

Exploring aspartic protease inhibitor binding to design selective antimalarials

Raitis Bobrovs*, Emils Edgars Basens, Laura Drunka, Iveta Kanepe, Sofija Matisone, Karlis Kristofers Velins, Victor Andrianov, Gundars Leitis, Diana Zelencova-Gopejenko, Dace Rasina, Aigars Jirgensons and Kristaps Jaudzems

Latvian Institute of Organic Synthesis, Aizkraukles 21, Riga, LV1006, Latvia
E-mail: raitis.bobrovs@osi.lv

Selectivity is a major issue in the development of drugs targeting pathogen aspartic proteases. Here we explore the selectivity determining factors by studying specifically designed malaria aspartic protease (plasmepsin) open-flap inhibitors. Metadynamics simulations are used to uncover the complex binding/unbinding pathways of these inhibitors, and describe the critical transition states in atomistic resolution. The simulation results are compared to experimentally determined enzymatic activities. Our findings demonstrate that plasmepsin inhibitor selectivity can be achieved by targeting the flap loop with hydrophobic substituents that enable ligand binding under the flap loop, as such behaviour is not observed for several other aspartic proteases. The ability to estimate compound selectivity before they are synthesized is of great importance in drug design, therefore, we expect that our approach will be useful in selective inhibitor design not only against aspartic proteases, but other enzyme classes as well.

Introduction

Aspartic proteases are wide spread in eukaryotes, and they are attracting attention as drug targets^[1,2]. This family of enzymes uses an activated water molecule bound to the catalytic dyad – two aspartic acid residues – to attack the scissile peptide bond in a substrate. The aspartic protease catalysed peptide bond cleavage can be prevented by inhibitors that replace this catalytic water molecule or in which the scissile peptide bond is replaced by a non-cleavable moiety. There are numerous inhibitors reported in literature targeting aspartic proteases involved in HIV/AIDS, malaria, Alzheimer's disease, etc^[3,4]. Over the recent years, aspartic proteases from the malaria parasite – plasmepsins (plms) – have gained particular interest from medicinal chemists^[5–13] due to increasing resistance of known antimalarials^[14,15].

An important issue when developing pathogen aspartic protease inhibitors is their selectivity over human enzymes. Plms share high sequence similarity with several human aspartic proteases, the most similar being the lysosomal enzyme cathepsin D (catD). Therefore, it is commonly used as a marker for cross-inhibition. Plm II and catD share 35% sequence identity, and even higher similarity is observed in the active site. Due to the high protein structural similarity, design of potent plasmepsin inhibitors that do not inhibit catD remains a challenging task.

Recent studies^[16] have shown that plm inhibitor selectivity over human aspartic proteases can be achieved by targeting the flap loop region of plms. Flap loop is a single long β -hairpin structure that lies perpendicularly over the aspartic dyad (see

Figure 1.A), and is believed to be involved in the substrate recognition^[17–20]. There have been numerous, mostly computational, studies focused on the characterisation of aspartic protease flap loop dynamics^[18,19,21–32]. It is commonly acknowledged^[25], that the flap-loop of aspartic proteases occupies a semi-open conformation in *apo* state. However, the hinge like motion of the flap loop is also capable of exposing a hydrophobic pocket (see

Figure 1.B) under the flap loop producing so called open-flap conformation^[33–37]. Inhibitors that bind partly under the flap loop (also known as open-flap inhibitors) have improved selectivity over human aspartic proteases, and the selectivity gain is attributed to differences in enzyme flexibility – plms easily reach open-flap conformation where an additional hydrophobic pocket is formed, whereas cathepsins either do not have or do not easily reach an open-flap conformation. The precise mechanism of enzyme reorganisation for the inhibitor binding to open-flap conformation, however, is not yet known. It is unclear what are the protein structural features and key protein-inhibitor interactions that enable flap loop opening, and lead to inhibitor binding under the flap loop. Such information would allow us to design inhibitors with improved selectivity not only against plms, but other aspartic proteases, and, possibly, other enzymes with highly mobile loops near the binding site.

An accurate description of the flap loop motion upon ligand binding at a molecular scale is essential to understand selectivity determining factors of open-flap inhibitors. Currently, however, there are no experimental methods that would enable such dynamic process characterisation at a molecular scale. Lately, *in-silico* techniques have emerged as a viable alternative to experimental methods and are becoming

increasingly effective in providing atomic-level description of protein–ligand interactions^[38–47]. Dynamic processes like these are typically studied using molecular dynamics (MD) simulations, however, ligand binding/unbinding is a rare event that can be hard to sample within a reasonable time^[48,49]. Therefore, enhanced sampling methods like metadynamics (MetaD) are utilised to accelerate the occurrence of rare events by depositing a history dependent biasing potential to selected degrees of freedom (also known as collective variables (CVs))^[39,40,50–56]. The bias potential within CV sub-space is built up by adding Gaussians along the sampled trajectory, and bias potential applied encourages the system to visit higher energy states and discourage the system from revisiting configurations already sampled. This approach leads to a significantly improved exploration of the simulation space, and, once the exploration is complete, an efficient reconstruction of the free energy surface (FES) as a function of the CVs used^[57,58].

MetaD in its various modifications has been applied in numerous ligand-protein binding studies^[38–47]. One of these MetaD variations is funnel metadynamics (FMetaD), where a funnel shaped restraint potential is applied to the system to reduce the space explored by the ligand once it is unbound, thus reducing computational cost even more^[46,59–63]. In FMetaD, the bias potential is typically applied to the distance between the ligand and the binding site, thus, sampling multiple binding/unbinding events within simulation. FMetaD allows identification of the ligand binding mode, absolute protein-ligand binding free energy, insight into the dynamics of the ligand binding, including the presence of alternative ligand binding modes and other energetically relevant states (e.g., transition states). Furthermore, it has been demonstrated that FMetaD can be used to study ligand binding process in challenging systems where binding site is buried or protein undergoes significant conformational changes upon binding^[60,64]. FMetaD has demonstrated to be able to predict ligand binding free energy with <1.5 kcal/mol accuracy^[60,61]. It is important to note, however, that FMetaD is an MD-based protocol and the correctness of the results depend on the accuracy of the protein and ligand force-field parameters, as well as system amino acid and ligand protonation state selection.

By examining known non-peptidomimetic plm inhibitors, we previously hypothesized^[16], that one of the key features enabling the ligand binding to the enzyme in open flap conformation might be hydrophobic substituent capable of aromatic interactions with Trp41, Tyr77 and/or Phe111 (plm II numbering) under the flap loop. To test this hypothesis, we designed and synthesized a series of 2-aminoquinazolin-4(3*H*)-one based^[65–67] inhibitors (see Figure 2) with hydrophobic substituents bearing benzene group at various locations, determined their activity against plm II, IV, V and cat D, and explored their binding pathway using funnel metadynamics simulations. This showed that the inhibitor binding mode can be estimated using metadynamics simulations, and information obtained can be used in selective inhibitor design.

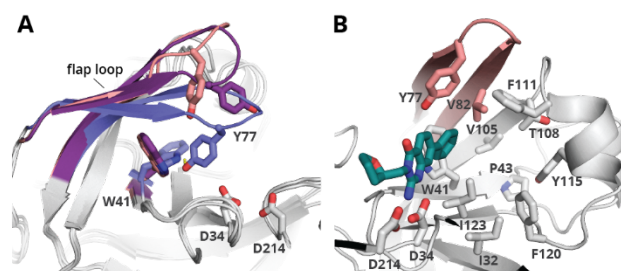


Figure 1. A Plm II flap loop conformations: blue – closed flap conformation in complex with peptidomimetic inhibitor (PDB ID 1LF4), purple and salmon – open flap conformation in complex with non-peptidomimetic inhibitors (PDB IDs 2BJU and 4Z22, respectively). Residues W41 and Y77, believed to play crucial role in flap loop dynamics, and catalytic dyad (D34 and D214) are shown as sticks (plm II numbering). Note that W41 changes its conformation upon flap opening/closing. B Hydrophobic flap pocket exposed upon flap loop opening (PDB ID 4Z22). Ligand is shown as green sticks, hydrophobic flap pocket residues as white sticks. Flap loop is in salmon. Hydrogens are omitted for clarity.

