Drug interaction profiles classification: Sørensen–Dice coefficient based on MM/GBSA energies to explore focal adhesion kinase inhibitors

Patricia A. Quispe,[†] Leandro Martínez Heredia,[†] Ignacio E. León,[†,‡] and Martin J. Lavecchia[*,†]

[†]*CEQUINOR (UNLP-CONICET, CCT La Plata, associated with CIC PBA), Departamento de Química, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, La Plata, Argentina*

[‡]*Cátedra de Fisiopatología, Departamento de Ciencias Biológicas, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, La Plata, Argentina*

E-mail: lavecchia@quimica.unlp.edu.ar

Abstract

Quantitative comparison of ligand-target interaction profiles is often challenging due to subjective interpretations in visual inspection or limitation to predefined interactions in analytical methods. Moreover, traditional analyses tend to overlook the dynamic nature of the interactions, focusing instead on mean energy values. By describing ligand-residue interactions as energy distributions, we account for these dynamics. Assuming a Gaussian distribution, the mean and standard deviation are sampled from molecular dynamics simulations. Using the Sørensen-Dice similarity index, we constructed a metric where the overlap between Gaussian interactions quantifies their

similarity. For ligand comparison, the average per-residue similarity is employed. This method allows for post-processing techniques like clustering and dimensionality reduction. We applied it to the focal adhesion kinase (FAK) type II inhibitors targeting the ATP binding site, identifying interaction patterns among the inhibitors.

Introduction

In drug discovery, it is crucial to understand the interaction between a small molecule and its target to identify interaction patterns, which dictate the efficacy and specificity of a drug and can contribute to a better design and optimization of potential drug leads. Experimentally, the determination of ligand-protein interactions can be challenging due to the complexity of the procedures and the high cost associated with them, among other factors. In this context, computer-aided drug design (CADD) plays a central role, as it employs computational tools to relate a drug's structure to its biological activity. These methodologies are often divided into two categories: ligand-based and structure-based. The latter includes molecular docking, molecular dynamics simulations (MD),[1] and molecular mechanics in combination with Poisson-Boltzmann or generalized Born surface area methods (MM/PBSA and MM/GBSA, respectively). Molecular docking is employed to propose molecular complexes between ligands and a biological target, while MD is used to simulate the time evolution of the system and to sample the conformations of each ligand within the target.[2] From the sampled conformations, MM/PBSA or MM/GBSA are used to estimate the binding free energy between the ligands and the target and to decompose the binding energy into separate contributions from each part of the target, like amino acids or nucleotides in the case of proteins and DNA, respectively.

Interaction fingerprints (IFPs) are used to condense the information from the ligand binding. The most common IFPs are one-dimensional vectors with binary values, where each component encodes the presence (1) or absence (0) of a predetermined interaction, such as hydrogen bonds (HBs), hydrophobic contacts, cation-π interactions, etc., for each residue in a

particular target. In recent years, IFPs have gained relevance as inputs for machine learning models or when used to calculate the likeness between ligand pairs employing similarity coefficients.[3] The IFPs used in these comparisons are usually based on interactions derived from predefined molecular patterns (e.g., the use of pairwise distances, angles, dihedrals angles, and atom types to identify hydrogen bonds, salt bridges, hydrophobic contacts, etc). These IFPs are restricted by the predefined interactions recognized and their definitions, and could be missing relevant stabilizing interactions between the ligand and the target. Moreover, these approaches are usually made over a particular ligand-target conformation and do not contemplate the dynamic nature of the binding complex.

Here, we propose a comparison based on fingerprints derived from ligand-residue interaction energies to provide more generality than an analysis based on predefined types of interactions. The interaction energies are extracted from a MD trajectory previously analyzed using the MM/GBSA method and condensed into an energy distribution so that energetic fluctuations caused by the dynamic binding are taken into account. The obtained fingerprint is then a vector composed of energy distributions for each ligand-residue pair. Finally, we propose a new similarity index, denominated SDC (vide infra) to perform ligands comparisons taking these fluctuations into account. The Kinase domain of the focal adhesion kinase (FAK), a previously studied target in our group, was chosen as a test case to implement the similarity analysis based on energetic distributions. FAK is a tyrosine kinase protein of 125 kDa (1052 residues) that has been extensively studied as a molecular target for the development of new antitumor therapies[4] because it is upregulated in both solid and non-solid tumors and plays a crucial role in the oncogenic signaling.

To date, eight inhibitors have been tested in clinical assays: PF-562271, GSK2256098, CEP-37440, PND-1186, Defactinib, BI 853520, Conteltinib, and APG-2449. In addition, other potential inhibitors have been designed based on these compounds to improve the antitumor potency.[5] In an attempt to compare their binding interaction patterns with FAK and quantify the similarity between the interaction profiles of clinical inhibitors, we applied

the energetic distributions based on the MM/GBSA decomposition energies. In addition, ten other inhibitors, which were subjected only to preclinical testing, were included in this study.

Materials and methods

Structure preparation

The experimental human FAK structures with inhibitors selected for this study were those deposited in the Protein Data Bank[6] (PDB) with the IDs 2ETM, 2JKK, 3BZ3, 4GU6, 6I8Z, 6YOJ, 6YQ1, 6YT6, 6YVS, and 4GU6. The 4GU6 structure[7] was chosen as the target for docking and MD simulations because it is complete around the ATP binding site, although it lacks two distant residues (444-445). The missing residues were built using the MODELLER server (ModLoop).[8,9]

Molecular docking studies

Molecular docking was performed to find and score protein-ligand binding poses on FAK using Smina 1.2.0 with the Vinardo scoring function,[10] and FRED, included in the OEDocking 4.2.0 suite, OpenEye.[11]

For SMINA runs, ligand PDBQT files were generated using prepare_ligand4.py[12] and then docked using a rigid receptor approach. A docking box (27.38 Å x 25.88 Å x 28.12 Å) was defined around the ATP binding site.

To perform molecular docking with FRED, multiple ligand conformations were generated with OMEGA[13] using the "pose" mode with default settings. The target structures were converted to receptor format using the pdb2receptor program included also in the OEDocking 4.2.0 suite. FRED was run with their default settings.[11] The pose selection was made using a consensus approach[14] with Smina and FRED. In this approach, we calculated the cross root-mean-square deviation (RMSD) between the best poses and considered an RMSD value

below 2 Å as a match. To validate the protocol, the docking results were compared for those inhibitors with an experimental structure with FAK.

Molecular dynamics simulations

The best poses obtained by docking were submitted to MD simulation. The complexes have a positive net charge, consequently, chlorides were added as counterions with Leap to achieve electroneutrality. The neutralized complexes were immersed in a box of TIP3P waters which extended up to 15 Å from the solute. The ligands were described using the Generalized Amber Force Field (GAFF)[15] with charges derived from the restrained electrostatic potential (RESP) approach, which were calculated with the Antechamber module[16] and Gaussian03 D.01,[17] while the target was described using AmberFF14SB force field.[18] The topology was assigned using Leap and Antechamber, which are included in the AmberTools 21 package.[19]

All MD simulations were run using the NAMD 2.14 software.[20] The van der Waals interaction cutoff distances were set at 9 Å and long-range electrostatic forces were computed using the particle mesh Ewald (PME) summation method with a grid size set to 1.0 Å. The 1–4 contributions were multiplied by a factor of 0.83 to match the AMBER force field requirements. Prior to simulations, the systems were subjected to a series of minimization steps: first applied only to hydrogen atoms, then extended to ligand and water molecules, followed by residue side chains, and finally the entire system was minimized. MD began by heating from 0 to 300 K over 30 ps and then continued at 300K for 40 ns. The constant temperature was maintained using Langevin dynamics with a damping coefficient of 2 ps, while pressure was kept constant at 1 atm through the Nose-Hoover Langevin piston method with a damping time constant of 50 fs. A time step of 1 fs was adopted for all simulations. Bonds involving hydrogen atoms of water were constrained using the SHAKE algorithm.[21] RMSD values were depicted as a line-style plot to determine the convergence and stability of simulations (Figure SM1).

MM/GBSA calculations

Ligand-FAK binding free energies were calculated for all complexes using the MM/GBSA method, which involves calculating the difference between the bound and unbound states of the protein and ligand.[21] The solvation-free energy (ΔG_{solv}) was separated into polar and non-polar contributions. The polar contribution to ΔG_{solv} was calculated using the generalized Born (GB) model[22] implemented in MMPBSA.py module,[23] with igb 2 as the selected model. The hydrophobic contribution to ΔG_{solv} was determined using the solvent-accessible surface area (SASA). The protein–ligand binding free energy was calculated using a single trajectory (for ligand, receptor,and complex) based on 1000 snapshots taken from the last 10 ns portion (10 ps interval) of the MD simulation trajectories. Additionally, the binding energy was decomposed into ligand–residue pairs to obtain a detailed representation of the ligands/FAK interactions and create interaction profiles. The selected residues were those within 5 Å of any ligand atom. These calculations were performed using a pairwise energy decomposition scheme (idecomp option 3) also with the MMPBSA.py module.

No entropic term is calculated, as we assume that the ligands are alike, and therefore their conformational entropy is of the same order.

Geometrical interaction fingerprints

In order to perform a complementary analysis, interaction fingerprints were computed using the Python library ProLIF (Protein-Ligand Interaction Fingerprints).[24] This library uses a predefined list of interactions, which are defined as two groups of atoms (belonging to the ligand and target) that satisfy geometrical constraints based on distances and/or angles. The ProLIF library generates interaction fingerprints by calculating the presence of specific atomic interactions between the protein and ligand, such as hydrogen bonding, hydrophobic contacts, and pi-stacking interactions. These interactions are represented as a binary vector, where each element corresponds to a specific interaction type. The resulting vector is then used as a fingerprint to represent the protein-ligand complex.

ProLIF can quantify the presence of different interactions during an MD trajectory and return an occupancy metric to indicate their prevalence. This methodology was applied on the last 10 ns of the molecular dynamics trajectories (1000 snapshots). The chosen active site included amino acids within 5 Å of any ligand.

FAK inhibitors selection

The ATP competitive inhibitors selection was done based on the availability of structural information, interaction patterns, and inhibitory activity in preclinical and clinical stages (see Table 1). The molecular structures of these inhibitors are shown in Figure 1. Also, a short description of selected compounds is given below, highlighting the most interesting features of each one.

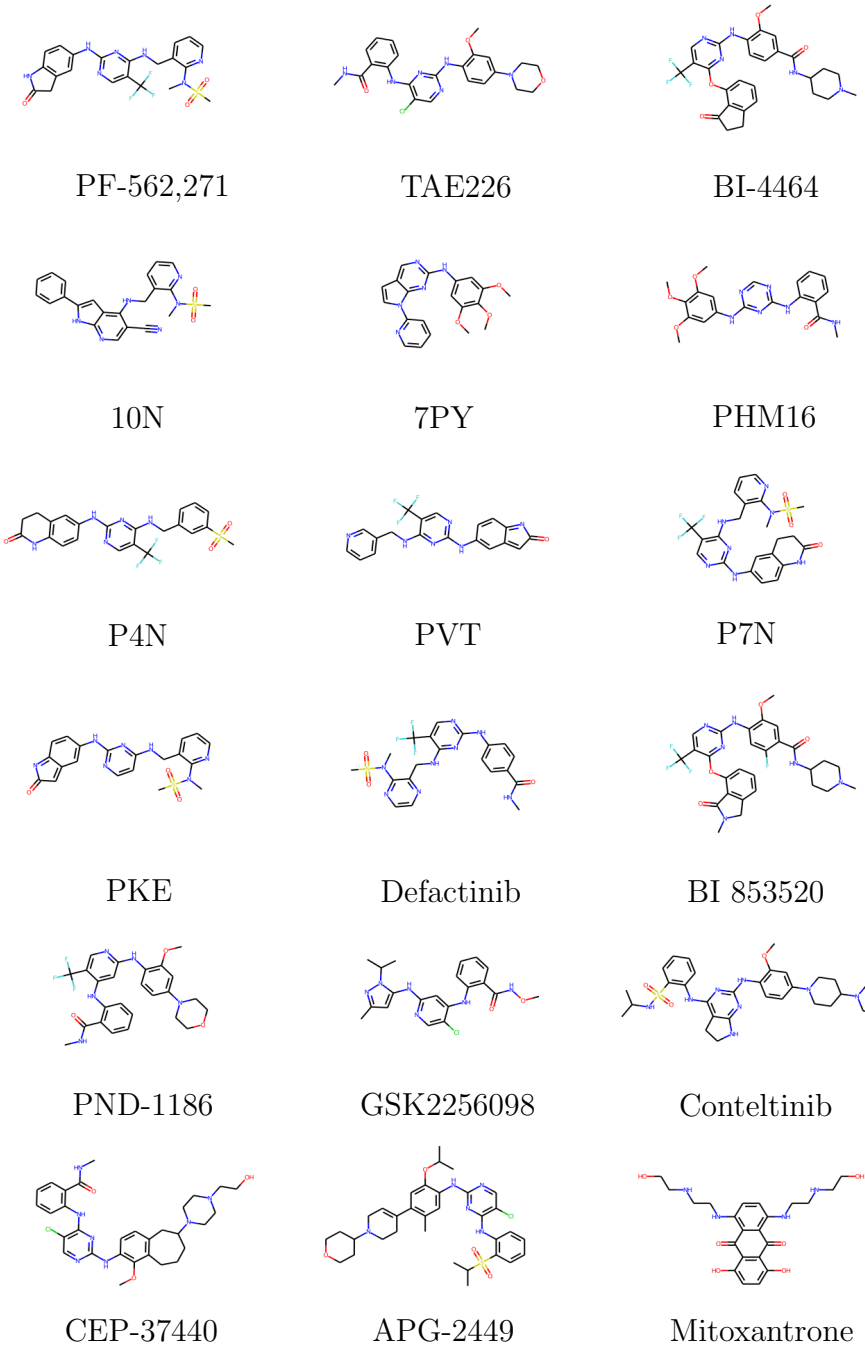


Figure 1: Molecular structure of the selected FAK kinase inhibitors.

