



OPEN ACCESS

EDITED BY

Konda Reddy Karnati,
Bowie State University, United States

REVIEWED BY

Soumya Basu,
Dr. D. Y. Patil Biotechnology &
Bioinformatics Institute, India
Kirubakaran Palani,
University of Texas MD Anderson
Cancer Center, United States

*CORRESPONDENCE

Stefan M. Ivanov,
✉ sivanov@addg-pharmfac.net

RECEIVED 28 November 2025

REVISED 10 January 2026

ACCEPTED 28 January 2026

PUBLISHED 27 February 2026

CITATION

Ivanov SM (2026) An efficient computational chemistry approach to generating negative data for drug discovery pipeline validation. *Front. Bioinform.* 6:1756279. doi: 10.3389/fbinf.2026.1756279

COPYRIGHT

© 2026 Ivanov. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

An efficient computational chemistry approach to generating negative data for drug discovery pipeline validation

Stefan M. Ivanov^{1,2,3*}

¹Faculty of Pharmacy, Medical University of Sofia, Sofia, Bulgaria, ²Centre of Excellence in Informatics and Information and Communication Technologies, Sofia, Bulgaria, ³Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences, Sofia, Bulgaria

Modern virtual high-throughput screening (VHTS) pipelines can be suboptimally validated, with no rigorous studies conclusively demonstrating that every one of their steps reliably adds increasing enrichment atop the baseline random hit rate. Moreover, what little benchmarking studies are available primarily focus on the docking aspect of the pipelines, which is usually only the beginning or near the beginning, and even there, authors tend to use flawed data sets that artificially inflate performance metrics. Herein, we present an alternative method to pipeline validation and data set generation that requires no additional experimental work and expenditure, yet offers vast amounts of negative data that can be used in VHTS pipeline validation. By randomizing ligands across published experimental structures and generating structural isomers of known binders, practically unlimited amounts of negative data can be generated. Such sets of positive and negative data points match closely in key molecular properties and are well suited to pipeline validation. Once such sets are generated, they are to be run through any proposed pipeline, assessing performance at every step. We stress the importance of using negative data of adequate quality and quantity in validation studies to definitively and verifiably demonstrate the utility of a given tool or workflow. Our goal is to help distinguish tools and pipelines that truly accelerate hit discovery and lead optimization from ones that promise to do so but actually do not, whereupon academia and industry can begin to tackle the many unaddressed medical needs of the 21st century.

KEYWORDS

cheminformatics, gnina, MM-PBSA, molecular docking, molecular dynamics simulation, negative data, recovery plots, VHTS

Introduction

The cost of developing new drugs ranges from hundreds of millions to billions of dollars (Sertkaya et al., 2024). Unsurprisingly, there is no shortage of conditions that lack treatment or an effective vaccine or require new and better drugs (Kundra et al., 2021; Sung et al., 2021; Qiu et al., 2009; Ivanov et al., 2020; Brooks, 2024; Draper et al., 2018; *The Lancet Gastroenterology and Hepatology*, 2021; Nasir et al., 2017). Preclinical development accounts for a significant share of the costs, time, and effort involved in modern drug development (Schlander et al., 2021). Indeed, it can take significant time, effort, and resources just to find a hit molecule for a particular target. The baseline random hit rates vary from protein to protein and from ligand set to ligand set, but most high-throughput

screens report hit rates well below 1%. For example, a compilation of 21 screening decks (Tran-Nguyen et al., 2020) deposited in the PubChem BioAssay database (Wang et al., 2017) shows a random hit rate as low as 0.018% against the κ_1 -opioid receptor—out of 284,220 compounds tested, only 51 proved active (assay ID 1777). Moreover, experimental screens where no actives have been found, i.e., have a hit rate of 0, are likely to not be reported at all. While hit rates around or even above 1% have been published (Tran-Nguyen et al., 2020), for any given protein target and commercially available set of small molecules, it would be reasonable to expect a random hit rate of 1% or lower, more likely 0.1% or lower, if and until sufficient evidence becomes available to suggest otherwise. Indeed, if random hit rates around 1% were typical, hit discovery would likely have been easy—for every 1000 random ligands tested, one would obtain about 10 different hits to kickstart a drug development program. Experimentalists have learned from experience that this many hits often require testing tens if not hundreds of thousands of molecules against the target of interest.

In an effort to streamline hit discovery and reduce costs, many practitioners have turned to *in silico* modeling and simulation which has given rise to computer-aided drug design (CADD (Niazi and Mariam, 2024)) and computer-aided molecular design (CAMD (Austin et al., 2016)). In a hit discovery context, structure-based virtual high-throughput screening (VHTS) pipelines (Carlsson and Lutten, 2024; Lionta et al., 2014) typically employ a funnel-shaped model where the search begins with a large number of small molecules, often commercially available collections, which may be filtered for pan-assay interference compounds (PAINS (Baell and Holloway, 2010)) and undesirable properties such as molecular volume or weight that are too large or too small for the binding site of the target (usually a protein molecule or a complex of proteins). The remaining ligands are then initially subjected to a crude and inexpensive computational technique to try to filter out the majority of the likely nonbinders, typically molecular docking (Zhang et al., 2022). Subsequent steps in the pipeline employ increasingly sophisticated, computationally costly, and theoretically rigorous techniques on a progressively decreasing pool of molecules, with the intermediate steps often involving end-state free energy techniques such as the linear interaction energy (LIE) method (Su et al., 2007; Brandsdal et al., 2003), molecular mechanics–generalized Born (MM-GBSA) or molecular mechanics–Poisson–Boltzmann (MM-PBSA) calculations (Genheden and Ryde, 2015; Maffucci and Contini, 2013; Maraf et al., 2024). The final stages of the pipeline may involve theoretically rigorous absolute binding free energy techniques (ABFEs (Aldeghi et al., 2018)) on one to several hundred molecules, the most promising of which are then purchased for experimental testing and verification.

