

In silico discovery and biological validation of ligands of FAD synthase, a promising new antimicrobial target

Isaias Lans¹, Ernesto Anoz-Carbonell^{2,3,4}, Karen Palacio-Rodríguez¹, José Antonio Aínsa^{3,4,5}, Milagros Medina^{2,3}, and Pilar Cossio^{1,6,*}

1 Biophysics of Tropical Diseases, Max Planck Tandem Group, University of Antioquia, 050010 Medellín, Colombia.

2 Departamento de Bioquímica y Biología Molecular y Celular, Facultad de Ciencias, Universidad de Zaragoza, Spain

3 Instituto de Biocomputación y Física de Sistemas Complejos (Unidades Asociadas BIFI-IQFR y CBsC-CSIC), Universidad de Zaragoza, Spain

4 Grupo de Genética de Micobacterias, Departamento de Microbiología, Pediatría, Radiología y Salud Pública. Facultad de Medicina, Universidad de Zaragoza, Zaragoza, Spain.

5 CIBER Enfermedades Respiratorias (CIBERES), Instituto de Salud Carlos III, Spain

6 Department of Theoretical Biophysics, Max Planck Institute of Biophysics, 60438 Frankfurt am Main, Germany.

* picossio@biophys.mpg.de

Abstract

New treatments for diseases caused by antimicrobial-resistant microorganisms can be developed by identifying unexplored therapeutic targets and by designing efficient drug screening protocols. In this study, we have screened a library of compounds to find ligands for the flavin-adenine dinucleotide synthase (FADS) -a potential target for drug design against tuberculosis and pneumonia- by implementing a new and efficient virtual screening protocol. The protocol has been developed for the *in silico* search of ligands of unexplored therapeutic targets, for which limited information about ligands or ligand-receptor structures is available. It implements an integrative funnel-like strategy with filtering layers that increase in computational accuracy. The protocol starts with a pharmacophore-based virtual screening strategy that uses ligand-free receptor conformations from molecular dynamics (MD) simulations. Then, it performs a molecular docking stage using several docking programs and an exponential consensus ranking strategy. The last filter, samples the conformations of compounds bound to the target using MD simulations. The MD conformations are scored using several traditional scoring functions in combination with a newly-proposed score that takes into account the fluctuations of the molecule with a Morse-based potential. The protocol was optimized and validated using a compound library with known ligands of the *Corynebacterium ammoniagenes* FADS. Then, it was used to find new FADS ligands from a compound library of 14,000 molecules. A small set of 17 *in silico* filtered molecules were tested experimentally. We identified five inhibitors of the activity of the flavin adenyl transferase mononucleotide of the FADS, and some of them were able to inhibit growth of three bacterial species: *Corynebacterium ammoniagenes*, *Mycobacterium tuberculosis*, and *Streptococcus pneumoniae*, where the former two are human pathogens. Overall, the results show that the integrative VS protocol is a cost-effective solution for the discovery of ligands of unexplored therapeutic targets.

Author summary

Developing cures for antimicrobial-resistant microorganisms is a pressing necessity. Addressing this problem requires the discovery of novel therapeutic targets -for example, bacterial proteins with no human homologues- and the development of cost-effective drug screening protocols. In this work, we tackled the problem on both sides. We developed an efficient and successful integrative computational protocol for screening inhibitory-molecules for unexplored targets. We used it to discover five novel inhibitors of flavin-adenine dinucleotide synthase (FADS), a promising protein target of pathogens causing *tuberculosis* and *pneumonia*.

Introduction

Given the growing incidence of infections caused by antimicrobial resistant pathogens, international institutions, such as the World Health Organization [1], have informed about the lack of potential therapeutic options for these pathogens, and have named a list of pathogens for which is critical to develop novel antimicrobial agents. The development of such treatments should involve efficient drug design protocols and the discovery of new molecular targets to fight antimicrobial resistance. A straightforward and effective way to increase the chances of success of the drug-screening pipeline is through the implementation of efficient virtual screening (VS) protocols. These methods provide powerful tools to reduce the costs of drug discovery by reducing the number of compounds to be tested in experimental trials [2–5]. Moreover, VS protocols increase the success rate (*i.e.*, active compounds found) and reduce the false negatives in high-throughput compound screening [2,3,6–12].

Efficient VS protocols have to be able to screen large compound libraries in short computational times. Therefore, these protocols usually implement a funnel-like strategy, which start from fast but less accurate methods (where a large number of molecules are filtered) and more accurate and time-consuming tools are used in the last steps [13–15]. Usually, pharmacophore-based tools are implemented in the first stages of VS, given their ability to quickly screen large compounds libraries [16–18]. While more sophisticated tools such as docking or molecular dynamics (MD) are implemented in the latter steps to predict ligand affinities [11,13,19–26].

Special attention deserve the tools used in the first steps of the VS, because these impact the ability of the protocol to explore large compound libraries and the chemical space of the compounds, such as the pharmacophore-based strategies [16–18,27–30]. Despite the usefulness of these methods, which accelerate the first steps of the VS, these strategies have some limitations. Several pharmacophore-based methods require knowing ligand or ligand-receptor structures for their training, limiting the chemical space of the filtered molecules to that associated with the training set [16]. Recently, the flexi-pharma method, a VS strategy that uses pharmacophores from MD conformations was developed to overcome these limitations [31]. However, in general, new protein targets, which have limited structural information available, such as the bifunctional enzyme flavin-adenine dinucleotide synthase (FADS), are challenging for the funnel-like VS strategies.

FADS is a potential target for drug design against antimicrobial-resistant organisms, such as the human pathogens *Mycobacterium tuberculosis* and *Streptococcus pneumoniae*. FADS is a bi-functional and bi-modular enzyme that catalyzes the synthesis of two essential co-factors: flavin mononucleotide (FMN) and flavin-adenine dinucleotide (FAD). These are essential for a large number of proteins participating in energy transformation or metabolic processes, in prokaryotic and eukaryotic organisms [32–34]. For FADS, the synthesis of FMN and FAD occurs at the C-terminal module (the RKF

module) and N-terminal module (FMNAT module) of FADS, respectively [35–38]. The RFK module has a similar structure and sequence to the equivalent enzyme in eukaryotic organisms. However, because the FMNAT module lacks both sequence and structural similarity with the equivalent enzymes in eukaryotic systems, the prokaryotic FADSs have emerged as potential antimicrobial targets [35, 39–41].

The most characterized FADS is the enzyme of *Corynebacterium ammoniagenes* (*CaFADS*), which is considered a good representative model for the FADS of the human pathogen *Mycobacterium tuberculosis* (*MtFADS*) [40, 41]. However, limited structural information about this enzyme is available. Moreover, no experimental structures of the FADS-FMNAT module in complex with substrates are reported. The only ligand-*CaFADS* FMNAT module structures correspond to theoretical or computational models [35, 42].

The objective of this work is to discover molecules able to inhibit the FMNAT activity of FADS using a novel integrative VS protocol. This promising protocol addresses some of the limitations found with traditional VS, for example, it does not require knowledge of ligands or ligand-receptor structures, and its attributes enable a better exploration of the chemical space of large compound libraries.

This manuscript is organized as follows. First, we describe the integrative computational protocol, which includes several filtering layers: *i*) flexi-pharma screening [31], *ii*) consensus docking screening [43], *iii*) MD sampling and scoring, and *iv*) compound activities measured by experimental assays. The protocol is tested and optimized using a library of 1993 compounds from which 39 compounds are true ligands of the *CaFADS* [41]. Subsequently, the optimized protocol is implemented over a library of 14000 compounds. A final list of 17 filtered compounds is tested experimentally. We discover that five molecules are able to inhibit the FMNAT-FADS activity, five bind to the FMNAT-FADS and five present growth inhibitory activity against *Corynebacterium amoniaenes*, *Mycobacterium tuberculosis* or *Streptococcus pneumoniae*. We conclude that the VS protocol and the new inhibitory compounds can contribute to further development of novel therapeutic strategies against antimicrobial-resistant pathogens such as *Mycobacterium tuberculosis* and *Streptococcus pneumoniae*.

Materials and methods

Computational Methods

FAD structure

The Cartesian coordinates of the FMNAT module (M1-H186) of *CaFADS* were taken from the crystal structure with PDB code 2X0K [35]. These were used for the MD simulations in the flexi-pharma or molecular docking stage.

Flexi-pharma method

The flexi-pharma protocol [31] has three substages: run an MD simulation of the ligand-free receptor target, generate a set of pharmacophores from each MD conformation, and assign a vote to each molecule every time it matches at least one pharmacophore from each MD conformation.

First, an MD simulation of ligand-free receptor is performed. To generate the pharmacophore set from each MD conformation, we use Autogrid4.2 [44] to calculate the affinity maps of several atom-types: hydrogen bond donor, hydrogen bond acceptor, hydrophobic, aromatic and charged atoms. Some atom-type affinity grids are first discarded if they show a flat distribution of the affinity values of the map (for this work, the affinity maps that have a histogram with kurtosis larger than 3). Then, we define a

grid-percentage threshold to determine the hotspots (clusters of grid-cells) for each atom type. The threshold is a percentage of the total number of cells in the grid with negative affinity energy. We cluster the selected cells to generate a pharmacophoric feature (given by a center, a radius of gyration, an atom type and in some cases a direction). A pharmacophore is built by combining three features. The pharmacophore set consists of all possible combinations of triplets of features from different active spaces (*i.e.*, centers of the affinity grids) together with a volume exclusion term. The Pharmer [45] program is used to screen the compound library with the created pharmacophore set.

Flexi-pharma gives a score to each compound by means of a voting strategy. If any molecule, from the compounds library, matches any pharmacophore obtained from a specific MD frame, then the molecule obtains a vote. The total number of votes is used as a score of the molecule. For more details about the flexi-pharma method see ref. [31].

Molecular docking

The molecular docking was carried out using the programs: Autodock4.2 [44, 46, 47], Vina [48], and Smina [49]. For these programs, the sampling space was defined using a grid box of $15 \times 15 \times 15$ Å centered at the oxygen of the amide group of the catalytic residue ASN125 [50]. The number of requested poses was 50.

Autodock4.2 [44, 46, 47] was used with a grid spacing of 0.25 Å. The search was performed using the Lamarckian genetic algorithm implemented in Autodock, with a starting population of 50 individuals, using 25000000 energy evaluations and 27000 generations. The resultant poses were clustered using the RMSD of the atomic positions, with a tolerance of 2.0 Å, using the default clustering method. In addition, to the sampling space and the number of poses, Vina [48] and Smina [49] were used with the default parameters. For Smina, the Vinardo [51] scoring function was used.