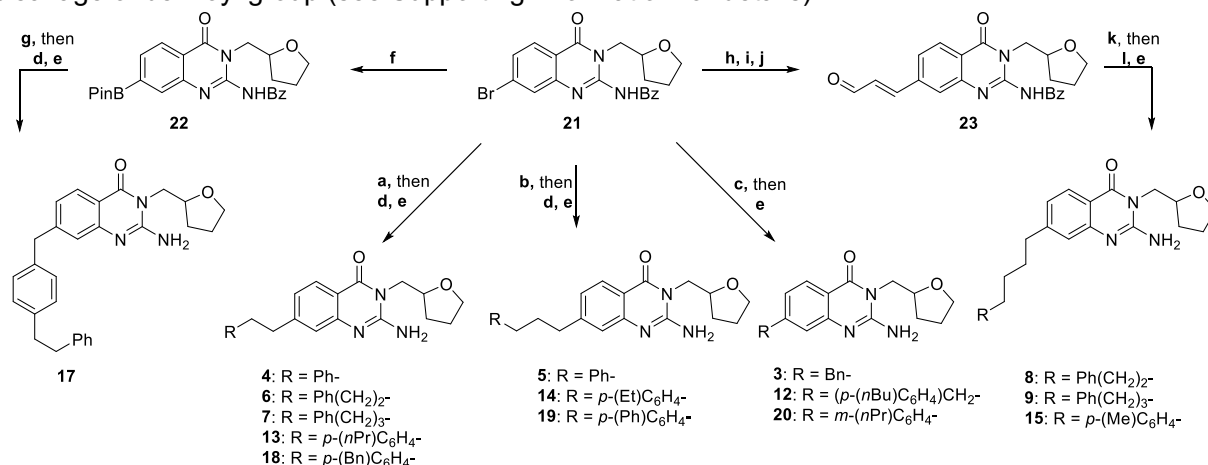
Results and Discussion

Compound design

The 2-aminoquinazolin-4(3*H*)-one derivatives were designed to have linear, hydrophobic flap pocket substituents of various alkyl chain length, benzene ring position and benzene ring count (see Figure 2). The initial compound subset varied in length of the linear alkyl linker between the ligand core and terminal benzene ring (**2-9**). Once the optimal flap pocket substituent length was identified, another ligand subset, where the substituent length was kept constant while the position of the benzene ring was varied, was synthesized (**11-15**). To evaluate the contribution of the terminal benzene ring on the inhibitor activity, compound **10**, bearing linear alkyl chain of the same length as the most active compound was synthesized. Moreover, several compounds bearing two benzene rings at various positions were synthesized to evaluate the impact of additional benzene ring in the flap pocket substituent (**16-19**). For comparison purposes, compound without flap pocket substituent (**1**) – the original fragment hit identified in our screening campaign^[67] – was prepared and characterised as well.

Chemistry

The synthesis of compounds **1**, **2**, **10**, **11**, **16** and **21** has been previously described^[65,67]. Compounds **3–9**, **12–15** and **17–20** were synthesized as shown in Scheme 1. Building block **21** was subjected to Sonogashira reaction with alkynes **24a–e**, followed by the saturation of alkyne and the cleavage of benzoyl group to obtain products **4**, **6**, **7**, **13** and **18**. Suzuki-Miyaura reaction of **21** with the corresponding vinyl dioxaborolanes **25a–c**, or aryl or alkyl dioxaborolanes **26a–c** with subsequent saturation of unsaturated bonds and the cleavage of benzoyl group gave products **5**, **14**, **19** and **3**, **12**, **20**. The target compound **17** was synthesized from pinacol borane derivative **22** and benzylbromide **27** using Suzuki-Miyaura type reaction with subsequent hydrogenation of alkyne moiety and benzoyl group cleavage. Coupling of **21** with propargyl alcohol (**28**) followed by oxidation of alcohol to aldehyde and reduction of alkyne to alkene provided **23**. Building block **23** was further converted into products **8**, **9** and **15** using the Wittig reaction, the saturation of double bonds and the cleavage of benzoyl group (see Supporting Information for details).



Scheme 1. Synthesis of 2-Aminoquinazolin-4(3H)-one Derivatives: a) RCCH **24a–e**, Pd(PPh₃)₄, Et₃N, CuI, DMF, rt; b) RCH₂CHCHBPin **25a–c**, Pd(PPh₃)₄, K₂CO₃, THF, H₂O, 80 °C; c) RBPIn **26a–c**, Pd(PPh₃)₄, K₂CO₃, dioxane, H₂O, 80 °C; d) H₂ (2 – 4 bar), Pd/C, THF, EtOH or CF₃CH₂OH, rt; e) N₂H₂·H₂O, EtOH, THF, 70 °C; f) (PinB)₂, Pd(dppf)Cl₂·CH₂Cl₂, NaOAc, dioxane, 80 °C; g) 1-(bromomethyl)-4-(phenylethynyl)benzene (**27**), Pd(dppf)Cl₂·CH₂Cl₂, Cs₂CO₃, dioxane, H₂O, 80 °C; h) propargyl alcohol (**28**), Pd(PPh₃)₄, Et₃N, CuI, neat, 70 °C; i) Mn₂O, acetone, rt; j) PPh₃, H₂O, THF, rt; k) RCH₂P(Ph)₃Br **29a–c**, NaH, DMSO, rt; l) H₂ (2 – 4 bar), Pd(OH)₂/C, MeOH.

Enzymatic activity

Compounds synthesized were tested for their inhibitory activity against digestive vacuole plms II and IV, endoplasmic reticulum protease plm V, and selectivity cross-marker – human catD (see SI for experimental details). The plm II inhibition potency (IC₅₀) of the compounds tested varied in a range from 3.5 μM down to 64 nM depending on the length of hydrophobic substituent and position of aromatic ring (see Figure 2 and Table S11 for full inhibition data). There were minor variations between plm II and plm IV inhibition potencies, whereas the potency against plm V and human catD was considerably lower. The catD inhibition potency followed plms II and IV inhibition pattern with 10 to 70 fold lower activities (55 to 2 μM), whereas potency against plm V varied in a narrow range (90 to 25 μM) and size of the flap pocket substituent seems to play no role in inhibitor activity against plm V.

More detailed analysis of structure activity relationships (SAR) of the inhibitor series indicates that installation of the flap pocket substituent greatly increases the inhibitor potency (15-fold activity increase from compound **1** to **2**). By increasing the distance between the ligand core (2-aminoquinazolin-4(3H)-one group) and the terminal aromatic ring, the inhibitory activity initially decreases, reaching lowest activity for inhibitors with methylene and ethylene linker (**3**, **4**), whereas further increase of linker length led to increase in inhibition potency. The highest potency was achieved for the compound bearing phenylpentyl group (**7**). Linker length increase beyond pentylene group led to reduced activity (**8**, **9**). The inhibition potency of the compound **10** suggests that presence of the terminal aromatic group in the flap pocket substituent is not necessary for high activity, as inhibition potency of both compounds (**7** and **10**) was comparable.

