

Allosteric-Activation Mechanism Of Bovine Chymosin Revealed By Bias-Exchange Metadynamics and Molecular Dynamics Simulations

Samiul M. Ansari,[†] Andrea Coletta,[‡] Katrine Kirkeby Skeby,^{‡,¶} Jesper Sørensen,^{‡,§}
Birgit Schiøtt,^{*,‡} and David S. Palmer^{*,†}

[†]*Department of Pure and Applied Chemistry, University of Strathclyde, Thomas Graham Building, 295 Cathedral Street, Glasgow, Scotland G1 1XL, UK*

[‡]*Center for Insoluble Protein Structures (inSPIN), Interdisciplinary Nanoscience Center (iNANO), and Department of Chemistry, Aarhus University, Langelandsgade 140, DK-8000 Aarhus, Denmark*

[¶]*Current Address: Digizyme Inc., Boston, MA, USA.*

[§]*Current Address: Dart NeuroScience LLC, 12278 Scripps Summit Dr, San Diego, CA 92131, United States*

E-mail: birgit@chem.au.dk; david.palmer@strath.ac.uk
Phone: +4587155975; +441415484178. Fax: +4586196199

Abstract

The aspartic protease, bovine chymosin, catalyses the proteolysis of κ -casein proteins in milk. The bovine chymosin- κ -casein complex is of industrial interest as the enzyme is widely employed in the manufacturing of processed dairy products. The apo form of the enzyme adopts a self-inhibited conformation in which the side chain of Tyr77 occludes the binding site. On the basis of kinetic, mutagenesis and crystallographic data, it has been widely reported that a HPHPH sequence in the P8-P4 residues of the natural substrate κ -casein acts as the allosteric activator, but the mechanism by which this occurs has not previously been elucidated due to the challenges associated with studying this process by experimental methods. Here we have employed two computational techniques, molecular dynamics and bias exchange metadynamics simulations, to study the mechanism of allosteric activation and to compute the free energy surface for the process. The simula-

tions reveal that allosteric activation is initiated by interactions between the HPHPH sequence of κ -casein and a small α -helical region of chymosin (residues 112-116). A small conformational change in the α -helix causes the side chain of Phe114 to vacate a pocket that may then be occupied by the side chain of Tyr77. The free energy surface for the self-inhibited to open transition is significantly altered by the presence of the HPHPH sequence of κ -casein.

Introduction

Bovine chymosin is an aspartic protease enzyme found in the stomachs of calves. Its native function is to selectively cleave κ -casein proteins in milk to initiate coagulation of casein micelles thereby aiding digestion.¹⁻³ This function is exploited in industry to induce milk-clotting in the manufacturing of processed dairy products. Over 20.5 million metric tonnes of cheese are currently produced each year using bovine chy-

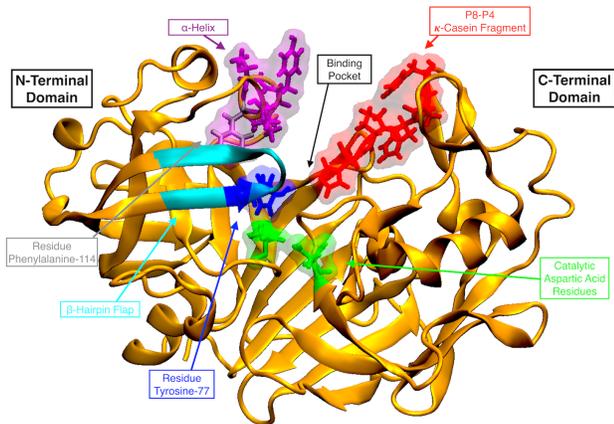


Figure 1: Depiction of the chymosin- κ -casein complex.

The recent discovery that camel chymosin is an effective clotting agent for bovine milk has led to renewed interest from industry in understanding the catalytic behaviour of both bovine and camel chymosin. The camel variant displays 70% higher clotting activity than bovine chymosin for bovine milk, while only having 20% of the unspecific proteolytic activity. By contrast, bovine chymosin is a poor coagulant for camel milk.⁵ Although both camel and bovine chymosin are now marketed for use in the food industry, several aspects of their biological function are not well understood at a molecular level. Furthermore, the reasons for the disparity in catalytic propensity between bovine and camel chymosin have not been fully explained. Here we focus on elucidating the mechanism that allows allosteric activation of bovine chymosin by the P8-P4 residues of the substrate κ -casein, as proposed by Visser *et al.*^{6,7} and widely reported in the literature.^{8,9}

Chymosin is a globular monomeric aspartic protease made up of 323 amino acids as depicted in fig. 1. It folds into a pseudosymmetric bi-lobal structure creating a binding cleft containing the catalytic residues Asp34 and Asp216. The enzyme exhibits 9% sequence identity between the N- and C-termini.¹⁰ The two catalytic aspartic acid residues are stabilised by a network of hydrogen bonds which incorporates two threonine residues, referred to as “*the fireman’s grip*”.¹¹ Chymosin contains

three disulphide bridges (Cys45-Cys50, Cys206-Cys210 and Cys249-Cys282) and five ion pairs (Arg59-Asp57, Arg157-Glu308, Arg157-Ile326, Arg307-Asp11 and Arg315-Asp138).^{10,12} The structure contains a single *cis-proline* residue, Pro23, found in the N-terminus.¹³ These features are also found in other aspartic proteases showing strong homology within the protein family.^{9,14–16}

The substrate of chymosin, κ -casein, is a 169 residue peptide found on the surface of milk serum aggregates called *casein micelles*.¹⁷ Chymosin selectively cleaves the Phe105-Met106 bond in κ -casein, which destabilises the casein micelles leading to the release of insoluble casein proteins causing milk clotting.^{8,18}

Crystal structures reveal that the side chain of Tyr77 in a β -hairpin flap region above the binding cleft in bovine chymosin can occupy two different positions.¹⁹ The side chain can be extended over the binding pocket occupying the position where κ -casein binds (referred to as the self-inhibited conformation) or it can be extended back into the β -hairpin flap (referred to as the open conformation).^{20,21} The self-inhibited to open transition is associated with a rearrangement of the β -hairpin flap, which becomes more puckered in the open conformation. It has been widely reported that apo-chymosin exists in the self-inhibited form and that this is converted into its active open form by allosteric activation by the P8-P4 fragment of κ -casein, His98-Pro99-His100-Pro101-His102, the so called “*His-Pro*” cluster.¹⁶ Evidence for this allosteric activation mechanism came originally from the experiments of Visser *et al.*⁷ and Gustchina *et al.*,¹⁶ who measured the catalytic rates for proteolysis of different fragments of κ -casein. They observed a ~ 200 fold reduction in catalytic rate for proteolysis of P2-P2’ or P3-P3’ fragments of κ -casein as compared to the native substrate. However, the reduction in catalytic rate was not observed when chymosin had been pre-incubated with the P8-P4 residues of κ -casein. Taken together with crystallographic data,²¹ which show that apo-chymosin occupies a self-inhibited conformation, the experiments carried out by Visser *et al.* suggest that the P8-P4

residues act as an allosteric activator. Further mutagenesis studies have demonstrated that all five of the residues in the *His-Pro* cluster are important for catalysis.^{6,7} The HPHPH cluster is conserved in many other mammalian κ -casein peptides including buffalo and goat, but in camel κ -casein the three histidine residues are mutated to arginines.^{1,5} Although the allosteric activation process has been widely discussed in the literature,^{1,7,16,19–21} the mechanism has not been elucidated at a molecular level due to the challenges associated with studying it by experimental methods. This has hindered the development of novel enzymes and enzymatic processes for the food industry.¹ Here we employ two computational techniques, molecular dynamics (MD) and bias exchange metadynamics (BEMD) simulations, to reveal the allosteric activation mechanism and its associated free energy surface. BEMD is an enhanced sampling technique that allows the efficient exploration of complex free energy landscapes. It is well suited to studying conformational/configurational transformations in biomacromolecules and has previously been used to study protein folding,^{22,23} protein-ligand recognition²³ and allosteric transitions.^{24,25}