Molecular dynamics simulations

The best pose for each compound from the docking stage, obtained with the Autodock4.2 program, was used as the initial conformation for the MD simulation. The compounds were protonated, at pH 7.0, using Open Babel software [52]. The PROPKA [53] module from the PDB2PQR software package [54, 55] was used to determine the protonation state of all ionizable groups at pH 7.0. The final models were solvated with a dodecahedral water box, centered at the geometric center of the complex. To neutralize the systems, Na⁺ ions were added when necessary. The AMBER99SB-ILDN [56] force field was used to model the protein with the TIP3P water model [57]. The GAFF force field [58] parameters were obtained for the compounds using Antechamber [58, 59]. ACPYPE [60] was used to change the topology files from amber to GROMACS [61, 62], which was used for all the MD simulations. The systems were minimized until the maximum force was ≤ 1000 kJ/mol-nm with the steepest descent algorithm. MD simulations were carried out with periodic boundary conditions. A spherical cutoff of 1.2 nm for the non-bonded interactions was applied together with a switch function acting between 1.0 and 1.2 nm. The non-bonded pair list was updated every 20 steps. The particle mesh Ewald method was used to compute long-range electrostatic force terms, and the leapfrog algorithm to propagate the equations of motion. All bond lengths and angles involving hydrogen atoms were constrained using the LINCS algorithm [63]. Equilibration consisted of 100 ps of NVT followed by 100 ps of NPT simulation at 310 K, with a time step of 2 fs was done. During equilibration the coordinates of protein and of ligand heavy atoms were restrained using a constant force of 100 kJ/mol-nm. Finally, MD simulations between 5 - 15 ns were carried out using the GROMACS 5.1.3 program [61, 62] with a time step of 2 fs, without restraints, in an isothermal-isobaric (NPT) ensemble at 310.15 K and 1 atm.

Exponential consensus ranking

A consensus methodology was used to combine the results from different scoring functions both for the docking and MD VS stages. We used an exponential consensus ranking (ECR) methodology [43]. This method assigns a score $p(r_i^j)$ to each molecule i for each scoring function j using an exponential function $p(r_i^j) = \exp(-r_i^j/\alpha)/\alpha$, which depends on the rank of the molecule (r_i^j) given by each individual docking program. α is the expected value of the exponential distribution, which we have set to 50% of the total molecules at each stage. The final score $P(i)$ is defined as the sum of the exponential functions for all of the programs, $P(i) = \sum_j p(r_i^j) = \frac{1}{\alpha} \sum_j \exp(-r_i^j/\alpha)$.

Validation metrics

The enrichment factor (EF_{x%}) is a measure of the change on the ligand/decoys proportion in a molecular dataset, after filtering it to the x%. EF_{x%} is defined as the ratio between ligands (Hits) found at a certain threshold (x%) of the best ranked compounds and the number of compounds at that threshold (N_{x%}) normalized by the ratio between the hits contained in the entire dataset (Hits_{100%}) and the total number of compounds N_{100%}:

$$EF_x = \frac{Hits^{x\%}}{N^{x\%}} \times \frac{N^{100\%}}{Hits^{100\%}}. \quad (1)$$

Values of EF_{x%} higher than 1 indicate an enrichment of the compound library.

The enrichment plot (EP) measures the performance of a filtering method at different levels of a compound library reduction. In an EP, the percentage of ligands found in the top x% of ranked compounds vs the top x% of filtered compounds is plotted [64].

To assess the error of the flexi-pharma EPs, a bootstrapping analysis with replacement was used. The selected MD frames were iteratively re-sampled with replacement 100 times. Thus, 100 EPs were obtained for each trajectory. From these the average and the error of the EPs were calculated (similarly as in ref. 31).

Experimental Methods

Chemicals

The selected compounds were acquired from Molport and dissolved in 100% DMSO to prepare stock solutions at 50 mM and 10 mM. According to the manufacturer indications, the purity of the compounds was >95%, and had been determined by high performance liquid chromatography (HPLC), thin layer chromatography (TLC), NMR, IR or basic titration.

Protein purification and quantification

CaFADS was produced as a recombinant protein in *Escherichia coli* BL21(DE3) and purified as previously described in ref. 40. Protein purity was tested by 15% SDS-PAGE. Protein content in pure samples (in 20 mM PIPES, pH 7.0) was quantified using the theoretical extinction coefficient (ϵ) 279 nm = 27.8 mM⁻¹·cm⁻¹.

Differential Scanning Fluorescence

Interaction of compounds with CaFADS was evaluated using fluorescence thermal denaturation, on the bases of the shifts in denaturation midpoints of thermal curves of the protein [65]. Denaturations were performed in a Stratagene Agilent Mx3005p qPCR

instrument (Santa Clara, US) following SYPR Orange (ThermoFisher Scientific) emission fluorescence (excitation at 492 nm and emission at 610 nm), which greatly increases when this probe binds to protein hydrophobic regions becoming solvent exposed upon thermal unfolding. Solutions containing 2 μM *CaFADS* with the studied compound in an increasing 5-250 μM concentration range (2% residual final concentration of DMSO) and 5xSYPR Orange in 20 mM PIPES pH 7.0, 10 mM MgCl_2 , with a 100 μL total volume were dispensed into 96-well microplates (BRAND® 96-well plates pure grade™). After an initial 1 min incubation at 25°C within the equipment, unfolding curves were registered from 25 to 100°C at 1°C·min⁻¹. Control experiments with *CaFADS* samples with/without DMSO were routinely performed in each microplate. For those compounds shifting midpoint denaturation temperature, T_m , the dissociation constant, K_d , was predicted by fitting the data to the equation [66]

$$\frac{\Delta T_m}{T_m} = \frac{NRT_m^0}{\Delta H_0} \ln \left(1 + \frac{[L]}{K_d} \right), \quad (2)$$

which estimates the extent of the ligand-induced protein stabilization/destabilization. $\Delta T_m = |T_m - T_m^0|$ with T_m^0 and T_m being the midpoint denaturation temperatures in the absence and the presence of ligand, respectively, and ΔH_0 the unfolding enthalpy of the protein in the absence of ligand.

Evaluation of the compound's ability to inhibit the *CaFADS* enzymatic activity

To determine the compound's ability to inhibit the RFK and/or the FMNAT activities of *CaFADS*, both enzymatic activities were quantitatively measured in the absence and presence of the compounds following previously described protocols [41]. Reaction mixtures contained 50 mM ATP, 5 μM RF in 20 mM PIPES, pH 7.0, 0.8 mM MgCl_2 , when assaying the RFK activity, and 50 mM ATP, 10 μM FMN in 20 mM PIPES, pH 7.0, 10 mM MgCl_2 when measuring the FMNAT reaction. Each compound was tested at 250 μM (0.5% residual final concentration of DMSO) for each of the two enzymatic reactions. The samples were pre-incubated at 25°C, the reaction was then initiated by the addition of ~ 40 nM *CaFADS* (final concentration) and allowed for 1 min. Finally, the reaction was stopped by boiling the samples for 5 min and the denatured protein was eliminated through centrifugation. The transformation of RF into FMN and FAD (RFK activity) and of FMN into FAD (FMNAT activity) was evaluated through flavins separation by HPLC (Waters), as previously described [37]. All the experiments were performed in triplicate. To evaluate the potency of compounds as inhibitors, we took advantage of the decrease in quantum yield of fluorescence when FMN is transformed into FAD, which allows to follow such transformation in a continuous system. Measurements were carried out using a microplate reader Synergy™HT multimode plate reader (Biotek) with BRAND 96-well plates pure Grade™. Reaction mixtures contained 5 μM RF or FMN, and 50 μM ATP in 20 mM PIPES, pH 7.0, 10 mM MgCl_2 , and the inhibitor compound in a 5-250 μM concentration range (2% residual final concentration of DMSO). Reactions were initiated through addition of 0.4 μM *CaFADS*, being the final reaction volume 100 μL . Flavin fluorescence (excitation at 440 nm and emission at 530 nm) was registered at 25°C, every 50 s during 15 min. The fluorescence change per time unit ($\Delta F/\Delta t$) was calculated as the slope of the resulting fluorescence decays recorded between 0 and 6 min (linear decay of the fluorescence). Controls which contained the reaction mixture without the enzyme and without any potential *CaFADS* inhibitory compound were included in the assay and referred as the 0% and 100% of enzymatic activity, respectively. IC_{50} was calculated as the concentration of compound required for a 50% inhibition of the enzymatic activity.

Determination of the antibacterial activity of the compounds

The minimum inhibitory concentration (MIC) of the inhibitors was determined by the resazurin serial broth microdilution method [67] according to the Clinical and Laboratory Standards Institute guidelines. Compounds were tested against a panel of bacterial strains including Gram positives, Gram negatives and acid fast bacteria (see Table SI 3). Serial 2-fold dilutions of the inhibitors were performed in cation-adjusted Mueller-Hinton broth (Difco) in 96-well polypropylene flat-bottom plates, with a final volume of 100 μL per well. Subsequently, liquid cultures of the bacterial strains in logarithmic phase were adjusted to 10^6 CFU/ml in Mueller-Hinton broth, and 100 μL of this suspension were added to each well, resulting in a final inoculum of $5 \cdot 10^5$ CFU/ml. Plates were incubated for 18 hours at 37°C . Then, 30 μL of 0.4 mM filter-sterilized resazurin (Sigma-Aldrich) was added to each well, and results were revealed after 4 h of further incubation at 37°C . When testing the compounds against mycobacteria, Middlebrook 7H9 (Difco) supplemented with 10% ADC (0.2% dextrose, 0.5% V fraction BSA and 0.0003% bovine catalase) (BD Difco) and with 0.5% glycerol (Scharlau) was used as culture media, and plates were incubated 4 days for *Mycobacterium smegmatis* and 7 days for *M. tuberculosis*. Resazurin (blue) is an indicator of bacterial growth, since metabolic activity of bacteria reduces it to resorufin (pink). The minimum inhibitory concentration (MIC) is the lowest concentration of compound that does not change the resazurin colour from blue to pink.

Evaluation of the cytotoxicity of the compounds in eukaryotic cell lines

The (4,5-dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromide (MTT) assay was used to determine the effect of the compounds in cell growth and viability of HeLa (ATCC®CCL-2) and A549 (ATCC®CCL-185) eukaryotic cell lines. Both cell lines were routinely cultured in high-glucose DMEM (Lonza) supplemented with 10% fetal bovine serum, 4 mM glutamine GlutaMAX™ (Gibco) and 1x non-essential amino acids (Gibco), under 5% CO_2 at 37°C in a humidified atmosphere. For the MTT assay, cells were initially seeded in 96-well flat-bottom plates at a density of 10^4 cells per well and cultured for 24 h. Cultures were routinely tested for mycoplasma presence. The compounds were dissolved in fresh culture medium, added in a 4-512 μM concentration range (1% DMSO final concentration), and incubated with the cells for 24 h. Finally, formazan crystals were dissolved with pure DMSO and MTT absorbance was measured at 570 and 650 nm. Untreated cells were included as control of 100% viability. Assays were done in quadruplicate.

Results and Discussion

Virtual screening protocol

The VS protocol aims to find active compounds, from large compound libraries, towards receptors for which little or no information about ligands or ligand-receptor structures is available. This is the case of the CaFADS. To achieve this goal, we implemented a funnel-like protocol with four filtering stages (Figure 1). It includes three main VS stages plus an experimental stage. In the following, we present the principal ideas for the integrative VS protocol.

Fig 1. Main stages of the funnel-like VS protocol. The protocol consists of four stages: *i*) flexible pharmacophore-based VS (flexi-pharma) [31], *ii*) docking and exponential consensus ranking (ECR-docking) [43], *iii*) MD simulations with consensus ranking -that includes a new Morse-based ligand-flexibility score-, and *iv*) biological experimental binding and activity assays. At each stage, the compound library was filtered. The protocol was optimized and validated over a library of 1993 compounds, which was previously tested over *CaFADS* [41]. On the left, we show an example of the reduction of this library going from 100% to 0.25% through the successive steps.

