JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

Integration of enhanced sampling methods with Saturation Transfer Difference (STD) experiments to identify protein druggable pockets.

Joana Magalhaes, Giannamaria Annunziato, Nina Franko, Marco Pieroni, Barbara Campanini, Agostino Bruno, and Gabriele Costantino

J. Chem. Inf. Model., **Just Accepted Manuscript** • DOI: 10.1021/acs.jcim.7b00733 • Publication Date (Web): 26 Feb 2018

Just Accepted

ACS Publications

Journal of Chemical Information and Modeling is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

Integration of Enhanced Sampling Methods with Saturation Transfer Difference (STD) Experiments to Identify Protein Druggable Pockets.

Joana Magalhães,†,■ Giannamaria Annunziato,†,■ Nina Franko,‡ Marco Pieroni,† Barbara Campanini,‡ Agostino Bruno,†,#,* and Gabriele Costantino†

†P4T group, Food and Drug Department, Parco Area Delle Scienze 27/A - 43124, Parma, Italy

‡Laboratory of Biochemistry and Molecular Biology, Food and Drug Department, Parco Area Delle Scienze 23/A - 43124, Parma, Italy

#Experimental Therapeutics Program, IFOM – The FIRC Institute for Molecular Oncology Foundation, Via Adamello 16 - 20139 Milano, Italy

■These authors equally contributed to the work

* Corresponding Authors.

E-mail: agostino.bruno@unipr.it; agostino.bruno@ifom.eu

Abstract

Saturation Transfer Difference (STD) is an NMR technique conventionally applied in drug discovery to identify ligand moieties relevant for binding to protein cavities. This is important to direct medicinal chemistry efforts in small-molecule optimization processes. However, STD does not provide any structural details about the ligand-target complex under investigation. Herein, we report the application of a new integrated approach, which combines enhanced sampling methods with STD experiments, for the characterization of ligand-target complexes that are instrumental for drug design purposes. As an example, we have studied the interaction between *St*OASS-A, a potential antibacterial target, and an inhibitor previously reported. This approach allowed us to consider the ligand-target complex from a dynamic point of view, unrevealing the presence of an accessory sub-pocket which can be exploited to design novel *St*OASS-A inhibitors. As a proof of concept, a small library of derivatives was designed and evaluated *in vitro*, displaying the expected activity.

Introduction

Saturation Transfer Difference (STD) is a relatively straightforward NMR technique[1] based on the exploitation of the nuclear Overhouser effect (nOe) observed during the interaction between a protein and a small molecule ligand.[2] STD methods can be applied only when particular conditions occur, that is when the residence time of the ligand into a protein binding site is long enough to permit spin diffusion through nOe, but, at the same time, short enough to allow ligand dissociation and interchange between bound and unbound ligand states within the timeframe of the NMR spin relaxation. Therefore, STD methods can be applied to weak ligands, usually with a K_d between 1 and 100 μM.

Under these conditions only ligand protons in close contact with the protein atoms will receive spin saturation by nOe and a difference spectrum can be recorded by subtracting the off-resonance (I_O) spectrum from the on-resonance (I_{SAT}) spectrum.[2] Since nOe occurs through space, only those protons close enough to protein atoms will show higher signal differences, while protons distant from protein atoms will display low or null signal differences.

STD has received interest in the field of drug design because it can be conveniently employed as a rapid and cheap screening technique.[2–4] Indeed, small molecules which do not bind to the target protein will not produce STD signals, since no spin saturation will be transmitted to the ligands; also aspecific binders can be easily eliminated, since the fast dissociation time will preclude effective spin saturation transfer. STD experiments only require a few milligrams of the soluble protein and the ligand, while the spectra can be produced in a few hours by using routine NMR instrument. Taken together, all these characteristics make STD a reliable tool for screening medium-sized libraries of fragments or small molecules.

The main limitation of this technique, which has hampered its wider use in classical drug design pipeline, is the lack of specificity for the obtained information. The detection of a STD signal only indicates that a given ligand binds to the protein, but no information about the nature of interactions

are provided and information about the interaction with specific binding pocket(s) can be obtained only if STD experiments are complemented with competition experiments. Therefore, STD can only be used for identifying hit structures and ligand moieties important for binding through Group Epitope Mapping (G.E.M.), which helps directing medicinal chemistry efforts in the hit expansion and hit to lead optimization process. Indeed, the identification of the binding epitopes can be used to enable complementary approaches. In one of such cases, when the 3D structure of the target protein is known, the experimentally recorded STD intensity can be used for guiding (or ranking) docking poses obtained by docking algorithms (CORCEMA,[5] PLANTS,[6,7]). In the common case of multiple binding poses with similar scores, the ambiguity can be solved by ranking the poses according to their coherence with STD signals. Operatively, this can be carried out by introducing a penalty function in the docking scoring function, which assigns low scores to those poses which are not in agreement with the experimental STD signals. The approach is however not largely applicable, since it requires at least preliminary knowledge about the binding pocket and the mode of interaction of the ligand with the target protein, a condition that seldom takes place. Taking into account all of these considerations, in this work we present a new method to integrate the STD technique with enhanced sampling methods, in order to broaden the scope of STD application to the hit optimization process.

While in the aforementioned approaches (CORCEMA,[5] PLANTS[6,7]) the STD signals are used to bias the docking function, we decided to use molecular dynamics (MD) to reconstruct the STD spectrum. The rationale behind this idea is that STD signals are the result of magnetization transfer from protein atoms to ligand protons (through space, not through chemical bonds). Therefore, only those ligand protons residing close enough and in contact with protein atoms, and for a time comparable with the spin saturation transfer timeframe will receive spin diffusion. The information derived from the observed nOe effects can be used to derive inter-atomic spatial information (e.g. distance between atoms).[8] Indeed, the higher the intensity of the STD signal, the closer is the ligand proton to the protein environment, and in turn higher is the number of contacts between that ligand

proton and the protein environment. This concept is similar to that already implemented in PLANTS,[6,7] where a defined distance cut-off is used to generate distance hierarchies so as to score the docking poses.

MD simulations have proven to be successful tools in classical drug discovery projects, allowing implementation of SBDD strategies that fully account for the dynamic properties of relevant protein-ligand complexes.[9-12] In this context, we reasoned that if the molecular dynamics of the ligand within the protein cavities is simulated for a sufficient amount of time, a distance cut-off can be used to define the number of contacts that each proton atom is establishing with the protein environment. The normalization of the number of contacts for the ligand proton characterized by the highest number of contacts (the proton closest to the protein atoms) provides with a measure useful to rank the ligand protons, which can be compared with the experimental STD.

In an ideal setting, if the simulation time is prolonged to infinitum, the ligand-protein complex will be able to explore every configurational microstate and the most stable state will be populated for a longer time, according to the Boltzmann distribution and ergodic hypothesis. This state will contribute to the largest extent to the number of contacts between ligand protons and protein atoms and this, in turn, will drive the computed STD profile. If the computed STD is in agreement with that experimentally determined, this means that the MD simulations converged correctly, sampling the configurational space, and thus it can be employed for further ligand optimization purposes. The obvious pitfall of this approach is that, to ensure convergence, an unbiased molecular dynamics simulation should cover the timeframe of the spin saturation transfer effect, which is in the range of milliseconds to seconds time. This limitation can be overcome by using enhanced sampling techniques (e.g. steered molecular dynamics,[13] umbrella sampling,[14] metadynamics[15] and its derived forms,[16]) which can be used to describe at atomic scale the behaviour of complex systems.[17–21] Among these, we decided to exploit Funnel-Metadynamics[22] (FM, a derived metadynamics form), an emerging technique that has proven to be successful in accelerating the number of back-and-forth events between the ligand bound and free states.[23–25] In FM, the introduction of a funnel-

shaped restraining potential limits the exploration of the solvated state outside the binding pocket, without affecting the sampling into the binding region, thus allowing the thorough characterization of the ligand binding process in a reasonably short wall clock time.[22] Most importantly, in the FM framework no *a priori* information about the ligand binding mode is required.

We thus used FM to reconstruct the relative STD profile for the real case of a protein-ligand complex and used the so obtained data to gain insights into the binding mode of a series of ligands.

We decided to test our integrated approach using as the reference system *O*-acetylserine sulfhydrylase (OASS), a PLP-dependent enzyme for which we have already produced a set of information, such as a consistent SAR for inhibitors and STD data.[26–29] OASS is a potential antibacterial target, as it catalyses the last step of the cysteine biosynthesis pathway in bacteria, and therefore the inhibition of the enzyme is expected to reduce the fitness of the bacteria under oxidative stress conditions.[30] Our group recently reported a series of OASS inhibitors based on the cyclopropyl-carboxylate scaffold.[26–28] While the 3D structure of *St*OASS is known, no X-ray crystal structure of *St*OASS in complex with small-molecule inhibitors has been reported so far and the binding mode of inhibitors can only be predicted by docking experiments. However our experience in the field indicates OASS as a particularly challenging system for docking algorithms, in view of the wideness and flexibility of the binding pocket.[26–28] Thus, we selected OASS-A as a reliable test case for our integrated STD/FM approach and we have selected one of our previously reported inhibitors as a test ligand.[31] Indeed, compound **12b**[31] ((1R,2S)-1-ethyl-2-phenylcyclopropanecarboxylic acid, Chart 1) has an experimental K_d of ≃ 13 μM and the STD profile with *St*OASS-A was previously reported,[31] making it a suitable ligand to be used to test our approach.

Therefore, we firstly proceeded to the reconstruction of the whole binding process between **12b** and *St*OASS-A, by applying the FM technique, which allowed us to accurately predict the binding free energy for such inhibitor (with a 0.95 kcal/mol difference with respect to the experimental one) and to obtain hundreds of thousands binding conformations. The binding conformations were rescored

by an in-house built software (STDc) able to compare the experimental STD profile with that recalculated for each FM-derived binding conformation. This approach also allowed us to consider the ligand-target complex from a dynamic point of view. Indeed, to better clarify the nature of the complex under investigation a small ensemble of protein-ligand complexes (3 structures) were extrapolated by means of STDc. These representative structures allowed us to rationalize the SAR profile of previously published *St*OASS-A inhibitors and, more importantly, to identify a newly druggable sub-pocket, which is conserved across different OASS isoforms and in different orthologues of the enzyme. These findings were used to synthesize a number of derivatives that displayed the expected activity. More importantly, the disclosure of a new druggable sub-pocket paves the way to further extend the chemical space of our class of inhibitors, addressing at the same time the so-called multi-parametric optimization process.[32]

Results and Discussion

Molecular Modelling.

Unbiased MD simulation. In a previous work we reported a docking derived complex between *St*OASS-A and compound **12b**.[31] The *St*OASS-A/**12b** complex was obtained by combining the complete relaxation and conformational exchange matrix analysis of saturation transfer (CORCEMA-ST) theory[5] and docking studies. Even if this complex was instrumental to rationalize preliminary SAR studies, it suffered from some limitations, which made the model useless in the view of further ligand optimization. Indeed, differences between the experimental STD profile and the calculated STD were observed.[31] Such differences were ascribed to the fact that some of the ligand moieties (particularly the aromatic ring) were in close proximity to a flexible loop delimiting the binding pocket, which can adopt different conformations in solution. STD signals are sensitive to protein flexibility, since they are the result of an average signal throughout time. On the contrary, in the docking experiments the conformation of the aforementioned loop was fixed. Therefore, to accurately take into account the effect of the protein dynamics on the calculated STD and to better

clarify the interaction mechanism of compound **12b**, we decided to consider the *St*OASS-A/**12b** complex from a dynamic point of view, by performing preliminary MD simulations studies. Unfortunately, preliminary 120ns MD simulations revealed that the selected complex was not stable under our simulation conditions, since the ligand adopts a different binding conformation after few ns (see Figure S1). Ligand instability is likely due to the fact that *St*OASS-A is a highly flexible enzyme that upon ligand binding can adopt significantly different conformations. So far, three *St*OASS-A X-ray crystal structures have been disclosed: (*i*) a full open ligand free conformation (pdb code: 1OAS) (*ii*) *St*OASS-A in an intermediate conformation in complex with chloride and sulfate anions (pdb code: 1FCJ (chain A)); and (*iii*) a K41A mutant close conformation (pdb code: 1D6S). In a previous work we demonstrated that *St*OASS-A can adopt different conformations with respect to the fully open state, depending on the presence and the nature of the inhibitors.[28] Such conformations are significantly different from those observed in the X-ray crystal structures. Thus, we cannot firmly exclude that compound **12b** can adopt a peculiar binding mode into the *St*OASS-A active site, which cannot be properly described by simply relying on the known X-ray crystal structures. In this context, enhanced sampling methods such as FM can be useful in elucidating the binding mechanism of a ligand to a target of interest, providing a better picture of the dynamic behavior of such complex.

