM/DMD is due to the construction of the QM-DMD boundary (see below). QM/DMD allows us to address the vast timescales that metalloenzymes operate under, from the chemical transformation to the large-scale motions that occur on the nano-, milli- and even second timescale. Solvation in QM/DMD is treated implicitly, both within DMD – in an averaged way, through parameterization of the force field, and within QM. The structures of the active site with the bound substrates, obtained from QM/DMD trajectories, will be subjected to geometric examination and charge density analysis, as well as the mechanistic study, for predictions of relative catalytic activity or inactivity.

A crucial aspect to assess, and eventually design for, is the electrostatic preorganization for stabilization of the rate-determining transition state (TS). A sensitive probe for this could be the charge density ($\rho(r)$) in the active site. We are, therefore, interested in the response of the charge density to the changes in the substrate and the enzyme, and observing how this response aligns with the changes in activity. We use the quantum theory of atoms in molecules (QTAIM) and recent extensions to QTAIM, which can be employed to assess bonding interactions based on the topology and the geometry of $\rho(r)$.[61] We will determine how close each active site model is to the transition state in terms of $\rho(r)$. This should correlate with enzyme efficiency.

The last computational component that we will use for the assessment of metalloenzymes is the mechanistic study by means of Density Functional Theory (DFT).

The three-fold computational approach will characterize proteins as catalytic or not catalytic with respect to given substrates in three different ways: in terms of structure (QM/DMD), in terms of the charge density (QTAIM), and in terms of the actual calculated mechanism and reaction barrier (DFT). The results of the three approaches have some mutual dependence. However, they are based on principally different physical observables. The agreement between the three predictions will increase our confidence in the results. We additionally note that the QM/DMD and QTAIM methods have not been previously applied to the computational metalloenzyme design problem.

Technical details

QM/DMD simulations. In QM/DMD simulations, the system is split into three regions: the QM-only, the DMD-only, and the shared QM-DMD region (Fig. 1B and C). DMD employs discrete step function potentials in place of continuous potentials in

traditional MD sample methods, which allows for much faster sampling.[48–52] DMD samples the entire protein except for the metal and its immediately coordinated shell, which is held frozen during the sampling. Due to this choice, there is no need for the parameterization of the force field to describe the metal, which is then free to change the coordination environment in the course of the simulation in the QM phase. In the DMD phase, the protein is subjected to a simulated annealing procedure that removes any clashes introduced by the PDB structure and prevents the protein from being trapped in local minima (thoroughly outlined in ref. 48). After each annealing phase, the protein is held at a constant temperature T of 0.10 kcal (mol k_B)$^{-1}$ ($\sim$50 K) for the remaining DMD portion of the simulation. This low temperature is fairly standard in DMD, as parameterized and extensively tested by the DMD developers to be efficient in providing well-equilibrated and extensively represented ensembles.[48] The DMD phase permits for 10 000 time units (t.u.), which is equivalent to $\sim$0.5 ns, and then moves on to the QM machinery. Designated QM-only and QM-DMD boundaries are cut from the protein, as shown in Fig. 1B. The QM part of the simulations consists of periodic relaxations following the *ab initio* gradients for nuclear motions of the larger active site that includes the metals, their ligands, and the substrate, and may include other important amino acids near the reactive center (shared QM-DMD region). The QM/DMD boundary is thus "breathing", and goes around the metal and its immediate coordinating atoms, or around the larger active site, depending on the stage of the simulation. The chosen active site leaves the dangling edges of the peptide and other residues to be managed by DMD, where the extended hydrogen-bonding network can be described. The presence of the shared QM-DMD region permits efficient communication between the QM and DMD regions where geometric information can be passed between the regions. Also, the DMD potential is slightly re-parameterized on-the-fly, placing the centers of the DMD potential wells as dictated by the QM optimization. The DMD and QM stages alternate to convergence, which is tracked by QM and DMD energies, and backbone root mean squared deviation (RMSD) (Fig. 2). It is typical for systems near their equilibrium to converge within 20 iterations, roughly corresponding to 10 ns,[48,53–55,58–60] and for systems undergoing more serious changes (as is expected in the present case) – on the order of 50–100 iterations, roughly corresponding to 50 ns of dynamics.[53–55]

The initial structure of CPA was obtained from the Protein Data Bank (PDB) (CPA PDB code: 6CPA[62]). The PDB contains a bound inhibitor, o-(((1r)-((*n*-phenylmethoxycarbonyl-l-alanyl)amino)-ethyl)hydroxyphosphono)-l-benzylacetic acid (ZAAP(O)F), which served as a scaffold to build the substrate, as done in other computational works.[46,55,56,60,63]

QTAIM analysis of native and mutant active sites. The electron charge density has been shown by Kohn to be near-sighted,[64a,b] meaning that $\rho(r)$ can be accurately reproduced using small cluster models in the active sites of enzymes, as long as the nuclear coordinates were obtained using full protein modeling, such as the QM/DMD methods used here. Therefore, we strive to capture the electrostatic preorganization of the active site using the charge density analysis methods of QTAIM and its extensions. In this work, we focus on the QTAIM results from the reactant state active site obtained from the lowest energy trajectory in the QM/DMD study. An investigation of how sensitive the charge density is to changes in the nuclear

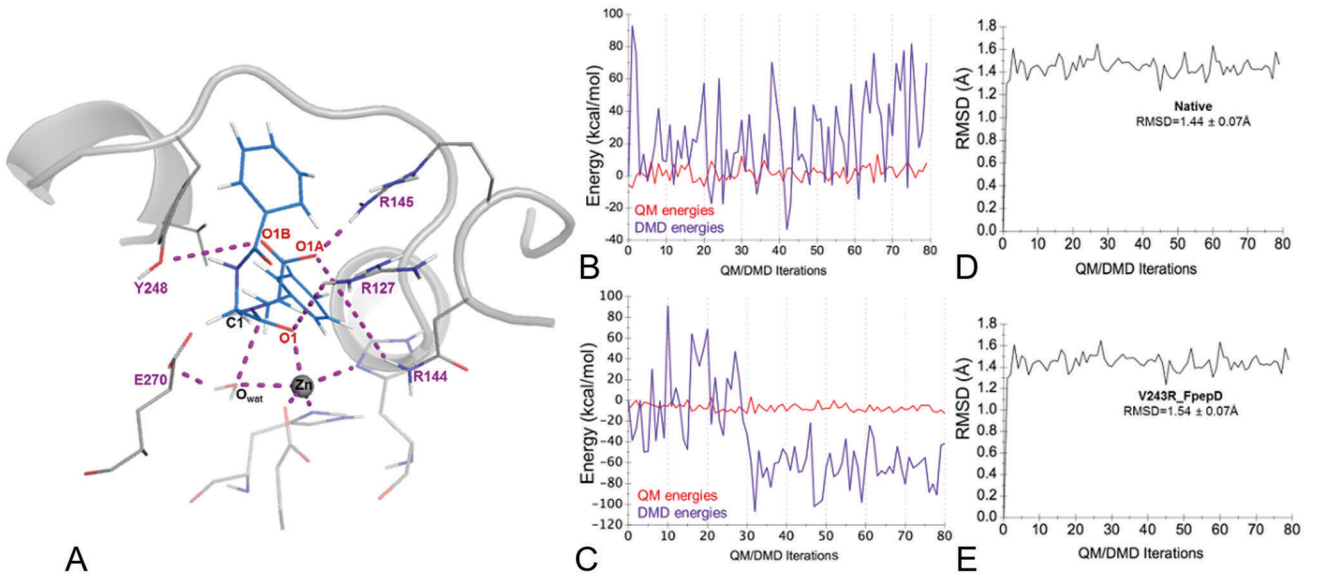

Fig. 2 (A) Representative structure of the binding pocket of native CPA and native substrate hippuryl-L-phenylalanine (hippuryl-L-Phe), coming from the QM/DMD trajectory. Important contacts for binding and positioning of the substrate and evaluating potential mutants are shown. Distances are given in Table 1. QM/DMD trajectories of the native (B) and mutant (C) peptidase/peptide systems, and the respective backbone root mean squared deviations (RMSDs) (D and E). The reference structure used for calculating the RMSDs is the crystal structure (PDB: 6CPA) with respective docked substrates that have undergone a partial QM optimization before any DMD sampling.

coordinates (such as by looking at a distribution of charge density functions from multiple QM/DMD trajectories) is currently underway, but beyond the scope of this paper.