Flexi-Pharma: pharmacophore filtering from ligand-free receptor conformations

Pharmacophore-based VS strategies are computationally efficient. These strategies are able to explore large compound libraries using pharmacophores: an ensemble of physico-chemical features that ensure the optimal interactions within the active site of a specific biological target [16]. Therefore, the first stage of the protocol implements a pharmacophore-based VS strategy [31]. The method flexi-pharma defines pharmacophores from ligand-free receptor conformations from MD simulations. It implements a rank-by-vote strategy, assigning a vote to each compound that matches an MD conformation. The use of multiple conformations allows for a better exploration of the pharmacophoric space. The voting strategy enables the filtering of the molecules at any percentage of the dataset. Details for the flexi-pharma strategy are presented in the Methods and in ref. 31.

ECR-docking: exponential consensus ranking of docking VS

The second stage of the protocol consists of a docking-based VS. Molecular docking aims to find the most favorable binding conformation of a molecule (*i.e.*, pose) upon binding to a pocket of a protein target [68,69], and assigns a docking score to each molecule. The docking score is an empirical or physics-based estimation of the affinity of the molecule towards the biological target. Therefore, with molecular docking, it is possible to screen and rank molecules from compound libraries. However, it has been shown that the docking results might be system or structure dependent [43,70], possibly due to algorithm-parameterization biases, which are trained over particular benchmark systems. To overcome this limitation, we use a consensus strategy that combines the results from different docking programs to obtain a consensus rank using a sum of exponential functions (ECR method) [43]. Below and in the Supplementary Information, we explain different docking-scoring alternatives that we used to find the optimal enrichment for the FMNAT-FADS ligand screening.

MD-ranking VS

MD simulations were used to estimate the compound affinities and the stability of the predicted complexes filtered from the docking stage. The MD starting configuration was selected from the best pose obtained with Autodock4.2 [44,46,47] in the previous stage. Inspired by conformational-prediction tools that take into account flexibility [71,72], we used two measures for the stability and affinity of the ligand bound to the receptor in the MD ensemble. The first measure generates a consensus rank using multiple scoring functions over the MD conformations. We called this scoring function-based rank (see below for details). The second measure is based on the root mean square deviation (RMSD) of the ligand's atomic positions along the MD trajectory. This is used with a Morse potential to define a score that measures the ligand's flexibility (see below for details). Finally, the scoring function-based rank and the Morse-based rank are

Fig 2. Morse-based score. A score that uses a Morse potential (Eq. 3) was implemented for scoring the flexibility of the ligand inside the pocket using MD simulations. The input variable is the standard deviation of the RMSD of the ligand's atomic positions around the binding site. Ligands that show large RMSD variations are considered very flexible -with dissociation tendencies (*i.e.*, unstable)- and their behavior is penalized (right of vertical black dashed arrow). Ligands with small RMSD fluctuations are considered rigid leading to a conformational penalization (left of black vertical dashed arrow). The Morse potential was implemented with a force constant $k = 1$ kcal/mol.nm², a depth of the well of $\omega = 1$ kcal/mol, and the minimum is localized at $r_0 = 0.242$ Å.

combined using the ECR method. We use this analysis to select the percentage of best-ranked molecules for the activities assays. In the following, we describe the scoring function-based rank and the Morse-based rank.

Scoring function-based rank. We used four scoring functions: Autodock4.2 [44, 46, 47], Vina [48], Vinardo [51] and CY score [73], which were calculated over each MD conformation. For each scoring function, an average score over all the conformations is calculated for each molecule. This can then be used to rank the molecules. The ranks from the four scoring functions are used with the ECR method [43] to obtain a consensus rank by combining their individual ranks.

Morse-based rank. We used the standard deviation of the RMSD of the ligand's atomic positions around the binding site as an indicator of the ligand's flexibility. Those ligands showing a small standard deviation of the RMSD at the binding site indicate a very rigid complex, which leads to a conformational penalization. On the other hand, molecules with high RMSD standard deviation, indicate a dissociation tendency, which leads to an affinity penalization. These behaviors can be characterized using a Morse potential (Eq. 3 and Figure 2)

$$V_M(r) = \omega(1 - \exp^{-a(r-r_0)})^2, \quad (3)$$

where ω is the depth of the well, r_0 is the position of the minimum, $a = \sqrt{\frac{k}{2\omega}}$ and k is a constant that defines the width of the well. For the Morse-base score, we used $V_M(r)$ from Eq. 3, where the dependent variable r is the standard deviation of the RMSD along the MD trajectory, and r_0 is the standard deviation corresponding to a normal distribution with null entropy (*i.e.*, $r_0 = 0.242$ Å). The Morse potential was implemented with a force constant $k = 1$ kcal/mol.nm² and a depth of the well $\omega = 1$ kcal/mol. Thus, RMSD values lower or higher than 0.242 Å are penalized with Morse-based score. We used this to rank the molecules according to the $V_M(r)$ score. We note that most molecules have a RMSD standard deviation greater than 0.242 Å, therefore, the parameters k and ω used in the score do not have a great impact in the final Morse-based rank.

VS parameter dependence

Although the presented VS protocol is sufficiently general to be applied over any receptor target, there are several parameters and setups that can be optimized. Moreover, because -in its complete form- it has not been tested, we considered it necessary to first validate and optimize the VS protocol over a benchmark library with known inhibitors of the *CaFADS* - FMNAT activity [41]. The results are presented in the following section.

Fig 3. Flexi-pharma VS stage over the Prestwick library $\text{\textcircled{R}}$. Average enrichment plot of the Prestwick library using the flexi-pharma stage over MD conformations of ligand-free *CaFADS*. The affinity-grid threshold value is 0.1% and 600 equidistant frames obtained from an MD of 60 ns were used [42]. The flexi-pharma number of votes for each molecule was used as a score to calculate the EPs. Bootstrapping analysis was performed by sampling with replacement 100 times to obtain the average EP and its standard deviation. The violet line shows the screening threshold ($\approx 30\%$) for the selection of molecules to be filtered and passed onto the second stage.

VS protocol FADS validation: Prestwick Chemical Library $\text{\textcircled{R}}$

To validate and optimize the VS protocol for screening potential ligands of *CaFADS*, a molecular library (Prestwick Chemical Library $\text{\textcircled{R}}$) was used. A previous study showed that 39 of its 1993 compounds are able to bind to the *CaFADS* with FMNAT inhibitory activity [41]. To study the performance, we measure the enrichment factor (EF) and the enrichment plots (EP) [64] (see the Methods). In the following, we present the results for each stage of the VS protocol applied over the Prestwick Chemical Library $\text{\textcircled{R}}$.

Flexi-pharma FADS optimization and validation

In Figure 3, we present the EP obtained after the application of the flexi-pharma stage over the Prestwick compound library $\text{\textcircled{R}}$. Since the flexi-pharma method uses MD conformations of the ligand-free receptor, we used 600 equidistant frames from 60 ns of MD of the ligand-free *CaFADS*, which was carried out in a previous study [42]. We applied the flexi-pharma VS as described in the Methods. We find an enrichment of the compound library, showing that the data (black line) are better than a random EP (red line). The vertical violet line shows the percentage of molecules selected to pass to the next stage. The selected list of compounds consists of 600 potential ligands ($\approx 30\%$ of the initial compound library) with 24 actual ligands, resulting in a EF of 2.0 for this stage. These results support the usefulness of flexi-pharma to enrich the Prestwick compound library.

The EPs showed in the Figure 3 involved several parameters, such as the affinity grid threshold value (*i.e.*, percentage of grid points with lowest grid energies), the active spaces and the number of features used to define the pharmacophores (see ref. 31). In that work, it was shown that the results are almost independent of the affinity-grid threshold. However, a large threshold implies a large number of features, which increases the pharmacophore set and the computational time to carry out the VS. Therefore, a good computational efficiency is obtained with small threshold values, while maintaining the performance. For this study, the threshold value of 0.1% is used. Because of the large size of the FMNAT active site, the pharmacophores were obtained from 7 active spaces (centered at NE2-H31, NE2-H57, CA-E108, CG-L110, ND2-N125, OG-S164 and CZ-R168 [50]).

ECR-docking FADS optimization and validation

The second stage of the VS protocol uses a docking-based strategy. Docking generates an optimal molecule-bound conformation with a corresponding score. However, some docking-program outcomes depend on the system of study [43,70]. Thus, a particular docking software can show good results for a receptor, however, it can show bad results for other receptors. For an untested receptor it is impossible to know, in advance, which docking software generates the best outcome. To overcome this limitation, we implement a modified version of the exponential consensus rank (ECR) strategy [43] using several docking programs.

Fig 4. ECR-docking VS stage over the Pretswick library \mathbb{R} . Enrichment plot using the ECR (black line) from the best Autodock4.2 pose that is re-scored with four scoring functions (Autodock4.2, Vina, Vinardo and CYscore). The shaded area encloses the best and worst behaviors for the individual scoring functions. The enrichment plot is normalized by the initial database values (39-ligands and 1993 compounds). The violet line shows the threshold for the selection of molecules for the third VS stage.

The top 600 molecules of the Pretswick library \mathbb{R} filtered from the flexi-pharma stage were docked, using several programs, to the FMNAT module using the crystallographic structure (PDB 2X0K) of *Ca*FADS. After several attempts (see the Supplementary Information), we found that the best EP was obtained by using the best pose from Autodock4.2 and re-scoring it with Autodock4.2 [44, 46, 47], Vina [48], Vinardo [51] (a function scoring implemented in Smina [49]) and CYscore [73] scoring functions. We note that molecules which best Autodock4.2 pose was not within the FMNAT-FADS active site (less than 5 Å of H31 and N125) [50] were discarded, reducing the list to 467 molecules (including 23 confirmed inhibitors). The ranks from each scoring function were combined in using an ECR methodology to obtain a consensus rank. The enrichment plot after this analysis is shown in the Figure 4 and SI 1 Fig. The top 100 molecules from this analysis contained 9 confirmed ligands, which represents a global EF% (*i.e.*, the EF% normalized to the initial compound library) of 4.6, showing a clear enrichment.

MD-ranking FADS optimization and validation

MD simulations of potential ligand-receptor complexes allows for a more accurate sampling of their conformational space. The stability of these conformations should give an estimate of the affinity of the potential ligand towards the receptor. As mentioned previously, the starting conformation was chosen from the best docking pose from Autodock4.2 obtained from the previous stage. For the Pretswick library \mathbb{R} , we found that by dividing the MD stage into two substages according to the simulation time (MD-ranking 1 and MD-ranking 2), there is a good trade-off between computational costs and performance. In the MD-ranking 1, we ran 5 ns for all the compounds filtered from the docking stage. We used the two measures over the MD conformations, scoring function-based and Morse-based, to rank of the potential ligands to the *Ca*FADS. The two results were combined using an ECR to obtain the EPs shown in Figure 5 (black line). Our results indicate that it is possible to obtain an enrichment of the library compound using MD and combining the two stability measured with using an ECR methodology. We note that, although the EP from the Morse-base rank shows better outcome than the EP from the ECR scoring functions-based (ECR-SF) (blue and green lines, respectively, in Figure 5), the use of both measures is relevant. The logic behind combining the two stability measures, lies in that the Morse-based rank gives only information about the conformational stability of the ligand but it does not contain direct information about any physico-chemical interactions. Whereas the scoring functions, *e.g.*, Autodock4.2 or Vina, include physics-based interactions which are relevant. Therefore, the scoring functions-based rank supplies an empirical contribution to the enthalpy and global entropy in the binding process. Thus, the two strategies should complement each other.