- **TAE226** consists of a core bis-anilino pyrimidine, and it was designed by Novartis Pharmaceutical Company to bind to the ATP-binding site of FAK, although it is also able to bind to other PTKs such as Insulin-like Growth Factor 1 Receptor (IGF-

Table 1: Summary of the selected inhibitors

Name/s	FAK Inhibitory Activity (IC50)	Clinical status	PDB ID
TAE226 (NVP-TAE226)	5.5 - 6.24 nM	Preclinical	2JKK
PF-562,271	1.5 nM	Preclinical	3BZ3
PHM16	400 nM	Preclinical	4C7T
BI-4464	17 nM	Preclinical	6I8Z
Mitoxantrone	–	Preclinical	–
7PY	200 nM	Preclinical	2ETM
10N	–	Preclinical	4GU6
P4N	–	Preclinical	6YOJ
PVT	–	Preclinical	6YVS
P7N	–	Preclinical	6YQ1
PKE	–	Preclinical	6YT6
GSK2256098	1.5 nM	Clinical Phase I	–
CEP-37440	2.3 nM	Clinical Phase I	–
BI 853520	1 nM	Clinical Phase I	–
PND-1186 (VS-4718)	1.5 nM	Clinical Phase I	–
Defactinib (VS-6063)	0.6 nM	Clinical Phase II	–
Conteltinib (CT-707)	1.6 nM	Clinical Phase I	–
APG-2449	–	Clinical Phase I	–

1R), Pyk2[25] and mutant forms of Epidermal Growth Factor Receptor (EGFR).[26–28] In fact, TAE226 exhibits more affinity by the mutant form of EGFR L858R/T790M than by EGFR wild-type form.[27] The crystal structure and molecular dynamics of this compound have confirmed the stabilization in the site by hydrogen bond, electrostatic, and hydrophobic interactions.[29]

- **PHM16** is a small inhibitor in both FAK and FGFR2, and it has been co-crystallized in a complex with FAK (PDB ID 4C7T) and it exhibits a binding mode proposed similar to TAE226/FAK complex.[30]

- **BI-4464** is a highly selective ATP competitive inhibitor a close analog of the 2-aminophenyl-4-phenoxypyrimidine BI 853520, a clinical inhibitor included in the dataset selected.[31]

- **Mitoxantrone** is a dihydroxyanthraquinone antibiotic used in the treatment of acute leukemia, lymphoma, prostate, and breast cancer, but also for severe late-stage multiple

sclerosis. This compound and its derivative A18 were proposed as FAK inhibitors through the combination of molecular docking and biological assays. Mitoxantrone was included in our studies because the anthraquinone core may bind to FAK in a different way from the ones occurring with other competitive FAK inhibitors.[32]

- **7PY** and **10N** are molecules with inhibitory activity over FAK and limited experimental studies. They were selected based on differences in interaction patterns in the crystalline structure. Particularly, 7PY does not form hydrogen bonds with Asp564[33,34]

- **PF-562,271** is a methane sulfonamide diaminopyrimidine of first-generation developed by the Pfizer pharmaceutical company. Although PF-562,271 is a dual inhibitor, that is, inhibits both FAK and Pyk2, it exhibits effects four times more powerful over the former.[35]

- **P4N**, **PVT**, **P7N**, **PKE** are structural analogs of PF-562,271, designed to assess essential features of molecules to FAK inhibition. They exhibit a wide range of activities toward FAK, but they are not in preclinical development as potential FAK inhibitors.[36]

- **GSK2256098** is a FAK inhibitor developed by GlaxoSmithKline Pharmaceutical Company and tested in phase I of clinical trials. GSK2256098 appears to be more selective to FAK than to other PTKs family members. It is approximately 1000 times more selective to FAK than to Pyk2.[37]

- **CEP-37440** is a potent and highly selective dual inhibitor of both FAK and anaplastic lymphoma kinase (ALK) with IC50 of inhibition of the catalytic activity of 2.3nM and 3.5nM respectively. The binding mode of CEP-37440 to FAK and ALK was predicted by molecular docking.[38]

- **BI 853520** is a potent and highly selective FAK inhibitor with an inhibition IC50 of 1 nM over recombinant FAK.[39]

- **PND-1186** also called VS-4718 is a small substituted pyridine FAK inhibitor of second-generation developed by Verastem Inc. It has been evaluated in phase I of clinical trials alone and in combination with antitumor drugs.[40]

- **Defactinib** (VS-6063, PF-04554878) is a second-generation inhibitor of FAK and Pyk2 developed by Pfizer Pharmaceutical Company. It has been demonstrated to be a potent inhibitor of FAK and Pyk2 activity with an IC50 of 0.6 nM (in both kinases) and it is 100-times more selective to them than to other members of the Protein Kinase Family (PTKs)[39]

- **Conteltinib** (CT-707) is a small FAK inhibitor (IC50: 1.6 nM) developed by Centaurus Biopharma that recently has been tested in clinical trials. The information about this molecule is limited, although it has achieved an objective clinical response.[41]

- **APG-2449** is another FAK inhibitor and recently, the results of a clinical trial carried out over patients with solid tumors have been released, with promising results.[42]

Results

Evaluating the performance of docking methodology: Redocking of crystalized compounds

In order to determine a suitable docking methodology, we conducted a redocking analysis of the crystalized FAK inhibitors. FRED and Smina were evaluated as docking software. The re-docking was carried out over ten crystal complexes, listed in Table SM1, selected from the PDB. The inhibitor was removed from the complex, and then a docking was performed over the FAK structure from PDB 4GU6 to recover the experimental pose. The RMSD between the best docking pose and the corresponding aligned experimental structure was used to assess performance. An RMSD value lower than 2 Å is considered as a good accuracy of the docking program to reproduce the experimental structure.[14] Both FRED and Smina

reproduced accurately the crystal structures of the inhibitors, except for 7PY with Vina, which yielded an RMSD value greater than 2 Å. Based on these results, we conclude that these programs adequately generate binding poses for this system, and these calculations are extensible to ligands that lack experimental structure. As both programs have successfully reproduced the experimental poses, we have decided to combine their outputs using a pose consensus approach to select the final pose. In this approach, the top three ranking poses from the two programs are compared and the match pose proposed by Cavasotto et. al[14] between them is finally selected for subsequent studies.

Docking of non-crystallized inhibitors

For the inhibitors GSK2256098, BI 853520, CEP-37440, PND-1186, Defactinib, Conteltinib, APG-2449, and Mitoxantrone (Table 1) there are no publicly available co-crystallized complexes with FAK despite preclinical studies demonstrating their interaction at the ATP binding site, and having undergone clinical trials (except mitoxantrone). Therefore, to propose the binding conformations to FAK, we employed consensus docking, as described in the Materials and Methods section.

All complexes show a similar position and orientation within the binding site; thus in principle, they could bind in a similar fashion.

However, it's important to note that even while two molecules may seem to interact similarly, relying solely on docking is insufficient for a comprehensive assessment of the binding strength and the identification of subtle differences in their binding modes.

Molecular dynamics and energetic analysis

To evaluate the stability of the proposed binding poses and explore specific binding patterns in the active site, the FAK inhibitor complexes were submitted to the MD protocol, with 40 ns of equilibration/production. All systems stabilized within 25 ns of simulation, as evidenced by the RMSD profiles, both for ligand and protein (Figure SM1). The RMSD

values showed that the crystallized inhibitors maintained their initial position. As expected, the non-crystallized inhibitors and/or surrounding residues showed small fluctuations during the early relaxation period of the MD simulation, but all of them reached stability similar to that of the crystallized ligands (Figure SM2). This finding strongly suggests that the selected poses are representative of stable complexes.

We used MM/GBSA to calculate the binding affinities (ΔEbind) of these complexes, as previously described. According to these results, the crystallized inhibitors show a ΔEbind in the range of -38.3 to -54.3 kcal mol^{-1} (see Table SM2). The eight inhibitors without structural information have energies of the same order as the crystalized inhibitors. Although the correlation is not perfect, a trend is observed between the interaction energy presented by the ligands with FAK and their inhibition activity (see Figure SM3).

Subsequently, per-residue energy interaction profiles were obtained using the MM/GBSA method over the last 10 ns of the MD trajectories. This approach allowed us to discern the energetic contributions of active site residues. The heatmap plot (see Figure 2) shows the contribution of each residue to the total binding energy for the eighteen inhibitors. A similar heatmap plot for the standard deviations is shown in Figure SM6). A limitation of this analysis is that it does not identify the specific types of interactions between the ligands and the active site.

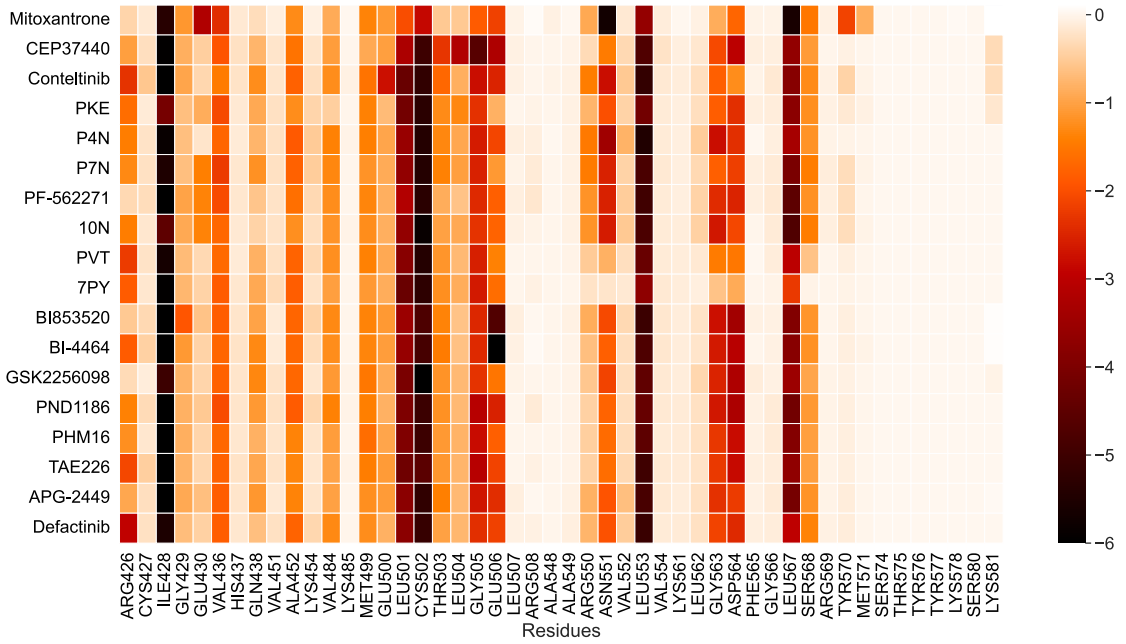


Figure 2: Heatmap of MM/GBSA mean binding energy contribution by residue for FAK inhibitors (values are in kcal mol^{-1})

MM/GBSA-based similarity quantification using the Sørensen–Dice coefficient

Ligand-protein binding is a complex process influenced by changes in the atomic dynamics of both the ligand and the protein. During a MD simulation, an ensemble of conformations is generated, each one representing a particular state of the ligand-target complex. The vast amount of information obtained is generally reduced to a single binding energy averaging over all the interaction energies using free energy estimation techniques. Or, at the very best, this mean energy is separated on contributions from each residue.

