Exploring the binding pathway of novel non-peptidomimetic plasmepsin V inhibitors

Raitis Bobrovs[1]*, Laura Drunka[1], Iveta Kanepe[1], Aigars Jirgensons[1], Amedeo Caflisch[2], Matteo Salvalaglio[3], Kristaps Jaudzems[1]

[1] Latvian Institute of Organic Synthesis, Aizkraukles 21, Riga, LV1006, Latvia
[2] Department of Biochemistry, University of Zurich, Winterthurerstrasse 190, CH-8057, Zurich, Switzerland
[3] Thomas Young Centre and Department of Chemical Engineering, University College London, London WC1E 7JE, United Kingdom

E-mail: raitis.bobrovs@osi.lv

Predicting the interaction modes and binding affinities of virtual compound libraries is of great interest in drug development. It reduces the cost and time of lead compound identification and selection. Here we apply path-based metadynamics simulations to characterise the binding of potential inhibitors to the *Plasmodium falciparum* aspartic protease plasmepsin V (plm V), a validated antimalarial drug target that has a highly mobile binding site. The potential plm V binders were identified in a high throughput virtual screening (HTVS) campaign and were experimentally verified in a fluorescence resonance energy transfer (FRET) assay. Our simulations allowed us to estimate compound binding energies and revealed putative transition states along binding/unbinding pathways in atomistic resolution. We believe that the method described allows the prioritisation of compounds for synthesis and enables rational structure-based drug design for targets that undergo considerable conformational changes upon inhibitor binding.

Introduction

Drug discovery relies on computer-aided drug discovery (CADD) methods to accelerate this time-consuming and costly process. In addition, CADD and particularly rational (structure-based) design has gained a notable role in identifying novel potentially active chemical scaffolds (through virtual screening, docking, and similarity search), as well as in hit-to-lead optimisation and fragment evolution.[1–3] Predicting the interaction modes and affinity of virtual compound libraries allow the prioritisation of compounds for synthesis or testing and rationalising structure-activity relationships.

A major challenge in applying structure-based drug design remains the accurate estimation of target protein-ligand binding free energies (BFEs). This is essential for lead optimisation since it is the most time and resources consuming step. Recent improvements in protein and ligand force fields, molecular dynamics (MD) codes, and enhanced sampling algorithms have made the calculation of relative and absolute binding free energies more accurate and accessible.[4–7] Currently, free energy perturbation (FEP) methods are preferred for relative binding free energy (RBFE) calculation, typically used in the hit-to-lead optimisation stage.[8,9] Lately, significant progress has also been made in the calculation of absolute binding free energies (ABFEs) using alchemical approaches.[10] These advances have made ABFE calculations more accessible and attractive at the hit identification stage as well. There are, however, several drawbacks that hamper the widespread use of perturbation-based methods. Most notably, calculating accurate binding energies for systems that undergo considerable conformational changes or systems that contain charged and/or non-congeneric ligands is impossible due to the inability to accurately sample pharmacologically relevant conformations and describe pocket solvation/desolvation.[11,12]

Alternatives to perturbation-based methods are collective variable-based free energy calculation methods, like metadynamics, which permit ABFE calculation along a physical

binding trajectory.[7,13–24] Metadynamics allows to sample target conformations relevant to the binding process and explore transition states along the binding pathway. Besides, it is usually advantageous to understand the inhibitor binding at an atomic scale (key protein-inhibitor and protein-protein interactions, ligand flexibility, solvation effects, etc.), as information on inhibitor binding/unbinding pathway, potential transition states can provide the basis for targeted lead compound development to design compounds with higher activity and specificity.[18,25,26] The ability of metadynamics to sample the complete binding pathway, including target flexibility upon ligand binding and unbinding, makes it an excellent choice in ligand binding studies. In contrast, alchemical methods only sample the bound and unbound states, and no information on potential transition states is obtained.
Drug targets containing flexible loops next to the binding site or having cryptic or allosteric binding pockets are not uncommon[27,28]. One such drug target class is aspartic proteases, in which a long β-hairpin structure covers the binding site, usually referred to as a flap or flap loop. It is a highly mobile region which is involved in substrate recognition[29]. Here we applied metadynamics to study the binding of potential inhibitors to the *P. Falciparum* aspartic protease plasmepsin V (plm V), a validated antimalarial drug target. The inhibitors were identified via high throughput virtual screening, and their plm V inhibitory potency was verified experimentally. The study shows the potential of metadynamics simulations to reveal the binding modes and energies of aspartic protease inhibitors and to facilitate the discovery of new inhibitor scaffolds.

Experiment design

Fig. 1 outlines the research workflow, and critical aspects are described below. In essence, the plm V structure that enables non-peptidomimetic inhibitor docking was prepared and used to identify novel inhibitor scaffolds via high throughput virtual screening. The inhibitors identified were verified experimentally, and metadynamics simulations were applied to characterise inhibitor binding modes and energies at an atomic scale.

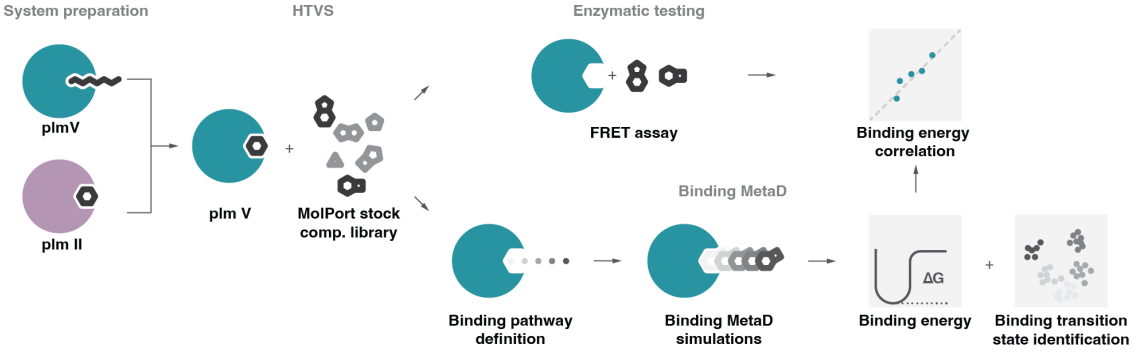

Fig. 1. Schematic research workflow

System preparation

Plasmepsin V (plm V) is a promising yet unexploited antimalarial drug target. It is an essential protease that processes proteins for export into the host erythrocytes and is located in the endoplasmic reticulum.[30] Plm V is a phylogenetically unique aspartic protease which shares little conservation with human proteases [31–33], making it less likely that its inhibitors will have selectivity issues. Most known plm V inhibitors are peptidomimetic[30,33–37]; for two of them, the crystal structures have been solved (PDB IDs: 4ZL4 and 6C4G; from *Plasmodium vivax*). Both structures[34,35] originate from complexes, where the flap loop (Tyr135-Gly147) is closed over the active site. While no plm V crystal structure in complex with a non-peptidomimetic inhibitor has been reported, it is known that binding of such inhibitors to other aspartic proteases can lead to conformations with the flap loop in a more open state (PDB codes 4Z22, 2BJU)[38]. Since we were interested in identifying non-peptidomimetic plm V inhibitors that could bind to the open flap conformation, we prepared a

docking model that would be able to accommodate such inhibitors. To do this, a moderately active plm V inhibitor (2-amino-4(3H)-quinazolinone DR720, IC_{50} = 64 μM), identified from our in-house aspartic protease inhibitor library[39–43], was modelled in the active site of plm V in a pose similar to that of another 2-amino-4(3H)-quinazolinone in complex with plm II (PDB ID 4Z22) (see Fig. 2). The Asp80 of the catalytic dyad was protonated, whereas Asp313 was deprotonated[44]. This system was subjected to molecular dynamics simulation with aspartic dyad and 2-amino-4(3H)-quinazolinone core intermolecular distances restrained (see Methods).