Analysis of the Ligand Unbinding Pathway and Free-Energy Calculation Using FM simulations. In order to obtain preliminary hints on the most probable ligand unbinding pathway, and, in turn, on how the funnel-restrained potential would be positioned in the next stages, Random Accelerated Molecular Dynamics (RAMD) were conducted.[33] This method allows for the identification of the most probable unbinding pathway offering two relevant advantages: (*i*) low computational cost and (*ii*) objective pathway search. RAMD was successfully applied to disclose unbinding pathways for several protein targets.[34–38] In Figure 1A a cartoon representation of the most probable unbinding pathway for compound **12b** is reported. In particular, in six out of ten RAMD simulations the ligand

egressed following similar unbinding pathways, in three simulations partial unbinding events, along the same unbinding route, were obtained. Finally, in one simulation no unbinding events were observed (Figure S2). After that, the funnel-restrained potential was positioned so that the cylindrical portion of the funnel could include the most probable unbinding pathway (Figure 1B). FM allows the well-characterization of the free-energy profile since it increases the number of binding events observed during the sampling.[22] Moreover, when FM is applied, no *a priori* information about the ligand binding mode is required, thus the exploration of the binding site is possible also considering slow protein motions and solvent effects.[22] At the end of the simulations, the free energy landscape can be derived from the history-dependent bias potential added on a few degrees of freedom of the system, namely collective variables (CVs). The FM simulations were performed for 1.5 μs applying the well-tempered metadynamics bias on two CVs: (*i*) the distance between the N backbone atom of Thr72 and the C atom of the carbonyl group of **12b** and (*ii*) the dihedral angle defined by N(Thr72)-N(Asn71)-C2(**12b**)-C1(**12b**) atoms (Figure 1C and Supplementary Information). The Gaussian deposition rate and heights were chosen so as to smoothly conduct FM simulations (see Materials and Methods section), thus allowing protein adaptation upon ligand binding. The statics of two furthers CVs, the projection along the main *z* axis of the funnel and the distance from that axis, were collected along the FM simulations to evaluate the conformational space explored by the ligand inside the funnel. As shown in Figure 2A-B, several re-crossing events (bound to unbound and *viceversa*) were observed (See also Supplementary Movie). Indeed, during simulation the ligand visits several times the bound (CV_1 < 14 Å) and the unbound states (33 <CV_1> 36 Å). Convergence of the calculations was ensured by monitoring the bias deposition and ΔG_b^0 (Figure 2C-D), as reported elsewhere.[23–25] FES (Figure 2B) can be reconstructed as a function of the bias added on the system and once the calculation is converged the absolute binding free-energy can be calculated as a function of the bias added on the system and the absolute binding free-energy can be calculated.

Absolute Binding Free-Energy. The quality and reliability of the free energy profile is tightly linked to the number of re-crossing events observed between different visited states.[22] As depicted in Figure 2A the system explores several times both bound (CV_1 < 14 Å) and unbound (33 < CV_1 > 36 Å) states. ΔG_b^0 was calculated as the free-energy difference of the system with the ligand in the bound and unbound state (Figure 2C) and from the potential mean force $W_{(z)}$ (Figure 2D) obtained from metadynamics (see Material and methods and Figure S3). The estimated ΔG_b^0 for compound **12b** converges to -12.15 ± 1.09 kcal/mol (Figure 2). Considering the analytical correction of 3.8 kcal/mol due to the funnel-restrained potential (see eqs. 1 and 2 in Material and Methods), the final free energy is -8.35 ± 1.09 kcal/mol. The calculated ΔG_b^0 is in good agreement with that derived from eq. 4 (see Material and Methods), which corresponds to -9.30 kcal/mol. Moreover, the -8.35 ± 1.09 kcal/mol value is in line with the experimental ΔG_b values measured for ligands having a K_d falling in the high nm-low μM range, which is typical for weak binders as **12b**.[22,39]

STDc re-scoring and hit-target complexes ensemble. We previously reported the experimental STD profile of compound **12b**.[31] In that work, the saturation received by the different ligand protons was expressed as Group Epitope Mapping (G.E.M.), e.g., the percentage of STD signals normalized with respect to the most saturated signal. The higher the value, the more intimate is the recognition of the ligand portion by the protein binding pocket.[40] Even if such data revealed relevant information about which ligand portions mostly interact with the enzyme active site, the structural details of the complex still remain elusive. Nevertheless, the experimental STD profile can be used to extract, from the FM derived bound state minima, the conformation(s) that closely resemble the experimental profile, providing us with reliable binding conformation(s). In principle, it might be possible to calculate the STD profile for each binding conformation derived from the FM bound state and to compare the calculated G.E.M. with the experimental ones. To date, two main methods have been reported in literature for the *in silico* calculation of the STD/G.E.M. profile. The first method is mainly based on the full calculation of the relaxation and conformational exchange

matrix (CORCEMA, and CORCEMA-ST).[5] Such method proved to be effective in identifying reliable binding modes for different ligand-protein complexes.[41–43] Nevertheless, CORCEMA and CORCEMA-ST have the limitation to be extremely time-consuming (several hours for one ligand-protein complex), and they consider the protein-ligand complexes from a static point of view, thus preventing their application for challenging ligand-protein complexes, where flexibility plays a dominant role. The second method is represented by the PLANTS docking scoring function, where an user-defined empirical scalar value (W_{STD}) is introduced to define several cut-off distances, which are then used to classify protein-ligand contacts with respect to the experimental G.E.M. values.[6,7] The PLANTS methods, although faster compared to CORCEMA and CORCEMA-ST, present relevant limitations, since (*i*) flexibility is not explicitly treated and (*ii*) it strongly depends on docking scoring functions, which are known to have limitations in ligand binding mode identification and ranking. We managed to overcome these limitations by developing STDc, an in-house built script able to process thousands of binding conformations in few hours. The binding conformations are then ranked by their root mean square errors (R.M.S.E) with respect to the experimental G.E.M. values. STDc finally provides the structure of the binding conformations with the closest G.E.M. values to the experimental ones and an ensemble of conformations which better resemble the experimental STD profile. In the latter case the ensemble of ligand binding conformations is projected as density surface into the enzyme active site, so as to better describe the binding cavity regions likely explored by the ligand during the STD experiments (for full details see Material and Methods and Supplementary Information).

First, we proceeded to the reconstruction of the FES along different CVs, using the reweighting algorithm developed by *Bonomi et al.*[44] This method allows the reconstruction of the FES using CVs different from those biased during the metadynamic simulations. The projection along the main *z* axis of the funnel and the distance from that axis were used as new CVs (Figure 3A). This procedure allowed us to better identify all the bound states falling into a minima of ≃ 2 kcal/mol (74747 structures). These structures were then clustered for a total of 462 families, which were

processed by STDc. In Figure 3B, the calculated *vs* the experimental G.E.M. values are reported for the single binding conformation, which closely resemble the experimental values and for the ensemble. The plot reported in Figure 3B highlights how the combination of the FM-STDc leads to a general improvement of the G.E.M. profile reconstruction. Indeed, the average deviation from the experimental G.E.M. profile (blue bars) is, respectively, equal to 0.11, 0.10 and 0.22 for the G.E.M. profile derived from the single conformation (orange bars), ensemble (cyan bars), and that derived from the docking studies (yellow bars), previously reported.[31] Therefore, by applying our protocol, we were able to computationally calculate a STD/G.E.M. profile with an error from the experimental profile reduced by 50% with respect to our previous work.[31] Most importantly, there is quite a good agreement between the experimental profile and that derived from the ensemble. This allowed us to collect a series of findings that previously could not be unravelled. It was possible to: (*i*) identify H_c as the proton that shows the strongest interaction with the enzyme active site, whereas in the previous work[31] this information was missing; (*ii*) better characterize the location of the pendant aromatic ring in position 2, for which it is possible to identify the following trend $H_{2,6}>H_{3,5}\simeq H_4$ (proton atoms of the aromatic ring), whereas in the previous work the signal of H_4 was not properly reproduced. This information is instrumental to correctly plan at which position of the aromatic ring functional groups can be introduced; (*iii*) accurately rank the signals of the protons of the cyclopropane ring with respect to their interaction with the active site, that can be described as such: $H_c>H_b>H_a$ (in our previous work[31] this was not possible due to the limitations of considering the protein as a rigid entity); and (*iv*) the CH_2 G.E.M. of the ethyl branch is properly reproduced, differently from the CH_3 that appears to be the only one which is completely miss-ranked. This is likely due to the great degree of freedom of this group, which can freely rotate inside the *St*OASS-A binding site, as highlighted by high standard deviation of the measurement.

The application of the FM-STDc approach allowed us to identify the binding mode reported in Figure 3C, whose stability was verified over 100ns of standard MD simulations (Figure 3D). The complex resulted significantly more stable with respect to that used at the beginning of our study

(Figure S1). From the analysis of the complex it can be appreciated how the ligand is well accommodated into the enzyme active site, establishing the following key interactions: (*i*) the carboxylic moiety H-bonds the NH proton of the backbone residue Thr72, Asn71 and the OH groups of Thr72 and Thr68 sidechains, in agreement with what previously hypothesized;[26,27,31] (*ii*) the pendant aromatic ring in position 2 protrudes in an accessory hydrophobic pocket previously described;[26] (*iii*) the pendant phenyl ring establishes a π-H-bond interaction with the Asn69 sidechain, (*iv*) the ethyl branch accommodates well into a hydrophobic pocket defined by the methyl group of Thr72 and Thr177 sidechains and the Lys-PLP co-factor. Of greater interest is the observation that during the FM simulations the enzyme readapts upon ligand binding, assuming a fully closed conformation (Figure 3E). Such a conformational change is typical for transaminases (fold-type I) and the members of the β-family of PLP-dependent enzymes (Fold type II). The latter class of enzymes is characterized by the presence of a large and a small flexible domain, separated by a hinge region, which rearranges upon substrate binding.[45–46] The structural comparison of the FM-STDc derived complex with other members of the same PLP-family for which X-ray crystal structures are available (Figure S4) supports the notion that the enzyme assumes a fully closed conformation upon ligand binding. The re-adaptation of the enzyme structure observed during the FM simulation was possible since we chose to perform a highly-smoothed FM protocol (see Material and Methods), allowing protein adaptation during ligand binding. Indeed, a similar conformational change was observed also in our previous work,[28] where we applied standard MD simulations, supporting the notion that such event is likely to occur at low energetic cost, if the ligand properly occupies the enzyme active site. So far, the X-ray crystal structure of wild-type *St*OASS-A in complex with a competitive inhibitor in its closed conformation has never been reported. To the best of our knowledge, our protocol allowed for the first time to propose a reliable model structure describing *St*OASS-A in its closed conformation in complex with an inhibitor. Such structural details are instrumental in order to aid further ligand optimization strategies, paving the way for the identification of improved inhibitors.

SAR profile of previously reported inhibitors. To properly validate our protocol, we firstly proceeded to the rationalization of the SAR profile of previously reported *St*OAAS-A inhibitors. The analysis of the G.E.M. profile derived from the ensemble reported in Figure 3B (cyan bars), and the projection of the density surface of the same ensemble (Figure 4A, see *STDc section in Material and Methods* for details) revealed that two positions can be exploited for functionalization: the CH_3 group on the ethyl branch and the aromatic ring on position 2 of the cyclopropane core. Regarding the modification of the methyl group, this was already exploited in our previous work,[31] leading to the identification of more potent inhibitors. Of great relevance is the observation that our protocol effectively identified such position as one of the most suitable to introduce bulkier substituents, strongly supporting the reliability of the method. In Supplementary Figure S5 the binding mode of previously reported OASS inhibitors[31] is depicted along with a short description of the SAR (See supplementary Information).

Design, Synthetic strategies and generation of derivatives. Once rationalized the SAR profile of previously reported inhibitors we looked for a robust proof that the structural details obtained by our integrated approach could be used to design new *St*OASS-A inhibitors with a desirable pharmacological profile.

The analysis of the density surface reported in Figure 4A revealed the presence of an additional accessory pocket, which can be filled by the introduction of bulkier substituents at the 3'-position. Noteworthy, modifications of the pendant phenyl ring are not reported so far, and our protocol identifies a possible region for further ligand derivatization and optimization. Moreover, the comparison of the X-ray crystal structures so far available for the OASS enzymes revealed that such pocket is conserved across different orthologues and isoforms (Figure S6).

