JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

ARTICLE

pubs.acs.org/JAFC

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

Jesper Sørensen,[†,‡] David S. Palmer,[‡,#] Karsten Bruun Qvist,[§] and Birgit Schiøtt*[,†]

[†]The Center for Insoluble Protein Structures (inSPIN) and the Interdisciplinary Nanoscience Center (*i*NANO) and [‡]Department of Chemistry, Aarhus University, Langelandsgade 140, DK-8000 Aarhus C, Denmark

[§]Cultures and Enzymes Division, Innovation, Chr. Hansen Inc., Bøge Allé 10-12, DK-2970 Hørsholm, Denmark

Supporting Information

ABSTRACT: Bovine chymosin has long been the preferred enzyme used to coagulate cow's milk, in the initial stage of cheese production, during which it cleaves a specific bond in the milk protein κ-casein. Recently, camel chymosin has been shown to have a 70% higher clotting activity toward cow's milk and, moreover, to cleave κ-casein more selectively. Bovine chymosin, on the other hand, is a poor clotting agent toward camel's milk. This paper reports a molecular modeling study aimed at understanding this disparity, based on homology modeling and molecular dynamics simulations using peptide fragments of κ-casein from cow and camel in both bovine and camel chymosin. The results show that the complex between bovine chymosin and the fragment of camel κ-casein is indeed less stable in the binding pocket. The results also indicate that this in part may be due to charge repulsion between a lysine residue in bovine chymosin and an arginine residue in the *P4* position of camel κ-casein.

KEYWORDS: chymosin, κ-casein, molecular dynamics, homology modeling

INTRODUCTION

Preparations of proteolytic enzymes, so-called rennets, have been used to initiate the first step of cheesemaking, the clotting of milk, for millennia. Historically, most of these enzyme preparations have been extracts from the stomachs of ruminant, mainly comprising chymosin and pepsin as clotting agents, but more recently proteases from plants or microorganisms have also been used.[1] Chymosin is an aspartic protease found predominantly in the stomachs of mammalian infants. The enzyme cleaves the milk protein κ-casein, which causes the milk to curdle.[2] The result of this process is that the milk is retained in the digestive system longer, allowing for increased absorption in the bowels. In contrast to other aspartic proteases (e.g., pepsin), chymosin exhibits a low general proteolytic activity.[3] Until recently, bovine chymosin has been considered the most suitable enzyme for clotting of bovine milk due to its high specificity for cleaving the peptide bond between *Phe105* and *Met106* in bovine κ-casein (κ-casein residues are given in italics in the text throughout).

Recently, commercial demand for an enzyme to produce cheese from camel's milk led to the expression of camel chymosin in *Aspergillus*.[4] Interestingly, it was later demonstrated that camel chymosin has a 70% higher clotting activity than bovine chymosin toward bovine milk; additionally, it was found to be more selective, with a general proteolytic activity 5 times lower than that of bovine chymosin.[4] The two enzymes have a high sequence identity (85%) and sequence similarity (94%), which in part could explain why camel chymosin is able to clot bovine milk, but not why it does this so effectively (a full sequence alignment is available in the Supporting Information). It is surprising, then, to note that bovine chymosin is a very poor clotting agent toward camel's milk.[4] This disparity is not understood at a molecular level, due to a lack of structural information about chymosin–κ-casein complexes. Understanding the selectivity holds industrial interest as bovine chymosin has been sold as a clotting enzyme in cheese manufacturing since it was first industrially purified and standardized in activity by the company Chr. Hansen Inc. in 1874.[1] Recently, camel chymosin has been successfully marketed as an alternative to bovine chymosin, offering higher yields and less bitter taste due to the lower proteolytic activity.[4,5]

There are, currently, four X-ray crystal structures of bovine chymosin,[6−9] whereas no structures of camel chymosin are available. Bovine chymosin has been crystallized in an open apo-form (pdb code 3CMS) and in what is believed to be a self-inhibited form (1CMS, 3CMS, 4CMS).[10] Bovine chymosin has also been cocrystallized with a bound norstatine-based inhibitor (1CZI).[9] No X-ray or NMR structure of either bovine or camel κ-casein is available yet.

Bovine chymosin is a globular protein with 323 amino acid residues,[2] which folds into a bilobal structure, with two similar β-barrel domains (see Figure 1). There is pseudosymmetry along a single cleft containing the catalytic residues Asp34 and Asp216 (numbering is according to the 1CMS crystal structure, a notation known as chymosin numbering). A β-hairpin flap is found above the catalytic residues, making contacts with the substrate upon binding. The overall structure is common for cellular pepsin-like aspartic proteases, among which there is a

Received: December 21, 2010
Revised: April 5, 2011
Accepted: April 8, 2011
Published: April 08, 2011

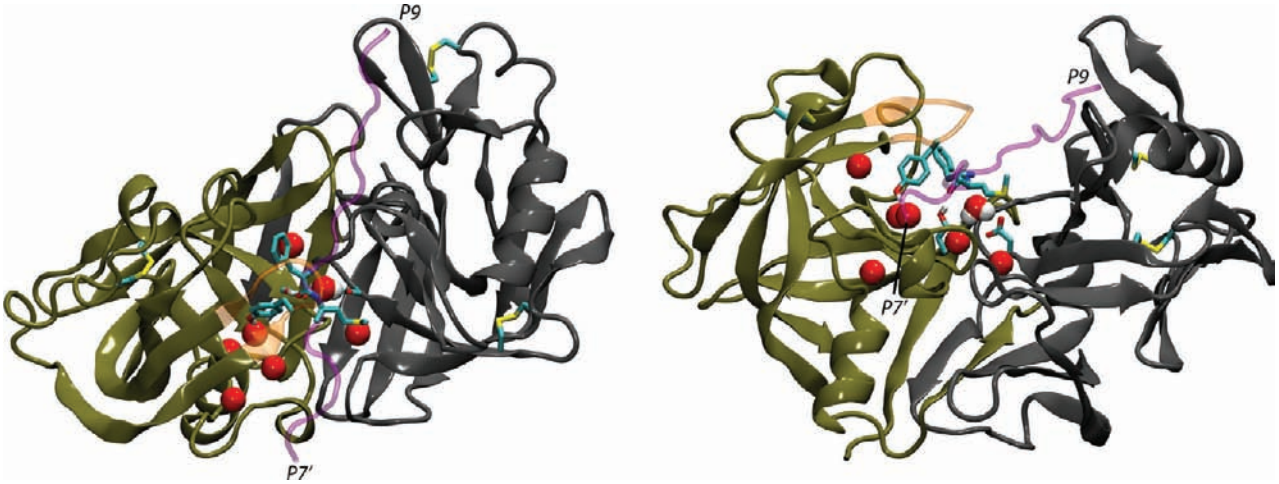


Figure 1. Top (left) and front (right) view of bovine chymosin complexed with the *P9*−*P7′* fragment (residues 97−112) of bovine κ-casein (nomenclature explained in the text). The N- and C-terminal domains have been colored in tan and gray, respectively. The main chain of bovine κ-casein from our previous model is shown in purple.[18] The *P1* and *P1′* residues are shown (in cyan) to illustrate the position of the scissile bond right above the two aspartic acids (in cyan) and the catalytic water molecule (in spacefilling). Tyr77 in the flap (in orange) is shown, as well as the cysteine bridges and the seven conserved water molecules (oxygens shown in spacefilling) in or close to the binding cleft.

high degree of structural homology.[11,12] It is, therefore, very likely that camel chymosin will be structurally very similar to bovine chymosin, especially given the high degree of sequence identity between the two species.[12] In both bovine and camel chymosin the catalytic residues Asp34 and Asp216 occur in conserved Asp-Thr-Gly motifs. The side chains of the catalytic aspartic acid residues are oriented toward each other, in an approximately planar geometry with a water molecule placed between the two catalytic residues in the three apo crystal structures of chymosin.

Four different types of casein are found in milk: α_{s1}-, α_{s2}-, β-, and κ-caseins. The role of κ-casein is to help solubilize the three other caseins, which are otherwise insoluble in milk serum, by promoting the formation of aggregates known as casein micelles. The κ-casein peptides cluster on the surface of these micelles, with their hydrophilic C-terminal ends pointing into the solvent. The catalytic action of chymosin is to cleave off the C-terminal part of κ-casein at the scissile bond (i.e., *Phe105*−*Met106* for bovine), which destabilizes the micelles, causing precipitation of the casein proteins. The bovine and camel κ-casein peptides contain different numbers of residues (169 and 162 residues, respectively).[13] The difference is in part due to an 8-residue deletion occurring close to the fragment that binds in the active site cleft (a full sequence alignment is available in the Supporting Information). It is unclear whether this deletion has any effect on enzyme activity; however, the deletion is not unique to the camel variant; it is also observed in pig and horse κ-casein. The sequence alignment of the κ-casein fragment that binds in the cleft is shown in Figure 2 for bovine, camel, and other domesticated animals. Due to the previously mentioned deletion, the scissile bond in camel κ-casein is *Phe97*−*Ile98*, whereas it is *Phe105*−*Met106* in bovine κ-casein.

