FEBS Letters

FEBS PRESS science publishing by scientists

RESEARCH ARTICLE

The structure of a *Lactobacillus helveticus* chlorogenic acid esterase and the dynamics of its insertion domain provide insights into substrate binding

Kellie K. Omori, Charles T. Drucker, Tracie L. S. Okumura, Nathaniel B. Carl, Brianna T. Dinn, Destiny Ly, Kylie N. Sacapano, Allie Tajii and Cedric P. Owens

Department of Chemistry and Biochemistry, Schmid College of Science and Technology, Chapman University, Orange, CA, USA

Correspondence
C. P. Owens, Department of Chemistry and Biochemistry, Schmid College of Science and Technology, Chapman University, 1 University Dr., Orange, CA 92866, USA
Tel: +1 714 997 6922
E-mail: cpowens@chapman.edu

(Received 24 April 2023, revised 31 July 2023, accepted 20 August 2023)

doi:10.1002/1873-3468.14731

Edited by Peter Brzezinski

Chlorogenic acid esterases (ChlEs) are a useful class of enzymes that hydrolyze chlorogenic acid (CGA) into caffeic and quinic acids. ChlEs can break down CGA in foods to improve their sensory properties and release caffeic acid in the digestive system to improve the absorption of bioactive compounds. This work presents the structure, molecular dynamics, and biochemical characterization of a ChlE from *Lactobacillus helveticus* (*Lh*). Molecular dynamics simulations suggest that substrate access to the active site of *Lh*ChlE is modulated by two hairpin loops above the active site. Docking simulations and mutational analysis suggest that two residues within the loops, Gln_{145} and Lys_{164}, are important for CGA binding. Lys_{164} provides a slight substrate preference for CGA, whereas Gln_{145} is required for efficient turnover. This work is the first to examine the dynamics of a bacterial ChlE and provides insights on substrate binding preference and turnover in this type of enzyme.

Keywords: chlorogenic acid; chlorogenic acid esterase; ferulic acid esterase; hydrolase; substrate specificity

Chlorogenic acid esterases (also referred to as chlorogenate esterases, ChlEs, EC 3.1.1.42) hydrolyze CGA, a water-soluble ester of caffeic and quinic acids. ChlEs are α/β hydrolases, a superfamily of enzymes that share a common α/β fold but differ considerably in terms of overall structure and substrate range [1,2]. That said, all known ChlEs use a Ser-Asp-His catalytic triad to catalyze ester bond hydrolysis [1,3–5]. ChlEs are found in both bacteria and fungi, however, ChlEs from bacterial and fungal species have low sequence similarity and are structurally dissimilar.

ChlEs can also cleave substrates that are similar to CGA, such as ethyl- and methyl ferulic acid [6,7]. Thus, ChlEs are closely related to ferulic acid esterases (FAEs, EC3.1.1.73) since their substrate ranges overlap. ChlEs are biotechnologically interesting since they have the ability to break down CGA in foods such as coffee and sunflower meal, thereby improving the foods' sensory properties [8–11]. Furthermore, ChlEs have been studied for their ability to mobilize caffeic acid in the gut and to break down CGA in fermented foods, thus improving their antioxidant content [12,13].

Despite their potential biotechnological utility and importance in mediating antioxidant uptake in the human digestive system, critical structural and functional properties of ChlEs are not well understood. Compared to fungal ChlE's, the bacterial enzymes are not well characterized. Although CGA hydrolysis has been demonstrated for several bacterial ChlEs [13,14], only a single ChlE has been both structurally and functionally characterized, *Lj*0536 from *Lactobacillus johnsonii* [4]. *Lj*0536 features a characteristic α/β insertion or lid domain that is located

Abbreviations
CGA, chlorogenic acid; ChlE, chlorogenic acid esterase; ECA, ethyl caffeic acid; EFA, ethyl ferulic acid; FAE, ferulic acid esterase; HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; MD, molecular dynamics; RMSD, root mean square deviation; RMSF, root mean square fluctuation.

above the active site cleft [4]. The insertion domain does not resemble that of fungal ChlEs [5], however, it is homologous to that of Est1E, a ferulic acid esterase from *Butyrivibrio proteoclasticus* [15]. *Lj*0536 is promiscuous toward aromatic esters since it can hydrolyze several hydroxycinnamic acid derivatives. However, it has a slight catalytic preference for CGA over the smaller substrate, ethyl ferulic acid, since the specificity constant, k_{cat}/K_m, for CGA hydrolysis is 1.4-fold higher for CGA [7]. Mutational analysis of the *Lj*0536 insertion domain showed that it is required for enzyme function. Deletion of the domain produces an inactive enzyme, and point mutations in the insertion domain around the active site that abolish hydrogen bonding between the enzyme and substrate are detrimental to both catalysis and binding [4].

Our group recently characterized a putative ChlE from *Lactobacillus helveticus* (*Lh*ChlE). *Lh*ChlE is very active against CGA and was used to break down CGA in sunflower flour to prevent sunflower flour cookie discoloration [9,10]. However, *Lh*ChlE's substrate preference was never unambiguously determined [16,17]. Moreover, its structure and the conformational dynamics of its insertion domain have not yet been studied.

The goal of this paper is to structurally and biochemically characterize *Lh*ChlE. To this end, we solved the structure of *Lh*ChlE and performed molecular dynamics simulations to study how conformational changes in the insertion domain influence substrate binding. Our results suggest that the movement of two insertion domain loops modulates the size of the entrance to the active site pocket and is thus important for substrate binding. Furthermore, docking simulations suggest that two residues on these loops, Lys_{164} and Gln_{145}, may stabilize CGA binding. Enzyme kinetics data demonstrate that *Lh*ChlE is not specific in its substrate range; however, it has a slight preference for CGA, like *Lj*0536. Mutational analysis indicates that Lys_{164} is important for imparting a slight CGA binding preference. In contrast, Gln_{145} does not contribute to establishing a preference for CGA. Instead, it is important for binding and catalysis for all substrates, likely because Gln_{145} is involved in conformational changes that are required during turnover for all substrates.

Materials and methods

Materials

Materials were purchased from Sigma-Aldrich (St. Louis, MO, USA), Thermo Fisher Scientific (Hampton, NH, USA), and VWR (Radnor, PA, USA). All reagents were ACS grade or equivalent. Primers were purchased from Integrated DNA Technologies (IDT, Coralville, IA, USA).

Cloning and mutagenesis

Wild-type *Lh*ChlE (Genbank ID: AHI12574.1) from *L. helveticus* H9 was used as a template to generate the K164A and Q145A mutants. Each mutant was made by site-directed mutagenesis using the Q5 Site-Directed Mutagenesis Kit from NEB (Ipswich, MA, USA) using the following primers:

K164A Forward: 5′ GGTGGGCAACGCGC TGGGCATGAAAGTGGG 3′
K164A Reverse: 5′ AGCGGCACCACATCCGGA 3′
Q145A Forward: 5′ CGGCAACACCGCGGGCG CGACCT 3′
Q145A Reverse: 5′ CGCAGCGCATCATCTTTC 3′

pET28a plasmids containing the esterase gene were cloned into BL21(DE3) cells, grown overnight in LB broth containing 50 $\mu g \cdot mL^{-1}$ kanamycin. The plasmid was then purified using a miniprep kit (Thermo Fisher Scientific). Plasmids were sent for sequencing (Genewiz, Cambridge, MA, USA) to confirm the presence of the mutation.

Protein expression, purification, and physical characterization

Protein expression and purification for WT, Q145A, and K164A *Lh*ChlE were carried out as described in Lo Verde *et al*. [16]. The gel filtration, circular dichroism, and thermal stability measurements were also carried out as described in Lo Verde *et al*. [16].

Protein crystallization

Crystallization of WT and K164A *Lh*ChlE were both performed using the hanging drop vapor diffusion method with a well volume of 1 mL and a drop size of 2 µL. The drop contained 1 µL of protein and 1 µL of well solution. For WT *Lh*ChlE, the well solution contained 0.1 M citric acid, pH 5.0, 10% PEG 6000, and 1.6 M lithium chloride. The protein concentration was 20 $mg \cdot mL^{-1}$. For K164A *Lh*ChlE, the well solution contained 0.1 M citric acid pH 5.0, 7.5% PEG 6000, 0.01 M zinc chloride, and 1.5 M lithium chloride. The protein concentration was 19 $mg \cdot mL^{-1}$. For both proteins, crystals formed after approximately 2 weeks, and crystals were harvested and cryo-protected in a solution containing equal volume well solution and 50% glycerol, and flash frozen in liquid nitrogen.

Data collection and structure determination

WT *Lh*ChlE was diffracted at ALS beamline 12-2 and K164A *Lh*ChlE at ALS beamline 5-02 using a wavelength of 1 Å. Indexing, scaling, merging, and truncation were done using the DIALS pipeline [18] in XIA2. For WT *Lh*ChlE, two datasets were merged prior to molecular replacement. The structure of WT *Lh*ChlE was solved by molecular

replacement with 3PF8 as the search model. WT *Lh*ChlE was used as the search model for K164A *Lh*ChlE. After molecular replacement, an initial structure was generated using PHENIX AUTOBUILD [19]. Structural refinement was conducted iteratively using PHENIX REFINE and COOT [20]. All protein structures were visualized in PYMOL (Schrödinger, Cambridge, MA, USA) [21].

Molecular dynamics methods

Molecular dynamics simulations of the enzymes were performed using GROMACS v. 2022.3 [22,23] using the CHARMM36 all-atom protein force field [24]. Proteins were situated 1 nm from the edges of a dodecahedral box and were solvated with TIP3P water molecules [25], yielding > 11 000 solvent molecules for each simulated protein. Sodium ions were added to neutralize the system. Initial energy minimization was performed using a steepest descent algorithm with a maximum of 10 000 steps. All simulations were performed using 2.0 fs time steps and the LINCS algorithm for constraining H-atom bonds [26]. Temperature coupling was implemented at 300 K using the modified Berendsen thermostat with stochastic velocity rescaling [27], with protein and nonprotein coupling groups using a time constant of 0.1 ps. Long-range electrostatics were calculated using the particle mesh Ewald method [28], and a long-range cutoff radius of 12 Å. The NVT and NPT simulations were run for 200 ps each using positional constraints on the heavy atoms. The NPT and production simulations were run at a pressure of 1.0 bar using the Parinello-Rahman barostat [29]. Production simulations lasting 200 ns were prepared using the NPT ensembles without positional constraints. Each protein was analyzed in triplicate.

