JCTC Journal of Chemical Theory and Computation

Effects of Calcium Binding on Structure and Autolysis Regulation in Trypsins. A Molecular Dynamics Investigation

Elena Papaleo, Piercarlo Fantucci, and Luca De Gioia*

Department of Biotechnology and Biosciences, University of Milano-Bicocca, P.za della Scienza 2, 20126 Milan, Italy

Received April 6, 2005

Abstract: The calcium ion was proposed to be involved in protein structure stabilization against thermal and proteolytic degradation, such as autolysis phenomena, in trypsin-like serine proteases. However, molecular details related to the role played by the metal ion are still largely unknown. Several molecular dynamics simulations of 6 ns have been used to investigate the dynamic behavior of bovine and salmon trypsins in calcium-bound and calcium-free forms, with the aim of evaluating the role of the calcium ion in trypsin three-dimensional structure and autoproteolysis propensity. It turned out that the calcium-free trypsins are characterized by a more flexible structure, revealing structure−function relationships connecting Ca^{2+} binding and autoproteolysis propensity. In particular, the removal of Ca^{2+} not only increases the flexibility of regions around its binding site, in the N-terminal domain, but also leads to channeling of the fluctuations to remote sites in the C-terminal domain, possibly involving the interdomain loop. Two primary autolysis sites are strongly influenced by calcium binding (R117 and K188) in bovine trypsin, whereas Ca^{2+} plays a less crucial role in salmon trypsin.

Introduction

Trypsin, a member of the serine protease family that is highly specific for cleavage at lysine and arginine residues, plays an important role in many biological processes. The large number of studies carried out on trypsins in their native form, as well as in the presence of a variety of ligands, allowed an understanding of fundamental issues related to substrate binding, specificity, and fine details of the catalytic mechanism.[1] Trypsins and several other eukaryotic serine proteases bind a Ca^{2+} ion, which is known to play an important role as a regulator of physiological functions and in the preservation of structural integrity in many proteins.[2] In particular, the calcium ion in trypsin-like serine proteases has been proposed to be involved in the stabilization against thermal and proteolytic degradation, such as autolysis phenomena.[3,4] The elucidation of factors regulating autolysis propensity is particularly relevant since the inability of trypsin to self-degradate has been linked to human hereditary pancreatitis, a genetic disorder due to inappropriate activation of trypsin within the pancreas.[5]

Even though several structural details related to Ca^{2+} binding to trypsins are known,[1] the detailed role played by Ca^{2+} in structure stabilization and autolysis is not yet fully understood. The calcium-binding loop runs from residue 69 to 80 (numeration of bovine trypsin, BT) and connects two antiparallel β strands in the N-terminal domain. The loop structure is maintained by hydrogen bonds formed with other parts of the protein. In addition, several water molecules form bridges between the loop and other protein regions.[3] The role played by Ca^{2+} in trypsin structure has been investigated using several spectroscopic techniques,[6−10] leading to the conclusion that the protein assumes a less-compact structure when the metal ion is removed. This behavior has been correlated to modifications in the environment of a conserved tryptophan residue (W141),[9] even though FT-IR spectroscopy did not show detectable perturbation of the protein secondary structure.[10]

The analysis of trypsin structures obtained by X-ray diffraction (mammalian and salmon trypsins) revealed that

* Corresponding author. Phone: +39 02 64483463. E-mail: luca.degioia@unimib.it.

10.1021/ct050092o CCC: $30.25
Published on Web 10/05/2005

the calcium-binding loop interacts with a protein portion usually described as the autolysis loop (residues 143−155, BT numeration).[3] This loop owes its name to its similarity with chymotrypsins, but it is still unclear if it constitutes a conserved primary autolysis site in trypsins. In fact, a well-known autoproteolytic site in the BT autolysis loop (K145) is not conserved in salmon trypsin (ST), where the closest putative autolysis site (K154) seems to be too buried and rigid to be a favorable autolysis site.[3] Moreover, several other autolysis sites have been characterized in mammalian trypsins.[11,12] The peptide bonds between R117−V118 and K145−S146 are simultaneously cleaved, leading to an enzyme form that retains activity.[12] Cleavage can also takes place at K60−S61[12] and K188−D189,[11] leading to enzyme inactivation. Finally, it was also shown that the autolysis of wild-type rat trypsin begins with cleavage at R117, and K60 is another relevant autolysis site.[13,14] Notably, the peptide segment between K60 and R117, which contains the so-called interdomain loop that connects the two globular protein domains, is part of the longest peptide portion in trypsin that does not include disulfide bridges.[1] It has been suggested that this region, which includes the Ca^{2+}-binding loop, may function as a built-in target for autolysis: the cleavage at any end of the peptide segment K60−R117 is followed by cleavages at several trypsin-sensitive sites between the two ends.[13] It was also proposed that Ca^{2+} could stabilize an autolysis-resistant conformation of the K60−R117 segment.[13,15] Interestingly, ST and other fish trypsins[3,16−18] are characterized by autolysis rates that are less dependent on the presence of calcium ions than those of mammalian trypsins. However, it has been noted that this difference does not necessarily indicate that Ca^{2+} has a lower stabilizing effect in ST, since slower autolysis rates could be due to the presence of fewer exposed autolysis sites.[3] As for the other factors that can affect autolysis propensity, Fontana et al.[19] have suggested that protein portions characterized by large flexibility (evaluated on the basis of B-factor values) are particularly susceptible to proteolytic attack, whereas Novotny and Bruccoleri[20] concluded that solvent accessibility is more important than flexibility in affecting proteolysis.

Molecular dynamics (MD) is a powerful tool used to understand the structure and functional features of a protein in atomic detail, and it can lead to significant insights into the atomic machinery underlying protein function.[21−23] In fact, MD simulations on trypsins from organisms adapted to different environmental conditions have been previously reported, revealing details about the relationship between protein dynamics and functions.[24,25] However, the role played by Ca^{2+} on the dynamic properties of trypsins has never been investigated.

With the aim of clarifying the role of calcium on trypsin structure and autoproteolysis propensity, we have carried out multiple 6 ns MD simulations of ST and BT proteins, in their native holo (calcium-bound) and apo (calcium-free) forms. Analogous approaches were applied successfully to investigate the role of calcium in α-lactalbumin and c-type lysozyme,[26] parvalbumin,[27] and calmodulin.[28,29]

Sequence and structure characteristics of ST and BT relevant to the proper analysis of results from the MD simulations are presented in the first part of the manuscript, whereas the results and analysis of MD simulations are discussed in the second section.

Methods

To collect known structures and sequences of mammalian and fish trypsins, the Protein Data Bank (PDB)[30] and the nonredundant sequence database were searched with Blast[31] using the amino acid sequences of BT and ST as probes. Multiple sequence alignments were generated with ClustalW[32] using default parameters.

MD simulations were performed using the GROMACS simulation software package,[33] implemented on a parallel architecture. The X-ray structures of native bovine (BT) and atlantic salmon (ST) trypsins (PDB entries 3PTB[34] and 2TBS,[35] respectively) were used as starting points for the computational investigation.

Initial structures for the apo forms have been obtained according to the following procedure: the Ca^{2+} ion has been removed from the PDB file, the apo-protein structure was optimized by molecular mechanics (1000 steepest descent cycles followed by 10 000 conjugate gradient steps), and the resulting structure was submitted for MD simulations.

MD simulations of the holo and apo forms have been carried out according to the following protocol. Protein structures, including crystallographic water molecules, were soaked in a dodecahedral box of 4523 (BT simulations) or 5244 (ST simulations) SPC water molecules.[36] The minimum distance between the solute (protein) and the box edges was set to 0.5 nm. The ionization state of charged residues was set to be consistent with neutral pH: Lys and Arg residues were positively charged, whereas Asp and Glu were negatively charged. The protonation state of the histidine residues was predicted using GROMACS tools and confirmed by visual inspection of the molecular environment of each histidine. To neutralize the overall charge of the system, a number of water molecules equal to the norm of the protein charge, and located within 4 Å of the protein surface, were replaced by Cl^- (BT) or Na^+ (ST) ions.

Initially, solvent molecules were relaxed by molecular mechanics (steepest descent method, 1000 steps). The optimization step was followed by 20 ps MD at 300 K (time step 1 fs) while restraining protein atomic positions using a harmonic potential. MD simulations were performed in the NPT ensemble at 283 and 310 K, applying periodic boundary conditions and using an external thermal bath[37] with a coupling constant τ of 0.1 ps for the protein and 0.002 ps for nonprotein groups. Pressure was kept constant (1 atm) by modifying the box dimensions. The time constant for the pressure coupling was set to 1 ps.[37] The LINCS algorithm[38] was used to constrain the lengths of bonds involving hydrogen atoms. Electrostatic interactions were calculated using the particle-mesh Ewald summation scheme,[39] with a cutoff of 0.12 nm for the separation of the direct and reciprocal space sum cutoffs. van der Waals and Coulomb interactions were truncated at 1.0 nm. The nonbonded pair list was updated every 10 steps, and conformations were stored every 2 ps. The time step was set to 2 fs. The coupling

constant of the external bath was set to 0.001 ps for both protein and nonprotein elements.

To improve the conformational sampling, two or three simulations, obtained initializing the dynamic run with different Maxwellian distributions of initial velocities, were carried out for each protein system. The rmsd (root-mean-square deviation), which is a crucial parameter in the evaluation of the stability of MD trajectories, was evaluated for main-chain atoms using as a reference the starting structure of the MD simulations.

MD trajectories characterized by different initial velocities have been merged to compute rmsd matrices, in which the rmsd values computed comparing all pairs of frames are organized in a two-dimensional map, which allows an evaluation of the resampling of similar substructures. The rmsd matrices representing merged trajectories have been processed using the Jarvis−Patrick method,[40] to extract information on possible clusters of conformations. The average structures of the clusters, defined as the protein structure with the smallest average distance to the other conformation belonging to the same cluster, have also been calculated.[33]

The secondary structure content has been calculated using the DSSP program.[41] The solvent accessibility degree has been evaluated with the NACCESS program[42] on structures collected from the trajectories every 20 ps. The root-mean-square fluctuation (rmsf) has been calculated on main-chain atoms at a residue base, using as a reference the average structure from the trajectories.