Rather than relying solely on average energies, consider including the full set of energies for each pose obtained throughout the simulation, and then describe the interaction energy as an energy distribution instead of a single value. This way, the fluctuations are (to some degree) considered in the analysis and not directly discarded. This distribution could be done on a per-residue basis, thus encoding an energy distribution for each ligand-residue

interaction, offering information about the strength of each interaction.

Additionally, to simplify the treatment, we hypothesize that for each ligand-residue interaction, the distribution of the energies during an MD simulation could be effectively represented by a Gaussian function, denoted as g(x). This distribution g(x) is defined by the mean energy (μ) and the standard deviation (σ), both obtained from the sampled energies. This way, the treatment only uses information from a MMGBSA.py run, without the need of additional runs. Alternatively, an energy distribution could be estimated from the sampled energies using, e.g., kernel density estimation (KDE), but we show below that the Gaussian assumption is sufficient and more efficient.

As a net result, our description takes into account an extra parameter, σ, to characterize the ligand-residue interactions, besides μ. This σ is related to the dispersion of energies during the simulation, and we understand it as an indicator of the lability of the interaction (Figure 3). A stable binding mode will have little movement, and therefore a small dispersion of the ligand-target interaction energies, while a more labile binding mode will produce deviations from the mean energy value as the ligand moves around the binding site.

Of course, this analysis is useful to a specific binding mode, and when multiple binding modes are possible or occur during the simulation, they should be analyzed separately.

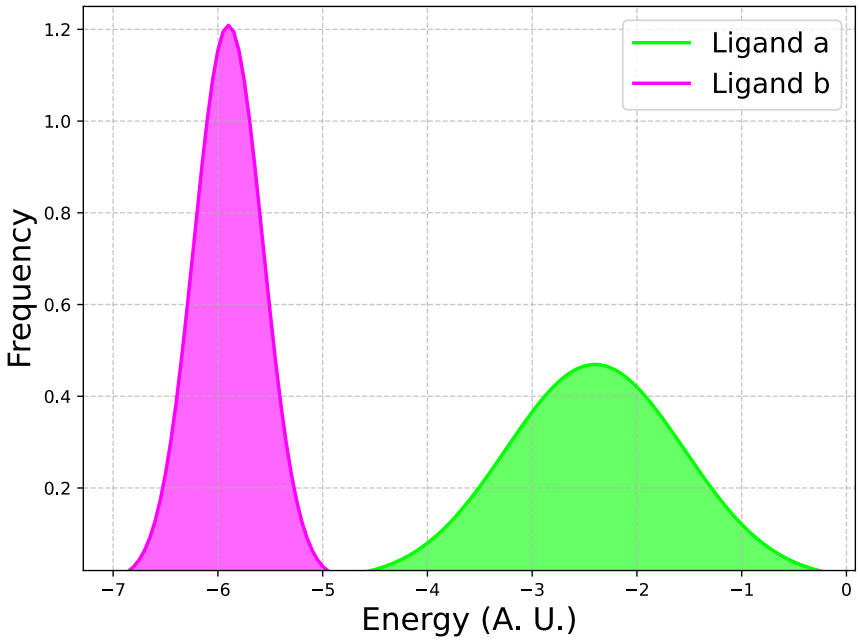


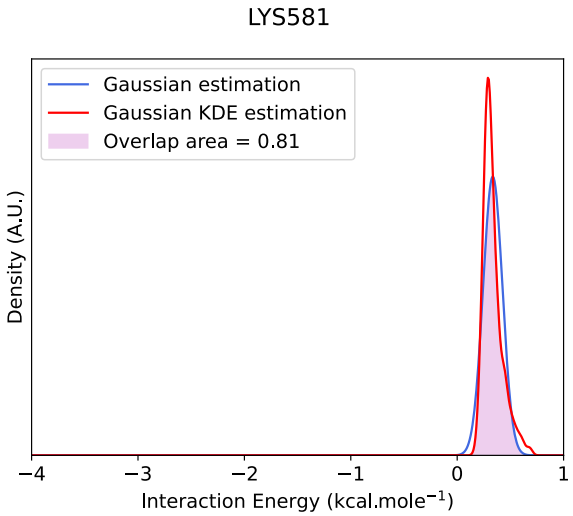
Figure 3: This scheme depicts the distribution of interaction energies of two ligands, a and b with the same residue of the target during an MD simulation. Leaving aside the absolute energy values, ligand b has a lower dispersion of energies around the mean energy value than ligand a. The b-target interaction is not easily perturbed by the movements during the MD in comparison to the a-target interaction. Therefore, we believe that the standard deviation, as a measure of dispersion around the mean value and in combination with the mean value, is valuable as an indicator of how labile an interaction is.

To test our assumptions, we first evaluated the reliability of describing the sampled ligand-residue interaction energies with a Gaussian distribution. In order to do this, we represented the original distribution of energies with a function K(x) obtained using KDE over the sampled energies. The KDE estimation was done using Gaussian Kernels as implemented in Scipy,[43] with Scott's rule. By calculating the overlap between g(x) and K(X), we get an indicator of the quality of the assumption, which we call Gaussianity. Therefore, a Gaussianity closer to 1 indicates a better description of the interaction with g(x).

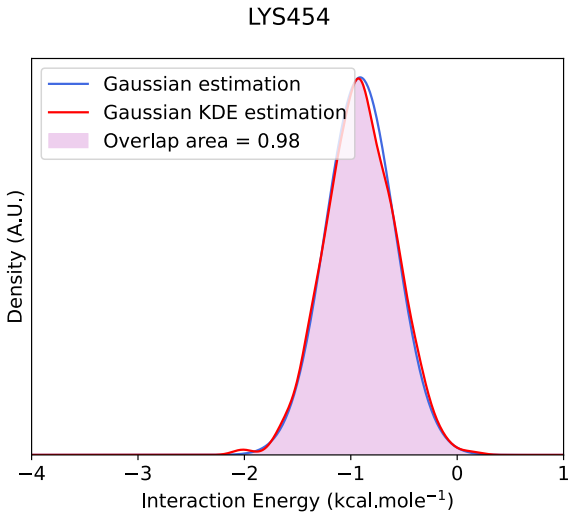
The molecule BI 853520 was randomly selected as a test case from the set, and the calculated overlap values ranged from 0.81 to 0.98 (see Figure SM4 and Table SM3). Figure 4 illustrates the comparison between the real energy distribution and its Gaussian approximation for two residues. Notably, the real energy distribution is highly similar to the Gaussian

approximation, and a pronounced overlap between the distributions is observed, highlighting a significant similarity.

In summary, we consider that using a Gaussian distribution to represent the distribution of ligand-residue interaction energies is an acceptable approximation.



(a)



(b)

Figure 4: Comparison between the Gaussian KDE estimation of the real interaction energy distribution and its Gaussian approximation for ligand BI 853520 with the residues (a) Lys581 and (b) Lys454. These residues have the greatest and smallest overlap between the KDE and Gaussian estimations (respectively) for this ligand, i.e., the gaussianity.

Then, to measure the similarity between two of these energy distributions g(x), which

represent the interaction of residue i with ligands a and b, $g(E_a^i)$ and $g(E_b^i)$, respectively, we use the superposition between the Gaussian functions (Figure 5). Since the energy distributions are normalized, intuitively, two populations (or samples) are similar when their distribution functions overlap.

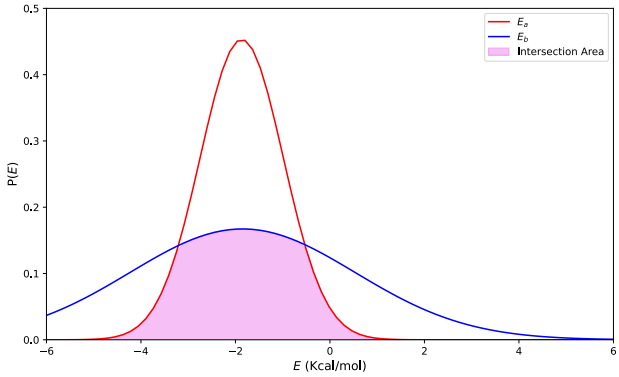


Figure 5: Overlap of two Gaussian functions, which represent the distribution of ligand-residue interaction energies for the same protein residue with different ligands a and b

As the functions $g(x)$ are normalized, this overlap is equivalent to the Sørensen-Dice Coefficient (SDC) between the distributions.[44] For two data sets, the SDC is twice the number of matching elements (A ∩ B) relative to the total number of elements (A+B) (Eq. 1). For two Gaussian distributions $g(E_a^i)$ and $g(E_b^i)$, which are normalized ($|A| = |B| = 1$), the SDC reduces to the intersection, i.e., the overlap between the distributions. The SDC coefficient ranges from 0 (no similarity) to 1 (perfect similarity).[45,46]

$$SDC(A,B) = \frac{2|A \cap B|}{|A| + |B|} \quad \therefore \quad \mathrm{SDC}_i(g(E_a^i), g(E_b^i)) = |A \cap B| = \int_{-\infty}^{\infty} g(E_a^i) \cdot g(E_b^i)\, dE \quad (1)$$

To compare the whole interaction profile of both ligands, the individual ligand-residue similarities (SDC_i) are averaged. The global similarity coefficient between two ligands, denoted as GSDC is then a value in the (0, 1) range (Eq. 2).

$$GSDC(A,B) = \frac{1}{N}\sum_{i} SDC_i(A,B), \quad i = \mathrm{Res}_1, \mathrm{Res}_2, \ldots, \mathrm{Res}_N \tag{2}$$

This GSDC serves as a quantitative indicator of the overall similarity in binding modes between the ligands across all residues, taking into account the dynamic nature of the interactions. The minimum value of $GSDC(A,B) = 0$ indicates that $g(E_a^i)$ and $g(E_b^i)$ are distinct for all residues, suggesting a low degree of similarity between the binding modes of the ligands with the target. On the other hand, a maximum value of $GSDC(A,B) = 1$ represents a perfect matching in binding modes between ligands, indicating identical behavior.

The SDC exhibits heightened sensitivity, especially when the standard deviation is low **(Figure 6)**. A smaller σ leads to a more sharply defined and concentrated Gaussian distribution around the mean energy value. The overlapping area may introduce errors, prompting us to establish a lower limit for σ at 0.3 units to mitigate potential calculation errors in the overlapping region.

The coefficient was calculated using a custom Python script developed in-house, available at`https://gitlab.com/gqc/gsdc_md`.