While molecular docking (Paggi et al., 2024) and dynamics (Bowers et al., 2006) have produced new knowledge (Ivanov et al., 2016; Ivanov et al., 2019a; Dodd et al., 2020; Yu et al., 2023), insights (Ivanov et al., 2020; Ivanov et al., 2019b; Bhakat and Söderhjelm, 2017; Kumalo et al., 2016; Childers et al., 2017; Atanasova et al., 2024; Marković et al., 2024; Spassov et al., 2024; Spassov, 2024; Atanasova, 2025; Philipova et al., 2025), and hits (Atanasova et al., 2022; Stavrakov et al., 2020), much remains to be desired. A thorough benchmarking analysis estimating the enrichment provided by

these techniques at every step of a VHTS pipeline is not available. Benchmarking studies have been published primarily focusing on docking (Bertrand and Triballeau, 2005). Yet, even there, much of the results remain suspect due to biases in the composition of the positive and negative data. For example, it has been pointed out that in the directory of useful decoys data set (DUD (Huang et al., 2006)), actives spread over a few dominant scaffolds (analog bias (Good and Oprea, 2008)) and that decoys exhibit net charges different from active molecules (Hawkins et al., 2008). Subsequent efforts have produced the directory of useful decoys data set-enhanced (DUD-E (Mysinger et al., 2012)), DUDE-Z (Stein et al., 2021), and maximum unbiased validation (MUV (Rohrer and Baumann, 2009)) sets which, nevertheless, suffer significant biases (Wallach and Heifets, 2018). For DUD-E, differences in key molecular properties such as polar surface area and hydrogen bond donor counts between actives and nonactives have been reported (Chaput et al., 2016). Moreover, the actives in DUD-E are biased toward higher potency than what is observed during prospective high-throughput screens (Tran-Nguyen et al., 2020). Analyses on commonly used data sets reveal that artificially inflated enrichments are often the result of analogue bias (Xia et al., 2015) or 2D bias (Vázquez et al., 2024). Imbalances in properties between actives and decoys deceptively inflate enrichment numbers and other performance metrics in benchmarking studies using these sets and overestimate the utility of the tools being assessed. This has been known for some time now and practitioners have become weary of overly optimistic reports (Wallach et al., 2015; Pereira et al., 2016), knowing that performance in a truly prospective screen is likely to be much lower than what is reported in the literature. This has provided the impetus for setting up the Critical Assessment of Computational Hit-finding Experiments (CACHE) challenges (Ackloo et al., 2022) where computational chemists are tasked with finding novel hits for targets of interest and optimizing them in a truly prospective setting. This produces an objective assessment of their methods' capabilities unlike *post hoc* reports where authors may omit or gloss over certain failings and exaggerate performance. We explicitly state that any deficiencies we point out here are not a criticism of the authors we cite who have done important foundational work in the field of VHTS. Rather, these serve as an illustration of the inherent difficulty of the problem of finely matching decoys to actives in a highly-dimensional property hyperspace.

So far, the standard approach to data set generation has been to experimentally measure different ligands against targets of interest. This is an expensive approach that inevitably results in differences between actives and inactives (and hence biases), as it is not possible to perfectly match binders to nonbinders in the vastness of chemical property space. Indeed, while authors may be somewhat successful in obtaining decoys with similar molecular weight or partition coefficient (logP) distributions to those in their active sets, it quickly becomes nearly impossible to obtain a good match when one also begins to consider other molecular properties such as number of rotatable bonds (henceforth also referred to as rotors), polar and nonpolar molecular surface area, net charge, number of hydrogen bond donors and acceptors, etc. Moreover, one of the key molecular properties—the number of rotatable bonds a ligand has—is a rather coarse metric. One could easily imagine a scenario where two molecules have the same number of rotatable bonds but which differ greatly in their flexibility (Wicker and Cooper, 2016),

for example, two small aromatic rings being separated by a given number of rotatable bonds versus two large aromatic systems being separated by the same number of rotors. Moreover, thus far no one has described a means to thoroughly measure and validate the performance of and value added by every step in a VHTS pipeline. In this work, we describe an efficient approach to producing negative data for VHTS validation. We also illustrate a means of evaluating performance at every step of the pipeline, which makes possible comparing different techniques and optimizing VHTS pipelines by swapping out less efficient techniques in favor of more productive ones. Moreover, our method allows validating modifications to VHTS pipelines as well as quantitatively comparing the performance of different pipelines.

Our approach is two-fold. First, rather than measuring ligand activities against a target of interest, we take actives from PDBbind (Wang R. et al., 2004) structures. PDBbind is the largest curated set of protein–ligand structures and affinities compiled from literature. To generate a corresponding negative set, we randomize ligands such that every ligand is paired with a noncognate protein structure. This is the only way to achieve a perfect match in chemical properties between positives and negatives, as the positive and negative ligand set contain the exact same molecules. Moreover, this approach has the added benefit that not only does every ligand serve as its own control, but so does every protein.

Second, given a protein–ligand crystal or NMR structure, we generate negative data by producing constitutional isomers (also referred to in chemistry as structural isomers) of the binder with MAYGEN (Yirik et al., 2021). Because actives and decoys have the same heavy atom content, they are much closer in key properties than any set that is obtained by experimentally testing for random actives and inactives which would in all likelihood be completely unrelated. Using the standard approach to data set construction, actives and decoys would likely have very different weights, numbers of rotors, polar and nonpolar molecular surfaces, or exhibit significant differences in some other key property. By generating structural isomers of the binder, many of these imbalances, such as molecular weight or molecular surfaces, are either greatly ameliorated or avoided completely (Ivanov, 2026). Ivanov (2026) explores property matching between actives, experimental decoys, and isomers at much greater depth; it also features example code to filter ligands for or against desired/undesired properties or property ranges.

Methods

PDBbind-derived set

We selected 29 monomeric structures with available dissociation constant (K_D) values from the refined set of PDBbind 2019, thus ensuring that we are working with high quality, reliable structures. For every protein–ligand structure, we generated 25 more structures by docking different noncognate ligands into the binding site of the protein with Gnina 1.0 (McNutt et al., 2021), using the cognate ligand as a reference to define the binding site. Typically, a random ligand can be safely assumed to be a nonbinder (decoy). However, some of the PDB structures in our

set were degenerate in that they contained the same protein (as evidenced by the Uniprot ID (Apweiler et al., 2004)). Therefore, we searched our set for instances where a ligand is docked to different structures of the same protein it is known to bind from the PDBbind structure it originally came from. This resulted in a final set of 29 cognate, binding protein–ligand pairs (the crystal structures), as well as 26 binding pairs from the randomizations. Out of the 725 randomized pairs, we have 26 binding and 699 nonbinding (decoy) pairs. All ligands were docked to the proteins from the 29 PDBbind structures. In 29 cases, this meant ligands were docked to the protein from the crystal structure they came from (redocking); in 26 cases, this meant docking to a different protein conformation (crossdocking). As an example, the ligand from the 5NK6 crystal structure (Heinzlmeir et al., 2017) was docked to the same protein–ephrin type-A receptor 2 – but from a different crystal (5IA3 (Heinzlmeir et al., 2016)) into the binding site defined by the ligand in 5IA3. All protein–ligand pairs and their binding status (binding/nonbinding) are given in [Supplementary Table 1](#).

All 754 docking runs (29 redockings and 725 crossdockings) were performed with Gnina, using 8 Monte Carlo chains during sampling and adding 4 Å of buffer space to the auto-generated box defined by the cognate reference ligand, and expanding the box by an additional 1 Å if needed to make sure the input ligand conformation can freely rotate within the box. Gnina provides 3 scores for every pose it outputs—the predicted affinity (in kcal/mol), a convolutional neural network (CNN) pose score—a neural network derived probability that the pose has a low root mean square deviation (RMSD) to the true binding pose—and a CNN affinity—a neural network estimate of the affinity, given in pK units. For the 754 structures originating from PDBbind (29 redockings and 725 crossdockings), we took the highest-scoring ligand pose for every protein–ligand pair using the CNN pose score and subjected the complexes to molecular dynamics (MD) simulations and MM-GBSA and MM-PBSA calculations.