We selected the top 50 molecules to be screened in the MD-ranking 2 stage. For this filtered set, we extended the MD time to 15 ns for each complex. These new conformations are scored similarly as before (*i.e.*, ECR combined scoring function-based and Morse-based ranking) and we select the top 5 molecules. The EFs for all the stages of the protocol are shown in Table 1. These results confirm that all stages of the VS

Fig 5. MD-ranking 1 VS stage over the Prestwick library $\text{\textcircled{R}}$. Enrichment plot obtained using an ECR methodology (black) from the combination of the ECR scoring function-based rank (ECR-SF) (green) and Morse-based rank (blue). 5 ns of MD for 100 complexes were carried out. We used 200 equidistant frames from the last 2 ns of MD simulation. The enrichment plot is normalized to the initial database (39-ligands and 1993 compounds). The violet line shows the threshold for the molecule selection for the MD-ranking 2 stage.

protocol increase the enrichment while saving computational resources.

422

Table 1. EFs obtained for each VS stages using the Prestwick library $\text{\textcircled{R}}$.

Stage	Stage EF	Global EF	Ligands	Molecules filtered
Flexi-pharma	2.0	2.0	24	600
ECR-Docking	2.3	4.6	9	100
MD-ranking 1	1.3	6.1	6	50
MD-ranking 2	1.7	10.2	1	5

Stage EF is the enrichment factor relative to the previous step, while the Global EF is the enrichment factor normalized to the initial database (39-ligands and 1993 compounds).

VS protocol FADS application: Maybridge database

423

Once the efficiency of the VS protocol was optimized with the Prestwick library $\text{\textcircled{R}}$, the protocol was implemented over the Maybridge compound library that contains 14000 molecules. A description of the VS protocol for this library is shown in SI 2 Fig. For the flexi-pharma filtering we selected 3000 molecules. Then, the best 600 were selected using the ECR-docking stage. Finally, from the two MD-ranking VS, we selected the top 30 molecules. We note that the difference in the screening between the Maybridge library with respect to the Prestwick library $\text{\textcircled{R}}$ is the percentage of molecules selected at each step. These numbers were changed to optimize the computational efficiency, since for just the MD-ranking stages 6 μ s of MD simulation time were used.

424

425

426

427

428

429

430

431

432

Experimental evaluation of the affinity and the inhibitory activity of selected molecules

433

434

Considering a range of properties for the 30 best VS-ranked compounds that relate to their potential drug-likeness, shown in Table SI 1 [74, 75], as well as their commercial availability, 17 compounds were chosen as virtual screening hits (VSH) to experimentally evaluate their performance. From the experimental assays 5 compounds were found as true ligands of *CaFADS*. Table 2 and Figure 6 show the dissociation constants (K_d) values in the range of 1.7 - 41 μ M.

435

436

437

438

439

440

Since binding of small molecules to a protein usually alters its thermal conformational stability, shifting the midpoint temperatures (T_m) of thermal denaturation curves [65], displacements in T_m induced by the different VSH appeared as a feasible approach to experimentally identify those binding *CaFADS* [76]. 5 compounds, out of the 17 selected, produced a dose-response T_m shift, ΔT_m , indicative of interaction with *CaFADS* (Figure 6A). Compounds C6 and C9 increased T_m by more than 3 degrees, indicating binding to the protein. In addition, C3, C5 and C18 shifted it to lower values (up to 2 and 6 degrees, respectively), suggesting that they produced a ligand-induced perturbation consistent with binding and destabilization of *CaFADS*. Fitting of the corresponding dose-response data to Eq. 2 that relates them to the

441

442

443

444

445

446

447

448

449

450

Table 2. *In vitro* performance of VS hits over the FMNAT and RFK+FMNAT *CaFADS* activities.

Compound	$K_d(\mu\text{M})^a$	FMNAT $\text{IC}_{50}(\mu\text{M})^b$	FMNAT % Res. act. 250 μM	RFK+FMNAT $\text{IC}_{50}(\mu\text{M})^b$	RFK+FMNAT % Res. act. 250 μM
C2			>95		88 \pm 2
C3	18 \pm 8	238 \pm 7	48 \pm 3	248 \pm 3	48 \pm 5
C5	1.7 \pm 0.7	53 \pm 1	6.0 \pm 1.4	83 \pm 2	14 \pm 1.3
C6	41 \pm 3	96 \pm 6	48 \pm 6		57 \pm 3
C7			>95		82 \pm 7
C9	6.4 \pm 1.2		56 \pm 6		76 \pm 3
C18	3.0 \pm 0.9	143 \pm 4	35 \pm 4	147 \pm 4	35 \pm 4
C26			72 \pm 6		84 \pm 4

The table includes dissociation constants (K_d) for the compounds altering thermal stability of *CaFADS*, concentration of compound causing 50% enzyme inhibition (IC_{50}) and residual activity at 250 μM of compound for the FMNAT and RFK+FMNAT activities of *CaFADS*. Thermal stability and activity experiments were carried out in 20 mM PIPES, pH 7.0, 10 mM MgCl_2 . *CaFADS* activities were assayed at 25°C. All samples contained 2% DMSO. (n=3, mean \pm SD). ^a Obtained from differential scanning fluorescence data and ^b kinetic measurements. For details see the Methods.

binding affinity, allowed to estimate the corresponding K_d values (Figure 6B, second column in Table 2). The data pointed to C5, C9 and C18 as the stronger binders. In addition, postulated the C5 >C18 >C9 >C3 >C6 affinity ranking with K_d values in the 1.7-41 μM range. The previous results support that compounds C5, C6, C9, C18 and C3 are actual ligands of *CaFADS*, highlighting the capacity of the VS protocol to find protein ligands for receptor targets.

Fig 6. *In vitro* assessment of VSHs ability to bind and to inhibit *CaFADS*.

A) Thermal denaturation curve for *CaFADS* (2 μM) observed by differential scanning fluorescence and T_m shifts observed in the presence of the compounds at 250 μM . Thermal stability curves are plotted against the normalized fluorescence signal. Experiments were carried out in 20 mM PIPES, pH 7.0, 10 mM MgCl_2 , 2% DMSO. B) Dependence of ΔT_m on the VSH concentration and data fit to Eq. 2. C) Dose-response curves for the FMNAT activity of *CaFADS* in the presence of representative VSHs. Experiments performed at 25°C in 20 mM PIPES, pH 7.0, 10 mM MgCl_2 , 2% DMSO, with 5 μM FMN and 50 μM ATP. Values derived from these representations are included in Table 2, such as the IC_{50} and % of remaining activity at 250 μM of the VSH. D) Comparison of the effects of the VSHs on the RFK and FMNAT activities of *CaFADS*. All the experiments were carried out at 25°C, in 20 mM PIPES pH 7.0, MgCl_2 (10 mM when assaying FMNAT activity and 0.8 mM when assaying RFK activity) at saturating concentrations substrates and in the presence of 250 μM of the VSH (2% DMSO, final concentration). Compound color code: Protein in the absence of VSH is shown in light gray, C3 is violet, C5 is red, C6 is green, C9 is blue, C12 is black (shown as control, neither binder nor inhibitor) and C18 is orange. Note that not all molecules are shown in all panels. In panel D, compounds different from the above mentioned are indicated in dark gray and calculated activity percentages are relative to the corresponding ones in absence of compounds. (n=3, mean \pm SD).

As our VS was directed towards the FMNAT active site of *CaFADS*, we then rated the power of the 17 VSHs as inhibitors of *CaFADS* ability to transform FMN into FAD. Hits were evaluated in terms of concentration of compounds causing 50% enzyme inhibition (IC_{50}), as well as of the percentage of remaining activity at the highest compound concentration assayed (250 μM) (Figure 6C, third and fourth columns in Table 2). 6 out of the 17 VSHs produced some inhibitory effect on the FMNAT activity.

These were compounds C3, C5, C6, C9, C18 and C26. Among them, C5 and C6 yielded IC_{50} values below $100 \mu M$, C5 $IC_{50} = 53 \pm 1 \mu M$ and C6 $IC_{50} = 96 \pm 6 \mu M$, with C5 inhibiting over 90% the FMNAT activity of *CaFADS* at the maximal compound concentration assayed. *CaFADS* is a bifunctional enzyme that in addition to the FMNAT N-terminal module holds an RFK C-terminal module that transforms riboflavin (RF) into FMN, producing the substrate of the FMNAT activity. The presence of this second module was not considered in our VS protocol, where only the ATP/FMN binding pocket of the FMNAT module was used as the active site. Nonetheless, since the RFK module also comprises active site binding of adenine and flavin nucleotides, its RFK activity might also be affected by the VSHs. Therefore, we also evaluated the ability of *CaFADS* to transform RF into FMN and subsequently into FMN. Figure 6D, compares the effect of the VSHs on both individual activities of *CaFADS*, RFK and FMNAT, showing that, under the assay conditions, the 17 hits produced minor effects on the RFK activity. In agreement, when evaluating the overall *CaFADS* activity the effect of the VSHs follows a similar trend to that when individually evaluating the FMNAT activity.

To assess the effect of VSHs on the growth of different bacteria, we determined their MIC (Table 3 and Table SI 2). Bacterial cells of *C. ammoniagenes*, *C. glutamicum*, *C. diphtheriae*, *M. tuberculosis*, *M. smegmatis*, *Streptococcus pneumoniae*, *E. coli*, *Listeria monocytogenes*, *P. aeruginosa*, *Salmonella thyphimurium*, *Staphylococcus aureus* and *Bacillus spp.* were grown in the presence of increasing concentrations of the selected VSHs. Among the VSHs, C2, C5, C6, C18 and C27 produced a detectable inhibition in the growth of *C. ammoniagenes*, being C5 (MIC = $32 \mu M$) the compound producing the largest antibacterial effect followed by C2 (MIC = $64 \mu M$). Interestingly, C5 is also the hit exhibiting the lowest IC_{50} for the FMNAT activity of *CaFADS* (Table 2). The five VSHs exhibiting antibacterial activity against *C. ammoniagenes* also had activity on the other *Corynebacterium* species analyzed, being particularly relevant the effects of C1 and C5 on *C. glutamicum* as well as of C5 and C18 on *C. diphtheriae*. Four of these five compounds, C2, C5, C6 and C27, as well as C3, had also antibacterial effect in the growth of *Mycobacterium* species, although they were in general less potent. In addition, C18 and C27 produced moderate MIC values ($64 \mu M$) for *L. monocytogenes* growth, C6 for *S. pneumoniae*, and C27 for *S. aureus*. It is also worth to note the inhibition of *Bacillus spp.* growth caused by C5 and C27 (MIC = $32 \mu M$).

