Pre-bound State Discovered in the Unbinding Pathway of Fluorinated Variants of the Trypsin-BPTI Complex Using Random Acceleration Molecular Dynamics Simulations

Leon Wehrhan and Bettina G. Keller*[a)]
Department of Biology, Chemistry, and Pharmacy, Freie Universität Berlin, Institute of Chemistry and Biochemistry, Structural Biochemistry, Arnimallee 22, Berlin, 14195 Germany

(Dated: 22 February 2024)

ABSTRACT

The serine protease trypsin forms a tightly bound inhibitor complex with Bovine Pancreatic Trypsin Inhibitor (BPTI). The complex is stabilized by the P1 residue Lys15, which interacts with the negatively charged amino acids at the bottom of the S1 pocket. Truncating the P1 residue of wildtype BPTI to α-aminobutyric acid (Abu) leaves a complex with moderate inhibitor strength, which is held in place by additional hydrogen bonds at the protein-protein interface. Fluorination of the Abu residue partially restores inhibitor strength. The mechanism with which fluorination can restore the inhibitor strength is unknown and accurate computational investigation requires knowledge of the binding and unbinding pathways. The preferred unbinding pathway is likely to be complex, as encounter states have been described before and unrestrained Umbrella Sampling simulations of these complexes suggest additional energetic minima. Here, we use Random Acceleration Molecular Dynamics to find a new metastable state in the unbinding pathway of Abu-BPTI variants from trypsin, which we call the pre-bound state. The pre-bound state and the fully bound state differ by a substantial shift in the position, a slight shift in the orientation of the the BPTI variants and change in the interaction pattern. Particularly important is the breaking of three hydrogen bonds around Arg17. Fluorination of the P1 residue lowers the energy barrier of the transition between fully bound state and pre-bound state and also lowers the energy minimum of the pre-bound state. While the effect of fluorination is in general difficult to quantify, here it is in part caused by a favorable stabilization of a hydrogen bond between Gln194 and Cys14. The interaction pattern of the pre-bound state offers insight into the inhibitory mechanism of BPTI and might add valuable information for the design serine protease inhibitors.

[a)] Electronic mail: bettina.keller@fu-berlin.de

I. INTRODUCTION

Proteases are enzymes that play a crucial role in the breakdown of peptides by catalyzing the hydrolysis of peptide bonds. Among them, serine proteases form a subgroup which catalyze this reaction via a serine residue in their active site. Serine proteases are found in most life forms, including bacteria, viruses, fungi, plants, and animals. They play essential roles in digestion, signal transduction, blood clotting, immune responses, and other cellular functions. In humans, serine proteases are important drug targets for many diseases including cardiovascular, cancer and infectious diseases.[1,2] An example of a serine protease is trypsin, a mammalian digestive enzyme. It has been widely used as model system for serine proteases, since it exhibits the most prevalent fold for proteases in humans and higher organisms.[3,4]

Protein-protein complexes involving trypsin are stabilized by a long positively charged residue located on the binding protein, the P1 residue, which reaches into the deep S1 binding pocket of trypsin (Schechter and Berger notation)[5]. The S1 pocket is lined with negatively charged residues which either bind directly to the positive charge or via water-mediated contacts[6,7]. Trypsin's catalytic site is located at the rim of the S1 binding pocket. Most proteins that bind to trypsin in this manner are cleaved at the C-terminal side of their P1 residue and thus act as substrates. Some proteins, despite binding to the S1 pocket, are not cleaved and instead act as inhibitors towards trypsin. Examples are Bovine Pancreatic Trypsin Inhibitor (BPTI), Antitrypsin or Serpins. Several mechanisms have been put forward to explain why these proteins are not hydrolyzed by trypsin, but instead form such a stable trypsin-inhibitor complex. Possibly, initial hydrolysis might take place, but re-ligation of the cleaved bond is fast and thus favoured over release of the hydrolyzed product[8]. In the clogged gutter mechanism[9], the hydrolyzed products are bound in a tightly and specific orientation to trypsin, such that product release is hindered.

We here focus on BPTI, which is an exceptionally well studied protein[9–13] and inhibits trypsin with an extraordinarily high binding affinity (the binding constant is $K_i = 5 \cdot 10^{-24} M$).[14] Its P1 residue is Lys15, which forms water mediated bonds to Asp189 and Ser190 at the bottom of the S1 pocket. The importance of Lys15 for the

binding process has been demonstrated by kinetic studies with BPTI mutants, where the K15A mutant BPTI shows dramatically decreased binding affinity.[14] Interestingly, the K15A mutant is also the only variant that has a significantly decreased association rate, highlighting the importance of the P1 residue for the trypsin-BPTI recognition.

While for complexes of proteins with small molecules, like the trypsin-benzamidine complex, the full energy landscape of the binding and unbinding process has been calculated,[15] the computational characterization of the binding-equilibrium in protein-protein complexes, like the BPTI-trypsin complex, is much more challenging[16,17]. The reasons for this include the slow movements of macromolecules along translational and rotational degrees of freedom. Also, the number of possible contact conformations of a protein-protein complex far exceeds that of a protein-small-molecule complex. In computational studies of protein-protein complexes additional restraints to the relative position and orientation may be applied to increase the sampling of the binding/unbinding process.[18] However this requires knowledge of the exact binding/unbinding path to obtain an accurate free energy profile and characterize relevant intermediate states.

Kahler *et al.*[19] studied the binding/unbinding process of wildtype BPTI with trypsin using unbiased simulations, seeded by umbrella simulations. They describe the binding/unbinding process as a two-step mechanism, in which trypsin and BPTI recognize each other first through Coulomb interactions and form encounter states before moving on to form the fully bound protein-protein complex.

Because of the extremely high binding constant, the binding/unbinding process of the wild-type BPTI-trypsin complex is particularly difficult to study computationally. We here study the unbinding process of BPTI variants where the Lys15 residue has been mutated to α-aminobutyric acid (Abu) and its mono- (MfeGly), di- (DfeGly) and tri-fluorinated (TfeGly) variants. These BPTI variants are not cleaved by trypsin, but instead act as moderate inhibitors with half-maximal inhibitory concentration (IC50) of $\mathrm{IC}_{50} = 4 \cdot 10^{-7}\,\mathrm{M}$ (Lys15Abu) and $\mathrm{IC}_{50} = 6 \cdot 10^{-8}\,\mathrm{M}$ (Lys15TfeGly)[20,21].

Interestingly, the inhibitor strength of the four BPTI variants systematically increases with increasing fluorination. An initial hypothesis based on crystal structures of the complexes suggested that this increase in binding affinity could be traced back to direct and specific interactions of the fluorine substituents with the water molecules in the S1 pocket[20]. In a recent computational study with Molecular Dynamics (MD) simulations, we did not find significant differences in the water structure or water-protein interaction strength across the four variants of the BPTI-trypsin complex[21], and thus could not confirm this hypothesis. However, a rough scan using umbrella sampling of the unbinding pathways hinted at a second free-energy minimum, next to the bound state. This pre-bound state was closer to the bound state than encounter states[19], which could also be identified in our scan of the unbinding pathway. The existence of a pre-bound state might offer insights into why BPTI acts as a inhibitor rather than a substrate to trypsin, and might open up new avenues for the design of trypsin inhibitors.

In this contribution, we investigate the unbinding path of the four BPTI variants Abu-BPTI, MfeGly-BPTI, DfeGly-BPTI and TfeGly-BPTI using Random Acceleration Molecular Dynamics (RAMD) simulations[22–25]. RAMD is an enhanced sampling method that applies an additional biasing force, which is randomly redirected throughout the simulation, to the center of mass of a ligand and thereby facilitates the exploration of curved unbinding pathways.[22] RAMD has been used frequently for complexes of proteins with small molecules, but, to the best of our knowledge, this is the first application of RAMD on a protein-protein complex. Our goal is to verify the presence of the pre-bound state and to explain its stability.

II. METHODOLOGY

A. Collective Variables

We constructed the collective variables describing the position and orientation of the BPTI variants with respect to trypsin from the positions of three reference points in trypsin (T1, T2, T3) and three reference points in (Abu, MfeGly, DfeGly, TfeGly)-BPTI (B1, B2, B3), adapted from ref. 18. The three reference points for the enzyme trypsin were defined to be the center-of-mass of the backbone of the whole enzyme (T1), the backbone of Val233-Ala241 (T2) and the backbone of Gln46-Leu67 (T3). The reference points of the ligands (Abu, MfeGly, DfeGly, TfeGly)-BPTI were defined to be the center-of-mass of the backbone of the whole ligand (B1), the backbone of Ala48-Thr54 (B2) and the backbone of Cys14-Ala16 (B3). The positions of the reference points in the starting structure are sketched in fig.2. The main collective variable is the center-of-mass distance r between the enzyme and the ligand (T1-B1). The angle θ_o (T1-B1-B2) and the dihedrals ϕ_o (T2-T1-B1-B2) and ψ_o (T1-B3-B1-B2) describe the orientation of the ligand with respect to the enzyme. The angle Θ_p (T2-T1-B1) and the dihedral Φ_p (T3-T2-T1-B1) describe the position of the ligand with respect to the enzyme. The collective variables r, Θ_p, Φ_p, θ_o, ϕ_o and ψ_o were calculated with Plumed[26,27] 2.8 for all simulations.

