Water Network in the Binding Pocket of Fluorinated BPTI-Trypsin Complexes - Insights from Simulation and Experiment

Leon Wehrhan,[1] Jakob Leppkes,[2] Nicole Dimos,[3] Bernhard Loll,[3] Beate Koksch,[2] and Bettina G. Keller*[1, a)]

[1)] *Department of Biology, Chemistry, and Pharmacy, Freie Universität Berlin, Institute of Chemistry and Biochemistry, Arnimallee 22, Berlin, 14195 Germany*

[2)] *Department of Biology, Chemistry, and Pharmacy, Freie Universität Berlin, Institute of Chemistry and Biochemistry, Arnimallee 20, Berlin, 14195 Germany*

[3)] *Department of Biology, Chemistry, and Pharmacy, Freie Universität Berlin, Institute of Chemistry and Biochemistry, Takustr. 6, Berlin, 14195 Germany*

Structural waters in the S1 binding pocket of β-trypsin are critical for the stabilization of the complex of β-trypsin with its inhibitor bovine pancreatic trypsin inhibitor (BPTI). The inhibitor strength of BPTI can be modulated by replacing the critical lysine residue at the P1 position by non-natural amino acids. We study BPTI variants in which the critical Lys15 in BPTI has been replaced by α-aminobutyric acid (Abu) and its fluorinated derivatives monofluoroethylglycine (MfeGly), difluoroethylglycine (DfeGly) and trifluoroethylglycine (TfeGly). We investigate the hypothesis that additional water molecules in the binding pocket can form specific non-covalent interactions to the fluorinated side chains and thereby act as an extension of the inhibitors. We report potentials of mean force (PMF) of the unbinding process for all four complexes and enzyme activity inhibition assays. Additionally, we report the protein crystal structure of the Lys15MfeGly-BPTI-β-trypsin complex (pdb: 7PH1). Both, experimental and computational data, show a step-wise increase in inhibitor strength with increasing fluorination of the Abu side chain. The PMF additionally shows a minimum for the encounter complex and an intermediate state just before the bound state. In the bound state, the computational analysis of the structure and dynamics of the water molecules in the S1 pocket shows a highly dynamic network of water molecules that does not indicate a rigidification or stabilizing trend in regards to energetic properties that could explain the increase in inhibitor strength. The analysis of the enthalpy and the entropy of the water molecules in the S1 binding pocket using Grid Inhomogeneous Solvation Theory confirms this result. Overall, fluorination systematically changes the binding affinity but the effect cannot be explained by a persistent water network in the binding pocket. Other effects, such as the hydrophobicity of fluorinated amino acids and the stability of the encounter complex as well as the additional minimum in the potential of mean force in the bound state, likely influence the affinity more directly.

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

I. INTRODUCTION

Hydrolysis of peptide bonds catalyzed by proteases is an ubiquitous process in all forms of life. This makes proteases important drug targets. However, drug design for these enzymes is difficult, because in proteases structural water molecules and hydration effects tend to play a critical role in ligand recognition and binding affinity[1–4].

β-trypsin is a serine protease that acts as digestive enzyme in the small intestine of mammals where it breaks down food proteins. It is itself a drug target, but it also serves as model system for the study of protease inhibitors. In serine proteases, the enzymatic hydrolysis of the peptide bond in the substrate proceeds via a catatytic serine residue[5–8], which is deprotonated by the combined action of an adjacent histidin and an aspartic acid. Deprotonation of the serine residue turns it into a strong nucleophile that can attack the carboxylic carbon atom in the peptide bond. The three critical residues serine, histidine and aspartic acid are called catalytic triad and are highly conserved in serine proteases[9]. In all serine proteases, the catalytic triad is located at the rim of a deep binding site called S1 binding pocket. Substrate specificity is realized by the amino acids within this binding pocket[9]. β-trypsin contains a negatively charged aspartic acid at the bottom of the S1 binding site. It thus recognizes positively charged amino acids, such as lysine and arginine[9], and hydrolyses the peptide bond at their C-terminal side[9].

β-trypsin is inhibited by multiple small molecules[10,11], most notably benzamidine[12–14]. Additionally, several proteins have evolved as natural inhibitors of serine proteases in general, and β-trypsin in particular. Among them, Bovine pancreatic trypsin inhibitor (BPTI) is likely the most studied protein trypsin inhibitor[5–8,15–19]. BPTI also inhibits other serine proteases, and has been used under the drug name "aprotinin" to prevent excessive blood loss during surgery by acting on serine proteases involved in blood coagulation[20].

BPTI is a small protein with only 58 residues and a molecular weight of 6.5 kDa[20]. It acts as a competitive inhibitor by inserting a lysine residue (P1 residue) into the S1 binding pocket and thus blocking the active site of β-trypsin. Interestingly, the lysine side-chain is too short to form a direct salt-bridge with the aspartic acid at the bottom of the S1 binding site. Instead, the binding is mediated by three water molecules that are trapped in the binding pocket[16,18]. One of the water molecules is located between the positively charged amino group of lysine and the negatively charged carboxyl group of aspartate. It accepts a

hydrogen bond from the amino group and donates a hydrogen bond to the carboxyl group, and, via this hydrogen-bond network, bridges the gap between lysine in BPTI and aspartate in β-trypsin.

Despite the water-mediated binding mode, wild-type (wt) BPTI-β-trypsin complex is one of the strongest protein-protein complexes found in nature, with a binding constant[15,17] of 5 to $6 \cdot 10^{-14}$ M and a corresponding binding free energy at $T = 300$ K of $\Delta G_{\text{bind}} = -76$ kJ/mol. But BPTI is not very specific, and also binds to several other proteases[15,17]. By mutating the P1 residue in BPTI, i.e. the lysine that inserts into the binding pocket of β-trypsin, one can tune the binding free energy between -76 kJ/mol and -20 kJ/mol[21–23]. However, the BPTI variants remain promiscuous.

One way to affect the specificity of a protein inhibitor is by introducing non-natural amino acids[24]. In small molecule inhibitors the use of fluorine substituents has become a valuable tool to tune affinity and selectivity[25–28]. In Ref. 29, some of us introduced fluorinated non-natural amino acids at the P1 position in BPTI and investigated the effect on the inhibitor strength and the structure of the complex. Substituting lysine by the shorter and aliphatic residue α-aminobutyric acid (Abu) drastically reduced the inhibitor strength. This is expected, because the short side chain of Abu only reaches to about mid-way in the S1 binding pocket, and because it is an aliphatic side-chain, it cannot be stabilized by hydrogen bonds or charge interactions in the binding pocket. However, when two fluorine substituents were added to the terminal methyl group of Abu (Lys15DfeGly-BPTI), the inhibitor strength increased. It increased even further when the methyl group of Abu was fully fluorinated (Lys15TfeGly-BPTI).

The crystal structures of Lys15Abu-BPTI-β-trypsin complex showed five trapped water molecules in the S1 binding pocket. Three water molecules occupied the same positions as in the wt-BPTI-β-trypsin complex, the other two water molecules were found in the space that had been generated by truncating the P1 residue. The crystal structure of the two fluorinated complexes also showed five water molecules at the same positions, however with much lower B-factors, which indicates reduced mobility. This suggested the following mechanism for the increased affinity of the fluorinated BPTI-variants: The fluorine substituents form specific non-covalent interactions with the two additional water molecules, which are not possible with purely aliphatic side-chains. The two water molecules can thus act as an extension of the inhibitor and mediate binding to β-trypsin. In short: the water-mediated binding in the

wt-BPTI-β-trypsin complex might be extended due to the presence of fluorine substituents[29].

Crystal structures only provide static pictures of the complexes. On the other hand, molecular dynamics (MD) simulations can sample the structural fluctuations of the proteins as well as the dynamics of the water molecules in the binding pocket. Here, we use a combination of MD simulations, inhibition assays, and X-ray crystallography to elucidate the interaction of the fluorinated BPTI with trypsin and the water molecules in the binding pocket.

II. RESULTS

A. Parametrization of the fluorinated variants of aminobutyric acid (Abu)

The following fluorinated variants of aminobutyric acid (Abu) are included in this study: monofluoroethylglycine (MfeGly), difluoroethylglycine (DfeGly) and trifluoroethylglycine (TfeGly) (see Fig. 1c). Fluorinated carbon atoms are not reproduced very accurately by the parameter sets of standard MD force fields. We therefore amended the AMBER14sb[30] force field and customized the parameters for the -CFH_2, -CF_2H and -CF_3 in MfeGly, DfeGly and TfeGly following the procedure suggested by Robalo *et.al.*[31,32] (see also section IV). We generated the bonded parameters using the GAFF with the Acpype[33] software. For the van-der-Waals interactions, we used the Lennard-Jones parameters for fluorine and hydrogen of fluoromethyl groups (H_F) from Robalo et al.[31,32], who had determined these parameters by fitting against hydration free energies and obtained very accurate results. For the Coulomb interactions, we calculated the atomic partial charges with a two stage RESP[34] fitting protocol, similar as in Ref. 31. In the second stage of this protocol, a conformational sample is generated by an MD simulation with preliminary charges, and the charges are then optimized based on this sample. We used a large sample of 200 structures generated by 100 ns of MD simulations for each amino acid and backbone conformation to ensure that the conformational ensemble of each of the amino acids it well represented. The thus obtained partial charges are reported in SI Fig. 1 and agree well with the charges from Refs. 31 and 32. In our forcefield, the partial charge of fluorine is slightly less negative than in the references, which leads to the partial charge of the highly fluorinated C_γ of TfeGly and DfeGly to be moderately less positive than in the references. The amended force field is included in the

supplementary material.