Table 3. *In vivo* performance of VS hits.

Compound	MIC (μM) <i>C. ammoniagenes</i>	MIC (μM) <i>M. tuberculosis</i>
C2	128	256
C3	>256	256
C5	32	128
C6	128	128
C27	128	256

VSHs which minimal inhibitory concentration (MIC) against *C. ammoniagenes* and *M. tuberculosis* is lower than $256 \mu M$.

In general, we observed that the VSHs showing inhibitory activity against *CaFADS* also inhibit the growing of *Corynebacterium* species. Thus, we can hypothesize that the growing inhibition effect should be caused by the *CaFADS* inhibition. In addition, those VSHs also inhibit the growing of the *Mycobacterium* species, supporting *CaFADS* as a representative model of the FADS of *Mycobacterium tuberculosis*. The fact that the compounds identified through VS have demonstrated some antimicrobial activity is an important result, even when this antimicrobial activity is moderate; historically, potential enzyme inhibitors identified through *in silico* or *in vitro* protein-binding

assays are mostly devoted of any antimicrobial activity, due to their inability to cross the high permeability barrier posed by the bacterial envelope [77].

Finally, we evaluated for the effect of VSHs on eukaryotic cell growth and viability. Compounds C2, C5 and C6 were not cytotoxic in HeLa and A549 cell lines, with IC_{50} (concentration of compounds causing the 50% inhibition of the cellular viability) above the maximal concentration evaluated (512 μ M). In contrast, C18 and C27 showed moderate cytotoxicity in both cell lines but only in the 256-512 μ M range, with complete viability being retained at lower concentrations. Thus, the compounds do not show high cytotoxic effects against eukaryotic cells, highlighting their potential use against prokaryotic pathogens.

Conclusion

We developed a VS protocol that is able to find ligands of an enzyme which does not require previous knowledge of ligands or ligand-receptor structures. The protocol is computationally efficient allowing for the screening of large compound libraries with moderate computational resources. The protocol was implemented over *CaFADS*, an enzyme that is considered a good model for FADSs of bacterial species that cause tuberculosis and pneumonia [40, 41].

The VS protocol involves a funnel-like strategy with filtering stages that increase in accuracy. In the first stage, we used the flexi-pharma method [31], a pharmacophore filtering strategy with ligand-free receptor conformations from MD. In the second stage, we used a consensus docking strategy to combine the results from different docking programs using the exponential consensus ranking (ECR) method [43]. In the third stage of MD-ranking, we developed a new score for the ligand's flexibility using a Morse potential. This score is combined with other scoring functions using the ECR method over the MD ensemble.

The protocol was optimized and validated over an experimentally-tested compound library with known ligands of the *CaFADS*. We then implemented the VS strategy over a unexplored compound library, resulting in a list of 17 compounds that were tested experimentally. Notably, we discovered five compounds able to bind to the *CaFADS*. One of these compounds shows significant inhibition of the FMNAT activity of *CaFADS*. In comparison to previous work [41], the computational protocol gives an enrichment of around 8 for the experimental stage. In addition, some of the new compounds show growth inhibitory activity against *Corynebacterium*, *Mycobacterium* or *Streptococcus* species, supporting the use of the integrative VS protocol for the initial stages of drug discovery.

Supporting information

SI 1 Fig. Enrichment plot for the different ECR-docking strategies. The violet line shows the EP from ECR combination of Autodock4.2, Vina and Smina docking results using the best pose from each program (as was done in ref. 43). We also studied the outcome when using the best pose for each molecule from the different programs: from Autodock4.2 (black), Vina (green) and Smina (blue), then re-scored it with Autodock4.2, Vina, Vinardo and CYscore, and these new scores were combined using an ECR methodology. We find that using the Autodock4.2 pose and re-scoring it with the other programs produces the best outcome.

SI 2 Fig. Funnel-like protocol implemented for the Maybridge compound library. The number of filtered molecules is shown on the left. The computational

protocol has several stages: first, a pharmacophore-based VS (flexi-pharma), then ECR-Docking, afterwards two MD stages (that depended on the simulation time) were used for ranking the compounds with a Morse-based score and an ECR combination of scoring functions. In the physico-chemical stage, we assessed a range of properties for the 30 best VS-ranked compounds that relate to their potential drug-likeness (Table SI 1), as well as their commercial availability, selecting 17 compounds for the experimental assays. 5 compounds were found to be ligands of *CaFADS*.

SI 3 Fig. Structure of the active VSHs towards *CaFADS* or the organisms tested. C3, C5, C6, C9 and C18 are able to bind to *CaFADS*. C3, C5, C6 and C18 cause 50% of FMNAT activity inhibition (IC_{50}) at concentration lower than 250 μ M. C3, C5 and C18 cause 50% RFK-FMNAT activity inhibition (IC_{50}) at concentration lower than 250 μ M. C3, C5, C6, C9, C18 and C26 have FMNAT residual activity < 95% at 250 μ M of compound. C2, C3, C5, C6, C7, C9, C18 and C26 have RFK+FMNAT residual activity < 95% at 250 μ M of compound. C2, C5, C6 and C27 have MIC values lower than 256 μ M against *C. ammoniagenes*, and or C2, C3, C5, C6 and C27 *Mycobacterium* species.

SI 1 Text Description of the different ECR-docking FADS optimization strategies.

SI 1 Table. Summary of properties for the *CaFADS* best ranked VS compounds. The table shows the VS rank, the Zinc and Maybridge Codes, the names and summary of physico-chemical criteria (values at pH 7.0) to evaluate their potential drug-likeness. Preferred criteria values are indicated on the top in green. Favorable criteria for each compound are highlighted in a green background, those in the limit are in a yellow background and those violating the criteria are in a red background. Compounds selected for experimental evaluation as virtual screening hits (VSH) are highlighted in red font in the first four columns.

SI 2 Table. Minimal Inhibitory Concentration (MIC) of VSHs against different microorganisms. Compounds were, in most cases, assayed in the 0-256 μ M concentration range. In some cases, they were assayed only in the 0-64 μ M concentration range, and if no effect was observed in these cases > 64 is shown. Best performing compounds are colored from red, orange, yellow to green.

SI 3 Table. Bacterial strains tested for VSH antibacterial activity

Author Contributions

I.L. and P.C. developed the theory. I.L. performed the computations. K.P-R tested some of the computational protocols. E.A., J.A. and M.M. developed conceptualization and analysis of experimental analysis. E.A performed the experiments. All authors wrote and revised the manuscript. P.C., J.A. and M.M. funding acquisition.

Funding:

This research was founded by Colciencias, University of Antioquia, Ruta N, Colombia, and the Max Planck Society, Germany. Also by the Spanish Ministry of Economy, Industry and Competitiveness (MINECO) [BIO2016-75183-P AEI/FEDER, UE to

M.M.], and the Government of Aragón-FEDER [Grupo de Referencia Biología Estructural (E35-20R to M.M.)].

591
592

Acknowledgments

593

I.L., K. P-R and P.C. were supported by Colciencias, University of Antioquia, Ruta N, Colombia, and the Max Planck Society, Germany. Some computations were performed in a local server with an NVIDIA Titan X GPU. P.C. gratefully acknowledges the support of NVIDIA Corporation for the donation of this GPU.

594
595
596
597

The authors would like to acknowledge the use of Servicios Generales de Apoyo a la Investigación-SAI, Universidad de Zaragoza. We thank M. Minjarez and M. Martinez-Júlvez for their collaboration in evaluating physico-chemical properties of VS hits. The authors declare no conflict of interest.

598
599
600
601

References

1. Tacconelli E, Carrara E, Savoldi A, Harbarth S, Mendelson M, Monnet DL, et al. Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. *Lancet Infect Dis.* 2018;18(3):318–327.
2. Rognan D. The impact of *in silico* screening in the discovery of novel and safer drug candidates. *Pharmacol Therapeut.* 2017;175:47–66.
3. Walters WP, Stahl MT, Murcko MA. Virtual screening—an overview. *Drug Discov Today.* 1998;3(4):160–178.
4. Leelananda SP, Lindert S. Computational methods in drug discovery. *Beilstein J Org Chem.* 2016;12(1):2694–2718.
5. Fradera X, Babaoglu K. Overview of methods and strategies for conducting virtual small molecule screening. *Curr Protoc Chem Biol.* 2017;9(3):196–212.
6. Schneider G. Automating drug discovery. *Nat Rev Drug Discov.* 2017;17(2):97–113.
7. Sliwoski G, Kothiwale S, Meiler J, Lowe EW. Computational methods in drug discovery. *Pharmacol Rev.* 2014;66(1):334–395.
8. Kalyaanamoorthy S, Chen YPP. Structure-based drug design to augment hit discovery. *Drug discov today.* 2011;16(17-18):831–839.
9. Spyrakis F, Cavasotto CN. Open challenges in structure-based virtual screening: Receptor modeling, target flexibility consideration and active site water molecules description. *Arch Biochem Biophys.* 2015;583:105–119.
10. Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov.* 2004;3(11):935–949.
11. Bajorath J. Integration of virtual and high-throughput screening. *Nat Rev Drug Discov.* 2002;1(11):882–894.
12. Lyne PD. Structure-based virtual screening: an overview. *Drug Discov Today.* 2002;7(20):1047–1055.

13. Tanrikulu Y, Krüger B, Proschak E. The holistic integration of virtual screening in drug discovery. *Drug Discov Today*. 2013;18(7-8):358–364.
14. Kumar A, Zhang KY. Hierarchical virtual screening approaches in small molecule drug discovery. *Methods*. 2015;71:26–37.
15. Bresso E, Furlan A, Noel P, Leroux V, Maina F, Dono R, et al. Large-Scale Virtual Screening Against the MET Kinase Domain Identifies a New Putative Inhibitor Type. *Molecules*. 2020;25(4):938.
16. Mortier J, Dhakal P, Volkamer A. Truly Target-Focused Pharmacophore Modeling: A Novel Tool for Mapping Intermolecular Surfaces. *Molecules*. 2018;23(8). doi:10.3390/molecules23081959.
17. Wang X, Shen Y, Wang S, Li S, Zhang W, Liu X, et al. PharmMapper 2017 update: a web server for potential drug target identification with a comprehensive target pharmacophore database. *Nucleic Acids Res*. 2017;45(W1):W356–W360.
18. Rohini K, Agarwal P, Preethi B, Shanthi V, Ramanathan K. Exploring the Lead Compounds for Zika Virus NS2B-NS3 Protein: an e-Pharmacophore-Based Approach. *Appl Biochem Biotechnol*. 2019;187(1):194–210.
19. Stumpfe D, Ripphausen P, Bajorath J. Virtual compound screening in drug discovery. *Future Med Chem*. 2012;4(5):593–602.
20. Chen L, Du-Cuny L, Moses S, Dumas S, Song Z, Rezaeian AH, et al. Novel inhibitors induce large conformational changes of GAB1 pleckstrin homology domain and kill breast cancer cells. *PLoS Comput Biol*. 2015;11(1).
21. De Vivo M, Masetti M, Bottegoni G, Cavalli A. Role of molecular dynamics and related methods in drug discovery. *J Med Chem*. 2016;59(9):4035–4061.
22. Zhao H, Caflisch A. Molecular dynamics in drug design. *European J Med Chem*. 2015;91:4–14.
23. Śledź P, Caflisch A. Protein structure-based drug design: from docking to molecular dynamics. *Curr Opin Struc Biol*. 2018;48:93–102.
24. Mortier J, Rakers C, Bermudez M, Murgueitio MS, Riniker S, Wolber G. The impact of molecular dynamics on drug design: applications for the characterization of ligand–macromolecule complexes. *Drug Discov Today*. 2015;20(6):686–702.
25. Durrant JD, McCammon JA. Molecular dynamics simulations and drug discovery. *BMC Biol*. 2011;9(1):71.
26. Duan L, Liu X, Zhang JZ. Interaction entropy: a new paradigm for highly efficient and reliable computation of protein–ligand binding free energy. *J Am Chem Soc*. 2016;138(17):5722–5728.
27. Cavasotto CN, Orry AJ. Ligand Docking and Structure-based Virtual Screening in Drug Discovery. *Curr Top Med Chem*. 2007;7(10):1006–1014.
28. Carlson HA, Masukawa KM, Rubins K, Bushman FD, Jorgensen WL, Lins RD, et al. Developing a dynamic pharmacophore model for HIV-1 integrase. *J Med Chem*. 2000;43(11):2100–2114.