QTAIM classifies charge density in terms of topological structure. Critical points (CPs), *i.e.* points where the gradient of the charge density vanishes, can be used to determine structural and reactivity information. There are four types of stable CPs: maxima at atomic nuclei (called nuclear CPs), cage CPs – local minima in $\rho(r)$ occurring inside atomic cages, saddle points occurring inside atomic rings (ring CPs), and a second type of saddle point which occurs between bonded nuclei, called bond CPs. The magnitude of the charge density at bond CPs has been found to correlate with bond strength.[65] Ridges of charge density (bond paths) are present between bond and nuclear CPs wherever a topological bonding interaction exists. QTAIM also provides an unambiguous approach to partitioning a molecule into its atoms or atomic basins, and by integrating the number of electrons in a basin a Bader charge can be calculated.[66] Recent developments to QTAIM allow for a finer partitioning of space into volumes bounding CPs other than nuclear CPs. For example, bond bundles are volumes that contain a single bond CP.[67,68] The shape, size, and valence electron count of bond bundles have been used to understand chemical reactivity and stability.[69,70]

The electronic charge densities of the native enzyme and V243R, hippuryl-L-aspartate mutant with the bound substrates (in the ES, TS, and EI states) were calculated using the Amsterdam Density Functional Package (ADF) version 2014.01[66,71–73] using similar computational parameters as in the QM mechanistic study. Single point calculations were performed on the reactant, transition state, and intermediate complexes using the nuclear coordinates from the stationary states obtained in the mechanistic study described below. A TPSS functional and a COSMO solvent model with a dielectric of 20.0 were utilized (a dielectric of 4.0 was also tested and showed no appreciable differences).[73] A double ζ quality basis set, DZP, was employed for all atoms except the metal, which was calculated using a triple ζ quality basis set, TZP.[74] The Bondalyzer add-on package in Tecplot[75] was then used to analyze the calculated charge densities.

Mechanistic studies of native and mutant CPA. The larger active site (Fig. 1B) was used for the mechanistic investigation on the cluster model, after protein equilibration. The QM calculations in the mechanistic study and within QM/DMD are done using *Turbomole*,[76] utilizing the Resolution of Identity[77] and multipole accelerated resolution of identity[78] to speed up the calculations. The unrestricted DFT[79] was used for QM, due to the size of the mechanistically important region. For most calculations, the exchange and correlation is treated with the TPSS functional[80,81] with a double ζ quality basis set, def2-SVP,[82] for H, C, N, O and S and a triple ζ quality basis set, def2-TZVPP,[83] for the metal, Zn. To account for solvation and screening effects introduced by the protein, the COSMO model with a dielectric of 20.0 was utilized. (4.0 was tested and showed no energetic or structural differences) The nature of each stationary point was confirmed using a full Hessian calculation, with selected parts being frozen to mimic the constraints imposed by the protein. The frozen parts were the Cα atoms with capping hydrogens, and the two N atoms on the peptide. The transition state was identified as one imaginary frequency with the normal mode going along the reaction coordinate, where the corresponding reactant and product states are truly two local minima structures connected by the computed transition state. The transition state of interest here is the nucleophilic attack of the zinc-bound water on the carbonyl carbon of the peptide. Experimentally explicating the specific mechanism for CPA is very difficult but many computational works[46,91] agree with the location and characterization of this transition state. Single point calculations with BP86,[84–86] B3LYP,[87,88] and TPSSh[89] functionals, and the def2-TZVPP basis set were performed on all stationary points (Table S1, ESI†). Natural population analysis (NPA)[90] calculations were done to get the partial charges on all the stationary points (listed in Table S2, ESI†).

III. Results and discussion

Choice of mutations

The mutations in the protein/peptide systems were chosen by intuition with the aim of achieving residue-specific interactions in the S′ subsite, while retaining the proper positioning of the substrate with respect to Zn and other catalytic residues. For the binding pocket, the following mutations were tried: (1) V243R, hippuryl-L-glutamate (2) V243K, hippuryl-L-glutamate, (3) V243R, hippuryl-L-aspartate, and (4) V243K, hippuryl-L-aspartate. With these choices, we strived to span the different molecular properties of available amino acids while redesigning a substrate binding pocket to switch from hydrophobic to hydrophilic. As mentioned earlier in a purely experimental approach, Edge *et al.* utilized site-directed mutagenesis to mutate away residues in the binding pocket to reverse the polarity of substrate/protein interaction.[20] In a similar spirit, we are mutating from bulky hydrophobic groups to charged ones, so as to completely alter the environment of the pocket, and hypothesize these interactions to be favorable. We assess the structure, and then the catalytic performance of these variants computationally.

Structural comparison of CPA native and mutant proteins

From the four mutant proteins tested using QM/DMD, two yielded stable RMSDs during the simulation, an appropriate hydrogen-bonding network in the binding pocket, and a reactive substrate orientation. For the V243R, hippuryl-L-glutamate (abbreviated V243R_FpepE) mutant, the reactive carbonyl of the peptide substrate loses coordination with the zinc multiple times throughout the duration of the simulations (Fig. S1A, ESI†), possibly due to the extra length of the Glu residue on the substrate. Without this key interaction, the enzyme is no longer active. In the V243K, hippuryl-L-aspartate (abbreviated V243K_FpepD) mutant, the substrate in the binding pocket rotated into a non-reactive orientation and maintained that position for the duration of the simulation (Fig. S1B, ESI†). These structural peculiarities justified the decision to remove them from our mutant search and focus on

the two remaining mutations, V243R, hippuryl-L-aspartate and V243R, hippuryl-L-glutamate, abbreviated V243R_FpepD and V243K_FpepE, respectively.

QM/DMD simulations of the V243R_FpepD and V243K_FpepE mutants both converged (simulation data shown for V243R_FpepD in Fig. 2C and E), with hydrogen bonds' placements and lengths in the substrate binding pockets that agree well with previous experimental and computational studies on the native CPA (Table 1).[46,47] In particular, the distance between the carbonyl oxygen and zinc, Zn–O1, (Fig. 1B) remained in a reactive range: 2.10 ± 0.15 Å, 3.08 ± 0.12 Å, and 3.03 ± 0.39 Å, for the native, V243R_FpepD, and V243K_FpepE variants, respectively. These results can be compared to previous computational studies that used a parameterized force field to model the Zn^{2+}, producing the Zn–O1 distance of 4.19 Å.[91] The hydrogen bond between the R127 and carbonyl oxygen bound to zinc, R127–O1, is a vital interaction for the catalytic activity of the enzyme.[46,63,91] The native CPA shows the R127–O1 bond distance to be 2.52 ± 0.46 Å (standard deviation resulted from QM/DMD sampling), comparable to the previously reported 2.64 ± 0.52 Å. The two mutant peptidases display larger deviations in R(R127–O1), and longer hydrogen bonds and interatomic distances within the binding pocket. The Zn–O1 distances are also longer for the mutant proteins as compared to the native, ~0.8 Å longer, most likely due to the longer distances in the peptide binding pocket. The V243R_FpepD variant is closer to the native CPA, with an R(R127–O1) distance of 3.46 ± 0.73 Å, while V243K_FpepE has an R(R127–O1) distance of 4.58 ± 0.69 Å, which is quite high. We thus predicted V243R_FpepD to be the best candidate to show catalytic activity. We also noted that the DMD energy for V243R_FpepD exhibits a switch to a new state at iteration 32 (Fig. 2C). This switch cannot be attributed to a single apparent change, such as a development of a new contact in the protein, and we therefore concluded that it is a collective result of multiple small adjustments that correspond to a more stable structure of the protein.