B. Inhibition assay and binding constants

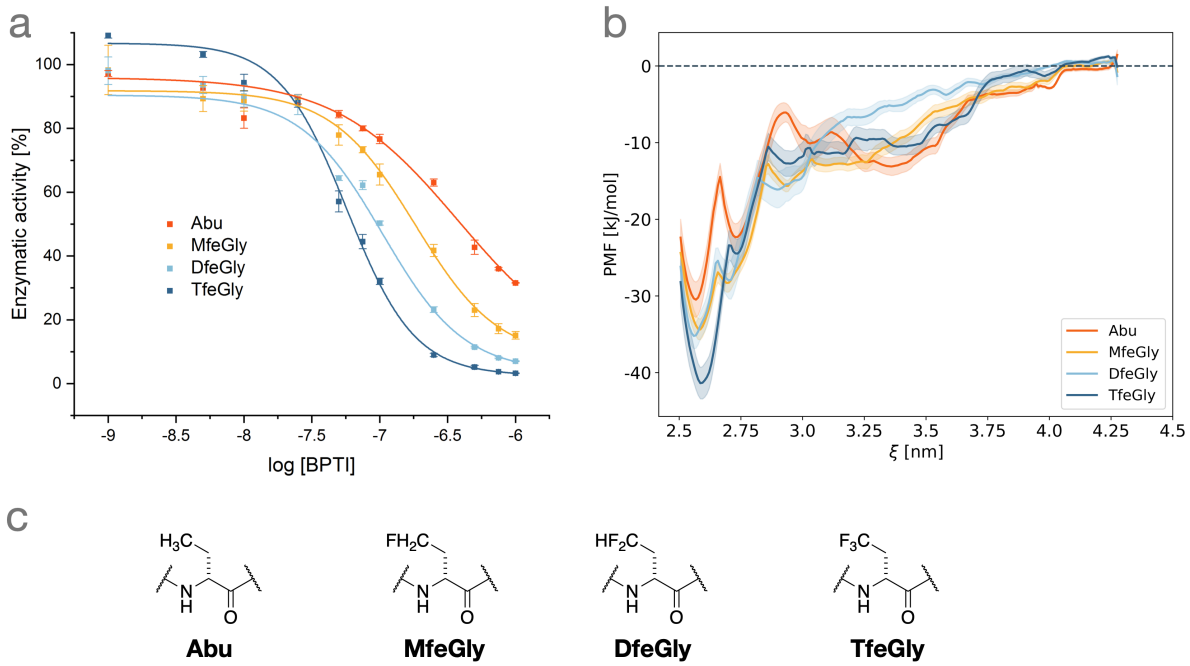

FIG. 1. Inhibition of β-trypsin by BPTI variants with Abu derivatives at P1 position. (a) The inhibition assay shows a stepwise increase of inhibitor strength with increasing fluorination of the Abu side chain. (b) The Potential of Mean Force, here as an estimation of the binding free energy, from umbrella sampling simulations of the unbinding process shows the same trend. (c) Chemical structures of Abu, MfeGly, DfeGly and TfeGly.

Next, we confirmed that fluorination of Lys15Abu-BPTI increases the inhibitor strength by repeating the inhibition assay from Ref. 29. We now additionally had the mono-fluorinated variant Lys15MfeGly-BPTI available and included it in the assay (Fig. 1.a). Furthermore, we now could determine the IC_{50} value of the unfluorinated variant Lys15Abu-BPTI. As before, fluorination of the methyl group in the Abu side chain lead to a stepwise increase of the inhibitor strength from $IC_{50} = 383$ nM ($3.83 \cdot 10^{-7}$ M) for the unfluorinated Lys15Abu-BPTI to $IC_{50} = 57$ nM ($5.7 \cdot 10^{-8}$ M) for the fully fluorinated Lys15TfeGly-BPTI. The IC_{50} for Lys15DfeGly-BPTI and Lys15TfeGly-BPTI are in good agreement with the previously reported values[29] (Tab. I).

We converted the IC_{50} values into binding constants and binding free energies (see Tab. I). This conversions depends on several assumptions, since the inhibition assay reports on enzyme activity and not on thermodynamic property[35,36]. However, within these assumptions, the binding free energy decreases almost linearly from -37.1 kJ/mol in Lys15Abu to -41.9 kJ/mol in the fully fluorinated Lys15TfeGly. Thus, the compounded effect of fluorinating the Abu sidechain is -4.8 kJ/mol, were each fluorine atom contributes about out -1.5 kJ/mol to stabilizing the fluorinated BPTI-β-trypsin complex. Overall, this is a small but reproducible and robust effect in the enzyme activity inhibition assay.

We find the same effect in our MD simulations. Fig. 1 shows the potential of mean force (PMF) for the unbinding of the four complexes, which we calculated by umbrella sampling and weighted-histogram analysis (34 umbrella windows and 1.7 μs total simulation time per PMF curve, see section IV). The minima at $\xi = 2.6$ nm corresponds to the fully bound state, and for $\xi > 4.1$ nm the complex is fully unbound. We approximated the binding free energy as the difference between the PMF at the unbound state and the PMF at the bound state (Tab. I). The PMF of the bound state decreases from the unfluorinated Lys15Abu-BPTI to the fully fluorinated Lys15TfeGly-BPTI, thus reproducing the effect from the inhibition assay. The values of the binding free energies from our PMFs are in good agreement with the experimental values.

The PMF show several interesting features. Most notably, the PMF of Lys15Abu-BPTI is more rugged than the PMFs of the fluorinated variants. Lys15Abu-BPTI has a broad minimum between $\xi = 3.0$ nm and the unbound state which corresponds to an encounter complex. There is a considerable barrier at $\xi \approx 3.0$ between this encounter complex and the bound state. In the fluorinated variants, we instead find a slowly decreasing PMF between $\xi = 3.0$ nm and the unbound state, which is separated only by a minimal barrier from the bound state. Second, the fully bound state in Lys15Abu-BPTI ($\xi = 2.6$ nm) is separated by a sharp and large barrier from a pre-bound state at ($\xi = 2.7$ nm). The fluorinated variants also show such a pre-bound state. However, the pre-bound state is not as stable (i.e. higher PMF relative to the fully bound state) and only separated by a small barrier from the fully bound state.

	Ref. 29		experiment, this work			sim, this work
system	K_a [M^{-1}]	ΔG_{bind} [kJ/mol]	IC$_{50}$ [M]	K_a [M^{-1}]	ΔG_{bind} [kJ/mol]	ΔG_{bind} [kJ/mol]
Lys15Abu			$(3.83^{+0.102}_{-0.102}) \cdot 10^{-7}$	$(2.91^{+0.3}_{-0.2}) \cdot 10^{6}$	$-37.1^{+0.1}_{-0.2}$	-30.49±2.31
Lys15MfeGly			$(1.83^{+0.109}_{-0.109}) \cdot 10^{-7}$	$(6.09^{+0.8}_{-0.5}) \cdot 10^{6}$	$-39.0^{+0.2}_{-0.3}$	-34.27±1.36
Lys15DfeGly	$3.88 \cdot 10^{7}$	-43.6	$(1.05^{+0.115}_{-0.115}) \cdot 10^{-7}$	$(1.06^{+0.2}_{-0.1}) \cdot 10^{7}$	$-40.4^{+0.3}_{-0.5}$	-35.22±1.67
Lys15TfeGly	$5.20 \cdot 10^{7}$	-44.3	$(5.73^{+0.115}_{-0.115}) \cdot 10^{-8}$	$(1.94^{+0.2}_{-0.1}) \cdot 10^{7}$	$-41.9^{+0.2}_{-0.3}$	-41.38±2.04

TABLE I. Binding constants and binding free energies derived from the inhibition assay. We assume $T = 300$ K.

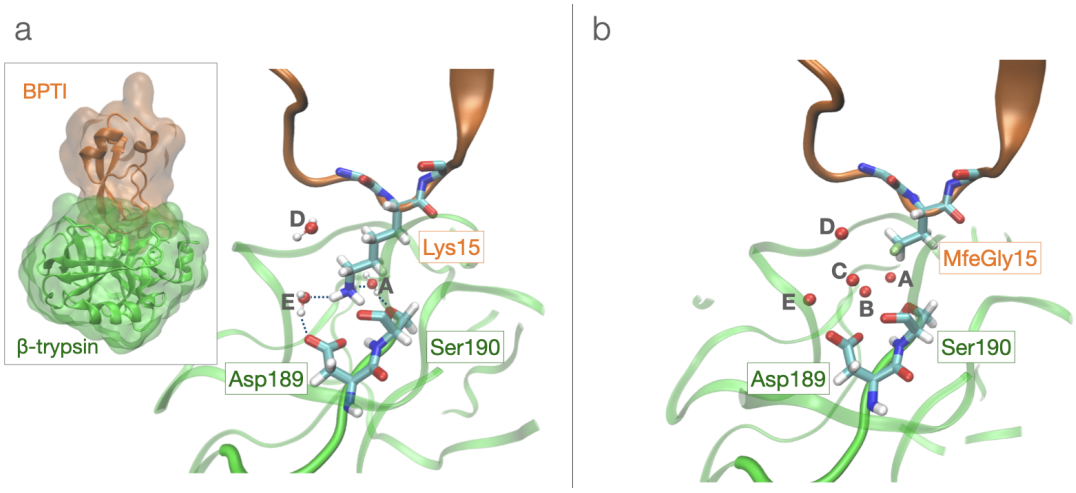

FIG. 2. Structure of the BPTI-β-trypsin complex. (a) Wild-type BPTI-β-trypsin complex (pdb: 3OTJ[18]). (b) Lys15MfeGly-BPTI-β-trypsin complex (this work, pdb: 7PH1).

C. Crystal structure of the Lys15MfeGly-BPTI-β-trypsin complex

Fig. 2.a shows the structure of the wild-type(wt)-BPTI in complex with β-trypsin[18]. β-trypsin has a deep S1 binding pocket, and the catalytic triad is located at the rim of this binding pocket. See Ref. 16 for a schematic drawing of the binding pocket including the catalytic triad. Asp189 at the bottom of the binding pocket governs the specificity of β-trypsin for positively charged side chains, such as lysine and arginine. wt-BPTI inserts the residue Lys15 into the binding pocket (P1 residue). However, the lysine side chain is not long enough to form a direct salt bridge to Asp189. Instead, the contact is mediated by hydrogen bonds to a trapped water molecule (E). Lys15 forms additional hydrogen bonds to a second trapped water molecule (A) and to Ser190 in β-trypsin. Finally, crystal structures of BPTI-β-trypsin complexes show a third water molecule (D) in the binding pocket close to the backbone of Lys15. The water molecules in Fig. 2 are labeled according to Ref. 29.