MAYGEN set

For the MAYGEN decoy generation, we began by taking the ligands from the 4QSW (Chaikuad et al., 2014), 5A5N (Demont et al., 2015), and 6EPU (Dolbois et al., 2020) crystal structures which all bind the same protein – ATPase family AAA domain-containing protein 2 (ATAD2). For every ligand, we generated 25 structural isomer decoys using MAYGEN (see [Supplementary Figure S1](#) and [Supplementary Table 1](#)). For this work, we chose MAYGEN because it is free and open source; any other proprietary or nonproprietary tool for generating constitutional isomers (Steinbeck, 2001; Badertscher et al., 2000; Nuzil et al., 1991; Gugisch et al., 2015) could be used. The ligands were converted from SMILES to sdf format and protonated corresponding to neutral pH with Open Babel (O’Boyle et al., 2011). SMILES strings for all MAYGEN ligands are given in [Supplementary Table 1](#). Finally, all 3 binders and 75 nonbinders were crossdocked to ATAD2 using the 4TU4 (Poncet-Montange et al., 2015) crystal structure and the binding site defined by the 4TU4 ligand which also binds the same site as the other three binders. Docking was performed with Gnina using identical settings as was done with the PDBbind set. However, here, for every

ligand we took the top-scoring pose according to Gnina affinity, CNN pose score, and CNN affinity. These were then subjected to molecular dynamics and MM-PB(GB)SA analysis. Gnina is a stochastic docking tool which means that different runs produce different predictions for the binding pose. Therefore, for both the PDBbind and MAYGEN set, we performed molecular docking in 4 independent replicas followed by molecular dynamics on the best (i.e., highest) scoring poses from every replica. For the PDBbind set, we performed 4 MD replicas starting from the highest scoring conformations by CNN pose score from the 4 docking runs; for the MAYGEN set, we performed 4 MD runs on the best scoring poses using all three Gnina descriptors or 12 MD runs in total for every ligand.

Molecular dynamics and MM-PB(GB)SA calculations

Ligand structures were processed with *antechamber* and *parmchk2* from Amber18 (Case et al., 2005) to generate Amber-compatible mol2 files containing AM1-BCC charges (Jakalian et al., 2000) and GAFF2.11 atom types (Wang JM. et al., 2004). Proteins were parameterized with the ff14SB force field (Maier et al., 2015) using *tleap*. Protein–ligand complexes were solvated with TIP3P water (Jorgensen et al., 1983) in a truncated octahedral box with a wall distance of 13 Å from the outermost atoms; NaCl (Joung and Cheatham, 2008) was added to neutralize the system net charge and reach a concentration of 150 mmol. The structures were then subjected to 50,000 steps of energy minimization with restraints of $3 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{Å}^{-2}$ on protein heavy atoms, followed by constant volume heating to 300 K over a period of 1 ns, a density equilibration phase of 1 ns, followed by 5 ns of equilibration at constant pressure and temperature (NPT conditions), maintained with the Berendsen barostat (Berendsen et al., 1984) and the Langevin thermostat, and a production phase of 25 ns of NPT simulation. Collision frequencies and pressure relaxation times were set as in previous work (Ivanov et al., 2019a). As before, a cutoff of 12 Å was used for both van der Waals and electrostatic interactions; long-range electrostatics beyond the real-space cutoff were calculated with the particle-mesh Ewald (PME (Darden et al., 1993)) scheme, and bonds to hydrogen were constrained with SHAKE (Ciccotti and Ryckaert, 1986), allowing for a 2 fs time step. All systems were simulated with heating, density equilibration, preproduction dynamics, and production dynamics under periodic boundary conditions; only the minimization stage was run under nonperiodic conditions. During production dynamics, frames were saved every 100 ps for a total of 250 frames per trajectory.

We then performed MM-GBSA and MM-PBSA calculations on the 250 protein–ligand frames from the production stage of every trajectory to compute the enthalpy of binding (ΔH) between the proteins and ligands. Full details about these calculations are given in our previous publication on RNase A–ligand binding (Ivanov et al., 2019a). Finally, actives and decoys were ranked by their docking and MM-PB(GB)SA scores. These were compared to assess their performance relative to each other.

Results

PDBbind-derived set

We begin by examining the rankings produced by docking and molecular dynamics on the PDBbind-derived set. It contains a total of 754 protein–ligand pairs, 55 of which are binding (29 redockings and 26 crossdockings) and 699 nonbinding for an active:inactive ratio of around 1:12. Scoring all 754 pairs and ordering them by score produces a monotonic score vs. rank function; marking the positions of the actives produces a plot henceforth referred to as a *recovery* plot. In an ordered list of scores, a perfect estimator would place all 55 binding pairs at the beginning of the list; a perfectly random predictor would space out actives evenly at a ratio of 1 active for every 12 or so decoys. Real scoring tools are assessed by their recovery capabilities—the more actives they place near the beginning of the list, the greater their recovery capacity and the greater their utility in a drug discovery setting. We first compare the rankings afforded by the Gnina affinity and CNN pose score from the four docking runs and compare them to each other as well as to the rankings produced by averaging across the four runs, as well as taking the extremum for every ligand from the four runs (Figure 1). Rather than being evenly spread throughout the ranking list, i.e., the X axis, the actives are predominantly concentrated near its beginning; this is more pronounced with the CNN pose score ranking. We also see that the rankings from the four replicas generally do not differ drastically; averaging scores across replicas or taking the best score for every ligand also do not appear to greatly enhance recovery. Crucially, both affinity and CNN pose score metrics rank many binding pairs well and place them near the beginning of the list. For this set, CNN pose score outperforms affinity in terms of recovery in that affinity places more actives in the middle or near the end of the list. The differences in recovery capacity become more apparent if we look at a violin plot of the ranks the different metrics assign to the actives in the 4 replicas (Figure 2).

The violin plot clearly demonstrates that for the PDBbind-derived set, of the three Gnina descriptors, CNN pose score achieves the best recovery. Of the three parameters, its distribution is most concentrated near 1 (the beginning of the list) – it has the lowest lying median (the white dots in the violin plots) and its interquartile range (the thick black bars in the plots) is the smallest and lowest-lying, i.e., it lies closest to 1. Conversely, this parameter has the lowest fraction of poorly ranked actives – we see that using CNN pose score, a small fraction of active pairs rank beyond 100. This is not the case with affinity and CNN affinity, which have similar distributions. Surprisingly, CNN pose score achieves superior recovery than either MM-GBSA or MM-PBSA which do not differ significantly in rank distributions.