To support the notion that the newly identified pocket (Figure 4) can be exploited to further increase the chemical space of our ligand class, we compared the selected binding mode and the

extrapolated density surface with the X-ray crystal structures of OASS-A from different organisms and in complex with inhibitors (Figure 4B-D and Figure S6). The analysis revealed that in other OASS from different organisms (e.g. *Haemophilus influenzae*[48] and *Mycobacterium tuberculosis*[49]) this accessory pocket is conserved and it can be occupied by bulkier and hydrophobic substituents. Encouraged by these results, we decided to set up a synthetic strategy to easily and quickly prepare a small focused library of derivatives bearing substituents of different nature in position 3', in order to prove this position as the most prone to modification and therefore validate our approach. The easiest strategy to modify compound **12b** in position 3' is the introduction of another aromatic function (Figure 4E) by means of a *Suzuki coupling* reaction. In this way, a small library of derivatives can be easily generated by using different boronic acid fragments. All the purchasable boronic fragments from Sigma-Aldrich and Maybridge were downloaded as sdf files and using the Reactor tool of ChemAxon a library of **12b** derivatives was generated.[50] A total of 113 derivatives were generated and docked into the *St*OASS-A conformation obtained from the STD-FM approach (See Material and Methods for details and Figure S7). Among the aforementioned derivatives only 4 compounds were selected, taking into account their docking scores and visual inspection (See Material and Methods for selection strategies and Figure S7). The dissociation constants for the selected compounds are reported in Table 1. Interestingly, compounds **7a**,**c** and **d** show comparable activities against *St*OASS-A, while **7b** shows a reduced ability to bind *St*OASS-A. To better characterized our newly designed inhibitor we also resolved the racemic mixture of compound **7a**, leading to compounds **10a** and **11a**, which show a slight affinity improvement if compared to **12b**. Interestingly, in our previous work[31] we have demonstrated that *St*OASS-A ligand requires a defined stereochemistry in order to be properly accommodated in the active site, since usually one of the enantiomers is more active than the other (**12b** *vs.* **12c**, Table 1).[31] We were pleased to notice that the modifications planned at position 3' of the phenyl ring allowed to surmount such stereo-chemical requirements, since both enantiomers display equal affinity towards OASS (**10a** and **11a** *vs.* **12b** and **12c**). At this regard, we can speculate that the introduction of bulkier substituents at the

3' position can lead to energetically favourable contacts with the protein active site, overcoming the stereo-chemical requirements of the pocket. Indeed, we have previously reported that *St*OASS-A preferentially binds compounds with specific stereochemistry, whereas the other enantiomers showed only residual affinity. This suggests the possibility that such geometrical requirements can be surmounted if proper ligand modifications are introduced. Since the chiral resolution of the racemate or an enantioselective synthesis are no longer required, a significant enhancement in the preparation of analogues can be reached to expand the series and refine the SAR.

Chemistry.

Styrene oxide **2** was obtained from the commercially available 3-bromostyrene **1** by oxidation with *m*-chloroperbenzoic acid, according to well established protocol (scheme 1).[51] Alkylation of triethyl phosphonoacetate **3** with bromoethane in the presence of sodium hydride afforded ethyl 2-(diethoxyphosphoryl)-3-(*p*-tolyl)propanoate **4** in good yield. Wittig-Horner reaction between precursors **2** and **4**, in anhydrous dimethoxyethane at 90 °C and in the presence of *n*-buthyllithium as the base, allowed us to obtain the required key intermediate *trans* ethyl 2-(3-bromophenyl)-1-ethylcyclopropanecarboxylate **5**. Suzuki-Miyaura cross-coupling reaction between the proper boronic acid and compound **5** gave derivatives **6a-d**, that were treated with LiOH at 100 degrees under microwave irradiation to afford the final compounds **7a-d** in good overall yields. Diastereoisomers **8a** and **9a** were obtained from racemic **7a**, that was reacted in dichloromethane at room temperature with chiral precursor (R)-(−)-2-Phenylglycinol, according to a standard amide synthesis procedure (scheme 2). After separation of the two diastereoisomers by flash column chromatography, hydrolysis in acidic conditions of **8a** and **9a** gave the desired enantiomers **10a** and **11a** in good yields.

Conclusion

We have herein reported for the first time the development of an integrated method based on the combination of FM experiments and STD approaches. This method can be used to identify reliable ligand-target complexes which can be exploited to drive and implement drug design strategies. As a test case, we have applied the newly developed FM-STD approach, to study the interaction of *St*OASS-A and a weak inhibitor previously identified by us. Several limitations of the STD method when used alone were clearly overcome by the use of this combined approach, which indeed allowed to (*i*) elucidate the binding mechanism of **12b** to *St*OASS-A and calculate the binding free energy for such complex which is in good agreement with the experimental one; (*ii*) identify a binding mode able to rationalize previously reported biological data; (*iii*) identify a sub-pocket conserved among different OASS orthologues and isoforms, which can be exploited to further functionalize this chemical class. Unexpectedly, this study led to the synthesis of inhibitors that, differently from the parent compounds, are not affected by strict stereochemical requirements, firmly simplifying the overall synthetic process.

The results obtained support the notion that besides classical drug design strategies, the STD-FM approach can be applied to relevant pharmaceutical targets for which only weak binders are reported and structural details about the ligand-protein complex are unknown. Moreover, our STDc script can be easily implemented in classical MD simulations or docking experiments as rescoring function in order to better rank different binding conformations of the same complex, and in turn, to identify reliable ligand-target complexes. Overall, these findings suggest that STD-FM can be of extremely high interest when applied to some of the most common medicinal chemistry cases, such as in fragment-based drug design approaches or right after carrying out HTS campaigns that furnishes a number of weakly binders. Moreover, STD-FM can be complementary applied in more challenging contexts such as the identification of binding cavities of modulators for intrinsically disordered proteins, for which conventional structure-based approaches cannot be applied.[52]

Nowadays almost all medicinal chemistry lab are integrated with NMR and computational facilities, where STD-FM could be easily performed without any further technological improvements or cost, providing at the same time useful structural hints to speed up the drug optimization process.

Experimental Section

Computational Methods.

*StOASS-A/**12b** complex generation and Plain MD simulation.* The initial *St*OASS-A/**12b** was obtained by combining the complete relaxation and conformational exchange matrix analysis of saturation transfer (CORCEMA-ST) theory[5] and docking studies, as previously reported.[26] In particular, the 1OAS X-ray crystal structure was used. The complex was prepared by applying the protein preparation protocol available in Maestro9.1.[53] The complex was further minimized using Macromodel,[53] the PRCG method, OPLSS-2005[54] as force field and water as implicit solvent model.

The system, thus obtained, was parameterized using *ff14SB* and *gaff* as force fields.[55] **12b** atomic single charges were computed using Gaussian 09, HF/6-31** as theory level, (http://gaussian.com) and fitted with restrained electrostatic potential (RESP). Atom type and parameters for **12b** were retrieved from *gaff*.[55] The **PLP** cofactor was simulated in its neutral state as previously reported,[28] the partial charges of the **PLP** were assigned using Gaussian 09, UB3LYP/6-31G* as theory level, (http://gaussian.com) and fitted with restrained electrostatic potential (RESP). Atom type and parameters for **PLP** were retrieved from *gaff*.[55] The system was solvated using leap and the TIP3P water model.

Plain MD simulations were performed using NAMD2.10[56] and a relaxation protocol based on six steps. The system was initially equilibrated performing 50ns in the NPT ensemble, after that a production phase of 120 ns was performed in the NVT ensemble.

Random Accelerated Molecular Dynamics simulations (RAMD). Recently RAMD have been successfully applied to disclose putative ligand unbinding pathways for relevant biological targets.[34–38] In the RAMD framework a randomly oriented force (*f*) is applied to the center of mass of a group of atoms (ligand) for a defined short amount of time-steps (*N*). Currently, RAMD can be performed in two schemes: (*i*) pure RAMD simulations in which the randomly-oriented acceleration is applied continuously, and (ii) combined RAMD-MD simulations in which RAMD steps alternate with standard MD steps. For the *St*OASS-A/**12b** complex the combined RAMD-MD simulations scheme was applied. The starting configuration was represented by the complex generated after the 120 ns of plain MD simulations, the constant force was applied on all the ligand heavy atoms and ten independent RAMD simulations were performed. Each ligand unbinding pathway was drawn by tracing the center of mass of the ligand heavy atoms using VMD1.9.2.[57] The funnel-restrained potential was applied so as to include the most probable unbinding route.

Funnel-metadynamics. The starting configuration for FM was represented by the last frame of the 120ns of the production phase. PLUMED1.3[57] plugin was used to run metadynamics calculations. Since the funnel-restrained potential is fixed in the space, the *St*OASS-A diffusion was avoided by imposing positional restrains on seven Cα atoms (Ala29, Ile44, Glu153, Val170, Val202, Val246, Ile297). The bias was added on a distance (CV_1) and dihedral angle (CV_2). For CV_1 and CV_2 Gaussian widths of 0.05 Å and 0.09 rad were respectively applied. The Gaussian deposition frequency of 1 Kcal mol^{-1} every 2 ps was initially applied, and gradually decreased on the basis of the adaptive bias with a ΔT of 2700 K. During FM simulations, an upper limit of 31 Å and a lower limit of -3 Å were imposed along the funnel *z* axis, so as to avoid ligand exit from the funnel-restrained potential. Trajectories were analysed with VMD1.9.2.[57]

Absolute binding free energy. In FM simulations when the ligand is inside the cone portion of the restraint potential, no external potential is applied, and the system works under the standard well-

tempered metadynamics framework. In this way, the potential does not affect the exploration of the enzyme active site by the ligand. When the ligand is in the unbound state $z > z_{cc}$ a cylindrical restraint potential is applied.[22] Using this protocol the binding constant K_b in presence of the restraint is given by

$$K_b = \pi R_{cyl}^2 \int_{site} dz e^{-\beta[W_{(z)} - W_{ref}]} \quad [1]$$

where πR_{cyl}^2 is the surface of the cylinder used as restraint potential while the potential $W_{(z)}$ and its value in the unbound state, W_{ref} can be obtained from the PMF (Potential Mean Force) obtained from FM calculation. $\beta = (k_b T)^{-1}$ is constant, where k_b is the Boltzman constant and T the temperature of the system. K_b and ΔG_b^0 are directly related through the formula reported below:

$$\Delta G_b^0 = -\frac{1}{\beta} \ln(C^0 K_b) \quad [2]$$

where ΔG_b^0 is the protein-ligand binding free energy, K_b the binding constant and $C^0 = 1/1660$ Å^{-3} is the standard concentration 1M. As reported in ref. 22 using eq. 1 in FM simulation eq. 2 can be rearranged

$$\Delta G_b^0 = \Delta G - \frac{1}{\beta} ln\left(\pi R_{cyl}^2 C^0\right) \quad [3]$$

$$\Delta G^0 = -RT ln K_{eq} \quad [4]$$

ΔG is the free energy difference between the bound and unbound states, and the absolute protein-ligand binding free energy (ΔG_b^0) is equal to ΔG minus the analytical correction in eq. 3. ΔG^0 was derived from eq. 4 using RT=0.57

The reweighting algorithm. To identify the main ligand binding conformations falling into a minima of ≃ 2 kcal/mol we applied a recently developed reweighting procedure.[44] This method allows the reconstruction of the FES using CVs different from those biased during the metadynamics simulations. Once the metadynamics simulation has converged, using the newly computed probability distribution, the FES can be reconstructed as a function of the newly selected CVs. The projection along the main *z* axis of the funnel and the distance from that axis were used as new CVs.

STDc. STDc is an in-house built script able to calculate the STD/G.E.M profile for a given ligand-target complex and to compare it with the experimental one. It is not a standalone script and it works in conjunction with *cpptraj*,[55] *catdcd*[59] and *plumed.*[58]

In the first step a MD trajectory is clustered using cpptraj and user defined parameters. In our case the average-linkage cluster algorithm with a R.M.S.D. cut-off of 1 Å (calculated on the ligand heavy atoms and the Cα atoms of the *St*OASS-A small domain and the flexible loop defined by 215-240 residues) were used. Cpptraj turns out with representative structures of each cluster obtained. Subsequently, every representative structures generated at the previous step is processed by plumed.[58] For each ligand hydrogen atom the contact with the protein environment are estimated by using eq. 5.

$$s_{(d)} = \sum_j \frac{1-\left(\frac{d_{ij}}{r_0}\right)^n}{1-\left(\frac{d_{ij}}{r_0}\right)^m}$$

[5]

The switching function reported in eq. **5** takes in consideration two groups of atoms, the ligand proton atom (*i*) and all the protein atoms (*j*). In this way $\mathbf{s}_{(d)}$ measures if the ligand proton atom *i* is

found at a certain distance d_{ij} from protein atoms j. When d_{ij} is less or equal to r_0 (distance cut-off) $\mathbf{s}_{(d)}$ is equal to 1 and the contact with the atoms is formed, while for $d_{ij} > r_0$ the function decays smoothly to 0 and the nature of the observed contacts become less relevant.

