BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

Biochemistry **2010,** *49,* 2563–2573 2563
DOI: 10.1021/bi902193u

Bovine Chymosin: A Computational Study of Recognition and Binding of Bovine κ-Casein†

David S. Palmer,‡ Anders U. Christensen,‡ Jesper Sørensen,§ Leyla Celik,§,∥ Karsten Bruun Qvist,⊥ and Birgit Schiøtt*,§

‡*Department of Chemistry, and* §*The Center for Insoluble Protein Structures* (*inSPIN*)*, Department of Chemistry, and the Interdisciplinary Nanoscience Center, Aarhus University, Langelandsgade 140, 8000 Aarhus C, Denmark, and* ⊥*Cultures and Enzymes Division, Innovation, Chr. Hansen A/S, Bøge Allé 10-12, 2970 Hørsholm, Denmark.* ∥*Current address: Department of Chemistry, Yale University, New Haven, CT.*

Received December 22, 2009; Revised Manuscript Received February 11, 2010

ABSTRACT: Bovine chymosin is an aspartic protease that selectively cleaves the milk protein κ-casein. The enzyme is widely used to promote milk clotting in cheese manufacturing. We have developed models of residues *97*−*112* of bovine κ-casein complexed with bovine chymosin, using ligand docking, conformational search algorithms, and molecular dynamics simulations. In agreement with limited experimental evidence, the model suggests that the substrate binds in an extended conformation with charged residues on either side of the scissile bond playing an important role in stabilizing the binding pose. *Lys111* and *Lys112* are observed to bind to the N-terminal domain of chymosin displacing a conserved water molecule. A cluster of histidine and proline residues (*His98-Pro99-His100-Pro101-His102*) in κ-casein binds to the C-terminal domain of the protein, where a neighboring conserved arginine residue (*Arg97*) is found to be important for stabilizing the binding pose. The catalytic site (including the catalytic water molecule) is stable in the starting conformation of the previously proposed general acid/base catalytic mechanism for 18 ns of molecular dynamics simulations.

Bovine chymosin is a mammalian aspartic protease found in the fourth stomach of calves, where it aids digestion by selectively cleaving the milk protein κ-casein (*1*). Since it was first purified industrially in 1874, the enzyme has been sold to initiate milk clotting in cheese manufacturing. Although bovine chymosin remains in common use, it has recently been demonstrated that the camel variant of the enzyme has 70% higher clotting activity and only 20% of the unspecific protease activity for bovine milk, which has led to it being successfully marketed as an alternative to the bovine enzyme (*2*). By contrast, bovine chymosin has a very low catalytic rate for proteolysis of camel κ-casein (*2*). The two enzymes have high sequence identity (85%), but the difference in catalytic efficacy is not well understood on a molecular level due to a lack of structural information about the chymosin−κ-casein complexes. Currently, four X-ray crystal structures of apo- or inhibitor-bound bovine chymosin are available (*3*−*6*), but there are no structural coordinates for apo camel chymosin or κ-casein bound complexes of either species.

The active bovine enzyme is formed by proteolytic cleavage of a 42-residue N-terminal propeptide from the zymogen in the acidic environment of the stomach. A single point mutation at residue 244 in a loop on the surface of the protein distinguishes form A (Asp) from form B (Gly) (residues are numbered as per PDB[1] entry 1CMS, a convention commonly referred to as chymosin numbering) (*7*). Chymosin A has a 20% greater milk clotting ability, but chymosin B is preferred in industry because it has a longer shelf life (*8*). The term chymosin C has been used both for what appears to be a degradation product of chymosin A and for a separate genetic variant (*9*, *10*). In the following, unless otherwise stated, the term chymosin refers to the B form of bovine chymosin.

Chymosin is a globular protein comprising 323 amino acids. The secondary structure is 13% helical (9 helices, 44 residues) and 48% β-sheet (29 strands, 158 residues) (*11*), and these features are organized to give a pseudosymmetry along a single cleft in which the catalytic Asp34 and Asp216 residues reside. Unlike viral aspartic proteases such as HIV protease, chymosin is monomeric, with a 9% sequence identity between N- and C-terminal domains (*12*). The catalytic Asp34 and Asp216 residues occur in conserved Asp-Thr-Gly motifs, with the Asp side chains orientated toward each other, in an approximately planar geometry. They are stabilized by a network of hydrogen bonds, which includes two threonine interactions coined the "fireman's grip" (*7*, *12*, *13*). A water molecule is observed between the catalytic Asp residues in all crystal structures of apo chymosin. The protein contains three disulfide bridges (Cys47−Cys52, Cys207−Cys211, and Cys250−Cys283) and a *cis*-proline residue (Pro25) that is conserved in mucorpepsin, endothiapepsin, and porcine pepsin (*12*, *14*).

The substrate of chymosin, κ-casein, is a 169-residue protein that helps to solubilize α_{s1}-, α_{s2}-, and β-caseins in milk serum by promoting the formation of aggregates referred to as casein

†This work was supported by grants from the Villum Kahn Rasmussen Foundation, the BioSys Network, The Danish Council for Independent Research | Natural Sciences, and the Centre for Theory in Natural Science, Aarhus University. Computations were made possible through grants from the Lundbeck Foundation, the Novo Nordisk Foundation, the Carlsberg Foundation, and the Danish Center for Scientific Computing.

*To whom correspondence should be addressed. Phone: +4589423953. Fax: +4586196199. E-mail: birgit@chem.au.dk.

[1]Abbreviations: PLOP, protein local optimization program; MCMM, Monte Carlo multiple minimum; PDB, Protein Data Bank; RMSF, root mean square fluctuation.

Published on Web 02/15/2010

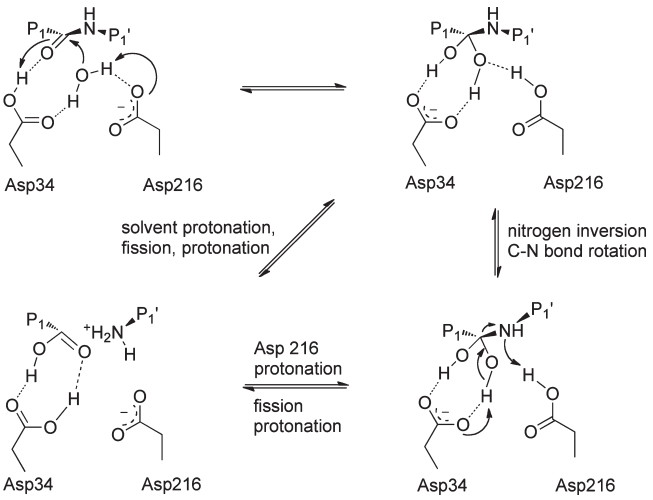
```
                P9      P1P1'   P7'
                |         ||    |
Cow:            RHPHPHLSFMAIPPKK    (97-112)
Buffalo:        RHPHPHLSFMAIPPKK    (97-112)
Goat:           RHPHPHLSFMAIPPKK    (97-112)
Camel:          RRPRPRPSFIAIPPKK    (89-104)
Pig:            RRPRPHASFIAIPPKK    (88-103)
```

FIGURE 1: The amino acid sequences of cow, buffalo, goat, camel, and pig κ-casein near the *Phe105-Met106* cleavage site. The residues on either side of the scissile bond are highlighted in red. Residues in camel and pig κ-casein that differ from the related residues in cow, buffalo, and goat κ-caseins are highlighted in green.

micelles. κ-Casein is found predominantly on the surface of casein micelles, which have raspberry-like heterogeneous structures with average diameters of ~200 nm (*15*). Chymosin destabilizes the casein micelles by cleaving κ-casein at the *Phe105*−*Met106* bond, which leads to milk clotting (κ-casein residues are given in italics in the text throughout). The amino acid sequences of κ-caseins from different species in the region of the cleavage site are given in Figure 1. By convention, a P*x* or P*x*′ nomenclature is used to denote κ-casein residues on either side of the cleavage site, e.g., *Ser104*, *Phe105*, *Met106*, and *Ala107* are referred to as P2, P1, P1′, and P2′, respectively. Similarly, the regions of chymosin that interact with the P2, P1, P1′, and P2′ residues are denoted S2, S1, S1′, and S2′ pockets, respectively. On the basis of the crystal structure of a chymosin−inhibitor complex (1CZI) and previous molecular modeling studies, κ-casein is thought to bind in an extended secondary structure (*16*, *17*). This is consistent with circular dichroism, solution NMR, and molecular modeling studies of unbound κ-casein, which show an extended structure in the region of the scissile bond (*18*, *19*). On the basis of geometric considerations and mutagenesis studies, it has been proposed that the P8−P7′ residues are located in the chymosin binding cleft during catalysis (*7*). Additionally, an arginine residue in the P9 position has been implicated in binding because it is conserved in bovine, camel, pig, buffalo, and goat chymosin (*20*). Accordingly, a variant of bovine κ-casein, in which the P9 position is occupied by a histidine, has been shown to be a poor substrate (*21*).