Having synthetic access to the monofluorinated amino acid MfeGly[37] and the corresponding variant of BPTI[38], we now determined the crystal structure of the Lys15MfeGly-BPTI-β-trypsin complex and deposited it in the Protein Data Bank (pdb: 7PH1). The crystal structure shows five water molecules in the binding pocket (Fig. 2). Water molecules A, D and E are located a the same positions as in the wt-BPTI-β-trypsin complex. Because the side chains of the P1 residue in the Lys15MfeGly-BPTI variants is considerably shorter than lysine, there is more space in the S1 binding pocket. This additional space is occupied by two additional water molecules, labeled B and C. This is in line with the previoulsy published crystal structures of the Lys15Abu-BPTI-β-trypsin complex and its di- and tri-fluorinated variants[29], which also show water molecules at the same five positions. See Ref. 29 for a detailed discussion of the distances between the trapped water molecules and the two proteins in the crystal structures.

As noted in Ref. 29, the B-factors of the water molecules in the Lys15Abu-BPTI-β-trypsin complex differ markedly from the B-factors in the fluorinated variants. The lower B-factors in the fluorinated variants indicate that the water molecules are less mobile in these variants and might act as a "hydration shell" or "contour line" of the fluorinated P1 residue which effectively acts as an extension of this residue[29] and thereby increases the binding affinity. The B-factors of water molecules in the mono-fluorinated Lys15MfeGly-BPTI-β-trypsin complex complete this picture. They are slightly higher than the B-factors of the di- and tri-fluorinated variants, but considerably lower than the B-factors of the complex with Lys15Abu-BPTI (Fig. 3. (bottom)).

D. Water network in the S1 binding pocket

a. The hypothesis. Based on the mobility of the water molecules in the crystal structures, the following mechanism to explain the influence of the fluorine substituents on the binding affinity seems plausible[29]: In contrast to Abu, the (partially) fluorinated methyl-group in MfeGly, DfeGly and TfeGly can form directed non-covalent interactions with water molecules B and C. These two water molecules form hydrogen bonds to water molecules E and A, and the resulting hydrogen bond network that bridges the gap between the truncated P1 side chain and Asp189. The stability of this hydrogen bond network increases with the number of fluorine atoms, and thereby increases the binding affinity.

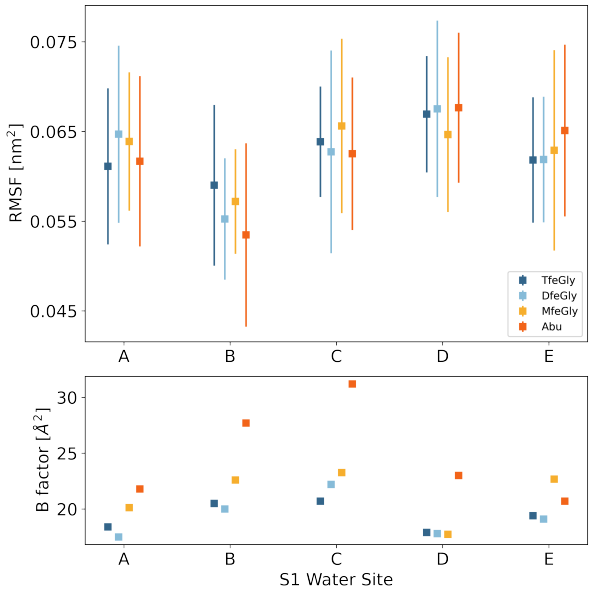

FIG. 3. (top) RMSF of water molecules with a lifetime longer than 200 ps at the hydration sites A-E in the β-trypsin S1 binding pocket. (bottom) B factors of the waters in the β-trypsin S1 binding pocket as reported in ref.29.

With MD simulations, we can test this hypothesis by monitoring the structure, the dynamics and the energy of the water molecules in the S1 binding pocket. Specifically, we would expect to see the following trends with increasing fluorination:

- increasing stability of the hydrogen bond network in the binding pocket
- decreasing fluctuations in the hydrogen bond network in the S1 binding pocket.
- increasing survival times of the water molecules in the binding pocket
- decreasing mobility of the water molecules within the binding pocket
- decreasing entropy of the waters in the S1 binding pocket
- a systematic trend in the water-water and the water-solute enthalpy

b. Structure of the water molecules in the S1 binding pocket. In all four complexes, the five positions A-E are occupied on average between 84 % and 95 % of the simulation time. The only exception is the Lys15MfeGly-BPTI-β-trypsin complex, in which positions D and E are populated less frequently (Fig. 4.b). Thus, on average the S1 water network of the solvated complexes resembles the water network in the crystal structures.

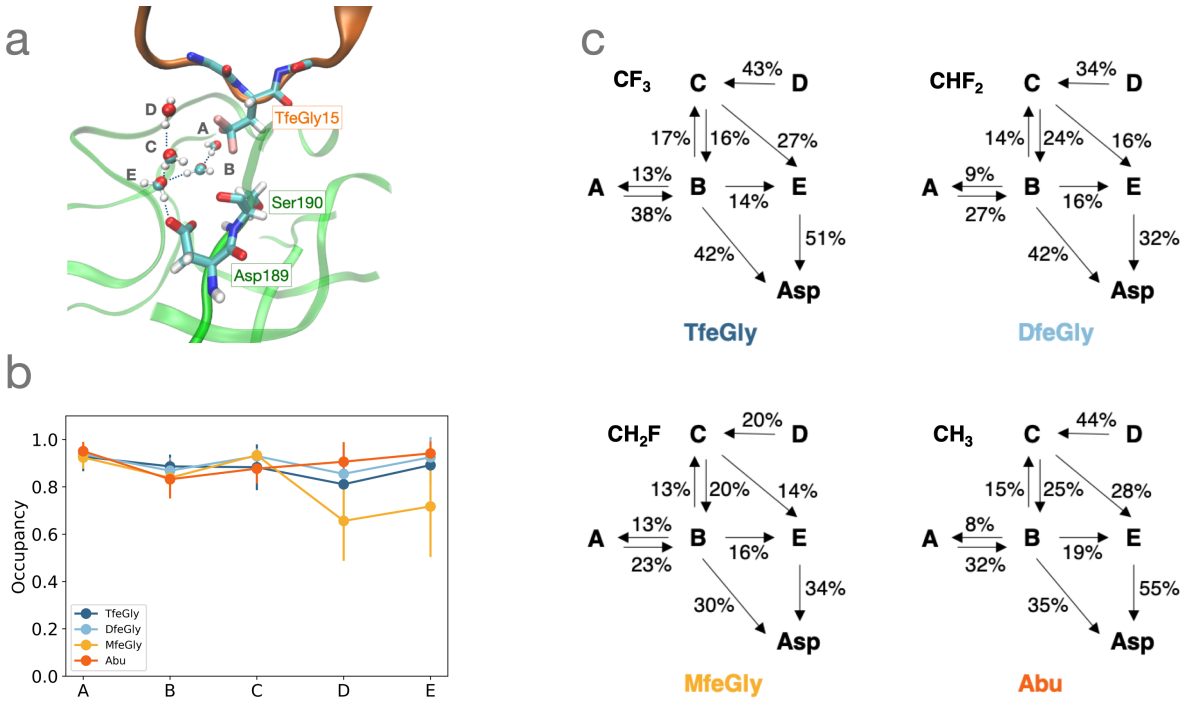

FIG. 4. Structure of the water molecules in the S1 binding pocket. (a) Simulation snapshot of the S1 binding pocket with waters A-E. (b) Occupancy of the hydration sites A-E, showing how often the respective hydration site is populated by at least one water molecule throughout the simulations. (c) Hydrogen bond network. Relative populations of the hydrogen bonds formed by waters A-E to each other and to the side chain of Asp189. The arrows point from the donor to the acceptor.

The waters in the narrow S1 pocket are in direct proximity to each other and can easily form hydrogen bonds. We calculated the frequency of the hydrogen bonds between waters A-E by counting in how many snapshots of our simulations a hydrogen bond is accepted from or donated to another of the S1 waters or Asp189 of β-trypsin. We find the same hydrogen bond network in all of the variants, with slightly to moderately varying relative populations (Fig. 4.c). The hydrogen bonds are the same as proposed in ref. 29.

The most frequent hydrogen bond is E–Asp189, which is in place for more than 50 % of the snapshots for the complexes of fully- and unfluorinated variants Lys15TfeGly-BPTI and Lys15Abu-BPTI. For the partially fluorinated complexes of Lys15DfeGly-BPTI and Lys15MfeGly-BPTI, it occurs just higher than 30 %. Asp189 also accepts a hydrogen bond from B. This hydrogen bond is more frequent in the higher fluorinated complexes

of Lys15TfeGly-BPTI and Lys15DfeGly-BPTI, where it is in place in 42 % of the snapshots, while it occurs for Lys15MfeGly-BPTI and Lys15Abu-BPTI 30 % and 35 %, respectively. The hydrogen bond D–C is again more often observed for the fully- and unfluorinated complexes. For Lys15MfeGly-BPTI, this can be explained by the fact of positions D and E being less frequently occupied. The other hydrogen bonds observed, namely A–B, B–A, B–C, C–B, B–E and C–E, are about equally frequent. Overall we do not observe a change in the structure of the water molecules in the S1 binding pocket or a trend in the stability of the hydrogen bond network with increasing fluorination.

To our surprise, we did not find any significant frequency of hydrogen bonds between the fluorine substituents and the water molecules at the hydration sites A-E. (We used the definition of a F-hydrogen bond from Ref. 39). The only exception is Lys15MfeGly-BPTI, in which the fluorine substituent occasionally forms hydrogen-bonds to a trapped water molecule (12.7 % of the frames), however in these cases the bonded water molecule cannot be assigned to any of the five hydration sites A-E. The absence of F-hydrogen bonds becomes plausible if one considers the orientations of the water molecules in the S1 binding pocket (Fig. 4.a). Water E acts as an hydrogen bond donor to Asp189 in β-trypsin and therefore faces with its oxygen towards BPTI. This enforces a similar orientation in the hydrogen-bonded water molecules A-D. Thus, the fluorine substituent, which would act as hydrogen bond acceptor, is in the vicinity of several water molecules, but the hydrogens in these water molecules face in the wrong direction.

c. Dynamics of the water molecules in the S1 binding pocket. Because the fluctuation around this static picture is rather large, we analyzed the dynamics of the water molecules in the binding pocket more closely. We calculated how long water molecules reside inside the S1 pocket, defined here as 5Å-spheres around the side chain ends of two critical amino acids at the bottom of the pocket, Asp189 and Ser190. The calculated survival probability is a measure of how likely it is, that a water molecule found in the S1 pocket at time t_0 will still be there at time $t_0 + t$ (Fig. 5.a). After an initial fast decay until $t \approx 50 - 100\,\mathrm{ps}$, which reflects local fluctuations across the boundary of the 5Å-spheres, the survival probability decays exponentially with mean lifetimes of 0.9 to 1.3 ns. Water molecules in the Lys15Abu-BPTI-β-trypsin complex seem to have a longer residence time than in the fluorinated BPTI variants. Note, however, that the difference does not exceed one standard deviation and might not be statistically significant.