Methods

Molecular Dynamics Simulations

Unrestrained MD simulations were performed for the six chemical systems described in table 1, which include the open and self-inhibited conformations of apo-chymosin and four replicas of the self-inhibited chymosin - P8-P4- κ -casein complex. The four replicas differed only in whether or not capping groups were applied to the κ -casein fragment and whether the side chain of His102 in κ -casein was modelled as protonated or neutral (with a proton on the N- δ atom of the imidazole ring). Each system was simulated four times: two simulations using the AMBER ff03²⁶ force field and two simulations using the AMBER ff99SB-ILDN²⁷ force field (6 systems \times 2 force fields \times 2 duplicates =

24 simulations in total). Duplicate simulations were started from the same coordinates but with atoms being assigned different initial velocities in different simulations.

Table 1: Six chemical systems used as input for the molecular dynamics simulations

ID	System	Apo/Holo ^a	Capping ^b	Proton ^c
A1	Open	Apo	-	-
B1	Self-inhibited	Apo	-	-
C1	Self-inhibited	Holo	Capped	No (HID)
C2	Self-inhibited	Holo	No Cap	No (HID)
C3	Self-inhibited	Holo	Capped	Yes
C4	Self-inhibited	Holo	No Cap	Yes

^a the holo complex is *P8-P4* κ -casein bound to chymosin; ^b His-Pro fragment can be capped at both ends or can be uncapped (no cap) at *P4His* position; ^c κ -casein fragment protonated at *P4His* position.

Input coordinates

Initial coordinates for the MD simulations were taken from our previous work.^{20,28–30} Only a brief summary of the steps used to prepare the input coordinates will be provided, since the details have previously been reported.^{20,28–30} In summary, the coordinates of chymosin were taken from the 3CMS crystal structure (Tyr77 is resolved in both open and self-inhibited forms). Chymosin coordinates were modified to: insert missing residues (Asn291-His292-Ser293); reverse the Val111Phe mutation; introduce disulphide bonds between Cys47-Cys52, Cys207-211 and Cys250-Cys283; assign amino acid protonation states appropriately for pH 6.5; include 16 conserved water molecules identified by Prasad *et al.*³¹ Since there are no crystal structures of chymosin- κ -casein complexes, the coordinates of the chymosin sensitive regions of κ -casein (residues P9-P7') in the complex were generated by a two step process.²⁰ Firstly, the P2-P2' residues were docked into the 3CMS structure of apo chymosin and relaxed by MD simulations. The binding pose has been shown to be in the correct geometry to allow proteolysis of the P1-P1' (Phe105-Met106) amide bond in κ -casein via the established reaction mechanisms.^{20,32–36} Secondly, the remaining residues were grown in the binding cleft using conformational search algorithms and MD

simulations. The resulting binding pose has previously been shown to be in good agreement with a crystal structure of a chymosin-inhibitor bound complex^{18,20} and with previous computational studies of the same system.⁸ Free energy calculations using this bound pose also agree with the results of experimental mutagenesis studies.^{20,28,30} In these previous models of the chymosin – P9-P7’ κ -casein complex, the sidechains of HisP8 and HisP4 in κ -casein were modelled as positively charged, while HisP4 was modeled as the neutral N- δ tautomer (since close contacts with the sidechain of Lys221 disfavor the protonated form).²⁰ These protonation states were assigned based on predictions from PROPKA2.0 and comparisons of binding energies computed using Poisson-Boltzmann solvent models.^{20,37} In the simulations reported here, we tested the positively charged form of HisP4 as well as the N- δ tautomer because HisP4 is more solvent exposed in the complex of chymosin – P8-P4 κ -casein. However, as demonstrated later, the HisP4 protonation state was not deemed to affect the conclusions. The chymosin – P8-P4- κ -casein complex that was simulated here was obtained by deleting the P9 and P3-P7’ residues and adding hydrogens or capping groups to complete the valency, as necessary. The self-inhibited complexes were obtained by copying the P8-P4- κ -casein residues from the open complex into self-inhibited, apo chymosin (after alignment on the chymosin coordinates), followed by relaxation of the coordinates of the complex by constrained minimisation and MD simulations, as described below.

System Preparation

Molecular dynamics simulations were performed in NAMD.³⁸ Each protein or protein-ligand complex was solvated by TIP3P³⁹ water molecules using the XLEAP module in AmberTools14.0.⁴⁰ Approximately 15000 water molecules were placed around the protein in a (rectangular cuboid) periodic box. All systems were neutralised then given an ionic strength of 0.07 mol dm⁻¹ using chloride and sodium ions as required.

Simulations

The solvated complexes were relaxed by conjugate gradient energy minimisation in four steps of 5000 iterations. In steps 1 to 3, the whole protein, the protein backbone, and the α -carbon atoms, respectively, were held fixed. All constraints were removed in the fourth step. The systems were gradually heated to 300 K in the NVT ensemble over 10 ps with the α -carbons held fixed, followed by a 4 ns equilibration at 300 K with no constraints.

Equilibration and production simulations were performed in the isothermal-isobaric (NPT) ensemble⁴¹ at 300 K and 1 atm. The pressure was regulated by the Nosé-Hoover Langevin piston pressure control⁴² with the piston set up to a target of 1.01325 bar, a period of 200 fs, a decay of 100 fs and a temperature of 300 K.⁴³ The temperature of the system was maintained by means of Langevin dynamics with the dampening coefficient set to 2 ps⁻¹, but not affecting hydrogens. Periodic boundary conditions were applied to the systems and electrostatic interactions were calculated by the particle mesh Ewald (PME) method.⁴⁴⁻⁴⁶ A cut-off distance of 10 Å was set for van der Waals’ interactions using a switching distance of 9 Å. The pair list was updated every 20 ns for atom pairs within 11 Å. The distances of all bonds between hydrogen atoms and hetero-atoms were constrained by the SHAKE algorithm.^{47,48} The velocity verlet algorithm was used to update the equations of motions every 2 fs, and snapshots were taken every 2 ps. For each system, a 2 ns equilibration was performed, followed by 80-100 ns of production dynamics.

Analysis

The open and self-inhibited forms of chymosin are distinguished by the N-C $_{\alpha}$ -C $_{\beta}$ -C $_{\gamma}$ (χ_{77}) dihedral angle in Tyr77 (fig. 2), which is approximately 300° in the open form and approximately 175° in the self-inhibited form (the dihedral angle is expressed on a scale from 0° to 360°, rather than the more common -180° to 180° scale, because it simplifies the resulting figures and free energy surface diagrams). In

the open form, the side chain of Tyr77 tucks into a pocket under the β -hairpin flap formed by residues 74 to 82 of chymosin, while in the self-inhibited form it occludes the binding site. Measuring the Tyr77 dihedral as a function of simulation time is therefore a convenient method to identify transitions between open and self-inhibited forms.