B. Molecular Dynamics General Methods

We ran all Molecular Dynamics (MD) simulations using Gromacs[28–30] software and our self-parameterized Amber14SB force field[21,31–33]. Energy minimizations were conducted with the steepest descend algorithm.

Equilibrations in the NVT ensemble were using a velocity rescaling scheme with a stochastic term[34] to keep the temperature at 300 K and harmonic restraints were applied on all protein heavy atom positions. Subsequent equilibrations in the NPT ensemble without restraints made use of the same velocity rescaling scheme with a stochastic term and the Parinello-Rahman barostat[35] to keep the temperature at 300 K and the pressure at 1.0 bar. Production MD simulations were run in the NPT ensemble at 300 K and 1.0 bar using the same thermostat and barostat. All MD simulations were performed with the leap-frog integrator and an integration time step of 2 fs. Bond lengths involving hydrogen atoms were kept constant using the LINCS[36] algorithm.

C. Starting Structure Preparation

Starting structures were generated from the crystal structure of the TfeGly-BPTI-trypsin complex (pdb code: 4Y11)[20]. Co-solutes and ions were deleted and appropriate hydrogen atoms were added to the crystal structure using the pdbfixer software. The three histidine side chains in the complex were protonated at N(ϵ) and N(δ). From this initial starting structure, the TfeGly residue was transformed into DfeGly, MfeGly and Abu, respectively, to yield one initial starting structure for every BPTI variant. For the RAMD simulations, the initial starting structures were placed inside a cubic box with periodic boundary conditions with 2.1 nm distance between the solute and the box edges and solvated in TIP3P[37] water. The systems were energy minimized and equilibrated in the NVT ensemble for 100 ps, followed by an equilibration in the NPT ensemble for 1 ns. To generate two more replicas for each of the complexes, two subsequent simulations of 10 ns were run with the equilibrated starting structures to yield the starting structures for the next replicas.

D. Random Acceleration Molecular Dynamics (RAMD)

We used Random Acceleration Molecular Dynamics (RAMD) to explore unbinding pathways of (Abu, MfeGly, DfeGly, TfeGly)-BPTI from trypsin using Gromacs2020.5-RAMD-2.0. Two pull groups were defined: one included all atoms of trypsin, the other included all atoms of (Abu, MfeGly, DfeGly, TfeGly)-BPTI. A random force acting between the two pull groups with a force constant of 3500 kJ/(mol $\cdot$ nm^2) was applied. The appropriate force constant was estimated by running single RAMD simulations of the TfeGly-BPTI-trypsin complex starting with a force constant of 250 kJ/(mol$\cdot$nm^2) and raising the force constant by 250 kJ/(mol $\cdot$ nm^2) every simulation until dissociation within 10 ns was achieved. Three starting structures for each of the four complexes of trypsin with Abu-BPTI, MfeGly-BPTI, DfeGly-BPTI and TfeGly-BPTI were generated as described above. For every of these replicas, ten RAMD simulations were run from the same starting structure, where the random seed of the random force was changed. The simulations were stopped after dissociation was achieved, the maximum length of the simulations was set to be 40 ns. In the beginning of the simulations, the direction of the biasing force was chosen at random. Throughout the simulations, after every 100 fs, the direction of the force was either retained, if the center of mass of the second pull group moved by more than 0.0025 nm, or changed randomly, if this was not the case. Snapshots were extracted every 2 ps.

E. Unbiased Molecular Dynamics of Pre-bound state

We ran unbiased Molecular Dynamics simulations to sample the fully bound state and the pre-bound state of the four complexes of trypsin with Abu-BPTI, MfeGly-BPTI, DfeGly-BPTI and TfeGly-BPTI using Gromacs2021.5[28–30], patched with Plumed[26,27] 2.8. For every complex and state, 20 simulations of 50 ns length were run, totalling 160 simulations with an aggregated length of 8 μs. Initial starting structures for the simulations of the fully bound state were generated from the pdb structure of TfeGly-BPTI as described above. The initial starting structures were placed in a cubic box with periodic boundary conditions with 1.5 nm distance between the solute and the box edges and solvated in TIP3P water. Then, 20 starting structures of every complex for the production MD simulations were generated by individually energy minimizing the systems, followed by equilibration in the NVT ensemble for 100 ps and equilibration in the NPT ensemble for 1 ns. Initial starting structures for the simulations of the pre-bound state were generated by extracting the coordinates of all protein atoms of 20 snapshots of one RAMD simulation of the TfeGly-BPTI-trypsin complex, when the system was in the pre-bound state. The TfeGly residue was transformed into DfeGly, MfeGly and Abu to yield initial starting structures for the other three complexes. The initial starting structures were individually placed in a box with periodic boundary conditions and energy minimized and equilibrated in the same way as the starting structures for the fully bound state. Production MD simulations were run for a length of 50 ns for every replica. Snapshots were extracted every 10 ps.

F. Analysis of Distances and Hydrogen Bonds

We calculated atomic distances and detected hydrogen bonds in simulation snapshots using the python package MDTraj[38]-1.9.4. Hydrogen bonds were detected using the Wernet-Nilsson criterion[39] implemented in MDTraj:

$$r_{DA} < 0.33\text{nm} - 0.00044 \cdot \delta^2_{HDA} \quad (1)$$

with the donor-acceptor distance r_{DA} and the angle between hydrogen atom, donor and acceptor δ_{HDA}.

Distances to the nitrogens in the guanidine moieties of arginine side chains were calculated by computing the distance of the respective interaction partner to all three nitrogen atoms of guanidine moiety and taking the minimum of these three distances for every simulation snapshot.

G. Umbrella Sampling

We conducted Umbrella Sampling using Gromacs2021.5[28–30], patched with Plumed[26,27] 2.8, based on the distance between the backbone oxygen of Phe41 (Phe41-O) of trypsin and the backbone nitrogen of Arg17 (Arg17-N) of the BPTI variants as the main collective variable ξ. Starting structures for the umbrella windows were generated starting from the fully bound crystal structure of the Abu-BPTI-trypsin complex as described above. An initial harmonic restraint with a force constant of 6276 kJ/(mol $\cdot$ nm^2) was placed at $\xi = 0.25$ nm. The system was equilibrated in the NPT ensemble with this harmonic restraint for 500 ps to yield the starting structure of the first umbrella window. Then, the harmonic potential was shifted by 0.05 nm and a new NPT equilibration of 500 ps was run to yield the starting structure of the next window. This procedure was repeated until ξ reached a value of 0.70 nm. Additional windows were added in-between at ξ values of 3.75 nm, 4.25 nm and 4.75 nm to achieve better sampling of the region of the free energy barrier. Finally there were 13 umbrella windows at the following positions of ξ (all in nm): 0.250, 0.300, 0.350, 0.375, 0.400, 0.425, 0.450, 0.475, 0.500, 0.550, 0.600, 0.650, 0.700. In each of the umbrella windows, a production MD simulation with a harmonic restraint with a force constant of 6276 kJ/(mol $\cdot$ nm^2) was run for a length of 30 ns. The potential of mean force profiles were calculated using binless WHAM[40,41]. Statistical uncertainty was estimated using a simplified Bootstrapping scheme: The simulations of every window were separated into five parts of 6 ns length. Then, for every window, five combinations of four of these parts were constructed by combining all parts but one. The WHAM calculation was performed on all of these five combinations and the mean and standard deviation of the resulting potential of mean force profiles was calculated.

III. RESULTS AND DISCUSSION

A. Protein-Protein Complex between Trypsin and BPTI Variants

We consider BPTI-variants in which the P1 residue is substituted by α-aminobutyric acid (Abu) and its mono- (MfeGly), di- (DfeGly) and tri-fluorinated (TfeGly), i.e., K15Abu, K15MfeGly, K15DfeGly, K15TfeGly; see fig.10a). The crystal structures of all four BPTI-variants (pdb codes: 4Y0Z, 7PH1, 4Y10, 4Y11, 4Y0Y) are similar to each other and to the wildtype complex[20,21]. The interaction strength and pattern between the P1 residue and the S1 pocket and water molecules within the S1 pocket did not differ significantly across BPTI variants, and thus did not explain the observed differences in the stability of the protein-protein complexes[21]. Fig. 1 shows the complex and binding interface of the TfeGly-BPTI complex as a representative.

Besides the S1-P1 interactions, the complex is stabilized by hydrogen bonds throughout the whole protein-protein interface. Fig.1b shows that the P1 residue TfeGly is held in place by seven hydrogen bond like contacts, most notably three backbone interactions holding the backbone carbonyl of the P1 residue in the oxyanion hole of the catalytic pocket. To the left in fig.1b, Arg39 of the BPTI variant can be seen in two alternative conformations, forming interactions to either the side chain or the backbone of Asn97 in trypsin.