We used n=6, m=12, r_0=3.0 so as for d_{ij}=10 Å $\mathbf{s}_{(d)}$ becomes 0 (See Supplementary Information and Figure S8). Indeed the distance cut-off for the long range interaction during the MD simulation was set to 10 Å, which is also the distance cut-off through which nOe effects is still observable in STD experiments.[2-6] At this point, a list reporting each ligand hydrogen atom with the corresponding total sum of contacts is stored and the values are normalized for the hydrogen atom with the highest number of contacts (the ligand proton atom closest to the protein atoms). For each hydrogen atom the R.M.S.E value is calculated with respect to the corresponding experimental value and the representative structures are ranked on the basis of their average R.M.S.E. Finally, STDc turns out with the representative structure having the lowest average R.M.S.E. and with an ensemble of representative structures extrapolated by means of a R.M.S.E. cut-off defined by the users. The ensemble of structures is then used to project a surface representation of the protein binding site where the ligands establishes protein contacts that mostly resemble the experimental STD/G.E.M. profile (See Supplementary Information and Figure S8).

Library Design and Compound selection for synthesis. Two library of commercially available boronic acid fragments were retrieved from Sigma-Aldrich (38) and Maybridge (164). These fragments were optimized with the Standardizer tool of ChemAxon, which allows for the geometrical optimization and protonation state evaluation. Duplicates were removed and the optimized fragments were conjugated to **12b** by means of Reactor and Suzuki coupling reaction. The point of substitution on **12b** was represented by the 3' position on the pendant phenyl ring. A total of 113 derivatives were obtained and used for the subsequent docking studies. The *St*OASS-A-**12b** binding conformation extracted with STDc was used as reference structure. The protein was prepared with the protein preparation tool of Maestro[53] and the docking grid was centred on **12b**.

The compound collection of derivatives was prepared with *ligprep*[53] each docking run was carried out with Glide[53] using the standard precision (SP) method, and the van der Waals scaling factor of nonpolar atoms was set to 0.8. For each derivative one binding conformation was generated. A total of 4 compounds were selected for synthesis and *in vitro* evaluation on the basis of the docking score and after visual inspection (Figure S7).

Chemistry

General Information. All the reagents were purchased from Sigma-Aldrich and Alfa-Aesar at reagent purity and, unless otherwise noted, were used without any further purification. Dry solvents used in the reactions were obtained by distillation of technical grade materials over appropriate dehydrating agents. Reactions were monitored by thin layer chromatography on silica gel-coated aluminum foils (silica gel on Al foils, Supelco Analytical, Sigma- Aldrich) at both 254 and 365 nm wavelengths. Where indicated, intermediates and final products were purified through silica gel flash chromatography (silica gel, 0.040−0.063 mm), using appropriate solvent mixtures.

^{1}H NMR and ^{13}C NMR spectra were recorded on a Bruker Avance spectrometer at 400 and 100 MHz, respectively, with TMS as internal standard. ^{1}H NMR spectra are reported in this order: multiplicity and number of protons. Standard abbreviation indicating the multiplicity was used as follows: s = singlet, d = doublet, dd = doublet of doublets, t = triplet, q = quadruplet, m = multiplet and br = broad signal. HPLC/MS experiments were performed with HPLC, Agilent 1100 series, equipped with a Waters Symmetry C18, 3.5 μm, 4.6 mm × 75 mm column; and MS, Applied Biosystem/MDS SCIEX, with API 150EX ion source. HRMS experiments were performed with LTQ Orbitrap XL Thermo.

All compounds were tested as 95−100% purity samples (by HPLC/ MS).

2-(3-bromophenyl)oxirane (2) m-CPBA (994 mg, 5.76 mmol) was added to a solution of 3-bromostyrene **1** (600.3 μL, 4.61 mmol) in chloroform (6.9 mL) stirring on ice. Reaction mixture was stirred for 3h on ice and 17h at room temperature. As the reaction wasn't complete further 1 eq (795mg, 4.61 mmol) of m-CPBA was added to the reaction mixture that was previously cooled down in an ice bath. After stirring for another 3h at room temperature it was stopped with the addition of a 1M solution of NaOH under ice cooling conditions. The product was extracted with chloroform and the combined organic layers were dried over anhydrous sodium sulphate, filtered and concentrated under reduced pressure to afford the product as a colourless oil in quantitative yield.

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.42 (d, *J* = 4.3 Hz, 2H), 7.21 (d, *J* = 4.8 Hz, 2H), 3.89 – 3.72 (m, 1H), 3.21 – 3.08 (m, 1H), 2.76 (dd, *J* = 5.5, 2.5 Hz, 1H).

Ethyl 2-(diethoxyphosphoryl)butanoate (4): was prepared as previously described and the analytical data matched with those already reported.[51]

***Trans*-ethyl-2-(3-bromophenyl)-1-ethylcyclopropane-1-carboxylate (5):** was prepared as previously described and the analytical data matched with those already reported.[51]

General Procedure for Suzuki-Miyaura Reaction. Method A:

Into a flask under argon stirred at room temperature was added a solution of compound **5** (1.00 equiv, 0.52 mmol) in a mixture of toluene/methanol/water (80/18/2 v/v, 40mL/mmol), tetrakis(triphenylphosphine)palladium (0.15 equiv, 0.08 mmol), phenylboronic acid (3.00 equiv, 1.56 mmol), and K_2CO_3 2M (1.6 equiv, 0.832 mmol). The reaction mixture was refluxed overnight at 110 °C and then filtered through a plug of celite. The filtrate was concentrated under reduced pressure and the residue purified by flash column chromatography on silica gel. Purity, yields and elution conditions are reported below.

***Trans*-ethyl 2-([1,1'-biphenyl]-3-yl)-1-ethylcyclopropane-1-carboxylate (6a):** Purified by flash chromatography (petroleum ether/ Ethyl acetate 99:1); colourless oil; 81% yield.

^{1}H NMR (300 MHz, $CDCl_3$) δ 7.66 – 7.54 (m, 2H), 7.51 – 7.29 (m, 6H), 7.18 (d, J = 7.6 Hz, 1H), 4.21 (dq, J = 7.1, 3.7 Hz, 2H), 2.87 (dd, J = 9.1, 7.1 Hz, 1H), 1.81 – 1.54 (m, 2H), 1.30 (t, J = 7.1 Hz, 3H), 1.22 (dd, J = 7.1, 4.6 Hz, 1H), 0.98 – 0.85 (m, 3H).

***Trans*-ethyl 2-(3-(2-aminopyrimidin-5-yl)phenyl)-1-ethylcyclopropane-1-carboxylate (6b):** Purified by flash chromatography (petroleum/ethyl acetate from 75:25 to 50:50 + 0.1% of thriethylamine); slightly yellow powder; 96% yield.

^{1}H NMR (300 MHz, $CDCl_3$) δ 8.52 (s, 2H), 7.54 – 7.28 (m, 3H), 7.19 (d, *J* = 7.0 Hz, 1H), 5.25 (s, 2H), 4.34 – 4.06 (m, 2H), 2.85 (t, *J* = 8.0 Hz, 1H), 1.77 – 1.55 (m, 2H), 1.30 (t, *J* = 7.1 Hz, 3H), 1.19 (dd, *J* = 6.9, 4.8 Hz, 1H), 0.92 – 0.83 (m, 3H).

General Procedure for Suzuki-Miyaura Reaction. Method B:

A microwave tube was charged with [1,1'-Bis(diphenylphosphino)ferrocene]dichloropalladium(II) (0.02 equiv, 0.0134 mmol), the proper boronic acid pinacol ester (2 equiv, 0.134 mmol), and caesium carbonate (2 equiv, 0.134 mmol). The tube was purged with argon three times and then a solution of compound **5** (1 equiv, 0.067 mmol) in a mixture of dimethoxyethane/water 5:1 (175 μL) was injected in the reaction mixture that was then heated in a microwave reactor for 30 min at 140°C at 100 Watts power. Afterwards, reaction mixture was filtered through a plug of celite, the filtrate was concentrated under reduced pressure and the residue purified by flash column chromatography. Purity, yields and elution conditions are reported below.

***Trans*-ethyl 2-(3-(1H-pyrazol-4-yl)phenyl)-1-ethylcyclopropane-1-carboxylate (6c):** Purified by flash chromatography (petroleum/ethyl acetate from 75:25 to 50:50 + 0.1% of triethylamine); slightly orange oil; 42% yield.

^{1}H NMR (300 MHz, $CDCl_3$) δ 7.67 – 7.28 (m, 5H), 7.04 (d, *J* = 6.5 Hz, 1H), 4.20 (ddd, *J* = 12.7, 7.1, 3.5 Hz, 2H), 2.87 – 2.69 (m, 1H), 1.77 – 1.59 (m, 2H), 1.35 – 1.24 (m, 4H), 1.17 (dd, *J* = 7.1,

4.7 Hz, 1H), 0.95 – 0.77 (m, 3H). HRMS (ESI): calculated for $C_{17}H_{20}N_2O_2$ [M+H] 285.15975 found 285.15948.

***Trans*-ethyl 1-ethyl-2-(3-(3-methyl-1H-pyrazol-4-yl)phenyl)cyclopropane-1-carboxylate (6d):** Purified by flash chromatography (petroleum/ethyl acetate from 75:25 to 50:50 + 0.1% of triethylamine); pale orange oil; 40% yield.

^{1}H NMR (300 MHz, $CDCl_3$) δ 7.69 (bs, 1H), 7.37 – 7.26 (m, 3H), 7.12 (d, *J* = 6.5 Hz, 1H), 4.20 (ddd, *J* = 12.7, 7.1, 3.5 Hz, 2H), 2.99 – 2.93 (m, 1H), 2.47 (s, 3H) 1.79 – 1.73 (m, 2H), 1.35 – 1.24 (m, 4H), 1.17 (dd, *J* = 7.1, 4.7 Hz, 1H), 0.95 – 0.77 (m, 3H).

HRMS (ESI): calculated for $C_{18}H_{22}N_2O_2$ [M+H] 298,1681 found 298,3860.

General procedure for ethyl ester hydrolysis: a solution of the ethyl ester compound (1 equiv, 0.59 mmol) in a mixture of THF/MeOH/H_2O in the proportions 3:1:1 was added to a microwave tube previously charged with LiOH (4 equiv, 3.54 mmol). Reaction mixture was heated in a microwave reactor at 100°C for 10 min. Then the solvent was evaporated and the residue was acidified with HCl 3N. Extractions with ethyl acetate were made and the combined organic layers were dried over anhydrous sodium sulphate, filtered and concentrated under reduced pressure to give a crude material, that is purified by flash column chromatography on silica gel to give the desired acid.

***Trans*-2-([1,1'-biphenyl]-3-yl)-1-ethylcyclopropane-1-carboxylic acid (7a):** Purified by flash chromatography (dichloromethane/methanol 99:1); orange oil; 83% yield.

^{1}H NMR (300 MHz, $CDCl_3$) δ 7.64 – 7.53 (m, 2H), 7.50 – 7.33 (m, 6H), 7.20 (d, *J* = 7.5 Hz, 1H), 3.00 (t, *J* = 8.2 Hz, 1H), 1.84 – 1.73 (m, 1H), 1.73 – 1.63 (m, 1H), 1.28 (ddd, *J* = 20.4, 10.6, 5.9 Hz, 2H), 1.04 – 0.79 (m, 3H).

^{13}C NMR (75 MHz, $CDCl_3$) δ 181.33, 140.94, 140.66, 136.89, 128.50, 128.36, 127.93, 127.88, 127.12, 126.90, 125.49, 33.14, 30.93, 21.34, 18.31, 11.38.

HRMS (ESI): calculated for $C_{18}H_{18}O_2$ [M+H] 265.12231 found 265.12341.

***Trans*-2-(3-(2-aminopyrimidin-5-yl)phenyl)-1-ethylcyclopropane-1-carboxylic acid (7b):**

Purified by trituration with diethyl ether; white powder; 50% yield.