A ligand subset (**11–15**) bearing the linkers with the same overall length as compound **7**, but varied in positions of aromatic ring was tested in order to understand, whether the activity of the inhibitor is affected solely by the length of the hydrophobic flap pocket substituent or position of the aromatic ring. The general inhibitory activity trend was the same as in initial inhibitor set, however, these inhibitors were 5 to 20 times more potent. The last ligand subset had terminal benzene ring fixed at a specific distance, and an additional aromatic ring was installed in the flap pocket substituent to introduce stacking interactions with Trp41, Tyr77, Phe111, Tyr115 and/or Phe120. Inhibitors bearing two benzene rings in the flap pocket substituent showed low micromolar/nanomolar potency, however, no clear SAR was observed for these compounds.

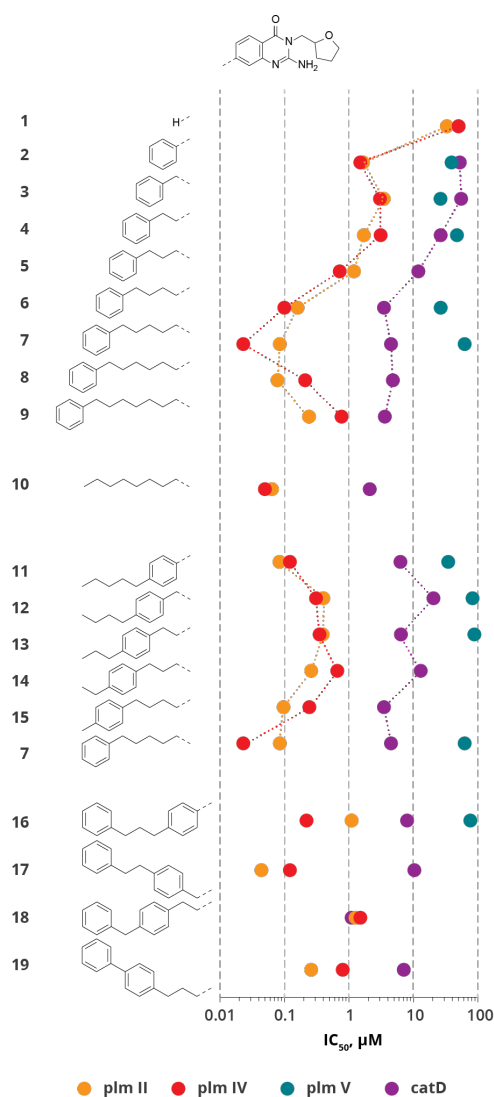


Figure 2. The 2-aminoquinazolin-4(3H)-one derivatives synthesized, their enzymatic potency (IC_{50}) against plm II, IV, V and catD, and compound numbering. Dashed line is added as a guide to the eye. IC_{50} data are given in logarithmic scale.

Metadynamics simulations

Schematic representation of the funnel shaped restraint potential used in our FMetaD simulations is given in Figure 3.A. The restraint potential was placed in a way that does not interfere with ligand binding and unbinding events, and its dimensions were chosen as small as possible to accelerate convergence, but large enough to avoid artefacts in the sampling. The funnel defining parameters are summarised in Figure 3.A. Here, the CV controlling the ligand binding/unbinding was the distance (d1) between the centre of mass of ligand transition state mimetic atoms and the centre of mass of the aspartic dyad Asp34 and Asp214 C γ atoms (see Figure 3.B,C). Distance CV easily discriminates between bound and unbound states, however, trajectory obtained can be reweighed to remap the free energy as a function of alternative CVs and gain more detailed insights, revealing bound minima, which are overlying in the maps corresponding to the initial CV choice^[58,68].

Several unbiased and exploratory MetaD simulations were first performed to get an idea of the preferred path of unbinding with statistical significance. During our preliminary simulations we observed that flap loop does not close completely upon ligand unbinding, and rotation of the Trp41, which is known to change conformation when bound to open-flap inhibitors^[37,67] (see

Figure 1.A), was not observed. To overcome this obstacle, we introduced PathCV^[69] that allows to study transitions between two or more states. Here we constructed the flap loop opening/closing pathway (including Trp41 rotation) by running trial Path MetaD simulations using X-ray crystal structures of closed and open flap conformations^[67,70] as end states (PDB IDs 1LF4 and 4Z22, respectively; see Figure 3.D,E). Equally spaced loop opening/closing pathway was constructed from the preliminary simulations and was further used in ligand binding/unbinding studies. Additionally, soft harmonic restraining walls were applied at both ends of the funnel axis to limit the exploration of the ligand inside the region enclosed by the funnel potential.

The overall structure of the protein was preserved throughout simulations, during which the loops and the binding site residues showed the necessary flexibility to allow binding and unbinding of the open-flap inhibitors.

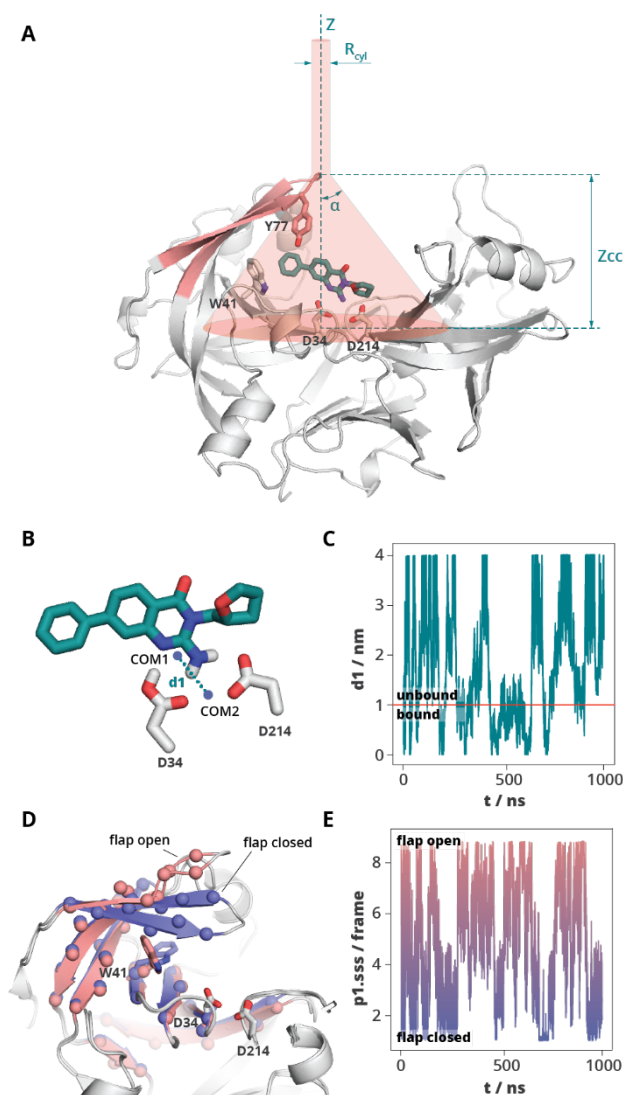


Figure 3. **A** Schematic representation of a funnel shaped restraint potential applied in funnel metadynamics simulations. R_{cyl} was set to 0.1 nm and funnel potential was limited to 4 nm from the origin (C α of Asp34) by a wall potential. The cone region was defined by a vertex height Z_{cc} of 2.0 nm from the origin and angle α was set to 0.6 rad. **B** Distance ($d1$) between the centre of mass of ligand transition state mimetic (COM1) and centre of mass of the C γ atoms of catalytic dyad residues (Asp34 and 214, COM2) was used as a CV in funnel metadynamics simulations. **C** The ligand-protein distance $d1$ as a function of the funnel metadynamics simulation time in the case of comp. 1-plm II system. Several binding and unbinding events are observed in given simulation. Red line indicates bound and unbound state threshold. **D** Borderline conformations used in metadynamic simulations to control flap loop opening and closing (purple – flap loop closed, salmon – flap loop open; PDB IDs 1FL4 and 4Z22, respectively). Atoms used in PathCV construction – C α atoms of residues involved in beta sheet secondary structures within 1.5 nm from Trp41 – are shown as spheres. Since Trp41 is known to change its conformation upon flap opening, all Trp41 heavy atoms were also added to PathCV. **E** The flap loop conformation along the flap loop opening/closing pathway as a function of the funnel metadynamics simulation time in the case of comp. 1-plm II system. Colour corresponds to the states given in subfigure D: purple – flap loop closed, salmon – flap loop open.