From the structural data obtained using QM/DMD, the specific interactions in the binding pocket can be monitored. CPA naturally has a slight preference for bulky hydrophobic groups. For the native enzyme, we track the interatomic distances of the C-terminal phenylalanine (CP3-6) and the pocket valine, V243 (CG1, CG2, CB) (Fig. 3 and Table 2). Ranging from 4.97 ± 0.94 Å to 7.39 ± 1.22 Å, QM/DMD structures show that the hydrophobic interactions are maintained throughout the simulation, with large standard deviations that could be a result of the non-specific nature of the binding pocket, accommodating a wide variety of hydrophobic residues (Table 2).

For the V243K_FpepE, the distance between the N of the NH_3 group of the pocket lysine, K243, and the O of the C-terminal glutamate side chain is 5.29 ± 0.64 Å. This distance being so large (not within a meaningful range for a strong H-bond), combined with the large R127–O1 distance, also suggests that this variant is not viable.

For the V243R_FpepD mutant, the shortest hydrogen bond between the peptide and R243 (H of arginine and OD2 of the peptide) is 3.61 ± 0.99 Å. All the hydrogen bonds and interatomic distances between the peptide and the pocket residue of the protein, specifically the two donors from the OD1 and OD2 of the peptide and the acceptors from the NH_2 groups from R243, fluctuate between ~3.6 and 5.8 Å (Fig. 3 and Table 2). This range of distances further validates this mutant protein as a potentially active enzyme, subject to further verification. QM/DMD sampling was done with the V243R mutant enzyme/native substrate system and yielded very unfavorable hydrogen-bonding distances that will presumably render the enzyme inactive. Therefore we hypothesize that the mutant enzyme will only be an active catalyst for substrates similar to FpepD.

Comparison of native enzyme and V243R_FpepD in the reactant states through QTAIM

Two key components to the catalytic activity of CPA are the correct positioning of the water molecule for nucleophilic attack and the stabilization of the partially charged substrate carbonyl. Fig. 4 shows QTAIM bond paths of interest for the enzyme–substrate complexes. In both the native CPA and the V243R_FpepD mutant, the water molecule is positioned near the substrate carbonyl due to coordination with the Zn ion and hydrogen bond donation to E270. The native enzyme has additional hydrogen bonding from the second water hydrogen to E72. In V243R_FpepD, E72 is doubly coordinated to the Zn ion, rather than acting as a hydrogen bond acceptor for the water molecule. The donation of the second hydrogen atom in the native enzyme results in a more negative Bader charge on the water oxygen (see Table 3). This should make the water

Table 1 Important interatomic distances in the substrate binding pocket of native and mutant CPA, compared to previous QM/MM simulation (ref. 91). Standard deviations come from averaging over the converged portions of the QM/DMD trajectories. In addition to the hydrogen-bonding interactions, zinc–peptide carbonyl oxygen (Zn···O1), zinc–water (Zn···O_{wat}) and water–peptide carbonyl carbon (O_{wat}···C1) distances are presented

Atomic interactions	Native (Å)	V243R_FpepD (Å)	V243K_FpepE (Å)	Ref. 91 (QM/MM ES)
R127 (1HH1)···O1	2.52 ± 0.46	3.46 ± 0.73	4.58 ± 0.69	2.64 ± 0.52
N144 (HND)···O1A	6.57 ± 1.29	5.45 ± 0.98	5.84 ± 0.46	1.85 ± 0.16
R145 (1HH1)···O1A	2.09 ± 0.34	3.03 ± 1.40	2.80 ± 0.50	1.77 ± 0.18
R145 (1HH1)···O1B	3.56 ± 0.69	4.35 ± 0.93	4.13 ± 0.69	1.77 ± 0.18
Y248 (HO)···O1B	4.62 ± 0.97	2.59 ± 0.76	2.05 ± 0.20	1.67 ± 0.11
Zn···O1	2.10 ± 0.15	3.08 ± 0.12	3.03 ± 0.39	4.19 ± 0.41
Zn···O_{wat}	2.35 ± 0.33	2.20 ± 0.00	2.20 ± 0.00	2.13 ± 0.08
O_{wat}···C1	2.45 ± 0.35	3.00 ± 0.03	2.69 ± 0.19	2.83 ± 0.22
E270 (OE1)···O_{wat} (H)	3.16 ± 0.75	2.13 ± 0.38	2.55 ± 0.60	1.77 ± 0.14
R127 (NH1)···O1	3.38 ± 0.36	4.01 ± 0.64	4.68 ± 0.51	n/a

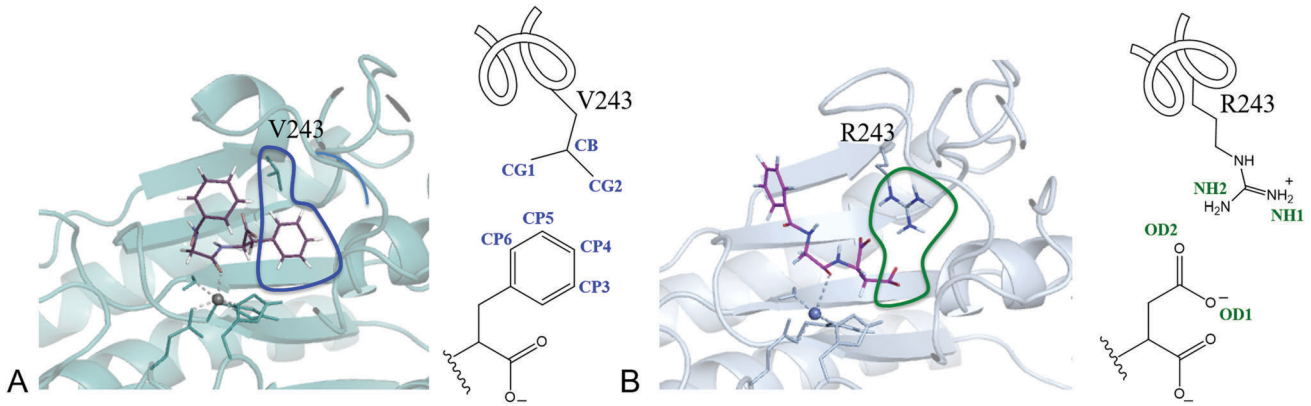

Fig. 3 3D structures and cartoon representations of truncated bonding pockets of the native CPA (A) and V243R_FpepD mutant (B).

