A Collective Variable for the Rapid Exploration of Protein Druggability

Citation for published version:
Cuchillo, R, Pinto-Gil, K & Michel, J 2015, 'A Collective Variable for the Rapid Exploration of Protein Druggability' Journal of Chemical Theory and Computation, vol. 11, no. 3, pp. 1292-1307. DOI: 10.1021/ct501072t

Digital Object Identifier (DOI):
10.1021/ct501072t

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Early version, also known as pre-print

Published In:
Journal of Chemical Theory and Computation

A Collective Variable for the Rapid Exploration of Protein Druggability

Rémi Cuchillo, Kevin Pinto-Gil, and Julien Michel*

EaStCHEM School of Chemistry, Joseph Black Building, West Mains Road, Edinburgh EH9 3JJ, United Kingdom

Received November XX, 2014

Abstract: An efficient molecular simulation methodology has been developed for the evaluation of the druggability (ligandability) of a protein. Previously proposed techniques were designed to assess the druggability of crystallographic structures and cannot be tightly coupled to molecular dynamics (MD) simulations. By contrast the present approach, JEDI (‘’Just Exploring Druggability at protein Interfaces’’), features a druggability potential made of a combination of empirical descriptors that can be collected "on-the-fly" during MD simulations. Extensive validation studies indicate that JEDI analyses discriminate druggable and nondruggable protein binding site conformations with accuracy similar to alternative methodologies, and at a fraction of the computational cost. Since the JEDI function is continuous and differentiable, the druggability potential can be used as collective variable to rapidly detect cryptic druggable binding sites in proteins with a variety of MD free energy methods. Protocols for applications to flexible docking problems are outlined.

* Corresponding author e-mail: mail@julienmichel.net

Keywords: druggability, ligandability, protein-ligand interactions, molecular dynamics simulation, free energy calculation

Introduction

The development of a new medicine is a long and expensive process subjected to high attrition rates.[1] Over the last decades, around 60% of drug discovery projects failed to identify viable leads able to modulate adequately the activity of a protein target.[2] Analyses of the sequenced human genome indicate that less than 50% of disease-involved genes code for druggable proteins.[3,4] A protein target found to be nondruggable late in the drug discovery process is a significant waste of time and expense in the pharmaceutical industry. Accordingly, an early assessment of druggability offers the opportunity to focus efforts on tractable targets, thereby reducing the rate of failure.[5] The concept of druggability is ambiguous because it has been used in many different fields to describe, in a different context, the properties of genes, proteins and ligands. In the context of structure-based drug design, protein druggability is often related to the ability of a therapeutic target to bind a drug-like small molecule, leaving aside many important facets of the drug discovery and development process such as selectivity, toxicology or pharmacokinetics.[3] Since druggability is closely linked to the notion of binding site in this specific context, the terms "bindability" or "ligandability" have also been proposed as alternatives.[6,7]

This report focuses on the use of computation for structure-based evaluation of protein druggability. The idea of relating binding site energetics to structural descriptors was explored as early as in 1985 with the Grid program of Goodford, and other related methods.[8-11] As interest in druggability developed in the last fifteen years, more recent efforts have focused on correlating directly structural descriptors to druggability. An early effort was contributed by Hadjuk and coworkers.[12] NMR-based fragment screening was used to develop a mathematical model for druggability measurements whereby structural descriptors were correlated to NMR hit-rates. The methodology is based on the assumption that a druggable cavity tends to bind more fragments than a nondruggable pocket. A second approach, called $\mathrm{MAP_{POD}}$, was published by Cheng et al. shortly after.[13] The authors proposed a scoring function to assess the maximal affinity between a small molecule and a binding site based on physicochemical and geometric features. This study also introduced a new category of proteins that are neither ‘druggable’ nor ‘nondruggable’, but are instead “difficult” to target with small molecules. The suggestion was that this category of proteins should be targeted with highly polar molecules administrated as pro-drugs. These early contributions have paved the way for a similar class of computational methods that aim to detect and evaluate potential binding sites at protein surfaces. For instance, the public dataset compiled for MAP_{POD} was used to parameterize Dscore, a druggability function coupled with the pocket detector SiteMap.[14,15] Dscore is a simple linear combination of three descriptors reflecting the volume, enclosure and hydrophobicity of the binding site. Schmidtke et al. have recently developed a fast methodology based on a new publically accessible dataset.[16] The approach features a logistic regression analysis to extract local and global hydrophobic descriptors of a protein pocket. Another recent structure-based approach published in the field is the Drugpred method.[17] The Drugpred scoring function is based on a large freely accessible non-redundant protein dataset and was derived using partial least-squares projection. Drugpred appears to be less sensitive to small binding site structural modifications that do not dramatically affect pocket properties.[17]

The above described methods were designed to assess the druggability of a crystallographic protein structure. However, it is well known that sometimes a few local structural rearrangements around a protein binding site can profoundly influence the binding affinity of a small molecule to its target.[18,19] Accordingly, a second class of druggability prediction algorithms based on molecular dynamics (MD) simulations have been proposed.[20-22] One of the first method based on classical molecular dynamics simulations was published by Seco et al.[20] In this grid-based approach, an explicit restrained MD simulation of a protein is performed in the presence of a given concentration of isopropyl alcohol. The binding propensities of the probe at the protein surface are then back-computed to evaluate spatially resolved binding free energies. A similar protocol was recently applied on different systems using several diverse probes.[22] The authors showed that probe molecules could induce both local and global structural rearrangements of the protein, leading to increases in target druggability. Nevertheless a frequent concern with these techniques is that the observed conformational changes reflect denaturation of the protein due to high probe concentrations. Thus judicious use of positional restraints is required to limit the occurrence of undesirable conformational changes. Also, probe diffusion necessary to compute binding propensities in buried cavities can be very slow with standard MD approaches.

To overcome the limitations of current MD based druggability prediction methods, this report introduces the JEDI algorithm (''Just Exploring Druggability at protein Interfaces''). JEDI has been designed to evaluate protein druggability "on-the-fly" during MD simulations without using organic probes or protein restraints. The druggability function relies on a set of geometric parameters describing the volume, the enclosure and the hydrophobicity of a binding site. The JEDI scoring function is fast, continuous and differentiable. Accordingly, the JEDI druggability descriptor can be used to construct artificial druggability potentials that are designed to bias sampling of protein binding site conformations similar to a training set with the aid of a druggability force applied during a MD simulation. JEDI has been implemented in the software PLUMED 1.3 to enable biased molecular dynamics simulations with a variety of free-energy calculations techniques and diverse popular MD engines.[23] The methodology was parameterized using the freely accessible Druggable Cavity Directory (DCD) dataset.[16] The sensitivity of the method to binding site conformational changes was tested with a compiled dataset of cryptic binding sites. Detailed druggability assessments have also been performed on the fly during unbiased and biased MD simulations of a test protein, VHL. This demonstrated the potential for JEDI analyses to detect cryptic druggable binding sites in proteins and to deliver conformations suitable as input for subsequent docking calculations.

Methods

Datasets. Protein structures were taken from the Non Redundant Druggability Dataset (NRDD) in the Druggable Cavity Directory compiled by Schmidtke et al.[16] A set of 63 unique proteins has been used to parameterize the JEDI scoring function. Each protein has been assigned by the authors of the original study an experimental druggability value from 1 to 10 (from less druggable to more druggable) according to its capability to bind a drug-like compound. The dataset can be further divided into three categories: non-druggable (*DCD_{score}* 1 to 4), difficult (*DCD_{score}* 5 to 7) and druggable (*DCD_{score}* 8 to 10). In order to benchmark JEDI against an existing methodology, druggability calculations were performed on the energy-minimized structures of the training dataset using the program fpocket.[16,24] A detailed list of the dataset is given in the supporting information, including druggability scores obtained with both approaches (SI Table S1). A validation dataset, called the hidden pocket dataset, has also been compiled. Each protein in this dataset is represented by two different structures that exhibit conformational variability in the binding site region that correlates with variations in the binding affinities of known ligands.

Protocol overview. JEDI is a grid-based approach. The methodology includes three major steps (Figure 1A). First, a region of interest where the druggability evaluation will be conducted must be defined. This area can be located anywhere in the protein structure in principle, but in this report efforts are focused on evaluating the druggability of regions known to contain a binding site. Thus spatial regions to analyze were defined from the position of known ligands. A large 3D cubic grid with 1.5 Å spacing between grid points is initially positioned around the region of interest. Next, only grid points within 6 Å of one ligand atom were retained. All protein heavy atoms within 3 Å of a grid point are then selected for druggability calculations and this set of atoms is referred as the 'binding site region'. This setup is then followed by either a single point calculation or MD simulations with druggability evaluated at regular intervals in unbiased simulations, or at each time-step for MD simulations biased with the JEDI potential. Every druggability assessment requires that the 'activity' of all grid points is evaluated, with grid points classified as inactive, partially active or fully active according to their geometric position in the binding site. Then, volume and hydrophobicity descriptors that depend on grid point activities and local geometric arrangements of protein atoms are computed in order produce a conformation-dependent protein druggability score.

To avoid errors in the druggability predictions due to diffusion of the protein over the course of an MD simulation, the Cartesian coordinates of the grid points are re-evaluated prior to each druggability assessment. Firstly, the distance vector between the center of mass of the protein atoms in the binding site region in the conformation at the *n*-th step of the MD simulation ($\boldsymbol{r}_{com,t=n}$) and the initial protein conformation ($\boldsymbol{r}_{com,t=0}$) is evaluated. Then, the rotation matrix that best fits the protein backbone atoms of the entire protein onto their coordinates at $t = 0$ is computed using the Kabsch algorithm.[25] Finally, the resulting translation vector and rotation matrix are used to transform the grid point Cartesian coordinates at $t = 0$ into grid point Cartesian coordinates at $t = n$.

Scoring Function. The JEDI druggability score is calculated as a linear combination of two partial-least squared derived descriptors reflecting the volume, and the hydrophobicity (eq 1).

$$JEDI_{score} = V_{druglike}(\alpha V_a + \beta H_a + \gamma) \tag{1}$$

where $V_{druglike}$, V_a and H_a represent respectively the drug-like volume descriptor, the pocket volume descriptor and the pocket hydrophobicity. α, β and γ are constants of the model derived by multiple linear regressions against a training set. All the descriptors presented below are based on spline functions such that the JEDI potential is continuous and at least twice differentiable. Two forms of spline functions have been used operating on variables v and k (Figure 1C). The first one turns "off" with v starting at k at v_{min}, reaching 0 at v_{min}+Δ (eq 2).

$$S_v^{off}(k, v_{min}, \Delta) = \begin{cases} k \; if \; m < 0 \\ k[(1-m^2)^2(1+2m^2)] \; if \; 0 \leq m \leq 1 \\ 0 \; if \; m > 1 \end{cases} \tag{2}$$

where $m = \frac{v - v_{min}}{\Delta}$. The second form turns "on" from 0 to k along an interval Δ (eq 3).

$$S_v^{on}(k, v_{min}, \Delta) = \begin{cases} 0 \; if \; m < 0 \\ k[1-(1-m^2)^2(1+2m^2)] \; if \; 0 \leq m \leq 1 \\ k \; if \; m > 1 \end{cases} \tag{3}$$

The active volume descriptor V of the binding site is given by equation 4:

$$V = \sum_{i=1}^{N} a_i \, V_g \tag{4}$$

where N is total number of grid points, V_g is the volume of space covered by a grid point. To capture the shape of the pocket, each grid point is assigned an activity score a_i between 0 and 1 (inactive to active), according to its geometric position inside the binding pocket (eq 5).

$$a_i = S_{BS_i}^{off}(1.0, BS_i, \Delta BS) S_{mind_i}^{on}(1.0, CC_{mind}, \Delta CC) S_{exposure_i}^{on}(1.0, E_{min}, \Delta E) \tag{5}$$

The first term of eq 5 gradually turns off grid points according their distances from the region of interest. This term is optional, but is useful to ensure that fluctuations in druggability scores are not unduly influenced by conformational changes that are remote from the protein region of interest. The minimum distance BS_i between a grid point i and the M atomic coordinates defining the binding site region is calculated as:

$$BS_i = \frac{\theta}{\ln\left(\sum_{j=1}^{M} \exp\left(\frac{\theta}{\|\boldsymbol{r}_{ij}\|}\right)\right)} \tag{6}$$

With θ=50.0 Å and $\boldsymbol{r}_{ij} = \boldsymbol{r}_{gi} - \boldsymbol{r}_{pj}$, where $\boldsymbol{r}_{gi}$ and $\boldsymbol{r}_{pj}$ are respectively the position vectors of grid point i and protein atom j belonging to the binding site region. The second term in equation 5 causes grid points that overlap with protein atoms to be gradually inactivated (Figure 1B). The minimum distance $mind_i$ between grid points and protein atoms is calculated with an equation similar to eq 6. The third term in equation 5 gradually inactivates solvent exposed grid points (Figure 1B).

$$exposure_i = \sum_{k=1}^{N} \left[S_{mind_k}^{off}(1.0, CC2_{min}, \Delta CC2) S_{\|\boldsymbol{r}_{ik}\|}^{on}(1.0, GP1_{min}, \Delta GP1) S_{\|\boldsymbol{r}_{ik}\|}^{off}(1.0, GP2_{min}, \Delta GP2) \right] \tag{7}$$

where $CC2_{min}$ /$\Delta CC2$ control the distance below which a grid point is considered as interacting with the protein. $GP1_{min}$ /$\Delta GP1$ and $GP2_{min}$ /$\Delta GP2$ are used to select grid points at a given distance interval from the grid point i in order to penalize solvent exposed grid points. With the default values presented in Table 1, a maximum of 44 grid points can be selected around a given grid point i and the maximum value of $exposure_i$ is 23.97 with the present parameterization. Thus to summarize a grid point achieves a high activity value (maximum 1) if it is neither too close nor too far from the protein atoms that form the binding site region of interest.