The mechanism of hydrolysis by aspartic proteases is widely thought to occur by nucleophilic attack on the peptide carbonyl carbon via a catalytic water molecule (*22*), but exact details of the mechanism remain contentious. On the basis of X-ray structural studies of an endothiapepsin−difluorostatone inhibitor complex, which was considered to be a transition state mimic, Veerapandian et al. proposed a mechanism in which the scissile bond carbonyl is protonated by Asp34 and concurrently undergoes nucleophilic attack by a water molecule, which is activated by the deprotonated Asp216 group (Scheme 1) (*23*). James et al. propose a similar mechanism based upon X-ray crystallographic studies of complexes between penicillopepsin and difluorstatine- or difluorostatone-containing peptides (*24*). This mechanism is consistent with the neutron diffraction studies (that were able to resolve the positions of the hydrogen atoms) carried out by Coates et al. on a hydroxyethylene-based inhibitor bound to endothiapepsin (EC 3.4.23.22), which suggested that Asp34 is deprotonated and Asp216 is protonated on the outer oxygen atom in the transition state (*25*). Recently, in the light of kinetic data and the *ab initio* studies of Piana and Carloni (*26*), Northrop has suggested a reaction mechanism that involves the formation of a low-barrier hydrogen bond between Asp34 and Asp216 (*27*). The mechanism includes a rearrangement of protons around a

Scheme 1: Reaction Mechanism Proposed by Veerapandian et al. (*23*)



10-membered cyclic intermediate that is proposed to occur by quantum tunneling (a thorough discussion of the different mechanistic proposals is presented in an excellent review article by Dunn) (*22*).

In one crystallographic structure of chymosin (*4*), a short β-hairpin loop above the catalytic Asp residues is resolved into two conformations, one of which has Tyr77 positioned in the S1 pocket. The observation that chymosin can adopt a conformation that occludes the S1 pocket is thought to indicate that the apo form of the enzyme is self-inhibited (*28*).

The crystal structures of chymosin and related aspartic proteases contain a large number of structural water molecules. From an analysis of 10 of these enzymes, 17 water molecules have been identified as conserved (*29*), including the catalytic water molecule, five molecules in the S1′−S7′ half of the binding cleft, and one in the S9−S1 end. A water molecule close to Ser35 is thought to take part in a hydrogen bond chain that helps to stabilize Asp34 (*28*). The same water molecule may also help to stabilize the open form of chymosin through a hydrogen bond to Tyr77 (*28*).

The importance of residues around the κ-casein scissile bond has been elucidated by mutagenesis studies. Human and porcine κ-casein (*Phe105-Ile106*) and rat and mouse κ-casein (*Phe105-Leu106*) are cleaved by bovine chymosin, which suggests that the *Phe105-Met106* bond in κ-casein is not essential to enzyme action. Indeed, a *Met106Phe* mutated bovine κ-casein is hydrolyzed 1.8 times faster than the wild-type protein by bovine chymosin (*30*). Analogues of κ-casein, in which the side chain of *Phe105* is replaced by Phe(NO_2) or cyclohexylamine, are also hydrolyzed by bovine chymosin, albeit at a reduced rate (*30*). The residues neighboring the cleavage site are known to influence the reaction rate. For instance, k_{cat} for cleavage of the P3−P3′ fragment is several hundred-fold higher than for P1−P1′ (*31*). Mutagenesis studies of the P3 position (which is a leucine in bovine κ-casein) indicate that catalysis is promoted by aliphatic residues, tolerated by hydrophilic residues, and disfavored by proline or positively charged residues (*32*). The fact that proline in the P3 position disfavors cleavage may provide a clue as to why bovine chymosin cannot cleave camel κ-casein, which has a proline residue at this position.

Of particular interest is the observation that the rate of hydrolysis is 20-fold higher for the P8−P4′ fragment of κ-casein

than for P3−P4′. Furthermore, when chymosin is incubated with a *His-Pro-His-Pro-His* pentapeptide (P8−P4 of κ-casein) prior to reaction, the catalytic rate for hydrolysis of P3−P4′ is observed to increase nearly 200-fold (*17*, *33*). It is suggested that the *His-Pro* cluster acts as an allosteric activator to convert self-inhibited chymosin to its open form. If correct, this might provide some clue as to why bovine chymosin is not an effective catalyst for camel κ-casein, which has *Arg-Pro-Arg-Pro-Arg* residues rather than *His-Pro-His-Pro-His* residues at the P8−P4 position. In the *His-Pro* cluster, both *His* and *Pro* residues are important for catalysis, and *His102Lys* mutations are particularly disfavored (*34*). Mutagenesis studies show that *Lys111* and *Lys112* promote catalysis but that the effect is only observed for *Lys112*, if *Lys111* is also present (*34*).

Few molecular modeling studies of chymosin−κ-casein complexes have previously been reported. Plowman et al. studied the *His98-Lys111* fragment of κ-casein in complex with both bovine chymosin and porcine pepsin by short molecular dynamics simulations using a distant-dependent dielectric model for solvent (*17*). The initial position of the ligand in the binding site cleft was obtained by computationally mutating a pepstatin inhibitor, which had been introduced by superposition of bovine chymosin and a rhizopuspepsin−pepstatin inhibitor complex. Harmonic restraints were applied to five atoms of the ligand during all simulations. The authors suggested that a *cis*-peptide bond between *His98-Pro99* was important to allow a favorable interaction between *His98* and Asp247 but later disputed this view based upon molecular dynamics of a longer peptide (*35*). Favorable electrostatic contacts were reported between *His102*···Glu245, *His100*···Asp279, and *Lys111*···Glu133 residues of κ-casein and chymosin, respectively. The simulations were carried out in implicit solvent, and no comment was made about the importance of catalytic or conserved waters for binding. Kashparov et al. used explicit solvent molecular dynamics of apo chymosin to suggest that the self-inhibited form of chymosin (observed in 3CMS A) is accessible in solution (*36*).

In this paper, we present unrestrained molecular dynamics simulation of bovine chymosin complexed with the P9−P7′ residues of bovine κ-casein. The proposed model is consistent with experimental evidence obtained from neutron and X-ray crystallographic studies and site-directed mutagenesis studies. In addition, the work helps to explain interesting features of chymosin binding: the importance of specific chymosin−κ-casein interactions in recognition; the dynamic properties of conserved water molecules in the binding cleft; the significance of Tyr77 and a conserved water molecule in chymosin self-inhibition.

MATERIALS AND METHODS

Overview. The P8−P7′ (and P9−P7′) fragment of bovine κ-casein has too many rotatable bonds to be docked directly by existing molecular docking software. To circumvent this problem, the docking procedure was carried out in two steps: (i) short central fragments of κ-casein were docked into open, apo chymosin (*4*); (ii) the best docked pose was used as a starting point from which to grow the P8−P7′ ligand in the binding site. The two-step docking strategy is acceptable for this complex because sufficient experimental evidence is available, from both mutagenesis studies and chymosin−inhibitor crystal structures, to accurately locate the central residues of κ-casein in the binding site. The second step of the docking procedure was carried out in duplicate using both the protein local optimization program (*37*, *38*) and Monte Carlo conformational search methodologies (*38*). Each minimum energy P8−P7′ pose was then used as a template to create P9−P7′ poses by adding the P9 residue and sampling the conformation of both P8 and P9 residues (a similar approach has been used in earlier chymosin modeling studies) (*17*). The stability of both the P8−P7′ and P9−P7′ minimum energy poses obtained from these methods was then studied with molecular dynamics simulations in the isothermal−isobaric (NPT) ensemble.

To provide a comparison to the dynamics of liganded chymosin and to investigate the flexibility of the β-hairpin flap region near Tyr77, additional molecular dynamics simulations were run of the open- and self-inhibited structures of apo chymosin (*4*). Furthermore, for comparison, additional simulations were performed of an inhibited structure with and without bound inhibitor (*6*).

Protein Preparation. Three X-ray crystallographic structures of apo bovine chymosin are available in the PDB (*39*): two (1CMS and 4CMS) are self-inhibited by rotation of Tyr77 into the S1 pocket (*3*, *5*); in the third (3CMS) both a self-inhibited form (A) and open form (B) are observed (*4*). 3CMS has the best crystallographic resolution (2.0 Å, *R*-factor 19.5%), and the open form was used in this work. We replaced three missing residues (Asn291-His292-Ser293, which occur in a loop region outside the binding site) by copying them from 1CMS and then minimizing residues Asn291-Lys295 (for 50 iterations of steepest descent, followed by 500 iterations of conjugate gradient) using the OPLS-2005 force field (*40*) in Macromodel 9.6 (*38*), with aqueous solvent modeled implicitly by the generalized Born/surface area (GB/SA) method (*41*), while holding the rest of the protein fixed. Additionally, the Val111Phe mutation that is present in the 3CMS chymosin structure was reversed. Disulfide bonds were introduced between Cys47−Cys52, Cys207−Cys211, and Cys250−Cys283, and hydrogen atoms were added and bond orders assigned using the Protein Preparation tool in Maestro 8.5 (*38*).

Conserved Water Molecules. Of the seven conserved waters in the binding cleft (*29*), the catalytic water and the five molecules in the S1′−S7′ end of the pocket are all present in the 3CMS structure of chymosin (Wat1125, Wat1113, Wat1007, Wat1006, Wat1005, and Wat1116). At the S8−S1 end, however, the 3CMS structure contains few crystallographic waters, and the conserved water molecule is missing. The conserved water molecule is observed in the 1CMS and 4CMS structures of chymosin, and therefore, we have incorporated it into our model (Wat645 from the 1CMS structure). This water molecule occurs in the base of the binding cleft approximately 7.5 Å from the catalytic water, in a position that would allow hydrogen bonding to Phe33(O) and Thr217(O). The remaining nonconserved water molecules in the binding cleft were removed prior to docking and growth of the κ-casein. However, water molecules whose original positions were more than 2.5 Å from the newly docked ligand were reinstated before the molecular dynamics simulations. All water molecules outside the binding cleft were retained. In this work, hydration sites were identified from a statistical analysis (*29*), but they may also be predicted by computational methods (*42*−*44*).