^{1}H NMR (400 MHz, MeOD) δ 8.83 (s, 2H), 7.50 (d, *J* = 7.7 Hz, 2H), 7.45 (dd, *J* = 18.9, 11.5 Hz, 1H), 7.30 (d, *J* = 7.6 Hz, 1H), 2.97 – 2.81 (m, 1H), 1.70 – 1.58 (m, 1H), 1.54 (dd, *J* = 13.9, 7.0 Hz, 1H), 1.36 (dd, *J* = 6.9, 4.9 Hz, 1H), 1.02 – 0.90 (m, 1H), 0.91 – 0.74 (m, 3H).

^{13}C NMR (101 MHz, MeOD) δ 178.20, 156.96, 140.07, 133.28, 130.67, 130.32, 128.16, 125.62, 124.92, 32.98, 32.42, 22.71, 18.22, 11.84.

HRMS (ESI): calculated for $C_{16}H_{17}N_3O_2$ [M+H] 282.12370 found 282.12488.

***Trans*-2-(3-(1H-pyrazol-4-yl)phenyl)-1-ethylcyclopropane-1-carboxylic acid (7c):**

Purified by flash chromatography (dichloromethane/methanol 99:1 + 0.1% of formic acid); slightly yellow powder; 21% yield.

^{1}H NMR (400 MHz, $CDCl_3$) δ 9.24 (s, 1H), 7.89 (s, 1H), 7.36 (d, *J* = 28.4 Hz, 3H), 7.09 (s, 1H), 2.96 (s, 1H), 1.75 (d, *J* = 23.1 Hz, 1H), 1.25 (dd, *J* = 20.5, 8.0 Hz, 2H), 0.94 (s, 3H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 180.49, 147.76, 146.79, 137.77, 133.41, 132.29, 128.94, 127.50, 127.26, 124.50, 33.22, 31.40, 21.92, 18.49, 11.85.

HRMS (ESI): calculated for $C_{15}H_{16}N_2O_2$ [M+H] 255.11280 found 255.11413.

***Trans*-1-ethyl-2-(3-(3-methyl-1H-pyrazol-4-yl)phenyl)cyclopropane-1-carboxylic acid (7d):**

Purified by flash chromatography (dichloromethane/methanol 99:1 + 0.1% of formic acid); light yellow powder; 25% yield.

^{1}H NMR (300 MHz, $CDCl_3$) δ 7.71 (bs, 1H), 7.35 – 7.23 (m, 3H), 7.15 (d, *J* = 6.5 Hz, 1H), 2.97 – 2.92 (m, 1H), 2.45 (s, 3H) 1.81 – 1.75 (m, 2H), 1.37 – 1.26 (m, 1H), 1.13 (dd, *J* = 7.1, 4.7 Hz, 1H), 0.93 – 0.74 (m, 3H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 181.47, 147.72, 147.65, 138.23, 134.23, 133.49, 128.94, 127.50, 127.26, 122.48, 33.22, 31.40, 21.92, 20.23, 18.49, 11.85.

HRMS (ESI): calculated for $C_{16}H_{18}N_2O_2$ [M+H] 270,1368 found 270,3320.

Synthesis of 2-([1,1'-biphenyl]-3-yl)-1-ethyl-N-((R)-2-hydroxy-1-phenylethyl)cyclopropane-1-carboxamide (8a and 9a):

(R)-(-)-2-Phenylglycinol (43.21 mg, 0.315 mmol), TBTU (67.42 mg, 0.21 mmol), EDC*HCl (60.4 mg, 0.315 mmol) and dry triethylamine (29.26 μL, 0.21 mmol) were added to a solution of compound **7a** in anhydrous dichloromethane, stirring on ice and under nitrogen. Reaction mixture was stirred for 1h at 0°C and about 6h at room temperature until consumption of the starting material as revealed by TLC. Then, a saturated solution of ammonium chloride was carefully added and extractions with dichloromethane were performed. The combined organic layers were dried over anhydrous sodium sulphate, filtered and concentrated under reduced pressure to be purified by flash column chromatography on silica gel using a gradient of ethyl acetate in petroleum ether (9:1 - 1:1).

8a: White solid (45% yield)

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.60 (d, *J* = 7.4 Hz, 2H), 7.53 – 7.28 (m, 10H), 7.17 (dd, *J* = 26.3, 8.5 Hz, 1H), 6.57 (d, *J* = 6.6 Hz, 1H), 5.14 (dd, *J* = 11.2, 5.0 Hz, 1H), 3.93 (d, *J* = 4.8 Hz, 2H), 3.02 (s, 1H), 2.95 – 2.83 (m, 1H), 1.67 (dt, *J* = 13.9, 7.0 Hz, 1H), 1.56 (ddd, *J* = 15.6, 12.3, 5.5 Hz, 1H), 1.24 – 1.08 (m, 2H), 1.00 – 0.88 (m, 3H).

9a: White solid (16% yield)

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.70 – 7.50 (m, 2H), 7.47 – 7.30 (m, 10H), 7.18 (dd, *J* = 13.3, 4.9 Hz, 1H), 6.55 (d, *J* = 6.6 Hz, 1H), 5.14 (dd, *J* = 11.4, 4.9 Hz, 1H), 3.90 (dd, *J* = 18.1, 5.3 Hz, 2H), 2.92 – 2.82 (m, 1H), 1.69 (dt, *J* = 43.0, 21.5 Hz, 1H), 1.55 (td, *J* = 14.7, 7.1 Hz, 1H), 1.26 – 1.08 (m, 2H), 0.94 (t, *J* = 7.4 Hz, 3H).

General procedure for amide hydrolysis:

The starting material (**8a**, **9a**) (1 equiv, 0.047 mmol) was solubilized in a mixture 1:1 of 3N H_2SO_4/ Dioxane and stirred overnight at 100°C. Water was carefully added to the reaction mixture after cool it down until room temperature. Extractions with dichloromethane were performed and the combined organic layers were dried over anhydrous sodium sulphate, filtered and concentrated under reduced pressure to be purified by flash column chromatography.

(-)-Trans-2-([1,1'-biphenyl]-3-yl)-1-ethylcyclopropane-1-carboxylic acid (10a): purified by flash column chromatography (dichloromethane/methanol (99:1); colourless oil; 56% yield.

^{1}H NMR (300 MHz, $CDCl_3$) δ 7.61 (d, *J* = 7.2 Hz, 2H), 7.53 – 7.30 (m, 6H), 7.23 (t, *J* = 8.6 Hz, 1H), 3.02 (t, *J* = 8.1 Hz, 1H), 1.87 – 1.75 (m, 1H), 1.75 – 1.66 (m, 1H), 1.39 – 1.23 (m, 2H), 0.95 (d, *J* = 2.4 Hz, 3H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 182.03, 141.38, 141.10, 137.32, 128.92, 128.78, 128.35, 128.29, 127.54, 127.32, 125.91, 33.58, 31.39, 21.76, 18.74, 11.79.

HRMS (ESI): calculated for $C_{18}H_{18}O_2$ [M+H] 265.12231 found 265.12332.

(+)-Trans-2-([1,1'-biphenyl]-3-yl)-1-ethylcyclopropane-1-carboxylic (11a): purified by flash column chromatography (dichloromethane/methanol (99:1); Colourless oil (42% yield).

^{1}H NMR (300 MHz, $CDCl_3$) δ 7.59 (d, *J* = 7.3 Hz, 2H), 7.53 – 7.32 (m, 6H), 7.20 (d, *J* = 7.6 Hz, 1H), 3.00 (t, *J* = 8.1 Hz, 1H), 1.86 – 1.74 (m, 1H), 1.74 – 1.63 (m, 1H), 1.37 – 1.24 (m, 2H), 0.99 – 0.76 (m, 3H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 181.25, 141.07, 140.80, 137.01, 128.61, 128.47, 128.04, 127.99, 127.23, 127.01, 125.60, 33.23, 31.04, 21.47, 18.42, 11.48.

HRMS (ESI): calculated for $C_{18}H_{18}O_2$ [M+H] 265.12231 found 265.12346.

Protein preparation and determination of dissociation constants

StOASS-A expressed in E. coli was purified by ion metal affinity chromatography as described previously.[60] The protein preparation was active and more than 95% pure based on SDS-PAGE analysis. The affinity of the selected compounds for StOASS-A was measured by a fluorimetric method published elsewhere.[61] Briefly, a solution containing 1 μM StOASS-A in 100 mM Hepes pH 7 was titrated with increasing concentrations of compound. The fluorescence emission intensity of the PLP cofactor at 500 nm upon excitation at 412 nm was collected after each addition, subtracted by the blank and normalized by the protein dilution. The dependence of the emission intensity at 500 nm on the concentration of the compound was fitted to a binding isotherm[25] to calculate the dissociation constant of the protein-ligand complex. As demonstrated elsewhere,[21,25] the fluorimetric method allows the calculation of the intrinsic dissociation constant of a competitive inhibitor for the enzyme and the calculated K_d is in very good agreement with the IC_{50} measured by activity assays.

Associated Content

The Supporting Information is available free of charge on the

ACS Publications website

Author Information

Corresponding Author

*E-mail: agostino.bruno@unipr.it; agostino.bruno@ifom.eu

Author Contribution

■J.M and G.A. contributed equally

Acknowledgments

The work described in this paper was carried out under the MSCA-ITN-2014-ETN project INTEGRATE (grant number 642620). All the authors thank Prof. Andrea Mozzarelli for critical reading and discussion. A.B. thanks CINECA consortium for the allocation of computational resources.

Abbreviations

STD, Saturation Transfer Difference; G.E.M., Group Epitope Mapping; MD, Molecular Dynamics; FM, Funnel-Metadynamics; OASS, O-acetylserine sulfhydrylase; PLP, pyridoxal 5′-phosphate; SAR, structure−activity relationship.

Authors will release the atomic coordinates and experimental data upon article publication. STDc will be released upon request.

References

(1) Meyer, B.; Peters, T. NMR Spectroscopy Techniques for Screening and Identifying Ligand Binding to Protein Receptors. *Angew. Chem. Int. Ed Engl.* **2011**, *42* (8)**,** 864–890.

(2) Viegas, A.; Manso, J.; Nobrega, F. L.; Cabrita, E. J. Saturation-Transfer Difference (STD) NMR: A Simple and Fast Method for Ligand Screening and Characterization of Protein Binding. *J. Chem. Educ.* **2011**, *88* (7), 990–994.

(3) Streiff, J. H.; Juranic, N. O.; Macura, S. I.; Warner, D. O.; Jones, K. A.; Perkins, W. J. Saturation Transfer Difference Nuclear Magnetic Resonance Spectroscopy as a Method for Screening Proteins for Anesthetic Binding. *Mol. Pharmacol.* **2004**, *66* (4), 929–935.

(4) Wang, Y.-S.; Liu, D.; Wyss, D. F. Competition STD NMR for the Detection of High-Affinity Ligands and NMR-Based Screening. *Magn. Reson. Chem. MRC* **2004**, *42* (6), 485–489.

(5) Jayalakshmi, V.; Krishna, N. R. Complete Relaxation and Conformational Exchange Matrix (CORCEMA) Analysis of Intermolecular Saturation Transfer Effects in Reversibly Forming Ligand-Receptor Complexes. *J. Magn. Reson. San Diego Calif 1997* **2002**, *155* (1), 106–118.

(6) Korb, O.; Möller, H. M.; Exner, T. E. NMR-Guided Molecular Docking of a Protein-Peptide Complex Based on Ant Colony Optimization. *ChemMedChem* **2010**, *5* (7), 1001–1006.

(7) Korb, O.; Stützle, T.; Exner, T. E. Empirical Scoring Functions for Advanced Protein-Ligand Docking with PLANTS. *J. Chem. Inf. Model.* **2009**, *49* (1), 84–96.

(8) Ni, F.; Scheraga, H. A. Use of the Transferred Nuclear Overhauser Effect To Determine the Conformations of Ligands Bound to Proteins. *Acc. Chem. Res.* **1994**, *27* (9), 257–264.

(9) Martínez-Rosell, G.; Giorgino, T.; Harvey, M.J.; de Fabritiis, G. Drug Discovery and Molecular Dynamics: Methods, Applications and Perspective Beyond the Second Timescale. *Curr Top. Med. Chem.* **2017**, *17*(23), 2617-2625.

(10) Ganesan, A.; Coote, M.L.; Barakat, K. Molecular Dynamics-driven Drug Discovery: Leaping Forward with Confidence. *Drug Discov. Today*. **2017**, *22*(2), 249-269.

(11) 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.