The active volume V is then converted in a pocket volume descriptor V_a using equation 8.

$$V_a = \frac{V}{V_{max}} \tag{8}$$

where V_{max} is the maximum active volume descriptor. This constant was set to be equal to the maximum active volume V calculated for protein binding sites in the ''druggable'' category of the DCD dataset. Accordingly, a cavity presenting the characteristics of a typical small-molecule binding site will have a typical V_a value in the interval [0.0,1.0]. In order to penalize overly large or overly small cavities that are not suitable for drug-like small molecules, the descriptor $V_{druglike}$ is also computed with eq 9.

$$V_{druglike} = S_V^{off}(1.0, V_{max}, \Delta V_{max}) S_V^{on}(1.0, V_{min}, \Delta V_{min}) \tag{9}$$

where V_{min} is equal to 0 Å^3 by default. Analysis of pockets from the DCD dataset suggested a ΔV_{min} value of 36 Å^3. For simplicity, the same value was used for ΔV_{max} . The effect is that cavities that differ substantially in active volume from those present in the training set will have a low value of $V_{druglike}$. In turn this will assign a low $JEDI_{score}$ to cavities that differ markedly from the training set. This parameter may be easily tuned if cavities for ligands that differ substantially from those present in the DCD training set are desired.

The active grid hydrophobicity function captures the average hydrophobicity of the active grid points and is given by eq 10:

$$H_a = \frac{1}{V} \sum_{i=1}^{N} (H_i a_i) \tag{10}$$

where the hydrophobicity score H_i of the grid point i is calculated as

$$H_i = \frac{apolar_i}{apolar_i + polar_i} \tag{11}$$

where $apolar_i$ and $polar_i$ are function of the number of apolar (C and S) and polar (O and N) protein atoms within the distance r_{hydro} defined by equations 12a and 12b:

$$apolar_i = \sum_{j=1}^{M_{apolar}} S_{\|r_{ij}\|}^{off}(a_i, r_{hydro}, \Delta r_{hydro}) \tag{12a}$$

$$polar_i = \sum_{j=1}^{M_{polar}} S_{\|r_{ij}\|}^{off}(a_i, r_{hydro}, \Delta r_{hydro}) \tag{12b}$$

where M_{apolar} and M_{polar} are the total number of apolar and polar protein atoms in the binding site region.

JEDI derivatives. Because the JEDI potential is based on functions that are continuous and differentiable, the gradient with respect to the Cartesian coordinates x_{p_j}, y_{p_j}, z_{p_j} of the j protein atoms in the binding site region can be calculated using the following equation:

$$\nabla JEDI_{score} = \sum_{j=1}^{M} \left(\frac{\partial JEDI_{score}}{\partial x_{p_j}} + \frac{\partial JEDI_{score}}{\partial y_{p_j}} + \frac{\partial JEDI_{score}}{\partial z_{p_j}} \right) \tag{13}$$

where *M* is the number of protein atoms in the binding site region, $\frac{\partial JEDI_{score}}{\partial x_{p_j}}$, $\frac{\partial JEDI_{score}}{\partial y_{p_j}}$ and $\frac{\partial JEDI_{score}}{\partial z_{p_j}}$ are the partial derivatives with respect to the Cartesian coordinates of protein atom j. The derivatives of the JEDI potential with respect to the grid Cartesian coordinates do not need to be calculated as the grid is frozen during MD time-steps. By application of the product rule:

$$\frac{\partial JEDI_{score}}{\partial x_{p_j}} = JEDI_{score} \left[\frac{1}{V_{druglike}} \frac{\partial V_{druglike}}{\partial x_{p_j}} + \frac{1}{\alpha V_a + \beta H_a + \gamma} \left(\alpha \frac{\partial V_a}{\partial x_{p_j}} + \beta \frac{\partial H_a}{\partial x_{p_j}} \right) \right] \tag{14}$$

Similar expressions can be derived for the partial derivatives with respect to y_{p_j} and z_{p_j}. A detailed derivation of all partial derivatives in equation 14 is given in the supplementary inforamation.

JEDI optimization. The parameters of the JEDI model were optimized using the python module PyEvolve.[26] After investigation, only the CC_{mind}, ΔE, $\Delta CC2$ and r_{hydro} variables were selected for optimization using a range of physically plausible values (SI Table S2). An elitist genetic algorithm was then iterated for 50 generations on a population of 40 individuals. The fitness function was defined to maximize the r^2 of $JEDI_{score}$ vs DCD_{score} values after a Partial Least Squares regression. Uncertainties in the $JEDI_{score}$ parameters were determined with 100 iterations of bootstrapping using a split of 0.7/0.3 for the training and validation sets.

MD simulations. Proteins, ligands and cofactors were prepared using the python script Protein Preparation Wizard developed by Schrödinger and available in Maestro.[27] First, missing hydrogen atoms were added to the structure to assign the appropriate bond number and formal charge. Then, proteins were manually verified to avoid incomplete side chains and steric clashes. Molecular dynamics simulations have been performed using GROMACS 4.5.5 combined with PLUMED 1.3.[23,28]

Simulations were carried out in implicit solvent using the Generalized Born model and the Onufriev-Bashford-Case method to calculate the Born radii with a cutoff of 20 Å.[29,30] An energy minimization was performed using the steepest descent algorithm to reach the convergence parameter of 300 kJ.mol^{-1}.nm^{-1} of maximum force change. Then, production runs of 50 ns were performed using a time step of 2.0 fs. Systems were maintained at a constant temperature of 310 K using a stochastic Berendsen thermostat with a coupling constant of 1.0 ps.[31] The force field Amber99sb-ILDN was used for the proteins and the GAFF force field has been used for ligands and cofactors.[32,33] The GAFF parameters for the ligands and the cofactors were obtained by using the software acpype, in combination with the antechamber utility from the AMBER12 software package.[34,35] A new atom type was created to represent grid points. To avoid interactions between the protein atoms and the grid points, the Lennard-Jones parameters σ and ε and the atomic partial charges of grid points were equal to zero. All grid points are frozen in space during energy minimization and molecular dynamics time-steps.

Umbrella sampling simulations. Several umbrella sampling calculations were performed using the following biasing potential:

$$U_{JEDI}\big(s(\boldsymbol{r})\big) = \kappa\big(s(\boldsymbol{r}) - s_0\big)^2 \tag{15}$$

where $s(\boldsymbol{r})$ is the $JEDI_{score}$ of protein binding site conformation $\boldsymbol{r}$, κ is the force constant of the biasing potential, and s_0 is a target value for $JEDI_{score}$.[36] Several biased MD simulations were performed by varying κ and s_0 for different systems. The resulting trajectories were clustered to identify the most likely conformations associated with a given set of (κ, s_0) values. The single linkage clustering approach as implemented in GROMACS was used to identify the most representative conformations of each resulting trajectory. The RMSD cluster cutoff was set to 1 Å. RMSD calculations were performed using the coordinates of heavy atoms constituting the binding site region, excluding atoms that can form symmetry equivalent conformations (e.g. Valine C_γ atoms). Finally, cluster homogeneity was manually checked.

Docking calculations. Several representative protein structures were extracted from the trajectories to perform docking calculations. The Maestro software was used to prepare input files for both receptors and ligands.[27] Protonation states of binding site Histidine residues were chosen to be consistent with those from the MD simulations (in particular, His110 and His115 were protonated on the ε-nitrogen atom). Docking calculations were performed with the software Autodock Vina and the Autodock/Vina plugin for pymol.[37,38] For each complex, the same docking grid was used, and up to twenty poses were generated. Different protocols featuring a fully rigid receptor or allowing side-chain flexibility of selected residues were used.

Results and Discussion

Choice of descriptors.

The druggability score of the JEDI methodology is based on a linear combination of structural descriptors characterizing the volume and the hydrophobicity of a cavity. The choice of those collective variables were influenced by the literature.[6,12,13,16,17,39] A rule-based method published by Perola et al. suggested five suitable descriptors: volume, depth, enclosure, percentage of charged residues and hydrophobicity. These descriptors summarize a general consensus fairly well.[40] After investigation, only two descriptors have been retained: the active volume and the hydrophobicity. An early version of JEDI was also including a descriptor capturing the degree of buriedness of the binding site. The buriedness, as described by Volkamer et al., was captured as the ratio between the number of hull grid points in contact with the protein surface and the total number of hull grid points.[39] Here hull grid points are defined as in Volkamer et al. and correspond to the outer layer of active grid points that define the shape of a binding site. After preliminary investigations, this descriptor was not found to contribute significantly to the druggability prediction. This is likely because the current definition of the active volume descriptor is penalizing solvent-exposed grid points and thus already accounts for buriedness. Consequently, shallow solvent exposed cavities have a lower active volume descriptor than buried enclosed closed cavities.

The results depicted in Figure 2 demonstrate that higher $JEDI_{score}$ values do correlate with a larger binding site active volume V and a larger hydrophobicity descriptor H_a.

Since the publication of the first large scale classification of protein binding sites by An et al,[41] numerous studies have been conducted in the field of pocket detection and analysis to improve understanding of the physicochemical properties that underlie protein-ligand interactions.[6,16,17,39] The average volume of a druggable binding site was evaluated around 600 Å^3,[41] with maximum values around 900-1200 Å^3.[39,40] These estimates are in line with those computed with JEDI; the average volume of a binding site represented by the total number of active and partially active ($a_i > 0$) grid points was found to be 496 ± 202 Å^3 , with a maximum value of 1019 Å^3. The results shown in Figure 2A depict the distribution of active volume (V) values for different categories of protein binding sites. As the active volume is the sum of the grid point volumes weighted by their activity, it is in general much smaller than the volume of the binding site. An average value for the whole dataset is $V = 125 \pm 60$ Å^3.

The JEDI hydrophobicity descriptor shares similarities with the descriptor used by by Eyrisch et al. [42,43] In accordance with previous literature studies, druggable binding sites tend to have higher average hydrophobicity values ($H_a = 0.72 \pm 0.03$) than non-druggable binding sites ($H_a = 0.60 \pm 0.04$). This descriptor was found to be the most significant contribution to the $JEDI_{score}$ values with a weight β almost five times larger than the α volume coefficient (Table 1). This observation is in a good agreement with the literature, where the apolar character of a cavity is usually the most important structural descriptor for druggability assessment.[13,39] Indeed, a single hydrophobicity descriptor has been shown in some instances to be sufficient to distinguish druggable proteins from nondruggable proteins. [16,24]

Druggability scoring of diverse protein structures.

The JEDI parameters were first optimized using multiple linear regressions and the elitist selection variant of the genetic algorithm methodology implemented in the python module PyEvolve.[26] JEDI druggability scores obtained at the end of the process are shown in Figure 3A. For comparison, fpocket was used to calculate the druggability score of each protein in the training dataset (Figure 3B). The results suggest that JEDI predictions are slightly more accurate than those obtained using fpocket with a r^2 of 0.63 ± 0.11 and 0.52 ± 0.13 respectively. Closer inspection of Figure 3A shows that JEDI discriminates fairly well sites categorized as ‘’undruggable’’ from those classified as ‘’druggable’’, but proteins in the ‘’difficult’’ category show a large scatter in $JEDI_{score}$ values. Clearly, the exact ‘’experimental’’ DCD druggability score assigned to a given protein can be debated, and this must be kept in mind when calibrating computational methods against this dataset. Additional tests were conducted by positioning the grid on buried or solvent exposed regions of the protein Malate Dehydrogenase (PDB 1BMD), where no apparent pockets were observed. The resulting $JEDI_{score}$ values where invariably lower than 1.5.