Protein flexibility was analyzed using root-mean-square deviation (RMSD) calculations of the protein backbone, selected residues, and the insertion domain compared to the initial protein conformation, along with root-mean-square fluctuation (RMSF) calculations of Cα atoms for each residue. Clustering analyses across triplicate simulations were performed using GROMACS's *cluster* module using the clustering algorithm described by Daura *et al.* [30], with an RMSD cutoff of 1.25 Å. Visualizations of simulations were prepared using PYMOL. Data visualizations were prepared using R v. 4.1.2 [31].

Docking simulations

All docking models were created using MOLSOFT ICM-PRO [32]. Default settings for ligand docking were used and thoroughness/effort was 3 for all complexes. The models with the lowest (best) ICM scores were selected for further analysis.

Kinetic characterization of *Lh*ChlE

The kinetic properties of WT, K164A, and Q145A *Lh*ChlE were determined using Michaelis–Menten assays. Hydrolysis of CGA, ethyl caffeic acid, and ethyl ferulic acid were all carried out at 21 °C using an Cary 60 UV-Vis spectrophotometer (Agilent Technologies, Santa Clara, CA, USA) in a buffered solution of 50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), pH 8. A stock solution of CGA was prepared by dissolving CGA in 50 mM HEPES pH 8.0. Stock solutions of ethyl caffeic acid and ethyl ferulic acid were made by dissolving each substrate in 50 mM HEPES pH 8.0, 5% ethanol. The concentration of each solution was calculated using the Lambert–Beer law using their respective extinction coefficients (Fig. S1). For all experiments, the enzyme concentration was 10 nM and the reaction volume was 0.8 mL. All reactions were monitored for 1 min at 0.05-min intervals from 250 to 400 nm. When the substrate concentrations were under 0.10 mM, the pathlength was 1 cm. For concentrations above 0.10 mM, the pathlength was 0.2 cm. The equations used for calculating product formation from the difference in molar absorptivity are listed in the Supporting Information. All kinetics data represents the average of at least five true independent replicates with at least three different batches of enzyme. Enzymatic velocities were calculated using a linear fit of the initial rates, and enzyme kinetics parameters were calculated by nonlinear curve fitting. All kinetic data analysis was conducted in PRISM 9 (Graphpad, Boston, MA, USA).

Results

Structure of WT *Lh*ChlE

The structure of WT *Lh*ChlE was solved by single-crystal X-ray diffraction to a resolution of 2.2 Å with R_{free}/R_{work} values of 0.21 and 0.18, respectively (Fig. 1A,B). The crystallography statistics are summarized in Table S1. The asymmetric unit of *Lh*ChlE contains three chains, with two chains forming a dimer (Fig. 1A). The third chain forms a dimer with its counterpart in a neighboring asymmetric unit. We previously established that the protein is dimeric in solution [16]. The *Lh*ChlE dimer interface was determined using the ccp4i program Areaimol [33]. It is large, approximately 2910 $Å^2$, as expected for a dimer. The overall structure of *Lh*ChlE is typical of the α/β hydrolase fold, containing at its core eight β-stands connected by α-helices on either side of the β-stands. The active site catalytic triad consists of Ser_{106}, His_{229}, and Asp_{201}, which sit toward the base of an elongated active site cleft. The active site volume was determined to be 283 $Å^3$ using MOLSOFT ICM-PRO [32].

The α/β insertion domain starts at Ala_{133} and ends at Leu_{180}. The domain has an α-ββ-ββ-α arrangement. The domain forms one face of the active site cleft (Fig. 1B) and also creates a flap at the top of the

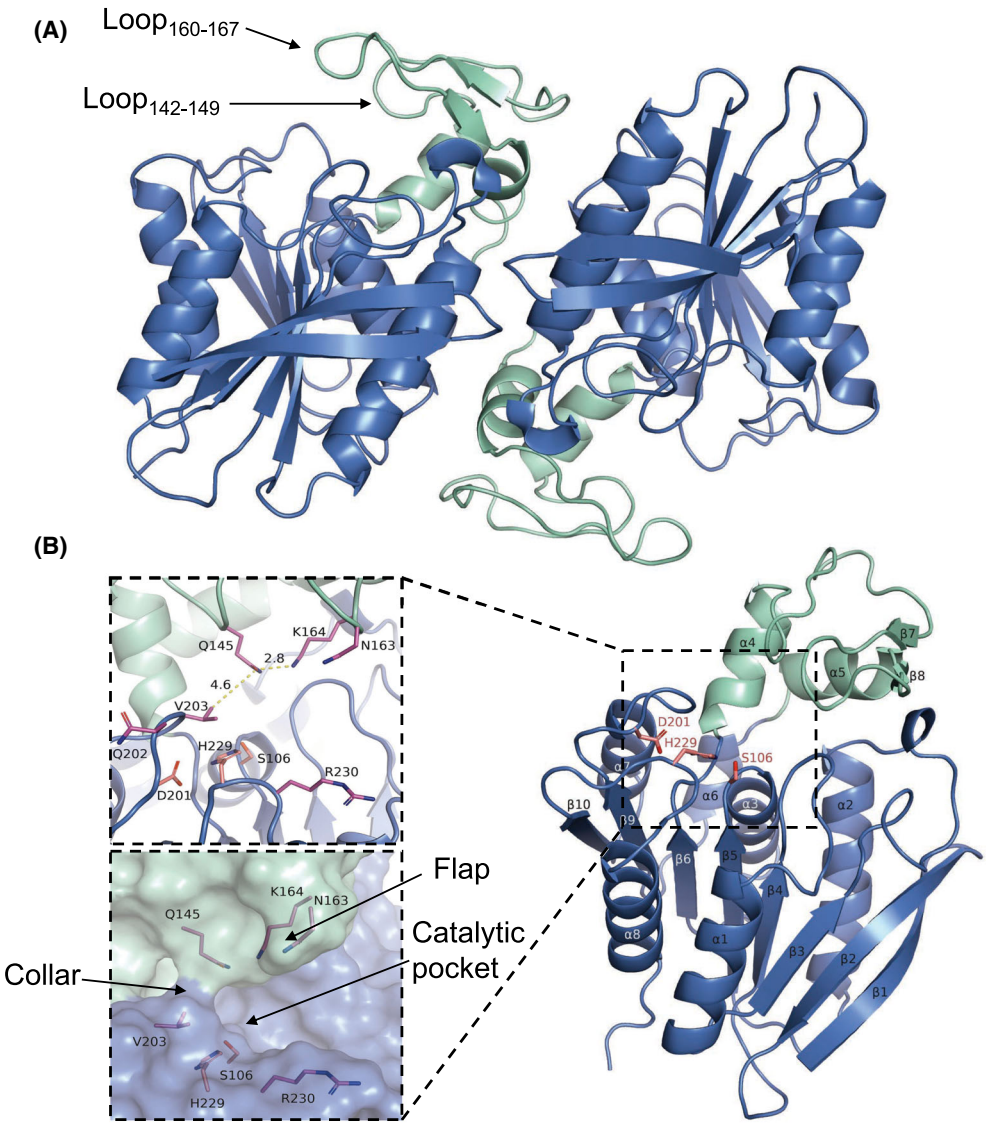

Fig. 1. (A) Structure of the *Lh*ChlE dimer. The insertion domain is highlighted in teal. Two important active site loops are also highlighted. (B) Structure of a protomer. Secondary structure elements are labeled on the figure. The insets offer close-up views of the active site as cartoon and surface representations. Active site residues are displayed in both inset. The surface representation highlights structural features that likely restrict substrate binding to the catalytic pocket. The views between the two insets are identical.

active site through an extended loop ($Loop_{160–167}$) that contains the bulky residues Asn_{163} and Lys_{164} at its tip. Compared to the rest of the protein, the insertion domain exhibits greater mobility, as suggested by its high B-factors (Fig. 2A).

Structural comparison with related bacterial chlorogenic acid and ferulic acid esterases

The structure of *Lh*ChlE is very similar to *Lj*0536 with an RMSD of 0.58 Å over the aligned Cα (Fig. 3A) based on a pair-wise alignment by TM-align [34]. The active site residues are highly conserved (Fig. 3B), as are the active site pocket volumes (283 $Å^3$ for *Lh*ChlE and 291 $Å^3$ for *Lj*0536). The largest structural difference between *Lh*ChlE and *Lj*0536 is $Loop_{160–167}$, as shown in Fig. 3A. This loop protrudes about 6 Å further over the active site in *Lh*ChlE compared to *Lj*0536. Like *Lh*ChlE, *Lj*0536 also has a lysine residue in its corresponding loop above the active site, however, due to the loop's smaller size, the lysine residue sits outside of the active site cleft. *Lh*ChlE is also very similar to two recently described ferulic acid esterases from *Lactobacillus acidophilus* (*La*FAE) [35],

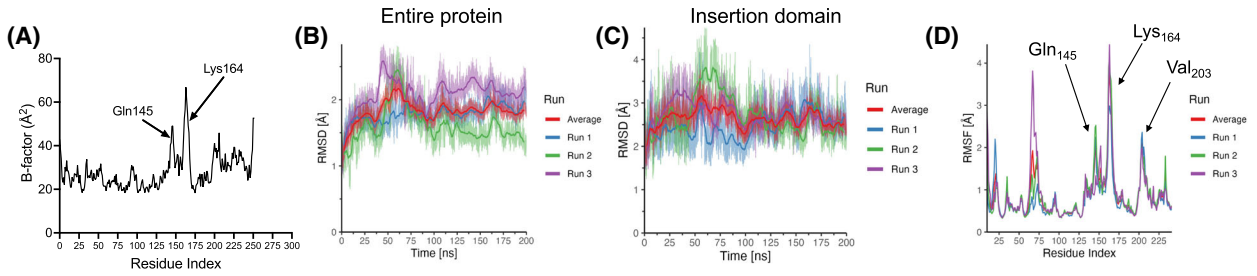

Fig. 2. (A) Plot of the B-factors by residue. The B-factor of the Cα of each residue was averaged over the three copies in the asymmetric unit. (B) RMSD per residue for triplicate MD trajectories for the entire protein and (C) the insertion domain (residues Ala_{133} to Leu_{180}), indicating the RMSD values are higher in the insertion domain. (D) RMSF values per residue for triplicate MD trajectories.