Ionizable Residues in Chymosin. Based upon the mechanism proposed by Veerapandian et al. (*23*), Asp34 was modeled as neutral with the proton on the outer oxygen atom (Oδ2), while Asp216 was modeled as deprotonated. Protonation states for noncatalytic residues were assigned using PROPKA2.0 (*45*) assuming a pH of 6.5. All Glu, Lys, and Arg residues were modeled as charged and all Cys and Tyr residues as neutral.

His76, His146, and His292 were modeled as charged histidinium ions, His55 was modeled as the ε-tautomer, and His181 was modeled as the δ-tautomer, based on an inspection of their local protein environments. The histidine residues are not deemed to be relevant to the specific conformation of the binding site. Hydrogen bonds were optimized by exhaustive sampling using the Protein Preparation Wizard in Maestro 8.5 before a restrained minimization was carried out to an rmsd of 0.3 Å (*38*).

Peptide Preparation for Docking. Symmetric (P1−P1′, P2−P2′, and P3−P3′) and unsymmetric (P2−P1′, P3−P1′, P1−P3′, and P1−P2′) fragments of bovine κ-casein were built using Maestro 8.5 (*38*). The P3−P3′ residues were modeled as neutral. P8−P7′ fragments of κ-casein were built from the best docked poses using Maestro 8.5 (*38*). Initially, *His98*, *His100*, and *His102* were modeled as charged histidinium ions and *Lys111* and *Lys112* as protonated on the side-chain nitrogen atom. Later it was decided to also include neutral tautomers of His102 (see Results below). All fragments of the κ-casein peptide were capped at the N- and C-termini with *N*-acetyl- and *N*-methylamide groups, respectively, in order to more appropriately account for the full-length peptide.

Docking of the Central Fragments. The central fragments of bovine κ-casein were docked into open chymosin (3CMS B) by Glide 5.0, using the Standard Precision (SP) GlideScore and a docking grid centered on the catalytic Asp residues (*38*, *46*). The poses with best GlideScore were then refined using the Extra Precision (XP) GlideScore function (*47*). Results for all docking simulations are provided in the Supporting Information.

Monte Carlo Conformational Searches. The best docked central fragment of κ-casein (in this case a P2−P2′ peptide; see the Results section) was extended in the binding site to give the P8−P7′ residues of κ-casein. Monte Carlo multiple-minima conformational searches (MCMM) were performed for each of the two ends of the peptide in their entirety, e.g., P8−P2′ and P2−P7′, in which P8−P2 and P2′−P7′ residues were permitted to change conformation, respectively. In each calculation, chymosin, the catalytic water, and either P1−P2′ or P2−P1′, as appropriate, were defined as frozen atoms, and a harmonic potential of 100 kJ mol^{-1} $Å^{-2}$ was applied to constrain the atoms of the amide group that lay on the boundary between the flexible and frozen atoms. Conformational searches were carried out in Macromodel 9.6 (*38*) using 10000 iterations of the MCMM procedure, in which each structure was minimized by the Polak−Ribiere conjugate gradient method to a gradient of 0.05 kJ mol^{-1} $Å^{-1}$, using the OPLS-2005 force field and a distance-dependent dielectric of $\varepsilon = 4.0$. All results within 1000 kJ mol^{-1} of the global minimum were written as output and clustered in Maestro 8.5 to check that a reasonable sampling of conformational space had occurred. The progress of the algorithm was monitored by plotting the potential energy of the best conformer against completed iterations of the algorithm until convergence.

Protein Local Optimization Program. PLOP is available through the Schrodinger suite of programs (*38*). A set of plausible ligand conformations are generated from a rotamer library, which are then clustered to give representative structures that are refined by side-chain optimization and energy minimization (*37*). The PLOP procedure was carried out in two parts. First, the P2−P2′ docked fragment of κ-casein was extended to P2−P7′ by carrying out a PLOP search of the P2′−P7′ residues. Second, the minimum energy P2−P7′ pose was extended to give the full P8−P7′ fragment using a similar PLOP search of the P8−P2 fragment of the peptide. The complete procedure was repeated twice, first keeping the chymosin structure rigid and second permitting chymosin side chains within 7.5 Å of the ligand to move freely. The minimum energy P8−P7′ pose for each protonation state of *His102* was finally used as a starting point to grow the P9−P7′ pose by a PLOP search in which only the conformations of the P9 and P8 residues were sampled.

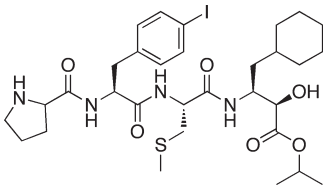


FIGURE 2: The chemical structure of the reduced bond inhibitor CP-113972, in which the hydroxyl group acts as a mimic of the peptide moiety in the transition state.

Parameterization of 1CZI Inhibitor. The parameters for the inhibitor CP-113972 (proline−iodophenylalanine−*S*-methylcysteine−cyclohexylnorstatine, which is a P4−P1′ mimic; see Figure 2) in the PDB entry code 1CZI were developed using antechamber (*48*) following tutorial B4 on the AMBER Web site by Ross Walker (*49*). The general AMBER force field (GAFF) (*50*) was used for atom typing, and AM1-BCC charges were calculated (*51*). A simulation of the ligand alone in explicit water was performed to check the stability of the molecule.

Molecular Dynamics Simulations. Protein−ligand complexes were solvated using the XLEAP module in AMBER 9 (*52*) to add TIP3P (*53*) water molecules within a distance of 10 Å around the protein in a periodic box. Approximately 31350 water molecules were added. The systems were neutralized and brought to an ionic strength of 0.07 mol dm^{-1} by adding sodium and chloride ions, as required. All simulations were run in NAMD (*54*) using the AMBER FF03 force field parameters developed by Duan and co-workers (*55*, *56*).

The solvated structures were relaxed by conjugate gradient energy minimization in four steps of 5000 iterations each. In steps 1−3, the whole protein, the protein backbone, and the α-carbon atoms, respectively, were held fixed. In the fourth step, all constraints were removed. The system was heated to 300 K by performing MD simulations in the NVT ensemble for 10 ps with the α-carbons atoms of the protein held fixed. To complete the relaxation and heating phase, a 10 ps MD simulation was run with the α-carbons held fixed, followed by a 10 ps simulation with no constraints.

Equilibration and production simulations were performed in the isothermal−isobaric (NPT) ensemble at 300 K and 1 atm. The Nosé−Hoover Langevin piston pressure control was used to regulate the pressure, with the piston target set to 1.01325 bar, the piston period at 200 fs, the piston decay set at 100 fs, and the piston temperature at 300 K (*57*−*59*). 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, and all electrostatic interactions were calculated using the particle mesh Ewald (PME) method (*60*−*62*). For van der Waals' interactions, a cutoff of 10 Å was set, using a switching distance of 9 Å. The pair list was updated every 20 steps including pairs of atoms within a distance of 11 Å. All of the hydrogen to heteroatom bond distances were held fixed using the SHAKE algorithm (*63*, *64*). The equations of motion were integrated every 2 fs using the velocity verlet algorithm, and snapshots were stored every 2 ps.

Complexes were equilibrated for 2 ns prior to 18 ns of production dynamics.

RESULTS

Docking Central Fragments of κ-Casein. Symmetric P1−P1′, P2−P2′, and P3−P3′ and unsymmetric P2−P1′, P3−P1′, P1−P3′, and P1−P2′ fragments of κ-casein were docked into the open form of chymosin. For each calculation, the quality of the docked poses was judged based upon the SP and XP GlideScores and the conformation of the catalytic site and peptide side chains with respect to existing experimental evidence (i.e., neutron and X-ray crystallographic studies of inhibitor-bound aspartic proteases and mutagenesis data). The P2−P2′ docked poses were found to be the most satisfactory, and therefore, in the interest of brevity, only these results will be discussed here. The results of all calculations are provided in the Supporting Information.

All 10 poses of P2−P2′ κ-casein with lowest GlideScore XP values had the *Phe105* side chain placed in the S1 pocket, which is an indication that the end-to-end alignment of the substrate in the binding pocket is correct. One pose was inconsistent with the proposed catalytic mechanisms and was discarded (pose eight when ranked by GlideScore XP, for which the distance between the carbonyl oxygen and the catalytic aspartates was greater than 8 Å). The remaining nine poses were tightly clustered with heavy-atom rmsds compared to the best scoring pose ranging from 1.5 to 2.0 Å. In each pose, the carbonyl oxygen was orientated toward the carboxyl hydrogen of Asp34, with distances ranging from 2.8 to 3.2 Å. The carbon atom of the peptide carbonyl group and the oxygen atom of the catalytic water were separated by distances ranging from 3.0 to 3.2 Å, while the angle at the carbonyl carbon subtended by the oxygen atoms of the catalytic water and the peptide carbonyl ranged from 69.5° to 71.5° for all poses. Although these angles are low in comparison to the Bürgi−Dunitz angle (*65*, *66*), these were later observed to relax during molecular dynamics simulations. The overall alignment of the ligand agrees very well with the proposed mechanism (*23*). The P2−P2′ pose with lowest GlideScore XP value was selected as the starting point to grow the full P8−P7′ peptide. The five best docked P2−P2′ poses are illustrated in Figure 3, showing a nearly perfect alignment of the central two residues.