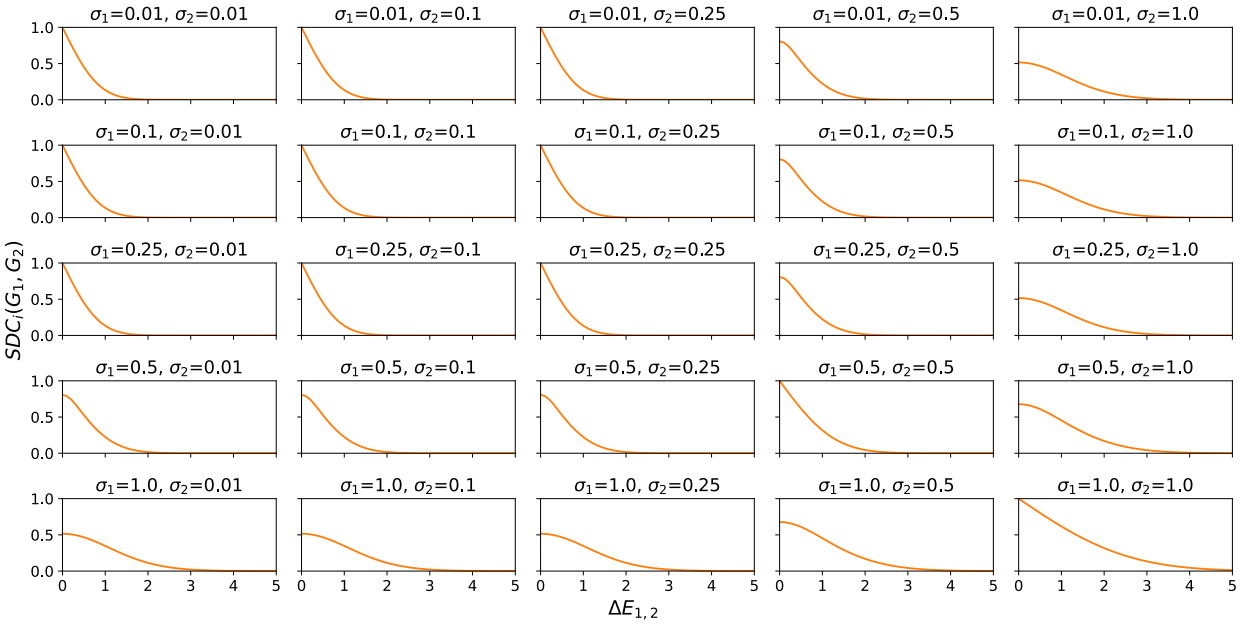


Figure 6: In each graph the SDC_i is calculated for a pair of Gaussian functions g_1 and g_2, with standard deviations σ_1 and σ_2, with means separated by $\Delta E_{1,2}$ units.

Finally, the selection of the Gaussian function to describe the energy distributions was suitable for the real behavior. Overall, these observations underscore the importance of carefully adjusting parameters to optimize the accuracy and reliability of the SDC in distribution analysis.

Use of the GSDC coefficient to cluster FAK inhibitors binding modes

Assessing the similarity of binding modes between a pair of ligands is a crucial aspect of the GSDC. Instead of simply comparing the structures of ligands, the GSDC delves into studying how these ligands bind to the receptor. In the context of FAK ligands, the GSDC was cross-calculated, obtaining values in the 0.62 to 0.90 range (see Figure 7). In specific instances, molecules with significantly different binding modes, as illustrated by Mitoxantrone, can be easily discerned, showcasing low similarity and a considerable distance from the overall group.

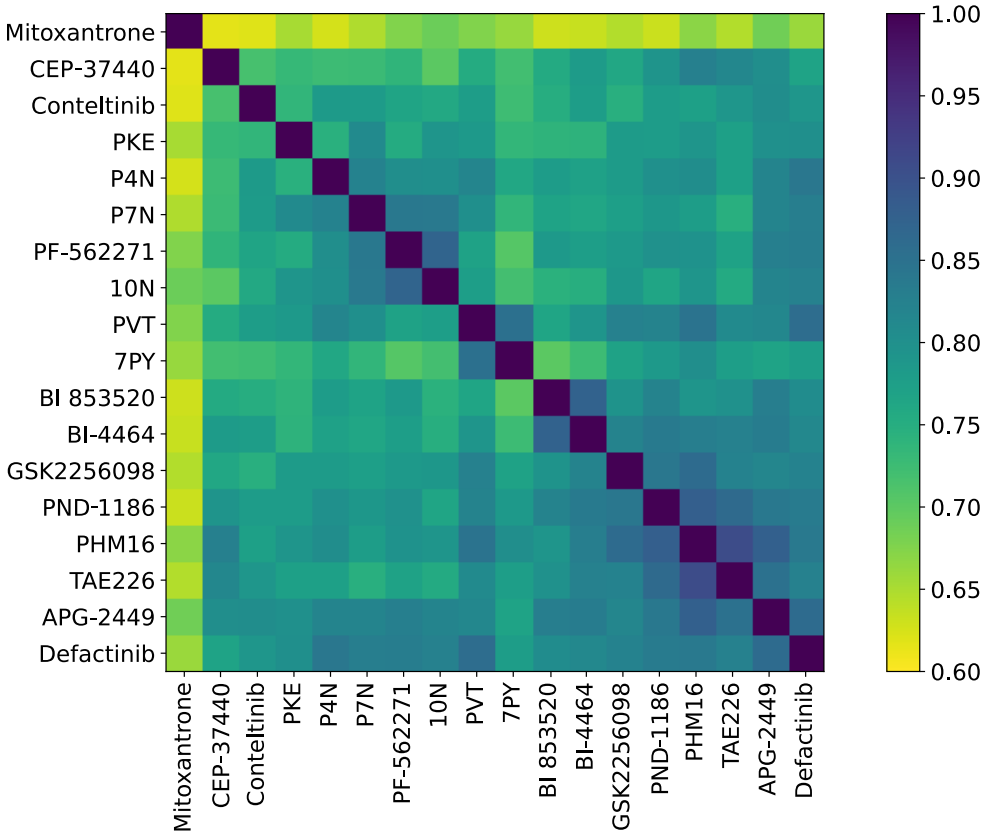


Figure 7: Global Sørensen–Dice coefficient values for FAK inhibitors. For GSDC values see Figure SM5.

As with other similarity measures, the GSDC determines amongst a group of objects which pairs are the closest, enabling the grouping or clustering of the data and easing the visualization and analysis of relationships within the data. To this end, we employed hierarchical clustering and dimensionality reduction methods.

In hierarchical clustering, each compound starts as its cluster and then the most similar clusters are merged until all compounds end up in the same and only cluster. In each step, the clusters with the closest points are merged (Figure 8.a).

In addition, we applied the t-distributed Stochastic Neighbor Embedding (t-SNE) and Multidimensional Scaling (MDS) dimensionality reduction algorithms to aid in the depiction of the data points in a two-dimensional space (Figure 8 b-c). These visualizations enhance the intuitiveness and readability of the relationships between inhibitors, but, as a lot of

information is lost in the projection, it should be used with care. In this regard, MDS focuses on the preservation of distances, losing some of the local information, while t-SNE tends to aggregate the closest points, with more inexact distances.

The compounds of each cluster share binding features that are responsible for differential interaction modes concerning other clusters. In our particular case, MDS and hierarchical clustering proved successful in isolating compounds with a low degree of similarity as Mitoxantrone, however, t-SNE was unable to maintain a significant distance between this compound and the others.

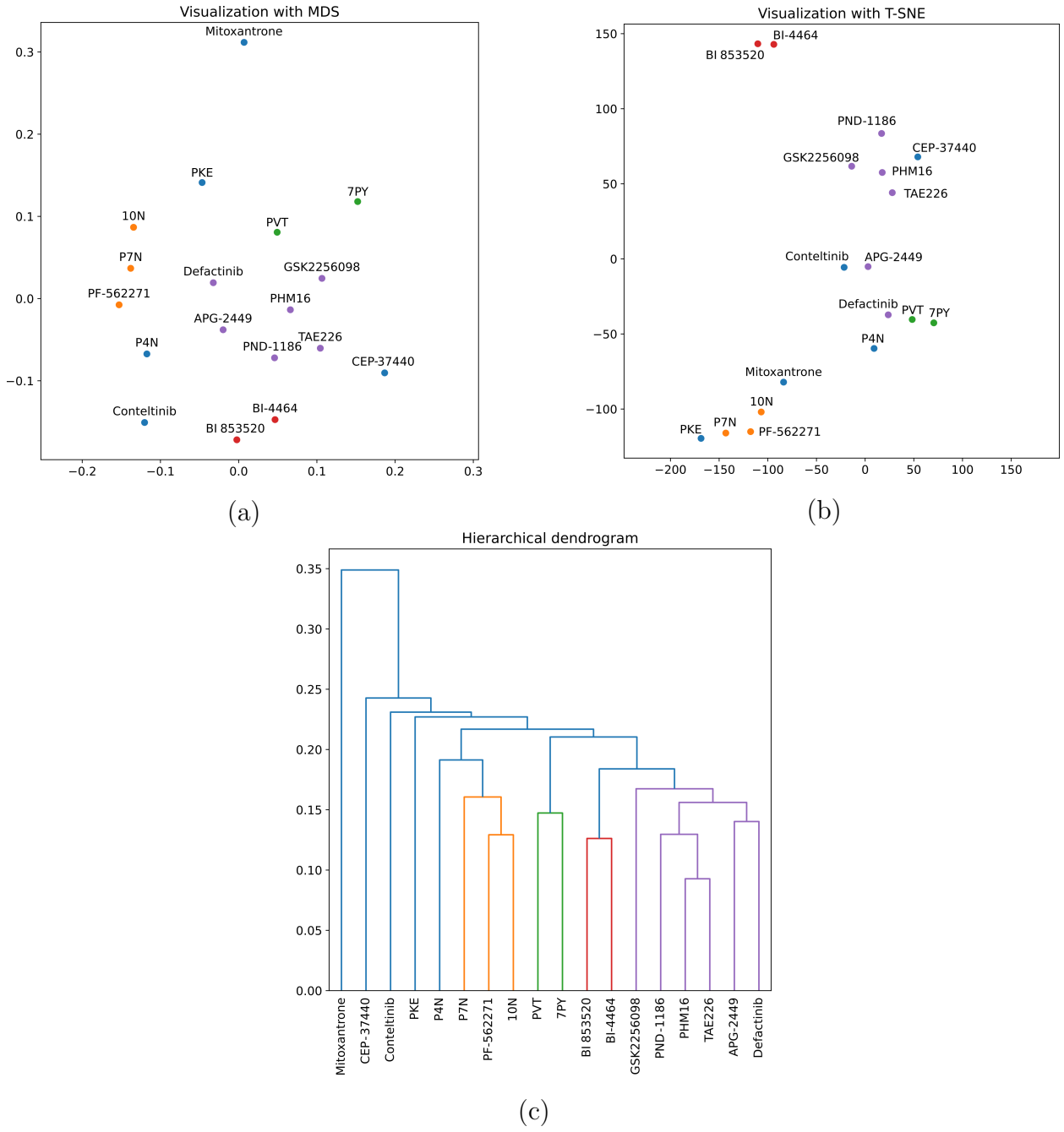


Figure 8: Comparison between the Gaussian KDE estimation of the real interaction energy distribution and its Gaussian approximation for ligand BI 853520 with the residues (a) Lys581 and (b) Lys454. These residues have the greatest and smallest overlap between the KDE and Gaussian estimations (respectively) for this ligand, i.e., the gaussianity.

To quantify the distinctiveness of a specific ligand-residue interaction, we define the per-residue SDC for ligand A and residue i ($RSDC_i(A)$) as the mean similarity of a residue

interaction to the interactions of all the other ligands with the same residue.

$$RSDC_i(A) = \frac{1}{N} \sum_{B \neq A} SDC_i(A, B) \tag{3}$$

It is a useful indicator of a characteristic interaction for a particular ligand (all SDC_i will be close to 0, and so the $RSDC_i$) and common interactions for a group of ligands (all SDC_i will be close to 1, and so the $RSDC_i$). The $RSDC_i$ is dependent on the ligands included in the analysis.

We calculated the RSDC_i for all the ligands and the selected residues (Figure 9). The RSDC_i sheds light on hierarchical clustering, identifying the particular interactions of cluster outliers, and common interactions within clusters. Residues His437, Val451, Lys485, Leu507-Ala549, Val554-Leu562, and Ser574-Ser580 were identified as regions where interactions are similar for all or most inhibitors. Regarding clusters (see Figure 8), Cluster O has particularly different interactions with residue Glu430, meanwhile, cluster R does it with residues Ile428, Gly429, and Glu506. Members of cluster V exhibit a differential interaction with Tyr577, a key residue within the activation loop that undergoes phosphorylation during the activation process of FAK. Unclustered ligands each have particular interactions that are not shared with any other ligand, such as Glu500 for Conteltinib or Val484 for PKE.

7PY and Mitoxantrone show markedly different interaction patterns from other inhibitors due to their distinct chemical features. 7PY shows variations in the Leu562-Ser568 region that comprises the DFG-out motif and Mitoxantrone has few common interaction points with the other inhibitors.