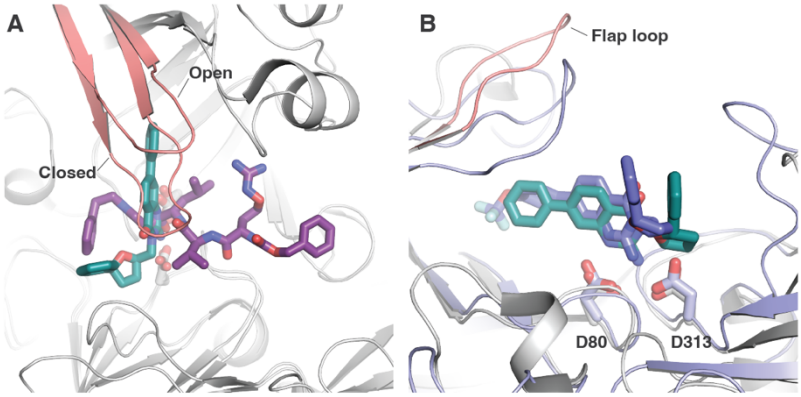

Fig. 2. ***A*** *Plm V-peptidomimetic inhibitor complex (PBD ID: 4ZL4) superimposed on plm II-non-peptidomimetic open flap inhibitor complex (PBD ID: 4Z22). The flap loop is shown in salmon. Peptidomimetic inhibitor WEHI-842 is shown as purple sticks, non-peptidomimetic inhibitor DR718A as teal sticks and catalytic aspartate sidechains as grey sticks. Hydrogens are omitted for clarity.* ***B*** *Plm V in complex with inhibitor DR720 after 100 ns restrained MD simulation superimposed on plm II-DR718A complex (PBD ID: 4Z22). Plm V is shown as a violet cartoon, and Plm II as a white cartoon. Catalytic aspartate sidechains are shown as sticks. The plm II flap loop is shown in salmon. DR720 is shown as violet sticks, and DR718A as teal sticks. Hydrogens are omitted for clarity.*

HTVS

The minimised protein structure was used in a high throughput virtual screening (HTVS) of the drug-like screening compounds from the MolPort stock compound library (~6 M comp.; 2020). To retain only the unique virtual screening hit scaffolds (i.e., the top-scoring compound from each cluster), the top scoring 3000 compounds were clustered by applying Tanimoto similarity metrics to linear molecule fingerprints. The top-ranked 300 compounds were visually inspected for their ability to form hydrogen bonds with the catalytic dyad and hydrophobic interactions with the flap pocket residues. Molecules showing internal strains or unsatisfied hydrogen bond donors were deprioritised.

Binding metadynamics

The inhibitor binding/unbinding pathway and respective binding energy were calculated using fully atomistic molecular dynamics simulations. This approach describes the movement of flexible protein parts near the binding site and considers the buried pocket solvation/desolvation during the inhibitor binding/unbinding event. Since inhibitor binding is a rare event and typically is far beyond the reach of typical atomistic simulations, we exploited an enhanced sampling method called metadynamics (metaD). MetaD is a method where molecular dynamics simulation is biased along a set of collective variables (CVs) using a history-dependent potential. The Gaussian-shaped potential is applied at regular intervals to escape any local minima along the CV space and visit previously unexplored regions in the CV space. Selecting the appropriate CVs for the process studied is crucial and often the most challenging step. The most apparent CV in the ligand-protein binding studies is the distance between ligand and binding site atoms. This CV alone, however, is unsuitable for systems where the protein binding site is highly flexible, as it has no control over the protein

binding site flexibility, and might result in situations where protein conformational space of interest is poorly explored. The intermolecular distance CV can be combined with CVs that describe protein flexibility.[22,45] However, since the computational cost to reconstruct the free energy surface (FES) grows exponentially with the number of CVs used, it is undesirable. To overcome the limitations above, we used the path CV that enables exploring complex multidimensional processes along a predefined pathway. In this method, two CVs describe the process: s – the progress along the predefined reference path, and z – the distance orthogonal to the reference path. Furthermore, introducing the CV z allows the exploration of configurations that differ from the reference path; thus, if the reference path provided is not entirely accurate, the system can deviate from it and discover more probable pathways.

Results

Virtual screening

Following the virtual screening approach described above, we selected and purchased 28 chemically diverse top-scoring compounds. These compounds were tested for plm V inhibition potency in a FRET-based assay, and 7 showed micromolar potency (IC_{50} in the range from 4.4 to 70 μM; see Fig. 3.; complete list of tested inhibitors in Table SI1). Additionally, the inhibition potency of three cross-inhibition markers, human cathepsin D (catD) and digestive plasmepsins II and IV, was measured for the compounds that showed measurable activity against plm V. Compounds MolPort-000-124-439, MolPort-002-904-606 and MolPort-046-754-050 showed considerable selectivity and no inhibition of cross-inhibition markers was observed at 100 μM concentration (see Table SI1). All of the hits identified had several common features: a) hydrogen bond donor and acceptor in a configuration that mimics the transition state of an enzymatic reaction and is capable of forming hydrogen bonds with catalytic dyad; and b) hydrophobic substituent forming hydrophobic interactions under the flap loop once the ligand core hydrogen bond donor and acceptor groups are interacting with the catalytic site. The transition state mimicking group alignment in the binding site matches other non-peptidomimetic aspartic protease inhibitors, where ligand hydrogen bond donor and acceptor bind co-planarly with aspartic dyad residues. The most commonly observed functional group interacting with the aspartic dyad was primary amide, whereas heterocyclic compound cores were represented by quinazoline-2-amine, pyrimidine-2-amine and triazoledione. Quinazolin-2-amines and pyrimidine-2-amines with substituents at various positions have been reported as aspartic protease inhibitors before[43,46], whereas triazoledione and primary amide-based compounds, to our knowledge, are novel. The hydrophobic substituents were predominantly aromatic systems interacting with flap pocket residues Tyr135, Tyr139, Ile145, Phe180 and Val188. In addition to the interactions mentioned above, compounds MolPort-023-187-757 and MolPort-046-754-050 interacted with the Ser316 sidechain, forming hydrogen bonds via sulphonamide and methoxy groups, respectively.