Fig. 1c shows the other side of the interface. On this side of the interface, Arg17 (P2' residue) of the BPTI variant forms interactions with its side chain to the backbone of His40 and with its backbone to the backbone of Phe41. The P4' residue Ile19 forms an interaction to the side chain of Tyr39 in Trypsin. This interaction has been described as important for the binding of BPTI to Trypsin, as Y39A mutants of Trypsin are less sensitive to BPTI.[3]

B. Collective Variables and Metadynamics with Restraints

To investigate the dissociation of the complexes between the BPTI variants and trypsin, we designed a set of collective variables, that describe the position and orientation of the (Abu, Mfegly, DfeGly, TfeGly)-BPTI with respect to trypsin, following Ref. 18. The collective variables are based on the backbone center of mass of the BPTI variants (B1) and trypsin (T1) and two additional points for each of the proteins (T2,T3 and B2, B3), defined as centers-of-mass of well-structured regions inside the proteins (fig. 2). The position of the BPTI-variant relative to trypsin is then given by the distance r between B1 and T1, the angle $\Theta_p = \angle$T2-T1-B1, and the dihedral angle $\Phi_p = \angle$T3-T2-T1-B1. The orientation of the BPTI-variant relative to trypsin is given by the angle $\theta_o = \angle$T1-B1-B2 and the dihedrals $\phi_o = \angle$T2-T1-B1-B2 and $\psi_o = \angle$T1-B3-B1-B2.

In an initial attempt to achieve a free energy surface of the binding and unbinding process of (Abu,TfeGly)-BPTI from trypsin, we performed restrained Metadynamics simulations, where we chose the center-of-mass distance between (Abu, TfeGly)-BPTI and trypsin as the main collective variable and used harmonic restraints to restrain the other collective variables to the values of the fully bound complex, which we extracted from the X-ray

crystal structure of the TfeGly-BPTI-trypsin complex (pdb code: 4Y11). Our efforts did not yield sufficient sampling of the binding and unbinding process, as after a single unbinding event, the ligand did not find back into the fully bound complex throughout 500 ns Metadynamics simulations, although their orientation and movement around the receptor was restrained (see fig. S1). We conclude, that the preferred binding and unbinding pathway has to be more complex than a simple movement on a straight line defined only by the center-of-mass distance, and likely contains intermediate states.

C. Random Acceleration Molecular Dynamics (RAMD)

To study the unbinding pathways of (Abu, MfeGly, DfeGly, TfeGly)-BPTI from trypsin, we performed Random Acceleration Molecular Dynamics (RAMD)[22,23,42] simulations. RAMD is an enhanced sampling method that applies an additional biasing force to the center-of-mass of a ligand in an otherwise unbiased MD simulation.[22]. If the unbinding process does not make progress despite the biasing force, the direction is of the this force is reoriented in a random direction at regular time intervals. The method was originally invented to discover unbinding pathways of buried protein ligands.[25].

For every of the four BPTI-variants, we generated three different starting structures and ran 10 simulations with a maximum length of 40 ns for each of these replicas. To achieve dissociation, we needed a high force constant of 3500 kJ/(mol·nm^2), which is about an order of magnitude higher than for protein-small molecule systems like benzamidine-trypsin.[22,23] This might be expected, as according to inhibition assays[21], our systems have a binding affinity of -37 kJ/mol to -41 kJ/mol, while benzamidine binds to trypsin with a binding affinity of -22 kJ/mol to -26 kJ/mol.[15] Moreover, as the complex is held in place by many hydrogen bonds, it is likely that some, if not most, of them have to be broken in a concerted way to achieve dissociation, which would result in a very steep free energy barrier, requiring a strong force to drive the system out of the bound state. Possibly, proteins in general need a higher force constant to be dissociated efficiently compared to small molecules. However, since this is - to the best of our knowledge - the first application of RAMD to a protein-protein complex, we do not have literature values for comparison.

For the TfeGly-BPTI-Trypsin complex, fig.3 shows the timeseries of the center-of-mass distance (r) as moving average with a moving window of 200 ps. The panels correspond to the three different starting structures, and we show the timeseries of the 10 simulations per starting structure in different colors. See SI (figs. S2-S5) for the corresponding timeseries of DfeGly-, MfeGly- and Abu-BPTI. The timeseries in fig. 3 first fluctuate around the center-of-mass distance of the fully bound complex at around 2.65 nm. Then they tend to transition to a state in which the center-of-mass distance fluctuates between 2.75 nm and 3.00 nm. The systems tends to remain in this state for tens of nanoseconds until they dissociate very rapidly. We call this intermediate state of the protein-protein complex the pre-bound state. We distinguish it from the fully bound state at 2.65 nm center-of-mass distance and from encounter states, which were investigated by Kahler *et al.*[19] and which would lie at center-of-mass distances around 3.00 nm[21].

The pre-bound state occurs in dissociation trajectories of all the four complexes of trypsin with (Abu, MfeGly, DfeGly, TfeGly)-BPTI. While in some of the simulations dissociation occurs without visiting the pre-bound state, we observe that in more than half of the trajectories for all BPTI variants, the moving average of the center-of-mass distance remains at least 1 ns between 2.75 nm and 3.00 nm, i.e. the pre-bound state is visited. Some trajectories did not dissociate after 40 ns of RAMD simulation, with some simulations ending in the pre-bound state and others ending in the fully bound state (see SI (tab. S1). The stability of the pre-bound state is remarkable, since throughout the RAMD simulations a strong biasing force designed to dissociate the protein-protein complex acts on the center of mass of the BPTI variant.

Once the system leaves the pre-bound state towards larger center-of-mass distances, the protein-protein complex rapidly dissociates. That is, we do not observe encounter complexes around or above 3.00 nm center of mass distance for any of the BPTI variants in our RAMD simulations. Encounter complexes are typically only weakly bound, and we assume that because of the strong biasing force encounter complexes rapidly dissociated in the RAMD simulations.

Inspecting the RAMD trajectories more closely, we find that the pre-bound state is not only characterized by an increase in center-of-mass distance r, but also by a significant shift of the system in the positional collective variables Θ_p and Φ_p compared to the fully bound state (see fig. S6). Fig. 4 shows that the positional variables Θ_p or Φ_p change along with the center-of-mass distance r when transitioning from the fully bound state to the pre-bound state. In the orientational variables, θ_o, ϕ_o, ψ_o, we do not find such a correlation, except for a slight shift in the θ_o angle (see. fig. S7).

The stability in presence of the biasing force and the systematic change in the positional variables indicate that the pre-bound state might be a chemically relevant state, which is stabilized by different interactions than the fully bound state and separated by a free-energy barrier from the fully bound state. As the fully bound state is held in place tightly by a number of hydrogen bond like interactions, it is likely that some of these interactions must be broken so the pre-bound state can be reached. In the fully bound state of the complex between Trypsin and (Abu, MfeGly, DfeGly, TfeGly)-BPTI there are twelve hydrogen bond like contacts, that can be found by visually inspecting the crystal structures and which are shown in fig. 1. We calculated the frequency with which these interactions are formed as a function of

the center-of-mass distance and present the histograms in fig. 5. The criterion for an interaction being in place was a heavy atom distance of less than 0.35 nm. Note that the histograms were generated from RAMD simulations, i.e. non-equilibrium simulations, and therefore do not represent equilibrium distributions.

One of the twelve interactions is only rarely populated in the fully bound state and not populated at all in the pre-bound state: the hydrogen bond between the side chain hydroxyl oxygens of Serine 197 (Ser197-OG) in trypsin and the amide hydrogen in the backbone of the P1 residue in the BPTI variants (X15-N), shown in orange in fig. 5.a. But since the overall change is small, this interaction is not suited to further define the pre-bound state. Also in fig. 5.a, we show the histograms of eight further interactions, which are present in the fully bound state as well as in the pre-bound state. Their population decreases with increasing center-of-mass distance, but since the change is gradual and there is still significant population in the interval 2.75 nm $< r <$ 3.00 nm, it is not plausible that this change in population constitutes a clear free-energy barrier between the fully bound state and the pre-bound state.

Fig. 5.b shows the histogram of three interactions which are highly populated in the fully bound state but rarely populated in the interval 2.75 nm $< r <$ 3.00 nm. These are the backbone-backbone interaction between Phe41 and Arg17, the interaction of the backbone of His40 with the side chain of Arg17 and the interaction between the side chain of Tyr39-OH and the backbone of Ile19. (compare fig.1c). The breaking of these three interactions likely contributes to the free energy barrier between the fully bound state and the pre-bound state.

D. Unbiased Simulation of the Pre-Bound and Fully Bound state

To further characterize the difference between the fully bound state and the pre-bound state, we ran 20 unbiased simulations of 50 ns each (i.e. 1 μs total simulation time) of the fully bound state and of the pre-bound state in all of the four complexes between trypsin and (Abu, MfeGly, DfeGly, TfeGly)-BPTI. The starting structures for the fully bound state were generated from the crystal structure and the starting structures for the simulations of the pre-bound state were generated from snapshots of the system in the pre-bound state from the RAMD simulations.