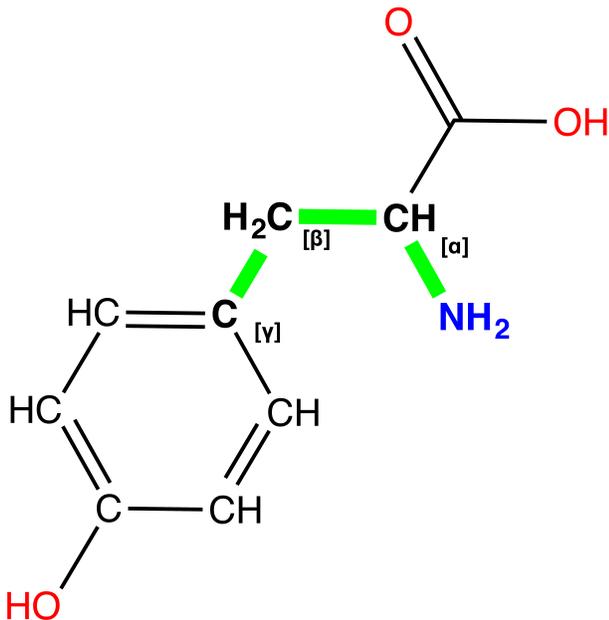


Figure 2: Tyr77 dihedral angle definition. Dihedral angle bonds in bold.

To provide further insight into the observed allosteric activation mechanisms, the dihedral angle in residue Phe114 ($C-C_\alpha-C_\beta-C_\gamma$), important residue-residue close-contacts, and hydrogen bonding networks stabilising Tyr77 were also measured as a function of simulation time. All measurements were automated using customised Tcl scripts in VMD.

BEMD – Bias-Exchange Metadynamics Simulations

Well-tempered Bias-Exchange Metadynamics simulations²³ of apo-chymosin (A1, B1 in table 1) and chymosin in complex with P8-P4 κ -casein residues (C1 in table 1) were performed in GROMACS-5.0.4 using the PLUMED-2.1 plug-in.⁴⁹ The Free-Energy Surfaces (FES) were reconstructed using two collective variables (CV):

1. The dihedral angle χ of Tyr77 (χ_{77} , defined by the N-C $_\alpha$ -C $_\beta$ -C $_\gamma$ atoms).
2. The number of contacts between the side chains of Tyr77 and Phe114, measured using the PLUMED implemented CV coordination number (CN) :

$$CN = \begin{cases} 1 & \text{if } r_{ij} \leq 0 \\ \sum_{ij} \frac{1 - \left(\frac{r_{ij}}{r_0}\right)^6}{1 - \left(\frac{r_{ij}}{r_0}\right)^{12}} & \text{if } r_{ij} > 0 \end{cases}$$

where $r_{ij} = |r_i - r_j| - d_0$, (r_i and r_j being the coordinates of Tyr77 and Phe114 atoms respectively). The values of the parameters d_0 and r_0 have been set to 4.0 and 3.0 Å, respectively.

In the case of apo-chymosin the BEMD simulations were performed using 4 replicas (one for each combination of the two structures A1 and B1 and the two CVs) while in the case of the chymosin - P8-P4 κ -casein complexes two replicas were used (one for each of the two CVs). In order to ensure the sampling of the free energy surface in the presence of the P8-P4 residues of κ -casein we included two piecewise linear/harmonic distance restraints: one between the terminal cap of P8 and N $_\delta$ of Asn241 and one between the terminal cap of P4 and O $_\gamma$ of Ser220. The BEMD simulations were subsequently analysed using the VMD plug-in “METAGUI”.⁵⁰

In order to characterize the correlation between the Tyr77 dihedral angle and the conformation of the protein we used the mutual information entropy⁵¹ value of χ_{77} dihedral and the protein secondary structure calculated as :

$$\mu_{res_i} = - \sum_{ss_i} \int \rho(\chi, ss_i) \log_2 \left(\frac{\rho(\chi, ss_i)}{\rho(\chi)\rho(ss_i)} \right) d\chi$$

where ss_i is the DSSP secondary structure⁵² and the probability densities were estimated from the metadynamics simulation using 144 and 8 bins for χ_{77} and ss_i respectively.

This statistical measure is a generalization of the linear correlation coefficient and gives an estimate of the extra-information gained using

the joint distribution function $\rho(\chi, ss_i)$ instead of the two single distributions $\rho(\chi)$ and $\rho(ss_i)$.

It can be shown that mutual information $\mu(a, b)$ between two random variables a and b can be expressed as $\mu(a, b) = (H(a) + H(b)) - H(a, b)$.

H is the Information Entropy[?] :

$$H(x) = - \sum_i p(x_i) \log_2(p(x_i))$$

where $p(x_i)$ is the probability of event x_i . Information entropy is a measure of the information content in the variable x , and of the number of "bits" needed to efficiently encode a time series of that random variable. Mutual Information $\mu(a, b)$ can be interpreted as a measure of the reduced number of bits needed to encode the information content in the joint distribution (a, b) with respect to the total amount needed to encode two single distributions for a and b separately.

Since it can be shown that $H(a, b) \leq H(a) + H(b)$ with the equality standing only in the case of a and b being independent, MI can be used as a generalized measure of correlation between a and b being not restricted to pure linear relationship between the two variables. The usual unit of measure for mutual info is the "bit".

Results and discussion

Molecular Dynamics

Apo-Chymosin

In none of the eight unrestrained MD simulations of *apo*-chymosin (>800 ns simulation time) was a transition between the open and self-inhibited forms of the enzyme observed. Analysis of the Tyr77 dihedral angle reveals a clear distinction between simulations started from either the open or self-inhibited forms (fig. 3).

Since the transition between open and self-inhibited forms was not observed in these simulations, it suggests that there is a high-barrier for rotation around the Tyr77 dihedral angle (and the associated rearrangement of the

β -hairpin flap) in the absence of the P8-P4 κ -casein pentapeptide, which agrees with the proposed allosteric activation method (further sampling of the dihedral angle is carried out using BEMD in the next section).

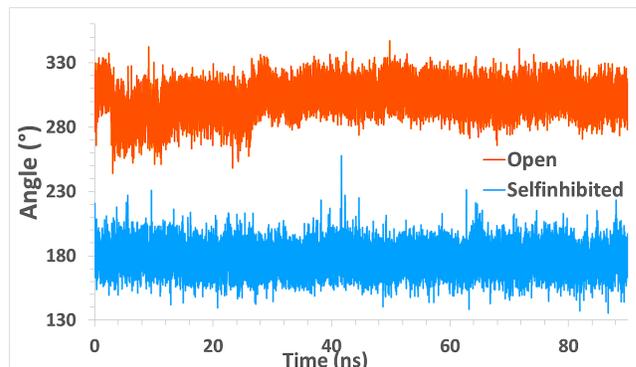


Figure 3: Tyr77 dihedral angle at open (orange, A1) and self-inhibited (blue, B1) conformations.