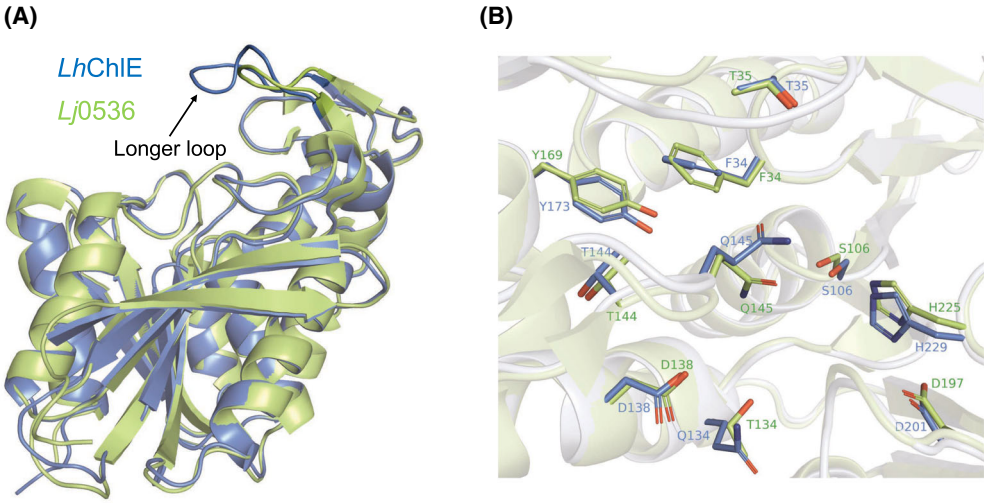

Fig. 3. Structural comparison of (A) protomers of *Lh*ChlE (blue) and *Lj*0536 (PDB ID: 3PF8, green). (B) Close up of the active sites of *Lh*ChlE (blue) and *Lj*0536 (green), highlighting the close structural similarity. The catalytic triad is composed of Ser_{106}, His_{229}, and Asp_{201} (*Lh*ChlE numbering).

Lentilactobacillus buchneri (*Lb*FAE) [36], and Est1E [15]. The RMSD values are 0.50, 1.77, and 1.36 Å, respectively, over the aligned Cα based on pairwise alignment with TM-align [34] (Fig. S2). The insertion domains in *Lh*ChlE, *La*FAE, and Est1E are nearly superimposable, sharing the same topology. The insertion domain in *Lb*FAE has a small α-helix connecting the two beta hairpins (α-ββ-α-ββ-α topology) but is otherwise identical to *Lh*ChlE. The next closest structural relative of *Lh*ChlE based on a DALI server search [37] is a human monoglyceride lipase (PDB: 3JW8), which features a similar core but has a different insertion domain topology.

Molecular dynamics simulations of *Lh*ChlE suggest that the insertion domain is mobile

The B-factors suggest that the insertion domain is flexible (Fig. 2A). Fluctuations in *Lh*ChlE were further examined using 200 ns molecular dynamics (MD) simulations with an explicit solvent model. Simulations were run on a single *Lh*ChlE protomer after confirming that MD trajectories for a protomer reflect those of the dimer (Fig. 2, Figs S3 and S4). Simulations converged, as evidenced by the stable radius of gyration (Fig. S4) and Root Mean Square Deviation (RMSD) values for the protein at the end of the simulation (Fig. 2B, Fig. S4).

Consistent with B-factor analysis, MD simulations suggest that the insertion domain contains the most flexible regions of the protein. The RMSD compared to the initial structure is higher by about 1 Å for the insertion domain than for the core of the protein, as shown in Fig. 2B,C. Furthermore, analysis of the root mean square fluctuation (RMSF), a measure that describes conformational variability during the trajectory, demonstrates that the most mobile part of the protein is the expended hairpin $Loop_{160–167}$ (Fig. 2D).

A second insertion domain loop between residues 142 and 149 ($Loop_{142–149}$) (Fig. 1A) that includes Gln_{145} at its tip also displays large RMSF values. Regions outside of the insertion domain that undergo large fluctuations are located around residues Glu_{18}, Phe_{72}, and Val_{203}. Val_{203} is part of the substrate binding cleft and sits across from Gln_{145}. Phe_{72} is located on a surface-exposed loop adjacent to the insertion domain, but it is not part of the active site, and Glu_{18} is located on a surface-exposed loop that is distant from both the insertion domain and the active site.

Analysis of insertion domain conformational dynamics

The flexibility of insertion or lid domains plays an important role in substrate binding and catalysis for many α/β hydrolases [1]. For example, conformational changes in the insertion domain are required for substrate recognition and binding in tannases [38,39], epoxide hydrolases [40], and lipases [41,42]. The insertion domain in *Lh*ChlE is similar to that of *Lj*0536 but does not resemble that of any α/β hydrolases with known conformation dynamics. Moreover, the structures of apo- and ferulic acid-bound *Lj*0536 are nearly identical around the active site and thus do not provide information on the conformational flexibility of the domain [4].

In the crystal structure of *Lh*ChlE, the insertion domain is likely in a closed conformation (Fig. 1B), suggesting *Lh*ChlE must undergo conformational changes to enable substrate binding. Entrance to the active site is restricted by two structural features. The first represents the narrowest section of the active site cleft and is located directly above the catalytic serine. It is formed by Gln_{145} and Val_{203}, whose sidechains are 4.6 Å apart in the crystal structure (Fig. 1B). Together, Gln_{145} and Val_{203} form a collar that likely must widen to allow the caffeic acid moiety of CGA to enter the catalytic pocket (Fig. 1B). The second restriction is at the mouth or entrance of the active site cleft and is between 7 and 9 Å wide. The bulky sidechains of Asn_{163} and Lys_{164} form a flap that sits above the active site. The closest residues to the flap on the opposite side of the cleft are Val_{203} and Arg_{230} (Fig. 1B). Furthermore, the flap is in close proximity to $Loop_{142–149}$ since Gln_{145} and Lys_{164} form a 2.8 Å hydrogen bond with each other (Fig. 1B). This leads to the hypothesis that the hydrogen bond between Gln_{145} and Lys_{164} may play a role in keeping the flap in a closed conformation by holding the $Loop_{160–167}$ over the active site.

To monitor structural changes that may influence substrate binding, we measured the distance between key amino acid pairs that gate the active site over the course of the molecular dynamics trajectory. The Cα distance between Gln_{145} and Val_{203} increases from 10 to 13 Å and their sidechain distances increase from 8 to 11 Å over the course of the trajectory (Fig. 4A, Fig. S5), causing a widening of the active site cleft. Furthermore, the trajectories suggest Gln_{145} and Lys_{164} drastically move apart after 50 ns (Fig. 4B) and that this movement is irreversible in the timescale of the simulation. The Cα distance between Gln_{145} and Lys_{164} increases from about 9 to 13 Å. More noticeably, the sidechain carbonyl and amine distance increases from ca. 6 to 15 Å. These conformational changes are illustrated in Fig. 4C, that shows *Lh*ChlE at the beginning and end of the trajectory.

Distances between additional residues at the active site entrance were also measured. Lys_{164} and Arg_{230} (located opposite Lys_{164}) remain separated by over 10 Å throughout the trajectory (Fig. S5). The distance between Asn_{163} and Arg_{230} fluctuates and the residue pair likely hydrogen bond reversibly during parts of the trajectory (Fig. S5), however, the residues do not appear to move closer together or further away, overall. Finally, we tracked the distance between Gln_{145} and Gln_{202} (the residue next to Val_{203}). These two residues move closer together during parts of the trajectory, coming within 9 Å of each other (Fig. S5). This suggests that at different times of the trajectory, either Val_{203} or Gln_{202} may represent the closest amino acid to Gln_{145}.

Taken together the MD simulations suggest that the insertion domain is flexible, likely to enable substrate entry to the active site. $Loop_{142–149}$ and $Loop_{160–167}$ are especially important in modulating the active site size since they experience large fluctuations and may participate in a gradual opening of the active site during the trajectory by moving apart.

*Lh*ChlE mutant generation and K164A *Lh*ChlE crystal structure

K164A and Q145A mutants were generated to study their role in *Lh*ChlE activity. Protein expression and purification for the mutants was identical to the WT (Fig. S6). Their physical properties (oligomeric state, secondary structure, and melting point) are very similar to the wild-type enzyme (Figs S6 and S7) as well. The structure of K164A *Lh*ChlE was solved to 2.1 Å and R_{free}/R_{work} values of 0.20 and 0.18, respectively (Table S1). It aligns with the wild-type *Lh*ChlE with an RMSD of 0.165 Å per protomer (Fig. 5). The electron density around Ala_{164} was well defined in the K164A mutant (Fig. S8). As expected, there are no significant differences in the sidechain orientation of