In what follows, we use the Schechter and Berger nomenclature to refer to the peptide residues and to the pockets of residues in the enzyme that interact with them.[14] For example, *Pn* and *Pn′* are used to denote κ-casein residues on either side of the scissile bond; for example, for camel κ-casein *Ser96*, *Phe97*, *Ile98*, and *Ala99* are referred to as *P2*, *P1*, *P1′*, and *P2′*, respectively (also see Figure 2). Similarly, Sn and Sn′ denote the corresponding

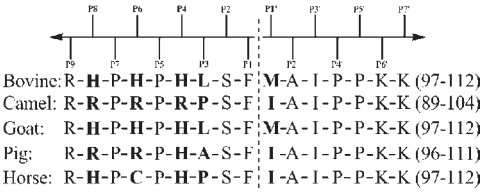


Figure 2. Amino acid sequences of cow, camel, goat, pig, and horse, near the P1−P1′ cleavage site. Positions with differences are in bold. Numbers in parentheses to the right are the sequence numbers of the aligned sequences of each variant.

residues forming the different pockets in the enzyme interacting with the peptide side chains. Due to the different numbering of residues in the two κ-casein variants, the Schechter and Berger nomenclature will be used (in italics) instead of residue numbers for κ-casein throughout the text, for example, *PheP1* rather than *Phe105*.

With one exception,[15] the substrates and inhibitors of aspartic proteases are found to bind in an extended conformation.[11] This is consistent with the X-ray crystal structure of chymosin inhibited by a norstatine-based inhibitor, which binds in the S4−S1′ pockets (1CZI).[9] A combined molecular modeling, circular dichroism, and solution NMR study of chymosin and κ-casein,[16,17] as well as X-ray structural analysis of many aspartic proteases,[11] confirms this.

Gilliland et al. and Newman et al. have used X-ray crystal structures to determine the composition of amino acids making up seven of the pockets (S4−S3′) in bovine chymosin.[6,8] It has been shown for bovine chymosin that the *P8*−*P7′* residues are located in the binding cleft during catalysis and that the *P9* residue, which is conserved in κ-casein, probably also binds in the active site cleft.[16,18]

The S1 pocket is hydrophobic in nature and more specific than the S1′ pocket.[19] If *PheP1* is replaced by Phe(NO_2) or cyclohexylamine analogues, the substrates are still hydrolyzed, although at a lower rate.[20] The S2 pocket is of low specificity, and it has been shown that it can accommodate tyrosine, valine, and serine residues in the substrate.[19] The specificity of the S2 pocket of bovine chymosin has been suggested to be due to the properties

of Lys221[21] and Gln288.[19] However, previous molecular modeling studies carried out by us indicated that the Lys221 side chain does not lie in the S2 pocket, but rather in the S4 pocket.[18] It has been suggested that the chemical characteristics of the *SerP2* position are crucial for enzyme action in bovine chymosin.[22] A hydrogen bond chain is formed from the *SerP2(OH)* to *Thr219-(OH)* and onward to Asp216(Oδ), assisting in keeping the two aspartic acids planar.[18] Binding in the S3 pocket is found to be more specific,[19] which is in accordance with the observation by Hill that protease activity is not present unless a residue occupies the *P3*-position.[22,23] It has been shown that isoleucine and valine are favored at *P3* in the substrate,[19] which is consistent with mutagenesis studies that have shown that catalysis is promoted by aliphatic residues, tolerated by hydrophilic residues, and disfavored by proline and positively charge residues at this position.[21,24] These observations are interesting as the *P3*-position in camel κ-casein is a proline residue, which may indicate why bovine chymosin only poorly clots camel's milk. Noting that the catalytic rate for hydrolysis of the *P3*−*P2′* fragment is slower than that of the full-length κ-casein protein and that the pentapeptide is not cleaved at pH 6.3, which is close to the pH of milk, Hill stated that other residues were needed to modulate the binding and subsequent cleavage; in particular, residues *HisP6* and *HisP4* were implicated as important of this modulating effect.[22] Interestingly, experiments have shown that the dipeptide H-Phe-Met-OH is not easily hydrolyzed by bovine chymosin, and neither are tri- or tetrapeptides containing the scissile bond.[23] The S8−S4 pockets have been shown to be very important for binding. It has been demonstrated that the rate of hydrolysis is ∼18-fold higher for the fragment *P8*−*P4′* compared to the *P3*−*P4′* fragment.[3,25] Moreover, it has been demonstrated that preincubating chymosin with the *P8*−*P4* pentapeptide and subsequently adding the remaining *P3*−*P4′* fragment increased the rate of catalysis by 200-fold compared to the catalytic rate of *P3*−*P4′* alone.[3,25] The increased catalytic rate suggests an increase in binding affinity for these additional amino acids residues, whereas the preincubation experiment suggests that a conformational change could be required for correct binding, which can be induced by the prebinding of the *P8*−*P4* fragment. At the tip of the flap above the catalytic residues sits Tyr77, which is believed to be involved in the activity of the enzyme, via a side-chain rotamer that in one conformation leads to self-inhibition of the apoenzyme.[10] In the 1CMS and 4CMS X-ray crystal structures of bovine chymosin the position of Tyr77 is such that it blocks an amino acid side chain from a substrate from entering the S1 and S3 pockets.[6,8] In the X-ray crystal structure with a norstatine-based inhibitor (1CZI), Tyr77 instead helps form the S1 pocket. Moreover, in 3CMS the position of Tyr77 in the crystal has been resolved in both of the aforementioned orientations (in a 60:40 ratio, respectively).[7] The *P8*−*P4* fragment does not directly interact with the Tyr77 residue, and therefore it has been suggested that the *P8*−*P4* fragment acts as an allosteric activator,[3,25] but the mechanism of activation has not yet been elucidated. A single-point mutation of the *P9* arginine residue in bovine κ-casein to a histidine residue has been demonstrated to result in a poor substrate for bovine chymosin.[26]

It is clear from the experimental evidence discussed above that pockets around the active site, which bind residues adjacent to the scissile bond, are involved in the specificity of the protease activity. Recently, we have developed a molecular model of bovine chymosin in a complex with a 16-residue fragment of bovine κ-casein (BOV/BOV) based on the X-ray crystal structures of bovine chymosin (we use a slash-separated nomenclature to distinguish the different chymosin/κ-casein complexes; for example, BOV/CAM refers to the complex of bovine chymosin with camel κ-casein). We used conformational search algorithms in combination with molecular dynamics simulations (20 ns) to develop the BOV/BOV model.[18] The model is in good agreement with the limited experimental data about the complexes, and it correctly predicts the existence of the fireman's grip hydrogen-bonding network and a stable active site with geometries appropriate for nucleophilic attack on the peptide bond carbonyl carbon by the catalytic water molecule. Additionally, the residues of κ-casein were shown to be correctly positioned around the active site in accordance with X-ray crystallographic structures and kinetic data.[18] In this paper we present homology models, of camel chymosin with camel κ-casein fragments (CAM/CAM) as well as the cross complexes between bovine and camel, based on our previous model. In our previous paper we presented 20 ns of molecular dynamics simulation for the BOV/BOV model. We now present 96 ns of molecular dynamics simulation for each of the four complexes. These models assist in explaining the differences in efficacy for the four complexes.

MATERIALS AND METHODS

Homology Modeling. Bovine and camel chymosin share a high sequence identity (85%) and sequence similarity (94%), which suggests that homology models developed with standard methods will be reasonably accurate,[27] similar to what is known for other aspartic proteases.[12] A sequence alignment between bovine and camel chymosin was performed using ClustalW[28] through The UniProt Knowledgebase (UniProtKB)[29,30] (the sequences have UniProtKB entry no. P00794 and Q9GK11, respectively). Additionally, a sequence alignment of bovine and camel κ-casein was performed (UniProtKB entry no. P02668 and P79139, respectively). The amino acid sequences of these fragments are given in Figure 2. Full sequence alignments of chymosin and κ-casein are available in the Supporting Information.

Homology models of BOV/CAM, CAM/BOV, and CAM/CAM were developed using MODELLER[50,51] 9v7, where putative models were ranked using the DOPE score. The energy-minimized BOV/BOV complex extracted after 17 ns of unrestrained molecular dynamics simulations in TIP3P water was selected as the template structure.[18] A single additional restraint was applied to align the side chains of the catalytic Asp residues in the camel chymosin structures (the Asp34(Cγ) to Asp216(Cγ) distance was restrained to 5.4 Å, which is the average separation observed in unrestrained molecular dynamics simulations of BOV/BOV). It is also comparable to distances observed in X-ray crystal structures of aspartic proteases. The coordinates of all conserved water molecules in BOV/BOV were included in the homology models. Side chains were manually rotated to favor hydrogen bonding between Ser104(Oγ)···Thr219(Oγ), Thr219(Oγ)···Asp216(Oδ2), and Ser37(Oγ)···Asp34(Oδ2), before MD simulations were started. The final homology models were validated using the structure assessment tools in the SWISS-MODEL workspace,[31] in particular, PROCHECK[32] and Qmean.[33] Additionally, the SolvX server[34] and ProQ server[35] were utilized to provide statistical measures of the reliability of the models.