Detailed structural analyses of accurate and inaccurate druggability predictions for representatives druggable and non-druggable protein binding sites is useful to characterize the strengths and weaknesses of the present approach. Four representative structures were chosen for this purpose (Figure 4), and JEDI descriptor values for these structures are shown in table 4. Figure 4A represents the binding site of a malate dehydrogenase in complex with the coenzyme NAD (PDB 1BMD). This enzyme has been classified as nondruggable due to the difficulty of finding a drug-like compound able to compete with NAD for access to the binding site. The binding affinity of several known nucleotide inhibitors have been previously determined by enzymatic assays.[44] The best competitive inhibitor is the cyclic nucleotide cAMP, presenting a K_i value 560 nM. If this protein is clearly evaluated as nondruggable by fpocket (*score* = 0.11), it remains challenging for other methodologies such as the NMR-based approach developed by Hadjuk and coworkers, which predicts the cavity as having an intermediate druggability.[12] This is in line with the observed $JEDI_{score}$ value for this system (5.1). The relatively high $JEDI_{score}$ is largely due to the relative large active volume V of the binding site (157 Å^3), which is in the range of V values typical for druggable sites (Table 3, first row). Thus, that malate dehydrogenase is not considered druggable in practice may be more a reflection of the difficulty for a drug-like molecule to compete with NAD at a ca. 300 μM expected intracellular concentration in mammalian cells,[45] rather than the occurrence of an unusually polar or shallow binding site. An example of a correct nondruggable prediction is depicted in figure 4B for the binding site of Inositol Polyphosphate (IP) phosphatase.[46] In addition to a small active volume due to a poor degree of enclosure, this small pocket presents a very low hydrophobicity score (Table 3, second row). This is mainly because of a Calcium ion in the binding site. A correctly predicted druggable cavity is shown in figure 4C. This mostly apolar well-enclosed pocket corresponds to the binding site of the S810L mutant mineralocorticoid receptor interacting with spironolactone (Table 3, third row).[47] This inhibitor has shown IC_{50} values in the range of 1.6 - 60 nM in a cell-based luciferase reporter assay.[48] Lastly, figure 4D depicts a druggable binding site that is incorrectly predicted to be ‘difficult’ to target. In addition to a high polarity caused by the presence of a zinc ion buried in the pocket, the binding site of carbonic anhydrase II is particularly small.[49] Most successful carbonic anhydrase inhibitors exploit direct interactions with the buried zinc ion. The present version of JEDI does not account for potentially favorable metal-ligand interactions and this explains the discrepancy between the $JEDI_{score}$ and DCD_{score} values (Table 3, fourth row).

Sensitivity to minor structural variations, and computational cost.

A potential concern at the outset of the project was that $JEDI_{score}$ values would be unduly sensitive to minor structural variations that are typically observed when crystal structures of the same protein are solved and refined independently. A major motivation for the development of JEDI was to observe variability in $JEDI_{score}$ between different structures of the same protein, only when conformational changes relevant for drug design are observed (e.g. a side-chain flip). This feature requires a subtle balance, on the one hand the methodology should not be too sensitive to very minor structural changes, but on the other hand it should be sufficiently sensitive to capture a fluctuation in druggability if the rearrangement is significant. The strategy here adopted was to evaluate the sensitivity of the $JEDI_{score}$ values for comparable conformations of the same protein interacting with different ligands. The structural similarity was quantified by means of $RMSD$ calculations on the backbone and

C_β atoms of the binding site atoms of each protein. Selected proteins for which *RMSD* values of the different structures were less than 0.5 Å were retained for further analysis. Additionally, visualization of the binding sites confirmed that there was no noticeable difference in binding site conformation between the different selected structures. Figure 5A shows the distribution of $JEDI_{score}$ values obtained by this analysis for a representative protein taken from the 'nondruggable', 'difficult' and 'druggable' categories of the DCD dataset. Although small fluctuations in $JEDI_{score}$ are observed in the case of the difficult and the druggable binding site, the results suggest nevertheless a good reproducibility and robustness to insignificant structural changes. By contrast the fpocket methodology sometimes exhibits substantial variations in druggability that complicates interpretation of the scores (Figure 5B). As an additional test of sensitivity, the dependence of the $JEDI_{score}$ values on the initial placement of the grid was assessed by evaluating the druggability of the same protein after translations of grid point coordinates by up to ±0.5 Å in the *x*, *y*, and *z* directions in Cartesian space. The druggability predictions were found to be quite insensitive to such translations, with fluctuations in the $JEDI_{score}$ values in the range of 0.1.

Next the computational cost of the JEDI calculations was assessed. An important consideration is that the calculations should not slow down too much molecular dynamics simulations. Benchmarks are shown in Table 4. If JEDI is used to monitor druggability values on the fly during an MD simulation, then it isn't necessary to evaluate druggability at every time-step, as snapshots between successive times-steps are highly correlated. With druggability evaluation every 1 ps the time incurred is negligible, unless the MD simulation is parallelized across multiple processors. Likewise, single-point druggability estimates of a protein structure are far faster than alternative methodologies that take seconds to minutes.[14,16,50] The implementation of MD simulation protocols biased with JEDI requires a druggability calculation at each time-step. In this case the performance loss is approximately a factor of 1.4 to 2.7, depending on the number of processors used to speed-up the evaluation of the non-bonded energies. Evidently, further gains in efficiency could be achieved by parallelizing key subroutines in the JEDI code. Alternatively, multiple time-step algorithms schemes for collective variables as proposed recently by Ferrarotti et al. could be used to decrease the computational cost further.[51] The relative efficiency is also influenced by the choice of an implicit solvent model for this study, which dramatically speeds up the evaluation of non-bonded energies. Overall, the performance was deemed acceptable, given scope for future improvements.

Application to a hidden pockets dataset.

Validation of the methodology was pursued by analysis of a set of six proteins known to adopt distinct binding site conformations in the presence of different ligands (Figure 6). In each instance, two conformations for each protein were selected for druggability assessments. Protein structures were aligned and a grid defined from the largest ligand was used to compute a $JEDI_{score}$ value for both conformations. In all instances the ligand atoms were ignored for druggability calculations. The results of this analysis are shown in Table 5. Human phenylethanolamine N-methyltransferase (hPNMT) is an enzyme involved in the synthesis of epinephrine from norepinephrine using the cofactor S-adenosyl- L-methionine (SAM) to methylate the primary amine of noradrenaline. Two different hPNMT inhibitors, **1** and **2**, have been reported to inhibit the enzyme with K_i values of 0.28 μM and 0.063 μM respectively (radiochemical assay).[52] It has been shown that these two ligands bind to different conformations of the hPNMT binding site (Figure 6A). Both compounds engage in significant hydrophobic interactions, but the larger ligand (**2**) positions a *p*-chlorophenyl group in a cavity that is hidden in the hPNMT/**1** complex. Formation of the enlarged cavity in hPNMT/**2** necessitates the rearrangement of the side-chain Lys57, as well as a small displacement of helix α3. The JEDI calculations were able to capture a favorable increase in druggability of ca. 0.8 units for the protein binding site conformation seen in hPNMT/**2** in comparison with hPNMT/**1**. The change in druggability is due to a favorable increase in both V and H_a (Table 5, first row).

The von Hippel-Lindau protein (pVHL) forms a complex with the proteins CUL2, Elongin B and C, and Rbx1. This complex is involved in the ubiquitination of the transcription factor hypoxia-inducible factor (HIF-1α), leading to proteasome-mediated degradation of HIF-1α.[53] Small molecules **3** and **4** have been reported to inhibit interactions between pVHL and HIF-1α with K_d values of 86.1 μM and 27.7 μM respectively (fluorescence polarization assay).[19] The ligands occupy the same binding site, but a different orientation of Arg107 is observed, giving rise to a slightly more enlarged cavity in VHL/**4** (Figure 6B). This translates into a slightly higher $JEDI_{score}$ value for VHL/**4** over VHL/**3**. This is because repositioning of Arg107 increased the value of H_α in VHL/**4**. However this is partially offset by a decrease in V_a. This is because the displacement of Arg107 exposes more grid points to the solvent, and as a consequence, grid points previously fully active become partially active (Table 5, second row).

Serine/threonine-protein kinase or polo-like kinase 1 (PLK-1) is an enzyme involved in the regulation of cell division., The PLK-1 inhibitor **5** binds with an IC_{50} = 730 nM (fluorescence polarization assay) to the ATP binding site, and also to a subpocket that has been called the adaptive pocket, whereas the inhibitor **6** shows an IC_{50} of 530 nM (kinase enzymatic assay) and binds to the native purine-pocket of the active site (Figure 6C).[54,55] However the larger active volume observed in the PLK-1/**5** bound conformation is mainly due to active grid points around the methylpiperazine moiety of **5.** These grid points are inactive in the PLK-1/**6** complex because they are too solvent exposed. The adaptive pocket seen in PLK-1/**6** is predicted to be less druggable than the native pocket seen in PLK-1/**5** by ca. 0.8 units (Table 5, third row).

Prostate specific membrane antigen (PSMA) is a glycoprotein overexpressed as a homodimer in many forms of prostate cancer. Compound **7** is an example of a first generation of PSMA inhibitors that binds the very polar binding site of PSMA with a K_i of 11 nM (fluorescence-based NAALADase assay).[56] More recently, compounds belonging to the class of antibody recruiting small molecules targeting prostate cancer (ARM-P) have been reported, and compound **8** binds PSMA with a K_i of 0.02 nM (enzymatic assay).[57,58] A crystallographic structure of the PSMA/**8** complex revealed that **8** binds to an open PSMA conformation that was not observed in the PSMA/**7** complex. The large difference in binding affinities between **7** and **8** appears to be well reproduced by a large difference in $JEDI_{score}$ values (Table 5, fourth row). However in this instance the active volume V is much larger than for a typical small molecule binding site and as a consequence the druggability score is strongly penalized by

$V_{druglike}$. This indicates that the predictions for PSMA should be treated with care as the binding site differs substantially from those present in the training set. Compound **8** is unusual because it is made of a long flexible linker connecting a moiety positioned in the buried PSMA active site (Figure 6D blue square), and another moiety positioned in the arene binding site at the protein surface (Figure 6D black square). The JEDI analysis was therefore repeated by splitting the initial grid in two regions to predict the druggability of each pocket independently. A first grid was placed around the active site, and a second was located around the DNP pocket. A low score was observed for the active site in both instances ($JEDI_{score}$ = 2.3 and 2.6 respectively), because of a very high polarity due to the presence of by several ions and polar and charged amino acids in the active site. The DNP pocket in PSMA/**8** does score slightly higher ($JEDI_{score}$ = 3.1) than the same region in the PSMA/**7** complex ($JEDI_{score}$ = 2.3) but the score remains small because the DNP pocket is relatively small. Thus the PSMA binding site is a good illustration of challenging conditions encountered when performing JEDI analysis of binding sites for ligand that depart from typical rule-of-five compliant small molecules.

HIV-RT is an enzyme playing a crucial role in the replication of the HIV virus. Several non-nucleoside RT inhibitors (NNRTIs) are available in the clinic for HIV treatments.[59-63] Druggability predictions were compared for the NNRTI-binding pocket of the apo structure of HIV-1 RT and in complex with **9** (Figure 6E). This compound belongs to the second generation of NNRTIs and inhibits wild type HIV-RT with an IC_{50} of 2.1 nM (antiviral assay).[64,65] The binding site of the HIV-RT/**9** complex was found to be one of the most druggable pocket analysed in this work. It is noteworthy that the NNRTI cavity is actually partially formed in the apo protein, and has an active volume of V = 192 A^3 The holo structure features an enlarged binding site and side chains rearrangements that increase the hydrophobicity H_a (Table 5, fifth row).

Interleukin-2 (IL-2) is a cytokine playing a crucial role in the regulation of white blood cells of the immune system. The small molecule **10** binds to a pocket only partially present in the apo structure. An additional cavity is present in the holo complex and it forms by displacement of two residues, Phe42 and Glu62 (Figure 6F).[66] A similar pocket volume descriptor is observed for both apo and holo forms of

IL-2. However this time, a higher druggability score was predicted in absence of ligand, because the hydrophobicity H_a is lower in the IL-2/**10** complex (Table 5, sixth row). This occurred because the motion of Phe42 and Glu62 promotes hydrogen bonding with Glu62 and Lys43, activating grid points close to polar atoms, thus decreasing hydrophobicity.

Overall, the methodology is clearly able to correlate fluctuations in druggability score with noteworthy binding site conformational changes that have the potential to impact structure-based ligand design activities. In five cases out of six, the conformation with the highest $JEDI_{score}$ corresponds to the conformation that binds the most tightly bound ligand. Careful interpretation of the results is needed when considering unusual protein-ligand complexes, such as PSMA/**8**. Quantitative correlation with binding affinities is not expected since the ligands differ. Further, druggability is not exclusively linked to binding affinity. PSMA is an example of a binding site for which ligands with very low K_i values are known (**7** and **8**), but the low predicted druggability score is adequate since most of the binding affinity is achieved by means of strongly polar ligand moieties positioned in the active site. These in turn translate into inauspicious drug-like properties, such as low cell permeability. [56,58]

On the fly evaluation of druggability during MD simulations.