The timeseries of the center-of-mass distance r for all of the unbiased MD simulations can be found in the SI (fig. S8). With very few exceptions, the systems remained ins their starting state throughout the whole simulation time. This indicates that both the fully bound state and the pre-bound state are stable on the timescale of 50 ns.

Fig. 6 compares the equilibrium distributions of the positional variables, r, Θ_p, Φ_p of fully bound state and of the pre-bound state. All four BPTI variants have similar distributions (different colors in fig. 6). However, the distribution differ significantly between the fully bound state and the pre-bound state (solid vs. dashed lines in in fig. 6). In the fully bound state, the systems adopt an average center of mass distance of 2.65 nm with a standard deviation of 0.04 nm, while in the pre-bound state the center of mass distance r amounts to a mean of 2.85 nm with a standard deviation of 0.05 nm. Likewise, the coordinates Θ_p and Φ_p shift to larger values in the pre-bound state. In all three positional coordinates, there is little overlap between the distributions of the fully bound state and the pre-bound state, confirming that the positions that the BPTI variants can occupy in these two states are distinct.

The distributions of the orientational variabiles, θ_o, ϕ_o, ψ_o, are included in the SI (fig. S.9). In each of the three variables we observe a systematic shift from the distributions of the fully bound state and to those of the pre-bound state, which is most pronounced for θ_o. However, the overlap between the fully-bound-state distributions and the pre-bound-state distributions is larger than for the positional variables. This indicates that the BPTI variant does not gain (much) orientational freedom when transitioning from the fully bound state to the pre-bound state.

Fig. 7 shows the relative population of the all hydrogen bonds between the two proteins with at least 0.1 relative population. For this analysis we merged the trajectories of the fully bound state of all four systems, and we merged the trajectories of the pre-bound state of all four systems. At this point this is justified, because the twelve interactions do not involve the side chain of the P1 residue and because we did not observe any significant difference in the positional and orientational variables across the four systems. We employed the Wernet-Nilsson criterion[39] in the MDtraj implementation to identify hydrogen bonds between Trypsin and the BPTI variants.

In the simulations of the fully bound state, we observe eleven hydrogen bonds with a relative population > 0.1. These are five hydrogen bonds, which are located around the S1 pocket, three hydrogen bonds of Arg39(compare fig.1b) and lastly the hydrogen bonds of Arg17 and Ile19 (compare fig.1c). We find the hydrogen bonds around the S1 pocket to also be present in the pre-bound state. The backbone-backbone interaction between Gly195 and the P1 residue has the same frequency in the pre-bound state as in the fully bound state, while the frequency of the neighboring interaction between the side chain of Gln194 and the backbone of Ala16 is lower in the pre-bound state, albeit with high statistical uncertainty. The frequency of two hydrogen bonds close to the S1 pocket, namely between the side chain of Gln194 and the backbone of Cys14 and the backbone-backbone interaction between Gly214 and Pro13 is higher in the pre-bound state, but again with high statistical uncertainty.

Three hydrogen bonds are frequently populated in the fully bound state but are virtually non-existent in the

pre-bound state, making these three broken hydrogen bonds a defining property of the pre-bound state. These are the same three hydrogen bonds that already showed a loss of population when transition from the fully bound state to the pre-bound state in the RAMD simulations (fig. 5.b). In the fully bound state, two of the hydrogen bonds are formed between between Arg17 in the BPTI variants and the backbone in trypsin, one between the side chain of Arg17 and the backbone of His40 and the other between the backbone of Arg17 and the backbone of Phe41. The third hydrogen bond is formed between the amide hydrogen of Ile19 in BPTI and the side chain of Tyr39 in trypsin. These hydrogen bonds are shown for the fully bound state in fig. 9.a. Fig. 9.b shows the same region in the pre-bound state. Side chains of Arg17 and Tyr39 have been reoriented and the three hydrogen bonds cannot be formed in the pre-bound state.

The hydrogen bonds of Arg39 in the BPTI variants with the backbone of trypsin are also more frequently populated in the fully bound state than in the pre-bound state (fig. 7). However, the drop in population is less pronounced than for the three hydrogen bonds discussed above.

The analysis so far shows that the dissocation of the protein-protein complex between Trypsin and (Abu, MfeGly, DfeGly, TfeGly)-BPTI proceeds via a pre-bound state which is stable at least on the timescale of 50 ns. The pre-bound state is characterized by a shift in the positional variables of BPTI and, to a lesser extent, by a shift in the orientational variables. To form the pre-bound state, three hydrogen bonds that are highly populated in the fully bound state, are broken.

E. Stabilizing Interactions in the Pre-Bound State

The analysis so far does not show why the breaking of the three hyrogen bonds results in a stable state that does not immediately revert back to the fully bound state. Fig. 7.b suggests that one of the factors contributing to the stability of the pre-bound state could be a cation-pi interaction that is formed by the now free Arg17 side chain of BPTI with the aromatic system of Tyr151 of trypsin.

We measured the distance distribution between the carbon atom of the guanidine moiety of Arg17 (Arg17-CZ) and the centroid of the aromatic ring of Tyr151 (fig.8a). While in the fully bound state, the distance can take a range of values between 0.3 nm and 0.8 nm, the distance in all simulation snapshots of the pre-bound state remains well below 0.45 nm. The broad distribution of the Tyr151-s-Arg17-CZ distance in the fully bound state shows that no specific bond is in place between the two residues. By contrast, the narrow distribution at low distances in the pre-bound state suggests the existence of a cation-pi interaction.

An interaction with the aromatic system of Tyr151 has so far not been described for the BPTI-trypsin complex. It is however present in X-ray crystallography structures of other trypsin inhibitors like Bdellastasin (pdb code: 1C9T), where a cationic lysine side chain at P2' position forms a cation-pi interaction with Tyr151[43] or Microviridin (pdb code: 4KTU) where a tyrosine at P2' position forms a t-shaped pi-pi interaction with Tyr151.[44]

To verify whether an interaction of the Arg17 side chain is indeed essential for the stabilization of the pre-bound state, we repeated the RAMD simulations for one replica of the Abu-BPTI-trypsin complex, where we mutated Arg17 in Abu-BPTI to alanine (fig.8b). The dissociation happens roughly on the same timescale as for the non-mutated Abu-BPTI variants. However, the pre-bound state is traversed rapidly on all ten of the unbinding trajectories. This supports the hypothesis that Arg17 is indeed essential for the stabilization of the pre-bound state.

F. The influence of the fluorine substituents

As a last step, we were interested in how the fluorine substituents in the BPTI variant influence the stability of the pre-bound state. To this end, we performed Umbrella Sampling between the fully bound state and the pre-bound state, where we used the newly identified interaction between the backbone amide of Arg17 in BPTI and the backbone oxygen of Phe41 in trypsin. Fig. 10a shows the resulting potential of mean force along this reaction coordinate. In all four systems, the potential of mean force exhibits two minima. The minimum around the a 0.3 nm corresponds to the fully bound state, whereas the minimum around 0.6 nm corresponds to the pre-bound state. In a previous study,[21] we investigated the interactions in the fully bound state and found no significant differences between the four BPTI variants. For the pre-bound state, we find that the barrier height between the two states for the unfluorinated Abu and the mono-fluorinated MfeGly is about 15 kJ/mol, while for the higher fluorinated DfeGly and TfeGly it is only about 10 kJ/mol. The minimum of the pre-bound state for Abu lies well above the minimum for the fully bound state. By constrast, in the TfeGly-BPTI complex, the pre-bound state is stabilized relative to the bound state. The partially fluorinated complexes lie in between. Thus, there is a clear effect of the fluorination on the energetic landscape between the fully bound state and the pre-bound state.

To find a possible mechanism for this stabilization, we revisited our hydrogen bond analysis for which we reported the aggregate statistics for all four BPTI variants in fig. 7 We re-analyzed for each BPTI-variant and found that for most interactions, the hydrogen bond populations did not differ significantly across the BPTI variants. A notable exception is the hydrogen bond between the side chain of Gln194 in trypsin and the backbone oxygen of Cys14 in the BPTI variants. This interaction can be observed in the fully bound state and also in

the pre-bound state, but it is more frequent in the pre-bound state of the fluorinated variants (MfeGly, DfeGly, TfeGly)-BPTI, while it is equally populated in the states of Abu-BPTI (see fig. S10). Gln194 and Cys14 are close to the side chain of the P1 residue (see fig.10b). When the hydrogen bond is formed, the side chain of Gln194 is in fact so close to the fluorine atoms that it appears plausible that the fluorine atoms with their negative partial charge help stabilize the NH2-end of the Gln194 side chain by providing an extra binding partner in addition to the backbone oxygen of Cys14.