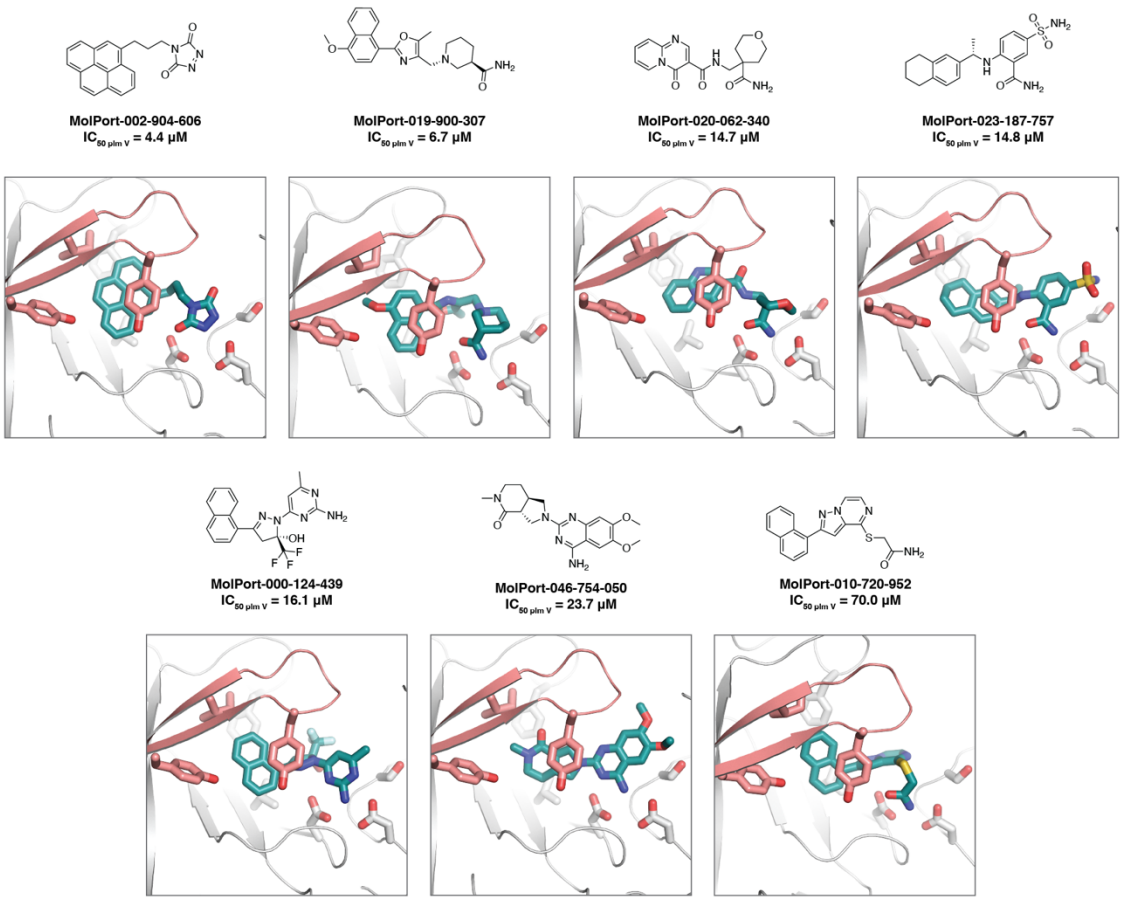

Fig. 3. Virtual screening hits against plm V. Compound molecular structure, MolPort ID and experimental IC_{50} value (µM) against plm V are given on the left; the docked pose of the respective compound in complex with plm V—on the right. Inhibitor and key amino acid residues are shown as teal and grey sticks, respectively. The flap loop is shown in salmon. Hydrogens are omitted for clarity.

Of the seven hits identified, commercially available analogues were purchased for MolPort-002-904-606, MolPort-000-124-439 and MolPort-046-754-050 (12, 24 and 15 analogues, respectively) due to their selectivity over human cathepsin D and higher medicinal chemistry development potential. The analogues selected had modifications in the expected flap pocket substituent position, whereas the transition state mimetic part of the molecule was not changed. The majority of the MolPort-002-904-606 and MolPort-000-124-439 analogues showed lower plm V inhibitory potency than the original hits (see Tables SI2, SI3), and only MolPort-019-894-150 was slightly more potent (IC_{50} (plm V) = 13.0±0.8 µM) than the parent compound. However, for MolPort-046-754-050, nearly all tested analogues showed higher plm V inhibition potency than the parent compound. The most active compound was compound MolPort-035-715-983 with an IC_{50} of 5.0±0.3 µM.

Binding energy and pathway calculation

The ligand binding/unbinding path used in our study as a reference pathway was generated from a trajectory obtained during an initial funnel metadynamics simulation (comp. DR720 unbinding from plm V, see Methods). It consisted of 15 equally spaced configurations, prepared using Plumed[47] pathtools. The Cα atoms of the binding site (within 15 Å from Ser87, which sits under the central part of the flap loop) and two ligand atoms were used to define a path describing both the binding site flexibility and ligand position during the binding process. The two ligand atoms defined within the path were hydrogen bond donor and acceptor groups interacting with catalytic aspartates. This configuration allowed us to use

the same path for all potential inhibitor simulations and specify ligand orientation with respect to the binding site. During the preliminary runs, it was observed that ligands occasionally would considerably deviate from the path provided. Therefore, upper wall restraint potential was introduced at RMSD 0.1 nm^2 for the CV *z*. This allowed the ligand to explore the binding pocket while restricting the space explored in the unbound state.
The path MetaD simulations were performed for all 28 ligands considered, ranging in simulation time from 1-3 μs. During these simulations, multiple ligand binding and unbinding events were observed, and ligands explored both bound and unbound states (frames 1-3 and 13-15 on the path, respectively; see Fig. 4.B and C). The unbiased probability distribution of the system was obtained through reweighting[48], which allowed us to construct the FES associated with the binding process (see Fig. 4.D) and calculate the ligand binding free energy. A randomly-selected, verified plm V inhibitor (MolPort-023-187-757; further referred as compound **1**) was used to optimise metaD parameters (see Methods) and is used as a representative case throughout the study.

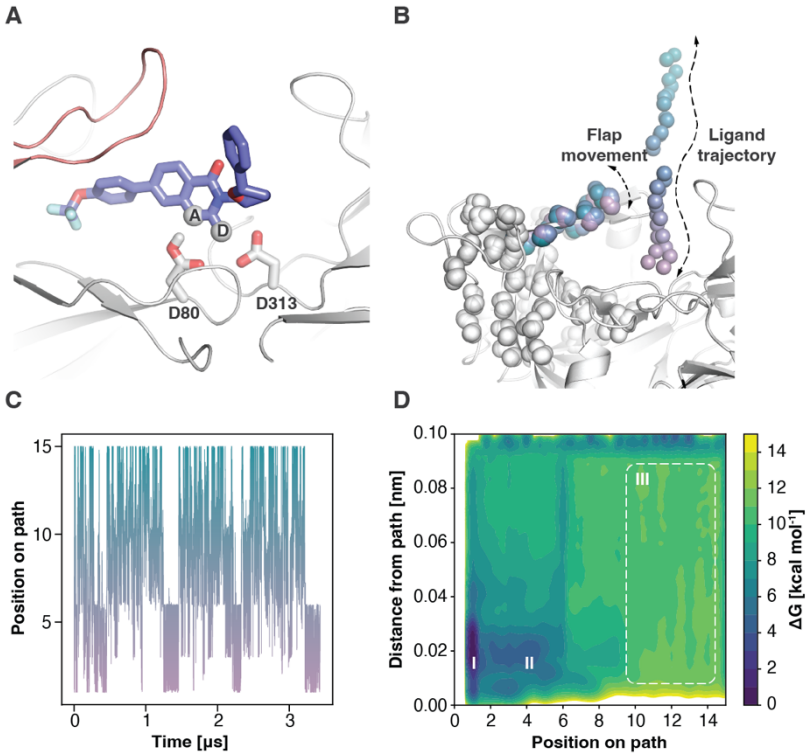

Fig. 4. ***A*** *Ligand transition state mimetic group hydrogen bond donor (D) and acceptor (A) atoms used to define ligand position on the binding/unbinding path depicted on plm V-DR720 complex. The flap loop is shown in salmon. DR720 is shown as violet sticks and catalytic aspartate sidechains as grey sticks. Hydrogens are omitted for clarity.* ***B*** *15 superimposed frames defining ligand binding/unbinding pathway. Atoms defining path (shown as spheres) are Cα atoms of residues involved in β-sheet secondary structures within 1.5 nm from Ser87, and ligand atoms A and D from subfigure A. Colour gradient shows progress along the binding/unbinding pathway, where purple conformation corresponds to bound state, teal—unbound.* ***C*** *System position on the binding/unbinding path as a function of metadynamics simulation time in the case of the plm V-comp.**1** system. Colours correspond to the states given in subfigure B: purple—bound state; teal—unbound.* ***D*** *FES of compound **1** binding to plm V obtained after reweighing[48] path metaD simulation. Isosurfaces are shown every 1 kcal/mol. The deepest FES basin I corresponds to the bound state, similar to the docked pose. Basin II corresponds to the putative transition state where the ligand interacts with the plm V flap loop, whereas state III corresponds to the unbound state where the ligand is not interacting with the protein.*