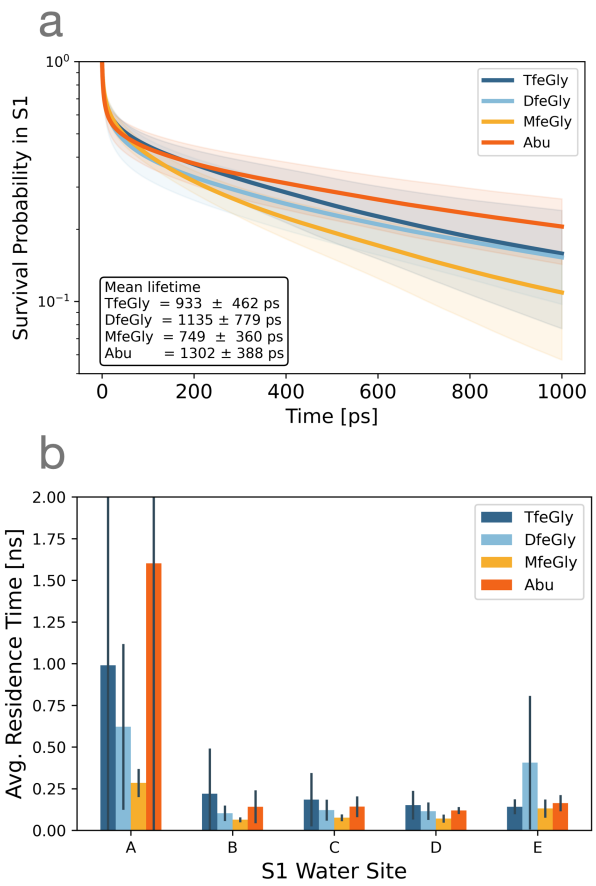

FIG. 5. Dynamics of the water molecules in the S1 binding pocket. (a) Average survival probability of an arbitrary water molecule in the S1 binding pocket. The mean lifetime is approximated by fitting a single-exponential decay function to the data after the initial burn phase. (b) Average residence time of a water molecule at the hydration sites A-E. The data is calculated from a total of 1 µs of simulation data for each of the protein complexes.

After determining the residence times for water molecules in the whole binding pocket, we were also interested in their mean residence time at the hydration sites A-E. As the water molecules which reside at these hydration sites exchange frequently throughout extensive MD simulations, keeping track of which water molecule resides at a specific hydration site for a given trajectory snapshot becomes a non-trivial task. We wrote a program to assign one water molecule for every simulation snapshot that resides at each of the hydration sites (see also section IV).

Fig. 5.b shows the average residence times of water molecules at the five positions A-E. In

all four variants, water molecules at position A have a significantly higher average residence time than water molecules at any of the four other positions, likely due to its position deeply buried in the S1 pocket. Water molecules at positions B and D have low residence times of only a few 100 ps, and are thus highly dynamic. It is unlikely that they play a structurally important role. Specifically, water molecules B and D, which are not present in the wild-type BPTI-β-trypsin complex but are hypothesised to play a structural role in the fluorinated variant of the complex, have low residence times of 50 to 250 ps.

When comparing the fluorinated variants of BPTI, we find that residence times tend to increase with increasing fluorination. However, the effect is small compared to the variance of the residence times. More importantly, the residence times in the fully fluorinated Lys15TfeGly-BPTI-β-trypsin are about as large as the residence times in the Lys15Abu-BPTI-β-trypsin complex. The partially fluorinated complexes show lower residence times.

The B-factors, that are used as a measure for the mobility in the crystal structure, cannot easily be calculated from MD simulations. Instead we calculated the root-mean-square fluctuations (RMSF) of water molecules with a residence time longer than 200 ps at their respective hydration site. The RMSF is closely related to the B-factor, and Fig. 3 shows the comparison. The water molecules in the five positions have similar RMSF values, where water molecules in position B have slightly lower RMSF values than the others. However, we do not find any trend in the RMSF with increasing fluorination.

Overall, fluorination does not lead to a rigidification or stabilization of the water network in the S1 binding pocket. Instead partial fluorination (Lys15MfeGly and Lys15DfeGly), leads to an increased mobility of the water molecules, at least locally within the binding pocket. As in the structural analysis, we find that Lys15Abu and the fully fluorinated Lys15TfeGly show similar behaviour.

d. Enthalpy and entropy of the water molecules in the S1 binding pocket. To evaluate the enthalpy and entropy of the trapped water molecules we used Grid Inhomogeneous Solvation Theory (GIST)[40] as implemented in the SSTMap[41–43] command-line tool. In a GIST analysis, the two proteins in the complex are kept rigid while the water molecules are allowed to move freely. The simulation box is then discretized using a regular grid, where the grid cells are called voxels. The water-water (E_{ww}) and the solute-water (E_{sw}) interaction energy as well as the translational (dST_{tr}) and rotational (dST_{tr}) entropic part of the free energy at $T = 300$ K is calculated for each voxel. The entropic terms are calculated relative

to the entropy in bulk water. Fig. 6 reports the Boltzmann-weighted averages of these four properties for each of the positions and for each of the four complexes. Neither the enthalpic nor the entropic terms show a trend with increasing fluorination. In particular the water-solute interactions of water molecules B and C which spacially replace the amino group of the lysine side chain do not change if the fluorination of the P1 methyl group changes. This indicates that the fluorinated methyl groups and the unfluorinated methyl group in Abu engage in similar interactions with these two water molecules and no specific interaction arises due to the fluorine substituents.

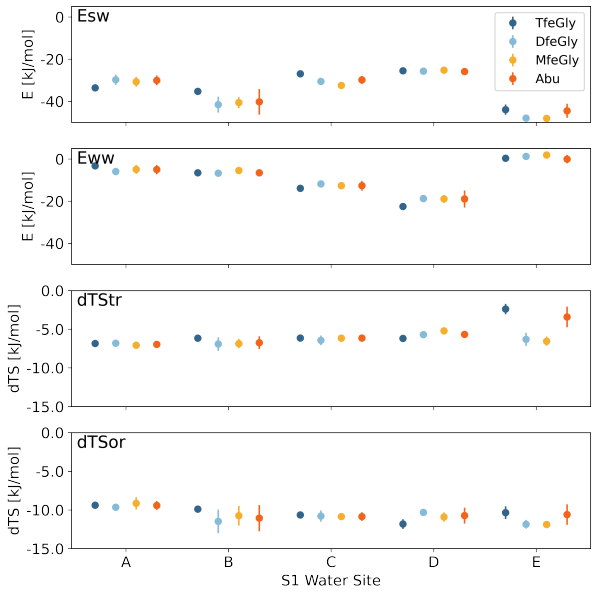

FIG. 6. The solute-water (E_{sw}) and the water-water (E_{ww}) interaction energy and translational (dST_{tr}) and rotational (dST_{tr}) entropy for S1 binding pocket waters, calculated from GIST.

E. Protonation state of His57

β-trypsin contains a histidine, His57, in the binding pocket which can be protonated at the two nitrogen atoms, N(δ) and N(ϵ), in its side chain. In the neutral form or unprotonated state, only one of the side chain nitrogen atoms is protonated, while in the protonated state both nitrogens are protonated. By itself, the pK$_a$ of a protonated histidine side chain is 6.0[44], and histidine is thus largely unprotonated at pH 7.

However, the pK$_a$ of histidine residues can be varied drastically by the protein environment[45–47].

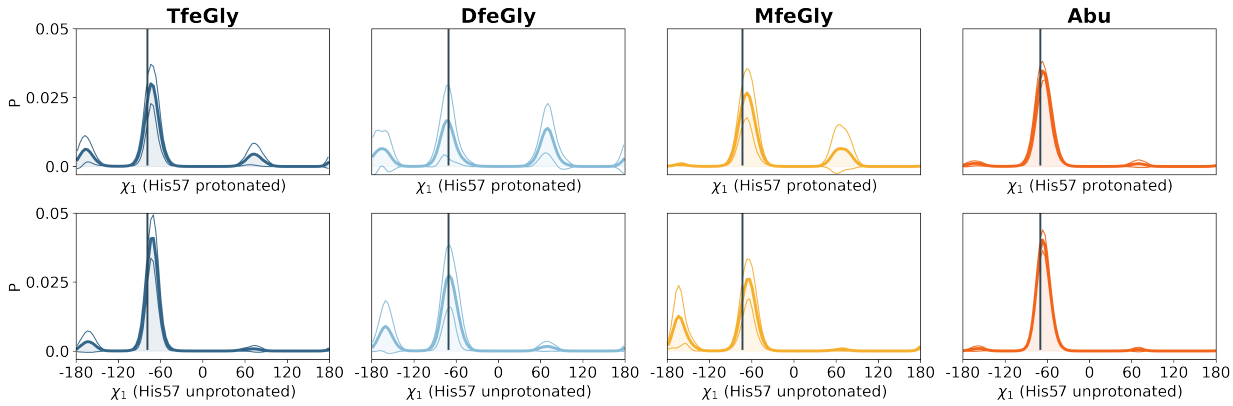

FIG. 7. Influence of the protonation state of His57 on the χ_1-torsion angles of the P1 residue. Histograms with the probability density P of the bins. The vertical line marks the value of the torsion angle from the respective pdb structure. *Top row:* His57 protonated. *Bottom row:* His57 unprotonated. Standard deviations are indicated as thin lines.

In the specific case of β-trypsin, His57 donates a hydrogen bond from N(δ) to the carboxyl group of Asp102. Since the N(δ)-proton is shared between His57 and Asp102, the equilibrium is shifted to the N(ϵ)-protonated state of His57 in the actual protein. That is the pKa of His57 is increased, and we expect it to be protonated. Therefore, the computational results presented so far were generated with His57 in the protonated state.