Tyr77 in its open conformation is found to be stabilised by a single water molecule. This stabilising water forms hydrogen bonds with Tyr77 and residues Ser37 and Asp39 of chymosin, as depicted in fig. 4. The water molecule has previously been shown to be conserved in crystal structures of aspartic proteases.³¹

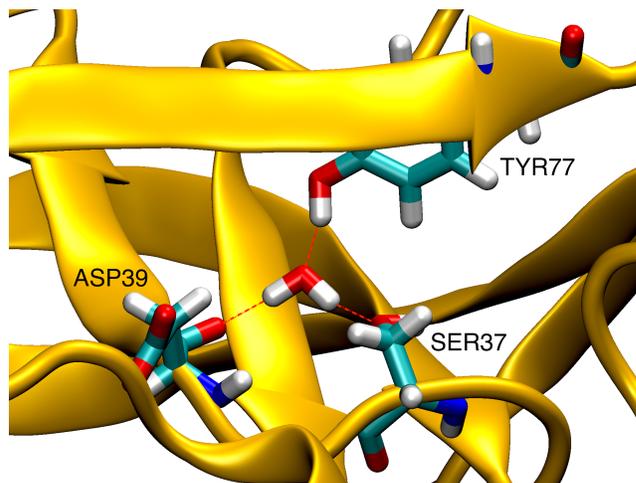


Figure 4: Open Tyr77 stabilisation

A different stabilising network is observed in the simulations in which the enzyme is in the self-inhibiting conformation. A single water molecule forms hydrogen bonds with Tyr77, Ser14 and Gly218, both of which reside in the binding pocket (fig. 5). The same hydrogen

bonding pattern is also observed for short recurring periods in the three simulations in which a change in Tyr77 conformation takes place. In these simulations, the water molecule was regularly displaced in short succession prior to the conformational change occurring, which suggests that the κ -casein fragment affects this hydrogen bonding pattern.

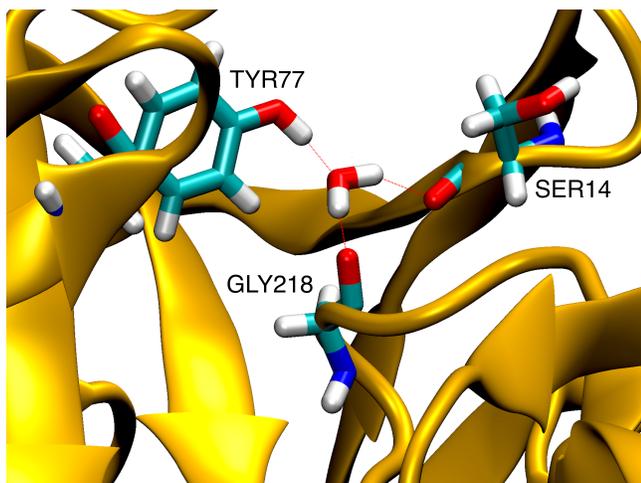


Figure 5: Self-inhibited Tyr77 stabilisation

Chymosin – P8-P4 κ -Casein

The allosteric transition from self-inhibited to open conformation in the presence of κ -casein fragment is expected to occur on a sufficiently long timescale to make it difficult to completely sample using regular MD simulations on current computational hardware (this sampling problem is addressed later by the use of bias-exchange metadynamics simulations). Nevertheless, the allosteric transition was observed in three of the regular MD simulations that included the κ -casein fragment (The C1 simulation using the AMBER ff99SB-ILDN force field and simulations C2 and C3 using the AMBER ff03 force field).

In the C1 simulation the dihedral angle of Tyr77 changes from self-inhibited to open at 70 ns in the 90 ns trajectory, shown in fig. 6. At 50 ns there is a deviation which lasts for 6 ns but this is not sustained and the dihedral angle of Tyr77 returns to self-inhibited until the change in form at 70 ns. The first conformational change at 50 ns is not sustained because

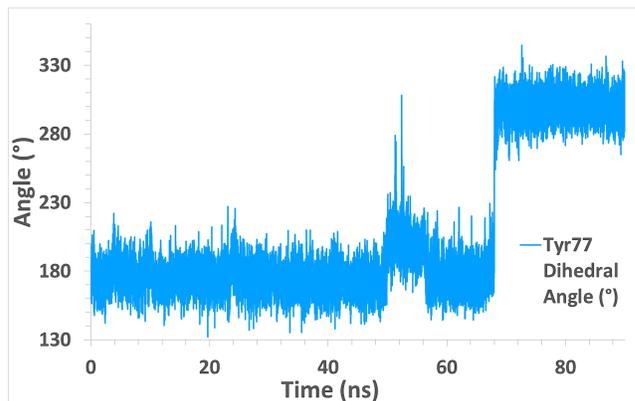


Figure 6: Tyr77 dihedral angle (blue) - AMBER ff99SB-ILDN Selfinhibited-HPHPH complex (C1).

residue Phe114 on the adjacent α -helix remains in the space that Tyr77 would occupy in its open conformation. The open conformation observed from 70 ns onwards is stabilised by the same hydrogen bonding pattern between Tyr77, water, Asp39 and Ser37 as observed in the simulation of apo-chymosin complex (and depicted in fig. 4).

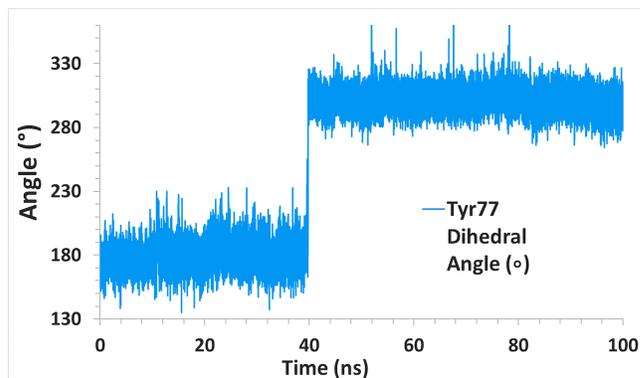


Figure 7: Tyr77 dihedral angle (blue) - AMBER ff03 Selfinhibited-HPHPH complex with no capping group (C2).

The change in form in the C2 simulation occurs at 40 ns where Tyr77 moves from a self-inhibiting position to its open conformation (fig. 7). This change is sustained for the remainder of the 100 ns trajectory. After the conformational change, Tyr77 is observed to make the same interactions that it does in the simulation of apo-chymosin in its open form (A1), including the hydrogen-bonding network between Tyr77, water, Asp39 and Ser37 that is depicted

in fig. 4.

The conformational change occurs early on in the 80 ns simulation of system C3 (see supporting information). At 6 ns Tyr77 moves from a self-inhibiting position to open and remains in this state for 18 ns. In this pose the water molecule in the hydrogen bonding network that stabilises the open conformation is continuously displaced and replaced in short succession, which suggests that the stable open conformation has not been fully reached. There is an increase in contact between Phe114 and Trp41 (located at the base of the β -flap) when Tyr77 is in its open conformation (self-inhibited:35% \rightarrow open:51%), which is the opposite of what is found in simulations C1 and C2, 37% \rightarrow 17% and 66% \rightarrow 32% respectively. This suggests steric interference inhibits the stabilisation of open Tyr77. The reverse transformation occurs at 24 ns where Tyr77 returns to its self-inhibited position concluding that the open form was not stabilised.

Analysis of the MD simulations provides an initial indication of the mechanism by which the κ -casein fragment induces allosteric activation. The κ -casein fragment interacts with the α -helical region of chymosin causing a sequence of changes, all of which must occur to give allosteric activation. The key changes include: (i) disruption of the hydrogen-bonding network between Tyr77, water, Ser14 and Gly218 that would otherwise help to stabilise Tyr77 in the self-inhibited form; (ii) interaction of the P8-P4 residues of κ -casein with the short α -helix in residues 112-116 of chymosin, which causes movement of the side chain of Phe114 such that it vacates the pocket that is occupied by the side chain of Tyr77 in the open conformation, (iii) rearrangement of the β -hairpin flap to allow rotation of the Tyr77 dihedral from its self-inhibited to open conformation. The steps in this pathway were not observed in their entirety in any of the simulations that did not show allosteric activation.