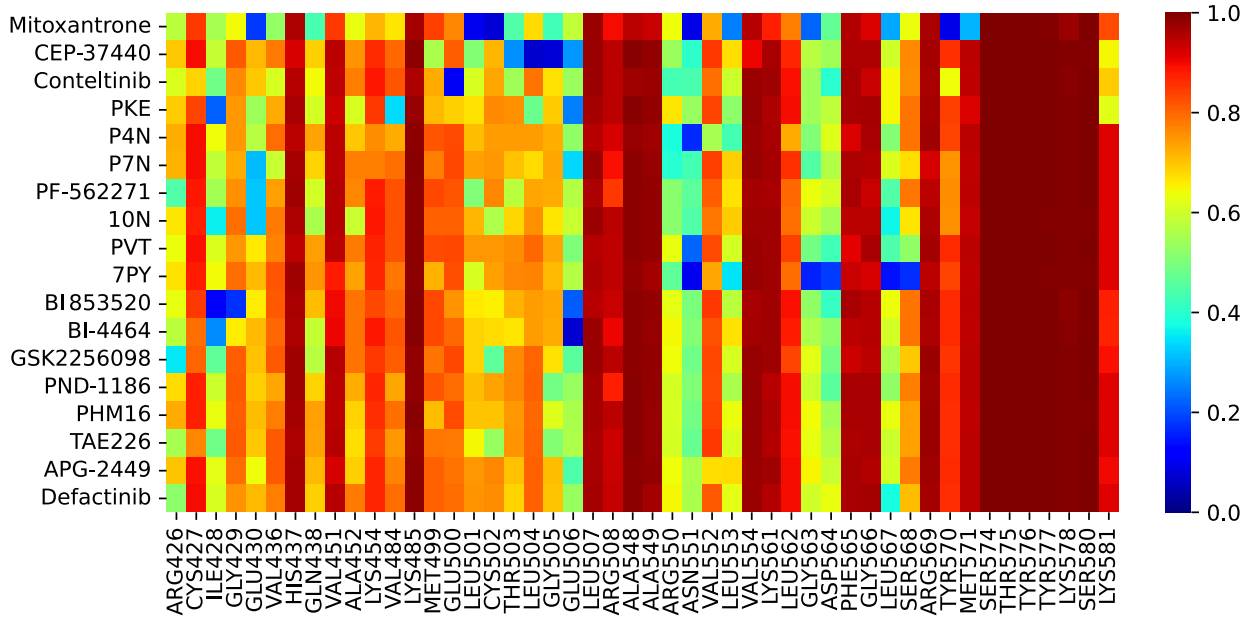


Figure 9: Per-residue $RSDC_i$ for all the ligands investigated. Regions with common interactions and regions where no interaction is similar can be discerned. Furthermore, distinctive interactions with residues for each ligand can be identified in blue, and, for some clusters similar characteristic interactions are found.

As demonstrated, the SDC can be employed to analyze and compare the binding modes of FAK inhibitors. Consequently, the visual analysis with the dimensionality reduction methods should be complemented with the numerical analysis of a similarity matrix for an optimal interpretation of clusters.

Geometric interaction analysis with ProLIF

A shortcoming of the purely energetic analysis is a lack of simple interpretability. Even if we separate MM/GBSA interaction energy into their individual components (electrostatic, van der Waals, polar and non-polar solvation), it's hard to understand and interpret the nature of the interaction taking place. For example, a highly electrostatic interaction could be due to an HB, salt bridge, cation-π interaction, etc. To this end, we complement the energetic analysis with a geometric analysis to interpret the obtained results.

We used ProLIF to perform a geometric analysis through the same MD trajectories.

In the ProLIF analysis, each frame is analyzed to identify predefined interactions and an occupancy is obtained for each interaction found. The occupancy represents the fraction of the analyzed frames in which a particular interaction is present, with values in the 0 to 1 range. The occupancy value of an interaction type serves as a measure of the residue's behavior with the ligand during the MD simulation (see Figure 10). The inhibitors exhibited interactions with residues of the active site, including hydrophobic moieties, HB donors and acceptors, and cationic interactions. HBs with the residues Asp564 and Cys502, as well as hydrophobic contacts with Ile428, Val436, Leu501, Gly505, Glu506, Asn551, Leu553, Gly563, and Leu567 were present in most ligands/FAK complexes. These findings were consistent with previous reports, where Cys502, Asp564, Leu553, and Leu567 were identified as important interactions with inhibitors in the ATP-binding site.[36]

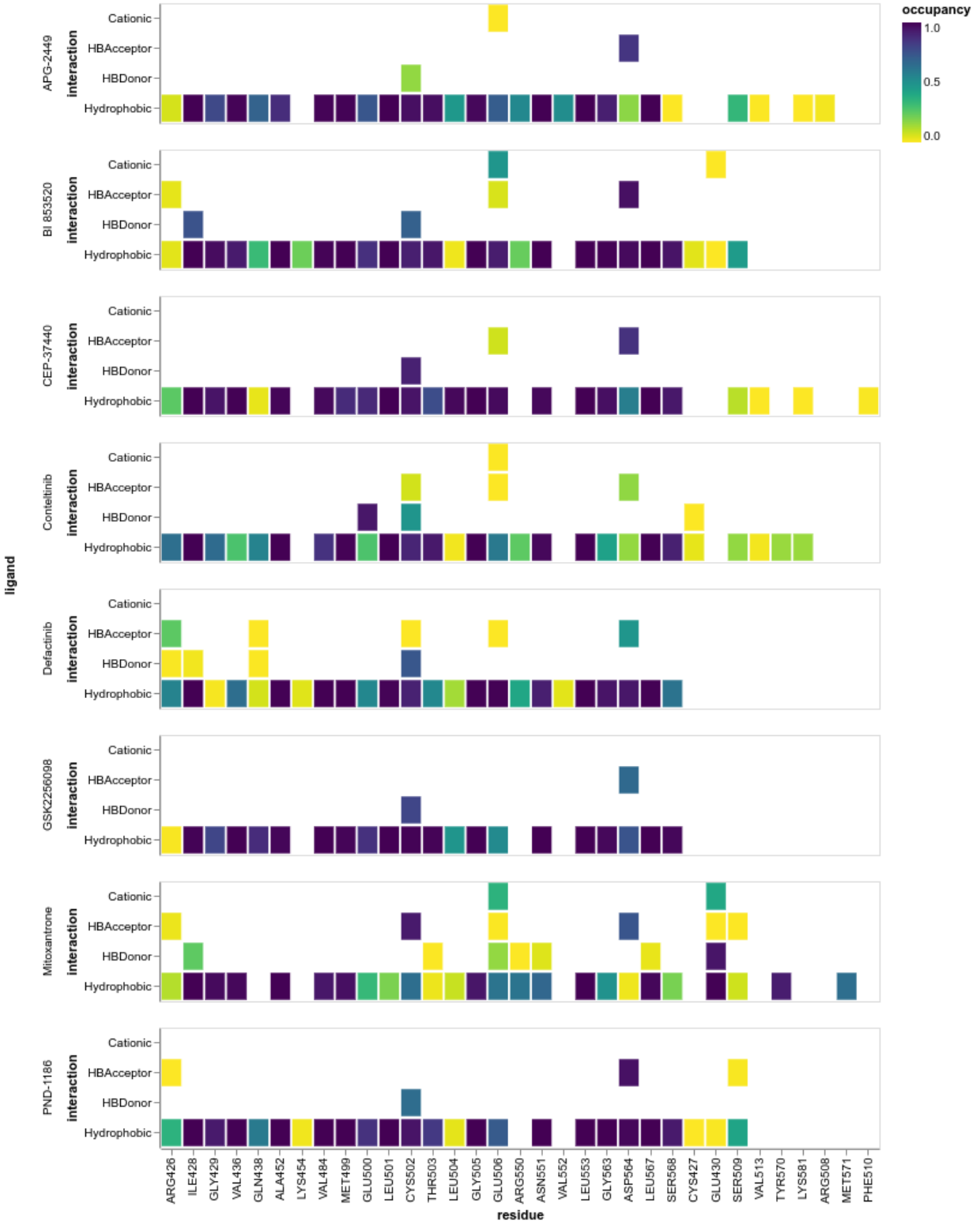


Figure 10: Occupancy values of the last 10 ns of MD simulations, after 40 ns of equilibration, for non-crystallized FAK inhibitors. The complete information for all ligands was processed and condensed onto a dynamic graph usable with any desktop browser, available at prolif_fingerprints.html (Supplementary material).

Conclusions

In this study, we introduced a set of metrics based on the Sørensen–Dice coefficient to quantify the similarity between ligand-protein interaction profiles by representing ligand-residue interactions as Gaussian energy distributions derived from MD simulations. This approach seeks to capture the dynamic nature of ligand binding, which is often overlooked in traditional analyses that focus on mean energy values or predefined interaction types. Our analysis showed that the proposed similarity metrics are sensitive to both the strength and variability of ligand-residue interactions. This is particularly valuable in cases where ligands exhibit similar binding energies but differ in the stability or dynamics of their interactions.

The GSDC is useful for clustering ligands based on their binding modes, highlighting both similarities and differences that might not be apparent from static structural comparisons alone. Our application of this methodology to FAK type II inhibitors revealed distinct interaction patterns among the inhibitors. In this case, the GSDC values ranged from 0.62 to 0.90, indicating varying degrees of similarity in binding modes. This analysis successfully identified ligands with significantly different binding modes, such as Mitoxantrone, which showed low similarity and distinct interaction profiles compared to other inhibitors.

Hierarchical clustering was successfully used to group ligands with similar binding characteristics while isolating those with unique interaction patterns. Dimensionality reduction techniques were applied to improve the visualization of these relationships. These methods allowed data points to be represented in a two-dimensional space, making it easier to interpret the relationships between inhibitors. However, it is important to use these techniques with caution, as the reduction in dimensionality inherently leads to the loss of information. MDS prioritizes preserving overall distances, potentially sacrificing local detail, while t-SNE focuses on clustering similar points together, but may distort distances between dissimilar points.

On the other hand, the $RSDC_i$ provides a detailed measure of similarity for individual residue interactions, allowing for a granular analysis of specific binding sites, which is

particularly useful in understanding the contributions of key residues to ligand binding. Additionally, with the complement of a geometric analysis, the nature of the key interactions can be assessed.

In summary, the proposed methodology provides a framework for the quantitative comparison of ligand-protein interaction profiles, taking into account the inherent dynamics of these complexes. This approach is particularly useful in the context of drug discovery, where understanding differences in binding modes can guide the design and optimization of more potent and selective inhibitors.

Acknowledgement

We would like to express our sincere gratitude to PhD student Julián F. Fernandez for his suggestions regarding this work. Part of this work used computational resources from CCAD – Universidad Nacional de Córdoba (`https://ccad.unc.edu.ar`), which are part of SNCAD – MinCyT, República Argentina. This work was partly supported by UNLP (EX005 and 11/X958) and Agencia (PICT-2021-I-A-00412, PICT 2019-2322) from Argentina. MJL and IEL are members of the Researcher Career of CONICET, Argentina. LMH and PAQ have a doctoral fellowship from CONICET.

Table SM1: RMSD values for re-docking tests.

Molecule	PDB	SMINA	FRED
PF-562,271	3BZ3	1.383	1.496
PVT	6YVS	1.923	1.782
PKE	6YQ1	1.513	1.379
TAE226	2JKK	0.774	0.804
10N	4GU6	0.730	1.508
P7N	6YT6	2.382	2.00
P4N	6YOJ	1.886	1.980
BI-4464	6I8Z	0.893	1.054
PHM16	4C7T	1.211	0.678
7PY	2ETM	1.345	1.630

Table SM2: FAK ATP binding site inhibitors. ΔE is the binding enthalpy calculated with MMPBSA.py. The experimental IC50 activity values were included for better comparison. The non-crystallized inhibitors are highlighted in boldface.

Molecule	ΔE_{bind} MM/GBSA (kcal/mol^{-1})	IC50 (nM)
PF-562,271	-46.3±3.0	1.5
TAE226	-49.3±3.0	5.5
BI-4464	-54.3±3.3	17
10N	-47.2±2.9	195
7PY	-38.3±2.6	200
PHM16	-45.2±2.8	400
P4N	-51.8±2.9	–
PVT	-42.0±3.5	–
P7N	-45.9±3.1	–
PKE	-40.8±3.0	–
Defactinib	-45.3±3.7	0.6
BI 853520	-58.0±3.8	1.0
PND-1186	-52.9±3.0	1.5
GSK2256098	-47.8±2.6	1.5
Conteltinib	-54.8±5.3	1.6
CEP-37440	-51.6±3.0	2.3
APG-2449	-49.1±3.0	–
Mitoxantrone	-36.0±3.7	–

Supporting Information Available

Table SM3: Overlapping between real and Gaussian approximation (Gaussianity).