Further tests were conducted with MD simulations of VHL. Druggability values were collected every ps over the course of a 50 ns simulation of apo VHL or VHL/**3**. The results are shown in Figure 7. The binding site druggability remained stable throughout the VHL/**3** simulation, with an average $JEDI_{score}$ of 7.8±0.6 which is consistent with the expected value from previous analyses (Table 5, row 2). Clustering analysis reveals only one major binding site conformation (76% of the trafectory), that is depicted in Figure 7C (right panel). By contrast, the apo simulation shows an average druggability score of 5.7±0.8. Numerous structurally different binding site conformations are sampled. In the present MD simulations, the apo binding pocket is quickly obstructed by the rearrangement of Tyr98 and His110, inducing a drop of druggability. This could reflect inaccuracies in the protein force field used for the present study. Dozens of clusters were identified and the most populated ($JEDI_{score}$ ca. 6.3) is present in

67% of the simulation (Figure 7C, left panel). This partially closed conformation is mainly stabilized by hydrogen bonds between the phenolic OH group of Tyr98 and the protein backbone. His110 is very flexible throughout the simulation. Surprisingly, significant side-chain rearrangements that partially block the binding site do not affect dramatically the $JEDI_{score}$ values. This occurs here because the shift in position for Tyr98 has created a new hydrophobic sub-pocket that contributes favorably to the $JEDI_{score}$. However this sub-pocket is now occluded by Tyr98 and disconnected from the rest of the binding site. Further, the rest of the VHL binding site is still partially present, including the central pyrrolidine binding pocket. Binding site conformations that correspond to extreme druggability fluctuations seen in the apo simulation are depicted in figure 7D. In general, the apo conformations that present high $JEDI_{score}$ values were found to be structurally similar to the VHL/**3** conformation seen in the crystal structure.

Biasing MD simulations with the JEDI potential.

Umbrella sampling simulations were performed for apo VHL and VHL/**3** using equation 15 and by varying force constant values for κ and target $JEDI_{score}$ values s_0. No reweighting of the biased simulations was performed, thus all results presented below correspond to equilibrium properties of the biased Hamiltonians. The results are depicted in Figure 8. Apo simulations were biased to achieve a $JEDI_{score}$ of 8, in expectation with the values previously observed for ligand bound complexes (Table 5, second row). Figure 8A (upper panel) indicates that the target druggability value is rapidly achieved in all instances. As expected fluctuations from the target value decrease with increased κ values. The trajectory obtained using κ = 2 000 kJ.mol^{-1}.nm^{-2} was subjected to further clustering. The most populated clusters (51% of the overall trajectory) are very similar to the VHL/**3** structure, with *RMSD* values always inferior to 2.0 Å. In the unbiased MD simulation of apo VHL, only 14% of the computed conformation exhibited an *RMSD* to the VHL/**3** conformation that was smaller than 2.0 Å. Some clusters still contain conformations with Tyr98 pointing inside the binding site, but the occurrence is greatly decreased. His110 was also found to be much less flexible. It is apparent that the ligand binding site is almost fully formed in the most populated cluster of the biased apo VHL simulation (Figure 8A, bottom part).

The umbrella sampling simulations of VHL/**3** were performed to encourage the binding site to adopt more druggable conformations. A reference value s_0 = 9 was selected based on the $JEDI_{score}$ of VHL/**4**. Figure 8B upper part shows that higher κ values are needed to achieve the desired s_0 value. This indicates that the conformations with high $JEDI_{score}$ values do not form spontaneously. The increase in $JEDI_{score}$ values that is achieved correlates largely with the position of Arg107. This amino acid initially closes the binding site, but with the present bias, it shifts rapidly to a solvent exposed position, thus causing an enlargement of the binding site. This motion was rarely observed in unbiased MD simulations.

Next, more significant structural rearrangements were sought by performing umbrella sampling simulations of apo VHL with s_0 = 3.0. Results obtained with κ = 2000 kJ.mol^{-1}.nm^{-2} are shown in Figure 9. Requesting such a low target druggability value forces VHL to largely collapse the binding site. Here the collapse is even more pronounced than observed in the unbiased apo VHL simulations, with the binding pockets of the isoxazole and pyrrolidine moieties completely masked. Consequently, the pocket volume descriptor V_a decreases, and the active volume V becomes sufficiently low such that the $V_{druglike}$ term penalizes the $JEDI_{score}$ values. The hydrophobicity descriptor H_α is stable during the biased simulation, with an average value slightly lower than observed in the unbiased apo VHL simulation. The closure of the binding site has totally or partially inactivated numerous grid points that were previously in a buried cavity, leaving only a few active grid points at the protein surface and near polar groups. An illustration of the most populated cluster (73% of the trajectory) is depicted in figure 9B.

Umbrella sampling simulations of apo VHL were also performed by setting s_0 = 10 and κ = 2000 kJ.mol^{-1}.nm^{-2} to encourage the exploration of conformations with high druggability. The results are presented in Figure 10A. As observed previously, the simulation is rapidly sampling conformations in the requested range of $JEDI_{score}$. As expected, V_a and H_a are almost always higher than in the previously described simulations. However, larger fluctuations are observed in both descriptors throughout the biased simulation. An increase in hydrophobicity H_a is always offset by a decrease of the active volume descriptor V_a and vice versa. Clustering analysis of the trajectory here reveals at least two significant distinct clusters (populations 18% and 8% respectively). The second cluster (Figure 10C) corresponds to a low V_a / high H_a binding site conformation that is significantly different from the VHL/**3** structure. The pyrrolidine pocket has collapsed and side-chains rearranged to expose hydrophobic groups to the surface. The first cluster (Figure 10B) corresponds to a conformation comparable to the VHL/**3** holo structure. Additionally, Arg107 has adopted a solvent exposed position that contributes favorably to the $JEDI_{score}$ as demonstrated previously (Table 5, second row). A significant difference that was not observed in previous simulations is the rearrangement of Arg69 in the left-hand side part of the binding site. This conformational rearrangement leads to a more extended cavity with high druggability scores. The flexibility of the left hand side pocket, has been recently discussed in the literature in the context of crystallographic structure analyses of multiple VHL ligand complexes,[67] and Galdeano et al. have suggested that additional interactions between ligands and this part of the binding site may facilitate the development of improved VHL ligands.

Docking ligands into JEDI computed conformations.

Several docking calculations were carried out to evaluate the utility of the conformational ensembles computed from the umbrella sampling simulations. Figure 11A depicts results obtained using the computed apo VHL conformation closest to the average conformation of the most populated cluster taken from an umbrella sampling simulation with s_0 = 10.0 and κ = 2000 kJ.mol^{-1}.nm^{-2} (Figure 10B, top). Ligand **3** was found to adopt a pose that bears a substantial similarity with the crystallographic position of the ligand (RMSD of 3.6 Å, VINA binding energy of -5.6 kcal.mol^{-1}). This is however not the top-scored pose which had a VINA binding energy of -6.2 kcal.mol^{-1}. Qualitatively the discrepancy with the crystallographic binding mode is mostly due to a shift of the isoxazole ring of **3** that is involved instead in stacking interactions with Tyr112. Closer inspection of the computed complex indicates that this binding mode is preferred because the computed ‘’left-hand side’’ VHL pocket that would normally host the isoxazole ring is too shallow. However, small fluctuations in pocket depth are apparent in snapshots that are present in the same cluster, and it is possible to manually select a snapshot with a left-hand-side pocket that more closely resembles the crystallographic structure (RMSD of the binding-site sidechains depicted in Figure 11A to the crystallographic conformation is 1.3 Å, see Figure S2 for details). Repeating docking calculations on this conformation (Figure 11B) yields indeed a well scored pose (VINA binding energy -6.4 kcal.mol^{-1}) that reproduces fairly well the crystallographic position of the ligand (RMSD of 2.1 Å) though this is again not the top-scoring pose which had a VINA binding energy of -7 kcal.mol^{-1}. As a control, the same docking protocol was also applied to the computed apo VHL conformation closest to the average conformation of the most populated cluster from an unbiased classical MD simulation (Figure 11C). As expected, the lowest-RMSD pose was significantly different from the crystallographic binding mode of **3** (RMSD of 5.4 Å, VINA binding energy -6.1 kcal.mol^{-1}). The docking calculations were repeated allowing side-chain flexibility of Tyr98, Ile109 but no improvements were observed. This is likely because significant conformational changes involving both side-chain and backbone atoms rearrangements are necessary to form the ligand binding site from the apo protein conformations sampled from the unbiased MD simulation. Conversely, little improvements was seen in the RMSD of the ligands docked into the JEDI computed conformations with the aid of a flexible side-chain docking protocol, presumably because the binding site is already largely formed.

Conclusions

A novel approach to assess protein binding site druggability has been developed. The fast, continuous and differentiable JEDI druggability estimator has been implemented in PLUMED and has been used as a collective variable in order to compute protein druggability at every integration step of a MD simulation.[23] While the use of MD to sample protein conformations limits throughput, it offers the advantage of sampling more accurate protein conformations than those that may be generated by alternative molecular modelling approaches. As discussed elsewhere, high accuracy is crucial if computed protein conformations are to be subject to follow-up virtual screens.[68] The methodology is able to distinguish nondruggable, difficult and druggable pockets (r^2 = 0.6±0.1), and is relatively insensitive to insignificant structural rearrangements in a binding site.

Some limits in the estimator were exposed, for instance neglect of potential metal-ligand interactions. This could be remedied with additional structural descriptors. In addition, the present scoring function is only calibrated for detecting cavities that bind drug-like small molecules. JEDI was tested additionally on a dataset of hidden pockets for structurally diverse protein targets. The results show a good ability for the approach to detect conformational changes that influence the druggability of a protein binding site. With the present version of the method, care must be taken when performing this analysis on binding sites for ligands that depart from typical rule-of-five compliant small molecules.

The main novelty of the approach lies in its potential to bias MD simulations with a JEDI force that will encourage a protein region to adopt conformations that match desired druggability scores. The results obtained through several umbrella sampling simulations of VHL indicate that JEDI enables the rapid sampling of 'holo-like' protein binding site conformations that are rarely seen in unbiased apo MD simulations. For structure-based drug design purposes this would be useful to identify tractable conformations in drug targets that may be otherwise considered undruggable from crystallographic analysis. An advantage of the approach over induced-fit docking/MD refinement protocols is that druggable cavities can be identified for targets that lack known ligands.[69] JEDI also enables biased simulations of protein-ligand complexes. For structure-based drug design purposes, this would be useful to identify enlarged cavities that could accommodate a larger analog of an existing ligand.

Further work will focus on replacing the GBSA implicit solvent model with explicit solvent models, and this is expected to improve the accuracy of the computed conformations.[70] Additional work is also desirable to identify the most efficient and accurate docking protocols to use in combination with JEDI computed conformations. Clustering of the biased simulations in VHL has identified in many instances several structurally distinct conformations that match a given target druggability value. That druggability is a degenerate collective variable was expected, and an exciting direction for this work is to couple the JEDI calculations with other collective variables to resolve distinct conformational states. This will facilitate the evaluation of the free energy of these hidden conformational states with respect to the native state conformation. This parameter is likely to be important for practical applications. Presumably the feasibility of targeting productively with a ligand a putative cryptic binding site hinges on an acceptable stability relative to the native state.[68] Several enhanced sampling methodologies could in principle be suitable to this end,[71] and progress towards this objective will be reported in due course.

Supporting Information

Table of computed $JEDI_{score}$, fpocket$_{score}$ and DCD_{score} for every protein in the DCD training set. Expressions for the derivatives of the JEDI potential. This material is available free of charge via the Internet at http://pubs.acs.org.

Acknowledgements

JM is supported by a Royal Society University Research Fellowship. RC was supported by a Marie Curie International re-integration grant from the European Commission (PIRG-GA-2010-276785). The research leading to these results has received funding from the European Research Council under the European Union’s Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement no. 336289.

References

(1) Kola, I.; Landis, J. *Nat. Rev. Drug Discov.* **2004**, *3*, 711–715.

(2) Brown, D.; Superti-Furga, G. *Drug Discov. Today* **2003**, *8*, 1067–1077.

(3) Hopkins, A. L.; Groom, C. R. *Nat. Rev. Drug Discov.* **2002**, *1*, 727–730.

(4) Russ, A. P.; Lampel, S. *Drug Discov. Today* **2005**, *10*, 1607–1610.

(5) Wyatt, P. G.; Gilbert, I. H.; Read, K. D.; Fairlamb, A. H. *Curr. Top. Med. Chem.* **2011**, *11*, 1275–1283.

(6) Sheridan, R. P.; Maiorov, V. N.; Holloway, M. K.; Cornell, W. D.; Gao, Y.-D. *J. Chem. Phys.* **2010**, *50*, 2029–2040.