IV. CONCLUSIONS

We applied several molecular simulation techniques to characterize the unbinding pathway of (Abu, MfeGly, DfeGly, TfeGly)-BPTI from trypsin. The BPTI variants likely dissociate via a curved pathway in a coordinate space that describes the relative position and orientation of the two proteins, as evidenced by restrained Metadynamics simulations in which the two protein do not rebind once they are dissociated. Using RAMD simulations[22–25] to accommodate this curved unbinding pathway, we identified a new metastable state on the unbinding pathway.

This pre-bound state is present on the unbinding pathway in all four variants of the BPTI-trypsin complex. In unbiased simulations it is stable for at least 50 ns. Since in an aggregated simulation time of 1 μs per BPTI variant, on pre-bound state only very rarely reverted to the fully bound state, we suspect that the average lifetime of the pre-bound state is in fact in the order of several 100 ns.

The pre-bound state is clearly distinct from the fully bound state in the positional coordinates from the fully bound state. The center-of-mass distance between the two proteins is increased by about 0.2 nm (from 2.65 to 2.85 nm) and the BPTI variants rotate by about 10° (0.2 rad) in Θ_p and by 10° (0.2 rad) around the dihedral angle Φ_p. There is little overlap between the distributions of the pre-bound state and fully bound state in these coordinates. We also observe a systematic shift in the orientational coordinates, but less pronounced.

The interaction pattern between the two proteins changes when transitioning from the fully bound state to the pre-bound state. These changes particularly involve Arg17 (P2' residue) and Arg39 in the BPTI variants. In the pre-bound state, the hydrogen bond of the Arg17 side chain to the backbone of trypsin is broken, but it is replaced by a cation-pi interaction between the guanidine moiety and a nearby trypsin tyrosine residue. Two further hydrogen bonds in the vicinity are also broken in this process, and the hydrogen bond between the side chain of Arg39 and the trypsin backbone becomes less populated. When we replaced Arg17 by an alanine residue in RAMD simulations, the protein-protein complex dissociated without spending time in the pre-bound state, which demonstrates that Arg17 is essential for the stabilization of this state. The structural rearrangements that stabilize the pre-bound state do not involve the P1 residue in BPTI or the negatively charged residues at the bottom of the S1 pocket of trypsin. It is thus likely that the unbinding of wildtype BPTI from trypsin proceeds via a similar pre-bound state.

In potentials of mean force (PMF), we find that fluorination of Abu lowers the free energy barrier between the fully bound and the pre-bound state and also lowers the free energy minimum of the pre-bound state. However, a quantitative interpretation of these one-dimensional PMFs is difficult. In particular, we suspect that the PMF might overstabilize the pre-bound state, as in some of the potentials the pre-bound state minimum is as low as the fully bound minimum. Nonetheless, the fluorine substituents on the P1 residue clearly have an influence on the stability of the pre-bound state. A possible, yet speculative, explanation is that the hydrogen bond between the side chain of Gln194 and Cys14 is stabilized by fluorine substituents in direct proximity of the side chain NH2 group of Gln194. Fluorine is known to have a wide range of possible effects on protein-inhibitor interactions, e.g. through hydrogen bonds[45], desolvation[32,33] or entropy[46], whose elucidation often requires in-depth computational studies. The differences in barrier height and stability of the pre-bound state in the fluorinated variants of BPTI are likely not only due to a single stabilizing interaction, like the Gln194-Cy14 hydrogen bond, but instead due to a combination of enthalpic and entropic effects.

Because of the large magnitude of the biasing force in the RAMD simulations, which is necessary to dissociate the protein-protein complexes, we did not observe encounter complexes in our simulations. We expect that encounter states do play a role in the binding and unbinding process of the (Abu, MfeGly, DfeGly, TfeGly)-BPTI-trypsin complexes[19]. However, the transition between the pre-bound state and these weakly bound encounter states should be characterized with other methods like Weighted Ensemble MD[47] and Molecular Rotational Grids[48].

The existence and structure of the pre-bound state invites speculation on the inhibitory mechanism of BPTI and its variants. After forming the initial Michaelis complex of a substrate with trypsin, the hydrolysis of the peptide bond proceeds via two steps. First the peptide bond is broken and the N-terminal part of the substrate (i.e., all residues from the N-terminus up to and including P1) forms a covalently bound acyl-enzyme intermediate. The C-terminal part of the substrate (i.e., all residues from P1' to the C-terminus) remains non-covalently bound and needs to dissociate before, in a second step, the acyl-enzyme intermediate can be hydrolyzed. Radisky and Koshland showed that, for a closely related serine protease complex, the initial formation of the acyl-enzyme intermediate is fast, but the release of the C-terminal part of the substrate is slow[9], such that the reaction re-

verts back to the intact peptide bond. This "clogged gutter" mechanism is further supported by a high-resolution structure of a cleaved BPTI variant with trypsin[13]. Our analysis showed that the interface between trypsin and the BPTI variants is stabilized by hydrogen bonds primarily from the C-terminal part of the BPTI variants (fig.7). Specifically, Arg17 which stabilizes the pre-bound state via a cation-pi interaction belongs to the C-terminal part. Thus, assuming that the clogged gutter mechanism applies to the BPTI-trypsin complex, these interactions likely contribute to stabilizing the C-terminal part within the protein complex.

Finally, our study shows that, to understand the stability of the (Abu, MfeGly, DfeGly, TfeGly)-BPTI-trypsin complex, one needs to consider two states, the fully bound state and the pre-bound state, which likely are in dynamic equilibrium. By mimicking the interactions in the pre-bound state one may open up additional ways to design serine-protease inhibitors.

V. DATA AND SOFTWARE AVAILABILITY

RAMD and unbiased MD simulations were performed with openly available Gromacs (https://www.gromacs.org), Gromacs-RAMD (https://github.com/HITS-MCM/gromacs-ramd/tree/release-2022) and Plumed (https://www.plumed.org) software. Starting structure preparation was done with the openly available pdbfixer (https://github.com/openmm/pdbfixer). Interaction distances and hydrogen bonds were analyzed using mdtraj (https://www.mdtraj.org), which is openly available. Force field parameters for the fluorinated amino acids and Abu have been released in a recent publication[21], all other force field parameters were taken from the openly available Amber14SB force field (https://ambermd.org). Protein structure starting files, MD parameter files and analysis scripts for all simulations are published on Github (https://github.com/leonwehrhan/Trypsin_BPTI_Prebound_RAMD_2024).

VI. ACKNOWLEDGEMENTS

We thank Ariane Nunes-Alves for fruitful discussions on RAMD simulations. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation): Project-ID 387284271-SFB 1349. The authors acknowledge access to high-performance computers via the Zentraleinrichtung FUB-IT of Freie Universität Berlin.

VII. BIBLIOGRAPHY

[1] B. Turk, "Targeting proteases: Successes, failures and future prospects," Nature Reviews Drug Discovery **5**, 785–799 (2006).

[2] M. Drag and G. S. Salvesen, "Emerging principles in protease-based drug discovery," Nature Reviews Drug Discovery **9**, 690–701 (2010).

[3] A. R. Batt, C. P. St. Germain, T. Gokey, A. B. Guliaev, and T. Baird Jr., "Engineering trypsin for inhibitor resistance," Protein Science **24**, 1463–1474 (2015).

[4] M. J. Page and E. D. Cera, "Evolution of Peptidase Diversity," Journal of Biological Chemistry **283**, 30010–30014 (2008), 18768474.

[5] I. Schechter, "Mapping of the Active Site of Proteases in the 1960s and Rational Design of Inhibitors/Drugs in the 1990s," Current Protein and Peptide Science **6**, 501–512 (2005).

[6] K. Kawamura, T. Yamada, K. Kurihara, T. Tamada, R. Kuroki, I. Tanaka, H. Takahashi, and N. Niimura, "X-ray and neutron protein crystallographic analysis of the trypsin–BPTI complex," Acta Crystallographica Section D: Biological Crystallography **67**, 140–148 (2011).

[7] M. Marquart, J. Walter, J. Deisenhofer, W. Bode, and R. Huber, "The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors," Acta Crystallographica Section B: Structural Science **39**, 480–490 (1983).

[8] M. Peräkylä and P. A. Kollman, "Why Does Trypsin Cleave BPTI so Slowly?" Journal of the American Chemical Society **122**, 3436–3444 (2000).

[9] E. S. Radisky and D. E. Koshland, "A clogged gutter mechanism for protease inhibitors," Proceedings of the National Academy of Sciences **99**, 10316–10321 (2002).

[10] J. P. Vincent and M. Lazdunski, "Trypsin-pancreatic trypsin inhibitor association. Dynamics of the interaction and role of disulfide bridges," Biochemistry **11**, 2967–2977 (1972).

[11] g.-i. family=Warshel, given=Arieh. and g.-i. family=Russell, given=Stephen., "Theoretical correlation of structure and energetics in the catalytic reaction of trypsin," Journal of the American Chemical Society **108**, 6569–6579 (1986).

[12] C. J. Farady and C. S. Craik, "Mechanisms of Macromolecular Protease Inhibitors," ChemBioChem **11**, 2341–2346 (2010).