29. Wieder M, Garon A, Perricone U, Boresch S, Seidel T, Almerico AM, et al. Common Hits Approach: Combining Pharmacophore Modeling and Molecular Dynamics Simulations. *J Chem Inf Model*. 2017;57(2):365–385.
30. Polishchuk P, Kutlushina A, Bashirova D, Mokshyna O, Madzhidov T. Virtual screening using pharmacophore models retrieved from molecular dynamic simulations. *Int J Mol Sci*. 2019;20(23).
31. Lans I, Palacio-Rodríguez K, Cavasotto CN, Cossio P. Flexi-pharma: a molecule-ranking strategy for virtual screening using pharmacophores from ligand-free conformational ensembles. under revision; see Supplementary Information. 2020;-.
32. Mayhew SG, Ludwig ML. 2 Flavodoxins and Electron-Transferring Flavoproteins. vol. 12; 1975. p. 57–118.
33. Mayhew SG, Tollin G. General properties of flavodoxins. *Chemistry and biochemistry of flavoenzymes*. 1992;3:389–426.
34. Lans I, Frago S, Medina M. Understanding the FMN cofactor chemistry within the Anabaena Flavodoxin environment. *Biochim Biophys Acta*. 2012;1817(12):2118–2127.
35. Herguedas B, Martínez-Júlvez M, Frago S, Medina M, Hermoso JA. Oligomeric State in the Crystal Structure of Modular FAD Synthetase Provides Insights into Its Sequential Catalysis in Prokaryotes. *J Mol Biol*. 2010;400(2):218–230.
36. Herguedas B, Lans I, Sebastián M, Hermoso JA, Martínez-Júlvez M, Medina M. Structural insights into the synthesis of FMN in prokaryotic organisms. *Acta Crystallogr, Sect D: Biol Crystallogr*. 2015;71.
37. Serrano A, Ferreira P, Martínez-Júlvez M, Medina M. The prokaryotic FAD synthetase family: a potential drug target. *Curr Pharm Des*. 2013;19(14):2637–48.
38. Serrano A, Sebastián M, Arilla-Luna S, Baquedano S, Pallarés MC, Lostao A, et al. Quaternary organization in a bifunctional prokaryotic FAD synthetase: Involvement of an arginine at its adenylyltransferase module on the riboflavin kinase activity. *Biochim Biophys Acta*. 2015;1854(8):897–906.
39. Krupa A, Sandhya K, Srinivasan N, Jonnalagadda S. A conserved domain in prokaryotic bifunctional FAD synthetases can potentially catalyze nucleotide transfer. *Trends Biochem Sci*. 2003;28(1):9–12.
40. Frago S, Martínez-Júlvez M, Serrano A, Medina M. Structural analysis of FAD synthetase from *Corynebacterium ammoniagenes*. *BMC Microbiol*. 2008;8(1):160.
41. Sebastián M, Anoz-Carbonell E, Gracia B, Cossio P, Aínsa JA, Lans I, et al. Discovery of antimicrobial compounds targeting bacterial type FAD synthetases. *J Enzyme Inhib Med Chem*. 2018;33(1):241–254.
42. Lans I, Seco J, Serrano A, Burbano R, Cossio P, Daza MC, et al. The Dimer-of-Trimers Assembly Prevents Catalysis at the Transferase Site of Prokaryotic FAD Synthase. *Biophys J*. 2018;115(6):988–995.
43. Palacio-Rodríguez K, Lans I, Cavasotto CN, Cossio P. Exponential consensus ranking improves the outcome in docking and receptor ensemble docking. *Sci Rep*. 2019;9(1):5142.

44. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem*. 2009;30(16):2785–91.
45. Koes DR, Camacho CJ. Pharmer: Efficient and Exact Pharmacophore Search. *J Chem Inf Model*. 2011;51(6):1307–1314.
46. Cosconati S, Forli S, Perryman AL, Harris R, Goodsell DS, Olson AJ. Virtual Screening with AutoDock: Theory and Practice. *Expert Opin Drug Discov*. 2010;5(6):597–607.
47. Forli S, Olson AJ. A Force Field with Discrete Displaceable Waters and Desolvation Entropy for Hydrated Ligand Docking. *J Med Chem*. 2012;55(2):623–638.
48. Trott O, Olson AJ. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem*. 2009;31(2).
49. Koes DR, Baumgartner MP, Camacho CJ. Lessons learned in empirical scoring with smina from the CSAR 2011 benchmarking exercise. *J Chem Inf Model*. 2013;53(8):1893–1904.
50. Serrano A, Frago S, Velázquez-Campoy A, Medina M. Role of key residues at the flavin mononucleotide (FMN): adenylyltransferase catalytic site of the bifunctional riboflavin kinase/flavin adenine dinucleotide (FAD) Synthetase from *Corynebacterium ammoniagenes*. *Int J Mol Sci*. 2012;13(11):14492–14517.
51. Quiroga R, Villarreal MA, Ballester P, Wong MH, Allain N, Laguerre M, et al. Vinardo: A Scoring Function Based on Autodock Vina Improves Scoring, Docking, and Virtual Screening. *PLoS One*. 2016;11(5):e0155183.
52. O’Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: An open chemical toolbox. *J Cheminformatics*. 2011;3(1):33.
53. Olsson MHM, Søndergaard CR, Rostkowski M, Jensen JH. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J Chem Theory Comput*. 2011;7(2):525–537.
54. Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res*. 2004;32:W665–W667.
55. Dolinsky TJ, Czodrowski P, Li H, Nielsen JE, Jensen JH, Klebe G, et al. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res*. 2007;35(Web Server issue):W522–5.
56. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, et al. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins*. 2010;78(8):1950–8.
57. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys*. 1983;79(926).
58. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber force field. *J Comput Chem*. 2004;25(9):1157–1174.

59. Wang J, Wang W, Kollman P, Case D. Antechamber, An Accessory Software Package For Molecular Mechanical Calculations; 2000. Available from: <http://ambermd.org/antechamber/antechamber.pdf>.
60. Sousa Da Silva AW, Vranken WF. ACPYPE - AnteChamber PYthon Parser interface. BMC Res Notes. 2012;5.
61. Berendsen HJC, van der Spoel D, van Drunen R. GROMACS: A message-passing parallel molecular dynamics implementation. Comput Phys Commun. 1995;91(1-3):43–56.
62. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX. 2015;1:19–25.
63. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: A linear constraint solver for molecular simulations. J Comput Chem. 1997;18(12):1463–1472.
64. Jain AN. Bias, reporting, and sharing: computational evaluations of docking methods. J Comput Aided Mol Des. 2008;22(3-4):201–212.
65. Velazquez-Campoy A, Sancho J, Abian O, Vega S. Biophysical Screening for Identifying Pharmacological Chaperones and Inhibitors Against Conformational and Infectious Diseases. Curr Drug Targets. 2016;17(13):1492–1505.
66. Cooper A, McAuley-Hecht KE. Microcalorimetry and the molecular recognition of peptides and proteins. Philos Trans R Soc A. 1993;345(1674):23–35.
67. Palomino JC, Martin A, Camacho M, Guerra H, Swings J, Portaels F. Resazurin microtiter assay plate: Simple and inexpensive method for detection of drug resistance in *Mycobacterium tuberculosis*. Antimicrob Agents Chemother. 2002;46(8):2720–2722.
68. Sousa SF, Ribeiro AJM, Coimbra JTS, Neves RPP, Martins SA, Moorthy NSHN, et al. Protein-Ligand Docking in the New Millennium – A Retrospective of 10 Years in the Field. Curr Med Chem. 2013;20(18):2296–2314.
69. Yuriev E, Holien J, Ramsland PA. Improvements, trends, and new ideas in molecular docking: 2012–2013 in review. J Mol Recognit. 2015;28(10):581–604.
70. Cavasotto CN, Abagyan RA. Protein Flexibility in Ligand Docking and Virtual Screening to Protein Kinases. J Mol Biol. 2004;.
71. Sarti E, Zamuner S, Cossio P, Laio A, Seno F, Trovato A. BACHSCORE. A Tool for Evaluating Efficiently and Reliably the Quality of Large Sets of Protein Structures. Comput Phys Commun. 2013;184(12):2860–2865.
72. Cossio P, Granata D, Laio A, Seno F, Trovato A. A simple and efficient statistical potential for scoring ensembles of protein structures. Sci Rep. 2012;2:1–8.
73. Cao Y, Li L. Improved protein-ligand binding affinity prediction by using a curvature-dependent surface-area model. Bioinformatics. 2014;30(12):1674–1680.
74. Lipinski CA. Lead-and drug-like compounds: the rule-of-five revolution. Drug Discov Today. 2004;1(4):337–341.

75. Ghose AK, Viswanadhan VN, Wendoloski JJ. A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases. *J Comb Chem.* 1999;1(1):55–68.
76. Villanueva R, Romero-Tamayo S, Laplaza R, Martínez-Olivan J, Velázquez-Campoy A, Sancho J, et al. Redox-and Ligand Binding-Dependent Conformational Ensembles in the Human Apoptosis-Inducing Factor Regulate Its Pro-Life and Cell Death Functions. *Antioxid Redox Sign.* 2019;30(18):2013–2029.
77. Lewis K. The Science of Antibiotic Discovery. *Cell.* 2020;181:29–45.