Molecular Dynamics Simulations. The XLEAP module in AMBER 9[36] was used to prepare protein−ligand complexes for molecular dynamics simulations. Disulfide bonds were introduced in bovine and camel chymosins between Cys47−Cys52, Cys207−Cys211, and Cys250−Cys283. Ionizable residues were modeled with the protonation

Table 1. Protonation States of the Ionizable Residues in Bovine and Camel Chymosins[a]

molecule	species	ionizable residues
chymosin	bovine	(COOH)-Asp34, ε-His55, (ImH+)-His76, (ImH+)-His146, δ-His181, (ImH+)-His292
	camel	(COOH)-Asp34, ε-His55, (ImH+)-His56, (ImH+)-His76, (ImH+)-His146, δ-His181

[a] States used in the molecular dynamics simulations. All Glu, Lys, Arg, and Asp residues that are not mentioned in the table were modeled as charged. All Cys and Tyr residues were modeled as neutral. Histidine residues were modeled with a hydrogen on either Nε (ε-His), Nδ (δ-His), or both nitrogens forming the histidinium tautomer (ImH+-His), which is positively charged.

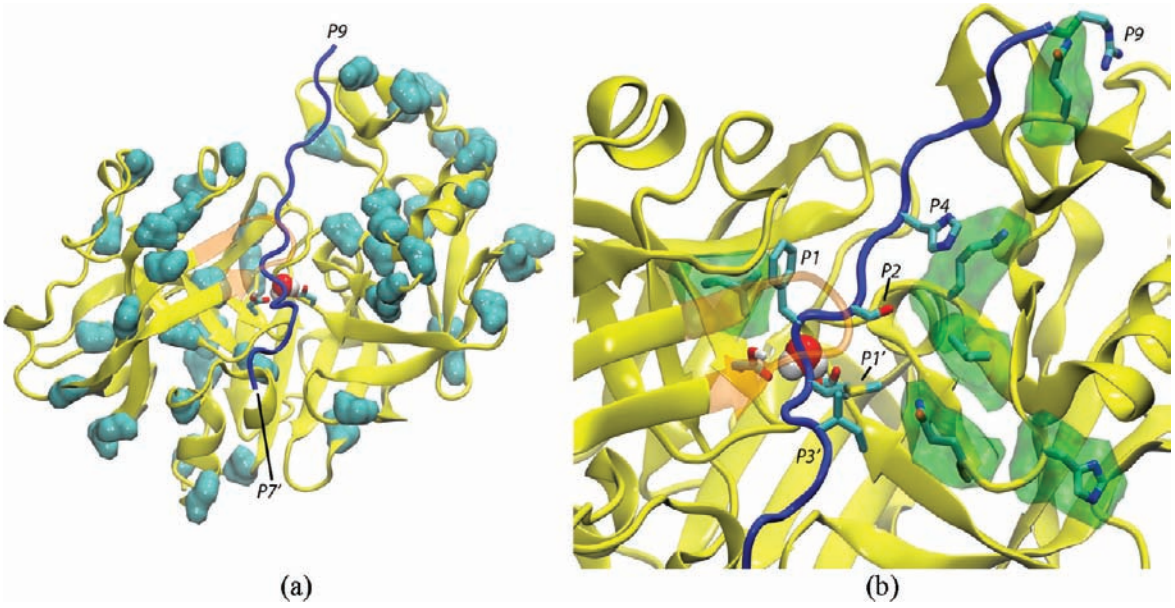


Figure 3. BOV/BOV complex where (a) the backbone atoms of amino acids that are different between bovine and camel chymosin are displayed by a cyan surface to illustrate that the differences are scattered in the structure and not localized in certain areas and (b) the differences in chymosin close to the binding cleft are shown along with the residues of the κ-casein fragment that interact with these.

states given in Table 1, which are consistent with our previous simulations of BOV/BOV.[18] The protonation states of camel residues not present in BOV/BOV were predicted with PROPKA 2.0.[37] Assuming a pH of 6.5, the overall charges of bovine and camel chymosin in our molecular models are −12 and −6, respectively. The *P8* and *P6* histidine residues in bovine κ-casein were modeled as histidinium tautomers, whereas the *P4* histidine was modeled as δ-His.[18] All conserved water molecules were included in the molecular dynamics simulations. TIP3P water molecules[38] were added within a distance of 10 Å around the protein in a periodic box. Between 11000 and 12000 water molecules were added to each system. The systems were neutralized and brought to an ionic strength of 0.07 mol dm^{-1}, to mimic the ionic strength in milk, by adding sodium and chloride ions, as required.

Molecular dynamics simulations were run in NAMD[39] using the AMBER FF03 force field parameters developed by Duan et al.[40] For BOV/CAM, CAM/CAM, and CAM/BOV a more elaborate minimization and equilibration protocol was employed compared to the previous BOV/BOV simulation (the procedure is provided in the Supporting Information). This extended protocol was used to ensure correct planar geometries for the catalytic residues in the minimization and equilibration phase while removing unfavorable interactions elsewhere in the complex introduced by the homology modeling. 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 at 100 fs, and the piston temperature at 300 K.[41−43] 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, and all electrostatic interactions were calculated using the Particle Mesh Ewald (PME) method.[44−46] 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.[47,48] The equations of motion were integrated every 2 fs using the Velocity Verlet algorithm, and snapshots were stored every 2 ps. Each of the four complexes was equilibrated for 4 ns, prior to 96 ns of production dynamics, totaling 400 ns of simulation time.

RESULTS

Sequence Analysis and Homology Models. The sequence alignments reveal a very high homology between the two variants of chymosin (see the Supporting Information) with no gaps in the alignments. Forty-eight residues are different between bovine and camel chymosin, of which 25 are titratable. The sequence alignments illustrate that the differences between the two enzyme variants are spread out in the sequence and are not localized in a particular region of the protein (see Figure 3a). Only six residues are different close to the κ-casein binding region. There are no differences in the residues directly adjacent to the catalytic residues, Asp34 and Asp216, and the conserved Asp-Thr-Gly motif is present in camel chymosin. Overall there are 24 residues that change charge, resulting in a charge

Table 2. Homology Modeling Scores and Structural Measures[a]

model	DOPE	RMSD (Å)	R_{gyr} (Å)	SASA (Å^2)	Ramachandran outliers
3CMS_B					Leu12, Asp13, Ser94, Gln189, Gln280,
BOV/BOV			19.85	14799	Asn10, Ser14, Ser94, Glu133, Tyr134, Gln189
BOV/CAM	−40593.17	0.23	19.97	15064	Ser94, Gln280, His292
CAM/CAM	−39768.48	0.21	19.99	14917	Ser10, Ser14, Ser94, Glu133, Tyr134
CAM/BOV	−40154.65	0.24	19.73	14816	Ser10, Ser14, Ser94, ASn95, Glu133, Tyr134

[a] Scores of each model and the template. DOPE scores calculated with MODELLER, RMSD values of the complexes calculated using the backbone Cα atoms aligned onto the template structure (BOV/BOV), radius of gyration, solvent accessible surface area, and a list of the outliers on the Ramachandran plots.

difference of +6 going from −12 in bovine chymosin to −6 in the camel variant. The following differences are observed in the binding sites of bovine and camel chymosin (the positions of the residues in chymosin are shown in Figure 3b). In the S9 pocket there is a Gln242Arg change adding a positive charge. In our model of BOV/BOV, Lys221 is positioned in the S4 pocket. A Lys221Val change in the S4 pocket removes a positive charge in the camel enzyme. In the S2 pocket a Val223Phe change alters the size of the residue in this position, although they are both nonpolar. A change occurs in the S1 pocket Leu32Val, which does not dramatically change the chemistry at this position and only slightly alters the size of the residue. In the S1′ and S3′ pockets the Gln294Glu change introduces a negative charge on the enzyme. Additionally, there is a His292Asn change near the S3′ pocket, which furthermore removes a positive charge from this position. The frequency of contacts between the residues in these pockets and κ-casein is described further in the section discussing the MD results. There is a positively charged area on camel chymosin in the region leading up to the S9 pocket. Casein micelles are overall negatively charged,[49] and this positively charged region could assist in the association of camel chymosin and casein micelles.

The sequence alignments of κ-casein showed significant differences. Overall, the sequence identity is 54%. There is an 8 residue deletion 7 residues before the part of κ-casein that binds to the enzyme cleft. The sequence identity of the 16 amino acids included in the modeling is 69%. The *P2*, *P1*, and *P2′*−*P7′* residues are invariant, and in our former work these were shown to be important for binding. In the *P8*, *P6*, and *P4* positions, histidine residues in bovine κ-casein are arginine residues in camel κ-casein. In pig κ-casein, *P8* and *P6* are also arginine residues, whereas *P4* remains a histidine residue. In goat, all three positions are histidine residues, whereas in horse *P8* and *P4* are histidine residues and *P6* is a cysteine residue. A leucine in bovine κ-casein at position *P3* is a proline residue in camel κ-casein. In goat the *P3*-position contains a leucine residue as in bovine, in pig the position is held by an alanine residue, and in horse a proline residue is present. Lastly, the *P1′* methionine in bovine κ-casein is an isoleucine in camel κ-casein. This isoleucine is also present in both pig and horse, whereas in goat it is a methionine residue as in bovine.