BEMD – Bias-Exchange Metadynamics Simulations

Free Energy Surface

To further investigate the influence of the P8-P4 κ -casein residues on the Tyr77 conformation, two bias-exchange metadynamics simulations were performed, one for the apo-enzyme and one for the chymosin – P8-P4 κ -casein fragment complex. The Free-Energy Surface (FES) as a function of χ_{77} and CN is reported in fig. 8-A1. The minima corresponding to the open and closed (self-inhibited) conformation of the apo-enzyme are highlighted with a green and a red spot respectively. The dihedral angles and the CN values of the stable states are reported in table 2. The open state minimum is found to have a low number of contacts between Tyr77 and Phe114 and a dihedral angle of $305^\circ \pm 5$ while the self-inhibited state minimum has a high number of contacts and a Tyr77 dihedral angle of $185^\circ \pm 5$. The path connecting the open and the self-inhibited states is divided in two sub-steps joined by an intermediate state with a Tyr77 dihedral value of $305^\circ \pm 5$ (equal to the open state), and a high coordination number between Tyr77 and Phe114 (similarly to the self-inhibited state). These findings provide further evidence that the small helix spanning from residue 112 to residue 116 of chymosin plays an important role in the interconversion from the self-inhibited to the open state of the enzyme.

Table 2: χ_{77} dihedral angle and CN values of the stable states observed in the apo-chymosin BEMD simulation. The errors are the bin-widths used to calculate free energy in the VMD plugin METAGUI

State	χ_{77}	CN
Open	$305^\circ \pm 5$	7 ± 5
Intermediate	$305^\circ \pm 5$	150 ± 5
Self-inhibited	$185^\circ \pm 5$	172 ± 5

A close-up of the structural change at the interface between the flap region and the small 112-116 helix in the three minimal states found in the apo-chymosin BEMD simulation is visible in fig. 8-B1 where a coordinate shift of the

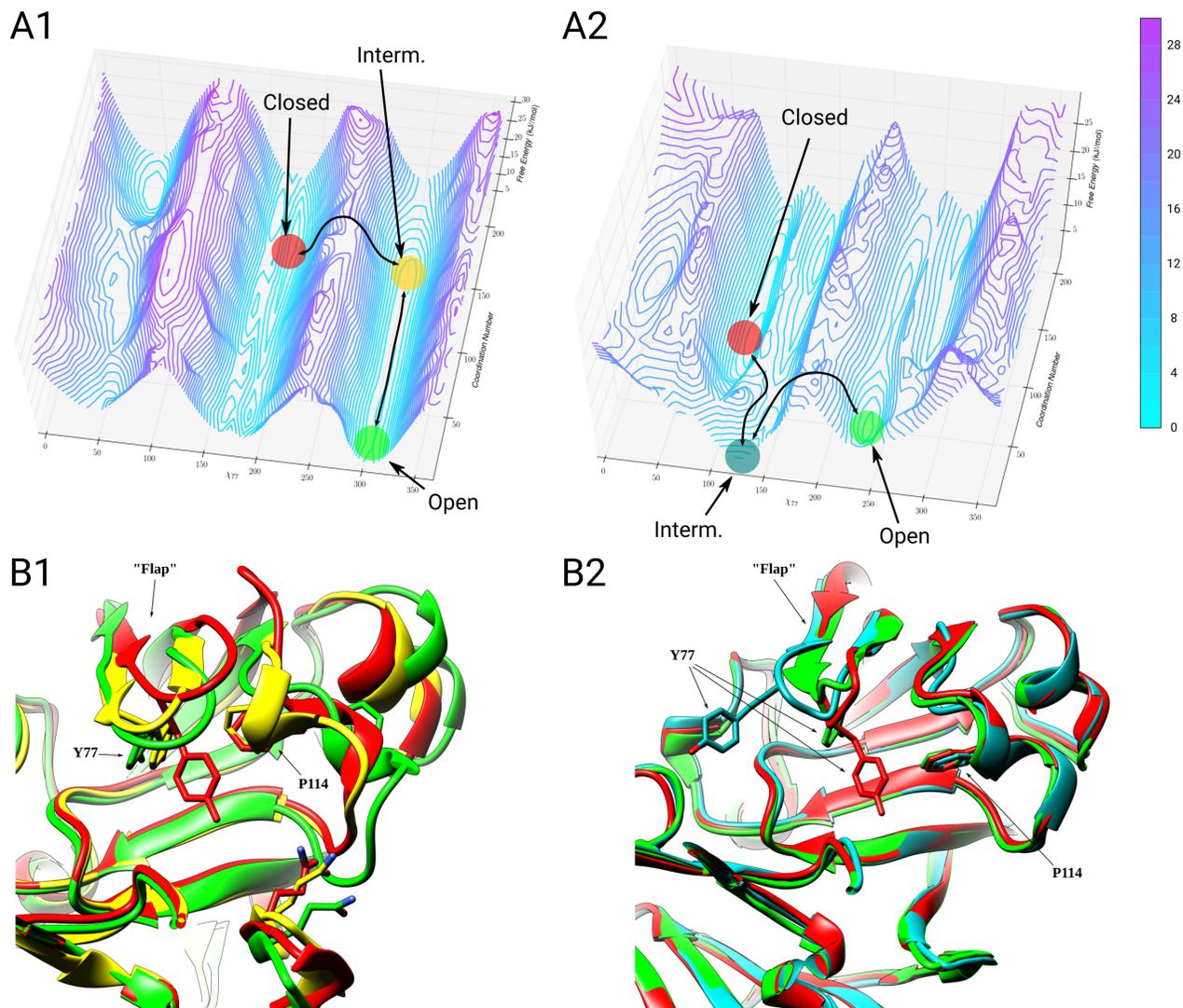


Figure 8: Top Panel: Free-Energy Surface (FES) as a function of χ_{77} and CN obtained from the BEMD simulation of apo-chymosin (A1) and chymosin – P8-P4 κ -casein complex (A2). The Open (green), Self-inhibited (red) and Intermediate (yellow) state of the enzyme are indicated with colored spots. A simplified picture of the transition path from Open to Self-inhibited state is reported as a black line Bottom Panel : Representative structure of the FES minima obtained from the BEMD simulation of apo-chymosin (B1) and chymosin – P8-P4 κ -casein complex. The enzyme is represented as ribbon and the coloring scheme is the same used in the Top Panel.

flap and the 2-turn helix is observed. In the first phase, the side chain of Tyr77 passes from the self-inhibited conformation (red) to the intermediate state (yellow) maintaining contact with Phe114, while the α -helix changes its conformation in concurrence with the β -flap region. The intermediate state observed here is considered unstable and is quickly transformed into the open conformation (green) in the trajectory. In the second phase the number of

contacts between Phe114 and Tyr77 is reduced and the small helix returns to a conformation close to the original. This is confirmed through residue contacts analysis carried out in simulations A1 and B1 systems used in regular MD, see supporting information. The findings are also in agreement with the regular MD simulations of holo chymosin, where a simultaneous movement of Tyr77 and Phe114 is observed as Tyr77 moves from its self-inhibited to its open

pose.