The visual analysis of protein structures was carried out using INSIGHT II tools (Biosym Technologies/Molecular Simulations, San Diego, CA) and visual molecular dynamics.[43]

In the following, the numbering of residues is referred to BT if not otherwise specified. The N- (16−135) and C-terminal (136−245) domains of trypsin are labeled as N and C, respectively. The β strands are numbered from β-1 to β-13, while the α helices are numbered from α-1 to α-4. BT and ST sequences are composed of 223 and 222 total amino acids, respectively.

Results and Discussion

Analysis of Trypsin Structures and Sequences. Structures of mammalian and fish trypsins collected in the PDB (59 entries; March 30, 2005) were initially analyzed to characterize the environment of the calcium ion. In addition, the BT sequence was aligned with trypsins from different organisms (Figure 1) to evaluate the conservation of calcium-binding residues. It turned out that the side-chain oxygen atoms of E70, E80, and E77, which are strictly conserved in all trypsins; the backbone oxygen atoms of N72 and V75 (less strictly conserved); and a water molecule define the Ca^{2+} coordination environment in the analyzed trypsins. The molecular environment of the calcium-binding loop (69−80) comprises the loop at the N-terminal end (16−28), the loop between Nβ-1 and Nβ-2 (37−39), part of the interdomain loop between Nβ-6 and Cβ-1 (110−133), and the autolysis loop (144−154) between Cβ-1 and Cβ-2 (Figure 2a). The conformation of the calcium-binding loop in BT and ST is extremely similar, as is the overall fold of the proteins. The main difference in the backbone conformation is localized in correspondence to the autolysis loop (143−155), which assumes significantly different conformations in the two proteins. In fact, the loop conformation observed in ST has been proposed to be stabilized by the ion pairs D150−K154, E21−H71, and K74−D153, which are typical of fish trypsins (Figure 1).[3]

Six disulfide bonds are present in ST and BT (Figure 1, Figure 2b): C22−C157 (connecting the N-terminal end to the Cβ-2 strand), C42−C58 (connecting Nβ-2 to the loop Nβ-3/Nβ-4), C128−C232 (connecting loop Nβ-6/Cβ-1 to Cβ-2), C136−C201 (connecting Cβ-1 to Cβ-4), C168−C182 (connecting Cβ-2 to Cβ-3), and C191−C220 (connecting the loop Cβ-3/Cβ-4 to the loop Cβ-5/Cβ-6). All disulfide bridges are conserved in mammalian and fish trypsins, with the exception of human trypsin II. Notably, only two disulfide bridges are localized in the N-terminal domain, and one of them connects the N- to the C-terminal domain.

Trypsins are proteolytic enzymes specific for arginine and lysine residues. In BT, there are 16 putative autoproteolytic sites (K60, R66, K87, K107, K109, R117, K145, K156, K159, K169, K188, K204, K222, K224, K230, and K239), while in ST, the putative autolytic sites are 12 (K23, K60, R62, R66, K74, R87, R90, K107, K110, K154, K188, and K230). Only a few sites, conserved in most of the sequences, are common to the two proteins (K60, R66, K/R87, K107, K188, and K230) (Figure 1). As mentioned in the Introduction, K60, K145, K188, and R117 were characterized as primary autolysis sites in trypsins.[15,16] It is relevant to observe that R117 is typical of mammalian trypsins, whereas K145 is present only in three trypsins (from bovine, porcine, and Antarctic fish; Figure 1). Most of the lysine and arginine residues are located in loop regions or in proximity of the ends of β strands or α helices, with the exception of K/R87 (in Nβ-5), BT K159 (in Cβ-2), BT K169 (in α-1), and BT K239 (in α-2). Interestingly, two of these residues, K/R87 and K159, are surrounded by other trypsin-sensitive sites, which are localized in loop regions and might be more susceptible to proteolysis.