<https://doi.org/10.1101/2020.04.22.055178>; this version posted April 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

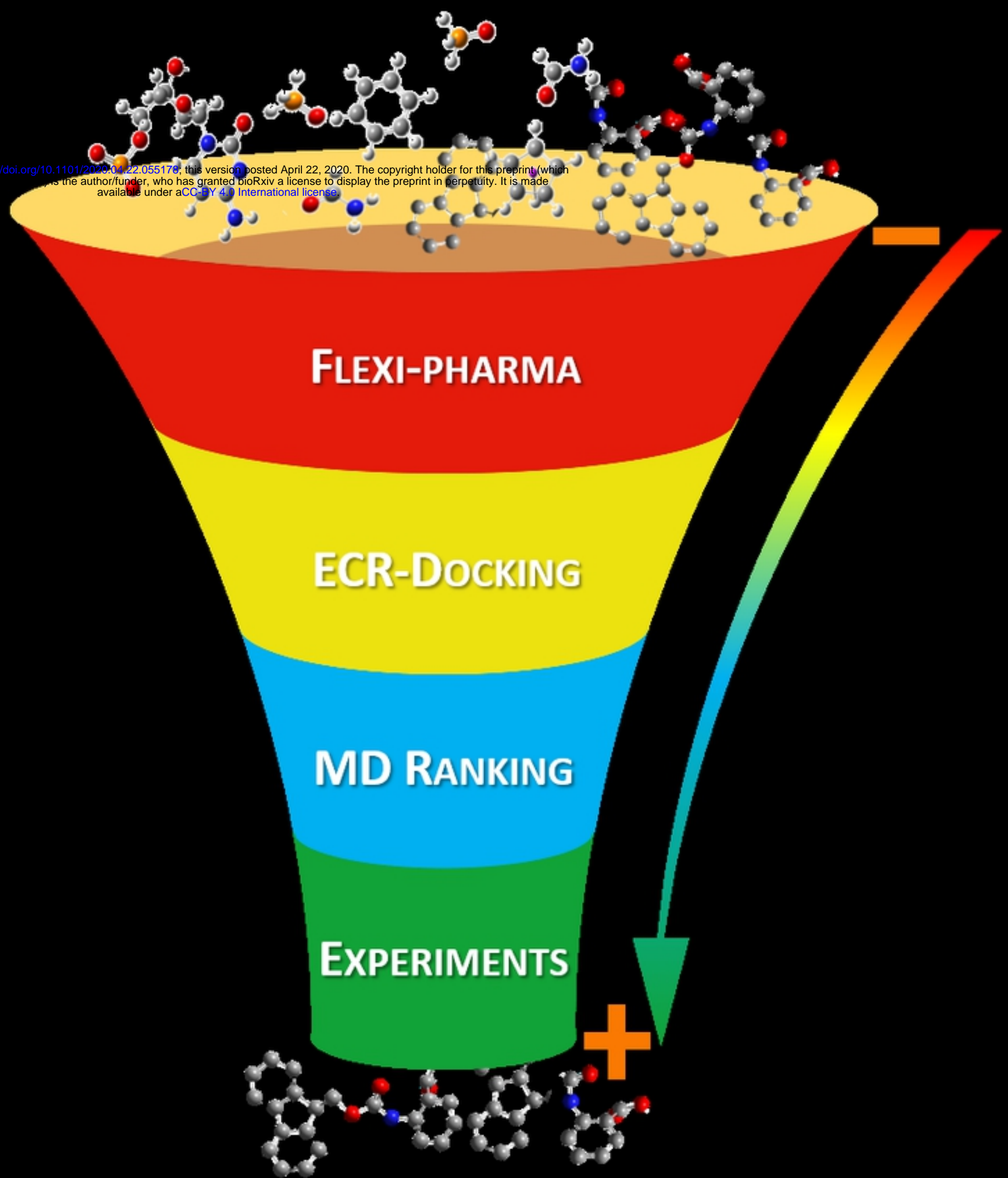


Figure 1

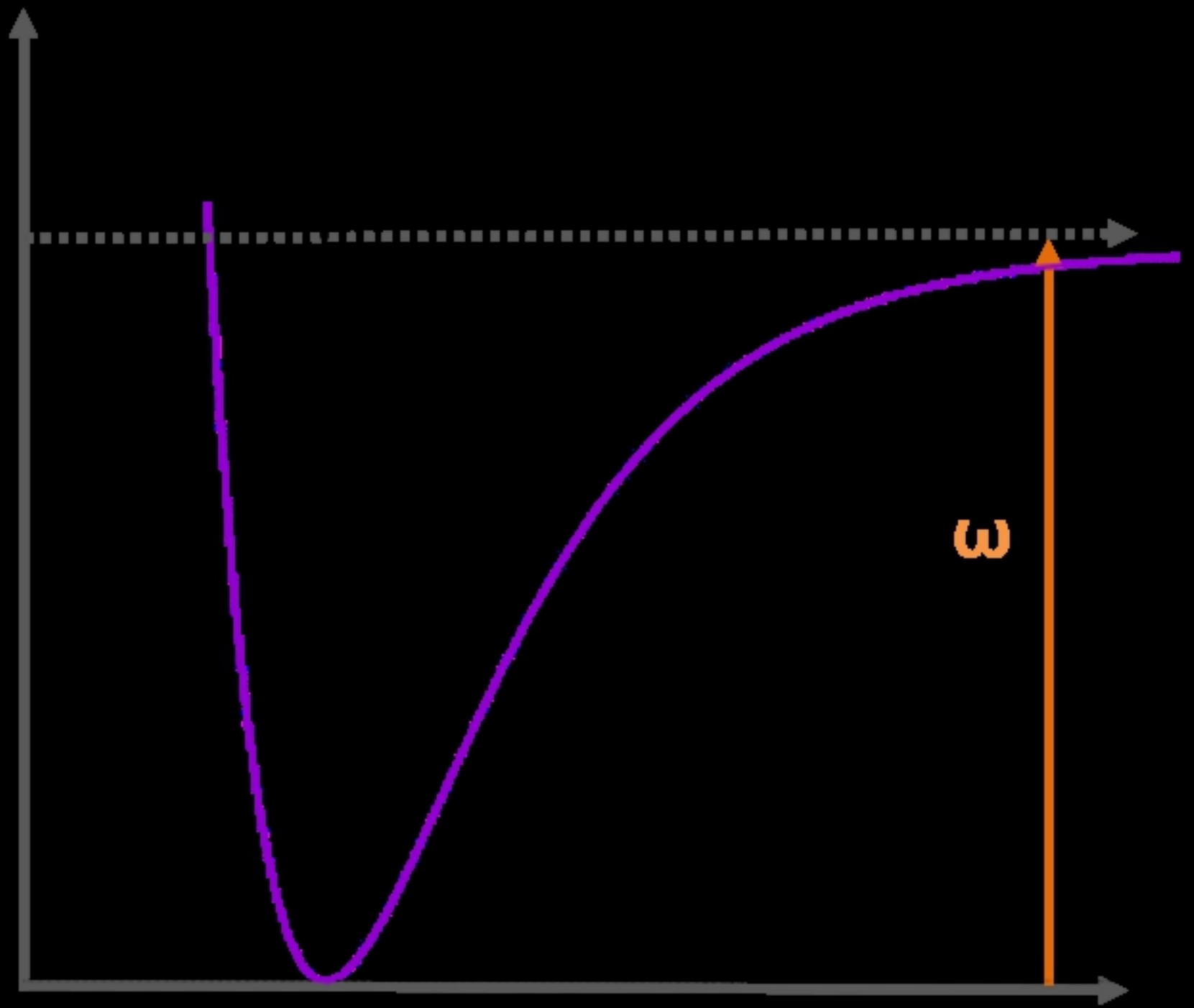


Figure 2

Flexi-pharma VS stage