[13] E. Zakharova, M. P. Horvath, and D. P. Goldenberg, "Structure of a serine protease poised to resynthesize a peptide bond," Proceedings of the National Academy of Sciences **106**, 11034–11039 (2009).

[14] M. J. M. Castro and S. Anderson, "Alanine Point-Mutations in the Reactive Region of Bovine Pancreatic Trypsin Inhibitor: Effects on the Kinetics and Thermodynamics of Binding to -Trypsin and -Chymotrypsin," Biochemistry **35**, 11435–11446 (1996).

[15] I. Buch, T. Giorgino, and G. De Fabritiis, "Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations," Proceedings of the National Academy of Sciences of the United States of America **108**, 10184–10189 (2011), 21646537.

[16] T. Siebenmorgen and M. Zacharias, "Computational prediction of protein–protein binding affinities," WIREs Computational Molecular Science **10**, e1448 (2020).

[17] Z. Wu, Q. Liao, and B. Liu, "A comprehensive review and evaluation of computational methods for identifying protein complexes from protein–protein interaction networks," Briefings in Bioinformatics **21**, 1531–1548 (2020).

[18] H.-J. Woo and B. Roux, "Calculation of absolute protein–ligand binding free energy from computer simulations," Proceedings of the National Academy of Sciences **102**, 6825–6830 (2005).

[19] U. Kahler, A. S. Kamenik, F. Waibl, J. Kraml, and K. R. Liedl, "Protein-Protein Binding as a Two-Step Mechanism: Preselection of Encounter Poses during the Binding of BPTI and Trypsin," Biophysical Journal **119**, 652–666 (2020).

[20]S. Ye, B. Loll, A. Ann Berger, U. Mülow, C. Alings, M. Christian Wahl, and B. Koksch, "Fluorine teams up with water to restore inhibitor activity to mutant BPTI," Chemical Science **6**, 5246–5254 (2015).

[21]L. Wehrhan, J. Leppkes, N. Dimos, B. Loll, B. Koksch, and B. G. Keller, "Water Network in the Binding Pocket of Fluorinated BPTI–Trypsin ComplexesInsights from Simulation and Experiment," The Journal of Physical Chemistry B (2022).

[22]D. B. Kokh, B. Doser, S. Richter, F. Ormersbach, X. Cheng, and R. C. Wade, "A workflow for exploring ligand dissociation from a macromolecule: Efficient random acceleration molecular dynamics simulation and interaction fingerprint analysis of ligand trajectories," The Journal of Chemical Physics **153**, 125102 (2020).

[23]D. B. Kokh, M. Amaral, J. Bomke, U. Grädler, D. Musil, H.-P. Buchstaller, M. K. Dreyer, M. Frech, M. Lowinski, F. Vallee, M. Bianciotto, A. Rak, and R. C. Wade, "Estimation of Drug-Target Residence Times by -Random Acceleration Molecular Dynamics Simulations," Journal of Chemical Theory and Computation **14**, 3859–3869 (2018).

[24]A. Nunes-Alves, D. B. Kokh, and R. C. Wade, "Recent progress in molecular simulation methods for drug binding kinetics," Current Opinion in Structural Biology **64**, 126–133 (2020-10).

[25]S. K. Lüdemann, V. Lounnas, and R. C. Wade, "How do substrates enter and products exit the buried active site of cytochrome P450cam? 1. Random expulsion molecular dynamics investigation of ligand access channels and mechanisms11Edited by J. Thornton," Journal of Molecular Biology **303**, 797–811 (2000).

[26]M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, R. A. Broglia, and M. Parrinello, "PLUMED: A portable plugin for free-energy calculations with molecular dynamics," Computer Physics Communications **180**, 1961–1972 (2009).

[27]G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, and G. Bussi, "PLUMED 2: New feathers for an old bird," Computer Physics Communications **185**, 604–613 (2014).

[28]M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers," SoftwareX **1–2**, 19–25 (2015).

[29]S. Páll, M. J. Abraham, C. Kutzner, B. Hess, and E. Lindahl, "Tackling Exascale Software Challenges in Molecular Dynamics Simulations with GROMACS," in *Solving Software Challenges for Exascale*, Lecture Notes in Computer Science, edited by S. Markidis and E. Laure (Springer International Publishing, 2015) pp. 3–27.

[30]S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, p. d. u. family=Spoel, given=David, B. Hess, and E. Lindahl, "GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit," Bioinformatics **29**, 845–854 (2013).

[31]J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling, "ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB," Journal of Chemical Theory and Computation **11**, 3696–3713 (2015).

[32]J. R. Robalo, S. Huhmann, B. Koksch, and A. Vila Verde, "The Multiple Origins of the Hydrophobicity of Fluorinated Apolar Amino Acids," Chem **3**, 881–897 (2017).

[33]J. Robalo and A. V. Verde, "Unexpected trends in the hydrophobicity of fluorinated amino acids reflect competing changes in polarity and conformation," Physical Chemistry Chemical Physics **21**, 2029–2038 (2019).

[34]G. Bussi, D. Donadio, and M. Parrinello, "Canonical sampling through velocity rescaling," The Journal of Chemical Physics **126**, 014101 (2007).

[35]M. Parrinello and A. Rahman, "Polymorphic transitions in single crystals: A new molecular dynamics method," Journal of Applied Physics **52**, 7182–7190 (1981).

[36]B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije, "LINCS: A linear constraint solver for molecular simulations," Journal of Computational Chemistry **18**, 1463–1472 (1997).

[37]W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," The Journal of Chemical Physics **79**, 926–935 (1983).

[38]R. T. McGibbon, K. A. Beauchamp, M. P. Harrigan, C. Klein, J. M. Swails, C. X. Hernández, C. R. Schwantes, L.-P. Wang, T. J. Lane, and V. S. Pande, "Mdtraj: A modern open library for the analysis of molecular dynamics trajectories," Biophysical Journal **109**, 1528 – 1532 (2015).

[39]g.-i. family=Wernet, given=Ph., D. Nordlund, U. Bergmann, M. Cavalleri, M. Odelius, H. Ogasawara, L. Näslund, T. K. Hirsch, L. Ojamäe, P. Glatzel, L. G. M. Pettersson, and A. Nilsson, "The Structure of the First Coordination Shell in Liquid Water," Science **304**, 995–999 (2004).

[40]G. Bussi and G. A. Tribello, "Analyzing and Biasing Simulations with PLUMED," in *Biomolecular Simulations: Methods and Protocols*, Methods in Molecular Biology, edited by M. Bonomi and C. Camilloni (Springer, 2019) pp. 529–578.

[41]Z. Tan, E. Gallicchio, M. Lapelosa, and R. M. Levy, "Theory of binless multi-state free energy estimation with applications to protein-ligand binding," The Journal of Chemical Physics **136**, 144102 (2012).

[42]S. K. Lüdemann, V. Lounnas, and R. C. Wade, "How do substrates enter and products exit the buried active site of cytochrome P450cam? 2. Steered molecular dynamics and adiabatic mapping of substrate pathways11Edited by J. Thornton," Journal of Molecular Biology **303**, 813–830 (2000).

[43]U. Rester, W. Bode, M. Moser, M. A. A. Parry, R. Huber, and E. Auerswald, "Structure of the complex of the antistasin-type inhibitor bdellastasin with trypsin and modelling of the bdellastasin-microplasmin system11Edited by I. A. Wilson," Journal of Molecular Biology **293**, 93–106 (1999).

[44]A. R. Weiz, K. Ishida, F. Quitterer, S. Meyer, J.-C. Kehr, K. M. Müller, M. Groll, C. Hertweck, and E. Dittmann, "Harnessing the Evolvability of Tricyclic Microviridins To Dissect Protease–Inhibitor Interactions," Angewandte Chemie International Edition **53**, 3735–3738 (2014).

[45]W. Pietruś, R. Kafel, A. J. Bojarski, and R. Kurczab, "Hydrogen Bonds with Fluorine in Ligand–Protein Complexes-the PDB Analysis and Energy Calculations," Molecules **27**, 1005 (2022).

[46]J. Wallerstein, V. Ekberg, M. M. Ignjatović, R. Kumar, O. Caldararu, K. Peterson, S. Wernersson, U. Brath, H. Leffler, E. Oksanen, D. T. Logan, U. J. Nilsson, U. Ryde, and M. Akke, "Entropy–Entropy Compensation between the Protein, Ligand, and Solvent Degrees of Freedom Fine-Tunes Affinity in Ligand Binding to Galectin-3C," JACS Au **1**, 484–500 (2021).

[47]D. M. Zuckerman and L. T. Chong, "Weighted Ensemble Simulation: Review of Methodology, Applications, and Software," Annual Review of Biophysics **46**, 43–57 (2017).

[48]H. Zupan, F. Heinz, and B. G. Keller, "Grid-based state space exploration for molecular binding," Canadian Journal of Chemistry **101**, 710–724 (2023).