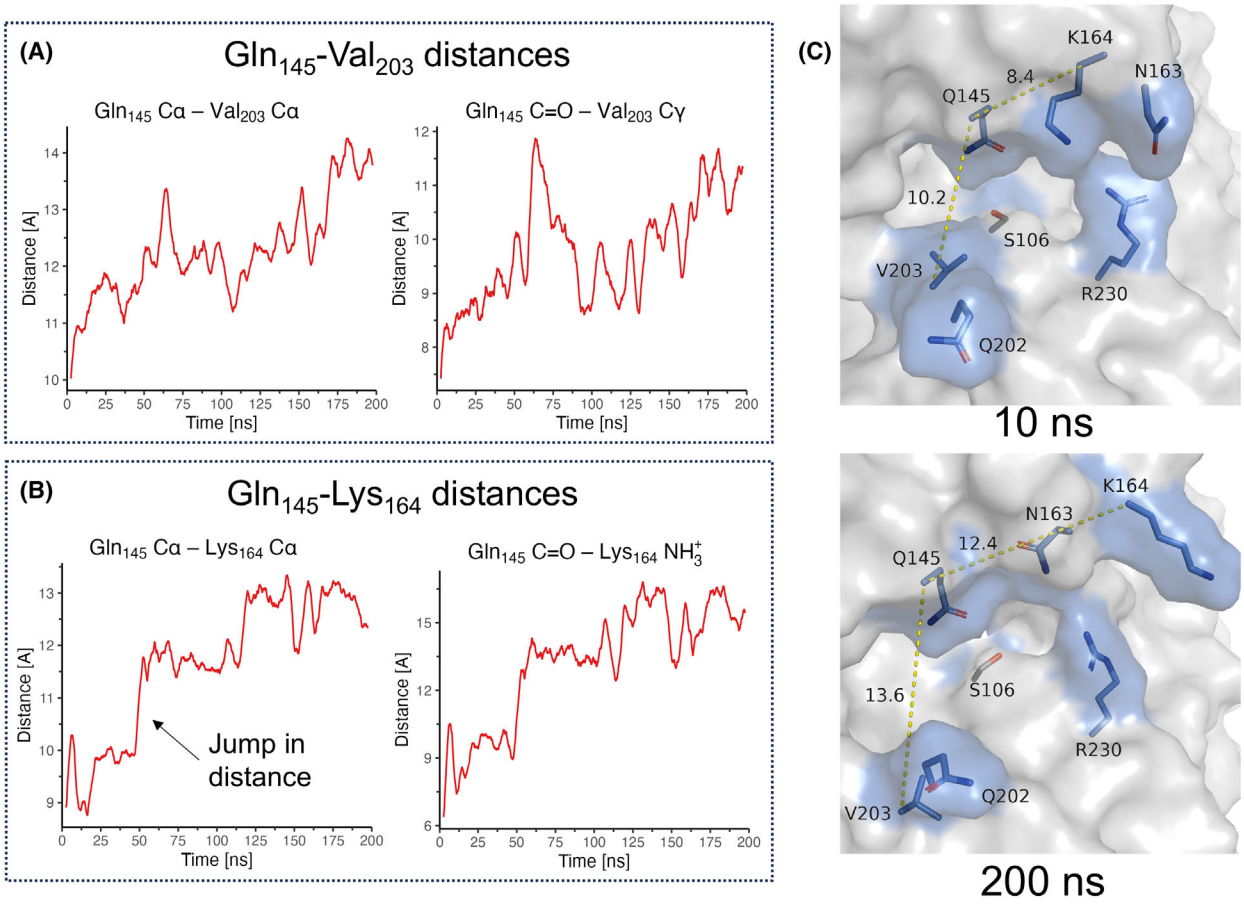

Fig. 4. Distances between Gln_{145} and (A) Val_{203} and (B) Lys_{164} (rolling 5 ns averages). The red line represents the average of three triplicate MD runs. The individual runs are displayed in Fig. S5. The trajectories show that Gln_{145} and Val_{203} move apart by about 3 Å (Cα and sidechain). More drastically, the sidechains of Gln_{145} and Lys_{164} move apart by 9 Å. (C) Snapshots of *Lh*ChlE at the start and end of the trajectory, illustrating the increase in access to the catalytic pocket. The distances that are marked are between the Cα. The catalytic Ser_{106} is shown for reference. Key residues that are discussed in the text are highlighted in blue.

loop and active site residues between the K164A- and WT *Lh*ChlE structures except for a single residue, Gln_{145}. Since this residue no longer can hydrogen bond with Lys_{164}, it moves horizontally approximately 2 Å compared to the wild-type structure (Fig. 5). The structure of Q145A *Lh*ChlE was determined by Robetta [43] using template-based modeling with the wild-type structure as the template. The model of the Q145A mutant and WT are nearly identical with an RMSD of 0.2 Å. No notable structural changes were detected around the active site, including in the loops containing residues 145 and 164 (Fig. 5).

Analysis of Q145A and K164A *Lh*ChlE molecular dynamics trajectories

Molecular dynamics simulations were run on Q145A and K164A *Lh*ChlE to determine how the mutations influence the conformational dynamics of the enzyme. As shown in Fig. 6A,B and Fig. S9, the overall RMSD values for both the entire protein and the insertion domain are much higher for Q145A *Lh*ChlE than the wild type, suggesting a greater departure from the original structure during the course of the trajectory. In contrast, K164A *Lh*ChlE displays only slightly higher RMSD values (Fig. 6A,B, Fig. S9). Similarly, fluctuations for Q145A *Lh*ChlE are higher than for the wild type in the insertion domain loop regions and around residue 145 and around Val_{203}, whereas the RMSF plot for K164A is similar to that of the wild type (Fig. 6C, Figs S10 and S11). We note that the variability in the RMSF and RMSD values for the insertion domain is large between the three replicate MD runs for the Q145A mutant (Figs S9 and S10). Such variability can be expected when predicting the motions of regions in a protein that are very flexible

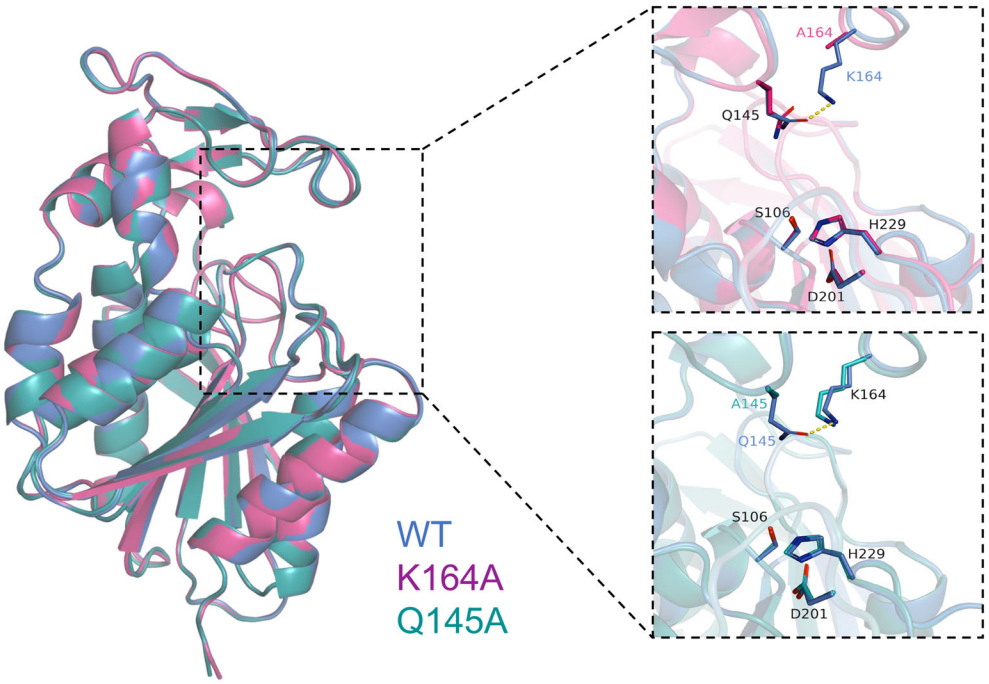

Fig. 5. Crystal structure of the K164A mutant and a model of Q145A LhChlE, suggesting that both proteins align very closely to the wild type. The top inset compares WT with K164A and the bottom inset compares WT with Q145A *Lh*ChlE.

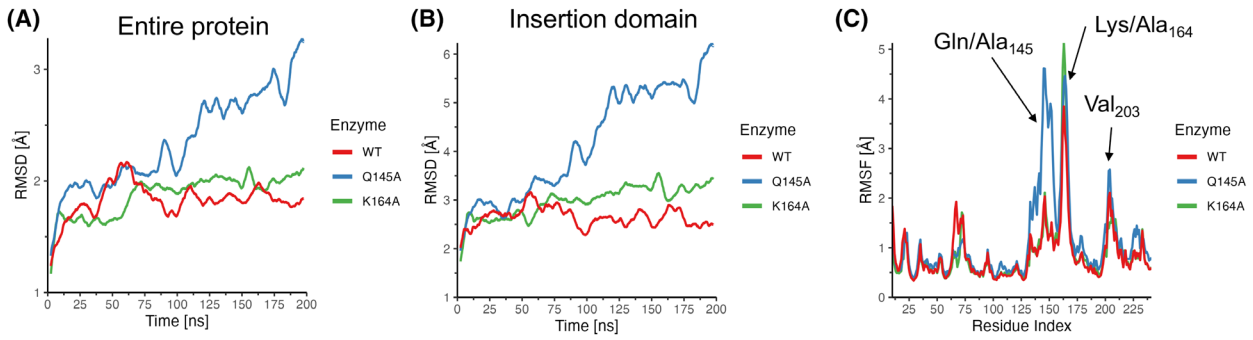

Fig. 6. Rolling average (5 ns averaging) RMSD values for WT compared to Q145A and K164A *Lh*ChlE for (A) the entire protein and (B) the insertion domain. (C) RMSF for the wild type and the two mutants. The individual replicates are found in Figs S9–S11.

[44]. The trends between Q145A and WT, however, hold true for all three Q145A runs. Taken together, the RMSD and RMSF plots suggest that the Q145A mutation causes the insertion domain of *Lh*ChlE to become more mobile.

Next, clustering analysis was performed to compare the major conformations that were sampled during the trajectories of wild type, Q145A, and K164A *Lh*ChlE, as shown in Fig. 7. Clustering was run with a binning RMSD of 1.25 Å. Only structures that represent at least 2.5% of the total population were analyzed (Fig. 7B, Fig. S12). Clustering analysis suggests that the loop mutations alter the distribution of active site conformations (Fig. 7A). Visual comparisons of the three most common conformations suggest that the Q145A mutation leads to a more open active site than WT. To compare open and closed conformations, we measured the dimensions of the active site at its narrowest points (Fig. 8) and also determined the surface accessible area of the catalytic triad as a proxy for accessibility using the GetArea server [45]. We chose this metric since comparing the size of the entire active site using either volume or total solvent accessible area was not descriptive. The pocket of the Q145A