Ligand binding free energy

The ligand binding free energy was calculated as a difference between the lowest energy bound state **I** and unbound state **III** (Fig. 4.D), where the ligand is completely solvent exposed and does not interact with the protein. Notably, the unbound region **III** is flat, indicating the presence of iso-energetic states once the ligand is sufficiently distant from the protein. The convergence of the binding energy was estimated by monitoring the evolution of the absolute binding free energy throughout the simulation (see Fig. 5.A and B). Fig. 5 shows that for the plm V-compound **1** system, free energy converges after ~1000 ns. Similar behaviour was observed for the rest of the inhibitors.

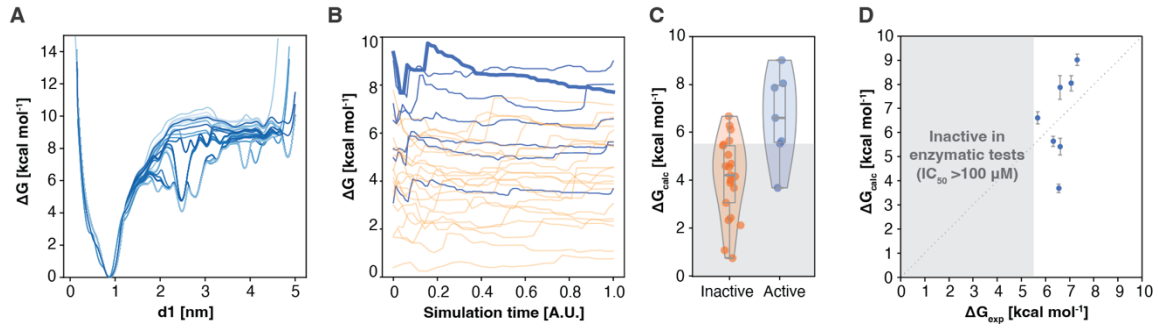

*Fig. 5. **A** The FES as a function of ligand core – aspartic dyad distance d1 obtained by reweighting the metadynamics simulation every 50 ns (for comp. **1**). The colour gradient indicates reweighting time: the lightest blue line - reweighting up to 50 ns; the darkest blue – reweighting up to 3.0 μs. The binding site region is up to ~1 nm, while fully solvated at >3 nm. **B** The binding free energy (ΔG) of the HTVS hits, and plm V is calculated every 50 ns throughout the simulation. Blue indicates verified inhibitors and orange – inactive compounds. The thick blue line represents reference comp. **1**. The compound **1** estimated binding energy of -7.8±0.5 kcal/mol is slightly higher than the experimental value (calculated from the IC_{50}) of ~-6.6 kcal/mol. **C** The violin plots of estimated binding free energy (ΔG_{calc}) for the experimentally verified active and inactive compounds. Grey shaded area indicates binding free energy corresponding to IC_{50} > 100 μM. **D** The estimated binding free energy (ΔG_{calc}) plotted against experimentally determined (ΔG_{exp}) values (calculated from IC_{50} values as $\Delta G_{exp}=-k_BT\ ln(IC_{50})$).*

The IC_{50} values were determined for the compounds with more than 50% plm V inhibition at 100 μM concentration. The binding energy was calculated from IC_{50} values as $\Delta G_{exp}=-k_BT\ ln(IC_{50})$ for the identified active compounds. The violin plots (Fig. 5.C) of the estimated binding free energy (ΔG_{calc}) indicate that PathMetaD binding energy estimate allows us to distinguish experimentally verified active and inactive compounds. Moreover, the scatter plot of the calculated and experimental binding energy (Fig. 5.D) demonstrates that PathMetaD binding energy estimate falls within ~1.3 kcal/mol of experimental data. Such accuracy is comparable to other MD-based binding energy calculation methods[49]. The one distinct outlier for the enzymatically verified plm V inhibitors is the binding energy estimate for the comp. MolPort-000-124-439 with a calculated binding free energy of -3.7 kcal/mol versus the experimental value of -6.5 kcal/mol. The calculated binding energies for the experimentally verified inactive compounds ranged from -0.76 to -6.68 kcal/mol. According to the above equation, these compounds' free energy of binding should be less negative than -5.5 kcal/mol (since IC_{50} > 100 μM). Four experimentally verified inactive compounds showed slightly better binding energy. The accuracy of the binding energy estimate is typically hindered by insufficient sampling and imperfect forcefield parameters. Here, the ligand extensively explored both bound and unbound states (see Fig. 4.C). Therefore, contributions from even minor parameter imprecisions in protein forcefield, ligand parameterisation and water model could cause the unexpected outcome.

Binding modes

Besides the lowest energy state **I** in Fig. 4.D, an additional basin (basin **II** in Fig. 4.D) was observed for most ligands, indicating the presence of transition state(s) along the ligand binding pathway. While path CVs easily discriminate between the bound and unbound states, it does not provide in-depth information on the binding mode(s) and possible transition states. Such information, however, can often be accessed by reweighing the trajectory and reconstructing the FES as a function of alternative CVs that allow decoupling states that otherwise would overlap in the initial CV space. Here, reweighting the trajectory to construct FES as a function of ligand-protein catalytic site distance *d1* and ligand torsion with respect to protein binding site *t1* (see Fig. SI1) did not provide clear information on key transition states along the binding pathway. The FES constructed had broad global minima corresponding to two similar binding modes – one in a pose similar to the docked one (basin **IA** in Fig.SI1.C) and another with hydrophobic flap pocket substituent slightly twisted to fill S3 pocket (basin **IB** in Fig.SI1.C). No information was obtained on intermediate state **II** observed in the biased path CV space (*s*, *z*) and one-dimensional FES (distance *d1*) at ~2.5 nm. Therefore, a dimensionality reduction method that allows the mapping of high-dimensional processes to low-dimensional space was employed. Here we exploited the non-linear dimensionality reduction algorithm sketch-map, based on metric multidimensional scaling and specifically designed to deal with atomistic simulation data [50,51]. This method accurately reproduces the relative position of adjacent basins and can often generate an intuitive representation of complex pathways. The 2D map generated (see Methods) and representative structures are shown in Fig. 6. Here, numerous well-localised basins were identified, and basins that contain structures that are expected to be next to each other on a binding pathway are located in close proximity and connections between these clusters become visible. For example, the binding mode with a ligand slightly twisted and hydrophobic substituent filling the S3 pocket (basin **IB**), as well as clusters representing transition states where the ligand is next to the catalytic dyad but interacting with a flap loop (basins **IC** and **ID** in Fig. 6.A) are located next to the bound state basin **IA**. Basins located further away from the bound state basins correspond to states where the ligand is outside the binding site. The most populated of these is basin **II**, where the inhibitor interacts with a flap loop tip and flexible loop on the opposite side of the binding groove[38]. This is the first ligand and protein encounter on the binding pathway. Interactions formed here guide the ligand deeper into the binding site to finally form hydrogen bonds with the catalytic dyad. Configurations corresponding to unbound ligand state (basin **III**) are widespread, indicating no preferred configuration once the ligand is away from the protein.