Residue	Overlapping
ARG426	0.838
CYS427	0.942
ILE428	0.962
GLY429	0.899
GLU430	0.878
VAL436	0.896
HIS437	0.946
GLN438	0.919
VAL451	0.965
ALA452	0.927
LYS454	0.979
VAL484	0.823
LYS485	0.911
MET499	0.943
GLU500	0.957
LEU501	0.962
CYS502	0.972
THR503	0.843
LEU504	0.926
GLY505	0.946
GLU506	0.897
LEU507	0.966
ARG508	0.971
ALA548	0.957
ALA549	0.952
ARG550	0.970
ASN551	0.941
VAL552	0.954
LEU553	0.925
VAL554	0.954
LYS561	0.964
LEU562	0.960
GLY563	0.965
ASP564	0.964
PHE565	0.963
GLY566	0.962
LEU567	0.976
SER568	0.906
ARG569	0.972
TYR570	0.909
MET571	0.900
SER574	0.975
THR575	0.970
TYR576	0.958
TYR577	0.840
LYS578	0.975
SER580	0.893
LYS581	0.813

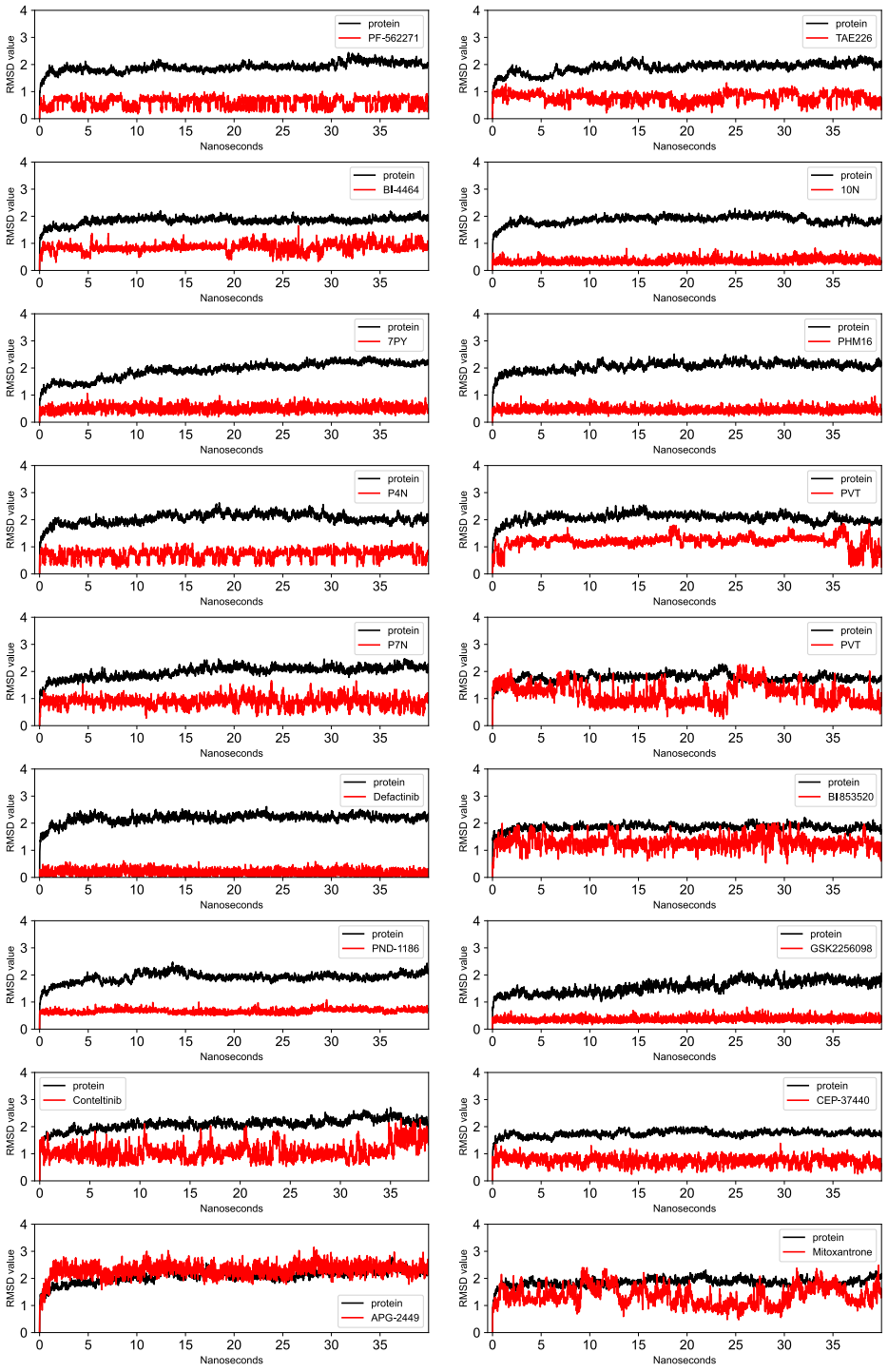


Figure SM1: RMSD of ligands and protein along 40ns MD trajectories.

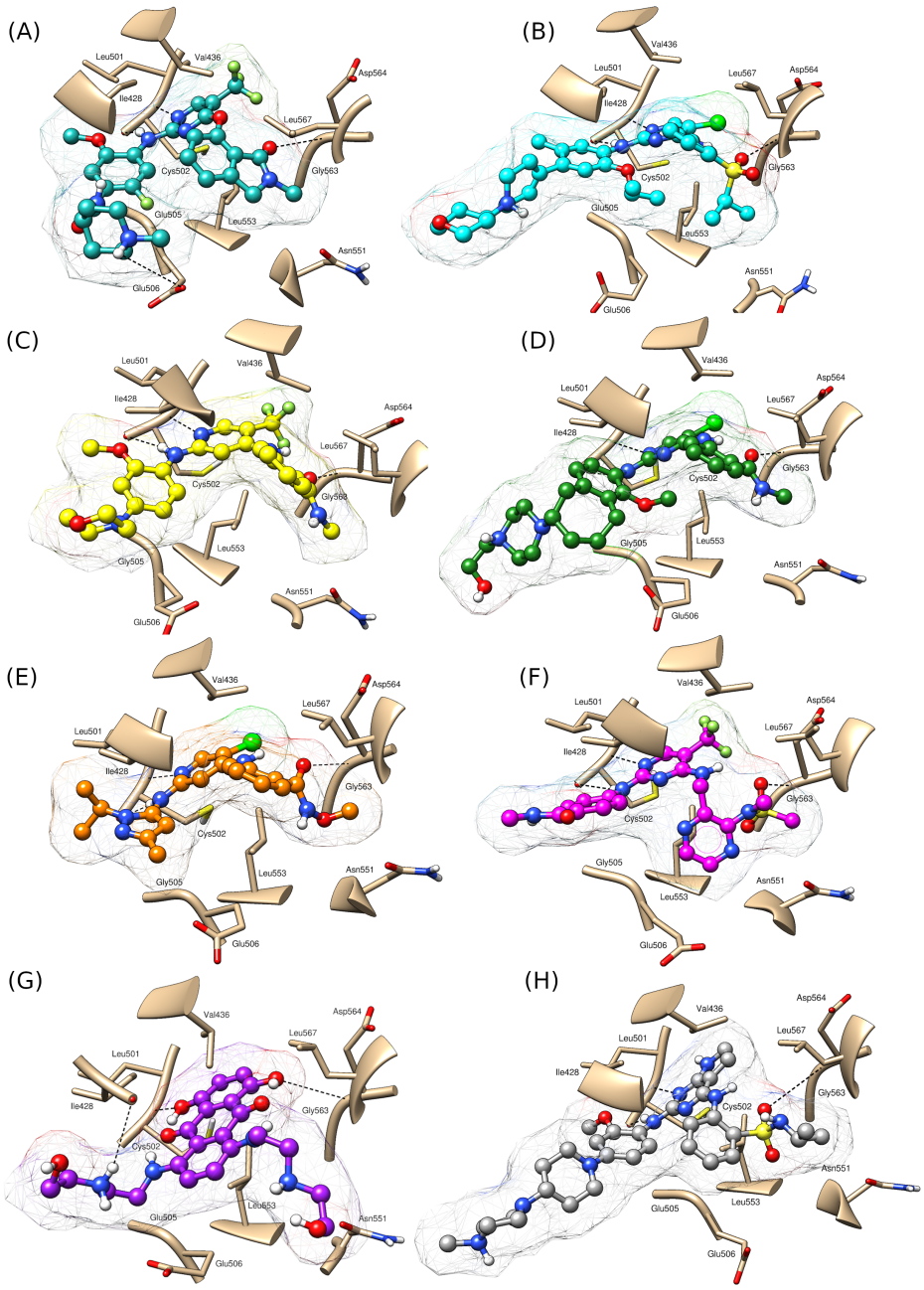


Figure SM2: Localization in ATP binding site of non-crystallized ligands after 40ns of MD (A) BI 853520 (B) APG-2449 (C) PND-1186 (D) CEP-37440 (E) GSK2256098 (F) Defactinib (G) Mitoxantrone (H) Conteltinib

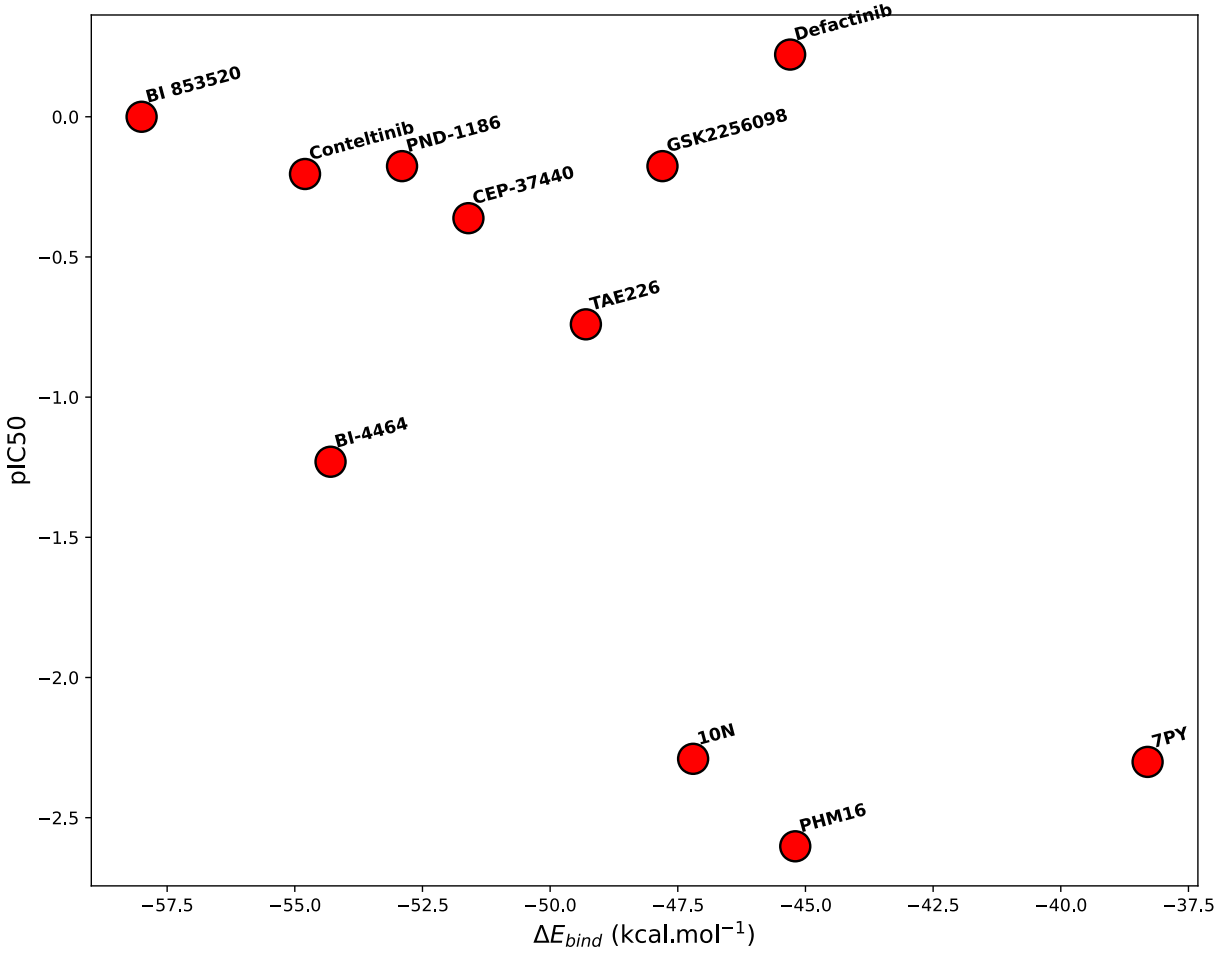


Figure SM3: pIC50 vs ΔEbind for the studied ligands. The cristalized and non-cristalized inhibitors have similar ΔEbind values. A subtle correlation is ovserbed between pIC50 and ΔEbind.