On the other hand the catalytic mechanism relies on a swift protonation and deprotonation of the His57 residue. We therefore additionally conducted simulation with His57 in the unprotonated state and tested the influence of the protonation state on the P1 residue and the water network.

Fig. 7 shows the χ_1 torsion angle profiles of the P1 residue in all four Lys15X-BPTI-β-trypsin complexes. The unflourinated Lys15Abu-BPTI and the fully fluorinated Lys15TfeGly-BPTI are not influenced by the protonation state of His57. The methyl group in Abu and the trifluoro-methyl group in TfeGly do not have an electric dipole moment and therefore do not interact with His57. Consequently, the P1 residues in Lys15Abu-BPTI and in Lys15TfeGly-BPTI have similar structural preferences. The χ_1-torsion angle remains largely in the gauche+ conformation, i.e the conformation that is also found in all four crystal structures. The methyl group (Lys15Abu-BPTI) and the trifluoro-methyl group (Lys15TfeGly-BPTI) can rotate freely (rotation around χ_2, data not shown) with equally populated gauche+, gauche- and trans conformations. The similar structure of the P1

residue in Lys15Abu-BPTI and Lys15TfeGly-BPTI might explain why the water network in these two variants tends to have similar structure and dynamics (Fig. 3 - Fig. 6).

The partially fluorinated variants, Lys15MfeGyl-BPTI and Lys15DfeGyl-BPTI, can deviate from gauche+ conformation of χ_1 torsion angle, which is found in the crystal structures. These side chains can possibly form dipole-dipole interactions with the backbone of a nearby loop or with the side chain of Ser195, which likely stabilizes the gauche- and the trans conformations. The conformational equilibrium between gauche+, gauche- and trans conformation is slightly influenced by the protonation state if His57 (Fig. 7).

We have also analyzed the water network in the binding pocket when His57 is unprotonated at N(δ). SI Fig. 4-8 in the supplementary material show the same analysis as presented in Figs. 3, 4.b, 4.c, 5.a, and 5.b for the protonated state. As in the protonated state, we do not find any trend that can explain the observed differences in the binding affinity. We do not find the water network and dynamics to be significantly influenced by the protonation state of His57.

III. CONCLUSION

We provide insights from experiments and computational simulation into the water network in the trypsin S1 binding pocket, in complex with Lys15Abu-BPTI and fluorinated analoga. We repeated the inhibition assay from Ref. 29, including the now synthetically available Lys15MfeGly-BPTI, and found a step-wise increase of inhibitor strength with increasing fluorination. In MD simulations, using umbrella sampling and our self-parametrized force field for fluorinated amino acids, we confirmed this step-wise increase in inhibitor strength and estimated a similar binding free energy. We determined the X-ray crystal structure of the complex of trypsin with Lys15MfeGly and found the pocket water configuration to be similar as in the complexes of Abu and the di- and tri- flourinated analoga.

Following the hypothesis of fluorine forming interactions with water molecules and thereby rigidifying the water network in the S1 binding pocket, we analyzed the structure and dynamics of the S1 water molecules computationally with MD simulations. We found the positional water structure seen in the crystal structure to be on average consistent with the protein crystal structure, but we observe the water molecules frequently exchanging positions and leaving or entering the binding pocket.

The water molecules form a hydrogen bond network that is similar for trypsin in complex with Lys15Abu-BPTI and its fluorinated analoga. Surpisingly, we do not find hydrogen bond like interactions between the fluorinated methyl groups and water molecules at the hydration sites A-E. We estimate a water molecule to stay in the S1 binding pocket for a mean lifetime of 0.9-1.3 ns and the exchanges at the specific hydration sites to be even faster, with mean residence times of 50-250 ps with no trend in regard to fluorination. Moreover the enthalpic and entropic properties of water in the region of the hydration sites also do not indicate increasing stability of water molecules with increasing fluorination.

We additionally tested whether the protonation state of His57 in the binding pocket has an influence on the water network. Just as in the protonated state, we did not find any trend when comparing the unprotonated Lys15Abu-BPTI to the partially fluorinated variants Lys15MfeGly-BPTI and Lys15DfeGly-BPTI and the fully fluorinated Lys15TfeGly-BPTI.

We conclude that we could not observe a rigidification or stabilizing effect of fluorination on the water network in the S1 pocket. Ref. 38 reports the van-der-Waals volumes of Abu and its fluorinated variants, which monotonously increase from $V_{\mathrm{vdW}} \approx 40\,\mathring{A}^3$ in Abu to $V_{\mathrm{vdW}} \approx 60\,\mathring{A}^3$ in TfeGly. The corresponding decrease of the B-factors of the water molecules in the X-ray crystal structure (Fig. 3.b) is thus possibly caused by a steric effect and not by the stabilization of the hydrogen bond network. Overall, the experimentally and computationally observed increase in inhibitor strength, respectively binding free energy, cannot be explained by a change in the structure or the dynamics of the water network. Other factors have to play a role.

Ref. 32 investigated the hydrophobicity of fluorinated residues and reports an increase of the hydration free energy from Abu to the fully fluorinated Lys15TfeGly of 1 kcal/mol = 4.2 kJ/mol. Since the P1 residue in BPTI is solvent-exposed, this effect will certainly affect the hydration free energy of the fluorinated BPTI-variants. Thus, the observed increase in inhibitor strength might partially be due to an increase in the free energy of the unbound BPTI, rather than a decrease in the free-energy of the bound state. But we do not think it can exclusively be attributed to the difference in hydrophobicity, since the P1 side chain is (partially) hydrated even in the binding pocket.

Furthermore, fluorination has a sizeable effect on the shape of the potential of mean force. We observe an encounter complex ($\xi > 3$ nm), whose stability seems to change

with increasing fluorination. A similar encounter complex has recently been discovered in the binding process of wt-BPTI to β-trypsin[19]. Additionally, we observe a pre-bound intermediate state whose stability sensitively depends on the fluorination. This pre-bound state does not seem to exist in the wt-BPTI-β-trypsin complex[19]. We plan to investigate the free energy landscape and the unbinding process more closely using dynamic reweighting techniques[48–50].

From a broader perspective, our study shows that fluorination is a sensitive tool to tune the affinity of a protein-protein complex. However, one cannot rationalize the observed effects by only considering the fully bound complex. Rather fluorine substituents also influence the solvation properties of the unbound ligand, the specific interactions within the fully bound complex, the stability of the encounter complex and the stability of intermediate states along the binding pathway. All these effects act together to generate the affinity of a fluorinated protein-protein complex.

IV. METHODS

A. Parametrization of the fluorinated Abu-variants

a. Parametrisation protocol. The AMBER14sb[30] force field was amended to include the parameters for the non-standard fluorinated amino acids and Abu. The parametrization process followed the protocol of Robalo et al.[31]. Bonded parameters were generated from GAFF with the Acpype[33] software. Lennard-Jones parameters for fluorine and hydrogen of fluoromethyl groups (H_F) were taken over from Robalo et al.[31,32]. Atomic partial charges were calculated using a two stage RESP fitting protocol. For each amino acid, two dipeptide starting structures Ace-[Abu-variant]-Nme were constructed. One starting structure with the ϕ and ψ backbone dihedrals corresponding to alpha helix and the other one corresponding to beta sheet secondary structure. The structures were geometrically optimized using a molecular mechanics steepest descend algorithm. A single point QM energy calculation was performed at the HF/6-31-G* level of theory and the electrostatic potential (ESP) surface was evaluated using Gaussian16[51]. Using the Antechamber[52] package, the ESP was fitted in a two stage RESP[34] procedure to obtain atomic partial charges. During the RESP fit, atomic charges of the backbone amine and carbonyl were fixated to the AMBER14sb

charge of the corresponding atoms for amino acids with neutral sidechains. The atomic partial charges of the Ace- and Nme-caps were fixated to the partial charges in Cornell et al.[53]. MD simulations of 100 ns each with constrained backbone atom positions were used to generate 200 conformations of each amino acid (100 conformations of alpha helix and 100 conformations of beta sheet). These conformations were submitted to a second RESP fit iteration under the same conditions as the first fit. The final atomic partial charges were calculated as the mean values of the second RESP fits.

b. MD simulations for second RESP fit iteration. The starting structures were placed into a cubic box of tip3p[54] water with a buffer length of 1.0 nm between the box edges and the solute. The simulations were conducted in the modified AMBER14sb force field with atomic partial charges from the first iteration. The system was energy minimized using a steepest descent algorithm and equilibrated in the NVT ensemble for 100 ps at 300 K, followed by an equilibration in the NPT ensemble for 100 ps at 300 K and at 1.0 bar. Production MD simulations of the equilibrated system were performed in the NPT ensemble for a length of 100 ns. A velocity-rescaling scheme[55] was used as thermostat and the Parinello-Rahman barostat[56] was applied. The backbone atoms were constrained throughout equilibration and production simulation to conserve the backbone dihedral structure. Out of the resulting trajectories, 100 structures of the alpha helix conformations and 100 structures of the beta sheet conformations were extracted in equally spaced time intervals of 1 ns.

B. MD simulations of the BPTI-Trypsin complexes and analysis

a. Unbiased simulations. MD simulations of the protein complexes were run with Gromacs 19-4[57–60] and the self-parametrized amended AMBER14sb force field described above. The crystal structures of the protein complexes (pdb code: 4Y11, 4Y10, 4Y0Z) were freed from water, co-solutes and ions and placed into a dodecahedric box with 1.0 nm buffer between the solute and the box edges. One of the fluorine atoms in 4Y10 was mutated to a hydrogen atom to yield the Lys15MfeGly-BPTI β-trypsin complex. The complexes were solvated in tip4p[54] water, energy minimized with a steepest descent algorithm and equilibrated in the NVT ensemble for 100 ps at 298 K, followed by an equilibration in the NPT ensemble for 100 ps at 298 K and 1.0 bar. A velocity-rescaling scheme[55] was applied as thermostat and the Parrinello-Rahman barostat[56] was applied. An initial simulation of 100 ns was run to

extract starting conformations for the production simulations in equally spaced timesteps of 10 ns to ensure decorrelation. The extracted starting conformations were prepared, solvated, minimized and equilibrated as described above. 10 production simulations of 100 ns were run for each complex in the NPT ensemble at 298 K and 1.0 bar. Snapshots were saved in 1 ps timesteps. After production run, the trajectories were aligned based on backbone RMSD with the pdb structure 4Y11 as reference. Further handling of the trajectories for analysis was generally done with the MDTraj[61] package.