MAYGEN-derived set

For the MAYGEN-derived set, we keep track of the individual actives which allows us to estimate how much their rankings vary across the 4 replicas. Figure 3 shows the recovery plots for the averaged Gnina descriptors and MM-GBSA and MM-PBSA scores with the 95% confidence intervals for the ranks and scores. It shows that the ligand from the 4QSW structure can be expected to be

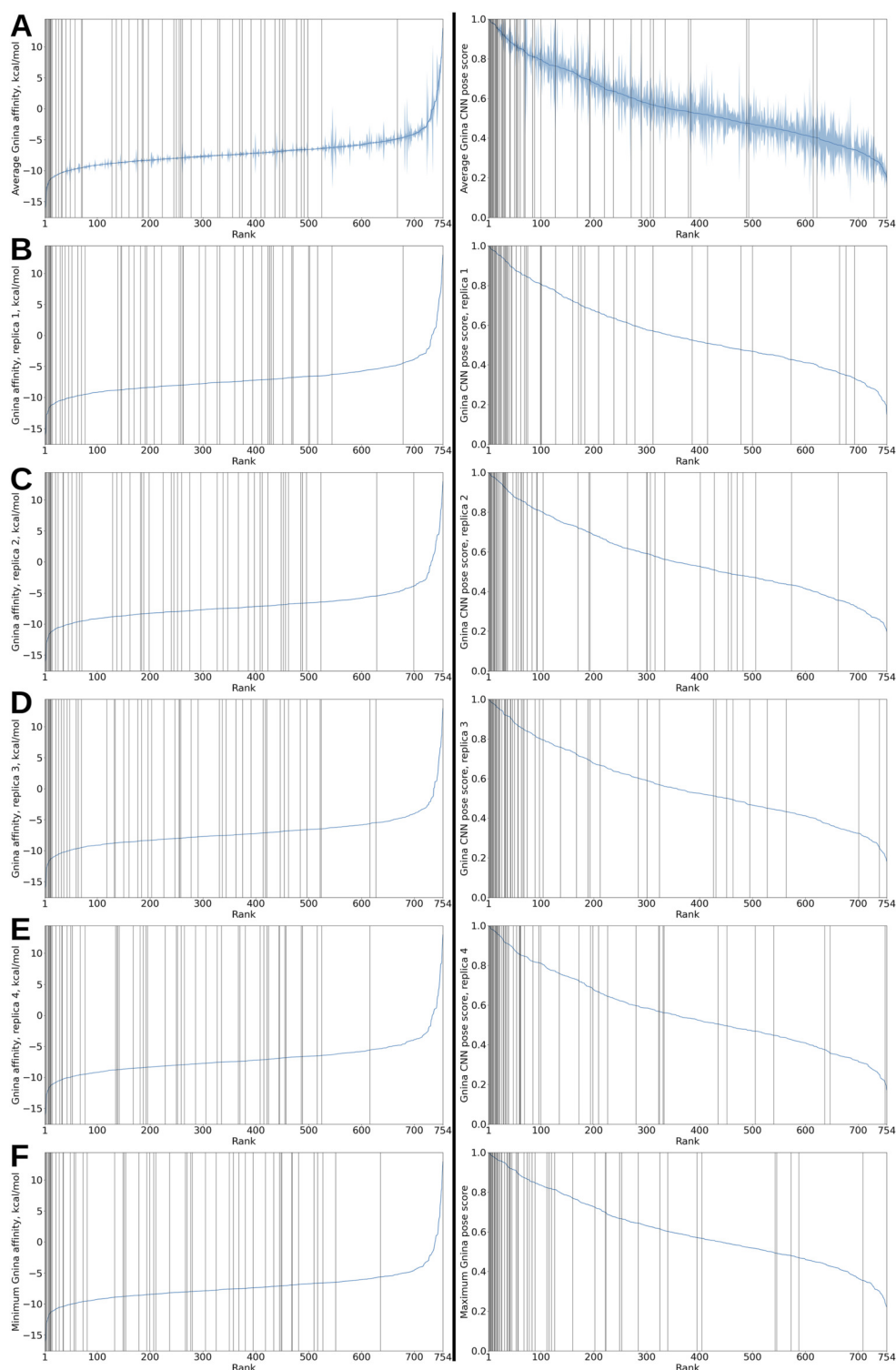


FIGURE 1

Recovery plots for Gnina affinity and CNN pose score on the PDBbind-derived set. In all panels, the vertical black bars indicate binding protein–ligand pairs. **(A)** Left. Protein–ligand pairs sorted by mean Gnina affinity, averaged over the 4 docking runs. The blue shaded region represents the score intervals at 95% confidence corresponding to every rank. Right. Protein–ligand pairs sorted by mean Gnina CNN pose score, averaged over the 4 docking runs. The blue shaded region represents the score intervals at 95% confidence corresponding to every rank. **(B)** Left. Protein–ligand pairs sorted by Gnina affinity from replica 1. Right. Protein–ligand pairs sorted by Gnina CNN pose score from replica 1. **(C)** Left. Protein–ligand pairs sorted by Gnina affinity from replica 2. Right. Protein–ligand pairs sorted by Gnina CNN pose score from replica 2. **(D)** Left. Protein–ligand pairs sorted by Gnina affinity from replica 3. Right. Protein–ligand pairs sorted by Gnina CNN pose score from replica 3. **(E)** Left. Protein–ligand pairs sorted by Gnina affinity from replica 4. Right. Protein–ligand pairs sorted by Gnina CNN pose score from replica 4. **(F)** Left. Protein–ligand pairs sorted by minimum Gnina affinity from the four replicas. Right. Protein–ligand pairs sorted by maximum Gnina CNN pose score from the four replicas.

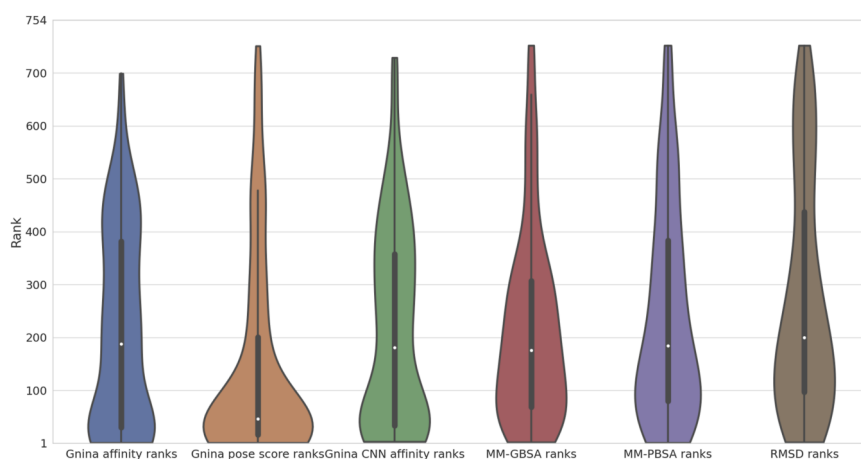


FIGURE 2
Ranking violin plots. The binding protein–ligand pair ranks from the four replicas when sorting by Gnina affinity, CNN pose score, CNN affinity, MM-GBSA, MM-PBSA, and RMSD are given. The RMSD is the ligand heavy atom RMSD at the end of production dynamics compared to at the end of minimization. The white dots represent the medians of the distributions; the thick black bars are the interquartile ranges; the thin black bars are the remaining data points.