(7) Edfeldt, F. N. B.; Folmer, R. H. A.; Breeze, A. L. *Drug Discov. Today* **2011**, *16*, 284–287.

(8) Goodford, P. J. *J. Med. Chem.* **1985**, *28*, 849–857.

(9) Guarnieri, F.; Mezei, M. *J. Am. Chem. Soc.* **1996**, *118*, 8493–8494.

(10) Stultz, C. M.; Karplus, M. *Proteins* **1999**, *37*, 512–529.

(11) Dennis, S.; Camacho, C. J.; Vajda, S. *Proteins* **2000**, *38*, 176–188.

(12) Hajduk, P. J.; Huth, J. R.; Fesik, S. W. *J. Med. Chem.* **2005**, *48*, 2518–2525.

(13) Cheng, A. C.; Coleman, R. G.; Smyth, K. T.; Cao, Q.; Soulard, P.; Caffrey, D. R.; Salzberg, A. C.; Huang, E. S. *Nat. Biotechnol.* **2007**, *25*, 71–75.

(14) Halgren, T. A. *J. Chem. Inf. Model.* **2009**, *49*, 377–389.

(15) Halgren, T. *Chem Biol Drug Des* **2007**, *69*, 146–148.

(16) Schmidtke, P. P.; Barril, X. X. *J. Med. Chem.* **2010**, *53*, 5858–5867.

(17) Krasowski, A.; Muthas, D.; Sarkar, A.; Schmitt, S.; Brenk, R. *J. Chem. Inf. Model.* **2011**, *51*, 2829–2842.

(18) Cozzini, P.; Kellogg, G. E.; Spyrakis, F.; Abraham, D. J.; Costantino, G.; Emerson, A.; Fanelli, F.; Gohlke, H.; Kuhn, L. A.; Morris, G. M.; Orozco, M.; Pertinhez, T. A.; Rizzi, M.; Sotriffer, C. A. *J. Med. Chem.* **2008**, *51*, 6237–6255.

(19) Van Molle, I.; Thomann, A.; Buckley, D. L.; So, E. C.; Lang, S.; Crews, C. M.; Ciulli, A. *Chem. Biol.* **2012**, *19*, 1300–1312.

(20) Seco, J.; Luque, F. J.; Barril, X. *J. Med. Chem.* **2009**, *52*, 2363–2371.

(21) Lexa, K. W.; Carlson, H. A. *J. Am. Chem. Soc.* **2011**, *133*, 200–202.

(22) Bakan, A.; Nevins, N.; Lakdawala, A. S.; Bahar, I. *J. Chem. Theory Comput.* **2012**, *8*, 2435–2447.

(23) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; Parrinello, M. *Comput. Phys. Commun.* **2009**, *180*, 1961–1972.

(24) Le Guilloux, V.; Schmidtke, P.; Tuffery, P. *BMC Bioinformatics* **2009**, *10*, 168–168.

(25) Kabsch, W. *Acta Crystallogr Sect A Cryst Phys Diffr Theor Gen Crystallogr* **1976**, *32*, 922–923.

(26) Butterfield, A.; Vedagiri, V.; Lang, E.; Lawrence, C.; Wakefield, M. J.; Isaev, A.; Huttley, G. A. *BMC Bioinformatics* **2004**, *5*, 1.

(27).

(28) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.

(29) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. *J. Phys. Chem. A* **1997**, *101*, 3005–3014.

(30) Onufriev, A.; Bashford, D.; Case, D. A. *Proteins* **2004**, *55*, 383–394.

(31) Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 014101.

(32) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. *Proteins* **2010**, *78*, 1950–1958.

(33) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.*

2004, *25*, 1157–1174.
(34) Sousa da Silva, A. W.; Vranken, W. F. *BMC Res. Notes* **2012**, *5*, 367.
(35) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. *J. Mol. Graph. Model.* **2006**, *25*, 247–260.
(36) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187–199.
(37) Trott, O.; Olson, A. J. *J. Comput. Chem.* **2010**, *31*, 455–461.
(38) Seeliger, D.; de Groot, B. L. *J. Comput. Aided Mol. Des.* **2010**, *24*, 417–422.
(39) Volkamer, A.; Kuhn, D.; Grombacher, T.; Rippmann, F.; Rarey, M. *J. Chem. Inf. Model.* **2012**, *52*, 360–372.
(40) Perola, E.; Herman, L.; Weiss, J. *J. Chem. Inf. Model.* **2012**, *52*, 1027–1038.
(41) An, J.; Totrov, M.; Abagyan, R. *Mol. Cell Proteomics* **2005**, *4*, 752–761.
(42) Eyrisch, S.; Helms, V. *J. Med. Chem.* **2007**, *50*, 3457–3464.
(43) Pérot, S.; Sperandio, O.; Miteva, M. A.; Camproux, A.-C.; Villoutreix, B. O. *Drug Discov. Today* **2010**, *15*, 656–667.
(44) Harris, D. G.; Marx, D. P.; Anderson, J. M.; McCune, R. W.; Zimmerman, S. S. *Nucleos. Nucleot. Nucl.* **2002**, *21*, 813–823.
(45) Yamada, K.; Hara, N.; Shibata, T.; Osago, H.; Tsuchiya, M. *Anal. Biochem.* **2006**, *352*, 282–285.
(46) Miller, G. J.; Wilson, M. P.; Majerus, P. W.; Hurley, J. H. *Mol Cell* **2005**, *18*, 201–212.
(47) Huyet, J.; Pinon, G. M.; Fay, M. R.; Fagart, J.; Rafestin-Oblin, M.-E. *Mol. Pharmacol.* **2007**, *72*, 563–571.
(48) Roll, D. M.; Barbieri, L. R.; Bigelis, R.; McDonald, L. A.; Arias, D. A.; Chang, L.-P.; Singh, M. P.; Luckman, S. W.; Berrodin, T. J.; Yudt, M. R. *J. Nat. Prod.* **2009**, *72*, 1944–1948.
(49) Di Fiore, A.; Pedone, C.; Antel, J.; Waldeck, H.; Witte, A.; Wurl, M.; Scozzafava, A.; Supuran, C. T.; De Simone, G. *Bioorgan. Med. Chem. Lett.* **2008**, *18*, 2669–2674.
(50) Volkamer, A.; Kuhn, D.; Rippmann, F.; Rarey, M. *Bioinformatics* **2012**, *28*, 2074–2075.
(51) Ferrarotti, M. J.; Bottaro, S.; Pérez-Villa, A.; Bussi, G. *J. Chem. Theory Comput.* **2015**, *11*, 139–146.
(52) Grunewald, G. L.; Seim, M. R.; Regier, R. C.; Criscione, K. R. *Bioorgan. Med. Chem.* **2007**, *15*, 1298–1310.
(53) Cockman, M. E.; Masson, N.; Mole, D. R.; Jaakkola, P.; Chang, G.-W.; Clifford, S. C.; Maher, E. R.; Pugh, C. W.; Ratcliffe, P. J.; Maxwell, P. H. *J. Biol. Chem.* **2000**, *275*, 25733–25741.
(54) Kothe, M.; Kohls, D.; Low, S.; Coli, R.; Cheng, A. C.; Jacques, S. L.; Johnson, T. L.; Lewis, C.; Loh, C.; Nonomiya, J.; Sheils, A. L.; Verdries, K. A.; Wynn, T. A.; Kuhn, C.; Ding, Y.-H. *Biochemistry* **2007**, *46*, 5960–5971.
(55) Elling, R. A.; Fucini, R. V.; Hanan, E. J.; Barr, K. J.; Zhu, J.; Paulvannan, K.; Yang, W.; Romanowski, M. J. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **2008**, *64*, 686–691.
(56) Wang, H.; Byun, Y.; Barinka, C.; Pullambhatla, M.; Bhang, H.-E. C.; Fox, J. J.; Lubkowski, J.; Mease, R. C.; Pomper, M. G. *Bioorgan. Med. Chem. Lett.* **2010**, *20*, 392–397.
(57) Zhang, A. X.; Murelli, R. P.; Barinka, C.; Michel, J.; Cocleaza, A.; Jorgensen, W. L.; Lubkowski, J.; Spiegel, D. A. *J. Am. Chem. Soc.* **2010**, *132*, 12711–12716.
(58) Murelli, R. P.; Zhang, A. X.; Michel, J.; Jorgensen, W. L.; Spiegel, D. A. *J. Am. Chem. Soc.* **2009**, *131*, 17090–17092.
(59) Grob, P. M.; Wu, J. C.; Cohen, K. A.; Ingraham, R. H.; Shih, C.-K.; Hargrave, K. D.; McTague, T. L.; Merluzzi, V. J. *AIDS Res. Hum. Retrov.* **1992**, *8*, 145–152.
(60) Romero, D. L.; Morge, R. A.; Genin, M. J.; Biles, C.; Busso, M.; Resnick, L.; Althaus, I. W.; Reusser, F.; Thomas, R. C.; Tarpley, W. G. *J. Med. Chem.* **1993**, *36*, 1505–1508.

(61) Romero, D. L.; Olmsted, R. A.; Poel, T. J.; Morge, R. A.; Biles, C.; Keiser, B. J.; Kopta, L. A.; Friis, J. M.; Hosley, J. D.; Stefanski, K. J.; Wishka, D. G.; Evans, D. B.; Morris, J.; Stehle, R. G.; Sharma, S. K.; Yagi, Y.; Voorman, R. L.; Adams, W. J.; Tarpley, W. G.; Thomas, R. C. *J. Med. Chem.* **1996**, *39*, 3769–3789.

(62) Esnouf, R. M.; Ren, J.; Hopkins, A. L.; Ross, C. K.; Jones, E. Y.; Stammers, D. K.; Stuart, D. I. *Proc. Nat. Acad. Sci. USA* **1997**, *94*, 3984–3989.

(63) Young, S. D.; Britcher, S. F.; Tran, L. O.; Payne, L. S.; Lumma, W. C.; Lyle, T. A.; Huff, J. R.; Anderson, P. S.; Olsen, D. B.; Carroll, S. S.; Pettibone, D. J.; O'Brien, J. A.; Ball, R. G.; Balani, S. K.; Lin, J. H.; Chen, I.-W.; Schleif, W. A.; Sardana, V. V.; Long, W. J.; Byrnes, V. W.; Emini, E. A. *Antimicrob. Agents Chemother.* **1995**, *39*, 2602–2605.

(64) Hsiou, Y.; Ding, J.; Das, K.; Clark, A. D.; Hughes, S. H.; Arnold, E. *Structure* **1996**, *4*, 853–860.

(65) Kertesz, D. J.; Brotherton-Pleiss, C.; Yang, M.; Wang, Z.; Lin, X.; Qiu, Z.; Hirschfeld, D. R.; Gleason, S.; Mirzadegan, T.; Dunten, P. W.; Harris, S. F.; Villaseñor, A. G.; Hang, J. Q.; Heilek, G. M.; Klumpp, K. *Bioorgan. Med. Chem. Lett.* **2010**, *20*, 4215–4218.

(66) Arkin, M. R.; Randal, M.; DeLano, W. L.; Hyde, J.; Luong, T. N.; Oslob, J. D.; Raphael, D. R.; Taylor, L.; Wang, J.; McDowell, R. S.; Wells, J. A.; Braisted, A. C. *Proc. Nat. Acad. Sci. USA* **2003**, *100*, 1603–1608.

(67) Galdeano, C.; Gadd, M. S.; Soares, P.; Scaffidi, S.; Van Molle, I.; Birced, I.; Hewitt, S.; Dias, D. M.; Ciulli, A. *J. Med. Chem.* **2014**, *57*, 8657–8663.

(68) Michel, J. *Phys. Chem. Chem. Phys.* **2014**, *16*, 4465–4477.

(69) Zhao, H.; Huang, D.; Caflisch, A. *ChemMedChem* **2012**, *7*, 1983–1990.

(70) Zhou, R. *Proteins* **2003**, *53*, 148–161.

(71) Abrams, C.; Bussi, G. *Entropy 2014, Vol. 16, Pages 163-199* **2013**, *16*, 163–199.

(72) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33–8– 27–8.