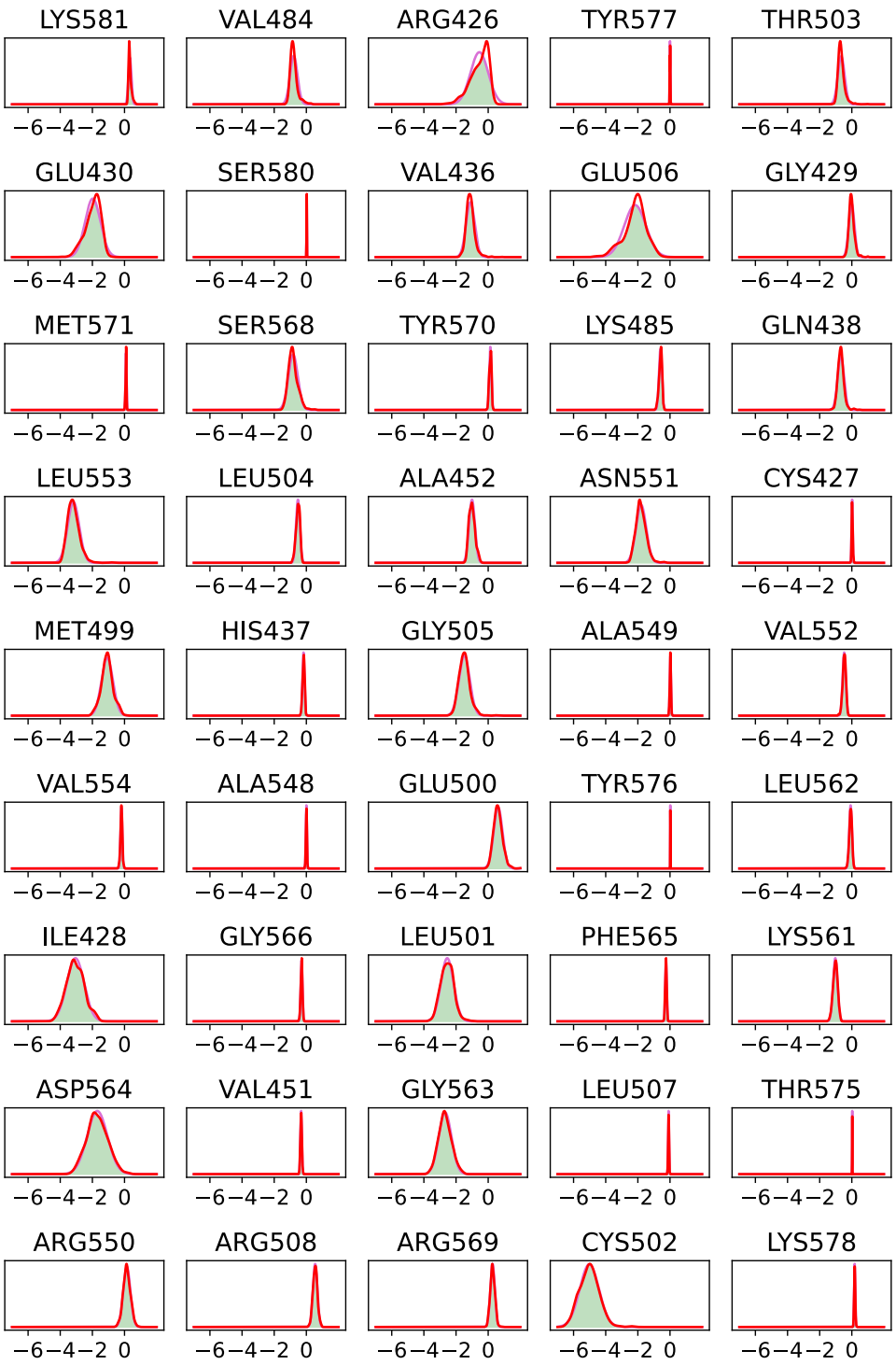


Figure SM4: Comparison between ligand-residue interaction energy distributions obtained from an MD simulation of BI 853520 in complex with FAK and the Gaussian approximation. Energies are shown on the x-axis. The subplots are arranged in descending order according to their overlapping values.

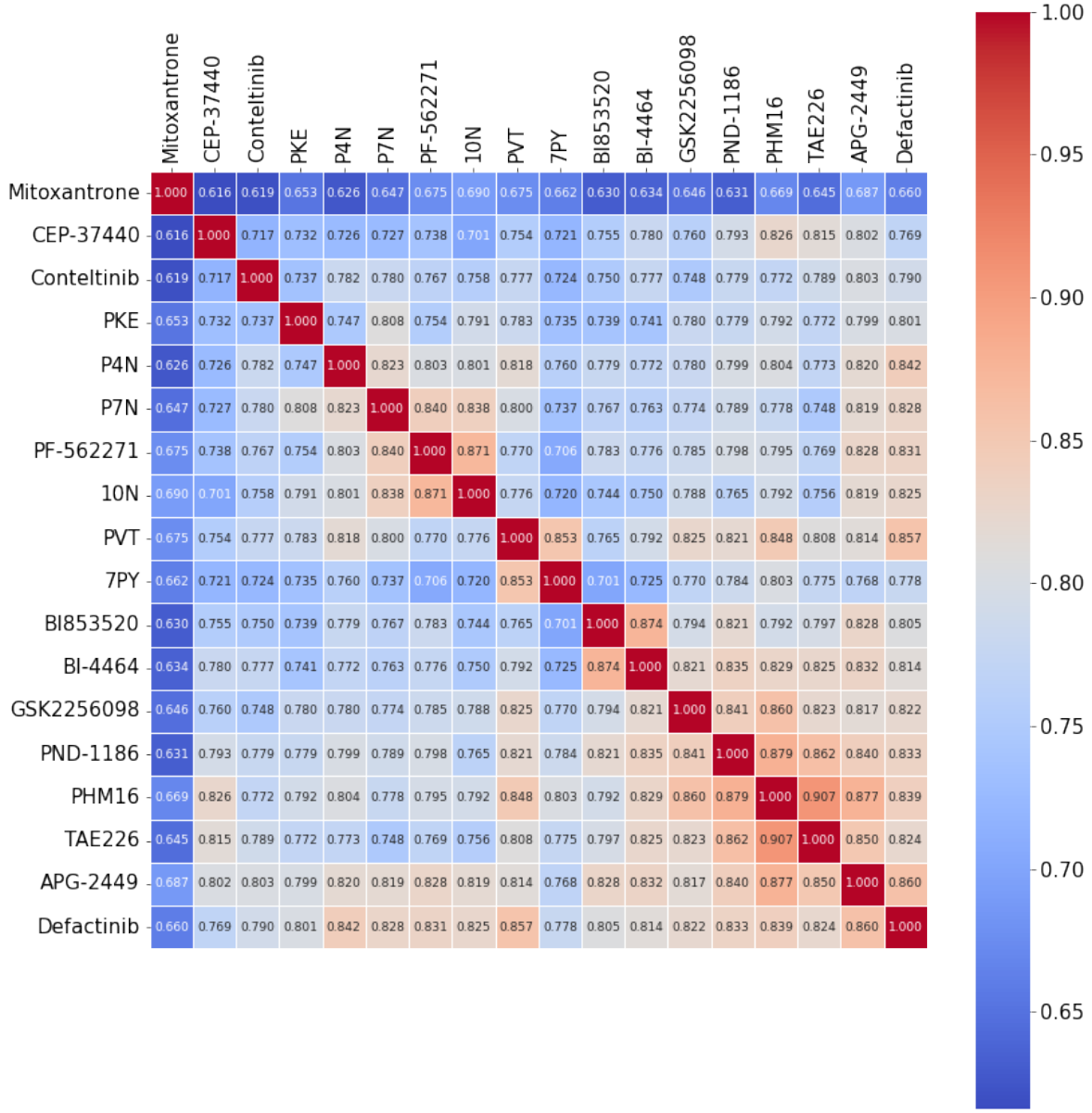


Figure SM5: GSDC values for FAK inhibitors.

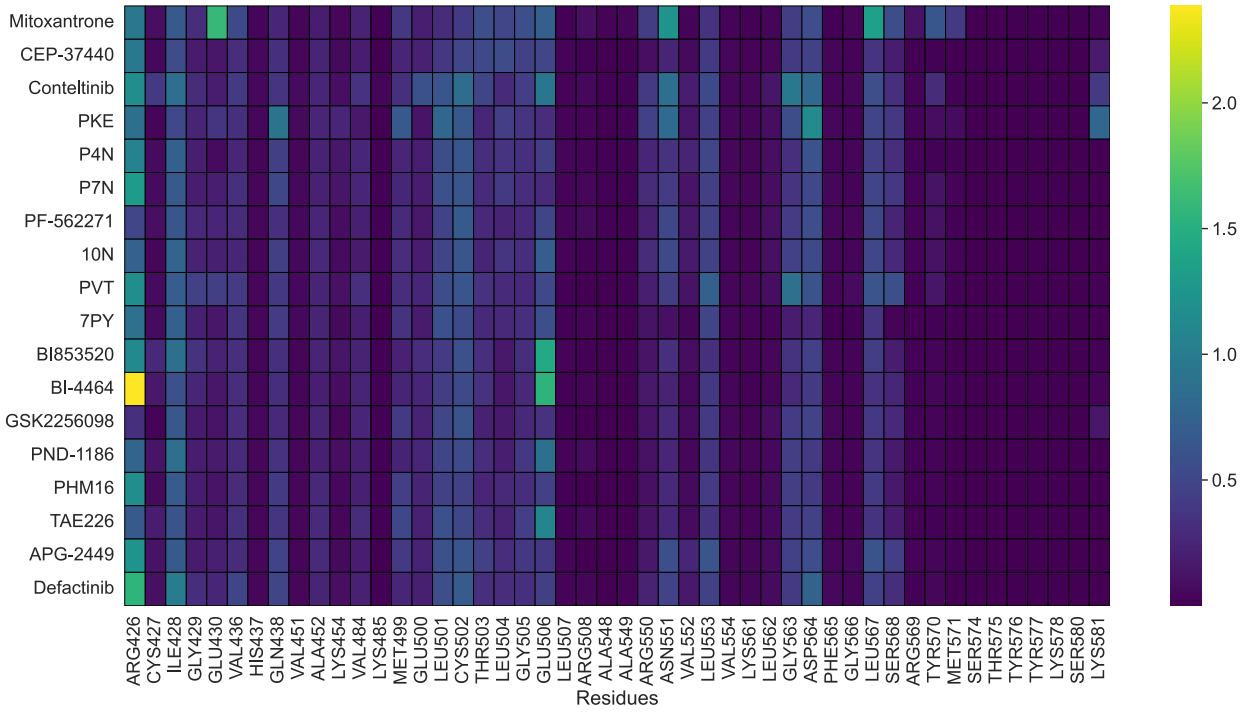


Figure SM6: Heatmap of standard deviations from MM/GBSA by residue for FAK inhibitors.

References

(1) Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opinion on Drug Discovery* **2015**, *10*, 449–461.

(2) Liu, K.; Kokubo, H. Exploring the stability of ligand binding modes to proteins by molecular dynamics simulations: a cross-docking study. *Journal of chemical information and modeling* **2017**, *57*, 2514–2522.

(3) Podlewska, S.; Bojarski, A. J. *Molecular Docking for Computer-Aided Drug Design*; Elsevier, 2021; pp 57–74.