Binding to plm II. The performed metadynamics simulations directly provide an estimate of where the system spends most of its time^[52,55]. The FES computed as a function of distance $d1$ and torsion ϕ (ligand alignment with respect to flap pocket axis, see Figure S11) indicate that there are several stabilised transition states along the ligand binding pathway (see Figure 4). Most of these states are observed for all of the plm inhibitors synthesized and tested. The deepest basin **A** in all simulations corresponds to the state where inhibitor is bound in a pose analogue to that found in crystal structure (PDB ID: 4Z22). The 2-aminoquinazolin-4(3*H*)-one core that mimics the transition state is involved in hydrogen bonding with Asp34 and Asp214, adjacent ligand core aromatic ring is interacting with the flap loop Tyr77 sidechain, whereas flap pocket substituent is buried under the flap loop. The dominant interactions under the flap loop depend on the flap pocket substituent length, with longest substituents interacting with Val105, Phe111 and/or Tyr115 sidechains deeper in the flap pocket, while shorter inhibitor substituents are interacting with Trp41, Phe111 and/or Phe120 sidechains. In all cases, ligand binding have disrupted the hydrogen bonding between Trp41 and Tyr77, which is common for aspartic proteases in closed flap conformation^[16,71]. Once this interaction is not present, enzyme flap loop is able to open-up and Trp41 is able to rotate to expose the hydrophobic depth of the hydrophobic pocket.

This behaviour is observed upon binding of all inhibitors, even those with short flap pocket substituents. It is worth noting that Tyr77 stacking interactions with inhibitor core in basin **A** were observed on both sides of the inhibitor (e.g. inhibitor locked in-between Tyr77 and Phe111, or inhibitor interacting with Tyr77 on one side and solvent exposed on the other).

Basin **B** corresponds to the binding mode where ligand core (2-aminoquinazolin-4(3*H*)-one group) is occupying pose similar to that of crystal structure, while the flap pocket substituent is not buried under the flap loop. Flap pocket substituent here is interacting with Tyr77 and Trp41, however, it is situated on the other side of these residues and is partly solvent exposed. Since this binding mode is observed when the flap loop is in semi-open conformation, we believe that this might be one of the transition states in the ligand binding and flap loop opening pathway. More specifically, the flap loop opening is initiated by the inhibitor binding in the substrate binding groove, and is followed by the inhibitor rearrangement where hydrophobic flap pocket substituent is guided underneath the flap loop while maintaining interactions with aspartic dyad. During this flap opening motion the common aspartic proteases hydrogen bonding network is disrupted and Trp41 rotates to expose the terminal end of the hydrophobic flap pocket and engage in aromatic stacking interactions with flap pocket substituent.

Besides previously described poses, where ligand core is interacting with catalytic dyad, two additional binding modes were identified for majority of the inhibitors. One, where aromatic group of the hydrophobic flap pocket substituent is wedged in-between Tyr77 and Phe111/Tyr115 while ligand core is solvent exposed (basin **C** in Figure 4), and another one, where the inhibitor is interacting with aromatic residues Phe11, Tyr17 in the binding groove (basin **D** in Figure 4). Ligand-protein interactions observed in basin **D** take place ~2 nm away from the aspartic dyad, therefore, it is unlikely that they play crucial role in inhibitor recognition and inhibitor binding, however, these intermediate states might play role in ligand guiding to the binding site.

Basin **C** corresponds to the state where inhibitor is situated closer to the binding site, however, does not interact directly with the aspartic dyad. Interactions here are dominated by the hydrophobic stacking between the inhibitor terminal benzene ring and Tyr77, Phe111 and/or Tyr115 sidechains in flap loop and α helix right next to the loop. These interactions are mostly observed when the flap loop is in closed conformation, and just like complex in basin **B**, could be a transition state on the flap loop opening pathway. Here flap loop opening is initiated by the hydrophobic interactions that direct the flap pocket substituent under the flap loop by successively interacting with aromatic residues Phe120, Tyr115, Phe111, Tyr77 and finally Trp41, while the transition state mimetic remains solvent exposed.

The presence of alternative binding modes observed in FMetaD simulations is supported by the X-ray diffraction data. Attempts were made to co-crystallise plm II and compounds synthesized (see SI for details), however, resolution for the majority of the crystals obtained were suboptimal. The only crystal that was solved was with compound **7** (PDB ID: 7QYH, to be released upon publication). Due to poorly defined electron density in the plm II binding site, it was not possible to solve the position of compound **7**, however, unmodelled electron density in the binding site indicates ambiguous flap pocket substituent binding (see Figure 5). Electron density map suggest that most probable phenylpentyl substituent placement is under the flap loop on either side of the Trp41, which matches FES basins **A** and **B** observed in simulations.

The states described above are observed for all inhibitors, however, there are some notable differences between inhibitors. The FES of the 2-aminoquinazolin-4(3*H*)-one without flap pocket substituent (**1**) is the smoothest one, indicating that the inhibitor binding energy is low and energetic penalty associated with the transition from one binding mode to another one is also low (~2 kcal/mol). Several energetically similar binding modes with negligible transition energies mean that the inhibitor is highly mobile and fluctuates between multiple competing binding modes, and, as a result, its inhibition potency is low. Introduction of more extensive flap pocket substituent reduces the accessible binding space and shifts the binding preference to the binding mode observed in the crystal structure. This results in improved inhibitor potency, since role of the alternative binding modes is diminished. This is in line with experimentally determined activity data, where inhibition potency increases as flap pocket substituent length increases, however, the impact of the flap pocket length increase is not linear. We believe that lower activity for compounds bearing phenylmethyl- and phenylethyl-substituents (**2**, **3**) is a result of sub-optimal terminal benzene ring placement in the flap pocket due to competitive stacking interactions. More specifically, shorter linker length prevents simultaneous ligand core stacking interactions with Tyr77 and/or Phe111, and terminal benzene ring interactions with Trp41 and/or Tyr115. Thus, compounds bearing methylene and ethylene linkers are able to form optimal interactions either with catalytic dyad and Tyr77/Phe111, or with Trp41/Tyr115, but not both. In addition, inhibitors with shorter flap pocket substituents bind to the open flap conformation, however, they are not able to fill the pocket completely, thus, leading to partly solvated hydrophobic pocket and/or introducing thermodynamic penalty due to partly filled cavity. The increase of the linker length enables simultaneous interactions for both groups, reaching optimal positioning with pentylene linker (comp. **7**). Moreover, the phenylpentyl substituent size is optimal to fill the flap pocket completely. Flap pocket substituents bearing longer alkyl chain linker can fit in the cavity (**8**, **9**), however, substituents that are too long in their fully extended conformation are expected to adopt energetically unfavourable conformations with internal *gauche* strains, leading to reduced potency^[72].