Protonation States of P8−P7′ Residues. The P8−P7′ fragment of κ-casein contains five titratable residues (*His98*, *His100*, *His102*, *Lys111*, and *Lys112*). Initially, all histidines were modeled as the charged histidinium ions and all lysines as the charged Lys(NH_3^+) ions. Once each distinct P8−P7′ fragment had been grown by PLOP and MCMM approaches, the closest contacts of each histidine residue in κ-casein were inspected. In all of the minimum energy poses, a close contact was observed between *His102* and Lys221 of chymosin, which we expect would disfavor protonation of this particular histidine residue. Furthermore, when the pK_a's of these residues in each complex were predicted with PROPKA 2.0 (*45*), the pK_a of *His102* was consistently shifted down by 1−2 units from the model pK_a of 6.5, whereas the pK_a's of *His98* and *His100* were almost unchanged. For this reason, the modeling procedure was repeated using the neutral δ- and ε-tautomers of *His102*, and the *His102* histidinium poses were disregarded. Interestingly, Lys221 is replaced by a valine residue in camel chymosin, which may be one of the causes of the observed differences in catalytic efficacy (the camel variant of the enzyme has 70% higher clotting activity and only 20% of the unspecific protease activity for bovine milk as compared to the bovine enzyme).

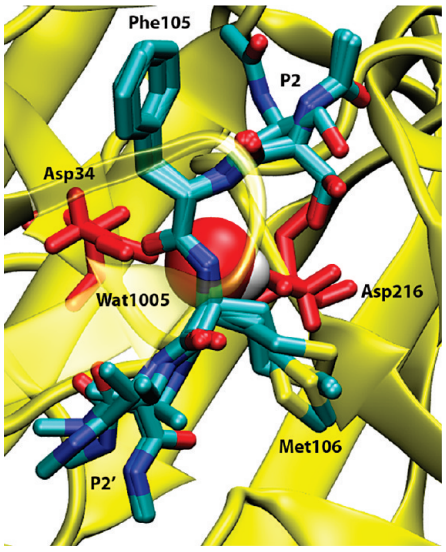


FIGURE 3: The five best docked poses of P2−P2′ κ-casein in chymosin. Asp34 (left) and Asp216 (right) of chymosin are shown in red on either side of the catalytic water molecule (vdW sphere representation). The P1 and P1′ backbone atoms are well aligned in all poses.

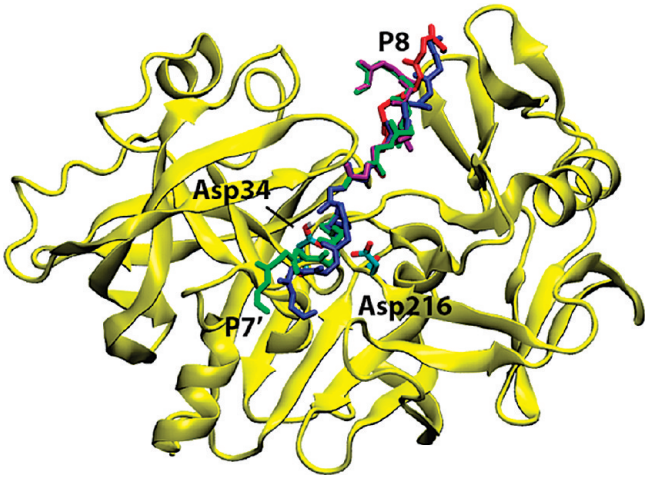


FIGURE 4: The P8−P7′ minimum energy poses of κ-casein bound to chymosin. The best poses of P8−P7′ κ-casein developed by PLOP are shown in blue and red for the δ-tautomer and ε-tautomers of *His102*, respectively. The minimum energy poses of P8−P7′ κ-casein developed by the MCMM method are shown in green and purple for the δ-tautomer and ε-tautomers of *His102*, respectively.

Building P9−P7′ Fragments with PLOP. The lowest energy P8−P7′ poses developed by the two-step PLOP calculation are illustrated in Figure 4. Both poses show an extended conformation in the binding cleft, with the P8−P1 end of the peptide making contacts with the C-terminal domain of chymosin and the P1′−P7′ fragment interacting with the N-terminal domain. The lowest energy P8−P7′ pose of both the δ- and ε-tautomer of *His102* was extended to give P9−P7′ fragments, using PLOP to resample the conformation of the P8 (*His98*) residue and the additional P9 (*Arg97*) residue.

Building P8−P7′ Poses with MCMM. In all minimum energy MCMM poses, a *cis*-peptide bond was observed between *His98* and *Pro99*, causing *His98* and the capping group to point into the binding cleft back toward the catalytic center. We consider that this was an artifact of the method caused by the

Table 1: Overview of All Molecular Dynamics Simulations[a]

		chymosin		
name	time (ns)	open/self-inhibited	apo/holo	ligand
A1	18	open	holo	P9−P7′ (*δ-His102*)
A2	18	open	holo	P9−P7′ (*ε-His102*)
B1	18	open	holo	P8−P7′ (*δ-His102*)
B2	18	open	holo	P8−P7′ (*ε-His102*)
C1	48	open	apo	n/a
C2	48	self-inhibited	apo	n/a
D1	18	open	holo	CP-113972

[a]The ligands in simulations A1, A2, B1, and B2 are peptide fragments of κ-casein; the P*x*−P*x*′ nomenclature refers to the included residues. Time is given in nanoseconds of preequilibrated production dynamics.

rigid-protein model used in the MCMM conformational search, which means that the side chains of the protein cannot relax around the substrate during the simulations; in particular, the side chains of Tyr11, Leu12, and Phe282 in the 3CMS crystal structure of apo chymosin occlude the region that might be occupied by P8 and P7 residues of κ-casein. For this reason, we did not pursue any further work with these structures. We note, however, that despite the *His98*−*Pro99 cis*-peptide bond, the MCMM poses have the same overall characteristics as the PLOP poses: the full peptide binds in an extended form, with the P8−P1 end contacting the C-terminal domain of chymosin and the P1′−P7′ contacting the N-terminal domain.

During the docking studies and PLOP/MCMM calculations the protein backbone was treated as rigid. Although some computational methods to include protein flexibility in either docking (*67*) or conformational searches (*38*) do exist, these were not applied here because analyses of the apo- and inhibitor-bound complexes of chymosin do not reveal large backbone conformational changes and because the docked poses were in good agreement with experimental data (as discussed in the Results section).

Molecular Dynamics Simulations. In total, seven systems were simulated by molecular dynamics (Table 1). The four bovine chymosin−bovine κ-casein complexes, generated by molecular docking calculations and PLOP (A1, A2, B1, and B2), were simulated for 18 ns to assess their stability. The mobility of a β-hairpin flap near the catalytic site was investigated by two 48 ns simulations initiated from the open (C1) and self-inhibited (C2) structures of apo chymosin, respectively. A complex of chymosin with a reduced-bond inhibitor (D1) was simulated as a general test of the selected methods.

In each of the simulations of the bovine chymosin−κ-casein complexes, κ-casein remained tightly bound to chymosin for 18 ns. The protein rmsd (calculated for backbone protein heteroatoms) fluctuated about a mean of 1.7 Å for simulations A1, A2, and B2 and around 1.9 Å for simulation B1, which indicates that the proteins are stable in these systems. Similarly, the protein RMSF values were below 2 Å for all residues in all four simulations, where RMSF is calculated as the root mean square fluctuation of the non-hydrogen atoms of each residue over the full trajectory. Asp34, Asp216, and Tyr77 were observed to have particularly small fluctuations, with values about 0.4, 0.5, and 0.5 Å, respectively.

The residues of κ-casein were observed to be stable, with RMSF values for all heteroatoms falling below 1.5 Å, in all simulations except B1. Here, it was observed that *His98* did not find a close contact with chymosin but rather protruded into the binding cleft, where it was partially solvated, resulting in a larger RMSF value of 2.5 Å. We attribute this observation to the absence of the conserved *Arg97* residue in P8−P7′ κ-casein, which makes stable interactions with the C-terminal domain of chymosin in both simulations A1 and A2. On the basis of this observation and the fact that *Arg97* is conserved in bovine, camel, pig, buffalo, and goat κ-casein, we suggest that the P9−P7′ poses (simulations A1 and A2) are more realistic models of the complex than the P8−P7′ poses. This observation is also supported by experimental evidence that shows that bovine milk containing a variant of κ-casein, in which the P9 position is occupied by a histidine, is a poor substrate for chymosin proteolysis (*21*). In simulations A1 and A2, essentially identical interactions are observed between the protein and the ligand, but in the latter the catalytic water was observed to leave the catalytic site after approximately 16 ns of production dynamics without being replaced. For this reason, all further references to our molecular dynamics simulations during the analysis will be to simulation A1, unless otherwise stated.

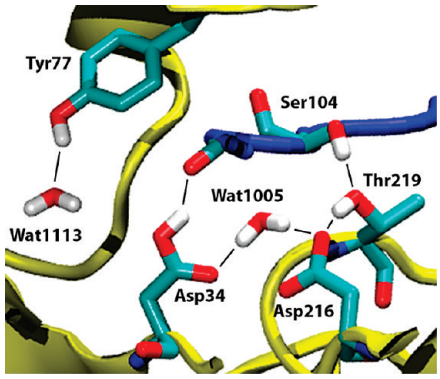


FIGURE 5: The conformation of the catalytic site after 17 ns of unrestrained molecular dynamics simulation (A1). The blue tube represents the backbone of κ-casein residues from the cleavage site toward the P9 residue (the P7′−P1′ residues are omitted). The N-terminal domain of chymosin is on the left-hand side of the figure.