Table 2 Important interatomic distances of residues in the binding pocket of native CPA and V243R_FpepD mutants. See Fig. 3 for atom label names

Native binding pocket	Interatomic distances (Å)	Mutant protein binding pocket	Interatomic distances (Å)
CG1 V243···CP3 PEP	7.09 ± 1.15	NH1 R243···OD1 PEP	5.17 ± 0.98
CG1 V243···CP4 PEP	6.18 ± 1.06	NH1 R243···OD2 PEP	5.07 ± 0.73
CG1 V243···CP5 PEP	4.97 ± 0.94	NH2 R243···OD1 PEP	4.34 ± 0.66
CG2 V243···CP3 PEP	7.39 ± 1.22	NH2 R243···OD2 PEP	3.73 ± 0.79
CG2 V243···CP4 PEP	6.38 ± 1.09	1HH1_ARG···OD1_PEP	4.93 ± 1.10
CG2 V243···CP5 PEP	5.15 ± 0.94	2HH1_ARG···OD1_PEP	5.85 ± 1.06
CG1 V243···CP3 PEP	7.09 ± 1.15	1HH2_ARG···OD1_PEP	4.16 ± 0.75
CG1 V243···CP4 PEP	6.18 ± 1.06	2HH2_ARG···OD1_PEP	4.52 ± 0.75
		1HH1_ARG···OD2_PEP	4.99 ± 0.80
		2HH1_ARG···OD2_PEP	5.82 ± 0.76
		1HH2_ARG···OD2_PEP	3.72 ± 0.93
		2HH2_ARG···OD2_PEP	3.61 ± 0.99

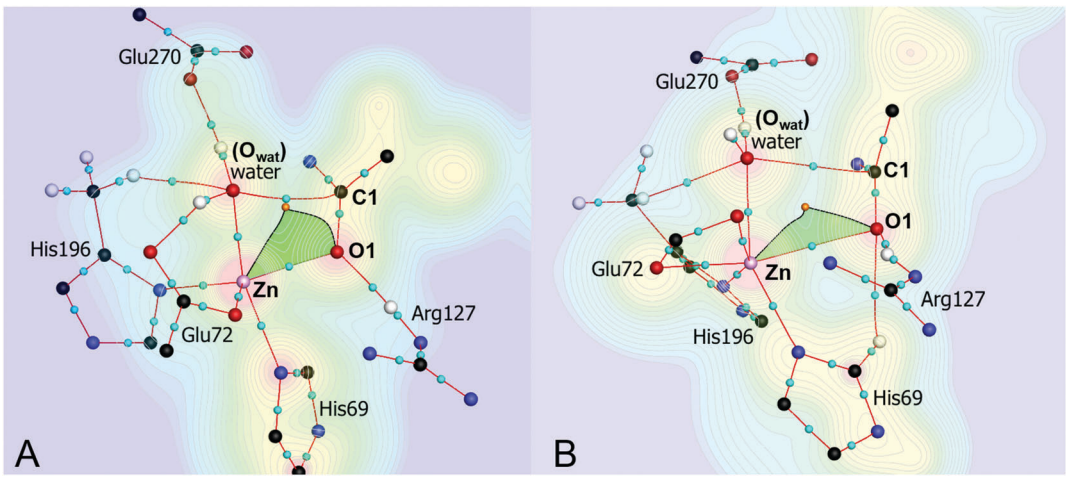

Fig. 4 Bond paths of interest in the native (A) and V243R_FpepD mutant enzyme (B), in the reactant state. Contours in $\rho(r)$ are drawn on a cut plane on a logarithmic scale from 10^{-3}–1 e Bohr^{-3}. Red lines indicate bond paths. The pictured portion of the Zn–O1 bond bundle is shaded green with black lines showing approximate edges. The following coloring scheme is used: Zn-purple, O-red, C-black, H-white, N-blue, bond CP-cyan, and ring CP-orange.

molecule a better nucleophile in the native enzyme as compared to V243R_FpepD.

The partially negative carbonyl oxygen on the substrate is stabilized through hydrogen bonding to R127 and coordination to the Zn center in both enzymes. The hydrogen bond from R127 to O1 in the reactants is not as strong in V243R_FpepD as compared to the native enzyme. This is indicated by the charge density at the H_R127–O1 bond critical point (CP), which has decreased from 0.0487 in the native enzyme to 0.0128 e Bohr^{-3} in the mutant protein (Table 4). Additionally, the O1–H_R127–N_R127 angle has decreased almost linearly in CPA from 175° to 147° in V243R_FpepD. However, there is an additional weak hydrogen bonding interaction between the carbonyl oxygen and H69 in the mutant enzyme.

Table 3 Bader charges of atoms involved in the chemical reaction of hydrolysis. See Fig. 4 for labeling

	ES		TS		EI	
	Native	Mutant	Native	Mutant	Native	Mutant
O_{wat}	−1.181	−1.135	−1.014	−1.038	−1.003	−1.003
C1	1.12	1.162	1.1	1.08	1.117	1.073
O1	−1.087	−1.103	−1.102	−1.102	−1.089	−1.085
Zn	1.314	1.308	1.297	1.304	1.25	1.262

Similarly, the native enzyme has greater stabilization of the carbonyl oxygen from the Zn ion as compared to V243R_FpepD. This can be seen in the amount of charge density at the Zn–O1 bond CP and also in the size of the Zn–O1 bond bundle. The charge density drops from 0.0575 in the native enzyme to 0.0183 e Bohr^{-3} at the bond CP in V243R_FpepD. A cut plane of a portion of the bond bundle in the plane of the Zn–O1-ring CP is shown in Fig. 4. These points define three vertices of the bond bundle that represents the volume of space the Zn–O1 bonding interaction occupies. The pictured portion of the bond bundle in the native enzyme is significantly larger than that in V243R_FPepD, indicating a more stabilizing bonding interaction.

In both the native CPA and the V243_FpepD mutant, the Zn ion and the hydrogen bond network are able to stabilize the partial charge on the substrate carbonyl, and promote the reaction. The slightly less stabilizing interactions from both hydrogen bonding and the coordination of the substrate carbonyl to the Zn ion in V243_FpepD suggest that the transition state shown in Fig. 1A will be more difficult to reach, predicting a higher activation energy.

Mechanistic study of CPA native, V243R_FpepD and V243K_FpepE mutants

Following structural and electronic comparisons of the native and mutant enzymes, we proceeded with a full QM mechanistic study of active site models. Previous works emphasize that the exact mechanism of CPA is not fully understood, it is particularly unclear as to whether the peptide hydrolysis is accomplished through a promoted water or anhydride mechanism. However, computations[44–46] have shown that the water-promoted mechanism has the lowest barrier for the rate-determining step, nucleophilic attack of peptide backbone carbonyl carbon by zinc-bound water (Fig. 1A). For mechanistic studies, the QM active site was extended to include the entire peptide, including the R127 residue in the binding pocket responsible for sequence specificity, and other residues known to stabilize the transition state (Fig. 1B and C). Also, only the rate-determining step is considered here due to the complication of the hydrogen source (zinc bound water or solvent) for the second step of the reaction: protonation of the backbone amide (Fig. 5). For the native and two mutant systems, a single structure from the QM/DMD trajectory was chosen based on the lowest QM and DMD energies.