The potency of the inhibitors with fixed flap pocket length and various benzene ring position (**11-15**, **7**) indicate that the flap pocket depth filling improves the inhibitor activity to sub-micromolar level. The fact that SAR with respect to benzene ring position for this inhibitor subset was identical to initial inhibitor subset indicate that

aromatic stacking and/or steric effects provided by the benzene ring play the same, decisive role in inhibitor activity. The additional terminal *n*-alkyl substituent for this inhibitor series improves the overall activity with respect to corresponding shorter substituents by filling the depth of the flap pocket, thus, reducing the thermodynamic penalty due to unfilled and/or partly solvated hydrophobic pocket^[72].

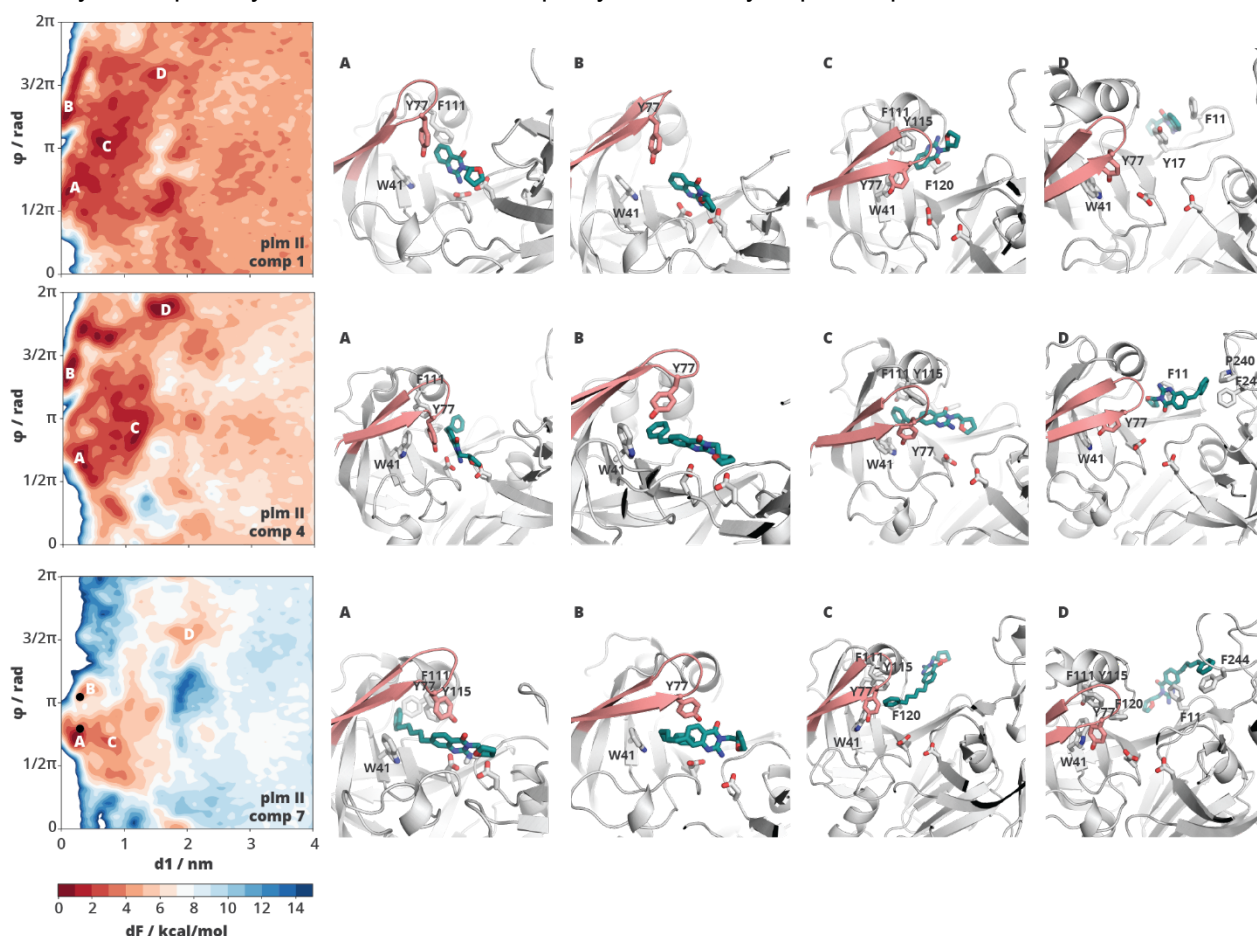


Figure 4. The FES of compounds **1**, **4** and **7** binding to plm II obtained using FMetaD simulations and reweighed^[58] as a function of the ligand core-catalytic dyad distance $d1$ (see **Figure 3.B**) and a torsion ϕ (ligand alignment with respect to flap pocket axis, see **Figure S11**). Isosurfaces are shown every 1 kcal/mol. Deepest FES basins are indicated as A, B, C and D, and corresponding binding modes are shown on the right. Binding mode in basin A matches the pose observed in the crystal structure 4Z22, whereas binding modes B, C and D are transition states on the ligand binding pathway. The binding modes estimated from the un-modelled electron density in the plm II-compound **7** co-crystal are indicated with black dots. Ligand is shown as green sticks, selected binding pocket residues and catalytic dyad residues as white sticks. Flap loop is in salmon. Hydrogens are omitted for clarity.

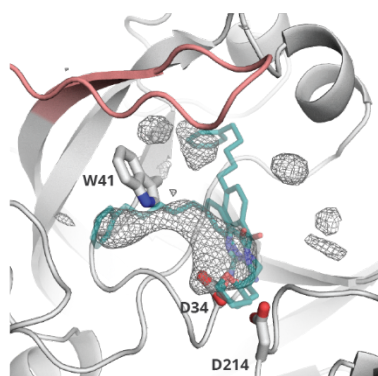


Figure 5. Electron density map (mFo-DFc at 3 σ) in the binding site of plm II and compound **7** co-crystal. Two alternative docked binding modes of compound **7** are shown (it was not possible to solve inhibitor position due to poor resolution). Ligand is shown as green sticks, selected binding pocket residue and catalytic dyad residues as white sticks. Flap loop is in salmon. Hydrogens are omitted for clarity.

Poor activity against catD and plm V. The FES of 2-aminoquinazolin-4(3*H*)-ones binding to plm V (Figure 6 top) and catD (Figure 6, bottom) are slightly different from those of plm II binding. The majority of the binding modes observed during the inhibitor binding to plm II are present also on the catD and plm V binding pathway, however, there is one noticeable difference – binding under the flap loop is negligible. For both enzymes, the energetically favoured binding mode is similar to that observed in basin **B** for plm II, where transition state mimetic group is interacting with aspartic dyad while the flap pocket substituent is located between the flap loop and unstructured region right next to it (see compound **7** binding modes **B** in Figure 6). Here hydrophobic flap pocket substituent is interacting with Phe74, Ile76 and/or Val144 of catD, and Tyr286, Tyr288 and/or Leu311 of plm V. Another, similarly favoured binding mode corresponds to the inhibitor bound in a binding groove in the opposite direction, where hydrophobic part of the inhibitor interacts with Phe126 and/or Phe131 of catD, and Leu179 and/or Tyr182 of plm V (see Figure 6, binding modes **A**). Despite plm V highly flexible flap loop tip (since Tyr139 is not involved in direct hydrogen bonding with protein core β sheet motif amino acids like plm II and catD, see Figure 7) inhibitor binding in a pose where flap pocket substituent is fully buried under the flap loop was not observed. Minor interactions under the plm V flap loop tip (basin **C**) were similar to those observed in basin **C** for plm II, where hydrophobic flap pocket substituent was stacked in-between Tyr139 and Phe180 while the transition state mimetic was solvent exposed. The rigidity at the plm V flap loop hinge region does not allow further flap opening, and as such inhibitor remains in non-optimal conformation. Inhibitor interactions under the catD flap loop tip were similar – terminal benzene ring interacted with Tyr78 and Phe126, however it did not lead to inhibitor binding under the flap loop. Similar binding behaviour was observed for all inhibitors.