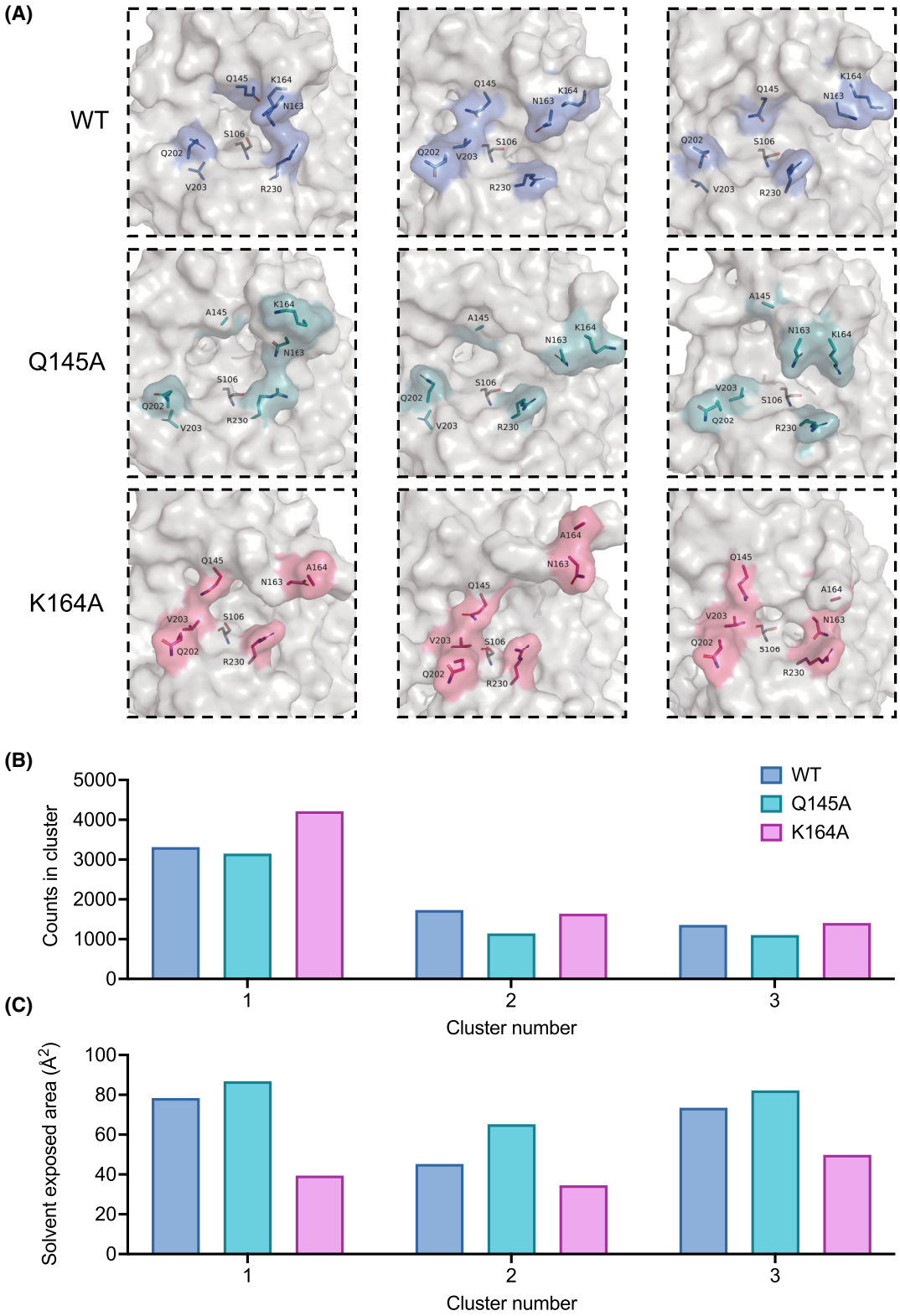

Fig. 7. (A) The top three conformations detected during the trajectory. Key residues lining the active site are highlighted, as is the catalytic serine. (B) Histogram of the three largest clusters. (C) Exposed surface area of the catalytic triad in the three largest clusters, suggesting that the triad is the most solvent exposed in the Q145A mutant.

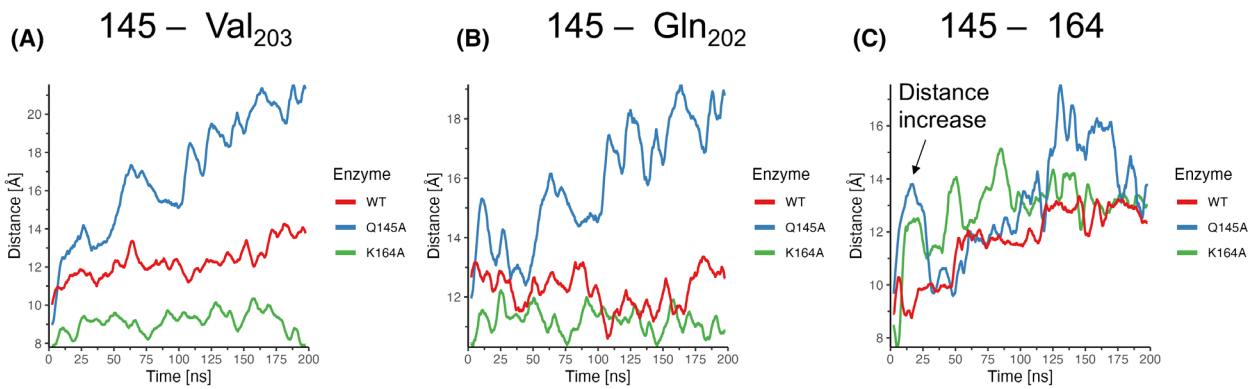

Fig. 8. Distances between key residue pairs for Q145A and K164A *Lh*ChlE. (A) Distance between Ala/Gln_{145} and Val_{203}, (B) Ala/Gln_{145} and Gln_{202} and (C) Ala/Gln_{145} and Ala/Lys_{164}. The solid lines represent 5 ns rolling average of triplicate MD runs. All distances are measured between Cα.

mutant becomes so wide that it is recognized as multiple binding cavities when using common pocket detection programs such as Molsoft or Caver. Compared to the wild type, the surface accessible area of the active site residues is higher for the Q145A mutant and smaller for the K164A mutant (Fig. 7C) for the top three clusters. This suggests that the active site for Q145A is more open than for either WT and K164A *Lh*ChlE.

Distance measurements between key residues at the entrance to the active site provide further evidence for an increase in flexibility around the active site of the Q145A mutant. Average distances are shown in Fig. 8 and the distances for individual replicates and between sidechains are presented in Fig. S13. The Cα distance between Ala_{145} and Val_{203} in the Q145A mutant increases to approximately 20 Å during the trajectory (Fig. 8A). Similarly, the Ala_{145} to Gln_{202} distance is much larger for Q145A than it is for wild-type *Lh*ChlE (Fig. 8B). These measurements suggest that the Q145A mutation causes the collar around the active site to open. This is consistent with the clustering results. As shown in Fig. 7, the collar is open in all top three clusters for Q145A but only in two for WT. The Q145A mutation also impacts the dynamics of residue Lys_{164}. In Q145A *Lh*ChlE, Lys_{164} rotates away from Gln_{145} during the first 20 ns of the trajectory, which is about 30 ns sooner than for the wild-type enzyme (Fig. 8C). Furthermore, Lys_{164} fluctuates slightly more in the Q145A mutant, compared to the wild type, as indicated by the higher RMSF values (Fig. 6C).

The K164A mutation leads to a rapid increase in the distance between residues Gln_{145} and Ala_{164} (Fig. 8C). Ala_{164} separates from Gln_{145} within the first 15 ns of the trajectory in the K164A mutant compared to 50 ns for the wild type. The K164A mutation also causes an overall increase in RMSF in $Loop_{160–167}$ (Fig. 6C). Interestingly, the K164A mutation impacts the Gln_{145}-Val_{203} and Gln_{145}-Gln_{202} distances. As seen in the clustering results, the collar remains closed in the top three conformations of K164A *Lh*ChlE (Fig. 7). This is reflected in the Gln_{145}-Val_{203} and Gln_{145}-Gln_{202} distances, which are shorter than for WT throughout most of the trajectory (Fig. 8, Fig. S13).

In summary, the MD trajectories of the mutants suggest that both $Loop_{160–167}$ and $Loop_{142–149}$ are critical in enabling substrate access to the active site and that Gln_{145} and Lys_{164} have important but different roles in modulating active site dynamics. Mutation of Gln_{145} causes an overall gain in flexibility of the active site, a widening of both the catalytic binding pocket (the collar), and an opening of the flap/mouth of the pocket. However, mutation of Lys_{164} has varying effects on different parts of the active site. While the mutation causes the flap above the active site to open, access to the catalytic residues remains more restricted since the collar remains closed.

Docking CGA to *Lh*ChlE suggests Lys_{164} and Gln_{145} stabilize CGA binding

To date, there are no experimental structures of CGA bound to a *Lh*ChlE; however, a *Lj*0536 structure with ethyl ferulic acid has been reported [4]. Our attempts to obtain substrate-bound structures by soaking *Lh*ChlE crystals with CGA yielded only apo-protein. Thus, we turned to docking studies using MOLSOFT ICM-PRO [32] to understand how substrates bind to *Lh*ChlE. To start, we tested the accuracy of our docking process by generating ethyl ferulic acid bound to *Lj*0536, which recapitulated the crystal structure [4] (Fig. S14). CGA docked deeply into the *Lh*ChlE active site cleft, with Gln_{145} and Val_{203} forming a collar

around the caffeic acid moiety (Fig. 9A,B). The caffeic acid moiety of CGA binds to the active site in a similar fashion as ferulic acid binds to *Lj*0536 in the crystal structure, and the ester is oriented so that it can be attacked by the hydroxide of the active site serine (Ser_{106}) (Fig. 9A). The docking model suggests that numerous hydrogen bonds stabilize CGA binding. Table 1 shows a list of key interactions proposed to be involved in CGA binding. Two noteworthy predicted hydrogen bonds between the protein and the caffeic acid moiety of CGA are from the sidechains of Asp_{138} and Tyr_{173} to the hydroxyl groups of the catechol ring. In addition, Phe_{34} may form a 3.9 Å π-stacking interaction with the catechol ring. The docking model suggests the quinic acid moiety is stabilized through hydrogen bonds with Thr_{35}, Gln_{145}, Lys_{164}, a possible hydrogen bond with Arg_{230}, and the main chain from two residues (Phe_{34} and Ala_{36}). The docking models of the smaller substrates EFA and ECA are shown in Fig. 9C,D. The models indicate that both EFA and ECA bind to *Lh*ChlE in the same general orientation as CGA; however, residues Thr_{35}, Lys_{164}, and Gln_{145} do not seem to interact with the smaller substrates.