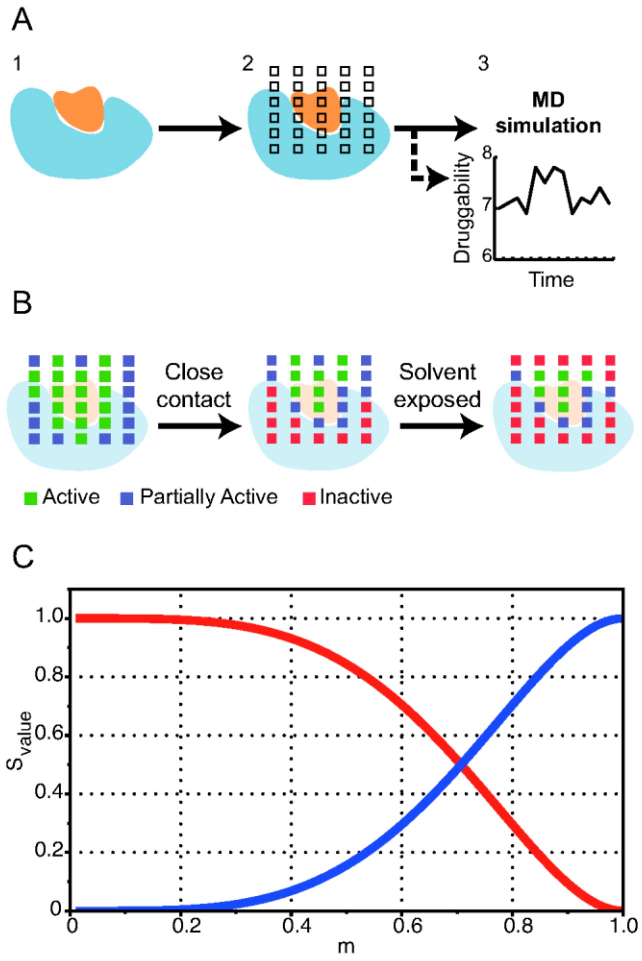


Figure 1. Overview of the JEDI protocol. A) The region of space for druggability assessment is determined and all atom models of the protein (and ligand if present) are prepared as for a conventional MD simulation (1). A grid with a 1.5 Å spacing is placed around the region of interest (2). A druggability assessment is performed either for the input structure only, or repeatedly over the course of an MD simulation (3). B) For every druggability evaluation, all grid points are assigned an initial activity according to their distance to the ligand in the input structure. Next, grid points overlapping with protein atoms in the binding site region are inactivated fully or partially. Finally, solvent exposed grid points are inactivated fully or partially. C) Graphical representation of the switching functions S_v^{on} (blue) and S_v^{off} (red) for k=1.0 and Δ=1.0.

Table 1. List of variables used to compute $JEDI_{score}$.

Symbol	Definition	Value
$V_{druglike}$	drug-like volume descriptor	[0.0,1.0]
V_a	pocket volume descriptor	[0.0,∞]
H_a	pocket hydrophobicity descriptor	[0.0,1.0]
V	active volume	[0.0,∞] Å^3
a_i	activity of the grid point *i*	[0.0,1.0]
H_i	hydrophobicity of the grid point *i*	[0.0,1.0]
M_{apolar}	number of C and S atoms in the binding site region	[0,∞]
$apolar_i$	Contact number with C and S atoms surrounding the grid point *i*	[0.0,∞]
M_{polar}	number of O and N atoms in the binding site region	[0,∞]
$polar_i$	Contact number with O and N atoms surrounding the grid point *i*	[0.0,∞]

Table 2. List of constants used to compute $JEDI_{score}$.

Symbol	Definition	Value
α	PLS derived volume coefficient	5.31
β	PLS derived hydrophobicity coefficient	24.29
γ	PLS derived constant according to α and β	-13.39
$V_{\mathbf{g}}$	grid resolution	1.5 Å^3
CC_{mind}	distance below which a grid point is fully in close contact with the protein	2.0 Å
ΔCC	distance interval over which a grid point is in partial contact with the protein	0.5 Å
E_{min}	minimum number of grid points between a distance of 2.5 Å and 3.5 Å from a grid point *i* interacting with the protein	10
ΔE	interval over which a grid point is considered as buried in the cavity	3
BS_{min}	minimum distance between a grid point and binding site atoms below which the maximal activity is fixed to 1	2.0 Å
ΔBS	distance interval over which the maximal activity is fixed to 0	6.0 Å
θ	constant used for minimum distance calculation	50.0 Å
$CC2_{min}$	minimum distance below which a grid point is overlapping the protein (for enclosure calculation)	0.15 Å
$\Delta CC2$	distance interval over which a grid point is in partial contact with the protein (for enclosure calculation)	0.14 Å
$GP1_{min}$	distance above which a grid point is considered for enclosure calculation	2.5 Å
$\Delta GP1$	distance interval over which a grid point is in partial contact with the protein	0.5 Å
$GP2_{min}$	distance below which a grid point is fully in close contact with the protein	3.0 Å
$\Delta GP2$	distance interval over which a grid point is in partial contact with the protein	0.5 Å
r_{hydro}	distance below which a grid point is fully in close contact with the protein (for hydrophobicity calculation)	4.0 Å
Δr_{hydro}	distance interval over which a grid point is in partial contact with the protein (for hydrophobicity calculation)	0.5 Å
V_{max}	volume below which $V_{druglike}$ is equal to 1	316 Å^3
ΔV_{max}	volume interval over which $V_{druglike}$ goes from 1 to 0	36 Å^3
V_{min}	volume below which $V_{druglike}$ is equal to 0	0.0 Å^3
ΔV_{min}	volume interval over which $V_{druglike}$ goes from 0 to 1	36 Å^3

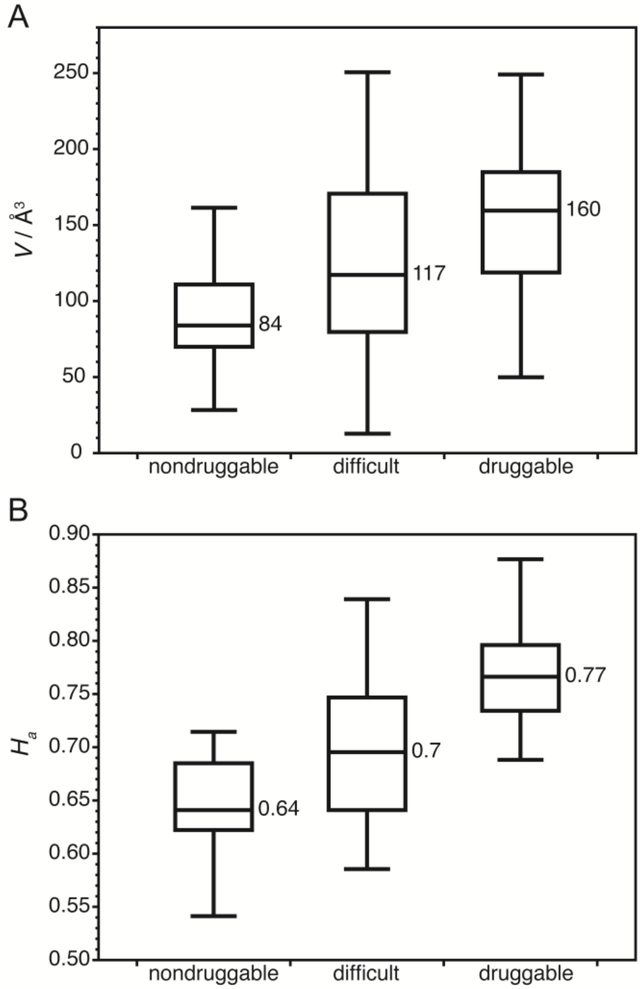


Figure 2. Boxplots of values of the (A) active volume V descriptor and (B) hydrophobicity descriptor H_a for the nondruggable, difficult and druggable systems of the training set. The box is defined using the first and the third quartile while the bar indicates the median. The edges of the boxplot represent the minimum and the maximum value observed for each category.

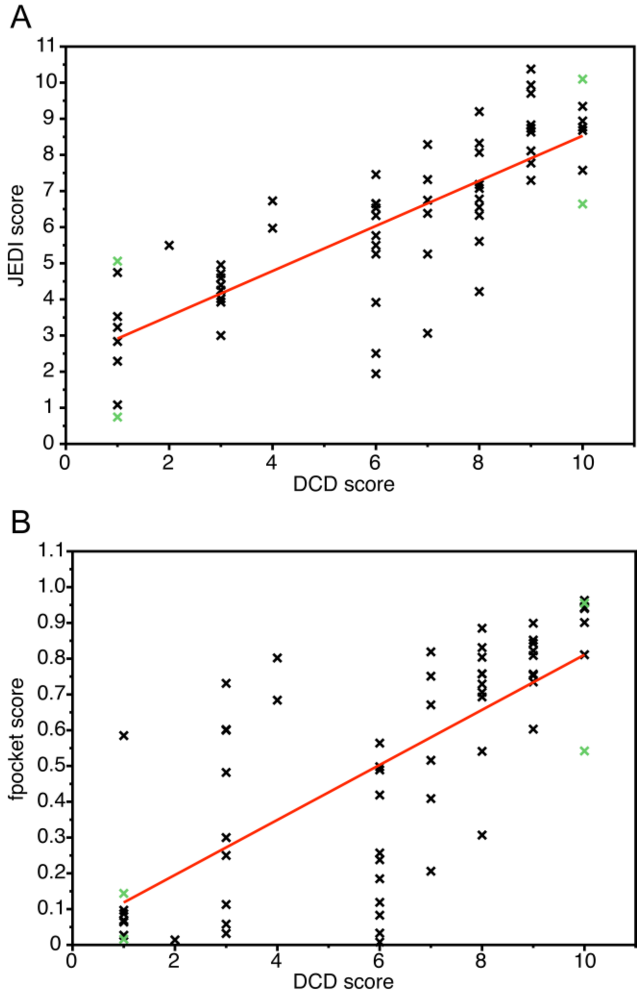


Figure 3. The correlation of computed druggability scores with DCD database druggability scores. A) Results for JEDI scores of the DCD training set. B) Results for fpocket scores of the DCD training set. Proteins discussed in the text, Figure 4 and Table 3 are represented with green crosses.

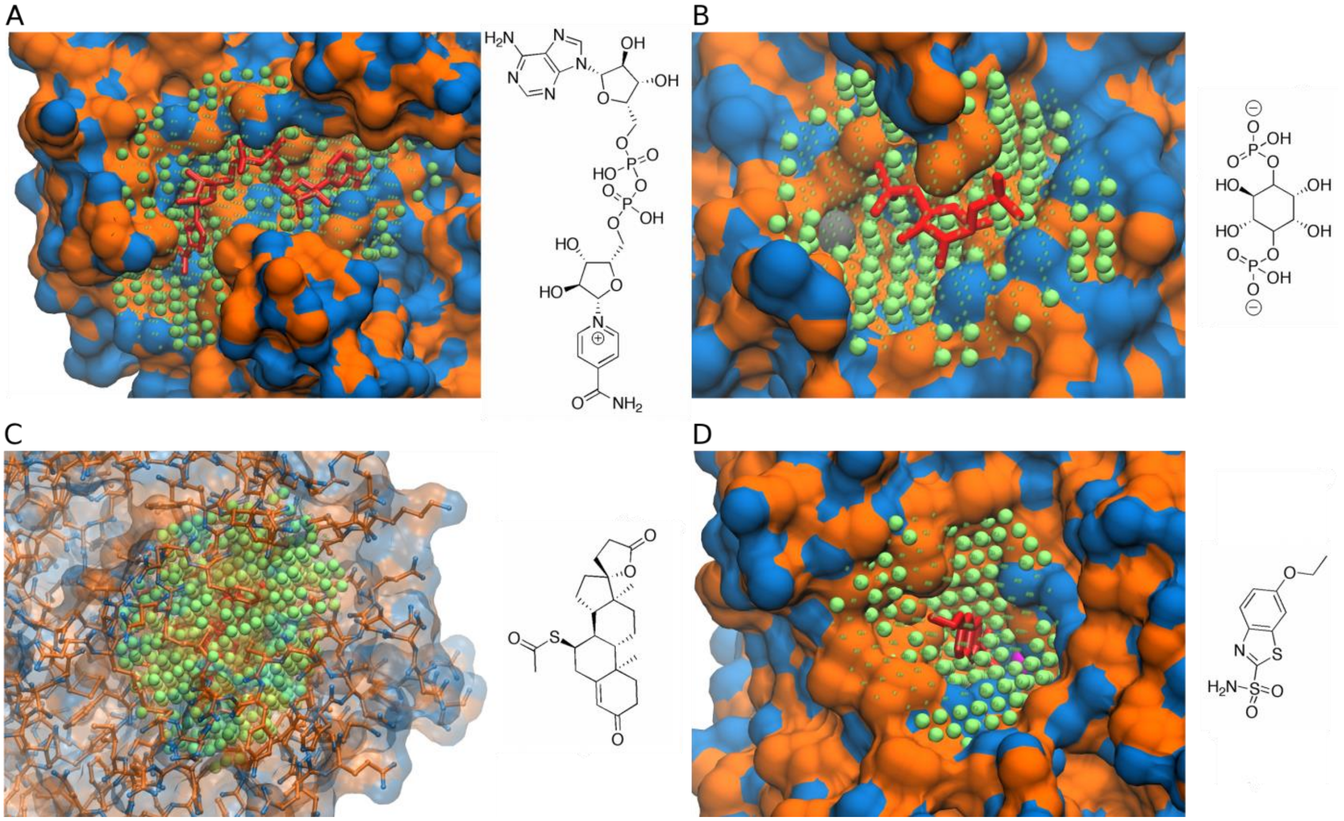


Figure 4. The relationship between JEDI druggability scores, binding site descriptors and ligand structures. A) Malate dehydrogenase is a nondruggable target predicted to have an intermediate druggability score. It is in complex here with the coenzyme NAD (PDB 1BMD). B) IP phosphatase is a nondruggable binding site that is predicted to have a low druggability score. It is here in complex with inositol(1,4)-bisphosphate (PDB 1I9Z). C) Mineralocorticoid receptor is a druggable target that is predicted to have a high druggability score. It is here in complex with spironolactone (PDB 2OAX). D) Carbonic anhydrase II is a druggable target that is predicted to have a low druggability score. It is here in complex with ethoxzolamide (PDB 3CAJ). The protein surface has been colored according to polar (blue) and apolar (orange) atoms. The 3D ligand conformations are represented in red licorice. Green dots symbolize grid points, and grid points with activity values $a_i > 0$ are depicted with small spheres. Calcium and zinc ions are respectively represented as grey and pink Van der Waals spheres. Pictures were prepared using the software VMD.[72]