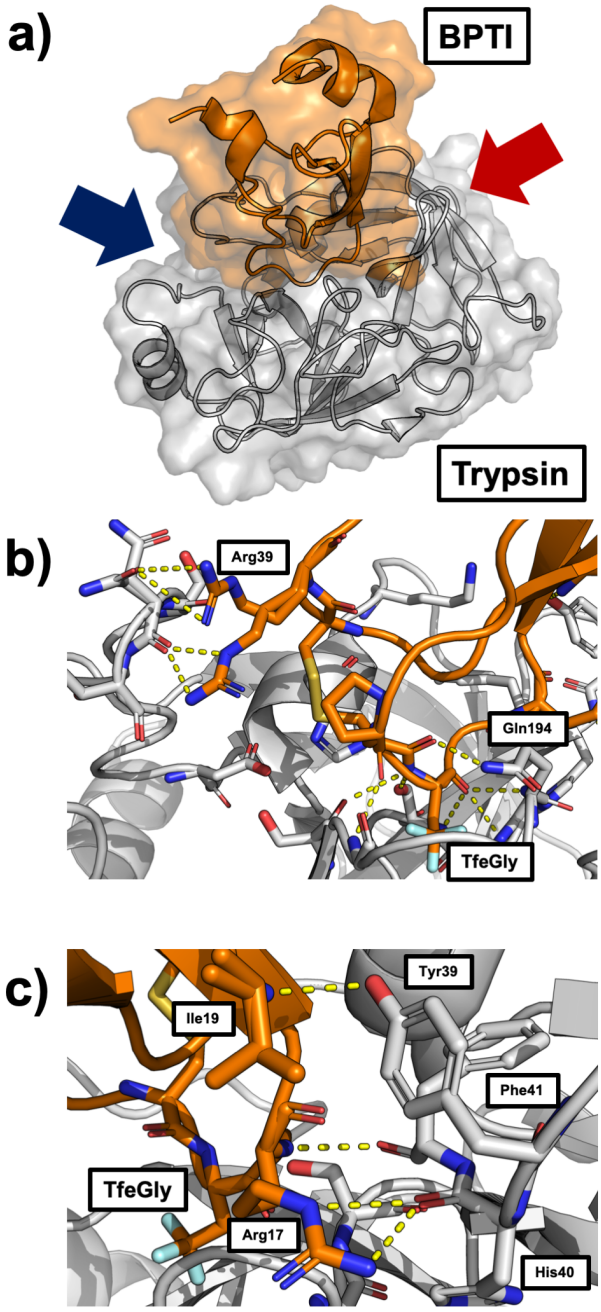

FIG. 1. (a) TfeGly-BPTI-trypsin complex with surface and cartoon representation (pdb code: 4Y11). (b) Protein-protein interface of the complex seen from the perspective of the blue arrow. The interactions around the S1 pocket are to the bottom right and the interactions of Arg39 are on the top left. (c) Protein-protein interface of the complex seen from the red arrow. The interactions of Arg17 (P2') and Ile19 (P4') are shown.

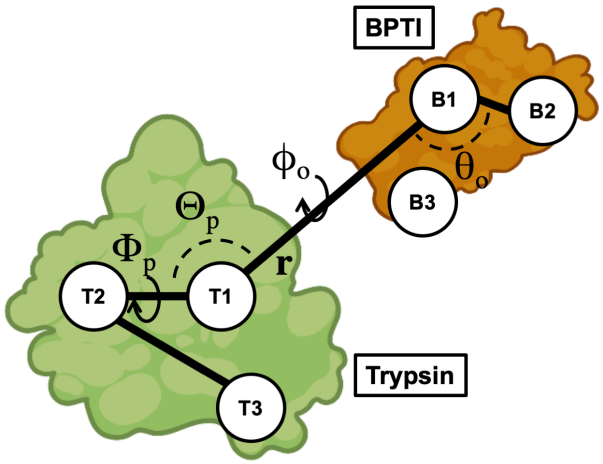

FIG. 2. Method to construct collective variables that describe position and orientation of the BPTI variants with respect to trypsin. The position relative to trypsin is described by Θ_p (T2-T1-B1) and Φ_p (T3-T2-T1-B1). The orientation is described by θ_o (T1-B1-B2), ϕ_o (T2-T1-B1-B2) and ψ_o (T1-B3-B1-B2). **r** (T1-B1) is the center-of-mass distance. Compare ref 18.

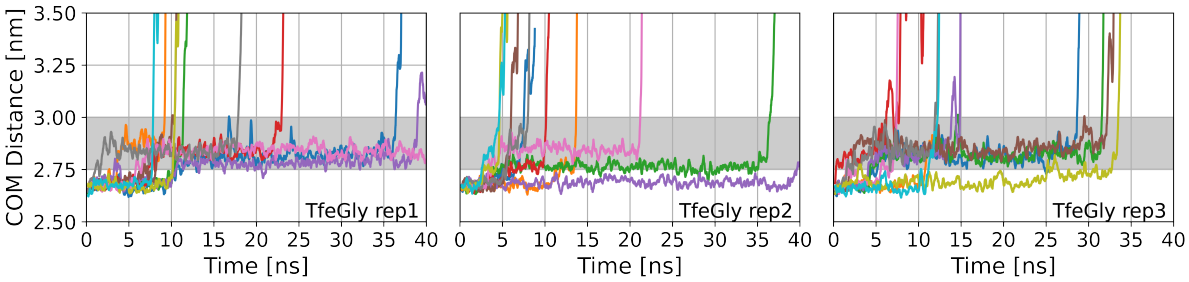

FIG. 3. RAMD dissociation timeseries of TfeGly-BPTI. COM = center-of-mass. Grey area shows the center of mass of the pre-bound state. Left panel: replica 1, middle panel: replica 2, right panel:replica 3. 10 RAMD runs per replica.

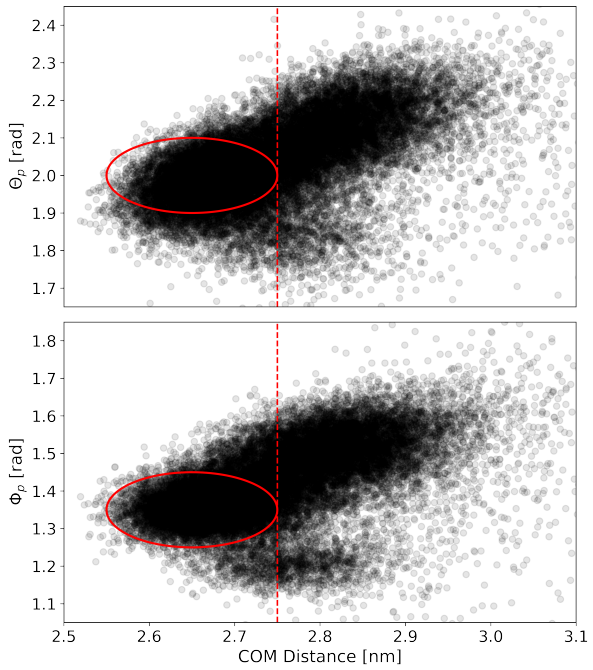

FIG. 4. Scatter plots of the center of mass distance and Θ_p (top) and Φ_p (bottom) for all combined RAMD simulations. The red circle marks the position of the fully bound state.

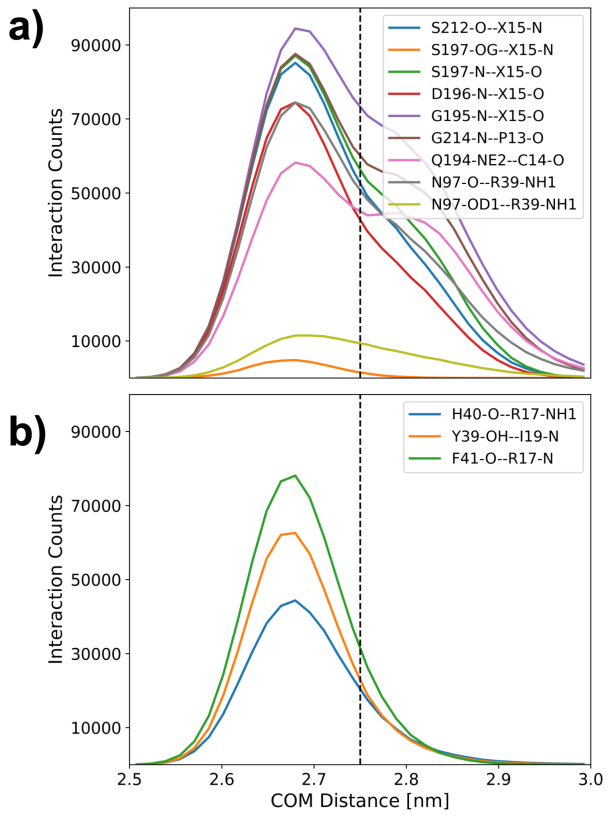

FIG. 5. Interaction histogram along center of mass distance in RAMD simulations. The criterion for an interaction to be in place was that the involved heavy atoms were separated by a distance less than 0.35 nm. The histograms were generated from the biased (non-equilibrium) RAMD simulations and therefore do not represent a Boltzmann distribution.