Homology modeling was used to generate the three models of complexes between chymosin and the κ-casein fragments, CAM/BOV, CAM/CAM, and BOV/CAM, using our previously reported BOV/BOV complex as a template.[18] The models, which have been selected on the basis of the DOPE scores, all have very low RMSD values compared to the template model (see Table 2). Similarly, the radius of gyration of these models is also very comparable along with the solvent accessible surface area (SASA). These low RMSD values and similar radius of gyration and SASA values are expected, as the sequence identity is very high (85 and 69% for the enzyme and the κ-casein fragment, respectively), and thus the structures are expected to be very similar.[27] To evaluate the structures, Ramachandran plots have been calculated for each model as well as for the X-ray structure and the template (see Table 2 and the Supporting Information). From the 3CMS PDB structure, there are five outliers in generously allowed regions. The residues (listed in Table 2) are all positioned in loop regions, where they are not expected to fall in regions of well-defined secondary structure. In template and homology models, all of the outliers are also located in loop regions. Of the outliers lying close to the binding cleft, His292 is located near the S1′ pocket, whereas the loop containing residues 10−14 is part of the S6 and S7 pockets and residues 133−134 are positioned in a loop helping to form the S6′ and S7′ pockets.

Table 3. Web-Based Model Evaluation Scores on the Three Homology Models and the Template for Comparison[a]

	SolvX	LGscore	MaxSub	Qmean
BOV/BOV template	−86.7	5.737	0.454	0.720
BOV/CAM	−85.0	4.359	0.328	0.768
CAM/CAM	−112.2	4.571	0.319	0.729
CAM/BOV	−113.2	4.207	0.314	0.720

[a] The structures were evaluated without hydrogen atoms.

The models were also evaluated using selected Web services (see Table 3). The SolvX method evaluates the models with respect to contacts between the complex and solvent on an amino acid residue basis, which is matched against values derived from a set of known structures.[34] Generally, the lower the SolvX score the better the model. The template and the BOV/CAM model have roughly the same SolvX score (−86.7 and −85.0, respectively), and similarly CAM/CAM and CAM/BOV have almost the same score (−112.2 and −113.2, respectively). The models with camel chymosin have a lower score than the bovine chymosin, reflecting the fact that more charged amino acid residues reside at the surface of camel chymosin. The ProQ service[35] evaluates the models using a neural network trained to predict correct models based on two scoring measures, the MaxSub[52] and LGscore.[53] On the basis of these two scores the models are placed in three categories. A value over 1.5 using the LGscore signifies a fairly good model, a score over 2.5 a very good model, and a score over 4 an extremely good model. The divisions between these categories using the MaxSub score are 0.1, 0.5, and 0.8, respectively. On the basis of the LGscore the

homology models of the complexes between chymosin and κ-casein fragments all fall in the extremely good category, and based on the MaxSub score the models fall in the high end of the fairly good category. Lastly, the Qmean method estimates the reliability of the model and scores the models between 0 and 1 on the basis of different structural descriptors, where 1 is the best possible score.[33] The Qmean scores for the models are very comparable to those for the template. Although there are minor discrepancies between these measures, they all signify that the models are of high quality, which was expected given the very high sequence identity. Adding to this the low RMSD values and lack of severe Ramachandran outliers, the models may be considered to be very good.

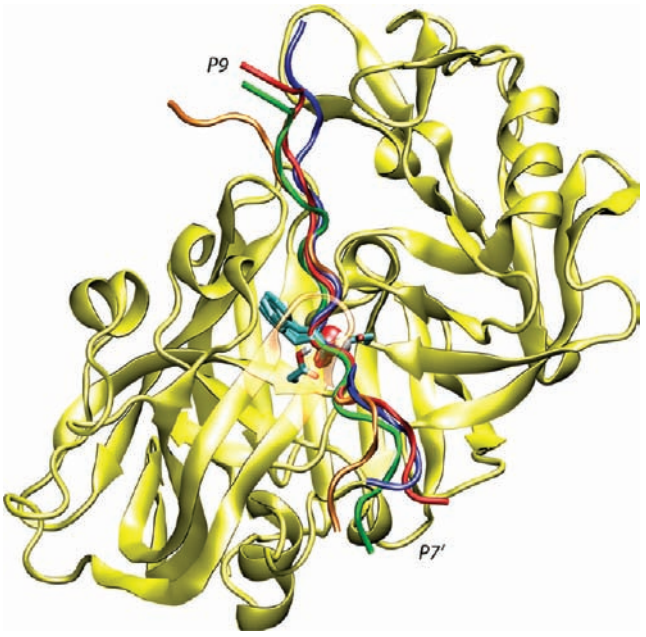


Figure 4. Bovine chymosin at 96 ns with the bovine κ-casein fragment bound (blue). The κ-casein fragment from the BOV/CAM (orange), CAM/CAM (red), and CAM/BOV (green) complexes are overlaid from their conformation in their respective complexes after 96 ns of MD. The catalytic Asp34 and Asp216 (cyan) along with the catalytic water molecule (space filling) are shown to mark the active site. The side chain on the PheP1 (cyan) from all four κ-casein fragments is also shown to mark its buried position next to the active site in the S1 pocket.

Molecular Dynamics. Ninety-six nanoseconds of production MD was performed on each of the four complexes. The κ-casein fragment remains tightly bound in the binding cleft during MD for all of the complexes. Figure 4 shows a snapshot of the four κ-casein peptides overlaid on the structure of bovine chymosin after 96 ns. The backbones of the κ-casein peptides show very similar conformations near the cleavage site. The RMSD values for the enzymes (calculated for Cα atoms and fitted on the first frame of the BOV/BOV simulations) reach equilibrated states at 1.2, 1.9, 1.5, and 2.0 Å for BOV/BOV, BOV/CAM, CAM/CAM, and CAM/BOV, respectively, indicating that all four complexes are stable.

The fluctuation of each residue for the BOV/BOV simulation is shown in Figure 5 (bottom panels), whereas the relative residue-based fluctuations compared to the BOV/BOV simulation are shown in the top panels for the three complexes. In general, the fluctuations are high in loop or turn regions, which is as expected. Overall, the dynamics of the enzymes are very similar, although there are a few parts where the dynamics are different, mostly in the C-terminus of chymosin. The BOV/CAM simulation reveals less flexibility of chymosin residues 9−16, which is a loop region that is close to the binding pocket. This can be explained by the conformational change of the N-terminus of the camel κ-casein fragment in this complex, as shown in Figure 4. The decrease in flexibility of residues 158−164 is also due to this conformational change as these residues are in a loop region right below the loop holding residues 9−16. Regions in the C-terminal domain of the enzyme in the BOV/CAM complex appear to be more flexible, which is also due to the changed position of the N-terminus of the κ-casein fragment, which no longer binds strongly to this region. Changes in the dynamics for this C-terminal region also occur in the other cross-complex, CAM/BOV, which can likewise be

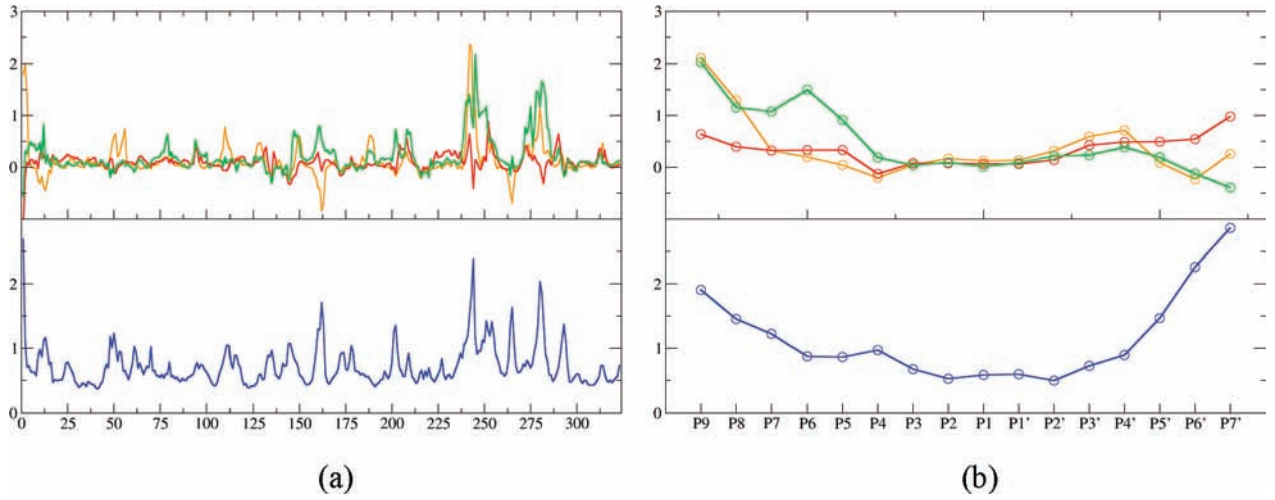
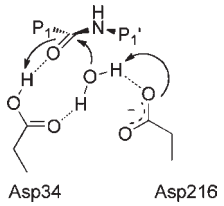


Figure 5. (a) Bottom panel is the RMSF graph of the enzyme in the BOV/BOV complex, and the top panel is the ΔRSMF graph of the enzyme in the BOV/CAM (orange), CAM/CAM (red), and CAM/BOV (green) complexes compared to the BOV/BOV complex. (b) Bottom panel is the RMSF graph of the κ-casein fragments in the BOV/BOV complex, and the top panel is the ΔRSMF graph of the κ-casein fragments in the other complexes compared to the BOV/BOV complex.