The solvent accessibility degree of the putative autoproteolytic sites, as well as their structural features, were also analyzed. The side chains of putative autolysis sites are generally well-exposed to the solvent (solvent accessibility degree larger than 30%), with the exception of R66, BT K107 and K156. The solvent accessibility degrees for the autolysis sites that have been characterized experimentally are 38.9% for BT K60, 39.9% for ST K60, 40.6% for BT R117, 60.7% for BT K145, 57.7% for BT K188, and 54.1% for ST K188. Notably, the solvent accessibility degree for K154, which has been proposed to functionally correspond to the BT K145 autolysis site,[3] is as large as 42.2%, in apparent disagreement with the hypothesis that the low propensity to autoproteolysis of this site could be due to solvent inaccessibility.[3] In fact, ST K154 is involved in a salt-bridge cluster with E21, H71, and D150, and therefore, its side chain could not be available to properly interact with the specificity pocket of trypsins. In a similar way, K74, which is located in the calcium-binding loop, can form an ion pair with D153.[3]

```
                                                                C22                 C42
Q8AV11Oncorhynchus_keta          ------------------------IVGGYECKAYSQPHQVSLNSGYHFCG 26
P35032Salmo_SalarII              -----------AAFAT----EDDKIVGGYECKAYSQPHQVSLNSGYHFCG 35
P35031Salmo_salarI               MISLVFVLLIGAAFAT----EDDKIVGGYECKAYSQTHQVSLNSGYHFCG 46
Q91515Fugu_rubripes              --------LIAAAYAAPI-DEDDKIVGGYECRKNSVAYQVSLNSGYHFCG 41
Q92099Paranotothenia_magellani   MRSLVFVLLIGAAFAT----EEDKIVGGKECSPYSQPHQVSLNSGYHFCG 46
Q9W7Q7Paralichthys_olivaceus     MRSLVFVLLIGAAFAM----EDDKIVGGYECTPYSQPHQVSLNSGYHFCG 46
Q9W7Q6Paralichthys_olivaceusII   ---LVFILLIGAAFAT----EDDKIVGGYECTPYSQPHQVSLNSGYHFCG 43
O93266Pleuronectes_americanus    MRSLVFVLLIGAAFAL----EDDKIVGGYECTPHSQAHQVSLNSGYHFCG 46
Q6RI79Tautogolabrus_adspersus    MRSLVFVLLIGAAFAI----DDDKIVGGYECTPHSQPHTVSLQSGYHFCG 46
P16049Gadus_morhuaI              MKSLIFVLLLGAVFAE-----EDKIVGGYECTKHSQAHQVSLNSGYHFCG 45
Q91041Gadus_morhuaII             MKSLIFVLLLGAVFAE-----EDKIVGGYECTRHSQAHQVSLNSGYHFCG 45
Q98TG9Engraulis_japonicusII      MRSLVFLVLLGAAFAE-----DDKIVGGYECQPYSQPHQVSLNSGYHFCG 45
Q98TH0Engraulis_japonicusI       MRPLVFLVLLGAAFAE-----DDKIVGGYECQAHSQPHTVSLNSGYHFCG 45
Q7SX90Brachydanio_rerio          MRSLVFLVLLGAAFALD----DDKIVGGYECQPNSQPWQASLNSGYHFCG 46
Q8QGW3Anguilla_japonica          MRSLVFILLLGVAVALD----DDKIVGGYECEPHSQPWQASLNAGYHFCG 46
Q7T1R8Pangasius_hypophthalmus    MRSLVLLLLVGACFALE----DDKIVGGYECTPYSQPWQVSLNVGYHFCG 46
Q6R670Oreochromis_aureus         MKYFILLALFAAAYAAPIED--DKIIGGYECAKNSVPYMVSLNIGYHFCG 48
Q6R671Oreochromis_niloticus      MKYFILLALFAAAYAAPIED--DKIIGGYECAKNSVPYMVSLNIGYHFCG 48
P35033Salmo_salarIII             ---------FAVAFAAPIDDEDDKIVGGYECRKNSASYQASLQSGYHFCG 41
P07477Homo_sapiensI              MNPLLILTFVAAALAAPFDDDD-KIVGGYNCEENSVPYQVSLNSGYHFCG 49
P07478Homo_sapiensII             MNLLLILTFVAAAVAAPFDDDD-KIVGGYICEENSVPYQVSLNSGYHFCG 49
P06872Canis_familiarisII         MNPLLILAFLGAAVATPTDDDD-KIVGGYTCEENSVPYQVSLNAGYHFCG 49
P00763Rattus_norvegicusII        MRALLFLALVGAAVAFPVDDDD-KIVGGYTCQENSVPYQVSLNSGYHFCG 49
Q792Z1Mus_musculusI              MSTLLFLALVGAAVAFPVDDDD-KIVGGYTCRENSVPYQVSLNSGYHFCG 49
P07146Mus_musculusII             MSALLILALVGAAVAFPVDDDD-KIVGGYTCRESSVPYQVSLNAGYHFCG 49
P00762Rattus_norvegicusI         MSALLILALVGAAVAFPLEDDD-KIVGGYTCPEHSVPYQVSLNSGYHFCG 49
P00760Bos_taurusI                ---FIFLALLGAAVAFPVDDDD-KIVGGYTCGANTVPYQVSLNSGYHFCG 46
P06871Canis_familiarisI          MKTFIFLALLGATVAFPIDDDD-KIVGGYTCSRNSVPYQVSLNSGYHFCG 49
P00761Sus_scrofaI                ---------------FPTDDDD-KIVGGYTCAANSIPYQVSLNSGSHFCG 34
P08426Rattus_norvegicusIII       MKALIFLAFLGAAVALPLDDDDDKIVGGYTCQKNSLPYQVSLNAGYHFCG 50
P70059Xenopus_laevisII           MKFLVILVLLGAAVAF---EDDDKIVGGFTCAKNAVPYQVSLNAGYHFGG 47
P19799Xenopus_laevisI            MKFLLLCVLLGAAAAF---DDD-KIIGGATCAKSSVPYIVSLNSGYHFCG 46
                                                     *:**  *   : .  .**: * ** *
```

```
                                      ;    C58 K60          Ca2+ b.loop
Q8AV11Oncorhynchus_keta          GSLVNENWVVSAAHCYKS--R-EVRLGEHNIKVTEGSEQFISSSRVIRHP 73
P35032Salmo_SalarII              GSLVNENWVVSAAHCYQS--RVEVRLGEHNIQVTEGSEQFISSSRVIRHP 83
P35031Salmo_salarI               GSLVNENWVVSAAHCYKS--RVEVRLGEHNIKVTEGSEQFISSSRVIRHP 94
Q91515Fugu_rubripes              GSLVNENWVVSAAHCYKS--RVVVRLGEHNIRANEGTEQFISSSRVIRHP 89
Q92099Paranotothenia_magellani   -SLVNENWVVSAAHCYKS--RVEVRMGEHHIRVTEGKEQFISSSRVIRHP 93
Q9W7Q7Paralichthys_olivaceus     GSLVNENWVVSAAHCYKS--RVEVRMGEHHIKINEGTEQFISSERVIRHP 94
Q9W7Q6Paralichthys_olivaceusII   GSLVNENWVVSAAHCYKS--RVEVRIGEHNLRVYEETEQFISSSRVIRHP 91
O93266Pleuronectes_americanus    GSLVNENWVVSAAHCYKS--RVEVRMGEHKIRVNEGTEQFVSSSRVIRHP 94
Q6RI79Tautogolabrus_adspersus    GSLVNEDWVVSAAHCYKS--RIQVRLGEHHIRVNEGTEQIISSSRVIRHP 94
P16049Gadus_morhuaI              GSLVSKDWVVSAAHCYKS--VLRVRLGEHHIRVNEGTEQYISSSSVIRHP 93
Q91041Gadus_morhuaII             GSLVSKDWVVSAAHCYKS--VLRVRLGEHHIRVNEGTEQFISSSSVIRHP 93
Q98TG9Engraulis_japonicusII      GSLVSDSWVVSAAHCYKS--RVEVRMGEHHIGMTEGNEQFIDSSRVIRHP 93
Q98TH0Engraulis_japonicusI       GSLVNENWVVSAAHCYKS--RVEVRLGEHHIGQNENTEQFIDSSRVIRHP 93
Q7SX90Brachydanio_rerio          GSLVSEYWVVSAAHCYKS--RVEVRLGEHNIVINEGTEQFITSEKVIRNP 94
Q8QGW3Anguilla_japonica          GSLVNENWVVSAAHCYKSPSRLEVRLGEHHIGLNEGTEQFIGASHVIRNP 96
Q7T1R8Pangasius_hypophthalmus    GSLINQNWVVSAAHCYQS--RIEVRLGEHNIQINEGTEQFISSSRVIRHP 94
Q6R670Oreochromis_aureus         GSLISSTWAVSAAHCYQS--SIQLRLGEHNIAVNEGTEQFISSSRVIRHQ 96
Q6R671Oreochromis_niloticus      GSLISSTWAVSAAHCYQS--SIQLRLGEHNIAVNEGTEQFISSSRVIRHQ 96
P35033Salmo_salarIII             GSLISSTWVVSAAHCYKS--RIQVRLGEHNIAVNEGTEQFIDSVKVIMHP 89
P07477Homo_sapiensI              GSLINEQWVVSAGHCYKS--RIQVRLGEHNIEVLEGNEQFINAAKIIRHP 97
P07478Homo_sapiensII             GSLISEQWVVSAGHCYKS--RIQVRLGEHNIEVLEGNEQFINAAKIIRHP 97
P06872Canis_familiarisII         GSLISDQWVVSAAHCYKS--RIQVRLGEYNIDVLEGNEQFINSAKVIRHP 97
P00763Rattus_norvegicusII        GSLINDQWVVSAAHCYKS--RIQVRLGEHNINVLEGNEQFVNAAKIIKHP 97
Q792Z1Mus_musculusI              GSLINDQWVVSAAHCYKS--RIQVRLGEHNINVLEGNEQFIDAANIIKHP 97
P07146Mus_musculusII             GSLINDQWVVSAAHCYKY--RIQVRLGEHNINVLEGNEQFVDSAKIIRHP 97
P00762Rattus_norvegicusI         GSLINDQWVVSAAHCYKS--RIQVRLGEHNINVLEGDEQFINAAKIIKHP 97
P00760Bos_taurusI                GSLINSQWVVSAAHCYKS--GIQVRLGEDNINVVEGNEQFISASKSIVHP 94
P06871Canis_familiarisI          GSLINSQWVVSAAHCYKS--RIQVRLGEYNIAVSEGGEQFINAAKIIRHP 97
P00761Sus_scrofaI                GSLINSQWVVSAAHCYKS--RIQVRLGEHNIDVLEGNEQFINAAKIITHP 82
P08426Rattus_norvegicusIII       GSLINSQWVVSAAHCYKS--RIQVRLGEHNIDVVEGGEQFIDAAKIIRHP 98
P70059Xenopus_laevisII           -SLINSQWVVSAAHCYKS--RIQVRLGEHNIALNEGTEQFIDSQKVIKHP 94
P19799Xenopus_laevisI            GSLITNQWVVSAAHCYKA--SIQVRLGEHNIALSEGTEQFISSSKVIRHS 94
                                  **:.. *.***.***:      :*:** ::   *  ** : :   * :
```

```
                                                R117      C128   C136
Q8AV11Oncorhynchus_keta                 NYSSYNIDNDIMLIKLSKPATLNTYVQPVALPSSCAPAGTMCTVSGWGNT 123
P35032Salmo_SalarII                     NYSSYNIDNDIMLIKLSKPATLNTYVQPVALPTSCAPAGTMCTVSGWGNT 133
P35031Salmo_salarI                      NYSSYNIDNDIMLIKLSKPATLNTYVQPVALPTSCAPAGTMCTVSGWGNT 144
Q91515Fugu_rubripes                     NYSSYNIDNDIMLIKLSKPATLNQYVQPVALPSSCAAAGTMCKVSGWGNT 139
Q92099Paranotothenia_magellani          NYSSYNIDNDIMLIKLSKPATLNQYVQAVALPSSCAPAGTMCTVSGWGST 143
Q9W7Q7Paralichthys_olivaceus            NYSSYNINNDIMLIKLREPATLNQYVQPVALPTSCAPAGTMCTVSGWGNT 144
Q9W7Q6Paralichthys_olivaceusII          NYSSYNINNDIMLIKLSEPATLNQYVQPVALPTSCAPAGTMCTVSGWGDT 141
O93266Pleuronectes_americanus           NYDSWNIDNDIMLIKLSKPATLNQYVKTVALPSSCAPAGTMCKVSGWGNT 144
Q6RI79Tautogolabrus_adspersus           RYSSYNIDNDIMLIKLSQSATLNQYVKTVALPTSCAPAGTMCKVAGWGNT 144
P16049Gadus_morhuaI                     NYSSYNINNDIMLIKLTKPATLNQYVHAVALPTECAADATMCTVSGWGNT 143
Q91041Gadus_morhuaII                    NYSSYNIDNDIMLIKLTEPATLNQYVHAVALPTECAADATMCTVSGWGNT 143
Q98TG9Engraulis_japonicusII             QYDSYNIDNDIMLIKLSKPATLNQYVQTVALPSSCAPAGTMCLVSGWGNT 143