placed by Gnina affinity between ranks 10 and 12 out of 78 total compounds (3 actives and 75 decoys) with 95% confidence, as estimated from the four docking runs. The 6EPU ligand can be expected to rank somewhere between 50 and 60; 5A5N consistently ranks poorly near the end of the list. Further, we obtain estimates for the variation in score at the corresponding ranks. For example, we see that for a compound to rank near the 50th spot by Gnina affinity, it needs to have a score between around -6.6 and -5.8 kcal/mol; the variation in scores increases near the end of the ranking. Conversely, there is much less variation in score at the beginning up to rank 10.

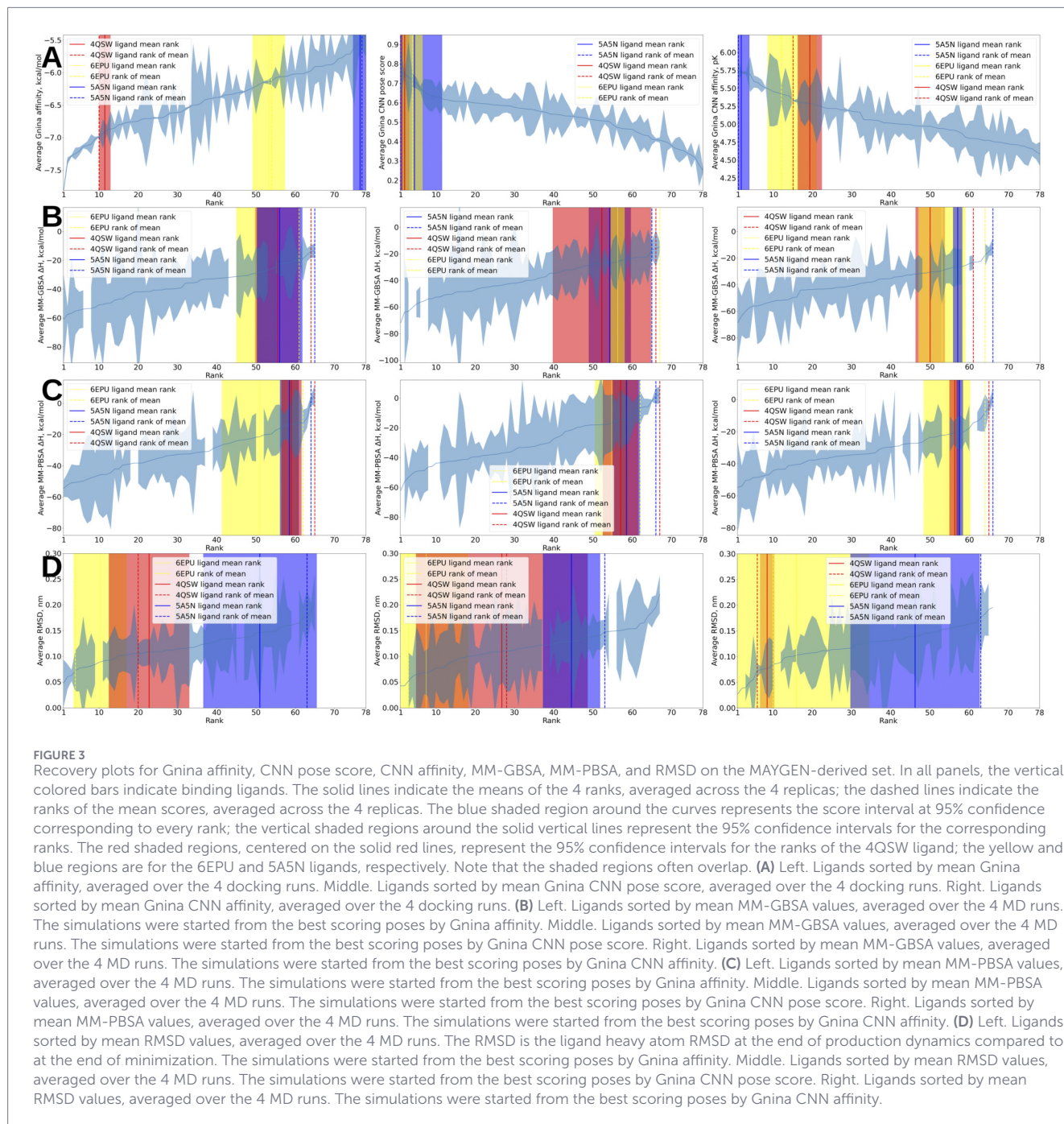
For the set we have here, with all three Gnina descriptors, there is a high degree of consistency between the means of the ranks from the 4 replicas (solid vertical lines in Figure 3) and the ranking produced by averaging scores and ordering the mean values (dashed vertical lines in Figure 3). Crucially, with the MAYGEN set, as well as with the PDBbind-derived set, CNN pose score is the top performer not only among the Gnina descriptors but overall—it ranks the three actives near the beginning of the ranking, unlike all other parameters which place one or all of the actives near the middle or the end. Although not all MD simulations succeeded and hence we do not have MM-GBSA and MM-PBSA scores for all ligands, it is evident the end-state techniques underperform—they clearly achieve inferior recovery to Gnina pose score, even on a smaller set of ligands. Moreover, MD-based scores show considerably more variation in score and ranking among the 4 replicas.

Discussion

First, consider the active structures from Supplementary Figure S1. Then consider their respective isomers. Looking at the 2D structures, one can have a high degree of confidence that not even one of the isomers is an active molecule or a false negative.

Similarly, for the PDBbind randomizations, one can be highly confident that the overwhelming majority of the randomized protein–ligand pairs do not bind each other with any appreciable affinity. Instead of synthesizing and experimentally testing every protein–ligand pair, the majority are discarded after an *in silico* evaluation, with experimental testing reserved only for the most promising compounds. These are some of the cost savings that justify and enable computer-aided molecular discovery and design. Remarkably, it is the high degree of selectivity and low random binding rates that enable modern pharmacotherapy (Gim and Vincent, 2013) in the first place. Indeed, were random binding rates much higher than 1% on average, it would likely be very difficult to administer a drug for a specific intended target without it perturbing some other important protein or pathway given that the human genome encodes around 20,000 proteins (Amaral et al., 2023) with additional splice variants (Wu et al., 2022) and macromolecules such as RNAs, DNA, and different macromolecular assemblies. To conclude this line of reasoning, we note that the primary utility of validation sets is in comparison—two or more pipelines are compared in terms of performance on the same data set or two or more modifications or variations of the same pipeline are tested. Recovery is evaluated using the known actives. Whatever false negatives may be present, be it due to alternative conformations of the primary binding site or any alternative binding sites, are treated equally as negatives until sufficient data become available to reclassify them as binders.