Table 3: χ_{77} dihedral angle and CN values of the stable states observed in the BEMD simulation of the chymosin – P8-P4 κ -casein complex. The errors are the bin-widths used to calculate free energy in the VMD plugin METAGUI

State	χ_{77}	CN
Open	$235^\circ \pm 5$	7 ± 5
Intermediate	$125^\circ \pm 5$	5 ± 5
Self-inhibited	$105^\circ \pm 5$	105 ± 5

When the protein is in complex with the κ -casein fragment a dramatic change in the FES is observed (cf. fig. 8-A2 and table 3). The minimum corresponding to the open state is shifted to a Tyr77 dihedral angle of $235^\circ \pm 5$ and a low coordination number with Phe114 while the minimum corresponding to the closed state is characterized by a shifted Tyr77 dihedral angle of $105^\circ \pm 5$ and a high coordination number. The systematic shift in the dihedral angles is possible because of a pronounced twisting of the β -hairpin flap in the BEMD simulations, which allows the side chain of Tyr77 to occupy the normal pockets in the open and self-inhibited conformations despite the change in angles. We believe that the twisting of the β -hairpin flap is more pronounced in the BEMD simulations than the regular MD simulations because the former allows a more thorough sampling of the conformational change. Nonetheless, the difference in the Tyr77 dihedral angles remains $\sim 130 \pm 5$ degrees in the BEMD simulations (similar to that observed in the MD simulations and the 3CMS crystal structure). The importance of the β -hairpin flap in aspartic proteases has previously been highlighted in studies of mammalian (chymosin, BACE) and viral (HIV-protease) enzymes.^{53,54} Interestingly, in the BEMD simulation of the complex, an intermediate state is found, but with a dihedral angle similar to the closed state ($105^\circ \pm 5$) and a low coordination number giving a different picture to what is observed in the apo-enzyme FES. Since free energy estimates are less accurate for higher-energy regions of phase-space (which are less well sampled during simulations), some caution must be exercised in esti-

ating barrier heights from the data in fig. 8-A1 and fig. 8-A2. Nonetheless, in the apo-enzyme, the open and intermediate states are clearly separated by a very low energy barrier while the intermediate and closed state are separated by a high energy barrier. By contrast, in the chymosin – P8-P4 κ -casein complex the closed and intermediate states are in the same kinetic basin while the intermediate and open states are divided by a high energy barrier.

From observations of the minimal structure states found for chymosin – P8-P4 κ -casein complex (fig. 8-B2), it appears that in the intermediate state (cyan) the side chain of Tyr77 is pointing away from Phe114 and the active site of chymosin. This intermediate state is nominally an active conformation because the side chain of Tyr77 does not occlude the binding site. The conformation is observed only fleetingly in the regular MD simulations, however. The coordinated motions of the small helix and the β -flap regions found in the apo-enzyme transition are less obvious in the holo-enzyme transition; the small helix where Phe114 resides, conserves its structure in all three states, which is in good agreement with what is observed in the regular MD simulations.

Mutual Information

The regions in the enzyme having a high mutual information (MI) value (cf. fig. 9-A and fig. 9-C) are those in proximity to Tyr77, roughly from residue 70 to 80 (β -flap region), residues 110 to 120 (where the small helix and Phe114 are found) plus some individual residues (148,162 and 163), a small loop (residues 240-246) and a β -hairpin (residues 276-283) constituting the binding site of P8-P4 κ -casein fragment on the chymosin C-terminal domain.

A general reduction of the mutual information between Tyr77 rotation and the enzyme domain change is observed in the chymosin – P8-P4 κ -casein complex (cf. fig. 9-B and fig. 9-D). The binding of the His-Pro cluster between the N-terminal β -hairpin and the C-terminal domain disrupts the communication network observed in the apo enzyme by modifying the conformation of the small α -helix, which, by a cascade

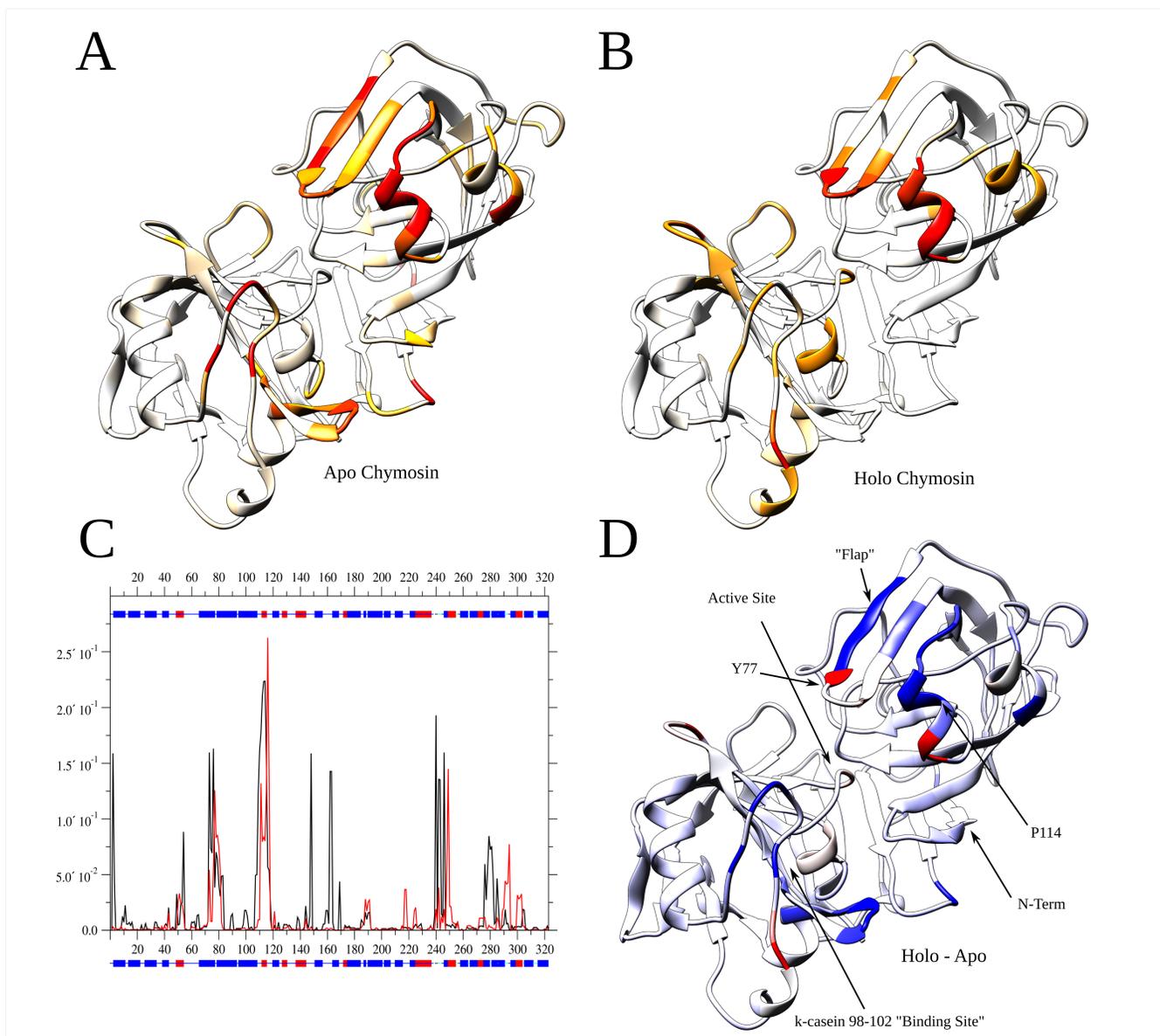


Figure 9: Mutual Information (MI) entropy between χ_{77} and secondary structure of chymosin residues. Panel A: Residues on the Apo-Enzyme system with a MI greater than 0.25 or 0.5 of the max value are colored in orange and red respectively. Panel B: MI of the Holo- Enzyme complex (same coloring scheme as Panel A). Panel C: Comparison of MI for Apo- (black line) and Holo-Enzyme (red line), as a function of residue number; Secondary structure on the residues (as observed in the PDB 3CMS) is reported (red: helix, blue: beta). Panel D: Change of MI upon P8-P4 fragment binding. Residues for which a decrease of MI is observed are colored in blue, while residues for which an increase of MI is observed are colored in red.

effect, allows the β -flap to deform and explore different paths for the conversion from the self-inhibited to the open state. The new path involves an intermediate state in the same kinetic basin as the self-inhibited state and, therefore, a higher rate of conversion between those two conformations. This intermediate state permits access to the κ -casein cleavage site on the chy-

mosin catalytic site.