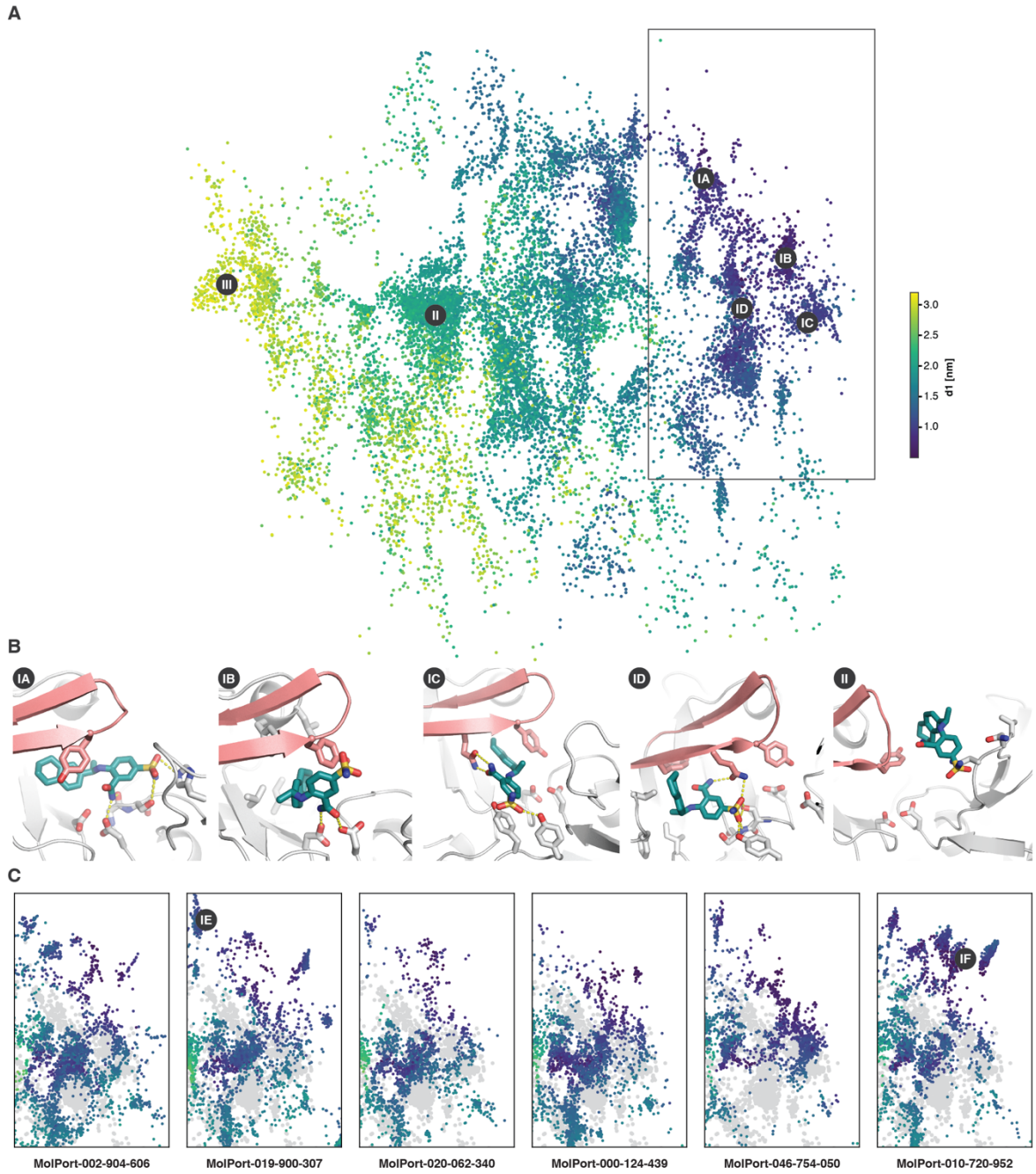

Fig. 6. ***A*** *Sketch-map of compound* ***1*** *binding to plm V was obtained by mapping the binding trajectory's three-dimensional space to 2D representation. The map is colour coded according to ligand core-catalytic dyad distance d1. The key states along the binding pathway are indicated as IA, IB, IC, ID and II, and the unbound state as III. The rectangle indicates the region where the inhibitor interacts with the binding site residues and is compared in subfigure C.* ***B*** *The corresponding binding modes along the binding pathway. The ligand is shown as green sticks, selected binding pocket residues and catalytic dyad residues are shown as grey sticks, and the flap loop is in salmon. Hydrogens are omitted for clarity.* ***C*** *Comparison of experimentally verified plm V inhibitor binding sketch-maps and reference map shown in subfigure A. Grey – reference sketch-map from subfigure A. Binding states not observed in the reference map are indicated as IE and IF (respective binding modes are shown in Fig. SI2). The colour coding matches subfigure A.*

Similar binding pathways are expected for all inhibitors identified. Therefore, trajectories of verified plm V inhibitors were projected on top of the reference sketch map discussed before (see Fig. 6.C). All inhibitors have several binding modes, all captured in the reference sketch

map. However, the population of each of them varies from ligand to ligand. The most populated basin for all ligands except MolPort-046-754-050 was **ID**, where the inhibitor is situated right next to the flap loop in a hydrophobic groove. For MolPort-046-754-050, the most populated were basins **IB** and **IC**, where ligand flap loop substituent is located under the flap and fills the S3 pocket. Basin **IC** was underrepresented in most ligands compared to the reference system, most likely because only the reference system ligand (comp. **1**) can simultaneously interact with flap loop Gln173 and nearby Tyr286. The basins of configurations corresponding to docked binding modes (**IA**) are slightly shifted towards the top right corner of the map for most of the inhibitors, indicating minor differences between the reference and other systems. Indeed, these binding modes are similar to the docked ones. However, the flap loop substituent is slightly twisted to partly interact in the hydrophobic region right next to the flap. Besides binding modes observed for the reference system, an additional scarcely populated binding mode was observed for most inhibitors (basin **IE** in Fig. 6.C). This binding mode corresponds to the transition state where the inhibitor core interacts with Gln183 and/or Gln184 in a substrate binding groove. Moreover, another binding mode was identified for MolPort-010-720-952, where the inhibitor transition state mimetic group is wedged between the flap loop and nearby α helix (basin **IF**). Here inhibitor hydrogen bond donor and acceptor group forms hydrogen bonds with His173 and Glu176, and the flap pocket substituent is involved in aromatic stacking with Phe180 (see Fig. SI2).[51]

Conclusions

We have presented here a method for inhibitor binding free energy estimation and putative transition state identification. The method was applied to characterise compound binding to validated antimalarial drug target *Plasmodium falciparum* aspartic protease plasmepsin V. The proposed approach utilises path metadynamics, where the process of interest is studied by exploring the space of unbinding paths starting from a predefined reference one. Here, the reference path that describes inhibitor binding contains not only the coordinates of the potential inhibitor but also the coordinates of selected binding site residue atoms to sample relevant target conformations. This is one of the most significant advantages of the proposed method, as it accurately characterises binding to flexible targets, which is typically problematic using free energy perturbation-based methods.