Cleavage Site. Veerapandian et al. and James et al. have proposed a catalytic mechanism in which the scissile carbonyl bond of κ-casein is protonated by Asp34 and concurrently undergoes nucleophilic attack by a water molecule, which is activated by the deprotonated Asp216 group (Scheme 1). In the simulated complex, the conformation of the active site comprising Asp34, Asp216, catalytic water, and κ-casein scissile bond was observed to be stable and in accordance with the proposed mechanism throughout the simulation (Figure 5). The average distance and angle of nucleophilic attack from the catalytic water to the *Phe105* carbonyl fluctuated around 3.11 ± 0.32 Å (median = 3.04 Å) and 88.7 ± 9.9° (median = 88.6°), respectively, with angles within 10° of the Bürgi−Dunitz angle sampled for 4.3 ns. The catalytic water molecule donated hydrogen bonds to Asp34(Oδ1) and Asp216(Oδ2) for the majority of the simulation, but a second orientation in which it accepted a hydrogen bond from Asp34(Oδ2), while donating one to Asp216(Oδ1), was also observed for a small fraction of the simulation. Although the latter orientation is in accordance with a proposed mechanism for catalysis in endothiapepsin (*68*), the distance between Thr219(Oγ) and the oxygen of the catalytic water fluctuated around 4.06 ± 0.55 Å (median = 4.01 Å) in the simulation, which is slightly too long for the hydrogen bond that is paramount to that mechanism. Thr219(Oγ) was observed to donate a very stable

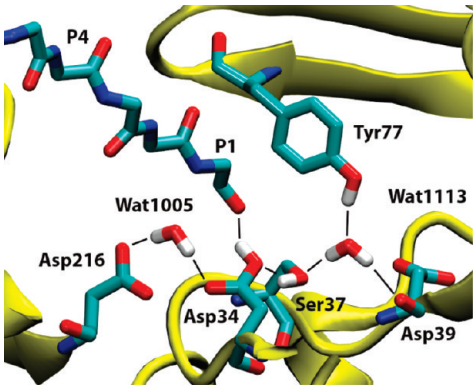


FIGURE 6: The catalytic site after 17 ns of molecular dynamics simulation (A1) showing the residues in a hydrogen bond chain from Tyr77(OH) (and Asp39(O)) to a conserved water molecule to Ser37-(Oγ) to Asp34(Oδ2). The backbone of residues P4−P1 of κ-casein are illustrated. The N-terminal domain of chymosin is on the right-hand side of the diagram.

hydrogen bond to Asp216(Oδ2), which supports the hypothesis that Thr219 is important for catalysis (Figure 5).

In particular, Thr219 probably helps to position the Asp216 side chain, which was observed to rotate less than Asp34 during the simulation. The Thr219(Oγ) side chain itself was stabilized by a hydrogen bond from *Ser104*(Oγ). Asp34(Oδ2) formed a predominantly stable hydrogen bond donated to *Phe105*(*O*) of κ-casein with an average heavy-atom distance of 3.19 ± 0.61 Å (median = 2.91 Å). The side chain of Asp34 is stabilized by a hydrogen bond chain connecting Tyr77(OH) and Asp39(O) with Wat1113, Ser37(Oγ), and Asp34(Oδ2) (Figure 6) (*28*). Small fluctuations were observed in these hydrogen bonds during the simulation, but the complete chain was present for more than 75% of the simulation, which supports the hypothesis that this conserved water is important for catalysis.

Fireman's Grip. Thr217(N) and Thr35(N) donate hydrogen bonds to Thr35(Oγ) and Thr217(Oγ), respectively, in what has been termed the firemans's grip (*13*). Additionally, Thr35(Oγ) and Thr217(Oγ) are stabilized by hydrogen bonds donated to Leu215(O) and Phe33(O), respectively. All four of these interactions were found to be present in the simulated complex but with relatively long hydrogen bonds (both the mean and median distance between heavy-atom donor and acceptor atoms were 3.2 ± 0.3 Å for all four hydrogen bonds).

P1−P1′ Residues. During the molecular dynamics simulation, the side chain of κ-casein *Phe105* pointed into the S1 pocket, where it made an edge-to-face stacking interaction with Tyr77 of chymosin and a parallel displaced face-to-face π-stacking interaction with Phe114. The pocket is hydrophobic in nature, which is consistent with the experimental observation that analogues of κ-casein, in which the side chain of *Phe105* is replaced by Phe(NO_2) or cyclohexylamine, are hydrolyzed by bovine chymosin at a much reduced rate (*30*). The side chain of *Met106* is not involved in any specific contacts but points out into the binding cleft between Tyr190, Ile297, and Ile214. The S1′ pocket is large enough to accommodate an aromatic side chain, and the nature of the pocket might explain why a *Met106Phe* mutant of bovine κ-casein has been hydrolyzed 1.8 times faster than the wild-type protein by bovine chymosin (*30*).

On either side of the catalytic water, backbone NH groups of Gly36 and Gly218 were observed to be close to Asp34(Oδ1) and Asp216(Oδ1), respectively. From consideration of the X-ray structure, it was anticipated that hydrogen bonds between these groups would be important for stabilizing the catalytic site, but both were found to be poorly aligned for much of the simulation.

β-Hairpin Flap. The β-hairpin flap that covers the catalytic site is involved in several interactions with bound κ-casein. Long hydrogen bonds or electrostatic interactions were observed between *Ser104*(*N*)···Thr79(Oγ), Thr79(N)···*Ser104*(*O*), and Gly78(N)···*Met106*(*O*). Following a temporary rearrangement of the flap, Thr79(Oγ) was also observed to act as a hydrogen bond donor to Ser332(O). The β-hairpin flap remained in the open form, with Tyr77 hydrogen bonded to the conserved water, for the full simulation, which supports the idea that a hydrogen bond chain from Tyr77 to Asp34 is important for catalysis (*28*).

P3−P3′ Residues. The dominant interactions of the P3, P2, P2′, and P3′ residues were observed to involve hydrogen bonds to the κ-casein backbone. *Leu103*(*O*) interacts with Ser220(NH) and *Ser104*(*N*) with Thr79(Oγ). The orientation of the Asp216 side chain is predominantly maintained by the conserved hydrogen bond between Thr219(Oγ) and Asp216(Oδ2) (see Figure 5). *Ser104*(Oγ) helps to stabilize Thr219 by donating a hydrogen bond to Thr219(Oγ), which might explain why serine is conserved in the P2 position of bovine, buffalo, pig, camel, and porcine κ-caseins. *Ala107* and *Ile108* made two and one backbone hydrogen bonds, respectively, between *Ala107*(*N*)···Gly36(O) and Tyr190(Oγ)···*Ala107*(*O*) and between *Ile108*(*N*)···His76-(O). The side chain of *Leu103* points into a gap between the backbone of Gln15 and the side chains of Phe119 in chymosin and *Phe105* in κ-casein. The orientation of the P3 side chain into this hydrophobic pocket may be one reason why mutagenesis studies have shown that catalysis is promoted by aliphatic residues, tolerated by hydrophilic residues, and disfavored by proline or positively charged residues (*32*). Interestingly, in camel κ-casein the P3 position is occupied by proline, which may help to explain why it is not cleaved by bovine chymosin.

P9−P4 κ-Casein Residues. *His98* forms a hydrogen bond to the side-chain oxygens of Asp279 through its Nδ proton and, for some of the simulations, to the side-chain oxygen of Tyr243 through the Nε proton. The carbonyl of *Pro99* lies close to the charged imidazolium ring system of *His98* and experiences some form of electrostatic effect. The Nε proton forms intermittent hydrogen bonds to Leu12(O), but these interactions are periodically lost when Leu12(O) accepts a hydrogen bond from Ser14(Oγ). The side chain of *His100* forms a stable electrostatic interaction with Asp13, while *His100*(*NH*) is hydrogen bonded to Thr284(Oγ). It has previously been demonstrated that *His102Lys* mutations are particularly disfavored (*34*), which is understandable from our simulated poses, in which *His102* makes a close contact with the aliphatic portion of the side chain of Lys221, while the $-NH_3^+$ headgroup is forced upward into a partially solvated pocket near Glu245. On the basis of steric and electrostatic considerations, a potential *Lys102* residue would not be expected to interact favorably with Lys221 of chymosin. As has already been mentioned, the proximity of Lys221 to residue P4 of κ-casein is one of the reasons for our decision to model the neutral forms of the P4 residue as well. Interestingly, the corresponding residue in camel chymosin is Val221, which may have important consequences for binding and catalysis of κ-casein. The chymosin−κ-casein complex is depicted in Figure 7 after 17 ns of MD simulations.