Q98TH0Engraulis_japonicusI              QYSSYNIDNDVMLIKLSTPATLNQYVQPVALPSRCASAGTMCLVAGWGNT 143
Q7SX90Brachydanio_rerio                 NYDSWDLDSDIMLIKLSKPATLNKYVQPVALPNGCAADGTMCRVSGWGNT 144
Q8QGW3Anguilla_japonica                 NYNSWDLDSDIMLIKLSKPATLNSYVQPVALPTRCAPAGTMCRVTGWGNT 146
Q7T1R8Pangasius_hypophthalmus           NYNSWTIDNDIMLIKLSQSASVNNYVQPVALPSSCPPAGTWCIVSGWGNT 144
Q6R670Oreochromis_aureus                SYNSYTLDNDIMLIKLSQPATLNSYVKTVSLPSGCAGAGTSCLISGWGNT 146
Q6R671Oreochromis_niloticus             SYNSYTLDNDIMLIKLSQPATLNSYVKTVSLPSGCAGAGTSCLISGWGNT 146
P35033Salmo_salarIII                    SYNSRNLDNDIMLIKLSKPASLNSYVSTVALPSSCASSGTRCLVSGWGNL 139
P07477Homo_sapiensI                     QYDRKTLNNDIMLIKLSSRAVINARVSTISLPTAPPATGTKCLISGWGNT 147
P07478Homo_sapiensII                    KYNSRTLDNDILLIKLSSPAVINSRVSAISLPTAPPAAGTESLISGWGNT 147
P06872Canis_familiarisII                NYNSWILDNDIMLIKLSSPAVLNARVATISLPRACAAPGTQCLISGWGNT 147
P00763Rattus_norvegicusII               NFDRKTLNNDIMLIKLSSPVKLNARVATVALPSSCAPAGTQCLISGWGNT 147
Q792Z1Mus_musculusI                     KFKKKTLDNDIMLIKLSSPVTLNARVATVALPSSCAAAGTQCLISGWGNT 147
P07146Mus_musculusII                    NYNSWTLDNDIMLIKLASPVTLNARVASVPLPSSCAPAGTQCLISGWGNT 147
P00762Rattus_norvegicusI                NYSSWTLNNDIMLIKLSSPVKLNARVAPVALPSACAPAGTQCLISGWGNT 147
P00760Bos_taurusI                       SYNSNTLNNDIMLIKLKSAASLNSRVASISLPTSCASAGTQCLISGWGNT 144
P06871Canis_familiarisI                 RYNANTIDNDIMLIKLSSPATLNSRVSAIALPKSCPAAGTQCLISGWGNT 147
P00761Sus_scrofaI                       NFNGNTLDNDIMLIKLSSPATLNSRVATVSLPRSCAAAGTECLISGWGNT 132
P08426Rattus_norvegicusIII              SYNANTFDNDIMLIKLNSPATLNSRVSTVSLPRSCGSSGTKCLVSGWGNT 148
P70059Xenopus_laevisII                  NYNSRNLDNDIMLIKLSTTARLSANIQSVPLPSACASAGTNCLISGWGNT 144
P19799Xenopus_laevisI                   GYNSYTLDNDIMLIKLSSPASLNAAVNTVPLPSGCSAAGTSCLISGWGNT 144
                                         :.   ::.*::****   . :.  : .:.**      .* . ::***.
```

```
                                        K145   K154 C157       C168          C182  K188 C191
Q8AV11Oncorhynchus_keta                 MSSTADK-NKLQCLNIPILSYSDCNNSYPGMITNAMFCAGYLEGGKDSCQ 172
P35032Salmo_SalarII                     MSSTADK-NKLQCLNIPILSYSDCNNSYPGMITNAMFCAGYLEGGKDSCQ 182
P35031Salmo_salarI                      MSSTADS-NKLQCLNIPILSYSDCNNSYPGMITNAMFCAGYLEGGKDSCQ 193
Q91515Fugu_rubripes                     MSSTADR-NKLQCLNIPILSDRDCENSYPGMITDAMFCAGYLEGGKDSCQ 188
Q92099Paranotothenia_magellani          QSSSADG-NKLQCLNIPILSDRDCDNSYPGMITDAMFCAGYLQGGKDSCQ 192
Q9W7Q7Paralichthys_olivaceus            MSSTANR-DMLQCLDLPILSDRDCENSYPGMITPAMFCAGYLEGGKDSCQ 193
Q9W7Q6Paralichthys_olivaceusII          MSSTDS--SRLQCLDLPILSERDCENSYPGMITNAMFCAGYLEGGKDSCQ 189
O93266Pleuronectes_americanus           MSSADNG-DLLQCLDIPILSFSDCNNAYPGMITDSMFCAGYLEGGKDSCQ 193
Q6RI79Tautogolabrus_adspersus           MSSSADG-DKLQCLNIPILSERDCDNSYPGMITDAMFCAGYLEGGKDSCQ 193
P16049Gadus_morhuaI                     MSSVADG-DKLQCLSLPILSHADCANSYPGMITQSMFCAGYLEGGKDSCQ 192
Q91041Gadus_morhuaII                    MSSVDDG-DKLQCLNLPILSHADCANSYPGMITQSMFCAGYLEGGKDSCQ 192
Q98TG9Engraulis_japonicusII             MS-NVSG-DKLQCLQIPILSDRDCKNSYPGMITESMFCAGYLEGGKDSCQ 191
Q98TH0Engraulis_japonicusI              MS-NVSG-DKLQCLQIPILSDRDCDNSYPGMITDAMFCAGYLEGGKDSCQ 191
Q7SX90Brachydanio_rerio                 MSSTADS-NKLQCLEIPILSDRDCNNSYPGMVTDTMFCAGYLEGGKDSCQ 193
Q8QGW3Anguilla_japonica                 MNPAVSG-DKLQCLEIPILSESDCSNSYPGMITSTMFCAGYLEGGKDSCQ 195
Q7T1R8Pangasius_hypophthalmus           MSSTADR-NKLQCVEVPILSDEDCNNSYPGMITKAMFCAGFLEGGKDSCQ 193
Q6R670Oreochromis_aureus                STSGSNYPDRLMCLNAPILSDTDCRNSYPGEITNNMFCAGFLEGGKDSCQ 196
Q6R671Oreochromis_niloticus             STSGSNYPDRLMCLNAPILSDTDCRNSYPGEITNNMFCAGFLEGGKDSCQ 196
P35033Salmo_salarIII                    SGSSSNYPDTLRCLDLPILSSSSCNSAYPGQITSNMFCAGFMEGGKDSCQ 189
P07477Homo_sapiensI                     ASSGADYPDELQCLDAPVLSQAKCEASYPGKITSNMFCVGFLEGGKDSCQ 197
P07478Homo_sapiensII                    LSSGADYPDELQCLDAPVLSQAECEASYPGKITNNMFCVGFLEGGKDSCQ 197
P06872Canis_familiarisII                LSSGTNYPELLQCLDAPILTQAQCEASYPGQITENMICAGFLEGGKDSCQ 197
P00763Rattus_norvegicusII               LSSGVNEPDLLQCLDAPLLPQADCEASYPGKITDNMVCVGFLEGGKDSCQ 197
Q792Z1Mus_musculusI                     LSSGVNNPDLLQCLDAPLLPQADCEASYPGKITKNMICVGFLEGGKDSCQ 197
P07146Mus_musculusII                    LSNGVNNPDLLQCVDAPVLPQADCEASYPGDITNNMICVGFLEGGKDSCQ 197
P00762Rattus_norvegicusI                LSNGVNNPDLLQCVDAPVLSQADCEAAYPGEITSSMICVGFLEGGKDSCQ 197
P00760Bos_taurusI                       KSSGTSYPDVLKCLKAPILSDSSCKSAYPGQITSNMFCAGYLEGGKDSCQ 194
P06871Canis_familiarisI                 QSIGQNYPDVLQCLKAPILSDSVCRNAYPGQISSNMMCLGYMEGGKDSCQ 197
P00761Sus_scrofaI                       KSSGSSYPSLLQCLKAPVLSDSSCKSSYPGQITGNMICVGFLEGGKDSCQ 182
P08426Rattus_norvegicusIII              LSSGTNYPSLLQCLDAPVLSDSSCKSSYPGKITSNMFCLGFLEGGKDSCQ 198
P70059Xenopus_laevisII                  LSSGTNYPDLLQCLNAPILTDSQCSNSYPGEITKNMFCAGFLAGGKDSCQ 194
P19799Xenopus_laevisI                   LSNGSNYPDLLQCLNAPILTNAQCNSAYPGEITANMICVGYMEGGKDSCQ 194
                                             .  . * *:. *:*.   *  :*** ::  *.* *:: *******
```

```
                                  C201          C220          C232
Q8AV11Oncorhynchus_keta           GDSGGPVVCNGELQGVVSWGYGCAEPGNPGVYAKVCIFNDWLTSTMATY- 221
P35032Salmo_SalarII               GDSGGPVVCNGELQGVVSWGYGCAEPGNPGVYAKVCIFNDWLTSTMATY- 231
P35031Salmo_salarI                GDSGGPVVCNGELQGVVSWGYGCAEPGNPGVYAKVCIFNDWLTSTMASY- 242
Q91515Fugu_rubripes               GDSGGPVVCNNELQGVVSWGYGCAERDHPGVYAKVCLFNDWLESTMASY- 237
Q92099Paranotothenia_magellani    GDSGGPVVCNGELQGVVSWGYGCAERDHPGVYAKVCLN-DWLETSMANY- 240
Q9W7Q7Paralichthys_olivaceus      GDSGGPVVCNGELQGVVSWGYGCAERDHPGVYARVCIFIDWLETTMASY- 242
Q9W7Q6Paralichthys_olivaceusII    GDSGGPVVCNGQLQGVVSWGYGCAQRDHPGVYAKVCIFIDWLERTMSSY- 238
O93266Pleuronectes_americanus     GDSGGPVICNGELQGVVSWGYGCAERGNPGVYAKVCLFNDWLESTMASY- 242
Q6RI79Tautogolabrus_adspersus     GDSGGPVVCNGQLQGVVSWGYGCAERDHPGVYAKVCLFNDWLESTMASN- 242
P16049Gadus_morhuaI               GDSGGPVVCNGVLQGVVSWGYGCAERDHPGVYAKVCVLSGWVRDTMANY- 241
Q91041Gadus_morhuaII              GDSGGPVVCNGVLQGVVSWGYGCAERDNPGVYAKVCVLSGWVRDTMASY- 241
Q98TG9Engraulis_japonicusII       GDSGGPVVCNGELQGIVSWGYGCAERDHPGVYAKVCLFNDWIDSTMAQYN 241
Q98TH0Engraulis_japonicusI        GDSGGPVVCNGELQGVVSWGYGCAERDHPGVYAKVCIFTDWLQSTMASN- 240
Q7SX90Brachydanio_rerio           GDSGGPVVCNGELHGIVSWGYGCAEKNHPGVYGKVCMFSQWIADTMRNN- 242
Q8QGW3Anguilla_japonica           GDSGGPVVCNGELQGVVSWGYGCAEQNHPGVYPKVCMFSDWLRTTMAST- 244
Q7T1R8Pangasius_hypophthalmus     GDSGGPVVCNGELQGIVSWGYGCAEKNHPGVYTKVCIFTDWIAQTIASN- 242
Q6R670Oreochromis_aureus          GDSGGPVVCNGQLQGIVSWGYGCAQRNRPGVYTKVCNYNSWISNTMANN- 245
Q6R671Oreochromis_niloticus       GDSGGPVVCNGQLQGIVSWGYGCAQRDRPGVYTKVCNYNSWISNTMANN- 245
P35033Salmo_salarIII              GDSGGPVVCNGQLQGVVSWGYGCAQRNKPGVYTKVCNYRSWISSTMSSN- 238
P07477Homo_sapiensI               GDSGGPVVCNGQLQGVVSWGDGCAQKNKPGVYTKVYNYVKWIKNTIAANS 247
P07478Homo_sapiensII              GDSGGPVVSNGELQGIVSWGYGCAQKNRPGVYTKVYNYVDWIKDTIAANS 247
P06872Canis_familiarisII          GDSGGPVVCNGELQGIVSWGYGCAQKNKPGVYTKVCNFVDWIQSTIAANS 247
P00763Rattus_norvegicusII         GDSGGPVVCNGELQGIVSWGYGCALPDNPGVYTKVCNYVDWIQDTIAAN- 246
Q792Z1Mus_musculusI               GDSGGPVVCNGQLQGIVSWGYGCAQKDNPGVYTKVCNYVDWIQNTIAAN- 246
P07146Mus_musculusII              GDSGGPVVCNGELQGIVSWGYGCAQPDAPGVYTKVCNYVDWIQNTIADN- 246
P00762Rattus_norvegicusI          GDSGGPVVCNGQLQGIVSWGYGCALPDNPGVYTKVCNFVGWIQDTIAAN- 246
P00760Bos_taurusI                 GDSGGPVVCSGKLQGIVSWGSGCAQKNKPGVYTKVCNYVSWIKQTIASN- 243
P06871Canis_familiarisI           GDSGGPVVCNGELQGVVSWGAGCAQKGKPGVSPKVCKYVSWIQQTIAAN- 246
P00761Sus_scrofaI                 GDSGGPVVCNGQLQGIVSWGYGCAQKNKPGVYTKVCNYVNWIQQTIAAN- 231
P08426Rattus_norvegicusIII        GDSGGPVVCNGQLQGVVSWGYGCAQKGKPGVYTKVCNYVNWIQQTVAAN- 247
P70059Xenopus_laevisII            GDSGGPVVCNGQLQGVVSWGYGCAQRNYPGVYTKVCNFVTWIQSTISSN- 243
P19799Xenopus_laevisI             GDSGGPVVCNGQLQGVVSWGYGCAMRNYPGVYTKVCNYNAWIQNTIAAN- 243
                                  *******:... *:*:**** ***
```