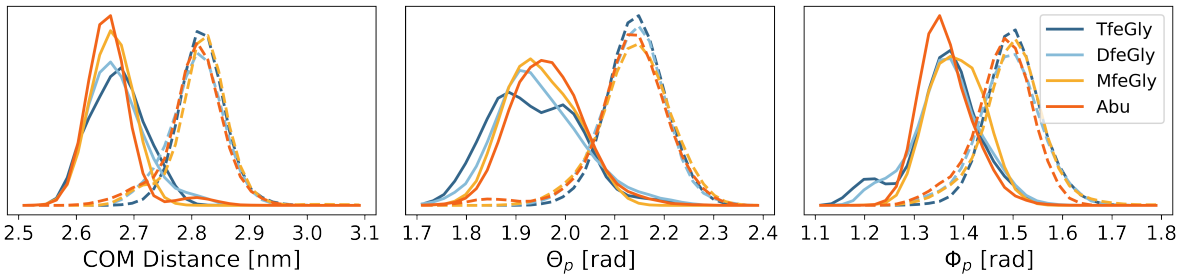

FIG. 6. Position of the BPTI variants in the unbiased simulations of the fully bound state (solid lines) and the pre-bound state (dashed lines) as described by the center of mass distance r (left), Θ_p (center) and Φ_p (right).

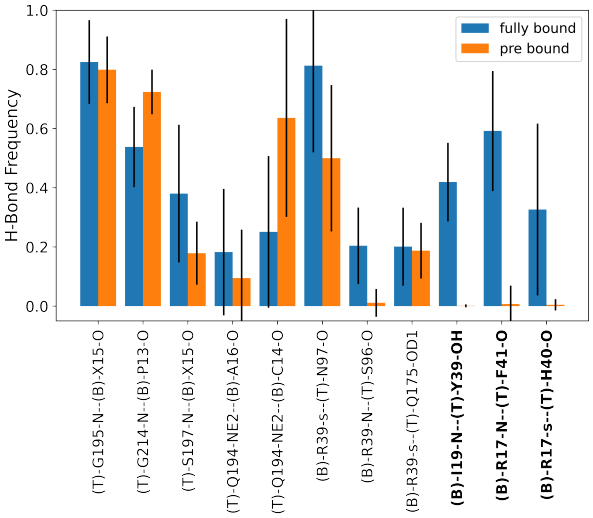

FIG. 7. Hydrogen bond frequencies of all combined unbiased simulations of the fully bound state and the pre-bound state. Residue name X = Abu, MfeGly, DfeGly or TfeGly, T = Trypsin, B = (Abu, MfeGly, DfeGly, TfeGly)-BPTI. O, N = heteroatoms in the backbone, OD1, OH, NE2 = heteroatoms in side chains. Hydrogen bonds are denoted as Donor-Acceptor. The side chain of arginine residues is denoted as "s", which means a hydrogen bond with any of the donors in the guanidine moiety.

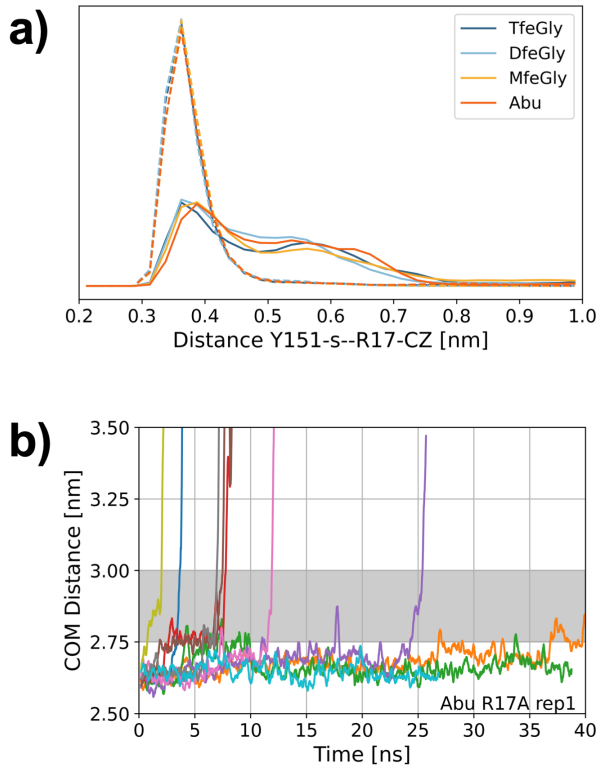

FIG. 8. (a) Distance between the centroid of the Tyr151 aromatic system (Y151-s) and the carbon of the guanidine moiety of Arg17 (R17-CZ) in the unbiased simulations of the fully bound state (solid line) and the pre-bound state (dashed lines). (b) RAMD dissociation trajectories of one replica of the R17A mutant of Abu-BPTI.

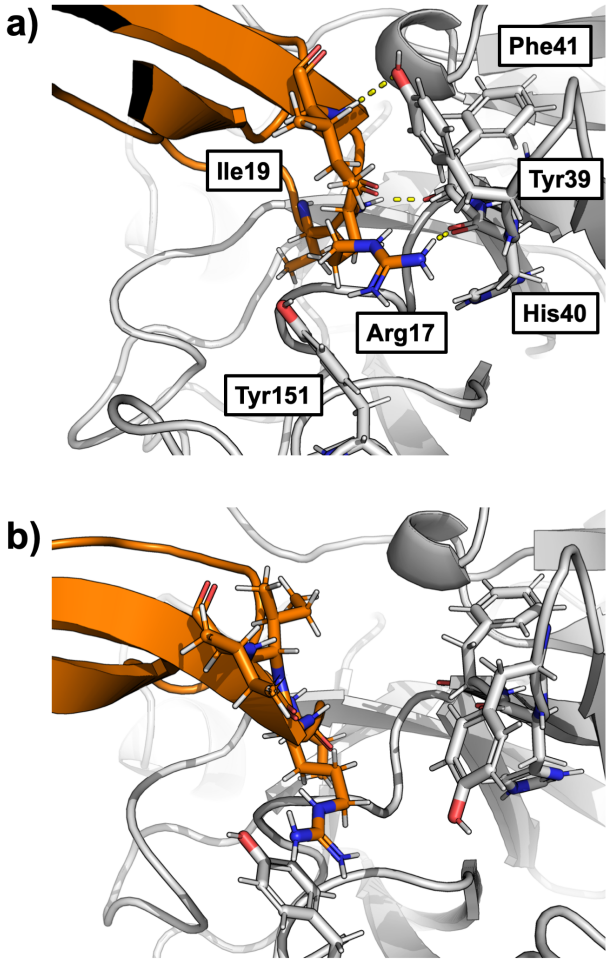

FIG. 9. Example snapshots from the unbiased simulations of the (a) fully bound and (b) pre-bound state. The figure shows a similar region as fig.1c.

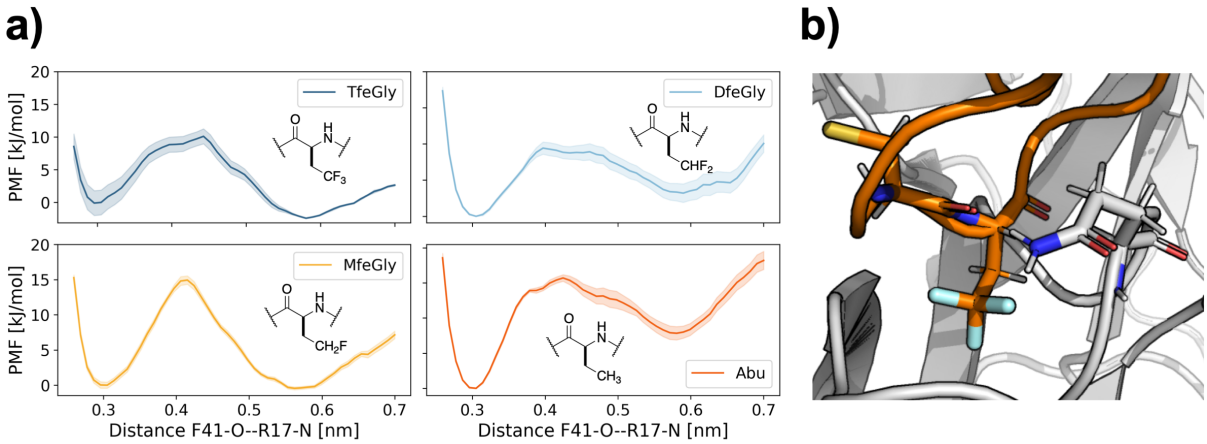

FIG. 10. (a) Potential of mean force profile of fully bound state and pre-bound state from Umbrella Sampling over the distance between the carbonyl oxygen of Phe41 (F41-O) and the backbone nitrogen of Arg17 (R17-N). (b) Interaction between Gln194 and Cys14 in the direct proximity of the TfeGly side chain.