Furthermore, metadynamics CV selection and parameter optimisation usually are tedious and time-consuming. Therefore, the fact that the proposed approach allows the same binding path for all compounds of interest greatly simplifies the simulation setup. The method also enables sampling complete inhibitor binding/unbinding pathways, including relevant transition states that might provide additional information for targeted lead compound development.

Moreover, we also demonstrated an approach to identify putative transition states that are shared between various structurally unrelated inhibitors (or are unique to a specific inhibitor) using the dimensionality reduction method sketch-map. This could be particularly useful in lead compound selection to identify inhibitors that meet a specific requirement. In addition to demonstrating the method's applicability, the study provides new, verified plasmepsin V inhibitor scaffolds that can be further developed into potential antimalarials. Overall, our results suggest that it is possible to use a path metadynamics-based approach to study systems where inhibitor binding is associated with considerable conformational changes. This technique can be easily deployed to determine accurate binding free energies. We believe that the approach described might prove useful in drug development pipelines where the binding of a large set of structurally diverse ligands needs to be characterised.

Supporting Information
The Supporting Information contains additional figures illustrating parameters used for MetaD simulation reweighing, relevant inhibitor-protein binding modes, enzymatic activity data, computational and experimental methods used.

Notes
Metadynamics simulations were performed using open-source software Gromacs2021 (https://www.gromacs.org/) patched with Plumed2.7.2 (https://www.plumed.org/). Plumed input files used in this work are available via PLUMED-NEST[52] (https://www.plumed-nest.org), the public repository for the Plumed consortium, using the project plumID: 23.019. Protein structures used in this study were retrieved from the RCSB protein data bank (https://www.rcsb.org/). Figures were prepared using PyMOL2.5.2 (https://pymol.org/2/), and data were plotted using matplotlib (https://matplotlib.org/).

Acknowledgements
R.B. acknowledges the European Regional Development Fund Project No. 1.1.1.2/VIAA/2/18/379 for financial support. Calculations were partly performed in Riga Technical University (RTU) High-Performance Computing (HPC) Centre facilities. Gareth Tribello is acknowledged for his help with the sketch-map calculation setup.

References
(1) Suryanarayanan, V.; Panwar, U.; Chandra, I. *Computational Drug Discovery and Design*; 2018; Vol. 1762. https://doi.org/10.1007/978-1-4939-7756-7.
(2) Kuhn, B.; Guba, W.; Hert, J.; Banner, D.; Bissantz, C.; Ceccarelli, S.; Haap, W.; Körner, M.; Kuglstatter, A.; Lerner, C.; Mattei, P.; Neidhart, W.; Pinard, E.; Rudolph, M. G.; Schulz-Gasch, T.; Woltering, T.; Stahl, M. A Real-World Perspective on Molecular Design. *J Med Chem* **2016**, *59* (9), 4087–4102. https://doi.org/10.1021/acs.jmedchem.5b01875.
(3) Wermuth, C. G. *The Practice of Medicinal Chemistry, 3rd Edition*, 3rd ed.; Wermuth, C. G., Ed.; Academic Press, 2008.
(4) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindah, E. Gromacs: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25. https://doi.org/10.1016/j.softx.2015.06.001.
(5) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc Natl Acad Sci U S A* **2002**, *99* (20), 12562–12566. https://doi.org/10.1073/pnas.202427399.
(6) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J Comput Chem* **2004**, *25* (9), 1157–1174. https://doi.org/10.1002/jcc.20035.
(7) Limongelli, V.; Bonomi, M.; Parrinello, M. Funnel Metadynamics as Accurate Binding Free-Energy Method. *Proc Natl Acad Sci U S A* **2013**, *110* (16), 6358–6363. https://doi.org/10.1073/pnas.1303186110.
(8) Williams-Noonan, B. J.; Yuriev, E.; Chalmers, D. K. Free Energy Methods in Drug Design: Prospects of “Alchemical Perturbation” in Medicinal Chemistry. *J Med Chem* **2018**, *61* (3), 638–649. https://doi.org/10.1021/acs.jmedchem.7b00681.
(9) Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; Romero, D. L.; Masse, C.; Knight, J. L.; Steinbrecher, T.; Beuming, T.; Damm, W.; Harder, E.; Sherman, W.; Brewer, M.; Wester, R.; Murcko, M.; Frye, L.; Farid, R.; Lin, T.; Mobley, D. L.; Jorgensen, W. L.; Berne, B. J.; Friesner, R. A.; Abel, R. Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. *J Am Chem Soc* **2015**, *137* (7), 2695–2703. https://doi.org/10.1021/ja512751q.

(10) Rizzi, A.; Murkli, S.; McNeill, J. N.; Yao, W.; Sullivan, M.; Gilson, M. K.; Chiu, M. W.; Isaacs, L.; Gibb, B. C.; Mobley, D. L.; Chodera, J. D. Overview of the SAMPL6 Host–Guest Binding Affinity Prediction Challenge. *J Comput Aided Mol Des* **2018**, *32* (10), 937–963. https://doi.org/10.1007/s10822-018-0170-6.
(11) Michel, J.; Essex, J. W. Prediction of Protein–Ligand Binding Affinity by Free Energy Simulations: Assumptions, Pitfalls and Expectations. *J Comput Aided Mol Des* **2010**, *24* (8), 639–658. https://doi.org/10.1007/s10822-010-9363-3.
(12) Lapelosa, M.; Gallicchio, E.; Levy, R. M. Conformational Transitions and Convergence of Absolute Binding Free Energy Calculations. *J Chem Theory Comput* **2012**, *8* (1), 47–60. https://doi.org/10.1021/ct200684b.
(13) Evans, R.; Hovan, L.; Tribello, G. A.; Cossins, B. P.; Estarellas, C.; Gervasio, F. L. Combining Machine Learning and Enhanced Sampling Techniques for Efficient and Accurate Calculation of Absolute Binding Free Energies. *J Chem Theory Comput* **2020**, *16* (7), 4641–4654. https://doi.org/10.1021/acs.jctc.0c00075.
(14) Bernetti, M.; Masetti, M.; Recanatini, M.; Amaro, R. E.; Cavalli, A. An Integrated Markov State Model and Path Metadynamics Approach to Characterize Drug Binding Processes. *J Chem Theory Comput* **2019**, *15* (10), 5689–5702. https://doi.org/10.1021/acs.jctc.9b00450.
(15) Raniolo, S.; Limongelli, V. Ligand Binding Free-Energy Calculations with Funnel Metadynamics. *Nat Protoc* **2020**, *15* (9), 2837–2866. https://doi.org/10.1038/s41596-020-0342-4.
(16) Deganutti, G.; Moro, S.; Reynolds, C. A. A Supervised Molecular Dynamics Approach to Unbiased Ligand–Protein Unbinding. *J Chem Inf Model* **2020**, acs.jcim.9b01094. https://doi.org/10.1021/acs.jcim.9b01094.
(17) Evenseth, L. S. M.; Ocello, R.; Gabrielsen, M.; Masetti, M.; Recanatini, M.; Sylte, I.; Cavalli, A. Exploring Conformational Dynamics of the Extracellular Venus Flytrap Domain of the GABAB Receptor: A Path-Metadynamics Study. *J Chem Inf Model* **2020**, *60* (4), 2294–2303. https://doi.org/10.1021/acs.jcim.0c00163.
(18) Bobrovs, R.; Basens, E. E.; Drunka, L.; Kanepe, I.; Matisone, S.; Velins, K. K.; Andrianov, V.; Leitis, G.; Zelencova-Gopejenko, D.; Rasina, D.; Jirgensons, A.; Jaudzems, K. Exploring Aspartic Protease Inhibitor Binding to Design Selective Antimalarials. *J Chem Inf Model* **2022**, *62* (13), 3263–3273. https://doi.org/10.1021/acs.jcim.2c00422.
(19) Mattedi, G.; Deflorian, F.; Mason, J. S.; De Graaf, C.; Gervasio, F. L. Understanding Ligand Binding Selectivity in a Prototypical GPCR Family. *J Chem Inf Model* **2019**, *59* (6), 2830–2836. https://doi.org/10.1021/acs.jcim.9b00298.
(20) Cavalli, A.; Spitaleri, A.; Saladino, G.; Gervasio, F. L. Investigating Drug-Target Association and Dissociation Mechanisms Using Metadynamics-Based Algorithms. *Acc Chem Res* **2015**, *48* (2), 277–285. https://doi.org/10.1021/ar500356n.
(21) Callea, L.; Bonati, L.; Motta, S. Metadynamics-Based Approaches for Modeling the Hypoxia-Inducible Factor 2α Ligand Binding Process. *J Chem Theory Comput* **2021**, *17* (7), 3841–3851. https://doi.org/10.1021/acs.jctc.1c00114.
(22) Tiwary, P.; Limongelli, V.; Salvalaglio, M.; Parrinello, M. Kinetics of Protein-Ligand Unbinding: Predicting Pathways, Rates, and Rate-Limiting Steps. *Proc Natl Acad Sci U S A* **2015**, *112* (5), E386–E391. https://doi.org/10.1073/pnas.1424461112.
(23) Tiwary, P.; Parrinello, M. From Metadynamics to Dynamics. *Phys Rev Lett* **2013**, *111* (23), 1–5. https://doi.org/10.1103/PhysRevLett.111.230602.
(24) Bertazzo, M.; Gobbo, D.; Decherchi, S.; Cavalli, A. Machine Learning and Enhanced Sampling Simulations for Computing the Potential of Mean Force and Standard Binding Free Energy. *J Chem Theory Comput* **2021**, *17* (8), 5287–5300. https://doi.org/10.1021/acs.jctc.1c00177.