Arg97 is expected to be important for binding because it is conserved in bovine, camel, pig, buffalo, and goat κ-casein and because a variant of bovine κ-casein that has a histidine residue in the P9 position is a poor substrate for bovine chymosin (*21*). At

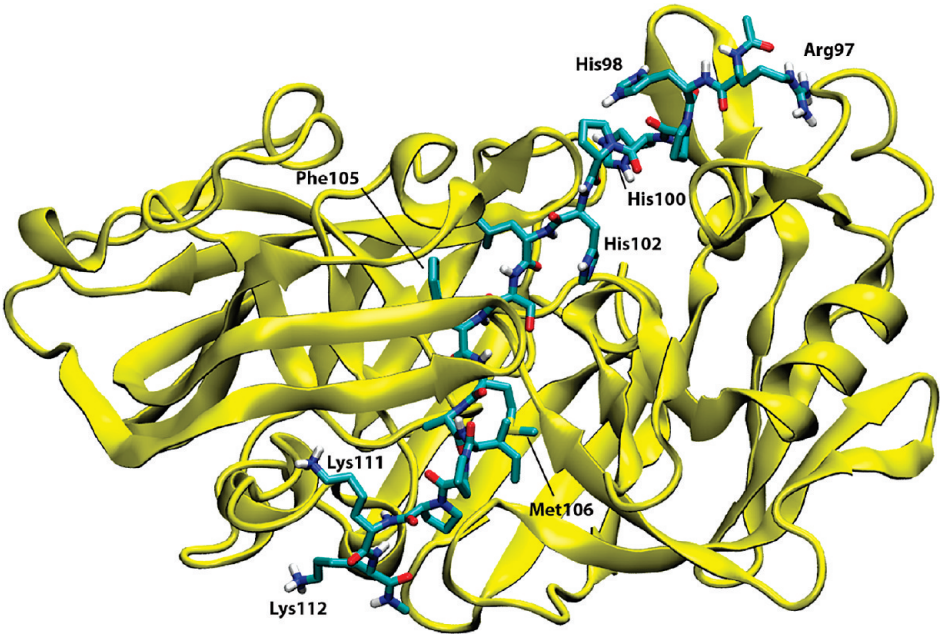


FIGURE 7: The chymosin−κ-casein complex after 17 ns of molecular dynamics simulation (A1).

the beginning of the simulation, *Arg97* interacts with Phe282, Gln280(O), and Gly281(O), before forming stable hydrogen bonds with the side-chain oxygens of Asp247 for the last 10 ns. In this final orientation, the "outer face" of *Arg97* was fully solvated, and the "inner face" was in close proximity to the side chain of Phe282. A hydrogen bond was also present between *Arg97*(*O*) and Asn241(Nδ).

The higher catalytic efficacy of chymosin A compared to chymosin B has previously been attributed to increased electrostatic interactions of the single mutated residue (Asp244 in form A, Gly244 in form B) with either *His98* or *His102* (*17*). In the simulation complex, *His102* is positioned close to Lys221 of chymosin, too far away to interact with residue 244. Either *His98* or *His100* could feasibly interact with residue 244, but the later would require a twisting of the peptide backbone as *His100* points into the binding cleft, making close contacts to Leu12(O) and electrostatic interactions with Asp13. *His98* interacts with the neighboring residue Tyr243 and is therefore well positioned to interact with the mutant Asp244 structure. We note, however, that a small rearrangement of the loop region in which residue 244 is located would be required to allow an optimum alignment of the interacting groups.

P4′−P7′ κ-Casein Residues. Mutagenesis studies have demonstrated that *Lys111* and *Lys112* in κ-casein promote catalysis but that the effect is only observed for *Lys112* if *Lys111* is also present (*34*). While the simulations do not allow us to speculate on the relative importance of the two residues, we do observe that both residues make specific contacts with chymosin throughout the simulation. The side-chain nitrogen of *Lys111* interacts with both Ser74(O) and Ser132(Oγ) through long hydrogen bonds or electrostatic interactions. The *Lys112*-(NH_3^+) group is stabilized by interactions with Pro128(O), Ala131(O), and Ser132(O). There is one *cis*-peptide bond in κ-casein in the simulated complex, between *Ile108* and *Pro109*, which may be important in positioning the two lysine residues.

Conserved Waters. Five of the seven conserved water molecules in the binding cleft were observed to be stable throughout the simulation (see Figure 8) without replacement by κ-casein or other water molecules. The catalytic water and Wat1113 have previously been implicated in catalysis and make stable hydrogen bonds, as have already been discussed (*28*). The remaining three waters are located in the base of the binding cleft near the S1′−S7′ pockets; Wat1116 is hydrogen bonded to Thr35(O), Leu215(O), and Trp191(N), Wat1006 to Ser38(N), Leu130(O), and Asp39(Oδ2), and Wat1007 to Gly124(O), Ser38-(O), and Val138(O). The fact that some of the conserved waters are not replaced is partly a function of their location: many occur in pockets where they would not be expected to be replaced on the time scale of a normal MD simulation. Interestingly, however, the conserved water molecules outside the binding cleft (on the surface of the protein) do undergo exchange with the bulk solvent, which suggests that the simulation is well equilibrated.

One conserved water molecule (Wat1125 in 3CMS) was displaced by κ-casein in the beginning of the simulation. The *Lys111*(NH_3^+) group moved from neighboring residues to form stable interactions with Ser74(O) and Ser132(Oγ), thereby displacing the water. The two serine residues are located on opposing flap regions in the N-terminal domain of the protein. The *Lys111* group maintained this new position for most of the simulation and was only occasionally displaced by solvent.

The single conserved water molecule in the P8−P1 end of the binding cleft (Wat645, imported from 1CMS) was not stable. After 2 ns of simulation, the water molecule moved to bulk solvent, and the neighboring residues (Tyr217(Oγ) and Phe33-(O)) rearranged to hydrogen bond to each other, preventing the lost water from being replaced.

We note that there are many examples of water molecules that are not conserved but which occupy stable positions in pockets in the protein. For instance, a water molecule makes a stable interaction to Gly124(O) and Phe139(N) throughout the simulation.

Molecular Dynamics of Inhibitor-Bound Chymosin from 1CZI. In PDB entry 1CZI, bovine chymosin is complexed with a reduced bond inhibitor, CP-113972. The inhibitor acts as an approximate transition state mimic, in which a hydroxyl group on the ligand donates a hydrogen bond to Asp216(Oδ2) and accepts a hydrogen bond from Asp34(Oδ2). As would be expected, the catalytic water observed in apo chymosin is not

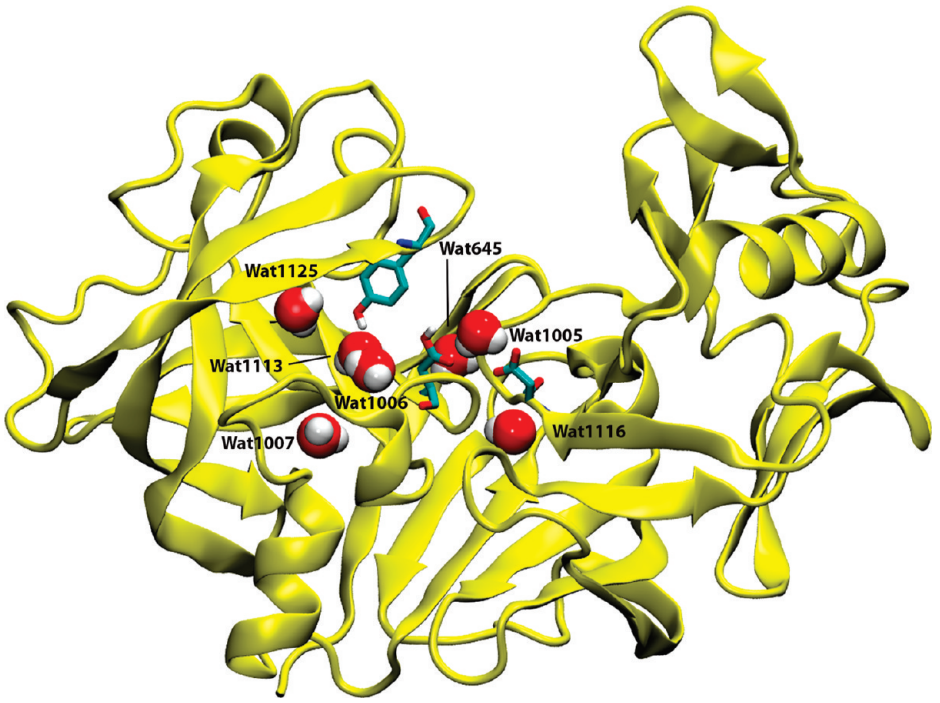


FIGURE 8: The seven conserved water molecules in the binding cleft of chymosin. During molecular dynamics simulation (A1), Wat1125 was displaced by *Lys111*, while Wat645 moved into the bulk solvent and was not replaced; the remaining five water molecules maintained their initial contacts. Tyr77, Asp34, and Asp216 are shown in tube representations.

present. In the simulations of this complex, most of the features found in the κ-casein simulations were observed to be present and stable, which validates the method and the choice of force field for these systems. Both the fireman's grip and the hydrogen bond chain from Tyr77 to Asp34 (via the conserved water molecule Wat369) were intact for the complete simulation, with all hydrogen bond distances fluctuating around 2.8 Å. The side chain of Asp34 twists approximately 80°, such that Asp34(Oδ2) is pointing into the S1−S8 end of the protein, rather than maintaining a coplanar conformation with the side-chain carboxyl group of Asp216.

Dynamics of the β-Hairpin Flap. We performed 48 ns of molecular dynamics simulations for both the open (C1) and self-inhibited (C2) forms of apo chymosin including all conserved water molecules and additional explicit solvent. The initial conformation of Tyr77 (i.e., open or self-inhibited) was maintained for the complete 48 ns in the respective simulations. In the open conformation, Tyr77 is stabilized by a hydrogen bond to a conserved water, as discussed earlier. In simulations of the self-inhibited form, Tyr77 was observed to make a stable hydrogen bond to Gly218(O), preventing it from changing conformation to the open form.