For the native enzyme, the Zn–O1 distance begins at 2.109 Å and continues to shorten to 2.035 Å and 2.012 Å, as the mechanism proceeds through the nucleophilic attack on the carbonyl C1 by the Zn-bound water, from the reactants to the TS and to the tetrahedral intermediate (EI) (all the distances are given in Table S3, ESI†). The O_{wat}–C1 distance begins at 2.636 Å, shortens to 1.853 Å in the TS, and ends at 1.500 Å in the EI, indicating a slightly late transition state. The barrier for the transition state is calculated to be 13.3 kcal mol^{-1} (15.8 kcal mol^{-1} calculated with B3LYP), in a fair agreement with the experimentally derived value of 15.7 kcal mol^{-1} calculated using transition-state theory (Fig. 5).[46,47] The zinc-bound water detaches from the zinc as the tetrahedral intermediate forms and the C1–O_{wat} distance decreases to 1.500 Å, forming a new bond. The O1–H_{R127} bond shortens from 1.661 Å to 1.590 Å in the transition state as the hydrogen bond increases in strength due to the buildup of charge on that carbonyl oxygen (see Table 5). Our geometric parameters agree well with a previous study on a similar system using a parameterized QM/MM.[44–46]

For the V243R_FpepD mutant, similar trends in varying interatomic distances are present throughout the reaction mechanism. The barrier for this transition is 21.6 kcal mol^{-1}, higher than the native, but still in the range of known barriers for systems that catalyze similar hydrolysis reactions.[46] The O1–H_{R127} bond is slightly elongated in comparison to the native CPA, starting at 2.247 Å and shortening to 1.904 Å in the intermediate. This could be due to the restructuring of the binding pocket in response to the V243R mutation, which has a slight perturbation on the hydrogen-bonding network. Overall, as was suggested by both structural and QTAIM analyses, the mutant protein appears to be less active, according to our calculations. We note now that the QTAIM analysis made predictions regarding reactivity based on the ground state electron density. This might allow using QTAIM for ranking

Table 4 Charge density ($\rho(r)$) at the bond and ring critical points, and interatomic distances (in parentheses) of the bonding interactions of interest for the rate-determining step

	ES		TS		EI	
	$\rho(r)$ at CPs (interatomic distance)		$\rho(r)$ at CPs (interatomic distance)		$\rho(r)$ at CPs (interatomic distance)	
	Native	Mutant	Native	Mutant	Native	Mutant
C1···O_{wat}	0.0166 (2.636)	0.0155 (2.675)	0.0953 (1.853)	0.0817 (1.919)	0.217 (1.500)	0.204 (1.523)
C1···O1	0.352 (1.285)	0.368 (1.268)	0.338 (1.312)	0.342 (1.306)	0.301 (1.364)	0.298 (1.369)
Zn···O1	0.0575 (2.109)	0.0183 (2.640)	0.0699 (2.035)	0.0599 (2.104)	0.0762 (2.011)	0.0746 (2.021)
Zn···O_{wat}	0.0558 (2.134)	0.0596 (2.088)	0.0327 (2.400)	0.0395 (2.294)	n/a (3.175)	n/a (2.946)
$H1_{R127}$···O1	0.0487 (1.661)	0.0128 (2.247)	0.0596 (1.590)	0.0205 (2.038)	0.0412 (1.749)	0.0287 (1.905)
Ring CP	0.0166	0.0129	0.0262	0.0269	n/a	n/a
H_{H69}···O1	n/a	0.0157 (2.220)	n/a	0.0160 (2.218)	n/a	0.0127 (2.379)

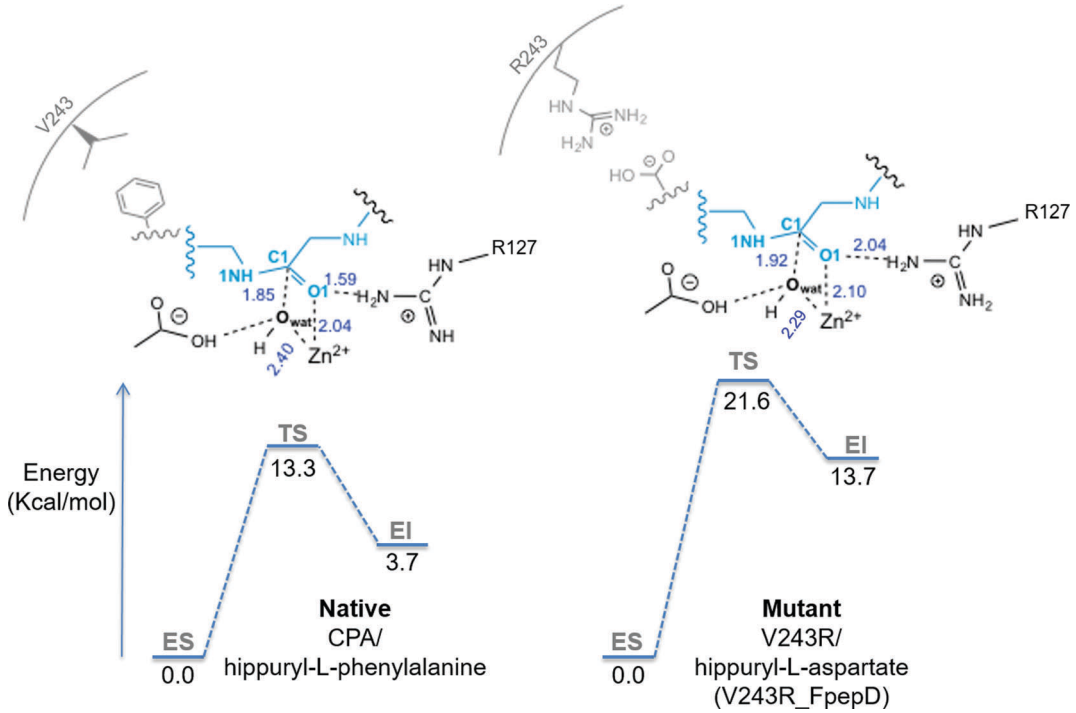

Fig. 5 Mechanism of native peptidase/peptide and mutant peptidase/peptide systems with representative transition state structures. Single point energies reported are calculated with TPSSh/def2-TZVPP for all atoms from TPSS/def2-SVP (C, O, N, H) and def2-TZVPP (Zn^{2+}) optimized structures.

Table 5 Geometric parameters of stationary points of the water-mediated hydrolysis of the native and V243R_FpepD mutant (in parentheses) obtained at the TPSS/def2-SVP (H, C, N, O) and def2-TZVPP (Zn^{2+}). Parameters are compared with ref. 46

	Native (Mutant – V243R_FpepD)			Ref. 46		
	ES	TS	EI	ES	TS	EI
Zn···O1	2.11 (2.64)	2.04 (2.10)	2.01 (2.02)	5.51	2.06	1.94
Zn···O_{wat}	2.13 (2.09)	2.40 (2.29)	3.18 (2.95)	1.93	2.19	3.30
C1···O1	1.29 (1.27)	1.31 (1.31)	1.36 (1.37)	1.24	1.30	1.37
O_{wat}···C1	2.64 (2.68)	1.85 (1.92)	1.50 (1.52)	5.38	1.88	1.37
N1···C1	1.66 (1.33)	1.59 (1.39)	1.75 (1.44)	1.34	1.37	1.44
O1···H_{R127}	1.66 (2.25)	1.59 (2.04)	1.75 (1.90)	1.83	1.89	1.67
O1···N_{R127}	2.70 (2.85)	2.65 (3.03)	2.74 (2.03)	n/a	n/a	n/a

related proteins according to their activity without costly TS calculations being required, as long as the general mechanism is known.

For the V243K_FpepE mutant, the mechanistic study yielded a reaction barrier for the rate-determining step of 28.7 kcal mol^{-1}. This barrier is too high, and therefore, we suggest V243K_FpepE to be considerably less active, as was also hypothesized based on its structure (see the discussion above).

Comparison of native enzyme and V243R_FpepD in the transition states through QTAIM

To provide further computational evidence for the relative catalytic abilities of the native and mutant enzyme, we performed QTAIM analysis on both the TS and EI complexes. As the reaction proceeds, the Zn–O_{wat} bonding interaction is eventually broken in the EI complex, opening the ring pictured in Fig. 4 (the EI image is provided in Fig. S3, ESI†). Topologically, the ring opening is achieved as the ring CP in the tetrahedral intermediate moves towards the Zn–O_{wat} bond CP in the TS (Fig. 6) until it ultimately annihilates the bond CP in the EI complex. While the ring CP in V243R_FpepD is geometrically closer to the Zn–O_{wat} bond CP in the reactant state, the charge density region between the ring and bond CP in the native enzyme is "flatter". The change in $\rho(r)$ between the ring-bond CPs is only 0.0392 e $Bohr^{-3}$ in the native enzyme while it is 0.0467 e $Bohr^{-3}$ in V243R_FpepD. In accordance with the electron-preceding picture,[92–95] it should be less energy intensive for the native enzyme ring CP to annihilate the bond CP than in the mutant enzyme, which agrees with our mechanistic study, and QTAIM predictions based on the reactant states.