(25) Lotz, S. D.; Dickson, A. Unbiased Molecular Dynamics of 11 Min Timescale Drug Unbinding Reveals Transition State Stabilizing Interactions. *J Am Chem Soc* **2018**, *140* (2), 618–628. https://doi.org/10.1021/jacs.7b08572.
(26) Tiwary, P.; Mondal, J.; Berne, B. J. How and When Does an Anticancer Drug Leave Its Binding Site? *Sci Adv* **2017**, *3* (5), 1–8. https://doi.org/10.1126/sciadv.1700014.
(27) Christopoulos, A. Allosteric Binding Sites on Cell-Surface Receptors: Novel Targets for Drug Discovery. *Nat Rev Drug Discov* **2002**, *1* (3), 198–210. https://doi.org/10.1038/nrd746.
(28) Durrant, J.; McCammon, J. A. Molecular Dynamics Simulations and Drug Discovery. *BMC Biol* **2011**, *9* (71), 1–9. https://doi.org/10.1002/cber.18940270364.
(29) Goldberg, D. E. Hemoglobin Degradation. *Curr Top Microbiol Immunol* **2005**, *295*, 275–291. https://doi.org/10.1007/3-540-29088-5_11.
(30) Sleebs, B. E.; Lopaticki, S.; Marapana, D. S.; O'Neill, M. T.; Rajasekaran, P.; Gazdik, M.; Günther, S.; Whitehead, L. W.; Lowes, K. N.; Barfod, L.; Hviid, L.; Shaw, P. J.; Hodder, A. N.; Smith, B. J.; Cowman, A. F.; Boddey, J. A. Inhibition of Plasmepsin V Activity Demonstrates Its Essential Role in Protein Export, PfEMP1 Display, and Survival of Malaria Parasites. *PLoS Biol* **2014**, *12* (7), e1001897. https://doi.org/10.1371/journal.pbio.1001897.
(31) Russo, I.; Babbitt, S.; Muralidharan, V.; Butler, T.; Oksman, A.; Goldberg, D. E. Plasmepsin v Licenses Plasmodium Proteins for Export into the Host Erythrocyte. *Nature* **2010**, *463* (7281), 632–636. https://doi.org/10.1038/nature08726.
(32) Boddey, J. A.; Hodder, A. N.; Günther, S.; Gilson, P. R.; Patsiouras, H.; Kapp, E. A.; Pearce, J. A.; De Koning-Ward, T. F.; Simpson, R. J.; Crabb, B. S.; Cowman, A. F. An Aspartyl Protease Directs Malaria Effector Proteins to the Host Cell. *Nature* **2010**, *463* (7281), 627–631. https://doi.org/10.1038/nature08728.
(33) Sleebs, B. E.; Gazdik, M.; O'Neill, M. T.; Rajasekaran, P.; Lopaticki, S.; Lackovic, K.; Lowes, K.; Smith, B. J.; Cowman, A. F.; Boddey, J. a. Transition State Mimetics of the *Plasmodium* Export Element Are Potent Inhibitors of Plasmepsin V from *P. Falciparum* and *P. Vivax*. *J Med Chem* **2014**, *57* (18), 7644–7662. https://doi.org/10.1021/jm500797g.
(34) Nguyen, W.; Hodder, A. N.; de Lezongard, R. B.; Czabotar, P. E.; Jarman, K. E.; O'Neill, M. T.; Thompson, J. K.; Jousset Sabroux, H.; Cowman, A. F.; Boddey, J. A.; Sleebs, B. E. Enhanced Antimalarial Activity of Plasmepsin V Inhibitors by Modification of the P2position of PEXEL Peptidomimetics. *Eur J Med Chem* **2018**, *154*, 182–198. https://doi.org/10.1016/j.ejmech.2018.05.022.
(35) Hodder, A. N.; Sleebs, B. E.; Czabotar, P. E.; Gazdik, M.; Xu, Y.; O'Neill, M. T.; Lopaticki, S.; Nebl, T.; Triglia, T.; Smith, B. J.; Lowes, K.; Boddey, J. A.; Cowman, A. F. Structural Basis for Plasmepsin v Inhibition That Blocks Export of Malaria Proteins to Human Erythrocytes. *Nat Struct Mol Biol* **2015**, *22* (8), 590–596. https://doi.org/10.1038/nsmb.3061.
(36) Gambini, L.; Rizzi, L.; Pedretti, A.; Taglialatela-Scafati, O.; Carucci, M.; Pancotti, A.; Galli, C.; Read, M.; Giurisato, E.; Romeo, S.; Russo, I. Picomolar Inhibition of Plasmepsin V, an Essential Malaria Protease, Achieved Exploiting the Prime Region. *PLoS One* **2015**, *10* (11), 1–35. https://doi.org/10.1371/journal.pone.0142509.
(37) Gazdik, M.; O'Neill, M. T.; Lopaticki, S.; Lowes, K. N.; Smith, B. J.; Cowman, A. F.; Boddey, J. A.; Sleebs, B. E. The Effect of N-Methylation on Transition State Mimetic Inhibitors of the Plasmodium Protease, Plasmepsin V. *Med. Chem. Commun.* **2015**, *6* (3), 437–443. https://doi.org/10.1039/C4MD00409D.
(38) Bobrovs, R.; Jaudzems, K.; Jirgensons, A. Exploiting Structural Dynamics To Design Open-Flap Inhibitors of Malarial Aspartic Proteases. *J Med Chem* **2019**, *62* (20), 8931–8950. https://doi.org/10.1021/acs.jmedchem.9b00184.