(12) Stornaiuolo, M.; Bruno, A.; Botta, L.; La Regina, G.; Cosconati, S.; Silvestri, R.; Marinelli, L.; Novellino, E. Endogenous vs Exogenous Allosteric Modulators in GPCRs: A Dispute for Shuttling CB1 among Different Membrane Microenvironments. *Sci. Rep.,* **2015**, *20*(5), 15453.

(13) Jarzynski, C. Nonequilibrium Equality for Free Energy Differences. *Phys. Rev. Lett.* **1997**, *78* (14), 2690–2693.

(14) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* **1992**, *13* (8), 1011–1021.

(15) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci.* **2002**, *99* (20), 12562–12566.

(16) Valsson, O.; Tiwary, P.; Parrinello, M. Enhancing Important Fluctuations: Rare Events and Metadynamics from a Conceptual Viewpoint. *Annu. Rev. Phys. Chem.* **2016**, *67* (1), 159–184.

(17) Bruno, A.; Scrima, M.; Novellino, E.; D'Errico, G.; D'Ursi, A. M.; Limongelli, V. The Glycan Role in the Glycopeptide Immunogenicity Revealed by Atomistic Simulations and Spectroscopic Experiments on the Multiple Sclerosis Biomarker CSF114(Glc). *Sci. Rep.* **2015**, *5*,

9200.

(18) Grazioso, G.; Limongelli, V.; Branduardi, D.; Novellino, E.; De Micheli, C.; Cavalli, A.; Parrinello, M. Investigating the Mechanism of Substrate Uptake and Release in the Glutamate Transporter Homologue Glt(Ph) through Metadynamics Simulations. *J. Am. Chem. Soc.* **2012**, *134* (1), 453–463.

(19) Lelimousin, M.; Limongelli, V.; Sansom, M. S. P. Conformational Changes in the Epidermal Growth Factor Receptor: Role of the Transmembrane Domain Investigated by Coarse-Grained MetaDynamics Free Energy Calculations. *J. Am. Chem. Soc.* **2016**, *138* (33), 10611–10622.

(20) Limongelli, V.; De Tito, S.; Cerofolini, L.; Fragai, M.; Pagano, B.; Trotta, R.; Cosconati, S.; Marinelli, L.; Novellino, E.; Bertini, I.; Randazzo, A.; Luchinat, C.; Parrinello, M. The G-Triplex DNA. *Angew. Chem. Int. Ed Engl.* **2013**, *52* (8), 2269–2273.

(21) Limongelli, V.; Marinelli, L.; Cosconati, S.; La Motta, C.; Sartini, S.; Mugnaini, L.; Da Settimo, F.; Novellino, E.; Parrinello, M. Sampling Protein Motion and Solvent Effect during Ligand Binding. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (5), 1467–1472.

(22) Limongelli, V.; Bonomi, M.; Parrinello, M. Funnel Metadynamics as Accurate Binding Free-Energy Method. *Proc. Natl. Acad. Sci.* **2013**, *110* (16), 6358–6363.

(23) Comitani, F.; Limongelli, V.; Molteni, C. The Free Energy Landscape of GABA Binding to a Pentameric Ligand-Gated Ion Channel and Its Disruption by Mutations. *J. Chem. Theory Comput.* **2016**, *12* (7), 3398–3406.

(24) 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), E386-391.

(25) Troussicot, L.; Guillière, F.; Limongelli, V.; Walker, O.; Lancelin, J.-M. Funnel-Metadynamics and Solution NMR to Estimate Protein-Ligand Affinities. *J. Am. Chem. Soc.* **2015**, *137* (3), 1273–1281.

(26) Amori, L.; Katkevica, S.; Bruno, A.; Campanini, B.; Felici, P.; Mozzarelli, A.; Costantino, G. Design and Synthesis of Trans-2-Substituted-Cyclopropane-1-Carboxylic Acids as the First Non-Natural Small Molecule Inhibitors of O-Acetylserine Sulfhydrylase. *MedChemComm* **2012**, *3* (9), 1111.

(27) Annunziato, G.; Pieroni, M.; Benoni, R.; Campanini, B.; Pertinhez, T. A.; Pecchini, C.; Bruno, A.; Magalhães, J.; Bettati, S.; Franko, N.; Mozzarelli, A.; Costantino, G. Cyclopropane-1,2-Dicarboxylic Acids as New Tools for the Biophysical Investigation of *O* -Acetylserine Sulfhydrylases by Fluorimetric Methods and Saturation Transfer Difference (STD) NMR. *J. Enzyme Inhib. Med. Chem.* **2016**, *31* (sup4), 78–87.

(28) Bruno, A.; Amori, L.; Costantino, G. Computational Insights into the Mechanism of Inhibition of OASS-A by a Small Molecule Inhibitor. *Mol. Inform.* **2013**, *32* (5–6), 447–457.

(29) Benoni, R.; Pertinhez, T.A.; Spyrakis, F.; Davalli, S.; Pellegrino, S.; Paredi, G.; Pezzotti, A.; Bettati, S.; Campanini, B.; Mozzarelli, A. Structural Insight into the Interaction of O-Acetylserine Sulfhydrylase with Competitive, Peptidic Inhibitors by Saturation Transfer Difference-NMR. *FEBS Lett.* **2016**, *590*(7), 943-953.

(30) Turnbull, A. L.; Surette, M. G. Cysteine Biosynthesis, Oxidative Stress and Antibiotic Resistance in Salmonella Typhimurium. *Res. Microbiol.* **2010**, *161* (8), 643–650.

(31) Pieroni, M.; Annunziato, G.; Beato, C.; Wouters, R.; Benoni, R.; Campanini, B.; Pertinhez, T. A.; Bettati, S.; Mozzarelli, A.; Costantino, G. Rational Design, Synthesis, and Preliminary Structure–Activity Relationships of α-Substituted-2-Phenylcyclopropane Carboxylic Acids as Inhibitors of *Salmonella Typhimurium O* -Acetylserine Sulfhydrylase. *J. Med. Chem.* **2016**, *59* (6), 2567–2578.

(32) Segall, M. D.; Yusof, I.; Champness, E. J. Avoiding Missed Opportunities by Analyzing the Sensitivity of Our Decisions. *J. Med. Chem.* **2016**, *59* (9), 4267–4277.

(33) Lüdemann, S. K.; Lounnas, V.; Wade, R. C. How Do Substrates Enter and Products Exit the Buried Active Site of Cytochrome P450cam? 1. Random Expulsion Molecular Dynamics

Investigation of Ligand Access Channels and Mechanisms. *J. Mol. Biol.* **2000**, *303* (5), 797–811.

(34) Carlsson, P.; Burendahl, S.; Nilsson, L. Unbinding of Retinoic Acid from the Retinoic Acid Receptor by Random Expulsion Molecular Dynamics. *Biophys. J.* **2006**, *91* (9), 3151–3161.

(35) Colizzi, F.; Perozzo, R.; Scapozza, L.; Recanatini, M.; Cavalli, A. Single-Molecule Pulling Simulations Can Discern Active from Inactive Enzyme Inhibitors. *J. Am. Chem. Soc.* **2010**, *132* (21), 7361–7371.

(36) Martínez, L.; Sonoda, M. T.; Webb, P.; Baxter, J. D.; Skaf, M. S.; Polikarpov, I. Molecular Dynamics Simulations Reveal Multiple Pathways of Ligand Dissociation from Thyroid Hormone Receptors. *Biophys. J.* **2005**, *89* (3), 2011–2023.

(37) Martínez, L.; Webb, P.; Polikarpov, I.; Skaf, M. S. Molecular Dynamics Simulations of Ligand Dissociation from Thyroid Hormone Receptors: Evidence of the Likeliest Escape Pathway and Its Implications for the Design of Novel Ligands. *J. Med. Chem.* **2006**, *49* (1), 23–26.

(38) Winn, P. J.; Lüdemann, S. K.; Gauges, R.; Lounnas, V.; Wade, R. C. Comparison of the Dynamics of Substrate Access Channels in Three Cytochrome P450s Reveals Different Opening Mechanisms and a Novel Functional Role for a Buried Arginine. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (8), 5361–5366.

(39) Freire, E. A Thermodynamic Approach to the Affinity Optimization of Drug Candidates. *Chem. Biol. Drug Des.* **2009**, *74* (5), 468–472.

(40) Mayer, M.; Meyer, B. Group Epitope Mapping by Saturation Transfer Difference NMR to Identify Segments of a Ligand in Direct Contact with a Protein Receptor. *J. Am. Chem. Soc.* **2001**, *123* (25), 6108–6117.

(41) Jayalakshmi, V.; Biet, T.; Peters, T.; Krishna, N. R. Refinement of the Conformation of UDP−Galactose Bound to Galactosyltransferase Using the STD NMR Intensity-Restrained CORCEMA Optimization. *J. Am. Chem. Soc.* **2004**, *126* (28), 8610–8611.

(42) Jayalakshmi, V.; Krishna, N. R. Determination of the Conformation of Trimethoprim in the Binding Pocket of Bovine Dihydrofolate Reductase from a STD-NMR Intensity-Restrained

CORCEMA-ST Optimization. *J. Am. Chem. Soc.* **2005**, *127* (40), 14080–14084.

(43) Yuan, Y.; Bleile, D. W.; Wen, X.; Sanders, D. A. R.; Itoh, K.; Liu, H.; Pinto, B. M. Investigation of Binding of UDP-Gal *F* and UDP-[3-F]Gal *F* to UDP-Galactopyranose Mutase by STD-NMR Spectroscopy, Molecular Dynamics, and CORCEMA-ST Calculations †. *J. Am. Chem. Soc.* **2008**, *130* (10), 3157–3168.

(44) Bonomi, M.; Barducci, A.; Parrinello, M. Reconstructing the Equilibrium Boltzmann Distribution from Well-Tempered Metadynamics. *J. Comput. Chem.* **2009**, *30* (11), 1615–1621.

(45) Fold Type Ii Pyridoxal 5'-phosphate Dependent Enzymes: Structure, Substrate Recognition and Catalysis. In *Biomolecular Forms and Functions*; WORLD SCIENTIFIC / INDIAN INST OF SCIENCE, INDIA, 2012; pp 398–416.

(46) Burkhard, P.; Tai, C.-H.; Ristroph, C. M.; Cook, P. F.; Jansonius, J. N. Ligand Binding Induces a Large Conformational Change in O-Acetylserine Sulfhydrylase from Salmonella Typhimurium. *J. Mol. Biol.* **1999**, *291* (4), 941–953.

(47) Raboni, S.; Spyrakis, F.; Campanini, B.; Amadasi, A.; Bettati, S.; Peracchi, A.; Mozzarelli, A.; Contestabile, R. Pyridoxal 5′-Phosphate-Dependent Enzymes: Catalysis, Conformation, and Genomics. In *Comprehensive Natural Products II*; Elsevier, 2010; pp 273–350.

(48) Salsi, E.; Bayden, A. S.; Spyrakis, F.; Amadasi, A.; Campanini, B.; Bettati, S.; Dodatko, T.; Cozzini, P.; Kellogg, G. E.; Cook, P. F.; Roderick, S. L.; Mozzarelli, A. Design of *O* -Acetylserine Sulfhydrylase Inhibitors by Mimicking Nature. *J. Med. Chem.* **2010**, *53* (1), 345–356.

(49) Poyraz, Ö.; Jeankumar, V. U.; Saxena, S.; Schnell, R.; Haraldsson, M.; Yogeeswari, P.; Sriram, D.; Schneider, G. Structure-Guided Design of Novel Thiazolidine Inhibitors of *O* -Acetyl Serine Sulfhydrylase from Mycobacterium Tuberculosis. *J. Med. Chem.* **2013**, *56* (16), 6457–6466.

(50) ChemAxon – Software for Chemistry and Biology (https://www.chemaxon.com)

(51) Pieroni, M.; Annunziato, G.; Azzali, E.; Dessanti, P.; Mercurio, C.; Meroni, G.; Trifiró, P.; Vianello, P.; Villa, M.; Beato, C.; Varasi, M.; Costantino, G. Further Insights into the SAR of α-Substituted Cyclopropylamine Derivatives as Inhibitors of Histone Demethylase KDM1A. *Eur. J.*

Med. Chem. **2015**, *92*, 377–386.

(52) Yu, C.; Niu, X.; Jin, F.; Liu, Z.; Jin, C.; Lai, L. Structure-based Inhibitor Design for the Intrinsically Disordered Protein c-Myc. *Sci Rep.* **2016**, *6*, 22298

(53) Schrödinger, LLC: New York, NY, 2011 (https://www.schrodinger.com)

(54) Jorgensen, W. L.; Schyman, P. Treatment of Halogen Bonding in the OPLS-AA Force Field; Application to Potent Anti-HIV Agents. *J. Chem. Theory Comput.* **2012**, *8* (10), 3895–3801.