Figure 1. Multiple sequence alignment of mammalian and fish trypsins. Sequences of mammalian and fish trypsins are indicated with their SwissProt code. The identical (*) and similar residues (: or .) are highlighted. The calcium-binding loop, the conserved cysteine residues, and the putative autolytic sites are highlighted in gray and boldfaced.

Molecular Dynamics. ST and BT were chosen as representative structures for fish and mammalian trypsins, respectively, since they have been thoroughly characterized experimentally.[1,3,16] Multiple 6 ns MD simulations of calcium-bound (holoBT) and calcium-free (apoBT) bovine trypsins, as well as calcium-bound (holoST) and calcium-free (apoST) atlantic salmon trypsins, were carried out at both 283 and 310 K, as outlined in the Methods. The choice of two different temperatures was motivated by the observation that BT and ST belong to mesophilic and psycrophilic organisms, respectively. However, results related to the effects of Ca^{2+} on the structure and dynamic properties discussed in the present contribution are essentially independent from the chosen temperature values (not shown). Therefore, only results for MD runs obtained at 310 K will be presented and discussed.

Several criteria have been used to evaluate convergence of the MD runs. Indeed, it is well-known that multiple trajectories help to identify recurring features and to avoid artifacts arising from the simulation procedure.[44] Therefore, to efficiently sample the conformational space, two or three independent MD simulations were carried out starting from the same atomic coordinates but using different initial velocities randomly taken from a Maxwellian distribution compatible with the simulation temperature. In the following, MD trajectories collected for the same system but characterized by different initial velocities are labeled as simX (as a subscript), where X runs from 1 to the number of MD simulations carried out for a particular system.

The rmsd value calculated for the main-chain atoms of ST and BT (Figure 3) reached a stable value after about 1.5 ns in all simulations, with the exception of $apoBT_{sim2}$ and $holoST_{sim1}$, which reach convergence after 2.7 ns. As a consequence, the analyses of MD trajectories have been carried out using only the last 4.5 or 3.3 ns.

The interatomic distance between the atoms of the coordinating residues (70, 72, 75, 77, and 80) and Ca^{2+}, as well as the protein gyration radius and total and potential energies of the system, are characterized by negligible fluctuations throughout the simulations, indicating stable trajectories (data not shown).

The removal of Ca^{2+} did not significantly affect the secondary structure content in BT and ST, possibly because of the presence of six disulfide bridges that stabilize the N-terminal (C42−C58) and C-terminal domains (C128−C232, C136−C201, C168−C182, and C191−C220) and cross-link the N- and the C-terminal domains (C22−C157). However, it was generally possible to observe a slight reduction of the average number of residues characterized by regular secondary structure in trajectories of the apo forms (Table 1). Since the N-terminal domain presents a fewer number of disulfide bridges, the secondary structure content was also evaluated separately for the two domains. The behavior of C- and N-terminal domains is similar, the only

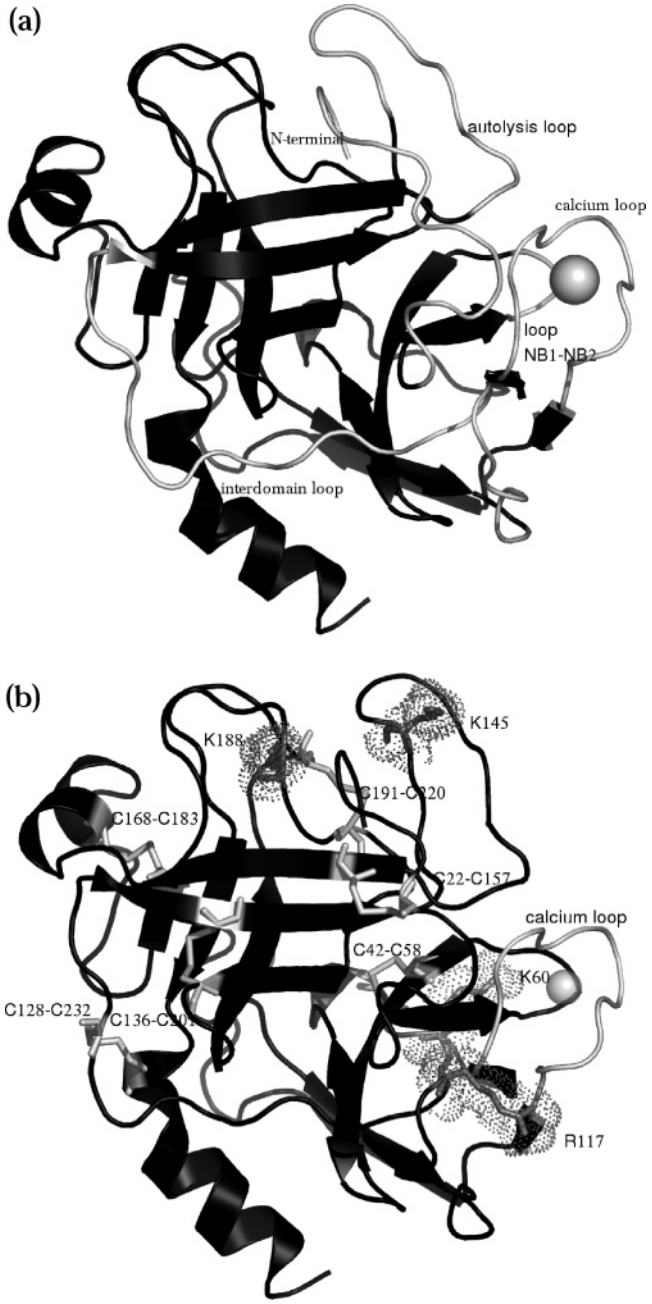

Figure 2. 3D structure of BT trypsin. (a) The calcium ion and the loop regions which surround the calcium-binding loop are shown in gray and light gray, respectively. (b) The calcium ion, the six conserved disulfide bridges, and the putative autolysis sites are shown as gray spheres, light gray sticks, and light gray sticks and dots, respectively. The secondary structure elements are shown as ribbons.

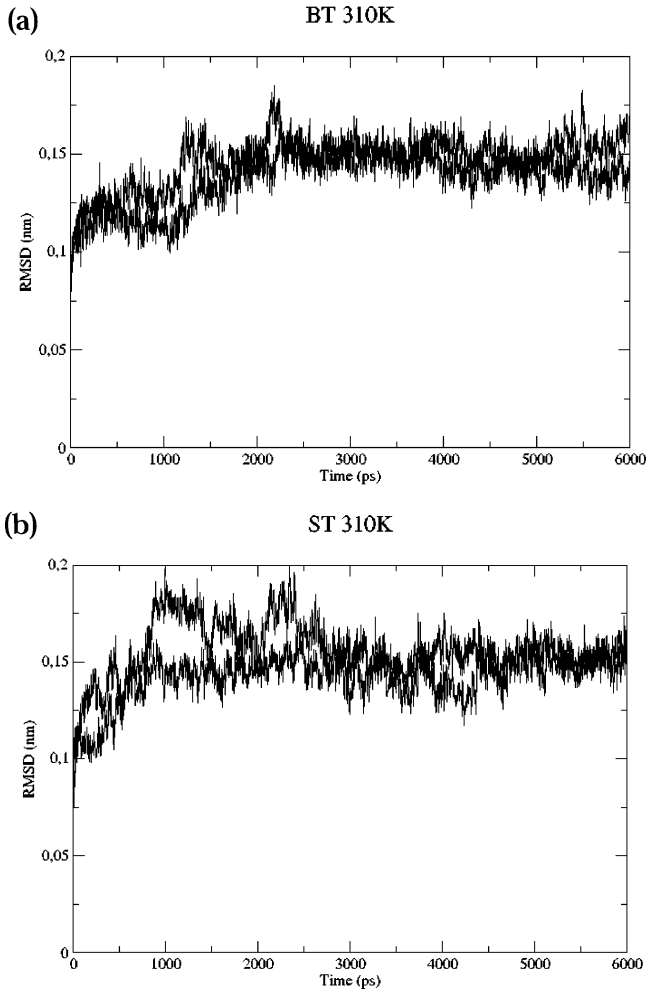

Figure 3. rmsd of main-chain atoms as a function of time. (a) rmsd as a function of time of BT at 310 K for the holo (black line) and apo forms (gray line). (b) rmsd as a function of time of ST at 310 K for the holo (black line) and apo forms (gray line). For the sake of clarity, only values relative to the simulation sim1 are shown.

appreciable difference being that the N-terminal domain in the apo forms is generally characterized by a lower number of residues in β-strand conformations (not shown).