DISCUSSION

The model developed here contains the P9−P7′ fragment of bovine κ-casein complexed with bovine chymosin. The methodology differs from earlier work (*17*) in the use of longer simulation times, unrestrained rather than restrained molecular dynamics, explicit rather than implicit solvation, and the inclusion of crystallographic and conserved waters. κ-Casein in both our model and the earlier work binds in an extended conformation in agreement with data obtained from circular dichroism and solution NMR. Furthermore, in both of these simulated complexes, the P9−P1 residues of κ-casein are observed to make close contacts with the C-terminal domain of chymosin, while the P1′−P7′ residues make contacts with the N-terminal domain. However, the two models do not agree about some critical interactions between chymosin and κ-casein. In the previous model, much importance was placed upon electrostatic interactions between positively charged regions of κ-casein and negatively charged regions of chymosin (*17*). *His98*, *His100*, and *His102* were restrained to retain close contacts with Asp247, Asp279, and Glu245, respectively, and a *His98-Pro99 cis*-peptide was found to be necessary to bring *His98* and Asp247 into close contact. Using unrestrained simulations, slightly different electrostatic contacts were observed. *Arg97*, *His98*, and *His100* made close contacts with Asp247, Asp279, and Asp13, respectively, with the *His98−Pro99* peptide bond in the trans conformation. The inclusion of Asp13 is interesting because it has not previously been implicated in binding. Furthermore, Glu245 made a stable interaction with Lys221 of chymosin, effectively screening the Glu245 to *His102* interaction (Glu245 and *His102* were observed to be separated by ∼6−8 Å). In camel chymosin, Lys221 is replaced by Val221, and in camel κ-casein the residue corresponding to *His102* is an arginine, which would be predominantly charged at pH 6.5. In the absence of other factors, both of these mutations would be expected to alter the electrostatic interactions in agreement with the observed kinetic data. First, the Lys221Val mutation would allow Glu245 to form stronger electrostatic interactions with residue *102* in κ-casein, which agrees with the observation that camel chymosin is a better catalyst than bovine chymosin for the hydrolysis of bovine κ-casein. Second, since the residue corresponding to *His102* in bovine κ-casein is an arginine in camel κ-casein, the longer side chain of the latter might be expected to allow the residue to form closer, stronger interactions with Glu245, which would agree with camel chymosin preferentially binding its native κ-casein. Third, the complex of camel κ-casein liganded to bovine chymosin would bring *Arg102* and Lys221 into close contact, which would be disfavorable on both steric and electrostatic grounds, and would agree with the observation

that bovine chymosin is a poor catalyst for the hydrolysis of camel κ-casein (*2*).

The P7′ residue of κ-casein, *Lys111*, was restrained in previous studies to form electrostatic interactions with Glu133. In the simulations presented here, the side chains of *Lys111* and *Lys112* formed hydrogen bonds with residues on either side of Glu133, but during the course of the simulation, both lysines intermittently left these contacts and made strong electrostatic contacts with the glutamate residue. On the basis of the contacts that *Lys111* and *Lys112* made with chymosin, there is little evidence from which to determine their relative importance for binding. However, *Lys111* was observed to displace a conserved water from between Ser74(O) and Ser132(Oγ), which would be expected to provide an entropic benefit and might provide some clue as to why *Lys111* has been observed to be more important for catalysis than *Lys112*.

The dynamics of the β-hairpin flap in chymosin are still not fully understood. From a naïve comparison of apo- and inhibitor-bound chymosin (*4*, *6*), it appears that the backbone of the β-hairpin flap does not change markedly upon binding, but it is not clear to what extent this conclusion is affected by crystal contacts. What is known is that Tyr77 can occupy either an open or self-inhibited structure (*4*). Our simulations of the apo chymosin suggest that both the open and self-inhibited forms are stable on a 48 ns time period. In addition, during the apo or holo simulations, we observe only small changes in the backbone of the β-hairpin flap, as compared to the crystal structure, which suggests that the flap dynamics are limited to the movement of Tyr77. If this is the case, then it is interesting to note that the *His-Pro* cluster, the proposed allosteric activator, binds to the C-terminal domain of chymosin, whereas the β-hairpin flap is in the N-terminal domain. The obvious question is whether, in this simple model of allosteric activation, binding of the *His-Pro* cluster to this region of the C-terminal domain is sufficient to switch the conformation of Tyr77 without otherwise changing the conformation of the protein.

Further studies are ongoing in our laboratory to understand the binding in both camel chymosin−camel κ-casein complexes and the bovine−camel cross-complexes and to elucidate the molecular basis of chymosin self-inhibition and allosteric activation.

SUPPORTING INFORMATION AVAILABLE

A table containing the most frequent contacts observed between each residue of κ-casein and chymosin during 18 ns of molecular dynamics simulation A1 and a table containing the results of the docking calculations. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Foltmann, B., Pedersen, V. B., Kauffman, D., and Wybrandt, G. (1979) The Primary Structure of Calf Chymosin. *J. Biol. Chem. 254*, 8447–8456.
2. Kappeler, S. R., van den Brink, H. J., Rahbek-Nielsen, H., Farah, Z., Puhan, Z., Hansen, E. B., and Johansen, E. (2006) Characterization of Recombinant Camel Chymosin Reveals Superior Properties for the Coagulation of Bovine and Camel Milk. *Biochem. Biophys. Res. Commun. 342*, 647–654.
3. Gilliland, G. L., Winborne, E. L., Nachman, J., and Wlodawer, A. (1990) The Three-Dimensional Structure of Recombinant Bovine Chymosin at 2.3 Å Resolution. *Proteins 8*, 82–101.
4. Strop, P., Sedlacek, J., Stys, J., Kaderabkova, Z., Blaha, I., Pavlickova, L., Pohl, J., Fabry, M., Kostka, V., and Newman, M. (1990) Engineering Enzyme Subsite Specificity: Preparation, Kinetic Characterization, and X-Ray Analysis at 2.0 Å Resolution of Val111Phe Site-Mutated Calf Chymosin. *Biochemistry 29*, 9863–9871.
5. Newman, M., Safro, M., Frazao, C., Khan, G., Zdanov, A., Tickle, I. J., Blundell, T. L., and Andreeva, N. (1991) X-Ray Analyses of Aspartic Proteinases. IV. Structure and Refinement at 2.2 Å Resolution of Bovine Chymosin. *J. Mol. Biol. 221*, 1295–1309.
6. Groves, M. R., Dhanaraj, V., Badasso, M., Nugent, P., Pitts, J. E., Hoover, D. J., and Blundell, T. L. (1998) 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. *Protein Eng. 11*, 833–840.
7. Chitpinityol, S., and Crabbe, M. J. C. (1998) Chymosin and Aspartic Proteinases. *Food Chem. 61*, 395–418.
8. Williams, M. G., Wilsher, J., Nugent, P., Mills, A., Dhanaraj, V., Fabry, M., Sedlacek, J., Uusitalo, J. M., Penttila, M. E., Pitts, J. E., and Blundell, T. L. (1997) Mutagenesis, Biochemical Characterization and X-Ray Structural Analysis of Point Mutants of Bovine Chymosin. *Protein Eng. 10*, 991–997.
9. Danley, D. E., and Geoghegan, K. F. (1988) Structure and Mechanism of Formation of Recombinant-Derived Chymosin C. *J. Biol. Chem. 263*, 9785–9789.
10. Rampilli, M., Larsen, R., and Harboec, M. (2005) Natural Heterogeneity of Chymosin and Pepsin in Extracts of Bovine Stomachs. *Int. Dairy. J. 15*, 1130–1137.
11. Kabsch, W., and Sander, C. (1983) Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers 22*, 2577–2637.
12. Newman, M., Safro, M., Frazao, C., Khan, G., Zdanov, A., Tickle, I. J., Blundell, T. L., and Andreeva, N. (1991) X-Ray Analyses of Aspartic Proteinases. IV. Structure and Refinement at 2.2 Å Resolution of Bovine Chymosin. *J. Mol. Biol. 221*, 1295–1309.
13. Pearl, L., and Blundell, T. (1984) The Active Site of Aspartic Proteinases. *FEBS Lett. 174*, 96–101.
14. Gilliland, G. L., Winborne, E. L., Nachman, J., and Wlodawer, A. (1990) The Three-Dimensional Structure of Recombinant Bovine Chymosin at 2.3 Å Resolution. *Proteins 8*, 82–101.
15. Dalgleish, D. G. (1998) Casein Micelles as Colloids: Surface Structures and Stabilities. *J. Dairy Sci. 81*, 3013–3018.
16. Groves, M. R., Dhanaraj, V., Badasso, M., Nugent, P., Pitts, J. E., Hoover, D. J., and Blundell, T. L. (1998) 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. *Protein Eng. 11*, 833–840.
17. Plowman, J. E., and Creamer, L. K. (1995) Restrained Molecular Dynamics Study of the Interaction between Bovine Kappa-Casein Peptide 98−111 and Bovine Chymosin and Porcine Pepsin. *J. Dairy Res. 62*, 451–467.
18. Plowman, J. E., Creamer, L. K., Liddell, M. J., and Cross, J. J. (1997) Solution Conformation of a Peptide Corresponding to Bovine Kappa-Casein B Residues 130−153 by Circular Dichroism Spectroscopy and ^{1}H-Nuclear Magnetic Resonance Spectroscopy. *J. Dairy Res. 64*, 377–397.
19. Plowman, J. E., Smith, M. H., Creamer, L. K., Liddell, M. J., Coddington, J. M., Gibson, J. J., and Engelbretsen, D. R. (1994) Proton Assignment and Structural Features of a Peptide from the Chymosin-Sensitive Region of Bovine κ-Casein Determined by 2D-NMR Spectroscopy. *Magn. Reson. Chem. 32*, 458–464.
20. Bairoch, A., Apweiler, R., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M. J., Natale, D. A., O'Donovan, C., Redaschi, N., and Yeh, L. L. (2005) The Universal Protein Resource (UniProt). *Nucleic Acids Res. 33*, D154–159.
21. Macheboef, D., Coulon, J. -B., and D'Hour, P. (1993) Effect of Breed, Protein Genetic Variants and Feeding on Cows' Milk Coagulation Properties. *J. Dairy Res. 9*, 373–374.
22. Dunn, B. M. (2002) Structure and Mechanism of the Pepsin-Like Family of Aspartic Peptidases. *Chem. Rev. 102*, 4431–4458.
23. Veerapandian, B., Cooper, J. B., Sali, A., Blundell, T. L., Rosati, R. L., Dominy, B. W., Damon, D. B., and Hoover, D. J. (1992) Direct Observation by X-Ray Analysis of the Tetrahedral Intermediate of Aspartic Proteinases. *Protein Sci. 1*, 322–328.
24. James, M. N., Sielecki, A. R., Hayakawa, K., and Gelb, M. H. (1992) Crystallographic Analysis of Transition State Mimics Bound to Penicillopepsin: Difluorostatine- and Difluorostatone-Containing Peptides. *Biochemistry 31*, 3872–3886.
25. Coates, L., Erskine, P. T., Wood, S. P., Myles, D. A., and Cooper, J. B. (2001) A Neutron Laue Diffraction Study of Endothiapepsin: Implications for the Aspartic Proteinase Mechanism. *Biochemistry 40*, 13149–13157.
26. Piana, S., and Carloni, P. (2000) Conformational Flexibility of the Catalytic Asp Dyad in HIV-1 Protease: An Ab Initio Study on the Free Enzyme. *Proteins 39*, 26–36.