(55) Amber Home Page (http://ambermd.org)

(56) NAMD - Scalable Molecular Dynamics (http://www.ks.uiuc.edu/Research/namd/)

(57) VMD - Visual Molecular Dynamics (http://www.ks.uiuc.edu/Research/vmd/)

(58) PLUMED (http://www.plumed.org)

(59) CatDCD - Concatenate DCD files (http://www.ks.uiuc.edu/Development/MDTools/catdcd/)

(60) Tian, H.; Guan, R.; Salsi, E.; Campanini, B.; Bettati, S.; Kumar, V.P.; Karsten, W.E.; Mozzarelli, A.; Cook, P.F. Identification of the Structural Determinants for the Stability of Substrate and Aminoacrylate External Schiff Bases in O-acetylserine Sulfhydrylase-A. *Biochemistry* **2010**, *49* (29), 6093-6103.

(61) Spyrakis, F.; Felici, P.; Bayden, A.S.; Salsi, E.; Miggiano, R.; Kellogg, G.E.; Cozzini, P.; Cook, P.F.; Mozzarelli, A.; Campanini, B. Fine Tuning of the Active site Modulates Specificity in the Interaction of O-acetylserine Sulfhydrylase Isozymes with Serine Acetyltransferase. *Biochim Biophys Acta.* **2013**, *1834* (1), 169-181.

Figure Captions

Figure 1. A) Most probable unbinding route identified by RAMD and used as reference pathway to collocate the funnel-restrained potential **B)**. **C)** CVs used to perform the FM experiments.

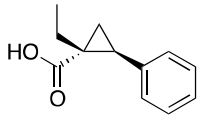
Figure 2. A) Evolution of the CV_1 along 1.5 μs of FM simulations. As shown in the plot, several recrossing events between the ligand bound and unbound states were observed, leading to the convergence in the estimation of the *St*OASS-A-**12b** binding free energy. **B)** The binding FES calculated using CV_1 and CV_2. Isosurfaces are shown 1.5 kcal/mol. The full binding minima is highlighted in the inset with a white star. The full unbinding state was considered in the region where the ligand did not feel neither the protein or the FM upper wall. **C)**. ΔG_b^0 was calculated as the free-energy difference of the system with the ligand in the bound ($CV_1 < 14$ Å) states and unbound ($33 < CV_1 > 36$ Å) states. **D)** Free energy profile projected as function of the position *z* along the funnel. The red arrow highlights the correction due to the cylindrical part of the funnel potential.

Figure 3. A) Binding FES calculated using the reweighting algorithm as a function of the funnel coordinates (projection on the *z* axis and distance from the axis), isosurfaces are shown 1.5 kcal/mol. The minima characterizing the full binding state was used to collect all the full binding conformations which were rescored by means of STDc. **B)** Experimental G.E.M. profile (blue bars) *vs.* that calculated form the single conformation with the lowest RMSE (orange bars), the ensemble (cyan bars), and that derived from the docking studies (yellow bars), previously reported.[25] **C)** Binding conformation of **12b** (orange sticks) into the *St*OASS-A active site (white cartoon and sticks, the PLP is depicted in cyan sticks), the structure used is that with the lowest RMSE. **D)** R.M.S.D of **12b** calculated during classical MD of the complex derived from STDc with the lowest RMSE. **E)** Comparison of the X-ray crystal structure of *St*OASS-A (pdb code: 1OAS) and the full binding conformation obtained in the FM experiments. Highlighted into the inset the catalytic loop which adopts two disparate conformations.

Figure 4. A) Surface projection of the ensemble extracted with STDc, **12b** is depicted as orange sticks. **B)** Superposition of the *St*OASS-A-**12b** complex (white cartoon and orange sticks), with the structure of the *Hi*OASS-A (pdb code: 3IQH, yellow sticks and cartoon) and *Mtb*OASS (pdb code: 3ZEI, blue sticks and cartoon). **C)** *St*OASS-A represented as white surface and in sticks the compounds co-crystallized in 3IQH and 3ZEI. **D)** Surface projection of STDc and in sticks **12b** the compounds co-crystallized in 3IQH and 3ZEI. **E)** Synthetic strategy used to generate **12b** derivatives substituted at postion 3'.

Charts.

Chart 1. Structure of compound **12b** ((1R,2S)-1-ethyl-2-phenylcyclopropanecarboxylic acid).



12b

Figures.

Figure 1.

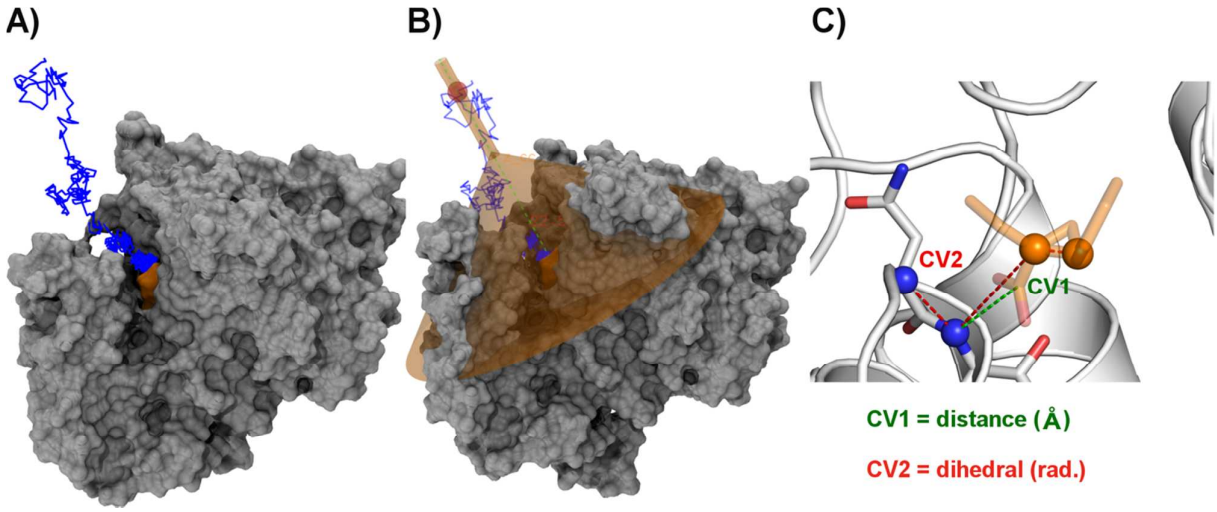


Figure 2.

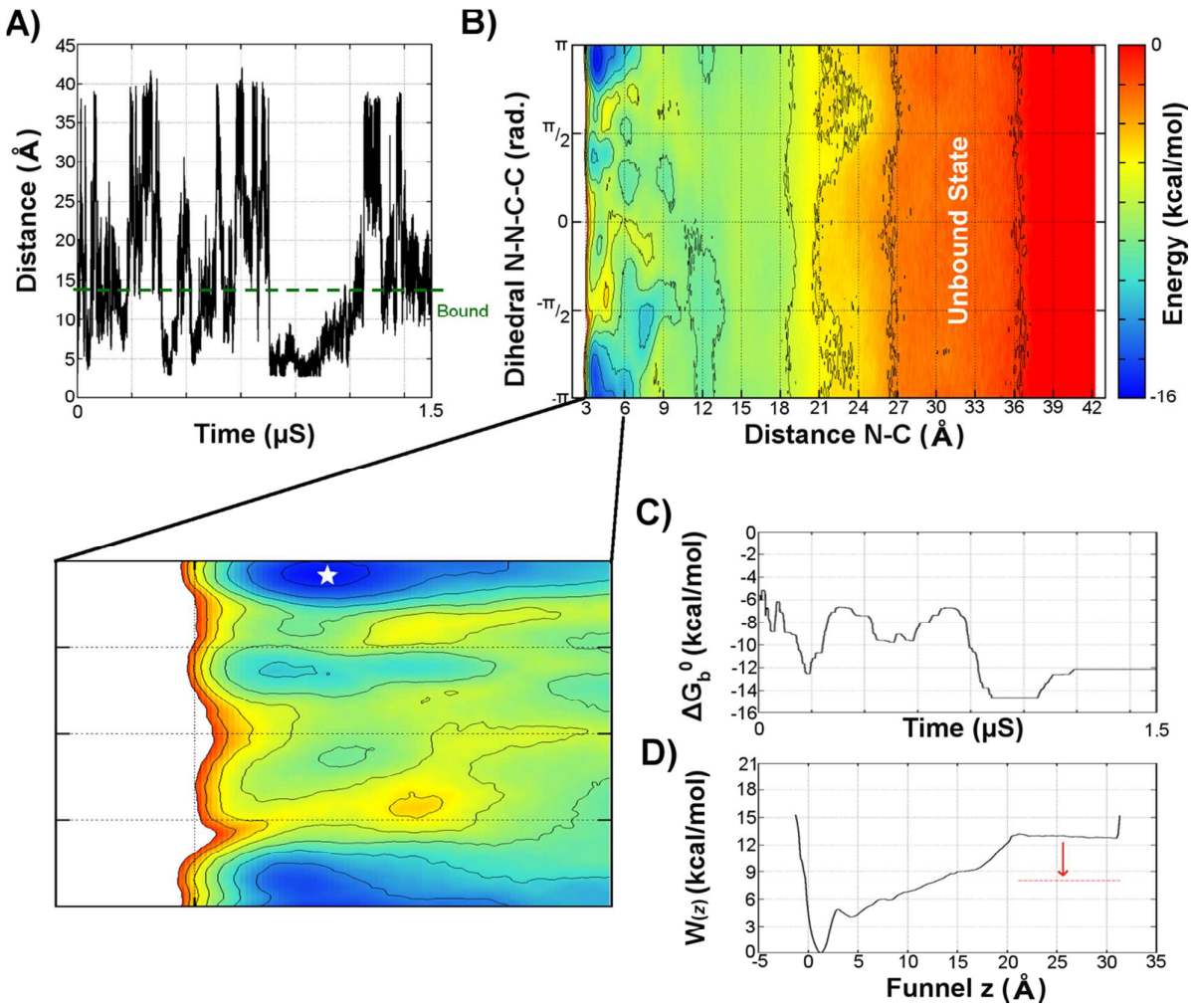

A)
Distance (Å)
Time (μS)
Bound
B)
Dihedral N-N-C-C (rad.)
Distance N-C (Å)
Unbound State
Energy (kcal/mol)
C)
ΔG_b^0 (kcal/mol)
Time (μS)
D)
W(z) (kcal/mol)
Funnel z (Å)

Figure 3.