To test the effect of the protonation state of His57 on the water dynamics, we produced one set of simulations with N(δ) and N(ϵ) of His57 protonated and one set of simulations where N(δ) is unprotonated.

b. Umbrella Sampling: Fig.1b. Umbrella Sampling simulations of the complexes of the four Abu-derived BPTI-β-trypsin complexes were run in 34 windows along the reaction coordinate, each with a production simulation time of 50 ns. The reaction coordinate was defined to be the center of mass distance of the protein heavy atoms. Starting structures for the sampling windows were extracted from pulling simulations where a harmonic potential with a spring constant of $k_{\text{pull}} = 1000\ \text{kJ} \cdot \text{mol}^{-1} \cdot \text{nm}^{-2}$ was attached to the BPTI derivative and moved away from β-trypsin (increasing the center of mass distance) at a speed of 0.01 nm/ps. The starting structures for the sampling windows were extracted every 0.05 nm along the reaction coordinate starting at 2.6 nm. Initially, the structures were equilibrated in the NPT ensemble for 100 ps with the Berendsen barostat, followed by a 1 ns equilibration using the Parinello-Rahman barostat, which was also used in the production runs. The thermostat employed was a velocity rescaling scheme[55]. A harmonic potential with a force constant $k = 6278\ \text{kJ} \cdot \text{mol}^{-1} \cdot \text{nm}^{-2}$ was applied to the sampling window simulations. For the pulling simulations, the starting structures from the unbiased MD runs of the complexes were solvated in tip4p water in a cubic box with 2.0 nm buffer between solute and box edges to ensure enough space for the pulling. The systems were minimized and equilibrated in the same way as the unbiased MD simulations. Potential of mean force (PMF) profiles were calculated following the weighted histogram analysis method[62] (WHAM) with the Gromacs-internal WHAM program. The binding free energy was calculated by the difference of the global minimum of the PMF profile and the PMF of the separated complex. To estimate statistical uncertainty, a simplified bootstrapping scheme was applied by separating the trajectory of every window into five parts of 10 ns and combining the parts in the following

way (beginning and starting time indicated in ns): (0-10, 10-20, 20-30, 30-40); (0-10, 10-20, 20-30, 40-50); (0-10, 10-20, 30-40, 40-50); (0-10, 20-30, 30-40, 40-50); (10-20, 20-30, 30-40, 40-50). PMF profiles of the bootstrapping samples were calculated and the mean and standard deviation were calculated. See supplementary material.

c. RMSF of long resident waters: Fig. 3. Using the assignment of water molecules to the hydration sites A-E, water molecules that reside longer than 200 ps at a hydration site were identified. The RMSF of these water molecules, while they reside at the respective site was calculated and the average was estimated for every complex and hydration site.

d. Assignment of water molecules to hydration sites: Fig. 4b. As the water molecules which are present at the S1 pocket hydration sites A-E exchange frequently throughout the MD simulations, one specific water molecule was assigned to each site on a frame-by-frame basis. The position of the hydration sites was based on the positions of the water oxygens in the respective pdb structure of the protein complex. A sphere around each of the positions of these water oxygens at site A-E with a radius of 2.0 Å was defined and kept fixated at this position throughout the trajectory. Iterating over the trajectory frames, the water molecules whose oxygen atom is placed inside the spheres were identified. As the spheres for A-E were overlapping, in case a water molecule appeared in more than one sphere, it was assigned to the one where it was closest to the center (Voronoi tesselation). The closest water to the center within each sphere was assigned to be present at the respective hydration site for the frame if there were more than one water inside the sphere. If there was no water in the sphere after Voronoi tesselation, the hydration site was regarded as unoccupied. The code for the assignment of water molecules is available at:
https://github.com/bkellerlab/track_waters.

e. Hydrogen bond network: Fig 4c. To minimize the need for large amounts of memory, water that was not in proximity to the S1 binding pocket was deleted from the trajectories before submitting them to the calculation of hydrogen bonds. The water was removed dynamically on a frame-by-frame basis. The code is available at:
https://github.com/bkellerlab/track_waters. Hydrogen bonds were estimated with the Wernet-Nilsson[63] method of the MDTraj[61] package. For every trajectory frame in particular, it was estimated if there are hydrogen bonds found between the waters at the hydration sites and to Asp189. The amount of frames in which the hydrogen bond was present was divided by the total number of trajectory frames to estimate the percentage

how frequent the respective hydrogen bond is in place. To find hydrogen bond like interactions between the fluorinated methyl groups of MfeGly, DfeGly and TfeGly and the S1 waters A-E, the distance between their oxygen atoms and the fluorine atoms, as well as the O-H-F angles were calculated for every trajectory frame. Using the definition in ref. 39 of a hydrogen bond like interaction, the frames in which the distance was smaller than 0.35 nm and the corresponding angle larger than 160° were counted as showing a hydrogen bond like interaction between fluorine and the respective water molecule. In a separate analysis, the distance and O-H-F angle of every water molecule coming close enough to the fluorine atoms was calculated for the trajectories to find the number of frames in which hydrogen bond like interactions are in place involving any water molecule, not just those that could be assigned to the S1 pocket hydration sites.

f. Survival probability in the S1 pocket: Fig 5a. The survival probability of water molecules in the S1 binding pocket was calculated using the waterdynamics[64] module of the MDAnalysis[65,66] package. The volume that belongs to the S1 pocket was defined as 5Å-spheres around the side chain ends of Asp189 and Ser190. The survival probability of a water molecule at time t in the S1 pocket, given for a lag-time τ, is defined as the probability that it is still present in the pocket at time $t + \tau$. The probability is averaged over the trajectory:

$$P(\tau) = \frac{1}{T} \sum_{t=1}^{T} \frac{N(t, t+\tau)}{N(t)}.$$

The survival probability up to a lag-time of 1 ns was calculated for each trajectory. An average was formed from the results for each complex and plotted in Fig. 5a. The mean lifetime t was approximated by fitting a single-exponential decay function of the form $P(\tau) = A \cdot e^{\frac{1}{t}\tau}$ to the survival probability of each trajectory after an initial burn phase of 100 ps to account for waters that have very short survival lifetime due to fluctuating in and out around the edge of the defined S1 pocket zone.

g. Mean residence time: Fig 5b. After assigning water molecules to the S1 pocket hydration sites as described above, exchanges of water molecules were counted. The total simulation time was divided by the number of exchanges to estimate the mean residence times of a water molecule at the particular hydration sites. To account for fluctuations on the edge of the spheres, simulation chunks of 20 ps were regarded, where the water molecule found most often at the hydration site in the chunk was assigned to that site for the whole

chunk. The resulting mean residence times were averaged over the trajectories for each complex.

h. GIST calculation: Fig. 6. Thermodynamic properties of water molecules were calculated based on Grid Inhomogeneous Solvation Theory (GIST). For the underlying simulations, the crystal structures of the protein complexes (PDB code: 4Y11, 4Y10, 4Y0Z) were prepared, solvated and equilibrated in the same way as the unbiased simulations of the BPTI-Trypsin complexes. Positional restraints were imposed on every protein heavy atom for the entire equilibration and production runs of 100 ns length. The GIST calculations were performed with the SSTMap[40–43] command-line tool. A cubic grid with side-length of 26 Å and grid spacing of 0.5 Å was centered on the C_γ atom of the BPTI P1 residue to capture the active site. The resulting voxel properties of E_{sw}, E_{ww}, dTS_{tr} and dTS_{or} were visually inspected in PyMol and clusters were identified around the S1 water hydration sites A-E of the crystal structures. Spheres of 2.0 Å radius were defined around the positions of the hydration sites (defined by the position in the crystal structure of 4Y11) and the voxels inside the spheres were grouped to the corresponding water site. Due to overlapping of the spheres, voxels were assigned to the closest water sites by Voronoi tessellation. The thermodynamic properties E_{sw}^{norm}, E_{ww}^{norm}, dTS_{tr}^{norm} and dTS_{or}^{norm} of the voxels were averaged by Boltzmann weighted averaging respectively for each voxel cluster group to obtain the mean magnitude of the thermodynamic property for a voxel assigned to the water site. The Boltzmann weights were determined by the number of water molecules N_{wat} from the SSTMap output. For each complex, four different simulations were equilibrated and run and the analysis was done on all four simulations to estimate statistical uncertainty.

i. Dihedral angle distribution: Fig 7 The dihedrals χ_1 and χ_2 were calculated from the unbiased MD simulations with the MDTraj[61] package. The histograms present an average distribution of the 10 trajectories for each complex. The standard deviation was added and substracted from the average to calculate the error lines. The dihedrals in the PDB crystal structures 4Y11, 4Y10, 7PH1 and 4Y0Z were measured for comparison to the dihedral distributions from the MD simulations.