Taking FMetaD simulation data into account, it is safe to assume, that poor plm V and catD inhibition is a result of inhibitor inability to bind under the flap loop, and enzyme inhibition is achieved by inhibitor binding in an alternative pose. These alternative binding modes, however, lack optimal (coplanar) 2-aminoquinazolin-4(3*H*)-one core placement (due to flap pocket substituent clashing with the plm V and catD flap loops) and, thus, ligand activity is reduced.

We believe that structural feature that plays crucial role in inhibitor binding under the flap loop is hydrophobic amino acid region at the flap pocket “entrance” between the flap loop and adjacent α helix. All enzymes have phenylalanine situated at the same place right next to the flap loop tip (Phe111 for plm II, Phe180 for plm V and Phe126 for catD, see Figure 8), which in unbound state can interact with flap loop tip Tyr and approaching inhibitor. Such transition states were observed in our simulations, and correspond to previously described states in basin **C**. Once initial interactions are established, the hydrophobic inhibitor substituent is then guided deeper in the flap pocket to interact with Tyr115 and Trp41 of plm II. In plm II, the Tyr115 sidechain is rigidly locked in the position by hydrogen bonding with Asp121 main chain, whereas Trp41 is known to change the conformation upon non-peptidomimetic inhibitor binding/flap opening. The flap pocket residue rearrangement results in a conformation where hydrophobic flap pocket substituent is locked between Tyr115 and Trp41. The hydrophobic flap pocket remains un-solvated during the flap pocket filling step, and required flap opening is aided by gradual flap pocket filling with hydrophobic inhibitor tail. In plm V and catD, the position of the Tyr115 sidechain is occupied by Glu176 and Gln121, respectively, and, therefore, hydrophobic interactions with the inhibitor are discouraged. This prevents ligand flap pocket substituent guiding deeper under the flap loop and efficient inhibitor binding.

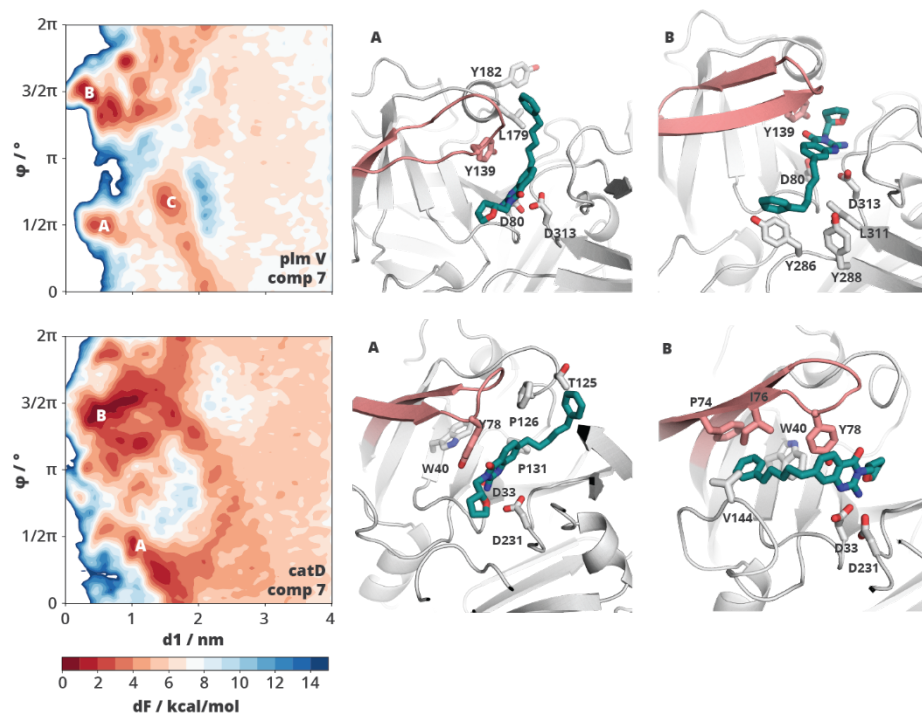


Figure 6. The FES of compound **7** binding to plm V (top) and catD (bottom) obtained using FMetaD simulations and reweighed^[58] as a function of the ligand core-catalytic dyad distance $d1$ and a torsion ϕ . Isosurfaces are shown every 1 kcal/mol. Deepest FES basins for each system are indicated as A and B, and corresponding binding modes are shown on the right. Ligand is shown as green sticks, selected binding pocket residues and catalytic dyad residues as white sticks. Flap loop is in salmon. Hydrogens are omitted for clarity.

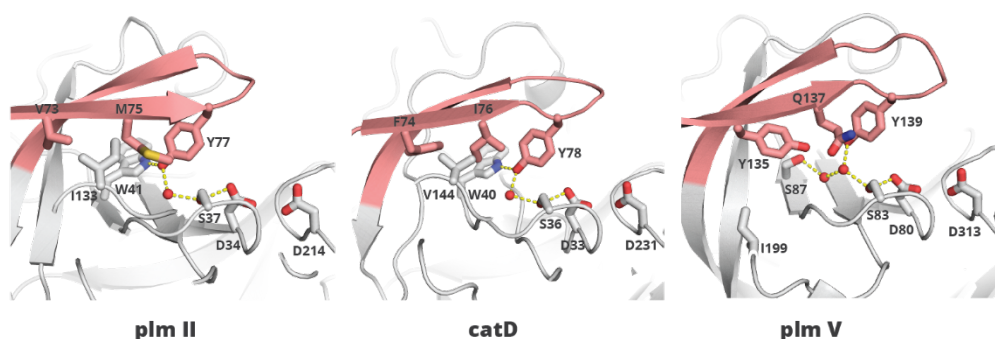


Figure 7. Hydrogen bonding network in the active site of plm II (PDB ID 1SME), catD (PDB ID 1LYA) and plm V (PDB ID 4ZL4). Protein backbone is depicted in cartoon representation, selected binding pocket residues are shown as white sticks, conserved water molecule as sphere. Hydrogen bonds are indicated with yellow dashed line. Flap loop is in salmon. Hydrogens are omitted for clarity.

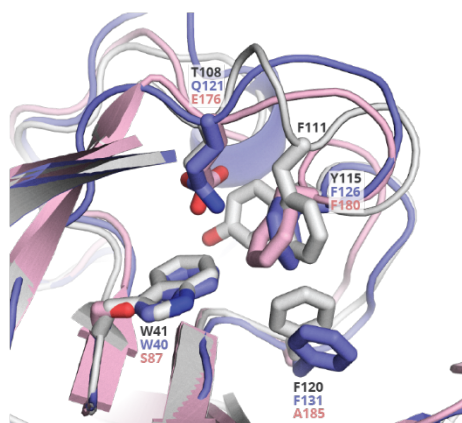


Figure 8. Superimposed plm II (white, PDB ID 1SME), plm V (pink, PDB ID 4ZL4) and catD (purple, PDB ID 1LYA) binding sites. Protein backbone is depicted in cartoon representation, selected binding pocket residues are shown as sticks. Hydrogens are omitted for clarity.