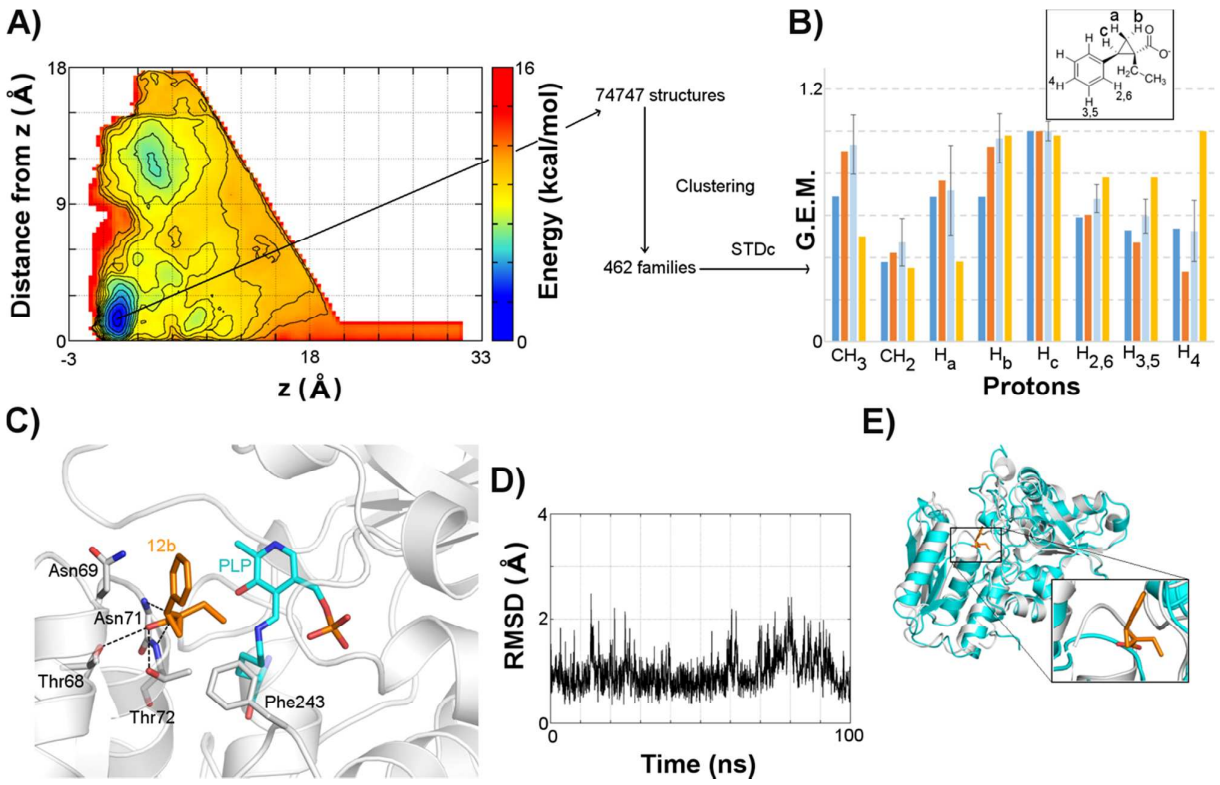
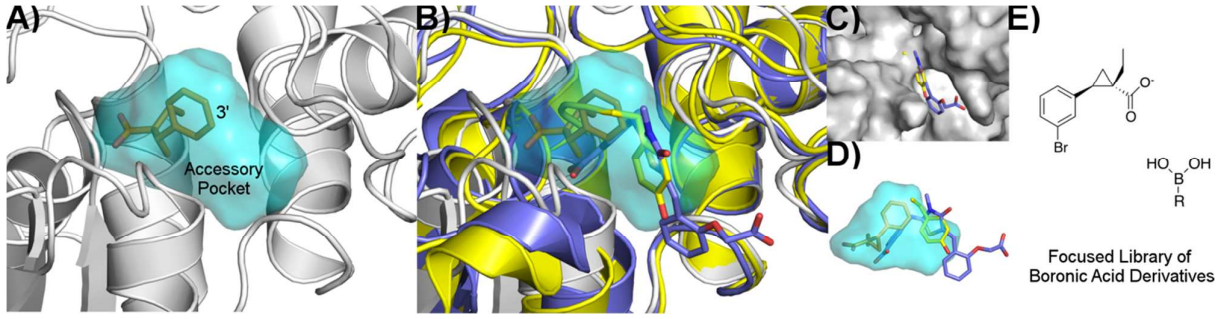
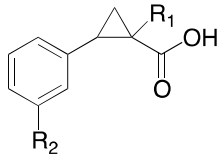


Figure 4.



Tables.

Table 1. Dissociation constants of the title compounds for *St*OASS-A.

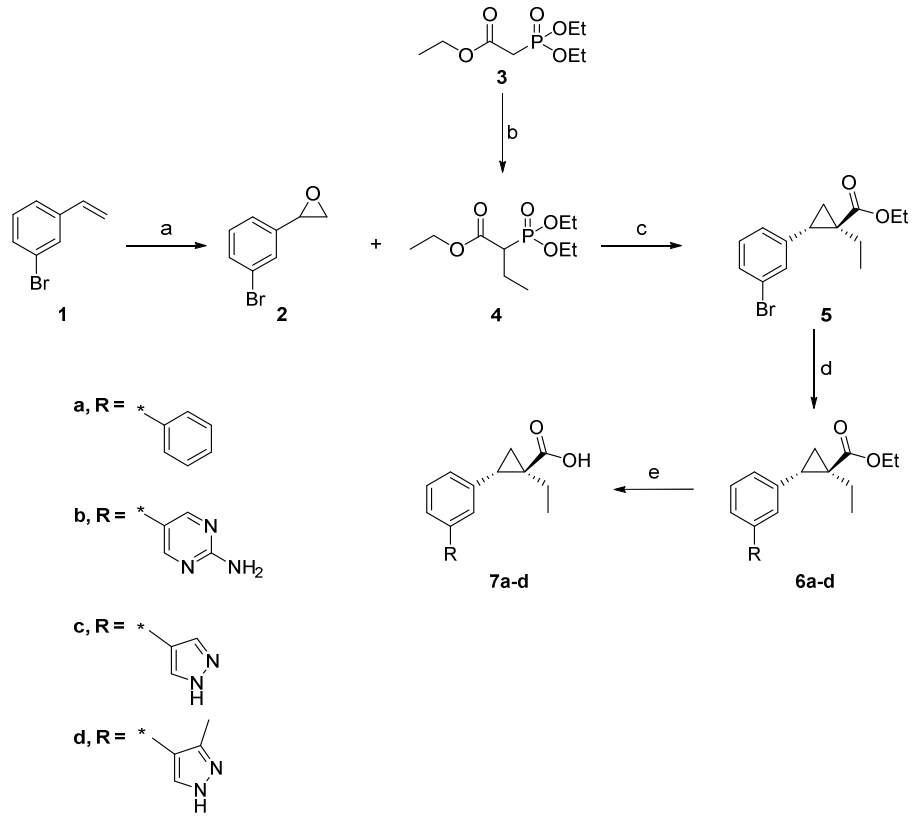


Cmpd	Stereochemistry	R_1	R_2	K_d (μM)
				***St*OASS-A**
12a*	Racemic	Et	H	15.1±1
12b*	1R,2S	Et	H	12.1±0.5
12c*	1S,2R	Et	H	1200±300
7a	Racemic	Et	phenyl	13.7±0.54
7b	Racemic	Et	2-amino-(5)-pyrimidyl	38.3±5.1
7c	Racemic	Et	1H-(4)-pyrazyl	18.9±1.0
7d	Racemic	Et	1H-3-methyl-(4)-pyrazyl	16.7±0.9
10a	(-)	Et	phenyl	9.2±1.1
11a	(+)	Et	phenyl	10.2±0.9

* Reported in ref. 25

Schemes.

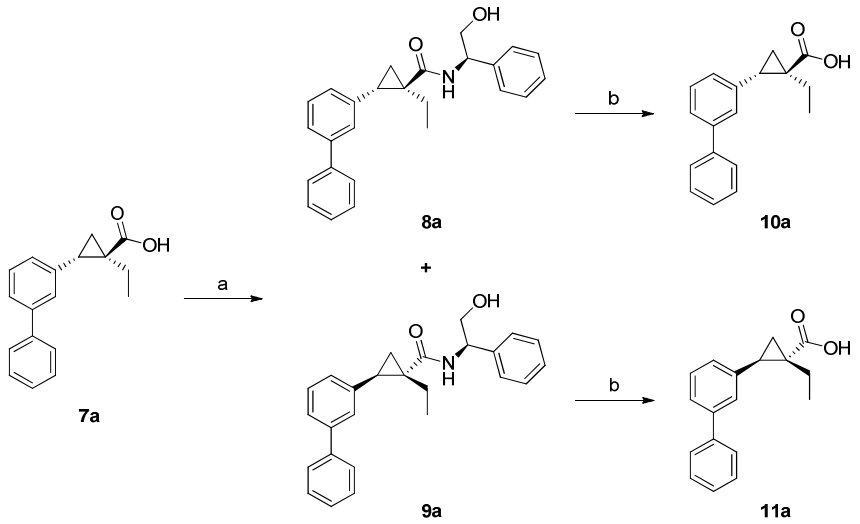
Scheme 1[a]



[a]*Reagents and conditions*. *a) m*-CPBA, $CHCl_3$, 0°C 3h, r.t. 18h, 98%; *b)* ethyl bromide, NaH, DME dry, 0°C -> r.t 2h, 60°C 3h, 80%; *c)* n-buthyllithium, DME dry, r.t. -> 90°C, 18h, 71%; *d)* tetrakis(triphenylphosphine)palladium, phenylboronic acid, K_2CO_3 2M, toluene/MeOH/H_2O, 18h, 110°C, 81-96%; *e)* LiOH, THF/MeOH/H_2O, MW 10 min, 100°C, 21-83%.

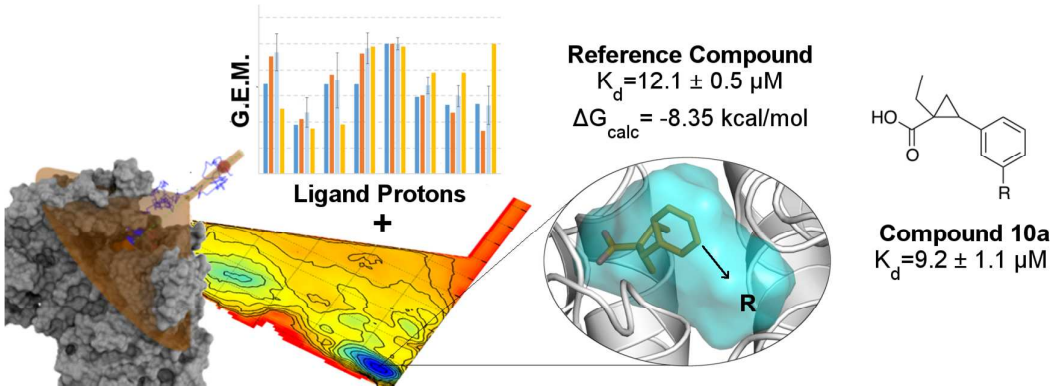
Scheme 2[a]

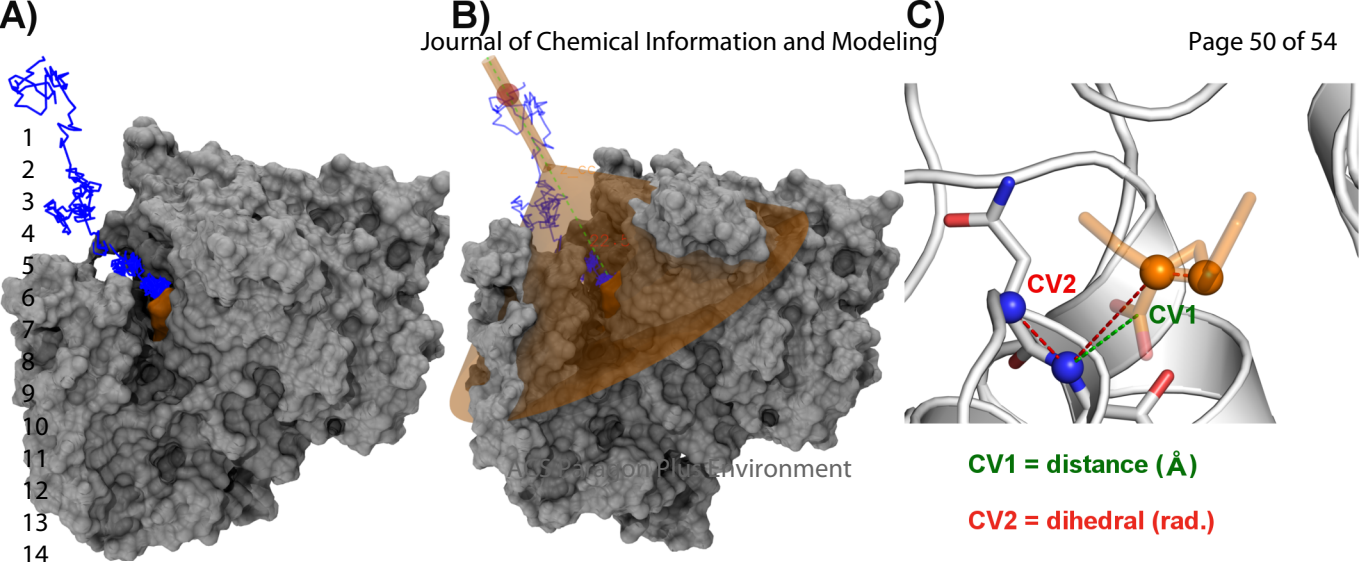
Scheme 2[a]



[a]*Reagents and conditions*. *a)* (R)-(-)-2-Phenylglycinol, TBTU, EDC*HCl, TEA, DCM, 0°C 1h,r.t. 6h, 60%; *b)* H_2SO_4 3N, dioxane, r.t., 18h, 42-56%.

Table of Contents Graphic


G.E.M.
Ligand Protons
+
Reference Compound
K_d=12.1 ± 0.5 μM
ΔG_{calc} = -8.35 kcal/mol
R
HO
O
R
Compound 10a
K_d=9.2 ± 1.1 μM


A)
B)
C)
CV2
CV1
CV1 = distance (Å)
CV2 = dihedral (rad.)