Conclusions

The conformational change occurring in the allosteric activation of bovine chymosin has been observed by both regular MD and BEMD sim-

ulations. In agreement with previous proposals based on kinetic, mutagenesis and crystallographic experiments,^{6,7,16,19,21} the simulations show that the HPHPH sequence from the P8-P4 residues of bovine κ -casein initiates a conformational change in the side chain of Tyr77 and the β -hairpin region of bovine chymosin. The allosteric activation mechanism occurs via the following steps: (i) the P8-P4 κ -casein fragment binds with chymosin and disrupts the hydrogen bonding network that stabilises the self-inhibiting pose of Tyr77 fig. 5; (ii) the P8-P4 κ -casein peptide interacts with the short α -helix in residues 112-116 of chymosin, which both allows the β -hairpin flap in residues 72 to 84 of chymosin to twist, and also causes the side chain of Phe114 to vacate the pocket that is occupied by Tyr77 in the open conformation; (iii) as Phe114 moves, Tyr77 simultaneously changes conformation from self-inhibiting to open and is stabilised by a hydrogen bonding network below the β -hairpin flap fig. 4. Subtle variations in the simulation trajectories suggest that allosteric activation is possible by multiple related pathways, but these all go through the general steps described above, which were observed in their entirety in all of the relevant MD and BEMD simulations.

Acknowledgement We thank Prof Alessandro Laio from the International School for Advanced Studies (SISSA) in Trieste, Italy, for his advice regarding the BEMD simulations. D.S.P. and S.A. are grateful for use of the EPSRC funded ARCHIE-WeSt High Performance Computer (www.archie-west.ac.uk, EP-SRC Grant No. EP/K000586/1). D.S.P. thanks the University of Strathclyde for support through its Strategic Appointment and Investment Scheme. A.C. post-doc fellowship is supported by the Lundbek Foundation (2014-813). A.C., K.K.S and B.S were supported by the Danish Council for Independent Research | Technology and Production Sciences (FTP-0602-01778B). A.C., K.K.S and B.S are grateful for us of the “Grendel” hybrid CPU/GPU cluster at the Centre for Scientific Computing Aarhus (www.cscaa.dk).

Supporting Information Available: Ad-

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

References

- (1) Dunn, B. M. Structure and mechanism of the pepsin-like family of aspartic peptidases. *Chem. Rev.* **2002**, *102*, 4431–4458.
- (2) Davies, D. R. The structure and function of the aspartic proteinases. *Annu. Rev. Biophys. Biophys. Chem.* **1990**, *19*, 189–215.
- (3) Foltmann, B. The primary structure of calf chymosin. *J. Biol. Chem.* **1979**, *254*, 8447–8456.
- (4) Dairyco, Datum World Dairy Product Production. 2015; <http://www.dairyco.org.uk/market-information/processing-trade/dairy-product-production/world-dairy-product-production/#.VWhk7RHbL4g>, Accessed on 10th May, 2016.
- (5) Kappeler, S. R.; van den Brink, H. J.; Rahbek-Nielsen, H.; Farah, Z.; Puhan, Z.; Hansen, E. B.; Johansen, E. Characterization of recombinant camel chymosin reveals superior properties for the coagulation of bovine and camel milk. *Biochem. Biophys. Res. Comm.* **2006**, *342*, 647–654.
- (6) Visser, S.; van Rooijen, P. J.; Schattenkerk, C.; Kerling, K. E. T. Peptide substrates for chymosin (Rennin). Kinetic studies with peptides of different chain lengths including parts of the sequence 101-112 of bovine κ -casein. *Biochim. et Biophys. Acta (BBA) - Enzymol* **1976**, *438*, 265–272.
- (7) Visser, S.; van Rooijen, P. J.; Schattenkerk, C.; Kerling, K. E. T. Peptide substrates for chymosin (Rennin) Kinetic

- studies with bovine κ -casein-(103-108)-hexapeptide analogues. *Biochim. et Biophys. Acta (BBA) - Enzymol* **1977**, *481*, 171–176.
- (8) Plowman, J.; Creamer, L. Restrained molecular dynamics study of the interaction between bovine kappa-casein peptide 98-111 and bovine chymosin and porcine pepsin. *J. Dai. Res.* **1995**, *62*, 451.
- (9) Blundell, T. L.; Jenkins, J. A.; Sewell, B. T.; Pearl, L. H.; Cooper, J. B.; Tickle, I. J.; Veerapandian, B.; Wood, S. P. X-ray analyses of aspartic proteinases: the three-dimensional structure at 2.1 Å resolution of endothiapepsin. *J. Mol. Biol.* **1990**, *1*, 919–941.
- (10) Newman, M.; Safro, M.; Frazao, C.; Khan, G.; Zdanov, A.; Tickle, I. J.; Blundell, T. L.; Andreeva, N. X-ray analyses of aspartic proteinases IV: Structure and refinement at 2.2 Å resolution of bovine chymosin. *J. Mol. Biol.* **1991**, *2*, 1259–1309.
- (11) Pearl, L.; Blundell, T. The active site of aspartic proteinases. *FEBS Lett.* **1984**, *174*, 96–101.
- (12) Gilliland, G. L.; Winborne, E. L.; Nachman, J.; Wlodawer, A. The three-dimensional structure of recombinant bovine chymosin at 2.3 Å resolution. *Prot.: Struc., Func. Gen.* **1990**, *3*, 82–101.
- (13) Chitpintyol, S.; Crabbe, M. Chymosin and aspartic proteinases. *Food Chem.* **1998**, *61*, 395–418.
- (14) Newman, M.; Watson, F.; Roychowdhury, P.; Jones, H.; Badasso, M.; Cleasby, A.; Wood, S. P.; Tickle, I. J.; Blundell, T. L. X-ray analyses of aspartic proteinase: V structure and refinement at 2.0 Å resolution of the aspartic proteinase from *Mucor pusillus*. *J. Mol. Biol.* **1993**, 260–283.
- (15) Cooper, J. B.; Khan, G.; Taylor, G.; Tickle, I. J.; Blundell, T. L. X-ray analyses of aspartic proteinases II: Three-dimensional structure of the hexagonal crystal form of porcine pepsin at 2.3 Å. *J. Mol. Biol.* **1990**, *3*, 199–222.
- (16) Gustchina, E.; Rumsh, L.; Ginodman, L.; Majer, P.; Andreeva, N. Post X-ray crystallographic studies of chymosin: the existence of two structural forms and the regulation of activity by the interaction with the histidine-proline cluster of kappa-casein. *FEBS Lett.* **1996**, *379*, 60–62.
- (17) Dalgleish, D. G. Casein micelles as colloids: surface structures and stabilities. *J. Dairy Sci.* **1998**, *81*, 3013–3018.
- (18) Groves, M.; Dhanaraj, V.; Badasso, M.; Nugent, P.; Pitts, J.; Hoover, D.; Blundell, T. A 2.3 Å resolution structure of chymosin complexed with a reduced bond inhibitor shows that the active site beta-hairpin flap is rearranged when compared with the native crystal structure. *Prot. Eng. Des. Sel.* **1998**, *11*, 833–840.
- (19) Andreeva, N.; Dill, J.; Gilliland, G. L. Can enzymes adopt a self-inhibited form? Results of X-ray crystallographic studies of chymosin. *Biochem. Biophys. Res. Commun.* **1992**, *184*, 1074–1081.
- (20) Palmer, D. S.; Christensen, A. U.; Sørensen, J.; Celik, L.; Qvist, K. B.; Schiøtt, B. Bovine chymosin: A computational study of recognition and binding of bovine κ -casein. *Biochemistry* **2010**, *49*, 2563–2573.
- (21) Strop, P.; Sedlacek, J.; Stys, J.; Kaderabkova, Z.; Blaha, I.; Pavlickova, L.; Pohl, J.; Fabry, M.; Kostka, V.; Newman, M. et al. Engineering enzyme subsite specificity: preparation, kinetic characterization, and X-ray analysis at 2.0 Å resolution of Val111Phe site-mutated calf chymosin. *Biochem.* **1990**, *29*, 9863–9871.