The evaluation of conformational sampling in MD trajectories is crucial to avoid a misleading interpretation of MD simulations.[44,45] A simple but efficient index of sampling is the extension of resampling in the phase space. The system revisiting similar structures is a good indication of the convergence of the simulation on structures that represent the conformational properties of the system. Therefore, multiple trajectories, obtained by different Maxwellian distributions of initial velocities, have been merged and analyzed. In particular, the data from rmsd matrices, obtained as outlined in the Methods, were used to carry out a cluster analysis and obtain further insight in the configurations visited by the system. Only one cluster has been identified in the MD trajectories for the holo forms, indicating that simulations initialized with different Maxwellian distributions converge to similar conformational basins. On the other hand, all the MD trajectories for the apo forms presented two different clusters, indicating a greater conformational freedom of calcium-free structures. However, when the secondary structure content of the average structures obtained from the cluster analysis (see Method) was analyzed, the only appreciable difference observed comparing the apo and holo forms was a greater disorder for the Nβ-2 and Nβ-5 strands in the apo forms, in agreement with the above observations about the secondary structure content of the N-terminal domain. The average 3D structures obtained from the conformational clustering were analyzed also to evaluate the structural features of some key regions of the protein: the

Table 1. Average Number of Residues in a Given Secondary Structure of BT and ST at 310 K in the Holo and Apo Forms, According to the DSSP Software[41] [a]

	BT				ST			
	$holo_{sim1}$	apo_{sim1}	$holo_{sim2}$	apo_{sim2}	$holo_{sim1}$	apo_{sim1}	$holo_{sim2}$	apo_{sim2}
structure	121.9 (5.7)	120.8 (5.2)	118.2 (5.3)	117.2 (5.3)	125.8 (5.5)	118.1 (5.2)	118.0 (5.9)	114.2 (7.0)
coil	58.6 (3.3)	58.8 (3.5)	60.9 (3.4)	60.0 (3.4)	55.6 (3.1)	61.3 (3.3)	61.0 (3.6)	63.1 (4.1)
β sheet	71.4 (3.3)	70.5 (3.4)	69.4 (3.7)	69.6 (3.6)	70.0 (3.2)	70.8 (3.4)	68.2 (3.3)	65.3 (4.1)
β bridge	6.6 (2.3)	5.0 (1.8)	5.0 (1.9)	7.2 (2.3)	6.4 (1.9)	4.8 (1.9)	6.2 (1.7)	6.2 (2.2)
bend	38.9 (4.1)	38.5 (3.8)	41.8 (3.8)	38.3 (4.1)	38.7 (4.4)	40.3 (4.2)	38.1 (4.4)	40.9 (4.8)
turn	26.1 (3.7)	26.2 (4.0)	25.4 (3.7)	26.5 (3.6)	30.1 (4.2)	23.0 (4.0)	27.1 (4.7)	24.9 (4.6)
helix	17.8 (2.0)	18.0 (2.1)	16.4 (1.5)	16.9 (1.5)	19.2 (1.7)	20.4 (1.2)	16.3 (1.3)	17.7(1.9)
3_{10} helix	3.6 (3.7)	3.9 (4.1)	5.2 (2.2)	4.5 (2.4)	1.9 (4.2)	2.4 (4.0)	4.8 (2.6)	3.8 (2.9)

[a] For the sake of clarity, only values relative to the simulations sim1 and sim2 are shown. Standard deviations are given in parentheses.

catalytic triad (H57, D102, and S195), the specificity pocket (D189, G216, and V226), and the hydrophobic pair I73−W141. The structural properties of the catalytic triad and of the specificity pocket are always very similar in the apo and holo forms. The same holds true for the hydrophobic pair I73−W141, which, according to experimental data,[9,10] might be implicated in the transmission of the structural modifications observed upon calcium removal. In fact, the interatomic distances between the C atoms of I73 and W141 are nearly constant during all MD simulations. This indicates that no relevant changes in the interaction between W141 and I73 take place upon calcium removal.

The rmsd per residue from the average structure (rmsf) was calculated to better evaluate the role of Ca^{2+} on protein flexibility (Figure 4).[47] The overall profiles of the rmsf deviations are quite similar for the apo and holo forms, as well as for ST and BT: the rmsf peaks are located in corresponding positions, whereas the intensity of the fluctuations is generally larger in the apo-form simulations. Regions characterized by a regular secondary structure show small fluctuations during the simulations, whereas pronounced fluctuations are observed for some loop regions.

To better highlight regions characterized by different flexibility in the apo and holo forms, the rmsf profile for each holo simulation was subtracted from the corresponding profile obtained for the apo form (rmsf-diff; Figure 4); values of rmsf-diff lower than zero indicate regions where the fluctuations are larger for the apo form. All proteins present a larger number of residues with rmsf-diff values lower than zero, confirming the previous observations indicating a larger flexibility for the apo forms. In particular, BT and ST show 127 and 135 residues (out of 223 and 222 total residues), respectively, with an rmsf-diff value lower than zero. BT and ST trajectories are characterized by similar rmsf-diff profiles, and most importantly, the lack of Ca^{2+} has effects not only on the N-terminal domain, where the calcium-binding site is localized, but also on the C-terminal domain.

Protein portions characterized by rmsf-diff values lower than −0.05 nm (greater flexibility in the apo form) or higher than 0.05 nm (greater flexibility in the holo form) have been highlighted on the average three-dimensional structures of BT (Figure 5). In the N-terminal domain, the regions characterized by enhanced fluctuations in the apo forms are the calcium-binding loop between Nβ-4 and Nβ-5 (in particular, the region 72−80), the loop at the N end (in particular the region 23−28), and part of the interdomain loop, between Nβ-6 and Cβ-1 (in particular, the regions 93−103 and 110−119) (Figure 5). Notably, all the latter regions surround the calcium-binding loop. Fluctuation intensity changes significantly in three regions of the C-terminal domain upon Ca^{2+} removal (Figure 5): residue 141 in ST and the pair of interacting loops between Cβ-3 and Cβ-4 (in particular, the region 184−191) and between Cβ-5 and Cβ-6

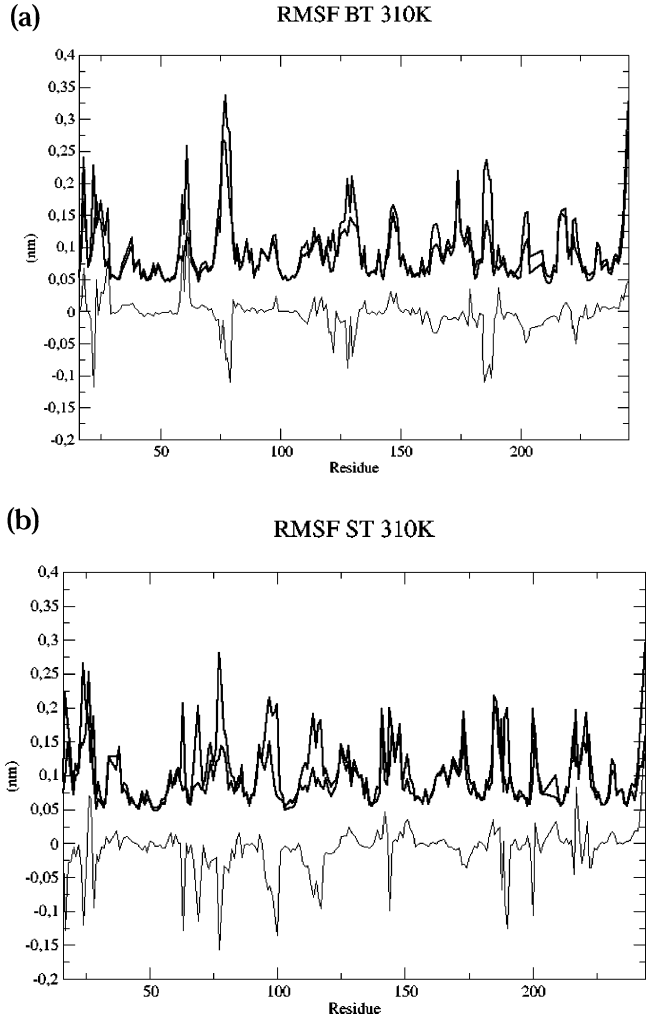

Figure 4. rmsf and rmsf-diff as a function of residue numbers. (a) rmsf values of each residue of the holo (black line) and apo forms (gray line) of BT at 310 K. (b) rmsf values of each residue of the holo (black line) and apo forms (gray line) of ST at 310 K. The rmsf-diff is indicated by a thick black line.

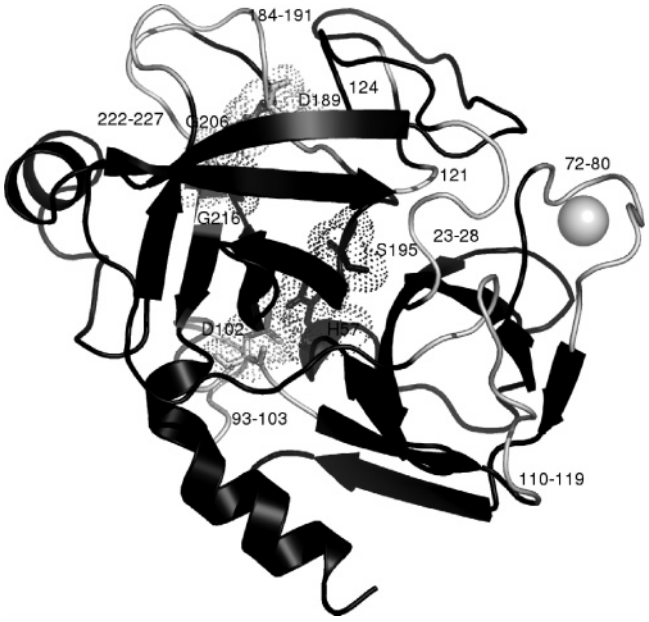

Figure 5. Regions characterized by greater fluctuation in the apo-trypsin structure. The calcium ion and the regions characterized by rmsf values greater in apo BT than those in holo BT are indicated in gray and light gray, respectively. The secondary structure elements are shown as ribbons. The catalytic triad (S195, H57, and D102) and the specificity pocket amino acids (G216, G206, and D189) are shown as sticks and dots.

(in particular, the region 222−227). The two latter loops are located in proximity of the substrate specificity pocket. Therefore, the removal of Ca^{2+} not only increases the flexibility of regions around its binding site in the N-terminal domain but also leads to channeling of the fluctuations to remote sites in the C-terminal domain. In particular, the analysis of the rmsf data in light of the three-dimensional structure of BT and ST suggests that the channeling could involve the interdomain loop, which is a long disordered region that connects the two protein domains.