Table 3. JEDI descriptor values for the structures depicted in Figure 5

Protein	V / Å ^{3}	H_a	$JEDI_{score}$	DCD_{score}
Malate dehydrogenase	157	0.64	5. 1	1
IP phosphatase	34	0.57	0.7	1
Mineralocorticoid receptor	236	0.80	9.7	10
Carbonic anhydrase II	85	0.76	6.6	10

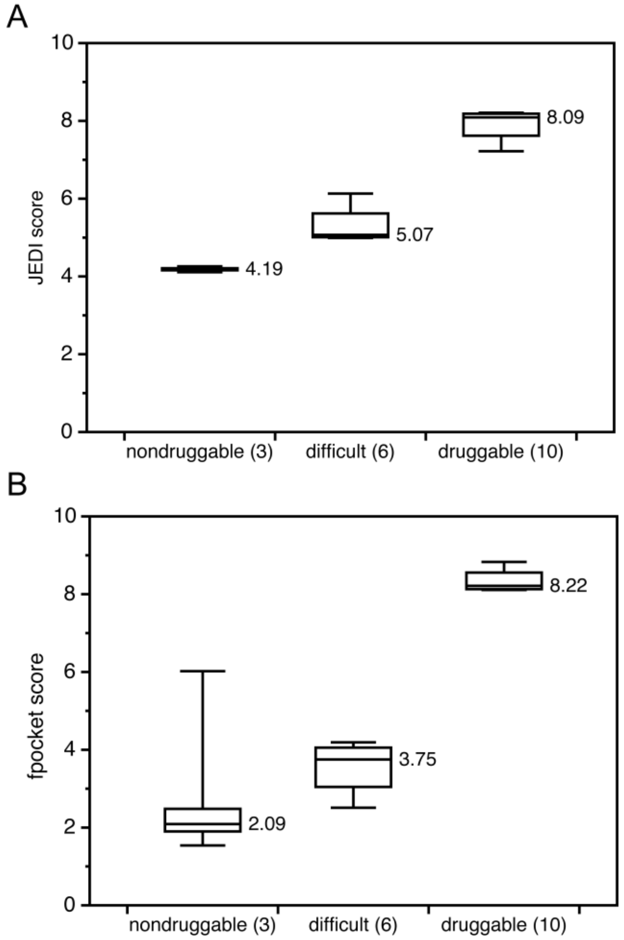


Figure 5. The sensitivity of druggability scores to small structural differences. The boxplots illustrate the fluctuations of the (A) JEDI and (B) fpocket druggability scores obtained from several highly similar conformations of a binding site for three different proteins. The DCD druggability score of each protein is given in parenthesis in the *x*-axis. The nondruggable, difficult and druggable systems selected for druggability assessment were respectively the dUTPase (PDB codes 1DUD, 1RN8, 1RNJ, 1SEH, 1SYL, 2HR6, 2HRM), the Kringle 1 domain of human plasminogen (PDB codes 1CEA, 1CEB, 2PK4, 1HPK) and the human sex hormone-binding globulin (PDB codes 1LHN, 1LHU, 1LHV, 1LHW). For the sake of consistency, only protein structures presenting a binding site identified by fpocket were selected.

Table 4. JEDI performance in ns/day for VHL (2278 atoms) and hPNMT (4057 atoms). The results were obtained using a cut-off of 20 Å for the neighbor list, and 100 ps simulations on an Intel Xeon E3-1270 v3 (3.5GHz) processor.

System	Number of processors	MD	MD/JEDI monitor mode	MD/JEDI bias mode
VHL	1	1.3	1.3	0.9
	2	2.5	2.5	1.1
	4	3.1	3.0	1.1
hPNMT	1	0.5	0.5	0.4
	2	0.8	0.8	0.5
	4	1.6	1.3	0.6

Table 5. JEDI descriptor values for the hidden pocket dataset.

Ligand	Protein	PDB code	JEDI score	V /A^3	H_a	$V_{druglike}$
1	hPNMT	1HNN	8.4	259	0.72	1.0
2		2G8N	9.2	276	0.74	1.0
3	VHL	3ZTD	8.2	118	0.80	1.0
4		3ZTC	8.5	114	0.82	1.0
5	PLK-1	2OWB	8.9	247	0.74	1.0
6		3DB6	8.1	223	0.72	1.0
7	PSMA	3IWW	0.0	493	0.54	0.0
8		2XEG	4.7	341	0.55	0.8
-	HIV-RT	1DLO	8.5	192	0.76	1.0
9		3M8P	9.6	213	0.78	1.0
-	IL-2	1M47	7.3	77	0.80	1.0
10		1M48	6.2	78	0.75	1.0

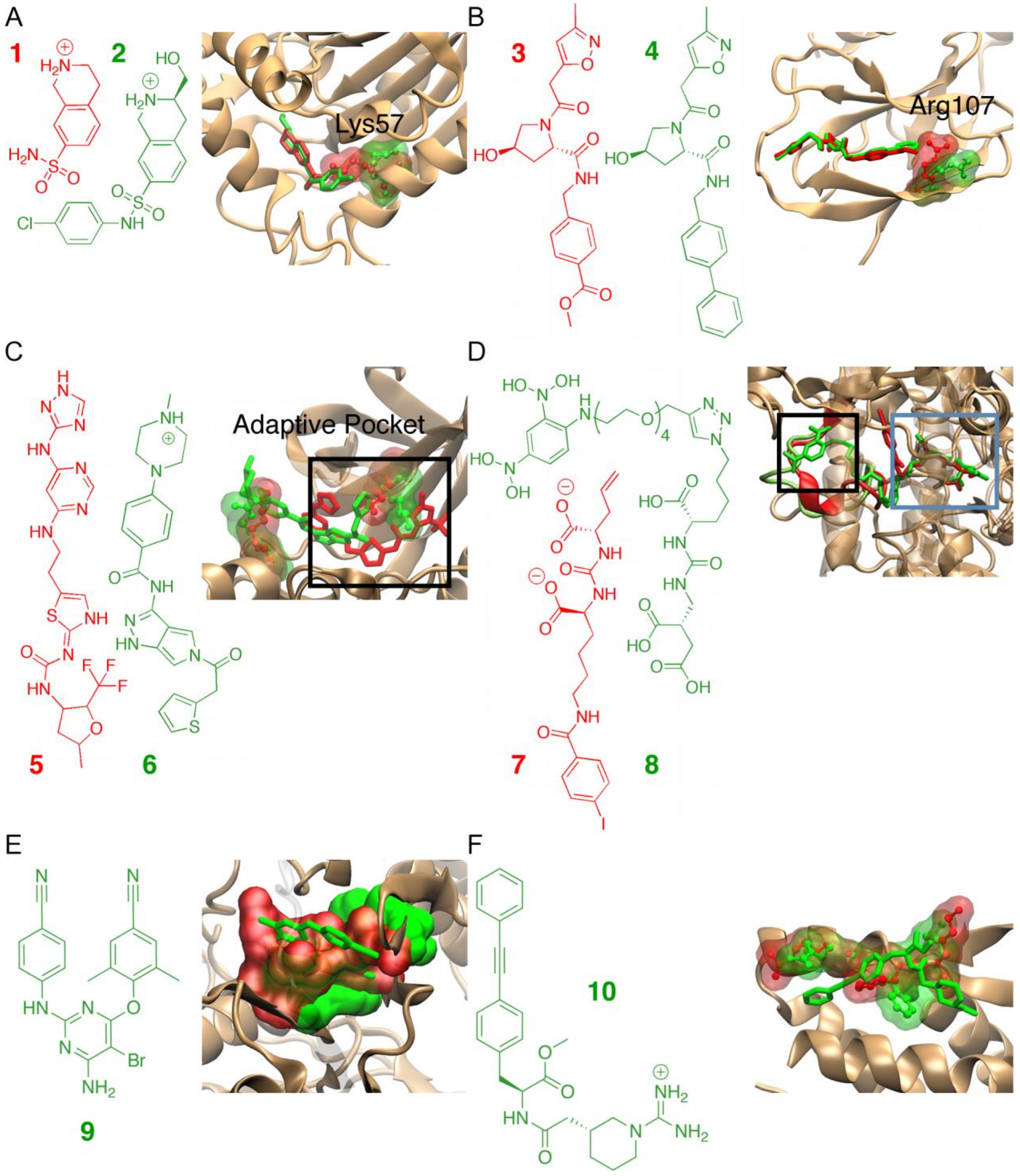


Figure 6. Conformational variability in the hidden pocket dataset. (A) hPNMT in complex with **1** or **2**, (B) VHL in complex with **3** or **4**, (C) PLK-1 in complex with **5** or **6**, (D) PSMA in complex with **7** or **8**, (E) HIV-1 in complex with **9**, (F) IL-2 in complex with **10**. Protein regions that are similar in both conformations are represented in brown. 3D structures of the ligands are displayed in licorice. Pictures were prepared using the software VMD.[72]

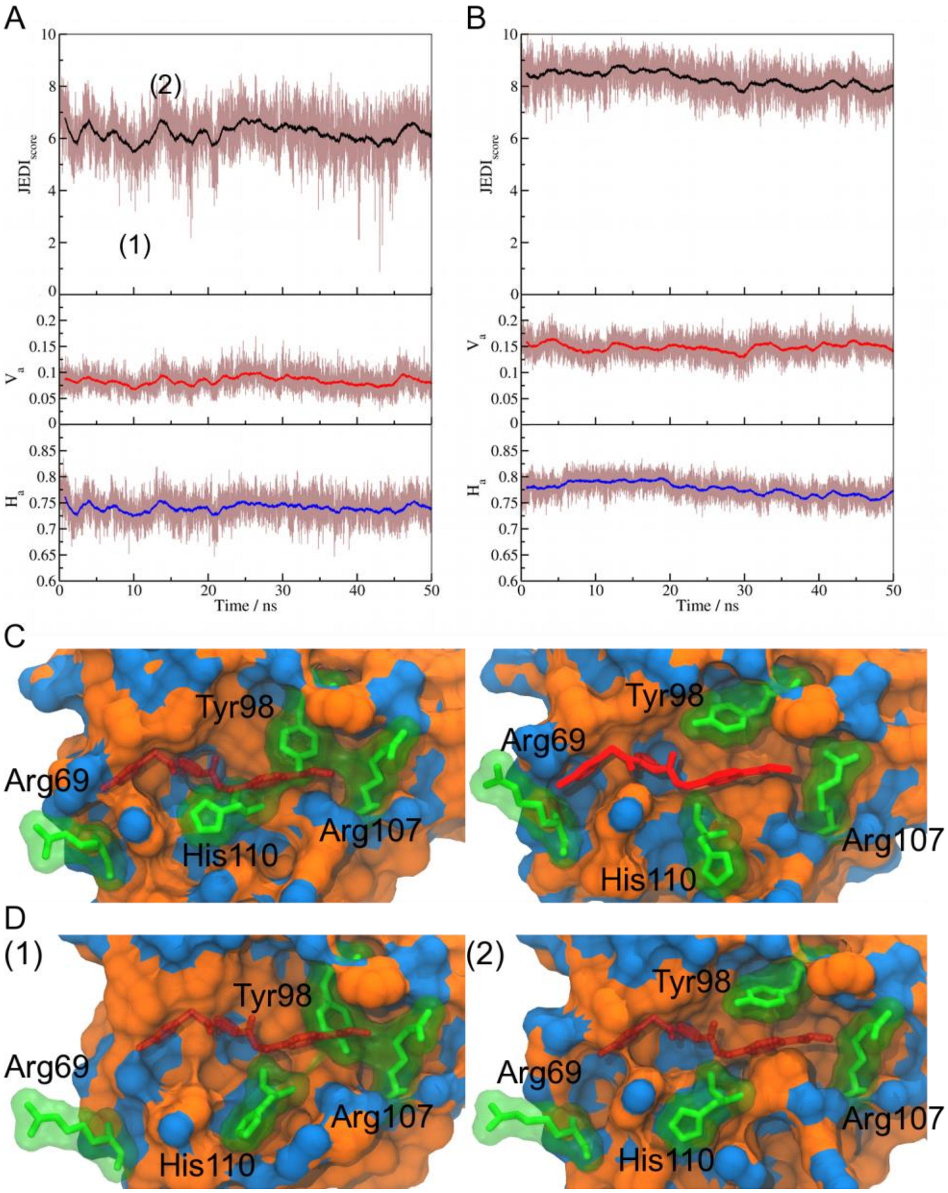


Figure 7. Druggability fluctuations during an MD simulations of apo VHL. Instantaneous values (thin lines) and 300 ps windowed averages (bold lines) of $JEDI_{score}$, V_a and H_a during an MD simulation are represented in black, red and blue respectively for (A) apo VHL and (B) VHL/**3**. (C) The most representative conformation of apo VHL (left) and VHL/**3** (right). (D) Representation of snapshots sampled at low (1) and high (2) druggability values in the apo VHL simulation (panel A). The protein surface was colored according to polar (blue) and apolar atoms (orange). Protein residues discussed in the text are highlighted in green sticks. The ligand is represented in red sticks. The crystallographic binding mode of the ligand is shown in transparent red. Pictures were prepared using the software VMD.[72]