27. Northrop, D. B. (2001) Follow the Protons: A Low-Barrier Hydrogen Bond Unifies the Mechanisms of the Aspartic Proteases. *Acc. Chem. Res. 34*, 790–797.
28. Andreeva, N., Dill, J., and Gilliland, G. L. (1992) Can Enzymes Adopt a Self-Inhibited Form? Results of X-Ray Crystallographic Studies of Chymosin. *Biochem. Biophys. Res. Commun. 184*, 1074–1081.
29. Prasad, B. V., and Suguna, K. (2002) Role of Water Molecules in the Structure and Function of Aspartic Proteinases. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 58*, 250–259.
30. Fox, P. F., Guinee, T. P., Cogan, T. M., and McSweeney, P. L. H. (2000) Fundamentals of Cheese Science, illustrated ed., pp 98−135, Springer, Berlin.
31. Dunn, B. M., and Hung, S. (2000) The Two Sides of Enzyme-Substrate Specificity: Lessons from the Aspartic Proteinases. *Biochim. Biophys. Acta 1477*, 231–240.
32. Dunn, B. M., Jimenez, M., Parten, B. F., Valler, M. J., Rolph, C. E., and Kay, J. (1986) A Systematic Series of Synthetic Chromophoric Substrates for Aspartic Proteinases. *Biochem. J. 237*, 899–906.
33. Gustchina, E., Rumsh, L., Ginodman, L., Majer, P., and Andreeva, N. (1996) 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. 379*, 60–62.
34. Visser, S., Slangen, C. J., and van Rooijen, P. J. (1987) Peptide Substrates for Chymosin (Rennin). Interaction Sites in Kappa-Casein-Related Sequences Located Outside the (103−108)-Hexapeptide Region that Fits into the Enzyme's Active-Site Cleft. *Biochem. J. 244*, 553–558.
35. Plowman, J. E., Creamer, L. K., Smith, M. H., and Hill, J. P. (1997) Restrained Molecular Dynamics Investigation of the Differences in Association of Chymosin to k-Caseins A and C. *J. Dairy Res. 64*, 299–304.
36. Kashparov, I. V., Russ, A. V., and Andreeva, N. S. (2002) Molecular Dynamics Analysis of Chymosin Conformations in Solution and in Crystalline Environment. *Mol. Biol.* (*Moscow*) *36*, 754–759.
37. Jacobson, M. P., Pincus, D. L., Rapp, C. S., Day, T. J., Honig, B., Shaw, D. E., and Friesner, R. A. (2004) A Hierarchical Approach to all-Atom Protein Loop Prediction. *Proteins 55*, 351–367.
38. Schrödinger LLC (2008) Schrödinger Suite 2008, Maestro Version 8.5, MacroModel Version 9.6, Glide Version 5.0, Prime Version 2.0.
39. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank. *Nucleic Acids Res. 28*, 235–242.
40. Jorgensen, W. L., Maxwell, D. S., and Tirado-Rives, J. (1996) Development and Testing of the OPLS all-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc. 118*, 11225–11236.
41. Still, W. C., Tempczyk, A., Hawley, R. C., and Hendrickson, T. (1990) Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *J. Am. Chem. Soc. 112*, 6127–6129.
42. Abel, R., Young, T., Farid, R., Berne, B. J., and Friesner, R. A. (2008) Role of the Active-Site Solvent in the Thermodynamics of Factor Xa Ligand Binding. *J. Am. Chem. Soc. 130*, 2817–2831.
43. Beuming, T., Farid, R., and Sherman, W. (2009) High-Energy Water Sites Determine Peptide Binding Affinity and Specificity of PDZ Domains. *Protein Sci. 18*, 1609–1619.
44. Michel, J., Tirado-Rives, J., and Jorgensen, W. L. (2009) Energetics of Displacing Water Molecules from Protein Binding Sites: Consequences for Ligand Optimization. *J. Am. Chem. Soc. 131*, 15403–15411.
45. Bas, D. C., Rogers, D. M., and Jensen, J. H. (2008) Very Fast Prediction and Rationalization of pKa Values for Protein-Ligand Complexes. *Proteins 73*, 765–783.
46. Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., Shaw, D. E., Francis, P., and Shenkin, P. S. (2004) Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem. 47*, 1739–1749.
47. Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., Sanschagrin, P. C., and Mainz, D. T. (2006) Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. *J. Med. Chem. 49*, 6177–6196.
48. Wang, J., Wang, W., Kollman, P. A., and Case, D. A. (2006) Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graph. Model. 25*, 247–260.
49. Walker, R. Tutorial B4: Simulating a pharmaceutical compound using antechamber and the Generalized Amber Force Field (http://ambermd.org/tutorials/), accessed 21st July 2009.
50. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004) Development and Testing of a General Amber Force Field. *J. Comput. Chem. 25*, 1157–1174.
51. Jakalian, A., Bush, B. L., Jack, D. B., and Bayly, C. I. (2000) Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: I. Method. *J. Comput. Chem. 21*, 132–146.
52. 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., Walker, R. C., Zhang, W., Wang, B., Hayik, S., Roitberg, A., Seabra, G., Wong, K. F., Paesani, F., Wu, X., Brozell, S., Tsui, V., Gohlke, H., Yang, L., Tan, C., Mongan, J., Hornak, V., Cui, C., Beroza, P., Mathews, D. H., Shafmeister, C., Ross, W. S., and Kollman, P. A. (2006) AMBER 9, University of California, San Francisco.
53. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys. 79*, 926–935.
54. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L., and Schulten, K. (2005) Scalable Molecular Dynamics with NAMD. *J. Comput. Chem. 26*, 1781–1802.
55. Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G., Zhang, W., Yang, R., Cieplak, P., Luo, R., Lee, T., Caldwell, J., Wang, J., and Kollman, P. A. (2003) A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations. *J. Comput. Chem. 24*, 1999–2012.
56. Lee, M. C., and Duan, Y. (2004) Distinguish Protein Decoys by using a Scoring Function Based on a New AMBER Force Field, Short Molecular Dynamics Simulations, and the Generalized Born Solvent Model. *Proteins 55*, 620–634.
57. Hoover, W. G. (1985) Canonical Dynamics: Equilibrium Phase-Space Distributions. *Phys. Rev. A 31*, 1695–1697.
58. Nose, S., and Klein, M. L. (1983) Constant Pressure Molecular Dynamics for Molecular Systems. *Mol. Phys. 50*, 1055–1076.
59. Martyna, G. J., Tobias, D. J., and Klein, M. L. (1994) Constant Pressure Molecular Dynamics Algorithms. *J. Chem. Phys. 101*, 4177–4189.
60. Ewald, P. P. (1921) Die Berechnung Optischer Und Elektrostatischer Gitterpotentiale. *Ann. Phys. 64*, 253–287.
61. Darden, T., York, D., and Pedersen, L. (1993) Particle Mesh Ewald: An $N * \log(N)$ Method for Ewald Sums in Large Systems. *J. Chem. Phys. 98*, 10089–10092.
62. York, D. M., Wlodawer, A., Pedersen, L. G., and Darden, T. A. (1994) Atomic-Level Accuracy in Simulations of Large Protein Crystals. *Proc. Natl. Acad. Sci. U.S.A. 91*, 8715–8718.
63. Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys. 23*, 327–341.
64. Weinbach, Y., and Elber, R. (2005) Revisiting and Parallelizing SHAKE. *J. Comput. Phys. 209*, 193–206.
65. Bürgi, H. B., Dunitz, J. D., and Shefter, E. (1973) Geometrical Reaction Coordinates. II. Nucleophilic Addition to a Carbonyl Group. *J. Am. Chem. Soc. 95*, 5065–5067.
66. Bürgi, H. B., Lehn, J. M., and Wipff, G. (1974) Ab Initio Study of Nucleophilic Addition to a Carbonyl Group. *J. Am. Chem. Soc. 96*, 1956–1957.
67. Sherman, W., Day, T., Jacobson, M. P., Friesner, R. A., and Farid, R. (2006) Novel Procedure for Modeling Ligand/Receptor Induced Fit Effects. *J. Med. Chem. 49*, 534–553.
68. Beveridge, A. J. (2009) An Ab Initio Study of the First Stage of Catalysis in the Monomeric Aspartic Proteinases. *J. Mol. Struct.* (*THEOCHEM*) *900*, 1–8.