Docking CGA to K164A *Lh*ChlE indicates that CGA binds to the mutant in an overall similar conformation as WT *Lh*ChlE (Fig. S15A). Furthermore, ECA and EFA also bind in a similar orientation (Fig. S16A). The most notable difference is the absence of the hydrogen bond between CGA and residue 164. In contrast, CGA docks to Q145A *Lh*ChlE in a different conformation (Fig. S15B). Compared to both WT and K164A *Lh*ChlE, CGA is docked into the active site less deeply and the hydrogen bonding network is different. The ester bond is also much further away from the catalytic triad. The different docking orientation is likely due to the opening of the collar around the ester in the mutant, which makes the active site less restricted. Similarly, ECA and EFA bind closer to the entrance to the active site (Fig. S16B). This is predicted to lead to the formation

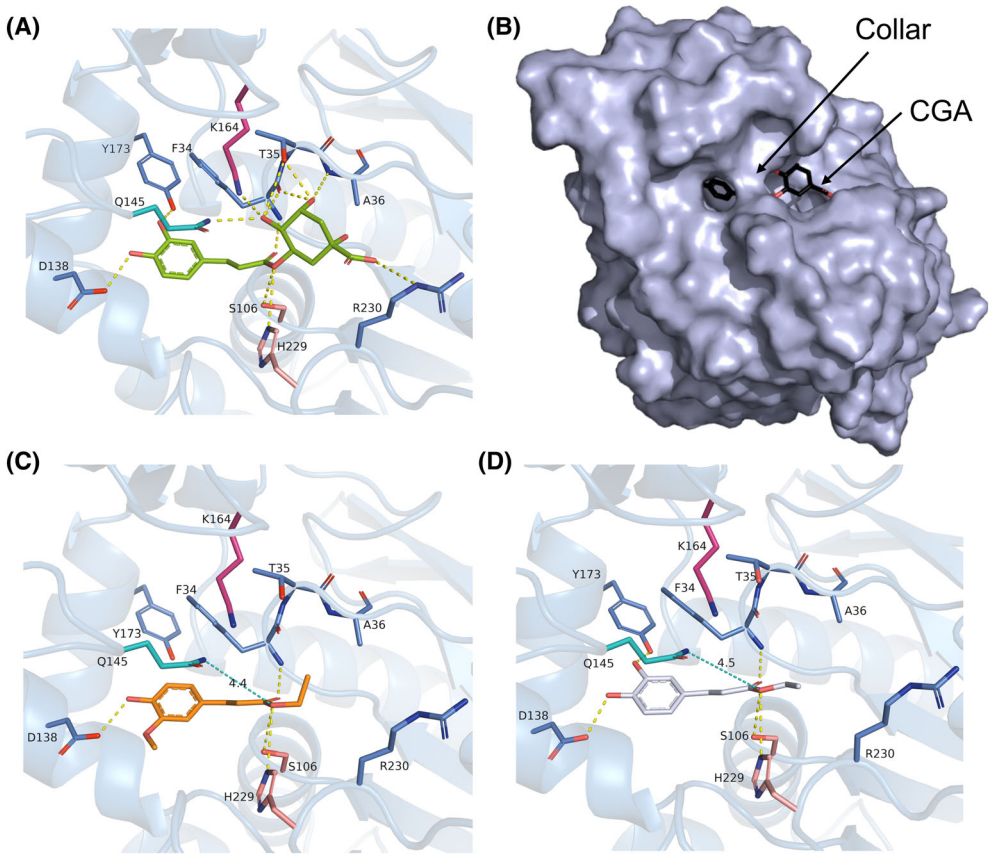

Fig. 9. Docking models for (A, B) CGA (C) Ethyl ferulic acid (EFA) and (D) Ethyl caffeic acid (ECA) binding to *Lh*ChlE. Key residues are shown as sticks and important H-bonds within 4 Å of the substrate are shown as dashed lines. The dotted line (teal) between Gln_{145} and the esters of ECA and EFA illustrates the residue is close to the ester bond but predicted to be outside of hydrogen bonding distance. The surface representation in (B) highlights the collar that encloses CGA.

Table 1. Wild-type *Lh*ChlE amino acids predicted to be involved in hydrogen bonding to CGA, EFA, and ECA.

	No. of H-bonds	Interacting amino acids	
		Side chain	Main chain
CGA	13	Y173, D138, S106, H229, Q145, K164, T35, R230	F34, A36
Ethyl ferulic acid	4	D138, S106, H229	F34
Ethyl caffeic acid	5	Y173, D138, S106, H229	F34

of a hydrogen bond between the ester and the side-chain of Lys_{164}.

Overall, MD simulation and docking results lead to the hypothesis that residues Lys_{164} and Gln_{145} are important for *Lh*ChlE function. Both residues directly interact with CGA, which likely increases substrate binding affinity. Furthermore, these residues may help impart a substrate preference for CGA over ECA and EFA. At the same time, both residues may influence conformational changes that likely are required for substrate binding and product release.

Enzyme kinetics of WT, K164A, and Q145A *Lh*ChlE

We hypothesize that Lys_{164} and Gln_{145} stabilize substrate binding to *Lh*ChlE and enhance CGA binding over ECA and EFA. These residues likely also influence substrate binding because of their role in modulating active site conformational changes. To test these hypotheses, the Michaelis–Menten parameters for CGA, ECA, and EFA hydrolysis were measured for the wild type and the two mutants. The measurements are shown in Fig. 10 and Table 2 and represent the averages and standard deviations from multiple independent triplicate runs.

For WT *Lh*ChlE, the kinetic constants for CGA hydrolysis are within the range of error of our previously published values (Fig. 10A and Table 2) [16]. The K_m values for ECA and EFA are higher than for CGA, and the k_{cat}/K_m values are lower. This indicates that *Lh*ChlE is more active against CGA than against ECA or EFA. The k_{cat} for ECA hydrolysis is higher than for CGA hydrolysis. We are unsure why ECA hydrolysis is faster than CGA hydrolysis but speculate that ECA's smaller size may allow for quicker product release.

The K_m of K164A mutant for CGA is 2.6-fold higher than the K_m of the WT, supporting our hypothesis that Lys_{164} is important for stabilizing CGA binding. The k_{cat} of the mutant increases 1.4-fold compared to the WT. The difference is statistically significant with a *P*-value of 0.022 over seven independent triplicate runs, so we are therefore confident that k_{cat} indeed increases in the K164A mutant. However, although k_{cat} increases, the mutant is nevertheless the worse enzyme for CGA hydrolysis as the k_{cat}/K_m value decreased 1.9-fold in the mutant. For both ECA and EFA hydrolysis, the values of K_m and k_{cat} between WT and K164A *Lh*ChlE are within error of each other. This further supports the hypothesis that Lys_{164} is involved in CGA but not ECA and EFA binding (Fig. 10B,C and Table 2). The increase in k_{cat} for CGA hydrolysis may be explained by the more open flap above the active site for K164A mutant, which may allow quinic acid release to occur more rapidly, and thus k_{cat} to increase. The observation that k_{cat} for ECA and EFA hydrolysis are not affected by the K164A mutation is somewhat surprising and suggests that the mutant's altered conformational dynamics in the Gln_{145}-Val_{203} collar do not influence catalysis.

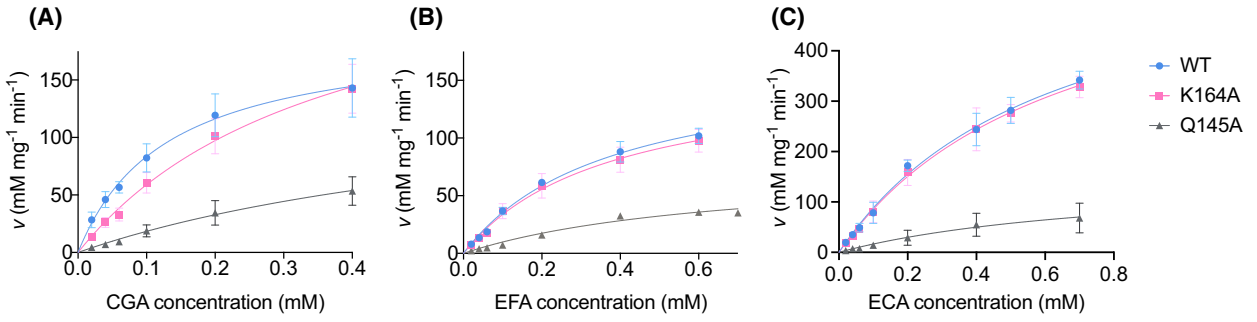

Fig. 10. Michaelis–Menten analysis for (A) CGA, (B) EFA, and (C) ECA hydrolysis. Each data point represents the mean and standard deviation of at least five independent triplicate measurements. The kinetic constants for these experiments are listed in Table 2. We note that the error for the K_m for ECA is large since saturation was not reached due to the enzyme's weak affinity for the substrate.

Table 2. Michaelis–Menten parameters (mean and standard deviation) for CGA, EFA, and ECA hydrolysis for WT, K164A, and Q145A *Lh*ChlE.

	K_m (mM)	V_{max} (μmol·min^{-1}·mg^{-1})	V_{max} (mM·min^{-1}·mg^{-1})	k_{cat} (s^{-1})	k_{cat}/K_m (mM^{-1}·s^{-1})
WT					
CGA	0.147 ± 0.041	164.6 ± 38.0	205.8 ± 47.6	82.6 ± 19.1	561.8
Ethyl ferulic acid	0.386 ± 0.046	144.9 ± 9.9	181.2 ± 12.4	72.7 ± 5.0	188.4
Ethyl caffeic acid	0.645 ± 0.112	499.9 ± 79.9	624.9 ± 99.8	250.8 ± 40.1	388.7
K164A					
CGA	0.384 ± 0.111	227 ± 49.4	283.8 ± 61.7	113.9 ± 24.8	296.5
Ethyl ferulic acid	0.38 ± 0.077	128.1 ± 15.2	160.1 ± 19.0	64.2 ± 7.6	168.9
Ethyl caffeic acid	0.564 ± 0.163	460.2 ± 117.1	575.2 ± 146.3	230.9 ± 58.7	409.0
Q145A					
CGA	0.623 ± 0.029	123.4 ± 22.7	154.3 ± 28.4	61.9 ± 11.4	99.4
Ethyl ferulic acid	0.704 ± 0.099	62.7 ± 9.1	78.4 ± 11.4	31.5 ± 4.6	44.7
Ethyl caffeic acid	0.855 ± 0.114	122.8 ± 44.3	153.5 ± 55.4	61.6 ± 22.2	72.1

The Q145A mutation has a significant effect on both the K_m and k_{cat} for CGA hydrolysis. The large increase in K_m confirms that Gln_{145} plays an important role in substrate binding by hydrogen binding to CGA. However, the Q145A mutation also causes the K_m to increase for ECA and EFA hydrolysis, and k_{cat}/K_m decreases significantly compared to the wild type for CGA, ECA, and EFA. Thus, unlike Lys_{164}, Gln_{145} does not appear to play a role in establishing a binding preference for CGA. The Q145A mutation is predicted to cause all three substrates to dock less deeply into the active site, providing an explanation for the increase in K_m for all three substrates. The Q145A mutation's effect on k_{cat} suggests that Gln_{145} likely has multiple roles. Gln_{145} is located close enough to the ester of the substrate (4.4 Å in the docking model) that it may be able to take part in catalysis, either by helping stabilize the transition state or by being involved in substrate positioning.