Table 4. Active Site Geometry[a]

	Asp34(Oδ2)···*PheP1*(O) (Å)	*PheP1*(C)···Wat(O) (Å)	*PheP1*(O)···*PheP1*(C)···Wat(O) (deg)
20 ns	3.19 ± 0.61 (2.91)	3.11 ± 0.32 (3.04)	88.7 ± 9.9 (88.6)
BOV/BOV	2.94 ± 0.43 (2.79)	3.55 ± 0.82 (3.11)	79.6 ± 16.6 (82.4)
BOV/CAM	2.72 ± 0.13 (2.70)	4.11 ± 0.58 (4.17)	50.4 ± 24.9 (46.3)
CAM/CAM	2.85 ± 0.34 (2.76)	3.16 ± 0.41 (3.06)	83.5 ± 13.2 (85.3)
CAM/BOV	3.12 ± 0.59 (2.87)	3.28 ± 0.23 (3.25)	85.7 ± 9.8 (85.8)



[a] Distances and angles in the active site. The numbers give the means, standard deviations, and medians (in parentheses) of each distance and angle.

Table 5. Hydrogen Bond Network Close to the Active Site[a]

	BOV/BOV	BOV/CAM	CAM/CAM	CAM/BOV
Thr217(N)···Thr35(Oγ)	3.28 ± 0.3 (3.23)	3.05 ± 0.2 (3.01)	3.26 ± 0.4 (3.13)	3.18 ± 0.5 (3.02)
Thr35(N)···Thr217(Oγ)	3.35 ± 0.3 (3.29)	3.31 ± 0.3 (3.26)	3.18 ± 0.2 (3.15)	3.25 ± 0.3 (3.18)
Thr217(Oγ)···Phe33(O)	2.80 ± 0.2 (2.78)	3.14 ± 0.4 (3.06)	2.89 ± 0.4 (2.77)	2.77 ± 0.2 (2.75)
Thr35(Oγ)···Leu215(O)	3.28 ± 0.6 (3.13)	4.15 ± 0.3 (4.16)	3.79 ± 0.8 (3.60)	4.00 ± 0.5 (4.01)
Thr35(Oγ)···Asp216(Oδ*)		2.76 ± 0.2 (2.74)	3.65 ± 0.7 (3.78)	2.88 ± 0.4 (2.73)
SerP2(Oγ)···Thr219(Oγ)	3.42 ± 0.9 (2.88)	4.03 ± 0.8 (4.27)	3.66 ± 0.9 (3.80)	3.31 ± 0.8 (2.86)
Thr219(Oγ)···Asp216(Oδ*)	2.64 ± 0.1 (2.63)	2.76 ± 0.26 (2.74)	2.69 ± 0.2 (2.67)	2.73 ± 0.2 (2.71)

[a] Heteroatom distances of hydrogen bond networks close to the active site. The numbers in each cell give the means, standard deviations, and medians of each distance. The Asp216(Oδ*) notation implies that either the δ1 or the δ2 carboxylate oxygen forms a persistent hydrogen bond.

explained by a slightly altered conformation of κ-casein close to this region in this complex. The conformational change, however, is not as dramatic as seen in the BOV/CAM complex (see Figure 4 as well as φ- and ψ-angles of the four κ-casein fragments provided in the Supporting Information). In general, the *P3*−*P2′* residues are very stable for all four complexes, whereas the residues further out in the binding cleft are found to be more flexible, indicating that they are not as tightly bound. As shown in the C-terminal part of the enzymes in the BOV/CAM and CAM/BOV complexes, the N-terminal part of the κ-casein fragment is more flexible in these structures. The dynamics of the *P4* residue is slightly depressed in the complexes with camel κ-casein. This could be due to the *P3* residue being a proline in camel κ-casein compared to a leucine in the bovine variant, which restricts the motion of the backbone of the *P4* residue.

Cleavage Site. The conformation of the active site for each complex was evaluated according to a proposed catalytic mechanism of Veerapandian et al. and James et al.,[54,55] in line with our previous paper on modeling the BOV/BOV complex.[18] The catalytic mechanism involves concurrent activation of the water molecule in the active site by Asp216, nucleophilic attack of the water on the carbonyl carbon atom, and protonation of the scissile carbonyl bond of κ-casein by Asp34 (see Table 4). The average distance and angle of nucleophilic attack between the catalytic water molecule and the *PheP1*(C), along with the hydrogen-bonding distance between Asp34(Oδ2) and *PheP1*-(O), are listed in Table 4 for all four complexes. The hydrogen bond between Asp34(Oδ2) and *PheP1*(O) is largely stable throughout simulation for all four complexes. The angle of nucleophilic attack is on average within 10° of the Bürgi−Dunitz angle for all but the BOV/CAM complex. The distance of attack is also longer for the BOV/CAM complex than for the other three complexes, because the position of the catalytic water molecule in the BOV/CAM simulation varies significantly throughout the simulation. This water molecule is observed to adopt two positions predominantly, either where it is placed below the two catalytic residues Asp34 and Asp216, forming a hydrogen bond chain Asp34(Oδ1)···H−O−H···Asp216(Oδ1), or where the catalytic water is hydrogen bonding to the *PheP1*(O) and either Asp216(Oδ1) or Asp216(Oδ2).

Hydrogen-Bonding Network. The average distances of the four hydrogen bonds making up the fireman's grip[56] are listed in Table 5 for all four simulated complexes. The fireman's grip is crucial for keeping the two lobes of the enzyme together and for stabilizing the enzyme active site. The two main hydrogen bonds in which Thr35(NH) and Thr217(NH) are hydrogen bond donors to Thr217(Oγ) and Thr35(Oγ), respectively, are present in all four simulated complexes. Although these hydrogen bonds are relatively long, the hydrogen bond angles (D−H···A) are very favorable (not shown). Thr217(Oγ) forms a hydrogen bond to Phe33(O), which is present for all complexes throughout the simulations. The hydrogen bond between Thr35(Oγ) and Leu215(O) is stable in the BOV/BOV complex. In both cross-complexes (BOV/CAM and CAM/BOV), an alternate hydrogen bond between Thr35(Oγ) and either oxygen in the carboxyl group of Asp216(Oδ) is formed. The side chain of Asp216 rotates and thus switches which oxygen atom Thr35 is bonded to, but the bond is persistent to either of the two. In CAM/CAM both of these bonds are seen, and thus the standard deviation of the average bond lengths is rather large for the two bonds. In our previous paper we observed a hydrogen bond chain from the *SerP2*(Oγ) to Thr219(Oγ) and from there continuing to Asp216(Oδ). The latter of the hydrogen bonds is present throughout all four simulations. The former is, however, more intermittent in the BOV/CAM and CAM/CAM simulations.

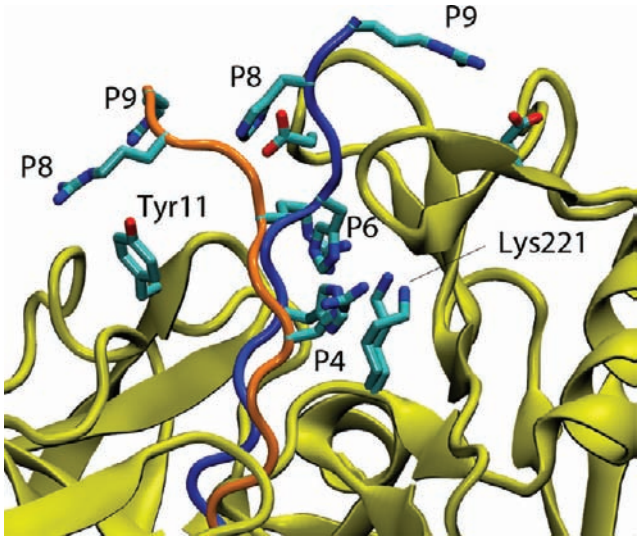


Figure 6. Backbone conformation of the camel κ-casein fragment (orange) in bovine chymosin at 96 ns compared to the bovine κ-casein fragment (blue). The P9, P8, P6, and P4 residues are shown on each variant to show their positions compared to each other. Hydrogen atoms are not shown.

Alternatively, *SerP2*(Oγ) forms a water-assisted hydrogen bond to the backbone carbonyl group of Lys221.