C. Enzyme activity inhibition assay: Fig.1a and Tab. I

a. Enzyme activity inhibition assay. For inhibitory assays increasing concentrations of BPTI variants (0, 1, 5, 10, 25, 50, 75, 100, 250, 500 750, 1000 nm) were incubated with 20 µL β-trypsin (100 nM, Sigma-Aldrich) in 200 mM Triethanolamine, 20 mM $CaCl_2$, pH 7.8 in a 96-well-plate for 1 h. 20 µL Nα-Benzoyl-L-arginine 4-nitroanilide hydrochloride (BApNA, Sigma-Aldrich) (1 mM) were added to 180 µL preincubated enzyme/inhibitor mixture. Hydrolysis of BApNA and formation of the product p-Nitroaniline (pNA) was monitored by measuring absorbance at 405 nm in an Infinite M200 microplate reader (Tecnan Group AG, Männedorf, Switzerland) for 30 min at 25 °C. Initial velocities were determined by plotting absorbance against reaction time and IC_{50} values by plotting residual enzyme activity against logarithmic inhibitor concentration using OriginPro 2020b (OriginLab Corporation, Northampton, MA, USA).

b. Conversion between IC_{50} and ΔG_{bind}. To relate the half maximal inhibitory concentrations (IC_{50}) measured in the inhibition assays to the binding free energy ΔG_{bind}, we consider the binding equilibrium of of the BPTI-variants Lys15X-BPTI to β-trypsin

$$\text{Lys15X-BPTI} + \beta\text{-trypsin} \rightleftharpoons \text{Lys15X-BPTI-}\beta\text{-trypsin}$$

The enzyme inhibition constant, or inhibitor dissociation constant, K_i for this equilibrium is defined to be equal to the dissociation constant of the inhibitor-enzyme complex

$$K_i = K_d = \frac{[\text{Lys15X-BPTI}][\beta\text{-trypsin}]}{[\text{Lys15X-BPTI}\beta\text{-trypsin}]}\,, \tag{1}$$

where $[A]$ denotes the concentration of species A. Thus, K_i has units of concentration. We related the half maximal inhibitory concentrations (IC_{50}) measured in the inhibition assays to the inhibitor dissociation constant via the Cheng-Prusoff equation[67]

$$K_i = \frac{\text{IC}_{50}}{1+\frac{[S]}{K_m}}\,, \tag{2}$$

where $K_m = 8.85 \pm 3.60$ mM is the Michaelis constant of β-trypsin. We determined the Michaelis constant of β-trypsin ourselves (see below). The substrate concentration was $[S] = 1$ mM in the inhibition assay in Fig. 1. With these values IC_{50} and K_i can be interconverted as $K_i = 0.898 \cdot \text{IC}_{50}$ and $\text{IC}_{50} = 1.113 \cdot K_i$.

Table I reports the association constant which is the inverse of the inhibition constant

$$K_a = \frac{1}{K_i}\,. \tag{3}$$

and the binding free energy

$$\Delta G_{\text{bind}} = -RT \ln \left[K_a \cdot c^0 \right] \tag{4}$$

where $R = 8.314$ J/(K mol) is the ideal gas constant, $T = 300$ K is the temperature, and $c^0 = 1\,\text{M}$ is the standard concentration. We obtained the uncertainties in K_a and ΔG_{bind} by propagating the uncertainties in IC_{50} and K_m.

c. Determination of Michaelis-Menten constant. To determine the Michaelis-Menten constant K_m, 20 µL Nα-Benzoyl-L-arginine 4-nitroanilide hydrochloride (BApNA, Sigma-Aldrich) were added in varying concentrations (0, 0.1, 0.2, 0.5, 0.7, 1.0, 1.2 mM; final concentrations of the measurement) to 180 µL β-trypsin (100 nM, final concentration, Sigma-Aldrich) in 200 mM Triethanolamine, 20 mM $CaCl_2$, pH 7.8 in a 96-well plate. Hydrolysis of BApNA and formation of the product p-Nitroaniline (pNA) was monitored by measuring absorbance at 405 nm in an Infinite M200 microplate reader (Tecan Group AG, Männedorf, Switzerland) for 10 min at 25 °C. Initial velocities were determined by plotting absorbance against reaction time. K_m was determined by using Enzyme Kinetics plug-in in OriginPro 2020b (OriginLab Corporation, Northampton, MA, USA).

D. Protein Crystallography

To form complex between trypsin and BPTI variants, 60 µL of 2.0 mM BPTI in 25 mM Tris, 10 mM $CaCl_2$, pH 7.4 were added to 50 µL of bovine trypsin in 25 mM Tris/HCl, 10 mM $CaCl_2$, pH 7.4. The solution was incubated for 12 h at 4 °C. β-trypsin-BPTI complex was purified by dialysis. Initial crystals were obtained by the sitting-drop vapor-diffusion method at 18 °C with a reservoir solution containing 2.2 M ammonium sulfate. Inter-grown crystals were used to prepare a seed stock. With a cat whisker, seeds were transferred to a freshly prepared crystallization drop. Well-formed crystals were soaked in 30 % (v/v) glycerol plus reservoir solution and frozen in liquid nitrogen. Data was collected at Berlin BESSYII, beamline 14.2. X-ray data collection was performed at 100 K. Diffraction data were processed with the XDS in space group I222 (Table SI-XX).[68] The structure of the Trypsin-BPTI-MfeGly was solved by molecular replacement with the coordinates of Trypsin-BPTI as search model using PHASER[69]. The structure was refined by maximum-likelihood restrained refinement and TLS refinement[70] using PHENIX[71,72] followed by iterative, manual model building cycles

with COOT[73]. Model quality was evaluated with MolProbity[74]. Figures were prepared using Visual Molecular Dynamics[75]. The atomic coordinates and structure factor amplitudes have been deposited in the Protein Data Bank under the accession code 7PH1 (Trypsin BPTI-MfeGly).

V. ACKNOWLEDGEMENTS

Gefördert durch die Deutsche Forschungsgemeinschaft (DFG): Projektnummer 387284271 – SFB 1349, Projektnummer 434130070 - GRK 2662. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation):Project-ID 387284271 – SFB 1349, Project-ID 434130070 - IRTG 2662. We acknowledge access to high-performance computers via the Zentraleinrichtung für Datenverarbeitung (ZEDAT) of Freie Universität Berlin. We acknowledge access to beamlines of the BESSY II storage ring (Berlin, Germany) via the Joint Berlin MX-Laboratory sponsored by the Helmholtz Zentrum Berlin für Materialien und Energie, the Freie Universität Berlin, the Humboldt-Universität zu Berlin, the Max-Delbrück-Centrum, the Leibniz-Institut für Molekulare Pharmakologie and Charité – Universitätsmedizin Berlin.

VI. SUPPLEMENTARY MATERIAL

- Amended force field AMBER14sb (amber14sb_amended.ff.zip) and plots of the partial charges and Lennard-Jones parameters (SI.pdf)
- Structure and dynamics of water in the β-trypsin S1 binding pocket for unprotonated His57 (SI.pdf)
- Crystallographic data (crystallographic_data.docx and SI.pdb)
- Code repository: https://github.com/bkellerlab/track_waters

REFERENCES

[1]R. L. Mackman, B. A. Katz, J. G. Breitenbucher, H. C. Hui, E. Verner, C. Luong, L. Liu, and P. A. Sprengeler, "Exploiting subsite s1 of trypsin-like serine proteases for selectiv-

ity: potent and selective inhibitors of urokinase-type plasminogen activator," Journal of medicinal chemistry **44**, 3856–3871 (2001).

[2]R. Abel, N. K. Salam, J. Shelley, R. Farid, R. A. Friesner, and W. Sherman, "Contribution of explicit solvent effects to the binding affinity of small-molecule inhibitors in blood coagulation factor serine proteases," ChemMedChem **6**, 1049–1066 (2011).

[3]J. Schiebel, R. Gaspari, T. Wulsdorf, K. Ngo, C. Sohn, T. E. Schrader, A. Cavalli, A. Ostermann, A. Heine, and G. Klebe, "Intriguing role of water in protein-ligand binding studied by neutron crystallography on trypsin complexes," Nature Communications **9** (2018), 10.1038/s41467-018-05769-2.

[4]F. Leidner, N. Kurt Yilmaz, J. Paulsen, Y. A. Muller, and C. A. Schiffer, "Hydration structure and dynamics of inhibitor-bound hiv-1 protease," Journal of chemical theory and computation **14**, 2784–2796 (2018).

[5]A. Warshel and S. Russell, "Theoretical correlation of structure and energetics in the catalytic reaction of trypsin," Journal of the American Chemical Society **108**, 6569–6579 (1986).

[6]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).

[7]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).

[8]C. J. Farady and C. S. Craik, "Mechanisms of macromolecular protease inhibitors," Chembiochem **11**, 2341–2346 (2010).

[9]J. M. Berg, J. L. Tymoczko, G. J. Gatto jr., and L. Stryer, "Stryer biochemie," Springer Spektrum (2018).

[10]F. Markwardt, H. Landmann, and P. Walsmann, "Comparative studies on the inhibition of trypsin, plasmin, and thrombin by derivatives of benzylamine and benzamidine," European Journal of Biochemistry **6**, 502–506 (1968).

[11]M. Renatus, W. Bode, R. Huber, J. Stuerzebecher, and M. T. Stubbs, "Structural and functional analyses of benzamidine-based inhibitors in complex with trypsin: implications for the inhibition of factor xa, tpa, and urokinase," Journal of medicinal chemistry **41**, 5445–5456 (1998).

[12]N. Plattner and F. Noé, "Protein conformational plasticity and complex ligand-binding

kinetics explored by atomistic simulations and markov models," Nature communications **6**, 1–10 (2015).

[13]S. M. Gopal, F. Klumpers, C. Herrmann, and L. V. Schäfer, "Solvent effects on ligand binding to a serine protease," Physical Chemistry Chemical Physics **19**, 10753–10766 (2017).

[14]L. W. Votapka, B. R. Jagger, A. L. Heyneman, and R. E. Amaro, "Seekr: simulation enabled estimation of kinetic rates, a computational tool to estimate molecular kinetics and its application to trypsin–benzamidine binding," The Journal of Physical Chemistry B **121**, 3597–3606 (2017).

[15]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).

[16]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).

[17]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 β-trypsin and α-chymotrypsin," Biochemistry **35**, 11435–11446 (1996).

[18]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 **67**, 140–148 (2011).

[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]"Wikipedia entry of aprotinin," `https://en.wikipedia.org/wiki/Aprotinin`, accessed: 2022-06-07.

[21]D. Krowarsch, M. Dadlez, O. Buczek, I. Krokoszynska, A. O. Smalas, and J. Otlewski, "Interscaffolding additivity: binding of p1 variants of bovine pancreatic trypsin inhibitor to four serine proteases," Journal of molecular biology **289**, 175–186 (1999).

[22]R. Helland, J. Otlewski, O. Sundheim, M. Dadlez, and A. O. Smalås, "The crystal structures of the complexes between bovine β-trypsin and ten p1 variants of bpti," Journal of molecular biology **287**, 923–942 (1999).

[23]A. Grzesiak, R. Helland, A. O. Smalås, D. Krowarsch, M. Dadlez, and J. Otlewski, "Substitutions at the p1 position in bpti strongly affect the association energy with serine

proteinases," Journal of molecular biology **301**, 205–217 (2000).

[24]S. Huhmann and B. Koksch, "Fine-tuning the proteolytic stability of peptides with fluorinated amino acids," European Journal of Organic Chemistry **2018**, 3667–3679 (2018), https://chemistry-europe.onlinelibrary.wiley.com/doi/pdf/10.1002/ejoc.201800803.