(4) Chuang, H.-H.; Zhen, Y.-Y.; Tsai, Y.-C.; Chuang, C.-H.; Hsiao, M.; Huang, M.-S.; Yang, C.-J. FAK in cancer: from mechanisms to therapeutic strategies. *International journal of molecular sciences* **2022**, *23*, 1726.

(5) Shanthi, E.; Krishna, M. H.; Arunesh, G. M.; Venkateswara Reddy, K.; Sooriya Kumar, J.; Viswanadhan, V. N. Focal adhesion kinase inhibitors in the treatment of metastatic cancer: a patent review. *Expert opinion on therapeutic patents* **2014**, *24*, 1077–1100.

(6) Berman, H. M.; Battistuz, T.; Bhat, T. N.; Bluhm, W. F.; Bourne, P. E.; Burkhardt, K.; Feng, Z.; Gilliland, G. L.; Iype, L.; Jain, S.; others The protein data bank. *Acta Crystallographica Section D: Biological Crystallography* **2002**, *58*, 899–907.

(7) Heinrich, T.; Seenisamy, J.; Emmanuvel, L.; Kulkarni, S. S.; Bomke, J.; Rohdich, F.; Greiner, H.; Esdar, C.; Krier, M.; Grädler, U.; others Fragment-based discovery of new highly substituted 1 H-pyrrolo [2, 3-b]-and 3 H-imidazolo [4, 5-b]-pyridines as focal adhesion kinase inhibitors. *Journal of medicinal chemistry* **2013**, *56*, 1160–1170.

(8) Fiser, A.; Do, R. K. G.; Šali, A. Modeling of loops in protein structures. *Protein science* **2000**, *9*, 1753–1773.

(9) Fiser, A.; Sali, A. ModLoop: automated modeling of loops in protein structures. *Bioinformatics* **2003**, *19*, 2500–2501.

(10) Quiroga, R.; Villarreal, M. A. Vinardo: A scoring function based on autodock vina improves scoring, docking, and virtual screening. *PloS one* **2016**, *11*, e0155183.

(11) McGann, M. FRED and HYBRID docking performance on standardized datasets. *Journal of computer-aided molecular design* **2012**, *26*, 897–906.

(12) El-Hachem, N.; Haibe-Kains, B.; Khalil, A.; Kobeissy, F. H.; Nemer, G. AutoDock and AutoDockTools for protein-ligand docking: beta-site amyloid precursor protein cleaving enzyme 1 (BACE1) as a case study. *Neuroproteomics: Methods and Protocols* **2017**, 391–403.

(13) Hawkins, P. C.; Nicholls, A. Conformer generation with OMEGA: learning from the data set and the analysis of failures. *Journal of chemical information and modeling* **2012**, *52*, 2919–2936.

(14) Scardino, V.; Bollini, M.; Cavasotto, C. N. Combination of pose and rank consensus in docking-based virtual screening: the best of both worlds. *RSC advances* **2021**, *11*, 35383–35391.

(15) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *Journal of computational chemistry* **2004**, *25*, 1157–1174.

(16) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of molecular graphics and modelling* **2006**, *25*, 247–260.

(17) Frisch, M. J. et al. Gaussian 03, Revision C.02. Gaussian, Inc., Wallingford, CT, 2004.

(18) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *Journal of chemical theory and computation* **2015**, *11*, 3696–3713.

(19) Case, D. A.; Aktulga, H. M.; Belfon, K.; Cerutti, D. S.; Cisneros, G. A.; Cruzeiro, V. W. D.; Forouzesh, N.; Giese, T. J.; Götz, A. W.; Gohlke, H.; others AmberTools. *Journal of chemical information and modeling* **2023**, *63*, 6183–6191.

(20) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *Journal of computational chemistry* **2005**, *26*, 1781–1802.

(21) Elber, R.; Ruymgaart, A. P.; Hess, B. SHAKE parallelization. *The European Physical Journal Special Topics* **2011**, *200*, 211–223.

(22) Mongan, J.; Simmerling, C.; McCammon, J. A.; Case, D. A.; Onufriev, A. Generalized Born model with a simple, robust molecular volume correction. *Journal of chemical theory and computation* **2007**, *3*, 156–169.

(23) Miller III, B. R.; McGee Jr, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. MMPBSA. py: an efficient program for end-state free energy calculations. *Journal of chemical theory and computation* **2012**, *8*, 3314–3321.

(24) Bouysset, C.; Fiorucci, S. ProLIF: a library to encode molecular interactions as fingerprints. *Journal of cheminformatics* **2021**, *13*, 72.

(25) Schultze, A.; Fiedler, W. Therapeutic potential and limitations of new FAK inhibitors in the treatment of cancer. *Expert opinion on investigational drugs* **2010**, *19*, 777–788.

(26) Kurio, N.; Shimo, T.; Fukazawa, T.; Takaoka, M.; Okui, T.; Hassan, N. M. M.; Honami, T.; Hatakeyama, S.; Ikeda, M.; Naomoto, Y.; others Anti-tumor effect in human breast cancer by TAE226, a dual inhibitor for FAK and IGF-IR in vitro and in vivo. *Experimental cell research* **2011**, *317*, 1134–1146.

(27) Otani, H.; Yamamoto, H.; Takaoka, M.; Sakaguchi, M.; Soh, J.; Jida, M.; Ueno, T.; Kubo, T.; Asano, H.; Tsukuda, K.; others TAE226, a bis-anilino pyrimidine compound, inhibits the EGFR-mutant kinase including T790M mutant to show anti-tumor effect on EGFR-mutant non-small cell lung cancer cells. *PLoS One* **2015**, *10*, e0129838.

(28) Liu, T.-J.; LaFortune, T.; Honda, T.; Ohmori, O.; Hatakeyama, S.; Meyer, T.; Jackson, D.; de Groot, J.; Yung, W. A. Inhibition of both focal adhesion kinase and insulin-like growth factor-I receptor kinase suppresses glioma proliferation in vitro and in vivo. *Molecular cancer therapeutics* **2007**, *6*, 1357–1367.

(29) Zhan, J.-Y.; Zhang, J.-L.; Wang, Y.; Li, Y.; Zhang, H.-X.; Zheng, Q.-C. Exploring the interaction between human focal adhesion kinase and inhibitors: a molecular dy-

namic simulation and free energy calculations. *Journal of Biomolecular Structure and Dynamics* **2016**, *34*, 2351–2366.

(30) Dao, P.; Jarray, R.; Smith, N.; Lepelletier, Y.; Le Coq, J.; Lietha, D.; Hadj-Slimane, R.; Herbeuval, J.-P.; Garbay, C.; Raynaud, F.; others Inhibition of both focal adhesion kinase and fibroblast growth factor receptor 2 pathways induces anti-tumor and anti-angiogenic activities. *Cancer letters* **2014**, *348*, 88–99.

(31) Popow, J.; Arnhof, H.; Bader, G.; Berger, H.; Ciulli, A.; Covini, D.; Dank, C.; Gmaschitz, T.; Greb, P.; Karolyi-Özguer, J.; others Highly selective PTK2 proteolysis targeting chimeras to probe focal adhesion kinase scaffolding functions. *Journal of Medicinal Chemistry* **2019**, *62*, 2508–2520.

(32) M Golubovskaya, V.; Ho, B.; Zheng, M.; Magis, A.; Ostrov, D.; G Cance, W. Mitoxantrone targets the ATP-binding site of FAK, binds the FAK kinase domain and decreases FAK, Pyk-2, c-Src, and IGF-1R in vitro kinase activities. *Anti-Cancer Agents in Medicinal Chemistry (Formerly Current Medicinal Chemistry-Anti-Cancer Agents)* **2013**, *13*, 546–554.

(33) Choi, H.-S.; Wang, Z.; Richmond, W.; He, X.; Yang, K.; Jiang, T.; Sim, T.; Karanewsky, D.; Gu, X.-j.; Zhou, V.; others Design and synthesis of 7H-pyrrolo [2, 3-d] pyrimidines as focal adhesion kinase inhibitors. Part 1. *Bioorganic & Medicinal Chemistry Letters* **2006**, *16*, 2173–2176.

(34) Heinrich, T.; Seenisamy, J.; Emmanuvel, L.; Kulkarni, S. S.; Bomke, J.; Rohdich, F.; Greiner, H.; Esdar, C.; Krier, M.; Grädler, U.; others Fragment-based discovery of new highly substituted 1 H-pyrrolo [2, 3-b]-and 3 H-imidazolo [4, 5-b]-pyridines as focal adhesion kinase inhibitors. *Journal of medicinal chemistry* **2013**, *56*, 1160–1170.

(35) Ortiz-Rivera, J.; Nuñez, R.; Kucheryavykh, Y.; Kucheryavykh, L. The PYK2 inhibitor

PF-562271 enhances the effect of temozolomide on tumor growth in a C57Bl/6-Gl261 mouse glioma model. *Journal of Neuro-oncology* **2023**, *161*, 593–604.

(36) Berger, B.-T.; Amaral, M.; Kokh, D. B.; Nunes-Alves, A.; Musil, D.; Heinrich, T.; Schröder, M.; Neil, R.; Wang, J.; Navratilova, I.; others Structure-kinetic relationship reveals the mechanism of selectivity of FAK inhibitors over PYK2. *Cell Chemical Biology* **2021**, *28*, 686–698.

(37) Zhang, J.; He, D.-H.; Zajac-Kaye, M.; Hochwald, S. N. A small molecule FAK kinase inhibitor, GSK2256098, inhibits growth and survival of pancreatic ductal adenocarcinoma cells. *Cell cycle* **2014**, *13*, 3143–3149.

(38) Ott, G. R.; Cheng, M.; Learn, K. S.; Wagner, J.; Gingrich, D. E.; Lisko, J. G.; Curry, M.; Mesaros, E. F.; Ghose, A. K.; Quail, M. R.; others Discovery of clinical candidate CEP-37440, a selective inhibitor of focal adhesion kinase (FAK) and anaplastic lymphoma kinase (ALK). *Journal of medicinal chemistry* **2016**, *59*, 7478–7496.

(39) Hirt, U. A.; Waizenegger, I. C.; Schweifer, N.; Haslinger, C.; Gerlach, D.; Braunger, J.; Weyer-Czernilofsky, U.; Stadtmüller, H.; Sapountzis, I.; Bader, G.; others Efficacy of the highly selective focal adhesion kinase inhibitor BI 853520 in adenocarcinoma xenograft models is linked to a mesenchymal tumor phenotype. *Oncogenesis* **2018**, *7*, 21.

(40) Xu, Q.; Kolev, V. N.; Wright, Q. G.; Shapiro, I. M.; Vidal, C. M.; Padval, M.; Keegan, M.; Paterson, D.; Horobin, J.; Pachter, J. A. Focal adhesion kinase (FAK) inhibitors VS-6063 and VS-4718 target cancer stem cells. **2013**,

(41) Song, P.; Zhang, X.; Yang, D.; Wang, H.; Si, X.; Zhang, L. Single-center study to determine the safety and efficacy of CT-707 in Chinese patients with advanced anaplastic lymphoma kinase-rearranged non-small-cell lung cancer. *Thoracic Cancer* **2020**, *11*, 1216–1223.

(42) Fang, D. D.; Tao, R.; Wang, G.; Li, Y.; Zhang, K.; Xu, C.; Zhai, G.; Wang, Q.; Wang, J.; Tang, C.; others Discovery of a novel ALK/ROS1/FAK inhibitor, APG-2449, in preclinical non-small cell lung cancer and ovarian cancer models. *BMC cancer* **2022**, *22*, 752.

(43) Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; others SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature methods* **2020**, *17*, 261–272.

(44) Pastore, M.; Calcagnì, A. Measuring distribution similarities between samples: a distribution-free overlapping index. *Frontiers in psychology* **2019**, *10*, 1089.

(45) Carass, A.; Roy, S.; Gherman, A.; Reinhold, J. C.; Jesson, A.; Arbel, T.; Maier, O.; Handels, H.; Ghafoorian, M.; Platel, B.; others Evaluating white matter lesion segmentations with refined Sørensen-Dice analysis. *Scientific reports* **2020**, *10*, 8242.

(46) Bero, S.; Muda, A.; Choo, Y.; Muda, N.; Pratama, S. Similarity measure for molecular structure: A brief review. Journal of Physics: Conference Series. 2017; p 012015.