The observed correspondence between values of $\rho(r)$ at bond and ring CPs in the reacting region combined with bond bundle analysis and the energetic barrier to the reaction outlines a potential strategy for predicting enzyme reactivity. $\Delta\rho(r)$ in the reactant states of related enzymes, for example designed computationally to catalyze the same reaction, serves as a probe of relative catalytic activity prior to the mechanistic study, as is shown here, and also was shown in a recent study of the metal-specificity in histone deacetylase.[96] This method is motivated by the Hammond postulate, which states that reaction barriers will be lower when the reactant state is "similar" to the transition state. The extensions to QTAIM used here provide a method for quantifying the similarity of reactant and transition states, through the amount of charge density at critical points that play an important role in the reaction coordinate.

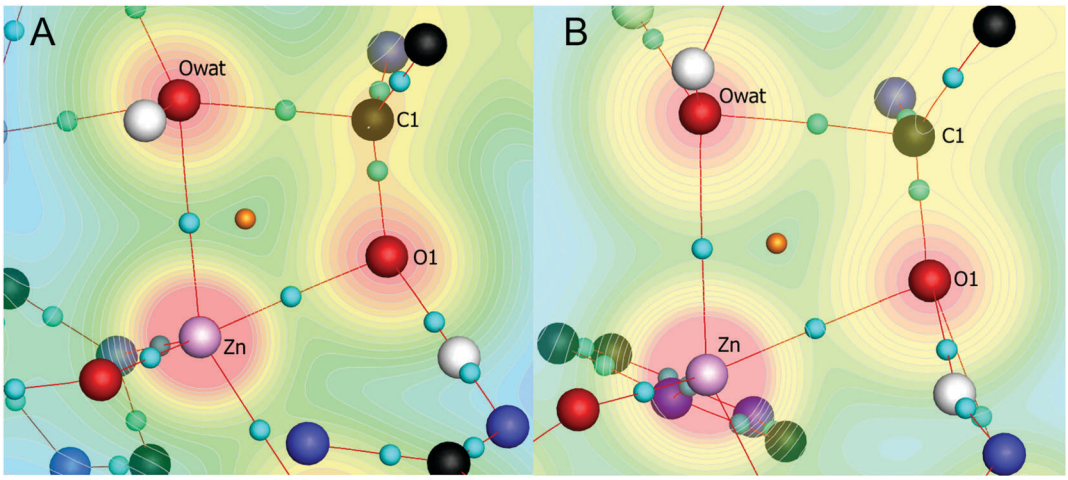

Fig. 6 Bond paths and critical points for both the native (A) and mutant protein (B) transition state. The ring CP (orange) has moved towards the Zn–O_{wat} bond CP in both enzymes. After the transition state, the ring CP converges with the Zn–O_{wat} bond CP, causing the two critical points to be topologically annihilated and the Zn–O_{wat} bonding interaction to break.

We have shown that in the case of CPA, $\Delta\rho(r)$ in the area of the reaction coordinate can be related to the reaction barrier for similar systems. This method can be directly used for the assessment of the activity of other related enzyme variants. The $\Delta\rho(r)$ probe is very sensitive and non-empirical, responding to the embedding in the protein charges, *i.e.* the electrostatic preorganization. Importantly, analysis can be performed on the reactant state alone, prior to performing more costly and often-tedious transition state calculations during the mechanistic study. While this method requires general knowledge of the reaction mechanism (*i.e.* knowing that the water–Zn bonding interaction will break as the reaction proceeds), it does not require full TS calculations. The only charge density data actually analyzed in this case are the difference in charge density between bond and ring CPs in the ES complex. Further in-depth investigation of the utility of this parameter for computational enzymology will be reported in the forthcoming more technical publication. In particular, we are currently investigating its sensitivity to the off-active site mutations.

IV. Concluding remarks

Multi-scale computational tools are vital to the success of designing bio-catalytic materials such as artificial metalloenzymes. The main point of this work is to bring together the necessary (and mostly novel) computational tools for metalloenzyme design, and to demonstrate their effectiveness on a trusted model system.

Our efforts begin with parsing experimentally determined X-ray structures of well-folded proteins into our fast and efficient hybrid dynamics software, QM/DMD, that employs a quantum mechanical description of the metal-containing active site coupled to sufficient backbone sampling reaching the nanosecond timescale. QM/DMD accurately models the small-scale electronic properties of the metal and large-scale protein motions induced by substrate binding and/or mutations to the protein sequence. Here we employed QM/DMD to redesign carboxypeptidase A (CPA) to selectively hydrolyze a specific peptide sequence. CPA is a Zn^{2+}-dependent exopeptidase that hydrolyzes C-terminal peptides with an aromatic or aliphatic side chain. The substrate docks into CPA *via* an extended hydrogen bond network, which is residue specific, and a hydrophobic pocket, where the interaction between the peptide substrate and protein is not residue specific. Through QM/DMD, we performed a series of *in silico* mutations to the hydrophobic pocket of CPA and the C-terminal residue of the peptide, to design a mutant protein/peptide system that would be catalytic. Additional analysis through extensions of QTAIM sheds light on the differences between the catalytic barriers in native and mutant CPA, particularly by investigating the topology and the geometry of the electronic charge densities of the two enzyme/substrate systems. Our application of QTAIM provides a direct probe of the electrostatic preorganization in the enzyme active sites, a critical component of enzymatic function, as was emphasized by Warshel and coworkers.[31,32] In this work, QTAIM results are shown to predict enzyme activity that agrees with the subsequent mechanistic study. Density Functional Theory (DFT) mechanistic studies were performed on cluster models, investigating specifically the water-promoted pathway for hydrolysis, of the native and various CPA mutants. The V243R_FpepD was selected for mechanistic analysis. For the rate-determining step of the hydrolysis, the native and V243R_FpepD mutant yield a barrier of 13.3 kcal mol^{-1} and 21.6 kcal mol^{-1}, respectively, in relative agreement with the experimental value of 15.7 kcal mol^{-1}[47] for the native enzyme. For the mutant protein, the catalytic activity appears to be lower, in agreement with structural and QTAIM predictions.

Finally, we would like to note that we chose CPA (in the native form) as a well-understood test case. The goal of the methodology employed in this work is to do future designs of artificial enzymes capable of non-physiological catalysis.

Based on the results of this work, we suggest that a combination of dynamic modeling with extensive sampling, QTAIM, and DFT can jointly offer a platform for redesigning natural enzymes.

We expect QTAIM to be transferrable for the assessment of reactivity, as it was already shown to be predictive in small molecules, and now in metalloenzymes. QM/DMD is one of the most affordable and reliable QM/MM methods, and is essential for sampling. The quality (and appropriateness) of DFT for the mechanistic work is in the hands of the researcher; it is where improvements can be made, and also where the limitations are known and to some degree – controllable. We note that, despite the competitive speed, our suite of methods cannot be used for massive screenings in design. It is too slow. Metalloenzymes is the area where cheap classical force fields would fail. Therefore, our hope is to have methods that are more accurate and yet affordable, for a higher success rate of design, and reduced need for brute-force screenings.