Conclusions

The characterisation of the malaria aspartic protease flap loop dynamics upon open-flap inhibitor binding provided several conclusions. First, several alternative binding modes are observed for majority of 2-aminoquinazolin-4(3*H*)-one based plm inhibitors. These binding modes are believed to be key transition states on the ligand binding pathway, and plays role in the ligand flap pocket substituent guiding under the flap loop and flap loop opening. Second, introduction of larger flap pocket substituent reduces the conformational space available by the ligand and shifts the binding preference to the binding mode observed in the crystal structure. This results in improved inhibitor activity, since role of the alternative binding modes is diminished. Third, poor 2-aminoquinazolin-4(3*H*)-one based inhibitor potency against plm V and catD is a result of ligands inability to bind under the flap loop. Binding modes identified suggest that the ligand binding under the flap loop in plm II is initiated by the hydrophobic interactions in-between the flap loop and adjacent α -helix; such interactions in plm V and catD are unfavourable.

Taken together, our findings confirm that plm inhibitor selectivity can be achieved by targeting the flap loop with hydrophobic substituents that enables ligand binding under the flap loop, as such behaviour is not observed for several other aspartic proteases. The insight on the molecular determinants of ligand selectivity between plm II, plm V and catD provide a clear indication for the structure-based rational design of selective inhibitors. The ability to estimate compound selectivity before they are synthesized is of great importance in drug design, therefore, we expect that approach used here will be a valid support in selective inhibitor design not only against aspartic proteases, but other enzymes as well.

Acknowledgement

This work was supported by the Latvian Council of Science, project No. Izp-2020/2-0012. RB acknowledges European Regional Development Fund project No. 1.1.1.2/VIAA/2/18/379 for financial support. Professor Emilio Parisini and Teodors Pantelejevs are acknowledged for their help with X-ray diffraction data acquisition and crystal structure refinement. We acknowledge Swiss Light Source for access to macromolecular crystallography beam line PX3 (proposal 20201060). Calculations were partly performed on Riga Technical University (RTU) High Performance Computing (HPC) Centre facilities.

- [1] G. H. Coombs, D. E. Goldberg, M. Klemba, C. Berry, J. Kay, J. C. Mottram, *Trends Parasitol.* **2001**, *17*, 532–537.
- [2] P. J. Rosenthal, **1999**, pp. 105–159.
- [3] B. Turk, *Nat. Rev. Drug Discov.* **2006**, *5*, 785–799.
- [4] J. Eder, U. Hommel, F. Cumin, B. Martoglio, B. Gerhartz, *Curr. Pharm. Des.* **2007**, *13*, 271–285.
- [5] P. Favuzza, M. de Lera Ruiz, J. K. Thompson, T. Triglia, A. Ngo, R. W. J. Steel, M. Vavrek, J. Christensen, J. Healer, C. Boyce, Z. Guo, M. Hu, T. Khan, N. Murgolo, L. Zhao, J. S. Penington, K. Reaksudsan, K. Jarman, M. H. Dietrich, L. Richardson, K.-Y. Guo, S. Lopaticki, W.-H. Tham, M. Rottmann, T. Papenfuss, J. A. Robbins, J. A. Boddey, B. E. Sleebs, H. J. Sabroux, J. A. McCauley, D. B. Olsen, A. F. Cowman, *Cell Host Microbe* **2020**, 1–17.
- [6] J. A. Boddey, *Science (80-)*. **2017**, *358*, 445–446.
- [7] A. S. Nasamu, S. Glushakova, I. Russo, B. Vaupel, A. Oksman, A. S. Kim, D. H. Fremont, N. Tolia, J. R. Beck, M. J. Meyers, J. C. Niles, J. Zimmerberg, D. E. Goldberg, *Science (80-)*. **2017**, *358*, 518–522.
- [8] P. Pino, R. Caldelari, B. Mukherjee, J. Vahokoski, N. Klages, B. Maco, C. R. Collins, M. J. Blackman, I. Kursula, V. Heussler, M. Brochet, D. Soldati-Favre, *Science (80-)*. **2017**, *358*, 522–528.
- [9] D. E. Goldberg, *Proc. Natl. Acad. Sci.* **2013**, *110*, 5283–5284.
- [10] B. E. Sleebs, M. Gazdik, M. T. O'Neill, P. Rajasekaran, S. Lopaticki, K. Lackovic, K. Lowes, B. J. Smith, A. F. Cowman, J. A. Boddey, *J. Med. Chem.* **2014**, *57*, 7644–7662.
- [11] B. E. Sleebs, S. Lopaticki, D. S. Marapana, M. T. O'Neill, P. Rajasekaran, M. Gazdik, S. Günther, L. W. Whitehead, K. N. Lowes, L. Barfod, L. Hviid, P. J. Shaw, A. N. Hodder, B. J. Smith, A. F. Cowman, J. A. Boddey, *PLoS Biol.* **2014**, *12*, e1001897.
- [12] C. Sedwick, *PLoS Biol.* **2014**, *12*, 1–2.
- [13] J. A. Boddey, A. N. Hodder, S. Günther, P. R. Gilson, H. Patsiouras, E. A. Kapp, J. A. Pearce, T. F. De Koning-Ward, R. J. Simpson, B. S. Crabb, A. F. Cowman, *Nature* **2010**, *463*, 627–631.
- [14] A. M. Dondorp, F. Nosten, P. Yi, D. Das, A. P. Phyo, J. Tarning, K. M. Lwin, F. Arie, W. Hanpithakpong, S. J. Lee, P. Ringwald, K. Silamut, M. Imwong, K. Chotivanich, P. Lim, T. Herdman, S. S. An, S. Yeung, P. Singhasivanon, N. P. J. Day, N. Lindegardh, D. Socheat, N. J. White, *N. Engl. J. Med.* **2009**, *361*, 455–467.
- [15] J. E. Hyde, *FEBS J.* **2007**, *274*, 4688–4698.
- [16] R. Bobrovs, K. Jaudzems, A. Jirgensons, *J. Med. Chem.* **2019**, *62*, 8931–8950.
- [17] K. Ersmark, B. Samuelsson, A. Hallberg, *Med. Res. Rev.* **2006**, *26*, 626–666.
- [18] Y. Cai, N. Kurt Yilmaz, W. Myint, R. Ishima, C. A. Schiffer, *J. Chem. Theory Comput.* **2012**, *8*, 3452–3462.
- [19] M. Mahanti, S. Bhakat, U. J. Nilsson, P. Söderhjelm, *Chem. Biol. Drug Des.* **2016**, 159–177.