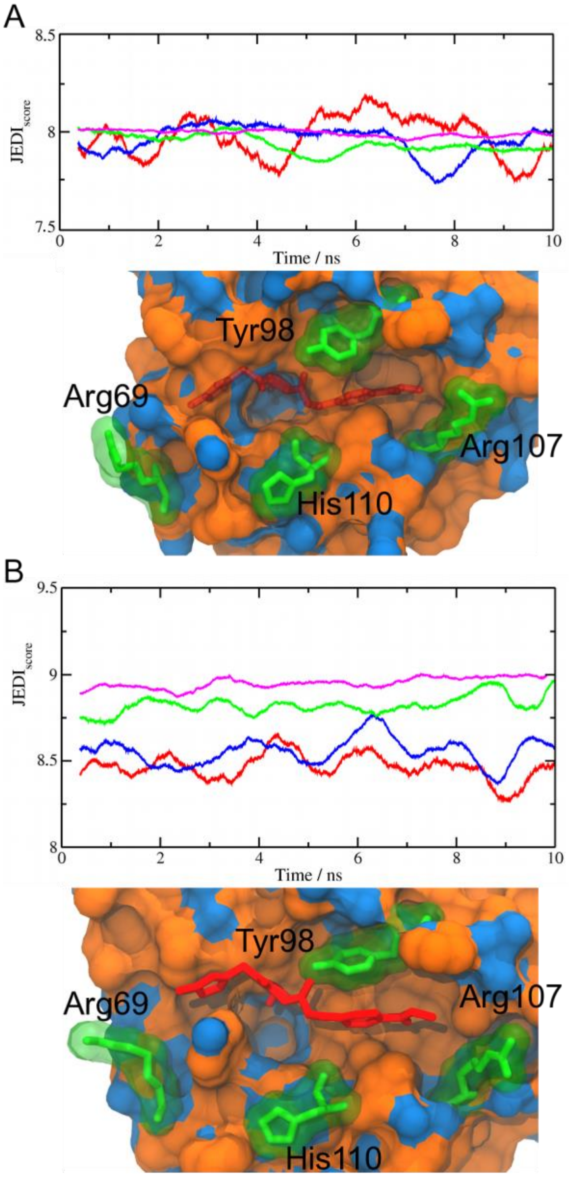


Figure 8. Druggability fluctuations during umbrella sampling simulations of (A) apo VHL and (B) VHL/**3**. For clarity, only the running averages are shown for four different spring constants (red: κ = 500 kJ.mol^{-1}.nm^{-2}, blue: κ = 1000 kJ.mol^{-1}.nm^{-2}, green: κ = 2000 kJ.mol^{-1}.nm^{-2}, magenta: κ = 5000 kJ.mol^{-1}.nm^{-2}). An illustration of the most populated cluster from the simulation performed with κ = 2000 kJ.mol^{-1}.nm^{-2} is depicted below each graph. All other symbols and representations are as in Figure 7.

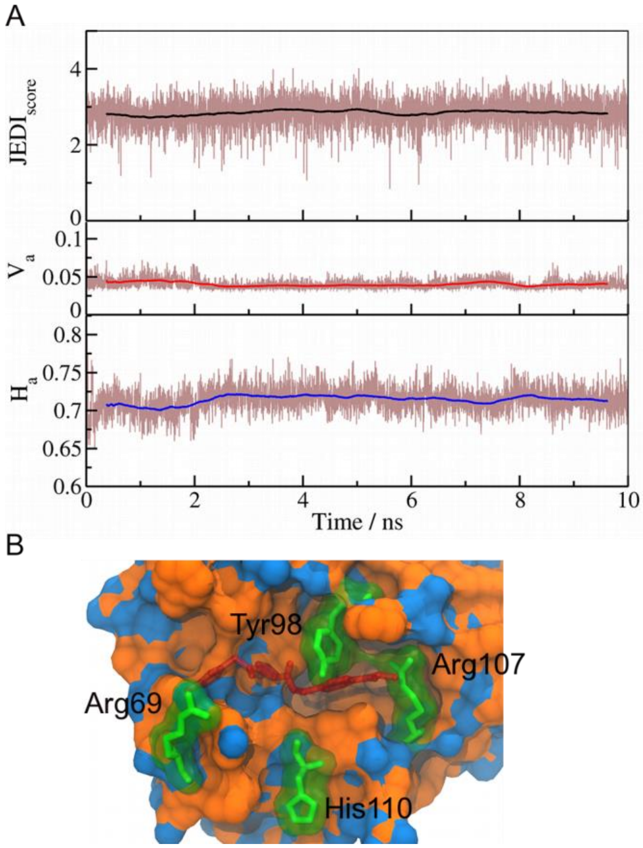


Figure 9. Druggability fluctuations during a biased simulation of apo VHL with s_0 = 3, κ = 2000 kJ.mol^{-1}.nm^{-2}. (A) Instantaneous values and running averages of $JEDI_{score}$, V_a and H_a. (B) Representative conformation of the most populated cluster identified in the simulation. All other symbols and representations are as in Figure 7.

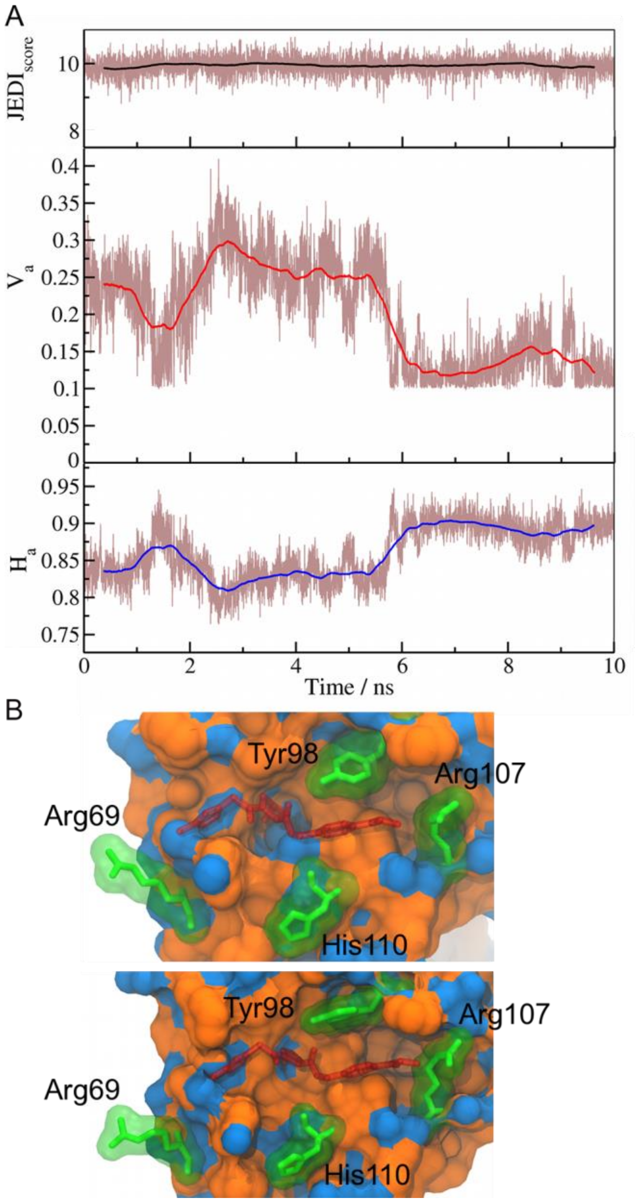


Figure 10. Druggability fluctuations in apo VHL umbrella sampling simulation with s_0 = 10 and κ = 2000 kJ.mol^{-1}.nm^{-2}. (A) Running averages and instantaneous values of of *JEDI$_{score}$*, V_a and H_a, (B) The most representative conformation of the first (top) and second (bottom) most populated clusters observed during the simulation. All other symbols and representations are as in Figure 7.

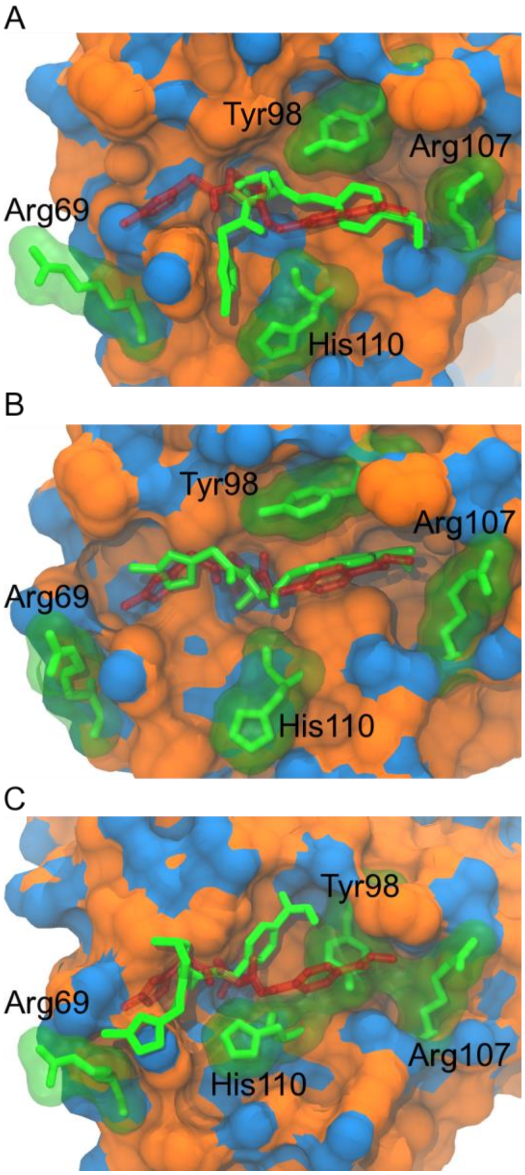
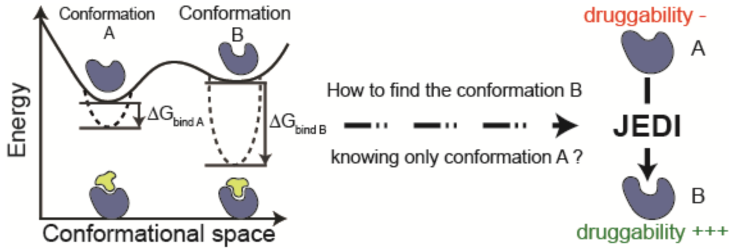


Figure 11. Ligand docking in JEDI computed VHL conformations. (A) Pose of **3** (green sticks) presenting the lowest RMSD with the ligand in the crystallographic structure, docked in the computed apo VHL conformation closest to the average conformation of the most populated cluster from an umbrella sampling simulation with s_0 = 10.0 and κ = 2000 kJ.mol^{-1}.nm^{-2}. (B) Same as (A) but with a receptor conformation manually selected from the most populated cluster. (C) Same as (A) but docked in the computed apo VHL conformation closest to the average conformation of the most populated cluster from an unbiased MD simulation. Results obtained using Vina.[37] The crystallographic pose is in red sticks. All other symbols and representations are as in Figure 7.

Graphical TOC


Conformation A
Conformation B
Energy
$\Delta G_{\text{bind A}}$
$\Delta G_{\text{bind B}}$
Conformational space
How to find the conformation B
knowing only conformation A ?
druggability -
A
JEDI
B
druggability +++