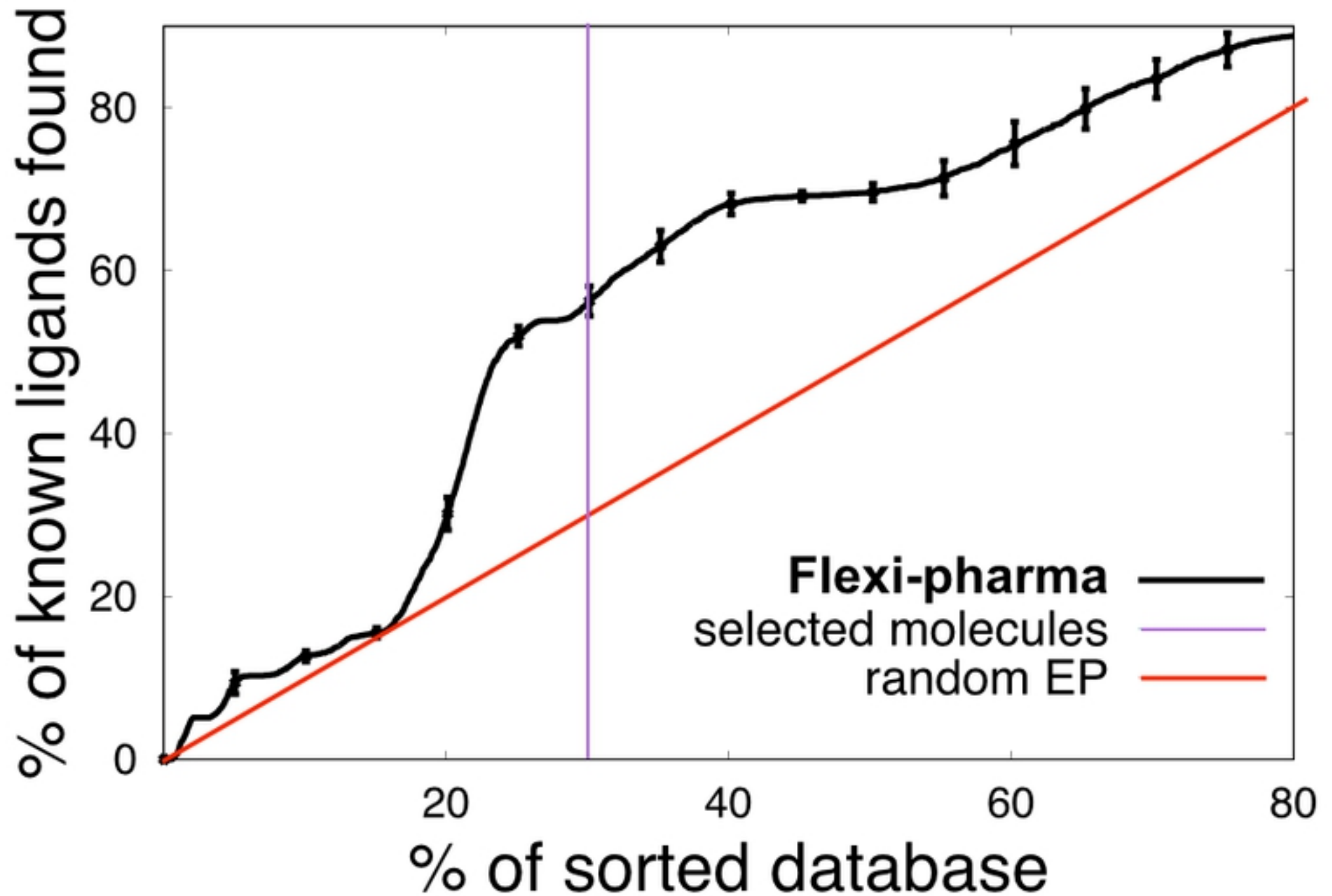


Figure 3

ECR-docking VS stage

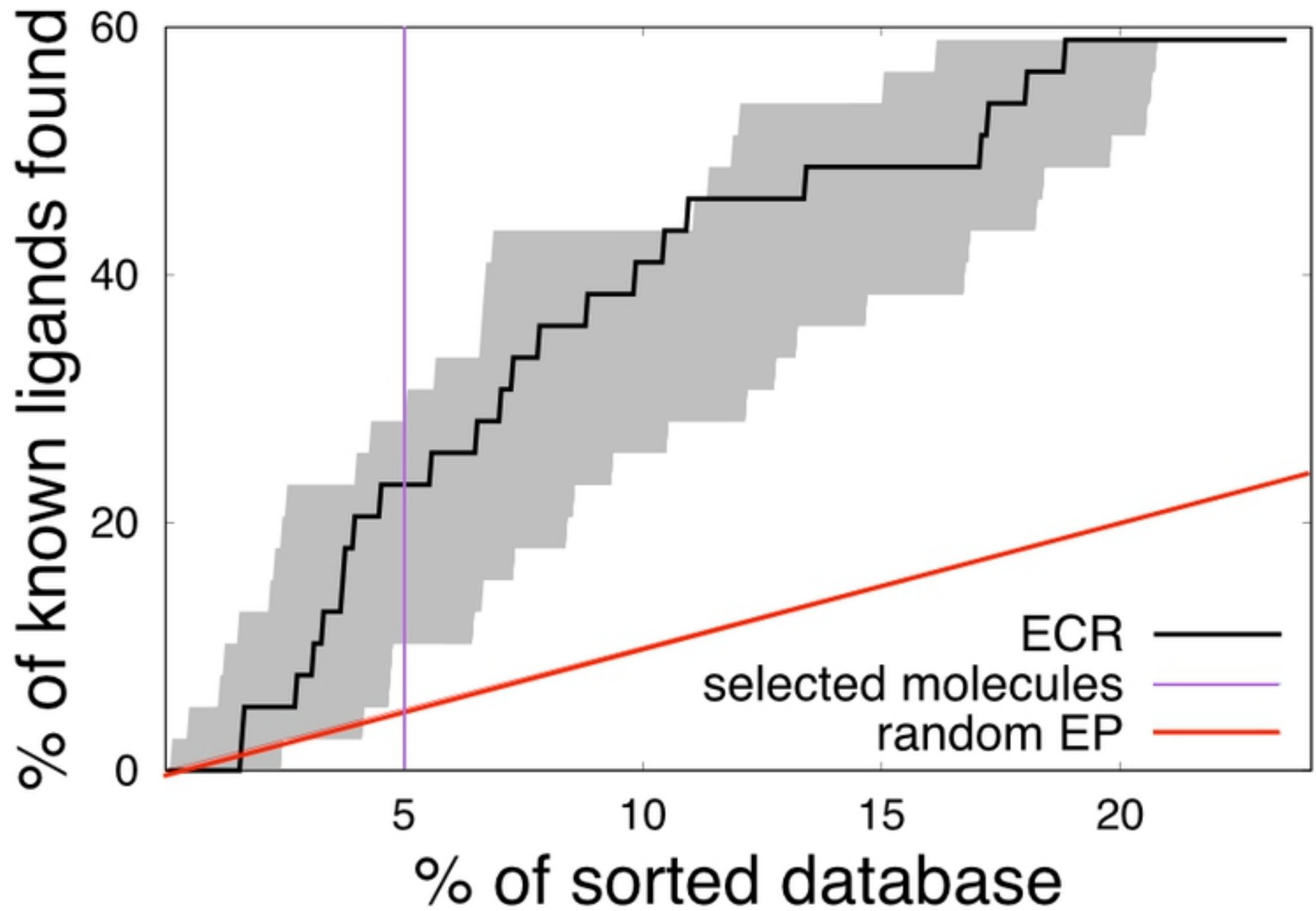


Figure 4

MD-ranking 1 VS stage

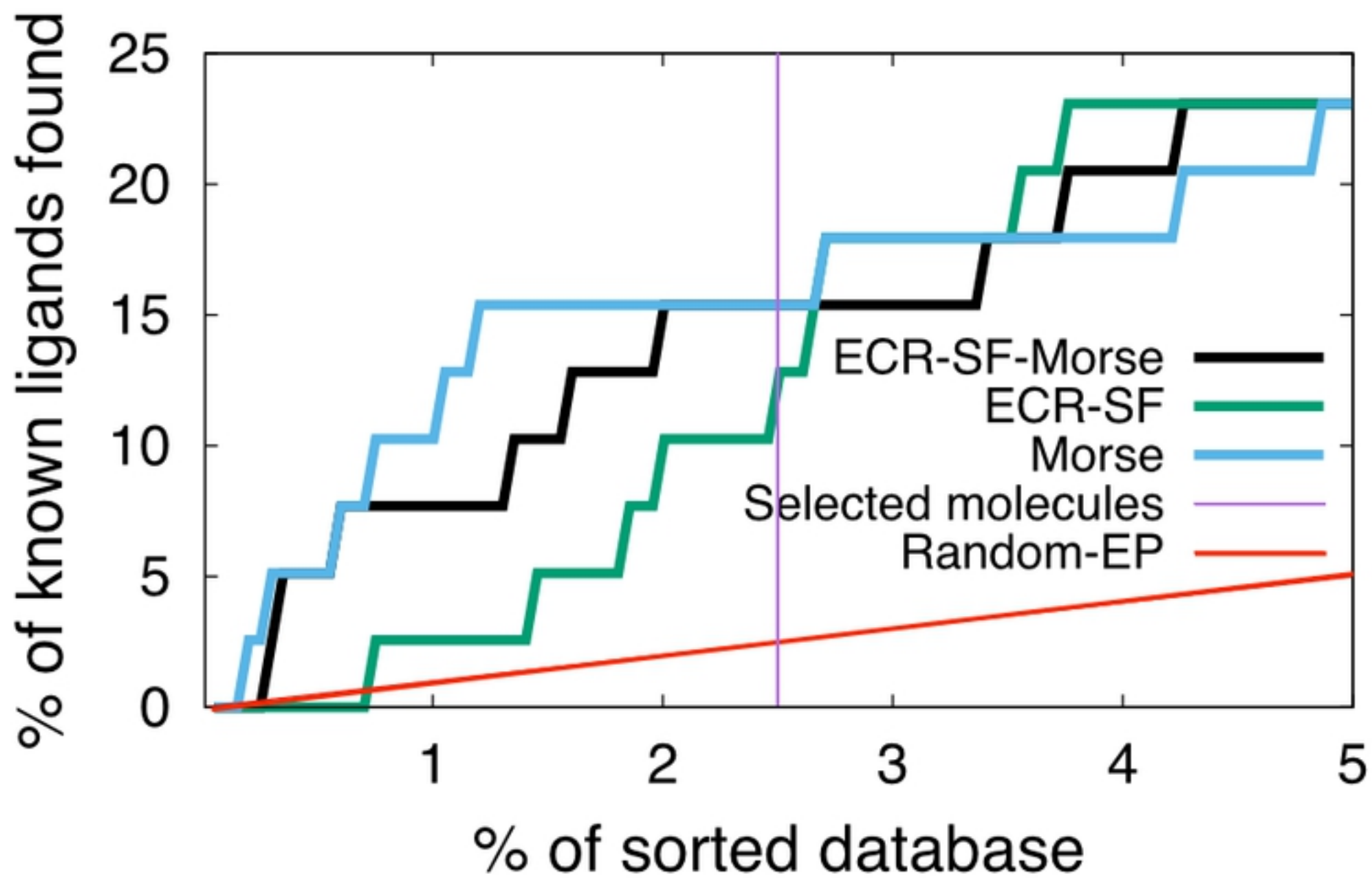


Figure 5

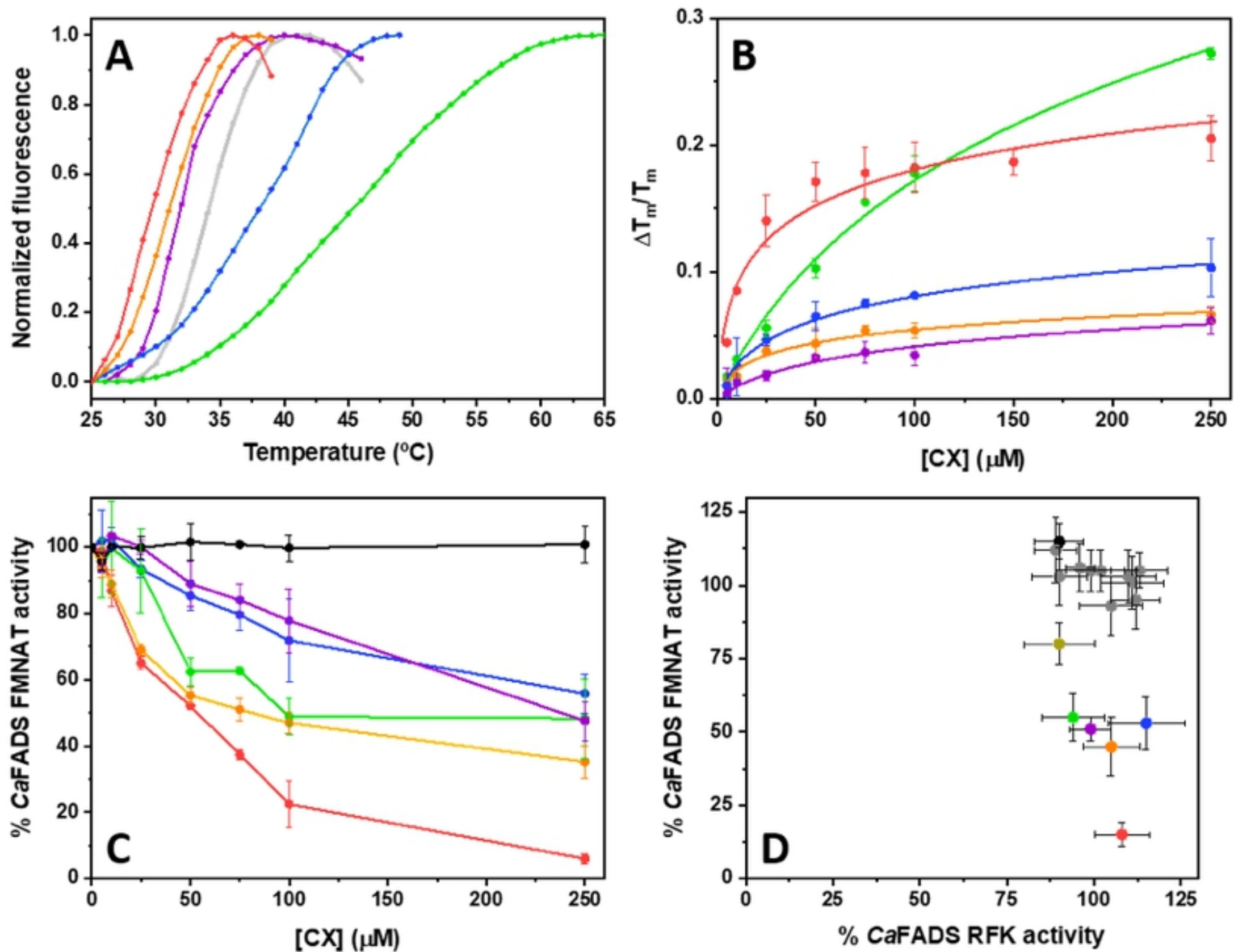


Figure 6