(39) Rasina, D.; Stakanovs, G.; Kanepe-Lapsa, I.; Bobrovs, R.; Jaudzems, K.; Jirgensons, A. Synthesis of 2-Aminopyridopyrimidinones and Their Plasmepsin I, II, IV Inhibition Potency. *Chem Heterocycl Compd (N Y)* **2020**, *56* (6), 786–792. https://doi.org/10.1007/s10593-020-02731-3.
(40) Zogota, R.; Kinena, L.; Withers-Martinez, C.; Blackman, M. J.; Bobrovs, R.; Pantelejevs, T.; Kanepe-Lapsa, I.; Ozola, V.; Jaudzems, K.; Suna, E.; Jirgensons, A. Peptidomimetic Plasmepsin Inhibitors with Potent Anti-Malarial Activity and Selectivity against Cathepsin D. *Eur J Med Chem* **2019**, *163*, 344–352. https://doi.org/10.1016/j.ejmech.2018.11.068.
(41) Kinena, L.; Leitis, G.; Kanepe-Lapsa, I.; Bobrovs, R.; Jaudzems, K.; Ozola, V.; Suna, E.; Jirgensons, A. Azole-Based Non-Peptidomimetic Plasmepsin Inhibitors. *Arch Pharm (Weinheim)* **2018**, No. May, 1–24. https://doi.org/10.1002/ardp.201800151.
(42) Rasina, D.; Stakanovs, G.; Borysov, O. V.; Pantelejevs, T.; Bobrovs, R.; Kanepe-Lapsa, I.; Tars, K.; Jaudzems, K.; Jirgensons, A. 2-Aminoquinazolin-4(3H)-One Based Plasmepsin Inhibitors with Improved Hydrophilicity and Selectivity. *Bioorg Med Chem* **2018**, *26* (9), 2488–2500. https://doi.org/10.1016/j.bmc.2018.04.012.
(43) Rasina, D.; Otikovs, M.; Leitans, J.; Recacha, R.; Borysov, O. V.; Kanepe-Lapsa, I.; Domraceva, I.; Pantelejevs, T.; Tars, K.; Blackman, M. J.; Jaudzems, K.; Jirgensons, A. Fragment-Based Discovery of 2-Aminoquinazolin-4(3H)-Ones As Novel Class Nonpeptidomimetic Inhibitors of the Plasmepsins I, II, and IV. *J Med Chem* **2016**, *59* (1), 374–387. https://doi.org/10.1021/acs.jmedchem.5b01558.
(44) Friedman, R.; Caflisch, A. The Protonation State of the Catalytic Aspartates in Plasmepsin II. *FEBS Lett* **2007**, *581* (21), 4120–4124. https://doi.org/10.1016/j.febslet.2007.07.033.
(45) Bobrovs, R.; Basens, E. E.; Drunka, L.; Kanepe, I.; Matisone, S.; Velins, K. K.; Andrianov, V.; Leitis, G.; Zelencova-Gopejenko, D.; Rasina, D.; Jirgensons, A.; Jaudzems, K. Exploring Aspartic Protease Inhibitor Binding to Design Selective Antimalarials. *J Chem Inf Model* **2022**. https://doi.org/10.1021/acs.jcim.2c00422.
(46) Geschwindner, S.; Olsson, L.-L.; Albert, J. S.; Deinum, J.; Edwards, P. D.; de Beer, T.; Folmer, R. H. A. Discovery of a Novel Warhead against β-Secretase through Fragment-Based Lead Generation. *J Med Chem* **2007**, *50* (24), 5903–5911. https://doi.org/10.1021/jm070825k.
(47) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New Feathers for an Old Bird. *Comput Phys Commun* **2014**, *185* (2), 604–613. https://doi.org/10.1016/j.cpc.2013.09.018.
(48) Tiwary, P.; Parrinello, M. A Time-Independent Free Energy Estimator for Metadynamics. *Journal of Physical Chemistry B* **2015**, *119* (3), 736–742. https://doi.org/10.1021/jp504920s.
(49) Gapsys, V.; Pérez-Benito, L.; Aldeghi, M.; Seeliger, D.; Van Vlijmen, H.; Tresadern, G.; De Groot, B. L. Large Scale Relative Protein Ligand Binding Affinities Using Non-Equilibrium Alchemy. *Chem Sci* **2020**, *11* (4), 1140–1152. https://doi.org/10.1039/c9sc03754c.
(50) Ceriotti, M.; Tribello, G. A.; Parrinello, M. Simplifying the Representation of Complex Free-Energy Landscapes Using Sketch-Map. *Proc Natl Acad Sci U S A* **2011**, *108* (32), 13023–13028. https://doi.org/10.1073/pnas.1108486108.
(51) Tribello, G. A.; Ceriotti, M.; Parrinello, M. Using Sketch-Map Coordinates to Analyze and Bias Molecular Dynamics Simulations. *Proc Natl Acad Sci U S A* **2012**, *109* (14), 5196–5201. https://doi.org/10.1073/pnas.1201152109.
(52) Bonomi, M.; Bussi, G.; Camilloni, C.; Tribello, G. A.; Banáš, P.; Barducci, A.; Bernetti, M.; Bolhuis, P. G.; Bottaro, S.; Branduardi, D.; Capelli, R.; Carloni, P.; Ceriotti, M.; Cesari, A.; Chen, H.; Chen, W.; Colizzi, F.; De, S.; De La Pierre, M.; Donadio, D.; Drobot, V.; Ensing, B.; Ferguson, A. L.; Filizola, M.; Fraser, J. S.; Fu, H.; Gasparotto,

P.; Gervasio, F. L.; Giberti, F.; Gil-Ley, A.; Giorgino, T.; Heller, G. T.; Hocky, G. M.; Iannuzzi, M.; Invernizzi, M.; Jelfs, K. E.; Jussupow, A.; Kirilin, E.; Laio, A.; Limongelli, V.; Lindorff-Larsen, K.; Löhr, T.; Marinelli, F.; Martin-Samos, L.; Masetti, M.; Meyer, R.; Michaelides, A.; Molteni, C.; Morishita, T.; Nava, M.; Paissoni, C.; Papaleo, E.; Parrinello, M.; Pfaendtner, J.; Piaggi, P.; Piccini, G.; Pietropaolo, A.; Pietrucci, F.; Pipolo, S.; Provasi, D.; Quigley, D.; Raiteri, P.; Raniolo, S.; Rydzewski, J.; Salvalaglio, M.; Sosso, G. C.; Spiwok, V.; Šponer, J.; Swenson, D. W. H.; Tiwary, P.; Valsson, O.; Vendruscolo, M.; Voth, G. A.; White, A. Promoting Transparency and Reproducibility in Enhanced Molecular Simulations. *Nat Methods* **2019**, *16* (8), 670–673. https://doi.org/10.1038/s41592-019-0506-8.

For Table of Contents Only