[25]A. Vulpetti and C. Dalvit, "Fluorine local environment: from screening to drug design," Drug discovery today **17**, 890–897 (2012).

[26]E. P. Gillis, K. J. Eastman, M. D. Hill, D. J. Donnelly, and N. A. Meanwell, "Applications of fluorine in medicinal chemistry," Journal of medicinal chemistry **58**, 8315–8359 (2015).

[27]N. A. Meanwell, "Fluorine and fluorinated motifs in the design and application of bioisosteres for drug design," Journal of medicinal chemistry **61**, 5822–5880 (2018).

[28]H. Mei, J. Han, K. D. Klika, K. Izawa, T. Sato, N. A. Meanwell, and V. A. Soloshonok, "Applications of fluorine-containing amino acids for drug design," European journal of medicinal chemistry **186**, 111826 (2020).

[29]S. Ye, B. Loll, A. A. Berger, U. Mülow, C. Alings, M. C. Wahl, and B. Koksch, "Fluorine teams up with water to restore inhibitor activity to mutant BPTI," **6**, 5246–5254.

[30]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).

[31]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).

[32]J. R. Robalo and A. Vila 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).

[33]A. W. Sousa da Silva and W. F. Vranken, "Acpype - antechamber python parser interface," BMC Research Notes **5**, 367 (2012).

[34]C. I. Bayly, P. Cieplak, W. Cornell, and P. A. Kollman, "A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the resp model," The Journal of Physical Chemistry **97**, 10269–10280 (1993).

[35]Z. Cournia, B. Allen, and W. Sherman, "Relative binding free energy calculations in drug discovery: recent advances and practical considerations," Journal of chemical information and modeling **57**, 2911–2937 (2017).

[36]J. Smirnovienė, L. Baranauskienė, A. Zubrienė, and D. Matulis, "A standard operating

procedure for an enzymatic activity inhibition assay," European Biophysics Journal **50**, 345–352 (2021).

[37]J. Leppkes, T. Hohmann, and B. Koksch, "Improved enantioselective gram scale synthesis route to n-fmoc-protected monofluoroethylglycine," Journal of Fluorine Chemistry **232**, 109453 (2020).

[38]J. Leppkes, N. Dimos, B. Loll, T. Hohmann, M. Dyrks, A. Wieseke, B. G. Keller, and B. Koksch, "Fluorine-induced polarity increases inhibitory activity of bpti towards chymotrypsin," RSC Chemical Biology (2022).

[39]J. R. Robalo, D. Mendes de Oliveira, P. Imhof, D. Ben-Amotz, and A. Vila Verde, "Quantifying how step-wise fluorination tunes local solute hydrophobicity, hydration shell thermodynamics and the quantum mechanical contributions of solute–water interactions," Phys. Chem. Chem. Phys. **22**, 22997–23008 (2020).

[40]T. Lazaridis, "Inhomogeneous fluid approach to solvation thermodynamics. 1. Theory," Journal of Physical Chemistry B **102**, 3531–3541 (1998).

[41]K. Haider, A. Cruz, S. Ramsey, M. K. Gilson, and T. Kurtzman, "Solvation Structure and Thermodynamic Mapping (SSTMap): An Open-Source, Flexible Package for the Analysis of Water in Molecular Dynamics Trajectories," Journal of Chemical Theory and Computation **14**, 418–425 (2018).

[42]K. Haider, L. Wickstrom, S. Ramsey, M. K. Gilson, and T. Kurtzman, "Enthalpic Breakdown of Water Structure on Protein Active-Site Surfaces," Journal of Physical Chemistry B **120**, 8743–8756 (2016).

[43]C. N. Nguyen, T. Kurtzman Young, and M. K. Gilson, "Grid inhomogeneous solvation theory: Hydration structure and thermodynamics of the miniature receptor cucurbit[7]uril," Journal of Chemical Physics **137**, 973–980 (2012).

[44]"Wikipedia entry of histidine," `https://en.wikipedia.org/wiki/Histidine`, accessed: 2022-06-26.

[45]S. P. Edgcomb and K. P. Murphy, "Variability in the pka of histidine side-chains correlates with burial within proteins," Proteins: Structure, Function, and Bioinformatics **49**, 1–6 (2002).

[46]E. Mehler, M. Fuxreiter, I. Simon, and B. Garcia-Moreno E, "The role of hydrophobic microenvironments in modulating pka shifts in proteins," Proteins: Structure, Function, and Bioinformatics **48**, 283–292 (2002).

[47]J.-O. Joswig, J. Anders, H. Zhang, C. Rademacher, and B. G. Keller, "The molecular basis for the ph-dependent calcium affinity of the pattern recognition receptor langerin," Journal of Biological Chemistry **296** (2021).

[48]L. Donati and B. G. Keller, "Girsanov reweighting for metadynamics simulations," The Journal of chemical physics **149**, 072335 (2018).

[49]S. Kieninger, L. Donati, and B. G. Keller, "Dynamical reweighting methods for markov models," Current opinion in structural biology **61**, 124–131 (2020).

[50]L. Donati, M. Weber, and B. G. Keller, "Markov models from the square root approximation of the fokker–planck equation: calculating the grid-dependent flux," Journal of Physics: Condensed Matter **33**, 115902 (2021).

[51]M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, "Gaussian˜16 Revision C.01," (2016), gaussian Inc. Wallingford CT.

[52]J. Wang, W. Wang, P. A. Kollman, and D. A. Case, "Antechamber: an accessory software package for molecular mechanical calculations," J. Am. Chem. Soc **222**, U403 (2001).

[53]W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, "A second generation force field for the simulation of proteins, nucleic acids, and organic molecules," Journal of the American Chemical Society **117**, 5179–5197 (1995).

[54]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).

[55]G. Bussi, D. Donadio, and M. Parrinello, "Canonical sampling through velocity rescaling,"

The Journal of Chemical Physics **126**, 14101 (2007).

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

[57]E. Lindahl, B. Hess, and D. van der Spoel, "GROMACS 3.0: a package for molecular simulation and trajectory analysis," Molecular modeling annual **7**, 306–317 (2001).

[58]S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess, and E. Lindahl, "GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit," Bioinformatics **29**, 845–854 (2013).

[59]S. Páll, M. J. Abraham, C. Kutzner, B. Hess, and E. Lindahl, "Tackling Exascale Software Challenges in Molecular Dynamics Simulations with GROMACS BT - Solving Software Challenges for Exascale," (Springer International Publishing, Cham, 2015) pp. 3–27.

[60]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).

[61]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).

[62]S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, "The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method," Journal of Computational Chemistry **13**, 1011–1021 (1992).

[63]P. Wernet, 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), https://www.science.org/doi/pdf/10.1126/science.1096205.

[64]P. Liu, E. Harder, and B. J. Berne, "On the calculation of diffusion coefficients in confined fluids and interfaces with an application to the liquid-vapor interface of water," Journal of Physical Chemistry B **108**, 6595–6602 (2004), arXiv:0311084 [physics].

[65]L. Rodney, J. Gregory, C. Dimitrios, and Z. Adam, "LA-UR-19-32421 Title : Intended for : Issued :," U.S. Department of Energy, Office ofEnergy Efficiency and Renewable Energy **13**, 1–30 (2019).

[66]N. Michaud-Agrawal, E. J. Denning, T. B. Woolf, and O. Beckstein, "MDAnalysis: A toolkit for the analysis of molecular dynamics simulations," Journal of Computational Chemistry **32**, 2319–2327 (2011).

[67]C. Yung-Chi and W. H. Prusoff, "Relationship between the inhibition constant (ki) and the concentration of inhibitor which causes 50 per cent inhibition (i50) of an enzymatic reaction," Biochemical Pharmacology **22**, 3099–3108 (1973).

[68]W. Kabsch, "Integration, scaling, space-group assignment and post-refinement," Acta Crystallographica Section D **66** (2010).

[69]A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, and R. J. Read, "*Phaser* crystallographic software," Journal of Applied Crystallography **40**, 658–674 (2007).

[70]M. D. Winn, M. N. Isupov, and G. N. Murshudov, "Use of TLS parameters to model anisotropic displacements in macromolecular refinement," Acta Crystallographica Section D **57**, 122–133 (2001).

[71]P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, *et al.*, "Phenix: a comprehensive python-based system for macromolecular structure solution," Acta Crystallographica Section D: Biological Crystallography **66**, 213–221 (2010).

[72]P. V. Afonine, R. W. Grosse-Kunstleve, N. Echols, J. J. Headd, N. W. Moriarty, M. Mustyakimov, T. C. Terwilliger, A. Urzhumtsev, P. H. Zwart, and P. D. Adams, "Towards automated crystallographic structure refinement with phenix. refine," Acta Crystallographica Section D: Biological Crystallography **68**, 352–367 (2012).

[73]A. Casañal, B. Lohkamp, and P. Emsley, "Current developments in coot for macromolecular model building of electron cryo-microscopy and crystallographic data," Protein Science **29**, 1055–1064 (2020).

[74]C. J. Williams, J. J. Headd, N. W. Moriarty, M. G. Prisant, L. L. Videau, L. N. Deis, V. Verma, D. A. Keedy, B. J. Hintze, V. B. Chen, S. Jain, S. M. Lewis, W. B. Arendall III, J. Snoeyink, P. D. Adams, S. C. Lovell, J. S. Richardson, and D. C. Richardson, "Molprobity: More and better reference data for improved all-atom structure validation," Protein Science **27**, 293–315 (2018).

[75]W. Humphrey, A. Dalke, and K. Schulten, "VMD – Visual Molecular Dynamics," Journal of Molecular Graphics **14**, 33–38 (1996).
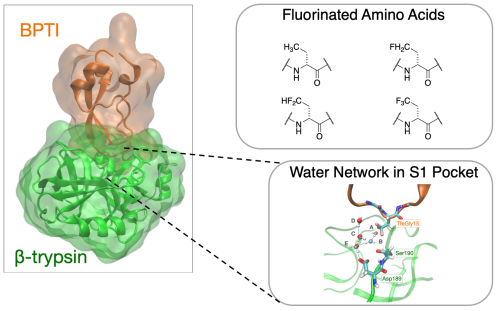
TABLE OF CONTENT FIGURE