Contacts between Chymosin and κ-Casein Fragments. As both variants of the κ-casein fragment bind in very similar conformations in both variants of chymosin, the regions of contact between the enzyme and the ligand are mostly the same. The greatest difference is the binding of the N-terminus of the κ-casein fragment in the BOV/CAM complex. This binds in a conformation that differs significantly from that of the other three complexes after 96 ns of MD. In the initial homology model, it binds in a slightly different manner from the BOV/BOV complex. After 13 ns, the camel κ-casein fragment forms an intermittent hydrogen bond between the *ArgP8* (donor) and Tyr11 (acceptor) of bovine chymosin (average distance *ArgP8*-(NH1)···Tyr11(Oγ) 4.29 ± 1.34 Å). The change in conformation is likely caused by the *HisP8Arg* mutation; both residues are modeled in their charged tautomer forms, but the side chain of arginine is longer than histidine. The movement of the *P8* residue pulls the *P9* residue along, and eventually it loses the interaction with the C-terminal domain of bovine chymosin. Importantly, although the binding of the *P9* and *P8* residues changes, the interactions in the remaining pockets are largely unchanged (see Figures 4 and 6). The binding of the camel κ-casein fragment compared to the bovine κ-casein fragment in bovine chymosin introduces a positive charge in the *P4* position, which intuitively seems unfavorable because the S4 pocket already holds a positive charge from Lys221. The position of the *P4* residue side chain is nonetheless roughly unaltered (see Figure 6). Instead, the positive charge on Lys221 is pushed farther away from the *P4* residue (a neutral His compared to Arg); this is evident in the frequency of close contacts between the side chains of these two residues, which drops from 92 to 57%. The frequencies are calculated by how often during the simulation the side-chain heavy atoms of the two residues are within 4 Å (for a complete scheme of frequency of interaction between *Px* residues and Sx enzyme pockets of all four complexes, see the Supporting Information). In BOV/BOV the pK_a of the *HisP4* residue calculated by PROPKA[57,58] is lowered by 2−3 units due to the interaction with Lys221, which precludes the *HisP4* residue from being in the histidinium tautomer (positively charged), and it is thus modeled as a neutral His in the δ-tautomer. Both *HisP8* and *HisP6* of bovine κ-casein on the other hand are modeled as charged; they are both surface exposed, and their pK_a values are unperturbed by their surroundings upon binding. Introducing an arginine in the *P4*-position causes the S4 pocket of bovine chymosin to change during the simulation. Also in the BOV/CAM complex the *P1′*-position is different compared to BOV/BOV with a *MetP1′Ile* variant change. The difference, which introduces a shorter aliphatic side chain and removes the electronegative and polarizable sulfur atom, results in a loss of the contact between the Gln294 residue and the *P1′* residue, which was present 45% of the time in the BOV/BOV simulation.

In the CAM/CAM complex, the S4 pocket contains a valine at position 221, compared to a lysine residue in bovine chymosin. This is reflected in the fact that *ArgP4* forms a constant contact with this residue (95% of the time during the simulation), compared to Lys221 and *ArgP4*, which interacted only 57% of the time in the BOV/CAM complex. Thus, removal of the positive charge increases the interaction frequency between these two positions, even though valine is much smaller and thus is unable to reach into the pocket as much as lysine can. In the S2 pocket a Val223Phe difference causes the *SerP2* to increase its frequency of interaction with the Phe223 residue, likely because it can now reach the *P2* residue side chain. In the S1 pocket the Leu32Val difference changes the interaction with the *PheP1* residue, as the Val32 residue sits lower in the S1 pocket and thus cannot reach the *PheP1* residue. In the S3′ pocket the interaction between *IleP3′* and Glu294 is almost nonexistent compared to Gln294 in bovine chymosin, which interacts with the *IleP3′* frequently.

The effect of the *LeuP3Pro* difference seems to have little impact on the structure. The position of the side chain is in the same pocket area (see the interacting residues in the Supporting Information). Both leucine and proline are hydrophobic residues, and thus this interaction is retained in the pocket. In BOV/BOV the backbone of *LeuP3* takes part in two hydrogen bonds with Ser220, Ser220(NH)···*LeuP3*(O), and *LeuP3*(NH)···Ser220(Oγ). The change from leucine to proline at *P3* does not seem to alter the formation of the former hydrogen bond in any of the four complexes. Proline residues do not hold a NH hydrogen atom, and thus the latter hydrogen bond cannot form, but no structural changes in the models during the MD simulations BOV/CAM and CAM/CAM are observed.

Conserved Water Molecules. Our previous model of the BOV/BOV complex incorporated all of the crystallographic waters from the 3CMS crystal structure and including the seven conserved water molecules identified from a comparative study of chymosin and nine related aspartic proteases.[59] The BOV/BOV simulation has now been extended to 96 ns, and four of the structurally conserved water molecules in the cleft, including the catalytic water molecule, are still stable throughout the entire simulation. One water molecule is exchanged (Wat1007) with the bulk solvent after about 80 ns, but a new water molecule enters the same position. The water molecule (Wat1125) was shown to be displaced by the *LysP6′* residue during MD simulations[18] and was not incorporated in the other three models. The displacement would be expected to provide an entropic benefit and might provide a clue to the observation that *LysP6′* is more

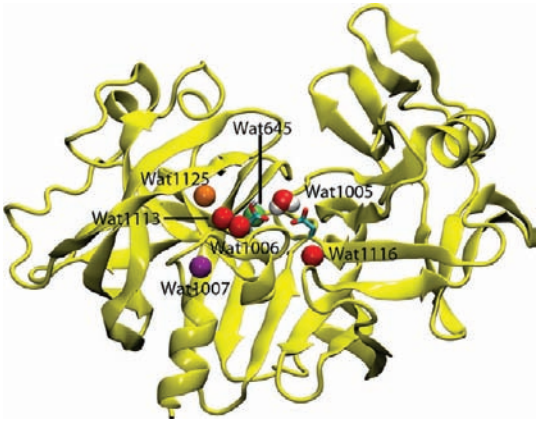


Figure 7. Structural water molecules in chymosin colored according to their behavior during MD of BOV/BOV. The red water molecules maintain their position during 96 ns of production MD; only the catalytic water molecule has it hydrogen atoms displayed. Wat645 (light green) moves away from its original position. Wat1007 (purple) is exchanged, but a water molecule always occupies this position. Wat1125 (orange) was displaced by *LysP6* during 20 ns MD of the BOV/BOV complex.

important for binding than *LysP7′*, although they are both shown to improve catalysis.[25] The last conserved water molecule (Wat645) was previously shown to become solvent exchanged[18] in accordance with the observation that this water molecule is present in only one chymosin crystal structure, in which an occupancy of 51% was found.[6] In the other three modeled complexes, the same four water molecules are also very stable, whereas Wat1007 is exchanged with the bulk solvent (see Figure 7).

DISCUSSION

Using, as a starting point, our previous model of bovine chymosin complexed with the *P9*−*P7′* fragment of bovine κ-casein[18] bound in the substrate cleft, we have developed models of the corresponding BOV/CAM, CAM/CAM, and CAM/BOV complexes using homology modeling. Homology modeling was necessary as there is currently no available structure of camel chymosin and no structural information about κ-casein is available when bound to chymosin. Comparing structures of homologous structures, Chothia and Lesk determined that proteins with a sequence identity of >50% have very low RMSDs and thus are very similar.[27] The sequence identity between bovine and camel chymosin is 85%, and the two enzymes are expected to have the same overall fold, something also observed in structures of the aspartic protease, cathepsin D, from bovine and human.[12] On the basis of the number of different or divergent residues the RMSD value of the main-chain atoms (N, C_α, C, O) should be around 0.53 Å for these structures based on a fitted measure by Chothia and Lesk.[27] The RMSD of the C_α atoms between the template structure and the three models is under 0.25 Å in each case, which is likely also due to the sequence similarity being 94%. Molecular dynamics simulations were performed on the three models for 96 ns, and the previous BOV/BOV simulation was extended from 20 to 96 ns. The RMSD values all converged at values of <2.0 Å. The κ-casein fragment binds in an extended conformation in all four cases in agreement with solution NMR and circular dichroism data[60] and with a wealth of crystal structures of aspartic proteases complexed with ligands and peptide mimics. We note that Thurn et al. have measured the radius of gyration of κ-casein in pure κ-casein micelles to be 70 Å,[61] which compared to the length of these fragments (around 45 Å) could suggest that some conformational change is required for binding. Thurn et al. have shown these micelles to have a sphere-like center and more flexible chains dangling at the periphery.[61] Whereas the structure of casein micelles, present in milk, is still being debated, it is generally agreed that κ-casein fragments are dangling from the surface of these micelles in a similar manner.[62] This may explain how this fragment is able to bind in an extended form, but no final conclusion can be drawn. Hydrogen-bonding networks, in particular the fireman's grip,[56] were stable throughout all four simulations. The *P9*−*P1* residues in the κ-casein fragment are observed to interact with the C-terminal domain of chymosin in the BOV/BOV, CAM/CAM, and CAM/BOV models, whereas this part of the fragment is more flexible in the BOV/CAM complex. That the camel κ-casein fragment is more flexible than the bovine κ-casein fragment when complexed to bovine chymosin could signify a lowered binding affinity, which was found experimentally by Kapeller et al. using undecapeptide analogues of the two κ-casein variants with bovine chymosin.[4] The *P1′*−*P7′* residues contact the N-terminal domain of chymosin similarly in all four complexes. The structure in the active site was evaluated on the basis of the geometry required in the first step of the reaction mechanism proposed by James et al.[55] The correct geometry was predominant in the BOV/BOV, CAM/CAM, and CAM/BOV simulations, whereas it was observed only intermittently in the BOV/CAM simulation. The BOV/CAM model, thus, seems to behave differently from the other models, which could be correlated with the experimental finding that bovine chymosin is poor at clotting camel's milk.[4]

The clotting process is sensitive to environmental factors such as pH, temperature, and salt concentration.[5] The clotting ability is determined by interactions between the enzyme and substrate, which assists the correct binding of the substrate in the active site, and therefore the pockets around the active site have been probed to determine their specificity.[19] The differences observed between the four models provide clues about which pockets contribute significantly to the binding specificity. In the BOV/BOV complex the *HisP4* interacts with the side chain of Lys221 in the S4 pocket. In the BOV/CAM complex the *ArgP4* and Lys221 positive charges are in close vicinity, causing the Lys221 side chain to move significantly deeper into the pocket away from the arginine. A *P4 His102Lys* mutant of bovine κ-casein has been shown to be a particularly disfavored substrate for bovine chymosin,[25] and thus it makes sense that *ArgP4* is not favored either. This interaction between positive charges is removed in the CAM/CAM complex, as residue 221 is a valine in the camel variant, with which arginine is observed to be in close contact. Finally, in the CAM/BOV complex, *HisP4* is also observed to interact with Val221. *HisP4* was modeled as neutral;[18] however, in the CAM/BOV complex, it would also be possible for *HisP4* to bind in its positively charged histidinium tautomer form, because no positive residues are found in the S4 pocket of camel chymosin.