Discussion

In this work, we solved the structure of *Lh*ChlE, examined its dynamic behavior, and characterized its kinetic properties. The structure of *Lh*ChlE is very similar to that of *Lj*0536, as is its moderate substrate preference for CGA over ethyl ferulic acid. To date, all characterized CGA esterases are relatively specific toward the hydrolysis of aromatic esters, and are not reactive toward linear, short, and long-chained esters [7]. This behavior is consistent with the relatively occluded catalytic triad in bacterial CGA esterases. Martínez-Martínez *et al.* [46] determined that esterase specificity correlates well with the surface accessibility of the catalytic triad. The catalytic triad of *Lh*ChlE in the crystal structure has a surface accessible area of only 40.1 Å^2, as determined in GetArea [45], classifying it as a specific esterase. In addition, specificity for binding to aromatic substrates is likely promoted by residues in the binding pocket that are able to engage in π-stacking, such as Phe_{34} and Tyr_{173}. That said, while bacterial CGA esterases are specific toward aromatic esters, they are able to cleave a wide range of aromatic substrates that are structurally similar to CGA as well as model compounds such as *p*-nitrophenyl ester derivatives.

One goal of this paper was to determine how *Lh*ChlE achieves a moderate substrate preference for CGA hydrolysis. Mutational analysis in *Lh*ChlE demonstrated that Gln_{145} is important for enzyme function. This aligns with previous observations in *Lj*0536 that showed a change in both V_{max} and K_m when using the model substrate 4-nitrophenolbutyrate [4]. However, Gln_{145} is required for effective catalysis of both CGA and ECA/EFA and is therefore not involved in enacting a CGA preference. In contrast, Lys_{164} is important for achieving a moderate binding preference for CGA. Although Lys_{164} plays a role in CGA binding, it is not required for *Lh*ChlE function, as the K164A mutant only displayed a small reduction in catalytic efficiency. For comparison, mutations of individual residues located on loops surrounding the active site of the well-characterized hydrolase chymotrypsin lead to orders of magnitude-level differences in substrate specificity [47]. Although Lys_{164} was not required for *Lh*ChlE function, a structure-based sequence alignment suggests that the residue is semi-conserved between *Lh*ChlE, *Lj*0536, Est1E, *La*FAE, and *Lb*FAE (Fig. S17), which suggests that it may be involved in establishing a moderate substrate binding preference whenever it is present.

In addition to Lys_{164}, other residues in *Lh*ChlE may be important in helping establish a CGA binding preference. Docking results suggest that residues Thr_{35} and

Arg_{230} are able to form hydrogen bonds with CGA. While neither of these residues is conserved, they may play a role in establishing a substrate binding preference since they are not predicted to interact with ECA or EFA. We intend on examining the importance of these residues on CGA binding in future work.

In addition to being structurally similar to *Lj*0536, *Lh*ChlE is highly homologous to the ferulic acid esterases *Lb*FAE, *La*FAE, and Est1E. The three latter enzymes are able to release ferulic acid from feruloylated arabinoxylan (F-AX) from the cell wall of plant biomass [15,35,36]. However, CGA hydrolysis was never tested in any of them. Given their high structural similarity with *Lh*ChlE and *Lj*0536, it would be interesting to determine if these enzymes are active against CGA and, conversely, if *Lj*0536 and *Lh*ChlE are able to hydrolyze ferulic acid from F-AX in biomass. To date, the most common classification method for FAEs relies on measuring enzyme activity against a range of model substrates that may not reflect the enzyme's true range [48–50]. Given the high structural similarity between bacterial ChlEs and FAEs, and the fact that bacterial ChlEs are capable of hydrolyzing ferulic acid esters, we hypothesize that bacterial FAEs and ChlEs may, in fact, be part of the same enzyme family.

This work is the first to examine the dynamics of a bacterial ChlE. Based on the crystal structure, we hypothesized $Loop_{142–149}$ and $Loop_{160–167}$ are important in modulating the dynamics of the insertion domain. This hypothesis is supported since mutation of either residue leads to altered flexibility. $Loop_{142–149}$ (the Gln_{145}-containing loop) is more important in modulating the dynamics of the insertion domain since a Q145A mutation leads to a more open conformation, both around the active site pocket and at its mouth. The effect of the K164A mutation is different. While the mutation causes $Loop_{160–167}$ to flip outward more readily, the active site collar remains in a closed conformation. Nevertheless, the K164A mutant did not display lower ECA and EFA activity compared to the wild-type enzyme, suggesting that the collar remains sufficiently flexible to accommodate substrates.

Overall, the conformational changes of the insertion domain broadly resemble those of other esterases that fluctuate between open and closed conformations based on lid domain motions [41,42]. Moreover, the proposed role of flexible insertion domain loops in modulating substrate specificity is reminiscent of a *Bacteroides* ferulic acid esterase that may modulate substrate binding through loop movements [51]. Since *Lh*ChlE lacks homology to any esterase whose dynamics are well-characterized, this work represents the first description of the conformational dynamics of a bacterial chlorogenic acid esterase. While we determined the dynamics of the apo-enzyme, we did not attempt to measure the dynamics of the substrate-bound enzyme or study how the enzyme's conformational flexibility relates to catalysis. Future work will examine these questions to shed greater light on the mechanism of these industrially relevant enzymes.

Acknowledgements

The authors thank Dr. Marco Bisoffi for the use of instrumentation in his labs and Drs. Andy Borovik, Nicholas Chim, and James Nowick for sharing their beamtime with our group. We are grateful to Dr. Lilian W. Senger for the fruitful discussion and acknowledge her collaboration on NIFA grant 2020-67018-31261 that supported part of this project. This work was further supported by National Science Foundation, Division of Chemistry grant 1905399, Research Corporation for Science Advancement Cottrell Scholar Award CSA-2020-112, and a Research Corporation for Science Advancement Cottrell Postbac Scholar award to CPO. KKO acknowledges support by the Chapman Center for Undergraduate Excellence.

Author contributions

KKO, CTD, TLSO, NBC, BTD, DL, KNS, AT, and CPO were involved in investigation and editing; KKO, CTD, and CPO were involved in conceptualization, supervision, and writing; CPO was involved in funding acquisition.

Data accessibility

The data that support the findings of this study are available from the corresponding author, cpowens@chapman.edu, upon reasonable request.

References

1 Bauer TL, Buchholz PCF and Pleiss J (2020) The modular structure of alpha/beta-hydrolases. *FEBS J* **287**, 1035–1053.
2 Suplatov DA, Besenmatter W, Svedas VK and Svendsen A (2012) Bioinformatic analysis of alpha/beta-hydrolase fold enzymes reveals subfamily-specific positions responsible for discrimination of amidase and lipase activities. *Protein Eng Des Sel* **25**, 689–697.
3 Benoit I, Asther M, Bourne Y, Navarro D, Canaan S, Lesage-Meessen L, Herweijer M, Coutinho PM, Asther M and Record E (2007) Gene overexpression and biochemical characterization of the biotechnologically