- (22) Piana, S.; Laio, A. A bias-exchange approach to protein folding. *J. Phys. Chem. B* **2007**, *111*, 4553–4559.
- (23) Baftizadeh, F.; Cossio, P.; Pietrucci, F.; Laio, A. Protein folding and ligand-enzyme binding from bias-exchange metadynamics simulations. *Curr. Phys. Chem.* **2012**, *2*, 79–91.
- (24) Corbi-Verge, C.; Marinelli, F.; Zafra-Ruano, A.; Ruiz-Sanz, J.; Luque, I.; Faraldo-Gómez, J. D. Two-state dynamics of the SH3-SH2 tandem of Abl kinase and the allosteric role of the N-cap. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, E3372–E3380.
- (25) Marinelli, F.; Kuhlmann, S. I.; Grell, E.; Kunte, H.-J.; Ziegler, C.; Faraldo-Gómez, J. D. Evidence for an allosteric mechanism of substrate release from membrane-transporter accessory binding proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, E1285–E1292.
- (26) Case, D. A.; Darden, T. A.; Cheatham, T. E. I.; Simmerling, C.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Pearlman, D. A.; Crowley, M. et al. AMBER 9. 2006.
- (27) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010**, *78*, 1950–1958.
- (28) Sørensen, J.; Palmer, D. S.; Schiøtt, B. Hot-spot mapping of the interactions between chymosin and bovine κ -casein. *J. Agr. Food Chem.* **2013**, *61*, 7949–7959.
- (29) Sørensen, J.; Palmer, D. S.; Qvist, K. B.; Schiøtt, B. Initial stage of cheese production: A molecular modeling study of bovine and camel chymosin complexed with peptides from the chymosin-sensitive region of κ -casein. *J. Agr. Food Chem.* **2011**, *59*, 5636–5647.
- (30) Palmer, D. S.; Sørensen, J.; Schiøtt, B.; Fedorov, M. V. Solvent binding analysis and computational alanine scanning of the bovine chymosin-bovine κ -casein complex using molecular integral equation theory. *J. Chem. Theory Comput.* **2013**, *9*, 5706–5717.
- (31) Prasad, B. V.; Suguna, K. Role of water molecules in the structure and function of aspartic proteinases. *Acta Crystallogr. Sect. D: Biol. Crystallogr* **2002**, *58*, 250–259.
- (32) Veerapandian, B.; Cooper, J. B.; Sali, A.; Blundell, T. L.; Rosati, R. L.; Dominy, B. W.; Damon, D. B.; Hoover, D. J. Direct observation by X-ray analysis of the tetrahedral "intermediate" of aspartic proteinases. *Protein Sci.* **1992**, *1*, 322–328.
- (33) Piana, S.; Sebastiani, D.; Carloni, P.; Parrinello, M. Ab initio molecular dynamics-based assignment of the protonation state of pepstatin A/HIV-1 protease cleavage site. *J. Am. Chem. Soc.* **2001**, *123*, 8730–8737.
- (34) Piana, S.; Bucher, D.; Carloni, P.; Rothlisberger, U. Reaction mechanism of HIV-1 protease by hybrid CarParrinello/Classical MD simulations. *J. Phys. Chem. B* **2004**, *108*, 11139–11149.
- (35) Northrop, D. B. Follow the protons: A low-barrier hydrogen bond unifies the mechanisms of the aspartic proteases. *Acc. Chem. Res.* **2001**, *34*, 790–797.
- (36) Coates, L.; Erskine, P. T.; Wood, S. P.; Myles, D. A. A.; Cooper, J. B. A neutron Laue diffraction study of endothiapepsin: Implications for the Aspartic Proteinase Mechanism. *Biochemistry* **2001**, *40*, 13149–13157.
- (37) Bas, D. C.; Rogers, D. M.; Jensen, J. H. Very fast prediction and rationalization of pKa values for protein–ligand complexes. *Proteins: Structure, Function, and Bioinformatics* **2008**, *73*, 765–783.

- (38) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comp. Chem.* **2005**, *26*, 1781–1802.
- (39) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (40) Case, D. A.; Berryman, J. T.; Betz, R. M.; Cerutti, D. S.; Cheatham, T. E.; III, T. A. D.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W. et al. AMBER 2015. 2015.
- (41) Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev.* **1985**, *31*, 1695–1697.
- (42) Nose, S.; Klein, M. L. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* **1983**, *50*, 1055–1076.
- (43) Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* **1994**, *101*, 4177–4189.
- (44) Ewald, P. P. Die berechnung optischer und elektrostatischer gitterpotentiale. *Ann. Phys.* **1921**, *64*, 253–287.
- (45) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An $N * \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (46) York, D. M.; Wlodawer, A.; Pedersen, L. G.; Darden, T. A. Atomic-level accuracy in simulations of large protein crystals. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 8715–8718.
- (47) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comp. Phys.* **1977**, *23*, 327–341.
- (48) Weinbach, Y.; Elber, R. Revisiting and parallelizing SHAKE. *J. Comput. Phys.* **2005**, *209*, 193–206.
- (49) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A. et al. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comput. Phys. Commun.* **2009**, *180*, 1961–1972.
- (50) Biarnés, X.; Pietrucci, F.; Marinelli, F.; Laio, A. METAGUI. A VMD interface for analyzing metadynamics and molecular dynamics simulations. *Comput. Phys. Commun.* **2012**, *183*, 203–211.
- (51) Lange, O. F.; Grubmüller, H. Generalized correlation for biomolecular dynamics. *Proteins* **2006**, *62*, 1053–1061.
- (52) Kabsch, W.; Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577–2637.
- (53) Mahanti, M.; Bhakat, S.; Nilsson, U. J.; Söderhjelm, P. Flap dynamics in aspartic proteases: A computational perspective. *Chem. Biol. Drug Des.* **2016**, *88*, 159–177.
- (54) Hornak, V.; Okur, A.; Rizzo, R. C.; Simmerling, C. HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 915–920.

Graphical TOC Entry