Previously, it has been suggested that Lys221 interacts with the *P2* residue.[21,56] We do not observe any direct interaction between the side chain of Lys221 and the *P2* residue of the κ-casein fragment in any of the four complexes. The backbone of Lys221 can interact with the *P2* residue via water-mediated hydrogen bonds, but this is not an interaction that is specific to

the lysine side chain. Suzuki et al. investigated the cleavage of a peptide with a glutamic acid at *P2*, the reasoning being that the positive charge on Lys221 should favorably interact with the negative charge on *GluP2*. Comparing native chymosin and an engineered Lys221Leu mutant resulted in no change in binding affinity for a synthetic *P5*−*P3′* peptide (Lys-Pro-Ile-Glu-Phe-Phe(4-NO_2)-Arg-Leu-OH),[63] thus supporting limited interaction between the side chains of these two residues.

Dunn et al. have suggested that the S3 pocket for aspartic proteases accounts for a lot of the specificity in the binding of substrates, where in particular the mammalian aspartic proteinases were more sensitive than the microbial aspartic proteinases.[24] They also suggested that a proline at the *P3* position was very unfavorable, using six different synthetic octapeptide substrates. This suggestion can help to explain why bovine chymosin is poor at clotting camel's milk, because camel κ-casein contains a proline exactly at *P3*. The paradox then is that camel chymosin does not have a problem clotting camel's milk, and there are no differences in the amino acids composing the S3 pocket between bovine and camel chymosin. The only difference we observe for *ProP3* is that only one hydrogen bond to the backbone is possible as the proline residue does not contain a hydrogen atom on the backbone nitrogen. It therefore could be that the synthetic octapeptide with a *ProP3* is not a good substrate, not just because of the proline residue alone but also the remaining residues in the substrate. Recently, Li et al. using the *Mucor pusilllus* pepsin suggested that the residues Leu13-Glu14-Glu15 found in a loop (sometimes referred to as the S10 loop, which is not to be confused with an S-pocket) determine the specificity in the S3 pocket and that in particular Leu13 makes the binding pocket more hydrophobic compared to the bovine chymosin counterparts Asp13-Ser14-Gln15.[64] In bovine chymosin, however, residue 13 does not form part of the S3 binding pocket, whereas in *M. pusilllus* pepsin the conformation of the loop containing these three residues makes a sharp turn and thus Leu13 helps form the pocket. Residue 13 is also a glutamic acid in pig pepsin, whereas in camel chymosin it is a glutamine and in *Endothia parasitica* proteinase it is an alanine residue. Thus, it seems, in line with observations by Dunn et al.,[24] that the residues forming the S3 pocket are different in microbial aspartic proteinases and in mammalian ones. Thus, it is more likely that altering the Gln15 to a more hydrophobic residue would have a higher impact on the binding affinity for mammalian aspartic proteinases and in particular bovine and camel chymosin.

In the S2 pocket a Val223Phe change between bovine and camel chymosin is located, making the pocket smaller. The *SerP2* is not affected by this in any of our camel chymosin models. *SerP2* is observed to form a hydrogen bond network through the Thr219(Oγ) to one of the carboxylates in Asp216, which is in accordance with earlier mutational studies by Pitts and Mantafounis on the importance of Thr219 to the activity of chymosin.[65]

The difference at S1, Leu32Val, between bovine and camel chymosin is of minor significance. It does not significantly alter the physical and chemical properties at this pocket, and the two residues are almost the same size, which is crucial as the pocket needs to accommodate the aromatic *PheP1*. Experimentally, it has been determined that aromatic residues are preferred at *P1* in bovine chymosin, whereas microbial aspartic proteases require a less specific binding and can accommodate both lysine and methionine.[19]

The *MetP1′Ile* change between bovine and camel κ-casein is also of minor significance, as it has been shown that bovine chymosin will cleave human and porcine κ-casein (*PheP1*−*IleP1′*) as well as rat and mouse κ-casein (*PheP1*−*LeuP1′*). Moreover, it has been demonstrated that a mutant of bovine κ-casein with *MetP1*−*PheP1′* as the scissile bond is cleaved 1.8 times more quickly than the wild type.[20] It has previously been suggested by Dalgleish that natural variations in the sequence of κ-casein are a factor in clotting ability,[66] and with the models developed here we can only support this view.

Further studies are currently ongoing in our laboratory aimed at estimating the important residues based on contributions to the binding free energy between chymosin and κ-casein fragments in the four complexes reported here as well as deciphering the allosteric activation.

■ ASSOCIATED CONTENT

Supporting Information. Material on the extended minimization and equilibration protocol used on the homology models before production MD (section 1), full sequence alignments of bovine and camel chymosin and κ-casein (section 2), Ramachandran plots of the 3CMS crystal structure and the four complexes between chymosin and κ-casein fragments (section 3), φ- and ψ-angles of the κ-casein fragments at 96 ns in each model (section 4), as well as tables describing the interactions between the chymosin and κ-casein and the frequency of these interactions (section 5). This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: +4589423953. Fax: +4586196199. E-mail: birgit@chem.au.dk.

Present Addresses

#Max Planck Institute for Mathematics in the Sciences, Inselstrasse 22, DE-04103 Leipzig, Germany.

Funding Sources

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 Sciences, 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.

■ ACKNOWLEDGMENT

We thank Heidi Koldsø for assistance with the homology modeling.

■ ABBREVIATIONS USED

PDB, Protein Data Bank; RMSD, root-mean-square deviation; RMSF, root-mean-square fluctuation; BOV/BOV, bovine chymosin−bovine κ-casein complex; BOV/CAM, bovine chymosin−camel κ-casein complex; CAM/CAM, camel chymosin−camel κ-casein complex; CAM/BOV, camel chymosin−bovine κ-casein complex; MD, molecular dynamics.

REFERENCES

(1) Harboe, M.; Broe, M. L.; Qvist, K. B. The production, action and application of rennets and coagulants. In *Technology of Cheesemaking*, 2nd ed.; Law, B. A., Tamime, A. Y., Eds.; Wiley-Blackwell: Oxford, U.K., 2010; pp 98−129.

(2) Foltmann, B.; Pedersen, V. B.; Kauffman, D.; Wybrandt, G. The primary structure of calf chymosin. *J. Biol. Chem.* **1979**, *254*, 8447–8456.

(3) 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 κ-casein. *FEBS Lett.* **1996**, *379*, 60–62.

(4) 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. Commun.* **2006**, *342*, 647–654.

(5) Bansal, N.; Fox, P. F.; McSweeney, P. L. Factors that affect the aggregation of rennet-altered casein micelles at low temperatures. *Int. J. Dairy Technol.* **2008**, *61*, 56–61.

(6) Gilliland, G. L.; Winborne, E. L.; Nachman, J.; Wlodawer, A. The three-dimensional structure of recombinant bovine chymosin at 2.3 Å resolution. *Proteins* **1990**, *8*, 82–101.

(7) Strop, P.; Sedlacek, J.; Stys, J.; Kaderabkova, Z.; Blaha, I.; Pavlickova, L.; Pohl, J.; Fabry, M.; Kostka, V. Engineering enzyme subsite specificity: preparation, kinetic characterization, and x-ray analysis at 2.0-Å resolution of Val111Phe site-mutated calf chymosin. *Biochemistry* **1990**, *29*, 9863–9871.

(8) 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**, *221*, 1295–1309.

(9) Groves, M. R.; Dhanaraj, V.; Badasso, M.; Nugent, P.; Pitts, J. E.; Hoover, D. J.; Blundell, T. L. A 2.3 Å resolution structure of chymosin complexed with a reduced bond inhibitor shows that the active site β-hairpin flap is rearranged when compared with the native crystal structure. *Protein Eng.* **1998**, *11*, 833–840.

(10) 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.

(11) Madala, P. K.; Tyndall, J. D. A.; Nall, T.; Fairlie, D. P. Update 1: Proteases universally recognize β strands in their active sites. *Chem. Rev.* **2010**, *110*, PR1–PR31.

(12) Metcalf, P.; Fusek, M. Two crystal structures of cathepsin D: the lysosomal targeting signal and active site. *EMBO J.* **1993**, *12*, 1293–1302.