relevant chlorogenic acid hydrolase from *Aspergillus niger*. *Appl Environ Microbiol* **73**, 5624–5632.
4 Lai KK, Stogios PJ, Vu C, Xu X, Cui H, Molloy S, Savchenko A, Yakunin A and Gonzalez CF (2011) An inserted alpha/beta subdomain shapes the catalytic pocket of *Lactobacillus johnsonii* cinnamoyl esterase. *PLoS One* **6**, e23269.
5 Hermoso JA, Sanz-Aparicio J, Molina R, Juge N, Gonzalez R and Faulds CB (2004) The crystal structure of feruloyl esterase A from *Aspergillus niger* suggests evolutive functional convergence in feruloyl esterase family. *J Mol Biol* **338**, 495–506.
6 Nieter A, Haase-Aschoff P, Kelle S, Linke D, Krings U, Popper L and Berger RG (2015) A chlorogenic acid esterase with a unique substrate specificity from *Ustilago maydis*. *Appl Environ Microbiol* **81**, 1679–1688.
7 Lai KK, Lorca GL and Gonzalez CF (2009) Biochemical properties of two cinnamoyl esterases purified from a *Lactobacillus johnsonii* strain isolated from stool samples of diabetes-resistant rats. *Appl Environ Microbiol* **75**, 5018–5024.
8 Bel-Rhlid R, Thapa D, Kraehenbuehl K, Hansen CE and Fischer L (2013) Biotransformation of caffeoyl quinic acids from green coffee extracts by *Lactobacillus johnsonii* NCC 533. *AMB Express* **3**, 28.
9 Pepra-Ameyaw NB, Lo Verde C, Drucker CT, Owens CP and Senger LW (2023) Preventing chlorogenic acid quinone-induced greening in sunflower cookies by chlorogenic acid esterase and thiol-based dough conditioners. *LWT* **174**, 114392.
10 Verde CL, Pacioles CT, Paterson N, Chin J, Owens CP and Senger LW (2023) Hydrolysis of chlorogenic acid in sunflower flour increases consumer acceptability of sunflower flour cookies by improving cookie color. *J Food Sci* **88**, 3538–3550.
11 Kraehenbuehl K, Page-Zoerkler N, Mauroux O, Gartenmann K, Blank I and Bel-Rhlid R (2017) Selective enzymatic hydrolysis of chlorogenic acid lactones in a model system and in a coffee extract. Application to reduction of coffee bitterness. *Food Chem* **218**, 9–14.
12 Leonard W, Zhang P, Ying D and Fang Z (2021) Hydroxycinnamic acids on gut microbiota and health. *Compr Rev Food Sci Food Saf* **20**, 710–737.
13 Fritsch C, Jansch A, Ehrmann MA, Toelstede S and Vogel RF (2017) Characterization of cinnamoyl esterases from different Lactobacilli and Bifidobacteria. *Curr Microbiol* **74**, 247–256.
14 Couteau D, McCartney AL, Gibson GR, Williamson G and Faulds CB (2001) Isolation and characterization of human colonic bacteria able to hydrolyse chlorogenic acid. *J Appl Microbiol* **90**, 873–881.
15 Goldstone DC, Villas-Boas SG, Till M, Kelly WJ, Attwood GT and Arcus VL (2010) Structural and functional characterization of a promiscuous feruloyl esterase (Est1E) from the rumen bacterium *Butyrivibrio proteoclasticus*. *Proteins* **78**, 1457–1469.
16 Lo Verde C, Pepra-Ameyaw NB, Drucker CT, Okumura TLS, Lyon KA, Muniz JC, Sermet CS, Were Senger L and Owens CP (2022) A highly active esterase from *Lactobacillus helveticus* hydrolyzes chlorogenic acid in sunflower meal to prevent chlorogenic acid induced greening in sunflower protein isolates. *Food Res Int* **162**, 111996.
17 Song YR and Baik SH (2017) Molecular cloning, purification, and characterization of a novel thermostable cinnamoyl esterase from *Lactobacillus helveticus* KCCM 11223. *Prep Biochem Biotechnol* **47**, 496–504.
18 Winter G, Beilsten-Edmands J, Devenish N, Gerstel M, Gildea RJ, McDonagh D, Pascal E, Waterman DG, Williams BH and Evans G (2022) DIALS as a toolkit. *Protein Sci* **31**, 232–250.
19 Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213–221.
20 Emsley P, Lohkamp B, Scott WG and Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486–501.
21 Delano WL (2002) The PyMOL Molecular Graphics System. Delano Scientific, San Carlos, CA.
22 Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE and Berendsen HJC (2005) GROMACS: fast, flexible, and free. *J Comput Chem* **26**, 1701–1718.
23 Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B and Lindahl E (2015) GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25.
24 Best RB, Zhu X, Shim J, Lopes PEM, Mittal J, Feig M and Mackerell AD (2012) Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain χ_1 and χ_2 dihedral angles. *J Chem Theory Comput* **8**, 3257–3273.
25 Jorgensen W, Chandrasekhar J, Madura J, Impey R and Klein M (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* **79**, 926–935.
26 Hess B, Bekker H, Berendsen HJC and Fraaije JGEM (1997) LINCS: a linear constraint solver for molecular simulations. *J Comput Chem* **18**, 1463–1472.
27 Bussi G, Donadio D and Parrinello M (2007) Canonical sampling through velocity rescaling. *J Chem Phys* **126**, 014101.
28 Essmann U, Perera L, Berkowitz M, Darden T, Lee H and Pedersen L (1995) A smooth particle mesh Ewald method. *J Chem Phys* **103**, 8577–8593.

29 Parrinello M and Rahman A (1981) Polymorphic transitions in single crystals: a new molecular dynamics method. *J Appl Phys* **52**, 7182–7190.
30 Daura X, Gademann K, Jaun B, Seebach D, van Gunsteren WF and Mark AE (1999) Peptide folding: when simulation meets experiment. *Angew Chem Int Ed Engl* **38**, 236–240.
31 R Core Team (2022) R: A Language and Environment for Statistical Computing in. R Foundation for Statistical Computing, Vienna.
32 Abagyan R and Totrov M (1994) Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. *J Mol Biol* **235**, 983–1002.
33 Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235–242.
34 Zhang Y and Skolnick J (2005) TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res* **33**, 2302–2309.
35 Jeon S, Hwang J, Do H, Le L, Lee CW, Yoo W, Lee MJ, Shin SC, Kim KK, Kim HW *et al.* (2023) Feruloyl esterase (LaFae) from *Lactobacillus acidophilus*: structural insights and functional characterization for application in ferulic acid production. *Int J Mol Sci* **24**, 11170.
36 Kasmaei KM, Kalyani DC, Reichenbach T, Jimenez-Quero A, Vilaplana F and Divne C (2022) Crystal structure of the feruloyl esterase from *Lentilactobacillus buchneri* reveals a novel homodimeric state. *Front Microbiol* **13**, 1050160.
37 Holm L (2020) Using Dali for protein structure comparison. *Methods Mol Biol* **2112**, 29–42.
38 Ferousi C, Kosinas C, Nikolaivits E, Topakas E and Dimarogona M (2023) Crystal structure of the *Fusarium oxysporum* tannase-like feruloyl esterase FaeC in complex with p-coumaric acid provides insight into ligand binding. *FEBS Lett* **597**, 1415–1427.
39 Ristinmaa AS, Coleman T, Cesar L, Langborg Weinmann A, Mazurkewich S, Branden G, Hasani M and Larsbrink J (2022) Structural diversity and substrate preferences of three tannase enzymes encoded by the anaerobic bacterium *Clostridium butyricum*. *J Biol Chem* **298**, 101758.
40 Mowbray SL, Elfstrom LT, Ahlgren KM, Andersson CE and Widersten M (2006) X-ray structure of potato epoxide hydrolase sheds light on substrate specificity in plant enzymes. *Protein Sci* **15**, 1628–1637.
41 Kumar R (2023) Structural dynamics and mechanistic action guided engineering of lipolytic enzymes. *J Cell Biochem* **124**, 877–888.
42 Barbe S, Lafaquiere V, Guieysse D, Monsan P, Remaud-Simeon M and Andre I (2009) Insights into lid movements of *Burkholderia cepacia* lipase inferred from molecular dynamics simulations. *Proteins* **77**, 509–523.
43 Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, Wang J, Cong Q, Kinch LN, Schaeffer RD *et al.* (2021) Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876.
44 Knapp B, Ospina L and Deane CM (2018) Avoiding false positive conclusions in molecular simulation: the importance of replicas. *J Chem Theory Comput* **14**, 6127–6138.
45 Fraczkiewicz R and Braun W (1998) Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *J Comput Chem* **19**, 319–333.
46 Martínez-Martínez M, Coscolin C, Santiago G, Chow J, Stogios PJ, Bargiela R, Gertler C, Navarro-Fernandez J, Bollinger A, Thies S *et al.* (2018) Determinants and prediction of esterase substrate promiscuity patterns. *ACS Chem Biol* **13**, 225–234.
47 Perona JJ and Craik CS (1997) Evolutionary divergence of substrate specificity within the chymotrypsin-like serine protease fold. *J Biol Chem* **272**, 29987–29990.
48 Dilokpimol A, Makela MR, Aguilar-Pontes MV, Benoit-Gelber I, Hilden KS and de Vries RP (2016) Diversity of fungal feruloyl esterases: updated phylogenetic classification, properties, and industrial applications. *Biotechnol Biofuels* **9**, 231.
49 Crepin VF, Faulds CB and Connerton IF (2004) Functional classification of the microbial feruloyl esterases. *Appl Microbiol Biotechnol* **63**, 647–652.
50 Hunt CJ, Antonopoulou I, Tanksale A, Rova U, Christakopoulos P and Haritos VS (2017) Insights into substrate binding of ferulic acid esterases by arabinose and methyl hydroxycinnamate esters and molecular docking. *Sci Rep* **7**, 17315.
51 Wefers D, Cavalcante JJV, Schendel RR, Deveryshetty J, Wang K, Wawrzak Z, Mackie RI, Koropatkin NM and Cann I (2017) Biochemical and structural analyses of two cryptic esterases in *Bacteroides intestinalis* and their synergistic activities with cognate xylanases. *J Mol Biol* **429**, 2509–2527.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Fig. S1. Experimentally determined extinction coefficients for the substrates and products used in this study.

Fig. S2. Structural alignment of *Lh*ChlE (blue), Est1E (2WTM, yellow), *Lb*FAE (7Z2X, salmon), and *La*FAE (7XRH, white), illustrating their structural similarity.

Fig. S3. (A) Radii of gyration, (B) RMSD, and (C) RMSF values for the *Lh*ChlE dimer from triplicate 50 ns simulations.
Fig. S4. Radii of gyration for (A) wild type, (B) K164A, and (C) Q145A *Lh*ChlE, suggesting simulations converged after 200 ns.
Fig. S5. Distance measurements between residues lining the entrance to the active site.
Fig. S6. SDS PAGE (10%) of (A) K164A and (B) Q145A at two different concentrations, demonstrating that the protein is pure. (C) Gel filtration chromatogram of WT, Q145A, and K164A *Lh*ChlE demonstrating that their molecular weight is the same.
Fig. S7. (A) Circular dichroism spectra of WT, Q145A, and K164A *Lh*ChlE indicating that the secondary structure is identical for both proteins. (B) Thermal denaturation curves indicate that the melting points for both proteins are similar to that of the wild type (65 °C).
Fig. S8. Electron density maps around the three Lys_{164} and Ala_{164} residues in the respective asymmetric units of the wild type and K164A mutant.
Fig. S9. RMSD values for Q145A and K164A *Lh*ChlE for the entire protein and the insertion domain for all three MD runs.
Fig. S10. RMSF values for the Q145A *Lh*ChlE mutant for all three MD runs.
Fig. S11. RMSF values for the K164A *Lh*ChlE mutant for all three MD runs.
Fig. S12. Full clustering analysis histogram for clusters that represent at least 2.5% of the total population.
Fig. S13. Distances between key residue pairs for WT, Q145A, and K164A *Lh*ChlE.
Fig. S14. Control for Molsoft ICM docking studies.
Fig. S15. CGA docking to (A) K164A and (B) Q145A *Lh*ChlE. The wild type enzyme and substrate docking pose is displayed in white, for comparison.
Fig. S16. EFA and ECA docking to (A) K164A and (B) Q145A *Lh*ChlE. EFA is orange whereas ECA is white.
Fig. S17. Structure-based sequence alignment using PROMAL3D [1] with *Lh*ChlE, Est1E (PDB ID: 2WTM), Lj0536 (PDB ID: 3PF8), *Lb*FAE (PDB ID: 7Z2X), LaFAE (PDB ID: 7XRH).
Table S1. Data collection and refinement statistics.