Acknowledgements

This work was supported by the NSF CAREER Award CHE1351968 (ANA), ONR Grant No. N00014-10-1-0838 (AM, MEE), and the NSF Graduate Fellowship #2011115747 (CEV). The authors additionally thank Prof. Joseph Loo and Hong Hanh Nguyen for helpful discussion.

References

1 A. M. Thayer, *Chem. Eng. News*, 2012, 13–18.
2 D. K. Ro, E. Paradise, M. Ouellet, K. J. Fisher, K. Newman, J. M. Ndungu, K. Ho, R. A. Eachus, T. S. Ham, J. Kirby, M. C. Y. Chang, S. T. Withers, Y. Shiba, R. Sarpong and J. D. Keasling, *Nature*, 2006, **440**, 940–943.
3 J. B. Siegel, A. Zanghellini, H. M. Lovick, G. Kiss, A. R. Lambert, J. L. St.Clair, J. L. Gallaher, D. Hilvert, M. H. Gelb, B. L. Stoddard, K. N. Houk, F. E. Michael and D. Baker, *Science*, 2010, **329**, 309–313.
4 L. Jiang, E. A. Althoff, F. R. Clemente, L. Doyle, D. Röthlisberger, A. Zanghellini, J. L. Gallaher, J. L. Betker, F. Tanaka, C. F. Barbas, D. Hilvert, K. N. Houk, B. L. Stoddard and D. Baker, *Science*, 2008, **319**, 1387–1391.
5 R. Lerner, S. Benkovic and P. Schultz, *Science*, 1991, **252**, 659–667.
6 D. Hilvert, *Annu. Rev. Biochem.*, 2000, **69**, 751–793.
7 C. Vasileiou, S. Vaezeslami, R. M. Crist, M. Rabago-Smith, J. H. Geiger and B. Borhan, *J. Am. Chem. Soc.*, 2007, **129**, 6140–6148.
8 D. Rothlisberger, O. Khersonsky, A. M. Wollacott, L. Jiang, J. DeChancie, J. Betker, J. L. Gallaher, E. A. Althoff, A. Zanghellini, O. Dym, S. Albeck, K. N. Houk, D. S. Tawfik and D. Baker, *Nature*, 2008, **453**, 190–195.
9 D. N. Bolon and S. L. Mayo, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 14274–14279.
10 B. I. Dahiyat and S. L. Mayo, *Protein Sci.*, 1996, **5**, 895–903.
11 A. Zanghellini, L. Jiang, A. M. Wollacott, G. Cheng, J. Meiler, E. A. Althoff, D. Röthlisberger and D. Baker, *Protein Sci.*, 2006, **15**, 2785–2794.
12 A. N. Alexandrova and W. L. Jorgensen, *J. Phys. Chem. B*, 2008, **113**, 497–504.
13 S. R. Gordon, E. J. Stanley, S. Wolf, A. Toland, S. J. Wu, D. Hadidi, J. H. Mills, D. Baker, I. S. Pultz and J. B. Siegel, *J. Am. Chem. Soc.*, 2012, **134**, 20513–20520.
14 Y. Yu, C. Cui, X. Liu, I. Petrik, J. Wang and Y. Lu, *J. Am. Chem. Soc.*, 2015, **137**, 11570–11573.
15 Z. Ke, S. Abe, T. Ueno and K. Morokuma, *J. Am. Chem. Soc.*, 2012, **134**, 15418–15429.
16 J. Bos and G. Roelfes, *Curr. Opin. Chem. Biol.*, 2014, **19**, 135–143.
17 T. Heinisch, M. Pellizzoni, M. Dürrenberger, C. E. Tinberg, V. Köhler, J. Klehr, D. Häussinger, D. Baker and T. R. Ward, *J. Am. Chem. Soc.*, 2015, **137**, 10414–10419.
18 J. Esselborn, C. Lambertz, A. Adamska-Venkates, T. Simmons, G. Berggren, J. Noth, J. Siebel, A. Hemschemeier, V. Artero, E. Reijerse, M. Fontecvae, W. Lubitz and T. Happe, *Nat. Chem. Biol.*, 2013, **9**, 607–609.
19 C. Buron, K. Sénéchal-David, R. Ricoux, J. P. Le Caër, V. Guérineau, P. Méjanelle, R. Guillot, C. Herrero, J. P. Mahy and F. Banse, *Chem. – Eur. J.*, 2015, **21**, 2188–12193.
20 M. Edge, C. Forder, J. Hennam, I. Lee, D. Tonge, I. Hardern, J. Fitton, K. Eckersley, S. East, A. Shufflebotham, D. Blakey and A. Slater, *Protein Eng.*, 1998, **11**, 1229–1234.
21 J. Kaplan and W. F. DeGrado, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 11566–11570.
22 S. D. Khare, Y. Kipnis, P. Greisen, R. Takeuchi, Y. Ashani, M. Goldsmith, Y. Song, J. L. Gallaher, I. Silman, H. Leader, K. L. Sussman, B. L. Stoddard, D. S. Tawfik and D. Baker, *Nat. Chem. Biol.*, 2012, **8**, 294–300.
23 P. Greisen and S. D. Khare, *Methods Mol. Biol.*, 2014, **1216**, 265–273.
24 V. Nanda, S. Senn, D. H. Pike, A. Rodriguez-Granillo, W. A. Hansen, S. D. Khare and D. Noy, *Biochim. Biophys. Acta, Bioenerg.*, 2016, **1857**, 531–538.
25 O. Maglio, F. Nastri, R. T. de Rosales, M. Faiella, V. Pavone, W. F. DeGrado and A. Lombardi, *C. R. Chim.*, 2007, **10**, 703–720.
26 M. Faiella, C. Andreozzi, R. T. de Rosales, V. Pavone, O. Maglio, F. Nastri, W. F. DeGrado and A. Lombardi, *Nat. Chem. Biol.*, 2009, **5**, 882–884.
27 A. J. Reig, M. M. Pires, R. A. Snyder, Y. Wu, H. Jo, D. W. Kulp, S. E. Butch, J. R. Calhoun, T. G. Szyperski, E. I. Solomon and W. F. DeGrado, *Nat. Chem.*, 2012, **4**, 900–906.
28 E. Brunk and U. Rothlisberger, *Chem. Rev.*, 2015, **115**, 6217–6263.
29 C. A. Rohl, C. E. M. Strauss, K. M. S. Misura and D. Baker, *Methods Enzymol.*, 2004, **383**, 66–93.
30 R. Das and D. Baker, *Annu. Rev. Biochem.*, 2008, **77**, 363–382.
31 A. Warshel and A. Papazyan, *Curr. Opin. Struct. Biol.*, 1998, **8**, 211–217.
32 S. D. Fried, S. Bagchi and S. G. Boxer, *Science*, 2014, **346**, 1510–1514.