- [20] A. M. Silva, A. Y. Lee, S. V. Gulnik, P. Maier, J. Collins, T. N. Bhat, P. J. Collins, R. E. Cachau, K. E. Luker, I. Y. Gluzman, S. E. Francis, A. Oksman, D. E. Goldberg, J. W. Erickson, *Proc. Natl. Acad. Sci.* **1996**, 93, 10034–10039.
- [21] N. Kurt, W. R. P. Scott, C. A. Schiffer, T. Haliloglu, *Proteins Struct. Funct. Genet.* **2003**, 51, 409–422.
- [22] R. Ishima, N. Kurt Yilmaz, C. A. Schiffer, *J. Biomol. NMR* **2019**, 73, 365–374.
- [23] W. Karubiu, S. Bhakat, L. McGillevie, M. E. S. Soliman, *Mol. BioSyst.* **2015**, 11, 1061–1066.
- [24] S. Bhakat, *RSC Adv.* **2021**, 11, 11026–11047.
- [25] J. A. Henderson, J. Shen, *J. Chem. Inf. Model.* **2021**, 1, DOI 10.1021/acs.jcim.1c01180.
- [26] R. Ishima, J. M. Louis, *Proteins Struct. Funct. Bioinforma.* **2007**, 70, 1408–1415.
- [27] D. Hamelberg, J. A. McCammon, *J. Am. Chem. Soc.* **2005**, 127, 13778–13779.
- [28] N. Okimoto, T. Tsukui, K. Kitayama, M. Hata, T. Hoshino, M. Tsuda, *J. Am. Chem. Soc.* **2000**, 122, 5613–5622.
- [29] S. N. Khan, J. D. Persons, J. L. Paulsen, M. Guerrero, C. A. Schiffer, N. Kurt-Yilmaz, R. Ishima, *Biochemistry* **2018**, 57, 1652–1662.
- [30] Y. Cai, W. Myint, J. L. Paulsen, C. A. Schiffer, R. Ishima, N. Kurt Yilmaz, *J. Chem. Theory Comput.* **2014**, 10, 3438–3448.
- [31] A. P. Wiley, S. L. Williams, J. W. Essex, *J. Chem. Theory Comput.* **2009**, 5, 1117–1128.
- [32] A. Badaya, Y. U. Sasidhar, *Int. J. Biol. Macromol.* **2018**, 118, 1696–1707.
- [33] M. Zürcher, F. Hof, L. Barandun, A. Schütz, W. B. Schweizer, S. Meyer, D. Bur, F. Diederich, *European J. Org. Chem.* **2009**, 1707–1719.
- [34] F. Hof, A. Schütz, C. Fähr, S. Meyer, D. Bur, J. Liu, D. E. Goldberg, F. Diederich, *Angew. Chemie - Int. Ed.* **2006**, 45, 2138–2141.
- [35] M. Zürcher, F. Diederich, *J. Org. Chem.* **2008**, 73, 4345–4361.
- [36] A. P. Huizing, M. Mondal, A. K. H. Hirsch, *J. Med. Chem.* **2015**, 58, 5151–5163.
- [37] L. Prade, A. F. Jones, C. Boss, S. Richard-Bildstein, S. Meyer, C. Binkert, D. Bur, *J. Biol. Chem.* **2005**, 280, 23837–23843.
- [38] B. R. Jagger, S. E. Kochanek, S. Haldar, R. E. Amaro, A. J. Mulholland, *Curr. Opin. Struct. Biol.* **2020**, 61, 213–221.
- [39] G. Bussi, A. Laio, *Nat. Rev. Phys.* **2020**, 2, 200–212.
- [40] P. Tiwary, V. Limongelli, M. Salvalaglio, M. Parrinello, *Proc. Natl. Acad. Sci. U. S. A.* **2015**, 112, E386–E391.
- [41] A. Cavalli, A. Spitaleri, G. Saladino, F. L. Gervasio, *Acc. Chem. Res.* **2015**, 48, 277–285.
- [42] F. Colizzi, M. Masetti, M. Recanatini, A. Cavalli, *J. Phys. Chem. Lett.* **2016**, 7, 2899–2904.
- [43] L. Callea, L. Bonati, S. Motta, *J. Chem. Theory Comput.* **2021**, 18, DOI 10.1021/acs.jctc.1c00114.
- [44] T. W. Whitfield, D. A. Ragland, K. B. Zeldovich, C. A. Schiffer, *J. Chem. Theory Comput.* **2020**, DOI 10.1021/acs.jctc.9b00781.
- [45] A. Dickson, *Biophys. J.* **2018**, 115, 1707–1719.
- [46] G. Mattedi, F. Deflorian, J. S. Mason, C. De Graaf, F. L. Gervasio, *J. Chem. Inf. Model.* **2019**, 59, 2830–2836.
- [47] N. Plattner, F. Noé, *Nat. Commun.* **2015**, 6, DOI 10.1038/ncomms8653.
- [48] R. A. Copeland, D. L. Pompliano, T. D. Meek, *Nat. Rev. Drug Discov.* **2006**, 5, 730–739.
- [49] S. Núñez, J. Venhorst, C. G. Kruse, *Drug Discov. Today* **2012**, 17, 10–22.
- [50] A. Barducci, G. Bussi, M. Parrinello, *Phys. Rev. Lett.* **2008**, 100, 1–4.
- [51] P. Tiwary, M. Parrinello, *Phys. Rev. Lett.* **2013**, 111, 1–5.
- [52] A. Laio, M. Parrinello, *Proc. Natl. Acad. Sci. U. S. A.* **2002**, 99, 12562–12566.
- [53] F. Pietrucci, *Rev. Phys.* **2017**, 2, 32–45.
- [54] M. Bernetti, M. Masetti, M. Recanatini, R. E. Amaro, A. Cavalli, *J. Chem. Theory Comput.* **2019**, 15, 5689–5702.
- [55] P. Tiwary, M. Parrinello, *J. Phys. Chem. B* **2015**, 119, 736–742.
- [56] G. Bussi, D. Branduardi, **2015**, 28, 1–49.
- [57] A. Barducci, M. Bonomi, M. Parrinello, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, 1, 826–843.
- [58] M. Bonomi, A. Barducci, M. Parrinello, *J. Comput. Chem.* **2009**, 30, 1615–1621.
- [59] L. Troussicot, F. Guillière, V. Limongelli, O. Walker, J. M. Lancelin, *J. Am. Chem. Soc.* **2015**, 137, 1273–1281.
- [60] V. Limongelli, M. Bonomi, M. Parrinello, *Proc. Natl. Acad. Sci. U. S. A.* **2013**, 110, 6358–6363.
- [61] S. Raniolo, V. Limongelli, *Nat. Protoc.* **2020**, 15, 2837–2866.
- [62] Z. F. Brotzakis, V. Limongelli, M. Parrinello, *J. Chem. Theory Comput.* **2019**, 15, 743–750.
- [63] F. Comitani, V. Limongelli, C. Molteni, *J. Chem. Theory Comput.* **2016**, 12, 3398–3406.
- [64] V. Limongelli, M. Bonomi, L. Marinelli, F. L. Gervasio, A. Cavalli, E. Novellino, M. Parrinello, *Proc. Natl. Acad. Sci. U. S. A.* **2010**, 107, 5411–5416.
- [65] D. Rasina, G. Stakanovs, O. V. Borysov, T. Pantelejevs, R. Bobrovs, I. Kanepe-Lapsa, K. Tars, K. Jaudzems, A. Jirgensons, *Bioorg. Med. Chem.* **2018**, 4, 1–13.
- [66] D. Rasina, G. Stakanovs, I. Kanepe-Lapsa, R. Bobrovs, K. Jaudzems, A. Jirgensons, *Chem. Heterocycl. Compd.* **2020**, 56, 786–792.
- [67] D. Rasina, M. Otikovs, J. Leitans, R. Recacha, O. V. Borysov, I. Kanepe-Lapsa, I. Domraceva, T. Pantelejevs, K. Tars, M. J. Blackman, K. Jaudzems, A. Jirgensons, *J. Med. Chem.* **2016**, 59, 374–387.
- [68] M. Invernizzi, M. Parrinello, *J. Phys. Chem. Lett.* **2020**, 11, 2731–2736.
- [69] D. Branduardi, F. L. Gervasio, M. Parrinello, *J. Chem. Phys.* **2007**, 126, DOI 10.1063/1.2432340.
- [70] O. A. Asojo, S. V. Gulnik, E. Afonina, B. Yu, J. A. Ellman, T. S. Haque, A. M. Silva, *J. Mol. Biol.* **2003**, 327, 173–181.
- [71] N. S. Andreeva, L. D. Rumsh, *Protein Sci.* **2008**, 10, 2439–2450.
- [72] M. Zürcher, T. Gottschalk, S. Meyer, D. Bur, F. Diederich, *ChemMedChem* **2008**, 3, 237–240.