(13) Brignon, G.; Chtourou, A.; Ribadeau-Duma, B. Preparation and amino acid sequence of human κ-casein. *FEBS Lett.* **1985**, *188*, 48–54.

(14) Schechter, I.; Berger, A. On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **1967**, *27*, 157–162.

(15) Li, M.; Phylip, L. H.; Lees, W. E.; Winther, J. R.; Dunn, B. M.; Wlodawer, A.; Kay, J.; Gustchina, A. The aspartic proteinase from *Saccharomyces cerevisiae* folds its own inhibitor into a helix. *Nat. Struct. Biol.* **2000**, *7*, 113–117.

(16) Plowman, J. E.; Creamer, L. K. Restrained molecular dynamics study of the interaction between bovine κ-casein peptide 98−111 and bovine chymosin and porcine pepsin. *J. Dairy Res.* **1995**, *62*, 451–467.

(17) Plowman, J. E.; Creamer, L. K.; Smith, M. H.; Hill, J. P. Restrained molecular dynamics investigation of the differences in association of chymosin to κ-caseins A and C. *J. Dairy Res.* **1997**, *64*, 299–304.

(18) 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.

(19) Chitpinityol, S.; Crabbe, M. J. C. Chymosin and aspartic proteinases. *Food Chem.* **1998**, *61*, 395–418.

(20) Fox, P. F.; Guinee, T. P.; Cogan, T. M.; McSweeney, P. L. H. Enzymatic coagulation of milk. In *Fundamentals of Cheese Science*; Aspen Publishers: Gaithersburg, MD, 2000; pp 98−135.

(21) Dunn, B. M.; Valler, M. J.; Rolph, C. E.; Foundling, S. I.; Jimenez, M.; Kay, J. The pH dependence of the hydrolysis of chromogenic substrates of the type, Lys-Pro-Xaa-Yaa-Phe-(NO_2)Phe-Arg-Leu, by selected aspartic proteinases: evidence for specific interactions in subsites S3 and S2. *Biochim. Biophys. Acta* **1987**, *913*, 122–130.

(22) Hill, R. D. Synthetic peptide and ester substrates for rennin. *J. Dairy Res.* **1969**, *36*, 409–415.

(23) Hill, R. D. The nature of the rennin-sensitive bond in casein and its possible relation to sensitive bonds in other proteins. *Biochem. Biophys. Res. Commun.* **1968**, *33*, 659–663.

(24) Dunn, B. M.; Jimenez, M.; Parten, B. F.; Valler, M. J.; Rolph, C. E.; Kay, J. A systematic series of synthetic chromophoric substrates for aspartic proteinases. *Biochem. J.* **1986**, *237*, 899–906.

(25) Visser, S.; Slangen, C. J.; van Rooijen, P. J. Peptide substrates for chymosin (rennin). Interaction sites in κ-casein-related sequences located outside the (103−108)-hexapeptide region that fits into the enzyme's active-site cleft. *Biochem. J.* **1987**, *244*, 553–558.

(26) Macheboef, D.; Coulon, J.; D'Hour, P. Effect of breed, protein genetic variants and feeding on cows' milk coagulation properties. *J. Dairy Res.* **1993**, *60*, 43–54.

(27) Chothia, C.; Lesk, A. M. The relation between the divergence of sequence and structure in proteins. *EMBO J.* **1986**, *5*, 823.

(28) Larkin, M. A.; Blackshields, G.; Brown, N. P.; Chenna, R.; McGettigan, P. A.; McWilliam, H.; Valentin, F.; Wallace, I. M.; Wilm, A.; Lopez, R.; Thompson, J. D.; Gibson, T. J.; Higgins, D. G. ClustalW and ClustalX version 2. *Bioinformatics* **2007**, *23*, 2947–2948.

(29) Uniprot Knowledgebase, 2010.

(30) 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.; Yeh, L. L. The Universal Protein Resource (UniProt). *Nucleic Acids Res.* **2005**, *33*, D154–159.

(31) Kiefer, F.; Arnold, K.; Künzli, M.; Bordoli, L.; Schwede, T. The SWISS-MODEL Repository and associated resources. *Nucleic Acids Res.* **2009**, *37*, D387–D392.

(32) Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **1993**, *26*, 283–291.

(33) Benkert, P.; Tosatto, S. C. E.; Schomburg, D. QMEAN: a comprehensive scoring function for model quality assessment. *Proteins* **2008**, *71*, 261–277.

(34) Holm, L.; Sander, C. Evaluation of protein models by atomic solvation preference. *J. Mol. Biol.* **1992**, *225*, 93–105.

(35) Wallner, B.; Elofsson, A. Can correct protein models be identified?. *Protein Sci.* **2003**, *12*, 1073–1086.

(36) Case, D. A.; Cheatham, T. E., III; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26*, 1668–1688.

(37) Powers, N.; Jensen, J. H. Chemically accurate protein structures: validation of protein NMR structures by comparison of measured and predicted pKa values. *J. Biomol. NMR* **2006**, *35*, 39–51.

(38) 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.

(39) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802.

(40) 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.; Kollman, P. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **2003**, *24*, 1999–2012.

(41) Hoover, W. G. Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A* **1985**, *31*, 1695.

(42) Nosé, S.; Klein, M. L. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* **1983**, *5*, 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.; 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. Comput. Phys.* **1977**, *23*, 327–341.
(48) Weinbach, Y.; Elber, R. Revisiting and parallelizing SHAKE. *J. Comput. Phys.* **2005**, *209*, 193–206.
(49) Darling, D. F.; Dickson, J. The determination of the zeta potential of casein micelles. *J. Dairy Res.* **1979**, *46*, 329.
(50) Fiser, A.; Sali, A. Modeller: generation and refinement of homology-based protein structure models. In *Methods in Enzymology*; Carter, C. W., Jr., Sweet, R. M., Eds.; Academic Press: New York, 2003; Vol.374, pp 461−491.
(51) Sali, A.; Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **1993**, *234*, 779–815.
(52) Siew, N.; Elofsson, A.; Rychlewski, L.; Fischer, D. MaxSub: an automated measure for the assessment of protein structure prediction quality. *Bioinformatics* **2000**, *16*, 776–785.
(53) Cristobal, S.; Zemla, A.; Fischer, D.; Rychlewski, L.; Elofsson, A. A study of quality measures for protein threading models. *BMC Bioinformatics* **2001**, *2*, 5.
(54) Veerapandian, B.; Cooper, J. B.; Scaronali, 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.
(55) James, M. N.; Sielecki, A. R.; Hayakawa, K.; Gelb, M. H. Crystallographic analysis of transition state mimics bound to penicillopepsin: difluorostatine- and difluorostatone-containing peptides. *Biochemistry* **1992**, *31*, 3872–3886.
(56) Pearl, L.; Blundell, T. The active site of aspartic proteinases. *FEBS Lett.* **1984**, *174*, 96–101.
(57) Li, H.; Robertson, A. D.; Jensen, J. H. Very fast empirical prediction and rationalization of protein pK_a values. *Proteins* **2005**, *61*, 704–721.
(58) Bas, D. C.; Rogers, D. M.; Jensen, J. H. Very fast prediction and rationalization of pK_a values for protein−ligand complexes. *Proteins* **2008**, *73*, 765–783.
(59) Prasad, B. V. L. S.; Suguna, K. Role of water molecules in the structure and function of aspartic proteinases. *Acta Crystallogr. D* **2002**, *58*, 250–259.
(60) Plowman, J. E.; Creamer, L. K.; Liddell, M. J.; Cross, J. J. Solution conformation of a peptide corresponding to bovine κ-casein B residues 130−153 by circular dichroism spectroscopy and ^{1}H-nuclear magnetic resonance spectroscopy. *J. Dairy Res.* **1997**, *64*, 377–397.
(61) Thurn, A.; Burchard, W.; Niki, R. Structure of casein micelles. 1. Small-angle neutron-scattering and light-scattering from β-casein and κ-casein. *Colloid Polym. Sci.* **1987**, *265*, 653–666.
(62) Dalgleish, D. G. On the structural models of bovine casein micelles—review and possible improvements. *Soft Matter* **2011**, DOI: 0.1039/C0SM00806K.
(63) Suzuki, J.; Hamu, A.; Nishiyama, M.; Horinouchi, S.; Beppu, T. Site-directed mutagenesis reveals functional contribution of Thr218, Lys220 and Asp304 in chymosin. *Protein Eng.* **1990**, *4*, 69–71.
(64) Li, T.; Wang, J.; Li, Y.; Zhang, L.; Zheng, L.; Li, Z.; Yang, Z.; Luo, Q. Structure of the complex between *Mucor pusillus* pepsin and the key domain of κ-casein for site-directed mutagenesis: a combined molecular modeling and docking approach. *J. Mol. Model.* **2011**, DOI: 10.1007/s00894-010-0869-3.
(65) Pitts, J. E.; Mantafounis, D. Protein engineering of chymosin − tailoring the pH profile. *Biochem. Soc. Trans.* **1990**, *18*, 920–921.
(66) Dalgleish, D. G. The enzymatic coagulation of milk. In *Advanced Dairy Chemistry*, 2nd ed.; Fox, P. F., Ed.; Elsevier Science Publishers: London, U.K., 1992; Vol.1, pp 579−619.