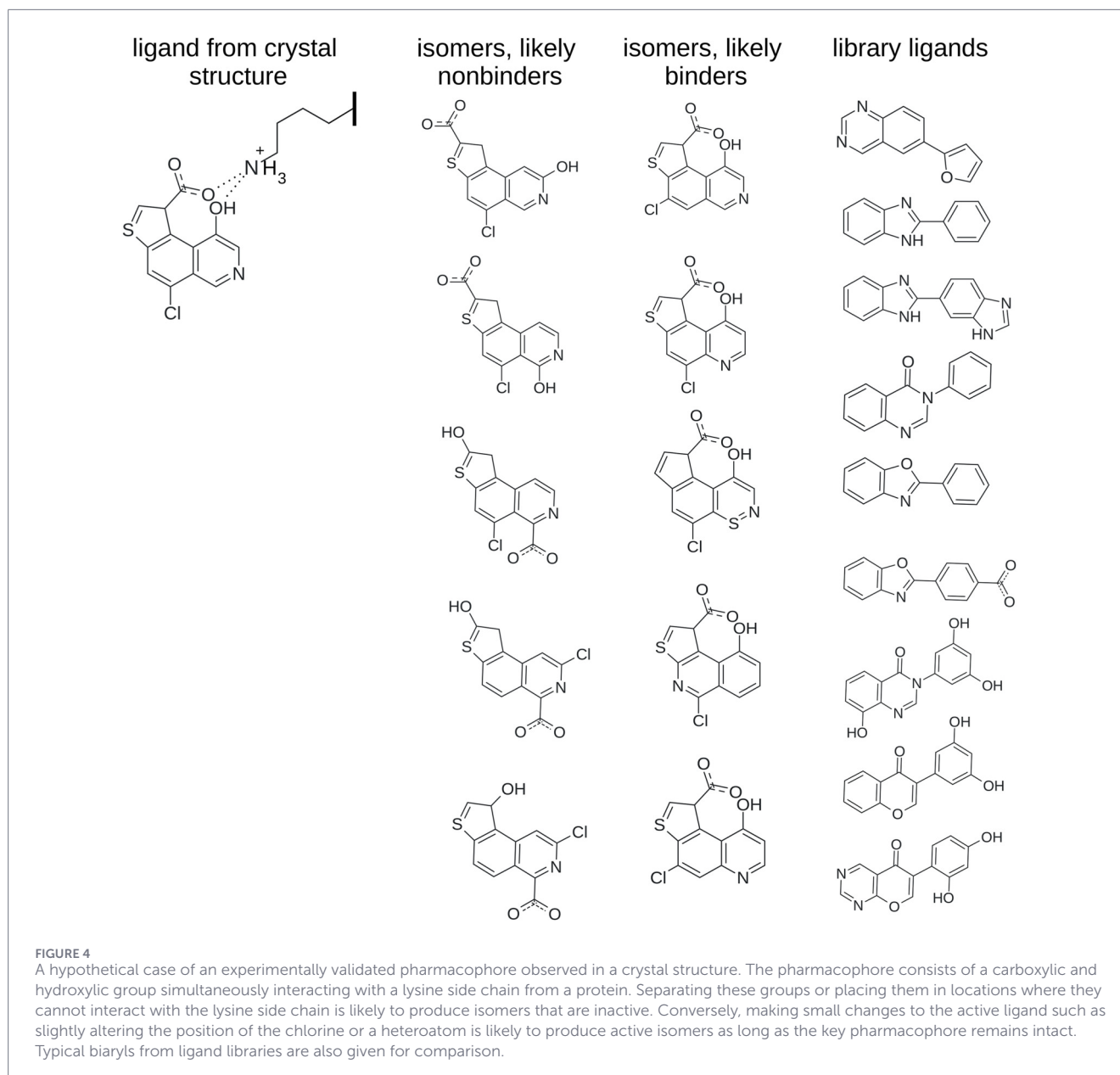
Generating isomers is particularly easy and useful when there is a clear, well-established and experimentally validated pharmacophore that is known to be essential for ligand activity. This scenario is illustrated in Figure 4 where we have a hypothetical pharmacophore consisting of a carboxylic and hydroxyl group simultaneously interacting with a lysine side chain. In such a scenario, it would be easy to generate nonactive isomers by placing the hydroxyl group and carboxylate far apart and away from the lysine side chain. Such ligands would be far more suitable



comparisons to the active ligand than the typical biaryls which are often found in drugs and druglike molecules (Dahlgren et al., 2013) and ligand libraries (Hussain et al., 2016; Boger et al., 1999; Stammers et al., 2023; Clark et al., 2023). Notably, in such cases, isomerization can likely generate active ligands as well. Given the active molecule in Figure 4, for example, shifting the position of the chlorine slightly or moving around one of the heteroatoms is not likely to abrogate activity, as long as the key pharmacophore remains intact.

Here, we present a rigorous means of estimating recovery and enrichment at every step of a VHTS pipeline in order to lay out a blueprint for future validation studies. The actives:inactives

ratio in our PDBbind set is around 1:12 (55 binding pairs to 699 nonbinding pairs), with the actives spread nearly equally between redocking and crossdocking cases; the active to inactive ratio in our MAYGEN set is 1:25. We heavily emphasize that the tools and data set we use here are merely examples. A vast number of alternative techniques and combinations thereof are possible. We cannot conclusively say which if any one of them is definitively superior to the rest and, if so, what that is. We can only provide a method to tackle these questions which is a large undertaking for the entire computational chemistry community. Our set is small due to limited computational resources; it only serves to illustrate the principle of evaluating performance for every step and variance in



rankings. In a prospective search, particularly in an industry setting, much larger sets with many more actives and decoys would have to be employed with dilutions up to 1:100 or better still 1:1000. Given access to cloud computing, such sets would be trivial to process (Lawrenz et al., 2015; Hande and Tamboli, 2024). Moreover, data sets would often have to be specifically prepared for the target at hand. For example, when targeting the pore of a GPCR ion channel, practitioners should assemble a large and diverse set of GPCR structures and binders. In such a scenario, globular proteins such as proteases or globins would be of little to no value and *vice versa*. Moreover, when targeting ion channels, practitioners would be advised to treat closed and open structures separately. Here, we can only provide general advice; we cannot hope to cover every eventuality. Ample positive data is available in PDBbind and in the literature. From then on, it is trivial to generate practically unlimited amounts of random protein–ligand pairs and isomers.