33 P. H. Pétra, *Methods Enzymol.*, 1970, **19**, 460–503.
34 G. R. Reeck, K. A. Walsh and H. Neurath, *Biochemistry*, 1971, **10**, 4690–4698.
35 J. E. Bodwell and W. L. Meyer, *Biochemistry*, 1981, **20**, 2767–2777.
36 B. L. Vallee, A. Galdes, D. S. Auld and J. F. Riordan, in *Zinc Enzymes*, ed. T. G. Spiro, Wiley, New York, 1983.
37 W. N. Lipscomb, *Proc. Natl. Acad. Sci. U. S. A.*, 1980, **77**, 3875–3878.
38 B. L. Vallee and D. S. Auld, *Faraday Discuss.*, 1992, **93**, 47–65.
39 B. L. Vallee and D. S. Auld, *Biochemistry*, 1990, **29**, 5647–5659.
40 A. Kilshtain-Vardi, M. Glick, H. M. Greenblatt, A. Goldblum and G. Shoham, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2003, **59**, 323–333.
41 A. Alex and T. Clark, *J. Comput. Chem.*, 1992, **13**, 704–717.
42 S. Alvarez-Santos, A. González-Lafont, J. M. Lluch, B. Oliva and F. X. Avilés, *Can. J. Chem.*, 1994, **72**, 2077–2083.
43 S. Alvarez-Santos, A. Gonzalez-Lafont, J. M. Lluch, B. Oliva and F. X. Aviles, *New J. Chem.*, 1998, **22**, 319–326.
44 A. Vardi-Kilshtain, G. I. L. Shoham and A. Goldblum, *Mol. Phys.*, 2003, **101**, 2715–2724.
45 J. Suh, T. H. Park and B. K. Hwang, *J. Am. Chem. Soc.*, 1992, **114**, 5141–5146.
46 D. Xu and H. Guo, *J. Am. Chem. Soc.*, 2009, **131**, 9780–9788.
47 S. J. Gardell, C. S. Craik, D. Hilvert, M. S. Urdea and W. J. Rutter, *Nature*, 1985, **317**, 551–555.
48 M. Sparta, D. Shirvanyants, F. Ding, N. V. Dokholyan and A. N. Alexandrova, *Biophys. J.*, 2012, **103**, 767–776.
49 N. V. Dokholyan, S. V. Buldyrev, H. E. Stanley and E. I. Shakhnovich, *Folding Des.*, 1998, **3**, 577–587.
50 N. V. Dokholyan, *Curr. Opin. Struct. Biol.*, 2006, **16**, 79–85.
51 F. Ding, W. Guo, N. V. Dokholyan, E. I. Shakhnovich and J. E. Shea, *J. Mol. Biol.*, 2005, **350**, 1035–1050.
52 F. Ding, D. Tsao, H. Nie and N. V. Dokholyan, *Structure*, 2008, **16**, 1010–1018.
53 M. Sparta and A. N. Alexandrova, *PLoS One*, 2012, **7**, e47172.
54 C. E. Valdez and A. N. Alexandrova, *J. Phys. Chem. B*, 2012, **116**, 10649–10656.
55 M. Sparta, C. E. Valdez and A. N. Alexandrova, *J. Mol. Biol.*, 2013, **425**, 3007–3018.
56 S. D. Khare, F. Ding and N. V. Dokholyan, *J. Mol. Biol.*, 2003, **334**, 515–525.
57 E. A. Proctor, F. Ding and N. V. Dokholyan, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2011, **1**, 80–92.
58 E. A. Proctor and N. V. Dokholyan, *Curr. Opin. Struct. Biol.*, 2016, **37**, 9–13.
59 S. Nedd, R. L. Redler, E. A. Proctor, N. V. Dokholyan and A. N. Alexandrova, *J. Mol. Biol.*, 2014, **426**, 4112–4124.
60 C. E. Valdez, M. Sparta and A. N. Alexandrova, *J. Chem. Theory Comput.*, 2012, **9**, 730–737.
61 R. F. W. Bader, *Atoms in Molecules: A Quantum Theory*, Oxford University Press, Oxford, 1990.
62 H. Kim and W. N. Lipscomb, *Biochemistry*, 1990, **29**, 5546–5555.
63 S. Wu, C. Zhang, D. Xu and H. Guo, *J. Phys. Chem. B*, 2010, **114**, 9259–9267.
64 (*a*) E. Prodan and W. Kohn, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 11635–11638; (*b*) W. Kohn, *Phys. Rev. Lett.*, 1996, **76**, 3168–3171.
65 R. J. Boyd and S. C. Choi, *Chem. Phys. Lett.*, 1986, **129**, 62–65.
66 C. Fonseca Guerra, J. G. Snijders, G. te Velde and E. J. Baerends, *Theor. Chem. Acc.*, 1998, **99**, 391–403.
67 M. E. Eberhart, *Philos. Mag. B*, 2001, **81**, 721–729.
68 T. E. Jones and M. E. Eberhart, *Int. J. Quantum Chem.*, 2010, **110**, 1500–1505.
69 T. E. Jones, M. E. Eberhart, S. Imlay and C. Mackey, *J. Phys. Chem. A*, 2011, **115**, 12582–12585.
70 T. E. Jones, *J. Phys. Chem. A*, 2012, **116**, 4233–4237.
71 G. te Velde, F. M. Bickelhaupt, E. J. Baerends, C. Fonseca Guerra, S. J. A. van Gisbergen, J. G. Snijders and T. Ziegler, *J. Chem. Phys.*, 2001, **22**, 931–967.
72 ADF2014. SCM, Theoretical Chemistry, Vrije Universiteit, Amsterdam, The Netherlands, http://www.scm.com.
73 A. Klamt and G. Schuurmann, *J. Chem. Soc., Perkin Trans. 2*, 1993, 799–805.
74 E. Van Lenthe and E. J. Baerends, *J. Comput. Phys.*, 2003, **24**, 1142–1156.
75 Tecplot. http://www.tecplot.com/ (2014).
76 2011, T. V., a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989–2007, Turbomole GmbH, since 2007, available from http://www.turbomole.com.
77 M. V. A Arnim, *J. Comput. Chem.*, 1998, **19**, 1746–1757.
78 M. Sierka, A. Hogekamp and R. Ahlrichs, *J. Chem. Phys.*, 2003, **118**, 9136–9148.
79 P. A. M. Dirac, *Quantum Mechanics of Many-Electron Systems*, 1929, **123**, 714–733.
80 J. P. Perdew and Y. Wang, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1992, **45**, 13244–13249.
81 J. Tao, J. P. Perdew, V. N. Staroverov and G. E. Scuseria, *Phys. Rev. Lett.*, 2003, **91**, 146401.
82 A. Schäfer, H. Horn and R. Ahlrichs, *J. Chem. Phys.*, 1992, **97**, 2571–2577.
83 F. A. Weigend, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297–3305.
84 (*a*) J. C. Slater, *Phys. Rev.*, 1951, **81**, 385–390; (*b*) S. H. Vosko, L. Wilk and M. Nusair, *Can. J. Phys.*, 1980, **58**, 1200–1211.
85 A. D. Becke, *Phys. Rev. A: At., Mol., Opt. Phys.*, 1988, **38**, 3098–3100.
86 J. P. Perdew, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1986, **33**, 8822–8824.
87 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.
88 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
89 V. N. Staroverov, G. E. Scuseria, J. Tao and J. P. Perdew, *J. Chem. Phys.*, 2003, **119**, 12129–12137.
90 A. E. Reed, R. B. Weinstock and F. Weinhold, *J. Chem. Phys.*, 1985, **83**, 735–746.
91 M. W. Y. Szeto, J. I. Mujika, J. Zurek, A. J. Mulholland and J. N. Harvey, *THEOCHEM*, 2009, **898**, 106–114.

92 A. Guevara-García, E. Echegaray, A. Toro-Labbe, S. Jenkins, S. R. Kirk and P. W. Ayers, *J. Chem. Phys.*, 2011, **134**, 234106.
93 T. E. Jones, M. E. Eberhart, S. Imlay and C. Mackey, *J. Phys. Chem. A*, 2011, **115**, 12582–12585.
94 P. W. Ayers and S. Jenkins, *J. Chem. Phys.*, 2009, **130**, 154104.
95 A. Guevara-García, P. W. Ayers, S. Jenkins, S. R. Kirk, E. Echegaray and A. Toro-Labbe, *Top. Curr. Chem.*, 2014, **351**, 103–124.
96 M. R. Nechay, N. M. Gallup, A. Morgenstern, Q. A. Smith, M. E. Eberhart and A. N. Alexandrova, *J. Phys. Chem. B*, 2016, **120**, 5884–5895.