The loop between Nβ-1 and Nβ-2, where the primary autolysis site K60 is located, is the only region surrounding the calcium-binding loop whose flexibility is not influenced by calcium removal, as a result of the presence of a disulfide bridge (C42−C58) that locally constrains the structure. The behavior of the autolysis loop, where BT K145 and ST K154 are located, can be explained by similar considerations. In fact, three disulfide bridges are located around this region: C22−C157, C136−C201, and C168−C192. Among the primary sites of autolysis in BT, R117, which is located in the interdomain loop, and K188 are the only residues characterized by significantly larger flexibility in the apo form. Notably, in ST, R117 is not conserved and the flexibility of K188 does not increase significantly upon Ca^{2+} removal. The latter observation is in good agreement with the experimental data indicating weak dependence of the autolysis rate on Ca^{2+} binding in ST.[16−18] The reliability of the rmsf analysis has been validated, splitting the MD trajectories (after equilibration) and computing the rmsf values for R117 and K188 separately for every trajectory portion (Table 2).

Table 2. rmsf Values for R117 and K188[a]

	R117	K188
holo $BT_{merged\ sim}$	0.101	0.064
holo BT_{sim1}	0.083	0.053
holo $BT_{sim1\ (first\ half)}$	0.057	0.048
holo $BT_{sim1\ (second\ half)}$	0.090	0.053
holo BT_{sim2}	0.117	0.069
holo $BT_{sim2\ (first\ half)}$	0.126	0.070
holo $BT_{sim2\ (second\ half)}$	0.055	0.063
holo BT_{sim3}	0.075	0.059
holo $BT_{sim3\ (first\ half)}$	0.094	0.061
holo $BT_{sim3\ (second\ half)}$	0.069	0.055
apo $BT_{merged\ sim}$	0.130	0.115
apo BT_{sim1}	0.154	0.107
apo $BT_{sim1\ (first\ half)}$	0.148	0.080
apo $BT_{sim1\ (second\ half)}$	0.123	0.090
apo BT_{sim2}	0.107	0.112
apo $BT_{sim2\ (first\ half)}$	0.096	0.090
apo $BT_{sim2\ (second\ half)}$	0.144	0.089
holo $ST_{merged\ sim}$		0.099
holo ST_{sim1}		0.088
holo $ST_{sim1(first\ half)}$		0.102
holo $ST_{sim1\ (second\ half)}$		0.071
holo ST_{sim2}		0.065
holo $ST_{sim2\ (first\ half)}$		0.064
holo $ST_{sim2\ (second\ half)}$		0.071
apo $ST_{merged\ sim}$		0.095
apo ST_{sim1}		0.051
apo $ST_{sim1\ (first\ half)}$		0.050
apo $ST_{sim1\ (second\ half)}$		0.051
apo ST_{sim2}		0.093
apo $ST_{sim2\ (first\ half)}$		0.089
apo $ST_{sim2\ (second\ half)}$		0.096

[a] The rmsf values for the merged trajectories are highlighted in bold. To evaluate the reliability of the rmsf analysis, single trajectrories have also been split and rmsf values have been computed for every trajectory portion. Values in nm. R177 values for ST are not reported because this amino acid is not conserved in fish trypsins.

As for other putative autoproteolysis sites, only ST K23, ST K74 (in the calcium-binding loop), BT K222, and BT K224 are localized in regions characterized by relevant rmsf differences between the apo and holo forms.

Since the susceptibility to autoproteolysis has been proposed to be related to both flexibility[19] and solvent accessibility of the autolysis sites,[20] the solvent-accessible area of putative autolysis sites has been monitored during MD simulations. The analysis of trajectories of the holo forms reveals that some putative autolysis sites located in loop regions (ST R62, ST K74, BT K109, ST K110, BT K204, BT K222, BT K224) present large solvent accessibility values when compared to primary autolysis sites (Table 3). In addition, the primary autolysis site BT R117, which is characterized by relatively high flexibility (Figure 4), presents low values of solvent accessibility.

The comparison of solvent accessibility values for primary autolysis sites in holo and apo simulations (Table 3) reveals that the removal of Ca^{2+} causes only small effects on some putative autolysis sites (K23, R62, R66, K110, and K154), whereas all other sites are essentially not affected by Ca^{2+} removal. In particular, solvent accessibility for BT K145 and ST K154 are not significantly different in the apo and holo

Table 3. Average Solvent Accessibility Degree (%) of Lysine and Arginine Side Chains for Putative Autolysis Sites in Holo and Apo BT and ST, According to the NACCESS Program[42] [a]

	BT		ST	
	holo	apo	holo	apo
K23	- - -	- - -	89.2 (10.0)	74.0 (17.7)
K60	**54.9 (8.5)**	**52.3 (8.1)**	**51.9 (6.9)**	**46.0 (7.6)**
R62	- - -	- - -	71.1 (10.5)	65.4 (7.6)
R66	21.9 (9.1)	17.9 (5.1)	27.2 (7.9)	19.0 (7.3)
K74	- - -	- - -	72.9 (5.8)	73.8 (6.1)
K87	64.2 (9.7)	63.0 (9.3)	63.0 (12.7)	61.3 (7.9)
R90	- - -	- - -	38.1 (7.7)	39.0 (6.1)
K107	26.5 (7.0)	27.5 (7.9)	31.8 (6.9)	26.7 (4.6)
K109	66.4 (6.2)	69.6 (6.5)	- - -	- - -
K110	- - -	- - -	81.0 (8.6)	72.3 (10.8)
R117	**20.3 (14.6)**	**14.7 (6.0)**	- - -	- - -
K145	**63.4 (6.1)**	**63.2 (5.8)**	- - -	- - -
K154	- - -	- - -	43.9 (6.1)	38.8 (7.2)
K156	26.6 (3.8)	25.8 (3.8)	- - -	- - -
K159	51.6 (5.8)	54.2 (5.7)	- - -	- - -
K169	48.8 (8.3)	52.0 (11.3)	- - -	- - -
K188	**40.5 (8.7)**	**51.5 (10.3)**	**45.0 (7.7)**	**52.2 (9.5)**
K204	56.0 (7.7)	55.4 (8.1)	- - -	- - -
K222	82.2 (7.8)	81.6 (8.2)	- - -	- - -
K224	40.1 (7.5)	37.1 (7.6)	- - -	- - -
K230	9.6 (3.0)	15.8 (7.5)	16.1 (6.0)	16.8 (4.9)
K239	72.0 (8.2)	70.8 (8.3)	- - -	- - -

[a] Standard deviations are given in parentheses. The primary autolysis sites are highlighted in bold, and - - - indicates that, in the corresponding position, lysine or arginine residues are missing.

Table 4. Average Solvent Accessibility Degree of Side Chains of W141 and Hydrophobic Core Residues (%) of BT and ST at 310 K in the Holo and Apo Forms, According to the NACCESS Program[a]

	BT		ST	
	holo	apo	holo	apo
W141	2.2 (1.3)	1.7 (1.3)	1.0 (0.9)	1.4 (1.0)
47	3.0 (2.6)	2.2 (2.2)	2.5 (3.0)	2.1 (2.1)
238	2.3 (2.7)	1.0 (1.8)	0.6 (1.0)	0.2 (0.4)
242	25.7 (6.3)	28.0 (5.8)	9.5 (4.4)	14.2 (3.8)
63	4.9 (3.9)	4.3 (3.5)	1.9 (1.7)	0.04 (0.2)
85	0.7 (1.8)	0.3 (1.3)	0.5 (1.1)	0.7 (2.0)
88	4.2 (6.7)	2.4 (4.4)	2.0 (2.7)	2.7 (4.1)
138	0.2 (0.5)	0.1 (0.2)	0.5 (0.7)	0.4 (0.9)
160	0.1 (0.5)	0.5 (1.3)	1.2 (1.4)	0.2 (0.3)
183	0.2 (0.4)	0.1 (0.3)	0.1 (0.3)	0.03 (0.2)

[a] Standard deviations are given in parentheses.

forms. This may be due to the presence of two stable ion pairs (K154−D151 and H71−E21, in ST) in their proximity. On the other hand, relevant differences in solvent accessibility are evident for BT R117 and K188 when comparing apo and holo trajectories.

Experimental data[6−10] indicate that trypsin assumes a more compact structure upon calcium coordination. In particular, the environment of W141 is affected by metal binding. Analyses of MD trajectories indicate that the solvent accessibility degree for W141 and amino acids forming the three trypsin core regions[48] (47, 238, 242; 63, 85, 88; 138, 160, 183) remains nearly unchanged upon calcium removal (Table 4). The only significant difference is observed for residue 242 (Ile in BT and Met in ST), which is more solvent-exposed and, consequently, less involved in hydrophobic interactions in the apo forms.

Conclusions

The present investigation has probed some effects of Ca^{2+} binding to trypsins, correlating conformational and dynamic properties to relevant functional features of the protein.

Several putative autoproteolytic sites (lysine and arginine residues) are present in ST and BT, and some of the sites are characterized by suitable features for proteolytic cleavage in terms of location in the loop regions, lack of ion pairs or disulfide bridges in their proximity, and solvent accessibility. Most of the potential trypsin-sensitive sites are located in the K60−R117 stretch, which does not include disulfide bridges and which was previously proposed as a built-in target for autolysis.[13]

The removal of Ca^{2+} affects the structural and dynamic properties of specific regions of trypsin. The effects caused by calcium removal on the three-dimensional structure are more pronounced in the N-terminal, where a decrease in the number of residues in β-strand conformations is observed, than in the C-terminal domain. In fact, the N-terminal domain presents a low number of disulfide bridges. Ca^{2+} removal increases the flexibility of regions around its binding site but also leads to channeling of the fluctuations to sites located in the C-terminal domain. This observation is in agreement with other general mechanisms by which the signal induced by metal binding is transmitted to remote regions in the 3D structure.[26−29] In particular, the long disordered interdomain loop, which connects the two globular trypsin domains, is proposed to be involved in the transmission of signals correlated to Ca^{2+} binding.