Ivanov (2026) showcases code to generate and process isomers on a large scale with MAYGEN. It demonstrates that isomerization can be an efficient tool to generate decoys that match the actives more closely than the experimental decoys in key molecular properties such as molecular weight, charge, surface, numbers of hydrogen bond donors, acceptors, and rotatable bonds. Crucially, it shows that isomerization can produce decoys that are better positioned than the experimental decoys in property hyperspace, i.e., closer to the actives (Ivanov, 2026).

We present a means of thoroughly and rigorously assessing the value added by every step in a VHTS pipeline and showcase how this matter should be investigated. Any pipeline should be tested and validated in a similar manner and every step in it should be modified until it reliably and consistently improves enrichment and recovery, otherwise it should be removed. Methods which underperform should be swapped out in favor of ones

which improve recovery compared to the previous step, which does not appear to be the case with Gnina and MM-GB(PB)SA in this instance. This is noteworthy because MM-GB(PB)SA results (Maraf et al., 2024; Poli et al., 2020; Kim et al., 2024; Sahakyan, 2021; Yazdani et al., 2023; Alturki et al., 2022; Ahmed et al., 2023; Okimoto et al., 2009) or even cruder metrics like atomic RMSDs or RMSFs (root mean square fluctuations (Menchon et al., 2018)) are employed as mid-to-late-stage filters in virtual high-throughput screens. Clearly, such pipelines should be validated and recalibrated using the approach we propose here. For the PDBbind set, for example, ranking ligands by RMSD is little different than ranking by MM-GB(PB)SA; for the MAYGEN set, ranking by RMSD arguably outperforms MM-GB(PB)SA, but exhibits a lot more variability from replica to replica. Here, MM-GBSA achieves recovery comparable to MM-PBSA, despite being a lower-level-of-theory technique, involving more simplifications, and costing about an order of magnitude less in calculations. This is consistent with our previous results on RNase A–ligand binding (Ivanov et al., 2019a), where MM-GBSA performed similarly to MM-PBSA on calculating relative ligand affinities. We once again stress that the tools we use here—MAYGEN, Gnina, and MM-GB(PB)SA—only serve as a story-telling device; they are mere illustrations of a more important principle. It is the *principle* that must be learned not the *example*. Use of the tools does not imply endorsement or recommendation, neither for, nor against. Indeed, the search for the optimal VHTS protocol very much remains ongoing.

While one can choose to omit either or both the isomerizations and randomizations during VHTS pipeline validation and use only set(s) one is interested in, we strongly encourage including both an isomerization and randomization set in the validation. By virtue of preserving heavy atom composition, isomerization sets are likely to be much less biased than sets of completely unrelated actives and inactives, although discrepancies and imperfections cannot be avoided completely. Thus, this is an astute means of *minimizing* bias. From this vantage point, we highlight two reasons for including the PDBbind randomizations in our analysis. First, results from the isomerization set (which is known to have ligand bias) must be compared to results from the PDBbind randomizations which do not have any ligand bias (because the same set of molecules serve as actives as well as inactives). Thus, if results are consistent between these two very different sets and whatever proprietary or publicly available set is being used, one has good reason to believe that what is being revealed is the true recovery capacity of the pipeline rather than some random outcome or the result of a bias in the isomerization or proprietary/open source set. Second, the protein–ligand randomizations constitute the ultimate benchmark for ABFE calculations which are believed to be approaching maturity. If this is truly the case, any sufficiently mature and capable ABFE pipeline (Ries et al., 2024; Heinzelmann and Gilson, 2021; Heinzelmann et al., 2024) should consistently outperform docking and end-state free energy techniques which involve multiple simplifications and approximations. Pairing ligands to different proteins and correctly ranking their affinities is a harder challenge and a more stringent test than ranking ligands with the same protein. For example, a benzodiazepine simulated with the steroid receptor or a steroid molecule simulated with the

benzodiazepine receptor should produce less favorable $\Delta G_{\text{calculated}}$ values than the cognate binding pairs. By their very nature, ABFE calculations are the only technique which should be expected to consistently succeed at this more stringent test. However, no such tests have thus far been carried out and it is presently unclear if current ABFE pipelines would pass even this conceptually simple check.

With deterministic sampling tools, e.g., OpenEye's OMEGA (Hawkins et al., 2010) and FRED (McGann, 2011), running the same task with the same settings multiple times produces identical results (within machine precision), so there is no added benefit, only added cost, to performing the same task more than once. This is not the case with stochastic sampling tools and algorithms, such as Gnina or Molsoft's ICM (Abagyan et al., 1994), which randomly sample possible ligand conformations and which might uncover new, more favorable, and previously unsampled conformations, given more runtime. It is then worth considering where molecular dynamics lies on the spectrum between fully deterministic and stochastic sampling techniques. In the hypothetical case of perfect sampling, where the entirety of available conformational space has been explored, MD results would be practically identical regardless of the starting conformation or velocities. Such simulations are said to be ergodic (Karplus and McCammon, 2002). In practice, however, MD runs often sample only parts of the available conformational space, with different runs reaching different subregions with varying degrees of overlap between the conformations sampled in different runs. This means that MD simulations are often quasi-ergodic rather than ergodic (Chipot, 2023). This makes it necessary to estimate the variability in results and rankings one can expect to obtain with stochastic (e.g., Gnina or ICM docking) or quasi-ergodic techniques (e.g., MM-GBSA or ABFE calculations). We demonstrate how this variability can be estimated. Our work on the MAYGEN set demonstrates that with MM-GBSA or MM-PBSA, rankings can vary up to 10 positions, potentially even more, which is a significant margin given that the set contains 78 molecules in total. Moreover, we see that RMSD is too crude of a metric—the rankings it produces can vary up to a third, potentially even half, of the span of the X axis. This is too wide a margin for a ranking to be considered reliable. Further, it is important to estimate how the variance in ranking scales with increasing dilution, as the actives:decoys ratio reaches and surpasses 1:50, 1:100, 1:1000, and beyond, reaching levels found in real-world screening decks. We leave this matter as the subject of future reports, by ourselves and hopefully others as well.