It was previously observed that the autoproteolysis rate at R117 in rat trypsin is strongly decreased upon Ca^{2+} binding, leading to the suggestion that metal binding can stabilize an autolysis-resistant conformation of the protein.[13] The present results reveal that R117 (in the interdomain loop) and K188 (in the C-terminal domain) are the only primary autolysis sites which are strongly influenced by Ca^{2+} binding in BT, disclosing the molecular relationship connecting Ca^{2+} binding to autoproteolysis propensity in mammalian trypsins. This conclusion is in nice agreement with a site-directed mutagenesis experiment, which indicated that R117 replacement in rat trypsin leads to proteins almost completely resistant to autolysis.[13,15] Remarkably, R117 is not conserved in fish trypsins and the flexibility of K188 is not significantly affected by Ca^{2+} removal, in agreement with the observed weak dependence of autolysis propensity on Ca^{2+} binding in fish trypsins.[16−18]

Acknowledgment. The authors thank L. Riccardi and R. Gonella Diaza for fruitful discussions and comments.

References

(1) Hedstrom, L. Serine protease mechanism and specificity. *Chem. Rev.* **2002**, *102*, 4501−4524.

(2) Burdette, S. C.; Lippard, S. J. Meeting of the minds: metalloneurochemistry. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3605−3610.

(3) Smalas, A. O.; Heimstad, E. S.; Hordvik, A.; Willassen, N. P.; Male, R. Cold adaptation of enzymes: structural comparison between salmon and bovine trypsins. *Proteins* **1994**, *20*, 149−166.

(4) Read, R. J.; James, M. N. Refined crystal structure of *Streptomyces griseus* trypsin at 1.7 Å resolution. *J. Mol. Biol.* **1988**, *200*, 523−551.

(5) Toth, M. S.; Graf, L.; Toth, M. Trypsinogen stabilization by mutation Arg 117→His: a unifying pathomechanism for hereditary pancreatitis? *Biochem. Biophys. Res. Commun.* **1999**, *264*, 505−508.

(6) Sipos, T.; Merkel, J. R. An effect of calcium ions on the activity, heat stability and structure of trypsin. *Biochemistry* **1970**, *9*, 2766−2775.

(7) Gabel, D.; Kasche, V. Autolysis of beta-trypsin. Influence of calcium ions and heat. *Acta Chem. Scand.* **1973**, *27*, 1971−1981.

(8) Bulaj, G.; Otlewski, J. Denaturation of free and complexed bovine trypsinogen with the calcium ion, dipeptide Ile-Val and basic pancreatic trypsin inhibitor. *Eur. J. Biochem.* **1994**, *223*, 939−946.

(9) Bulaj, G.; Otlewski, J. Ligand-induced changes in the conformational stability of bovine trypsinogen and their implications for the protein function. *J. Mol. Biol.* **1995**, *247*, 701−716.

(10) Buono, R. A.; Prestrelski, S. J.; Liebman, M. N.; Byler, D. M. Infrared spectroscopic studies of calcium binding to inhibited beta-trypsins. *Biochim. Biophys. Acta* **1994**, *204*, 124−128.

(11) Smith, R. L.; Shaw, E. Pseudotrypsin. A modified bovine trypsin produced by limited autodigestion. *J. Biol. Chem.* **1969**, *244*, 4704−4712.

(12) Maroux, S.; Desnuelle, P. On some autolyzed derivatives of bovine trypsin. *Biochim. Biophys. Acta* **1969**, *181*, 59−72.

(13) Varallyay, E.; Pal, G.; Patthy, A.; Szilagyi, L.; Graf, L. Two mutations in rat trypsin confer resistance against autolysis. *Biochem. Biophys. Res. Commun.* **1998**, *243*, 56−60.

(14) Kaslik, G.; Patthy, A.; Balint, M.; Graf, L. Trypsin complexed with alpha 1-proteinase inhibitor has an increased structural flexibility. *FEBS Lett.* **1995**, *370*, 179−183.

(15) Li, X. F.; Nie, X.; Tang, J. G. Anti-autolysis of trypsin by modification of autolytic site Arg117. *Biochem. Biophys. Res. Commun.* **1998**, *250*, 235−239.

(16) Outzen, H.; Berglund, G. I.; Smalas, A. O.; Willassen, N. P. Temperature and pH sensitivity of trypsins from Atlantic Salmon in comparison with bovine and porcine trypsin. *Comp. Biochem. Physiol. B* **1996**, *115*, 33−45.

(17) Hjelmeland, K.; Raa, J. Characteristics of two trypsin type isozymes isolated from the artic fish capelin. *Comp. Biochem. Physiol. B* **1982**, *71*, 557−562.

(18) Reeck, G. R.; Neurath, H. Pancreatic trypsinogen from the African lungfish. *Biochemistry* **1972**, *11*, 503−510.

(19) Fontana, A.; Fassina, G.; Vita, C.; Dal Zoppo, D.; Zamai, M.; Zambonin, M. Correlation between sites of limited proteolysis and segmental mobility in thermolysin. *Biochemistry* **1986**, *25*, 1847−1951.

(20) Novotny, J.; Bruccoleri, R. E. Correlation among sites of limited proteolysis, enzyme accessibility and segmental mobility. *FEBS Lett.* **1987**, *211*, 185−189.

(21) Norberg, J.; Nilsson, L. Advances in biomolecular simulations: methodology and recent applications. *Q. Rev. Biophys.* **2003**, *36*, 257−306.

(22) Wang, W.; Donini, O.; Reyes, C. M.; Kollman, P. A. Biomolecular simulations: recent developments in force fields, simulations of enzyme-catalysis, protein−ligand, protein−protein, and protein−nucleic acid noncovalent interactions. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 211−243.

(23) Hansson, T.; Oostenbrink, C.; van Gusteren, W. Molecular dynamics simulations. *Curr. Opin. Struct. Biol.* **2002**, *12*, 190−196.

(24) Heimstad, E. S.; Hansen, L. K.; Smalas, A. O. Comparative molecular dynamics simulation studies of salmon and bovine trypsins in aqueous solution. *Protein Eng.* **1995**, *8*, 379−388.

(25) Brandsdal, B. O.; Heimstad, E. S.; Sylte, I.; Smalas, A. O. Comparative molecular dynamics of mesophilic and psycrophilic protein homologues studied by 1.2 ns simulations. *J. Biomol. Struct. Dyn.* **1999**, *17*, 493−506.

(26) Iyer, L. K.; Qasba, P. K. Molecular dynamics simulation of α-lactalbumin and calcium binding c-type lysozyme. *Protein Eng.* **1999**, *12*, 129−139.

(27) Cates, S. M.; Teodoro, M. L.; Phillips, G. N. Molecular mechanism of calcium and magnesium binding to parvalbumin. *Biophys. J.* **2002**, *82*, 1133−1146.

(28) Yang, C.; Jas, G. S.; Kuczera, K. Structure, dynamics and interaction with kinase targets: computer simulations of calmodulin. *Biochim. Biophys. Acta* **2004**, *1697*, 289−300.

(29) Komeiji, Y.; Ueno, Y.; Uebayasi, M. Molecular dynamics simulations revealed Ca(2+)-dependent conformational change of calmodulin. *FEBS Lett.* **2002**, *521*, 133−139.

(30) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissing, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.

(31) Altschul, S. F.; Madden, T. L.; Schaffer, A. A.; Zhang, Z.; Miller, W.; Lipman, D. J. Gapped BLAST and PSI−BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389−3402.

(32) Higgins, D.; Thompson, J.; Gibson, T.; Thompson, J. D.; Higgins, D. G.; Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalities and weight matrix choice. *Nucleic Acids Res.* **1994**, *22*, 4673−4680.

(33) van der Spoel, D.; van Buuren, A. R.; Apol, E.; Meulenhoff, P. J.; Tieleman, D. P.; Sijbers, A. L. T. M.; Hess, B.; Feenstra, K. A.; Lindhal, E.; van Drunen, R.; Berendsen, H. J. C. *Gromacs User Manual version 3.1.1*, * University of Groningen: Groningen, The Netherlands, 2002. www.gromacs.org.

(34) Marquart, M.; Walter, J.; Deisenhofer, J.; Bode, W.; Huber, R. The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors. *Acta Crystallogr., Sect. B.* **1983**, *39*, 480.

(35) Smalas, A. O.; Hordvik, A.; Hansen, L. K.; Hough, E.; Jynge, K. Crystallization and preliminary X-ray crystallographic studies of benzamidine-inhibited trypsin from the North Atlantic Salmon (Salmo salar). *J. Mol. Biol.* **1990**, *214*, 355−358.

(36) Berendsen, H. J. C.; Postma, J. P. M.; van Gusteren, W. F.; Hermans, J. *Interaction models for water in relation to protein hydratation*; Reidel: Dordrecht, The Netherlands, 1981; pp 331−342.

(37) Berendsen, H. J. C.; Postma, J. P. M.; van Gusteren, W. F.; Di Nola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.

(38) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraahje, J. G. E. M. LINCS: a linear constraint solver for molecular interactions. *J. Comput. Chem.* **1997**, *18*, 1463−1472.

(39) Darden, T. A.; York, D. M.; Pedersen, L. G. Particle mesh Ewald. An N log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.

(40) Jarvis, R. A.; Patrick, E. A. Clustering using a similarity measure based on shared nearest neighbors. *IEE Trans. Comput.* **1973**, *11*, 1025−1034.

(41) Kabsch, W.; Sander, C. Dictionary of Protein Secondary Structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577−2637.

(42) Hubbard, S. J.; Thortnon, J. M. *NACCESS*; 1993.

(43) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33−38.

(44) Tsai, J.; Levitt, M.; Baker, D. Hierarchy of structure loss in MD simulations of src SH3 domain unfolding. *J. Mol. Biol.* **1999**, *291*, 215−225.

(45) Hess, B. Convergence of sampling in protein simulations. *Phys. Rev. E.* **2002**, *65*, 031910/1−031910/10.

(46) Smith, L. J.; Daura, X.; van Gusteren, W. F. Assesing equilibration and convergence in biomolecular simulations. *Proteins* **2002**, *48*, 487−496.

(47) Note that the analysis of single trajectories led to similar results, with some differences in the intensity of the peaks in the rmsf plots.

(48) Leiros, H. K.; Willassen, N. P.; Smalas, A. O. Structural comparison of psycrophilic and mesophilic trypsins. Elucidating the molecular basis of cold-adaptation. *Eur. J. Biochem.* **2000**, *267*, 1039−1049.

CT050092O