We also stress that the method we lay out here should be used to first calibrate VHTS pipelines on the target of interest, i.e., to assess the baseline enrichment and recovery metrics at every step given the specific target and ligand deck at hand. Among practitioners, it is known that different classes of proteins can vary greatly in terms of difficulty for hit discovery - kinases tend to be relatively easy (by the standards of hit discovery), proteases are usually more difficult, metalloenzymes are typically very difficult. Indeed, hit rates can vary by up to several orders of magnitude from target to target and from ligand deck to ligand deck. Thus, one should not expect good hit rates from one target to immediately translate to another. Rather, a baseline should be established for every class of target, ideally for every protein target at hand. Then, this baseline should be used to validate

any modifications to the pipeline. Hydrogen mass repartitioning, for example, is often employed to speed up MD simulations (Hopkins et al., 2015). Whether or not it affects recovery, in what way, and in what context, however, have thus far not been explored. Other avenues which warrant further exploration are the influence of adding ligand solvation energies (Ivanov, 2024a) to pipelines which do not have them, ligand strain energies (Tong and Zhao, 2021) and rototranslational entropies (Irudayam and Henchman, 2009; Chang et al., 2007) to name a few, many more can be envisaged and tested. Ranking and recovery analysis, as presented here, can also be used in force field development (and method development more broadly) as a test that is orthogonal (and hence independent) from the metrics used to tune and validate force fields or water models such as enthalpies of melting or evaporation (Warshel and Lifson, 1970), melting points of water (Xiong et al., 2020), charge distributions (Izadi et al., 2014), radial distribution functions (RDFs, usually the oxygen–oxygen RDF for water models (Izadi and Onufriev, 2016)), QM-derived energies and parameters (Tian et al., 2020), etc. Results from both the PDBBind and MAYGEN sets clearly demonstrate that the Gnina CNN pose score metric achieves superior recovery to Gnina affinity and Gnina CNN affinity. This is the exact same outcome that was observed during Gnina's development (McNutt et al., 2021), despite the fact that the Gnina developers used different proteins, ligands, and protein–ligand pairs from the ones used here. In fact, the Gnina developers did not even have access to the MAYGEN decoys used herein. This work, therefore, corroborates and validates the developers' decision to make pose score the default selection criterion using a test that is very different from the one used during Gnina's development (McNutt et al., 2021). More generally, it illustrates that testing and validation should be done using at least two orthogonal sets, and shows that robust and reliable conclusions can only be drawn after arriving at the same outcome from very different starting points.

Supplementary Figure S2 shows a breakdown of the rankings afforded by some of the terms that go into the final MM-GBSA and MM-PBSA results and compares them to the final MM-GBSA and MM-PBSA rankings. The figure lists the rankings by polar solvation energy (EGB), nonpolar solvation energy (ESURF), total gas phase (G GAS), and total solvation energy (G SOLV) from the MM-GBSA calculations. For MM-PBSA, it lists polar solvation (EPB), nonpolar solvation (ENPOLAR), total solvation (G SOLV), as well as the dispersion terms (EDISPER) that go into the ENPOLAR terms in MM-PBSA, as well as the gas phase energies (G GAS) which are the same for MM-GBSA and MM-PBSA. We see that the combination of the terms that produces the final MM-GBSA and MM-PBSA results and rankings has superior recovery to the individual terms, thereby offering an orthogonal verification of the end-state free energy techniques and their underlying theory. It has also not evaded our notice that some of the least theoretically rigorous terms in MM-GB(PB)SA—the solvation terms and the polar solvation terms in particular—have some of the lowest recovery capacities. Furthering the previous line of reasoning, we conjecture that using more rigorous solvation schemes (Ivanov, 2024a) and adding a term for entropy (potentially more than one, e.g., ligand configurational entropy (Chang et al., 2007), side-chain conformational entropy (Doig and Sternberg, 1995),

rototranslational entropy (Fogolari et al., 2018), etc.) that is (are) of sufficient quality and theoretical rigor should further improve recovery to the point where it consistently outperforms docking, potentially becoming competitive with alchemical ABFEs (Lee et al., 2020; Macchiagodena et al., 2020; Cournia et al., 2020; Aldeghi et al., 2016).

We caution the novice practitioner to carefully consider the output from VHTS tools and pipelines and not overrely on and overtrust their output. As an example, note that the majority of protein–ligand pairs in both the PDBbind- and MAYGEN-derived sets have a predicted Gnina affinity below -6 kcal/mol (see the left panels in Figure 1 and the left panel in Figure 3A). Despite the fact that the results are given in kcal/mol, these should not be viewed as absolute binding free energies (which would correspond to dissociation constants below 40 μ M which is appreciable binding) but relative to each other at best. Indeed, given the inherently low positive rates for random protein–ligand pairs, it is likely that not even one protein–ligand pair other than our designated positives have any appreciable affinity for each other.

With the present report, we hope to focus the attention of the research community on the need for sufficient and high quality negative data. Again, we do not view our current data sets as definitive or conclusive; we are merely laying out the blueprint for larger studies with positive:negative ratios exceeding 1:100, hopefully 1:1000. As a further means of making the validation more stringent and rigorous, protein structures could be replaced by alternative ones (Hahn et al., 2022) such that only crossdocking is performed. It is well established that redocking is an easier task than crossdocking (Röhrig et al., 2023). This is also evident from our own work, where ranking by any parameter produces superior recovery in cases of redocking versus crossdocking (Supplementary Figure S3). To make the test even harder and the validation still more stringent, the pipeline could be initiated using apo- rather than holostructures (Zhang et al., 2023).

Furthermore, high quality negative data is needed not only to accurately estimate the capabilities of VHTS pipelines in a hit discovery setting; it is also necessary to fairly estimate the capabilities of relative binding free energy (RBFE) methods in hit-to-lead progression and lead optimization. Strikingly, examples of machine learning algorithms trained exclusively on positive data have been published (Jiménez et al., 2018). This is completely unrealistic in drug discovery where the supermajority of ligands are decoys, i.e., do not actually bind the target. If all ligands an ML model or artificial intelligence (AI (Blaschke et al., 2020; Guo et al., 2021; Arús-Pous et al., 2020; Fialková et al., 2021; Guo et al., 2023; Sundin et al., 2022; Moore et al., 2022; Wang et al., 2022)) encounters during training are binders, this reinforces the utterly incorrect implicit supposition that every ligand the model encounters is a binder, even if only a weak one (Leckband and Sivasankar, 1999). Similarly, instances of RBFE or even ABFE studies conducted on data sets that do not have any negative data points have been published (Aldeghi et al., 2016; Moore et al., 2023; Ross et al., 2023; Aldeghi et al., 2017). Worse still, in many cases most of the ΔG_{bind} values are much lower than 0 kcal/mol ($\Delta G_{\text{bind}} \ll 0$ kcal/mol) or are very few in number and span a narrow affinity range (Fratev et al., 2018). Studies like these do not assess the tools'

ability to identify nonbinding congeners. We, therefore, recommend that negative data of sufficient quality and quantity in the form of nonbinding congeners be added to the best practices requirements for carrying out protein–ligand RBFE benchmarks (Hahn et al., 2022) and commend reports that have data points with ΔG_{bind} below 0 but close to 0, as well as data points with $\Delta G_{\text{bind}} > 0$ kcal/mol (Wade et al., 2022; Zhang et al., 2021). We emphasize that for RBFE studies, nonbinding congeners should be used as negative data not random isomers.

Our work demonstrates that despite the boastful reports of hit discovery using VHTS that abound in industry and academia alike, VHTS hit discovery is still very much reliant on the underlying random hit rate for the screening deck at hand. At best, VHTS pipelines only provide enrichment on top of that. Note, for example, that in the MAYGEN set, many of the industry-wide tools and techniques fail to place even 1 active in the top 10% (which is 8 ligands in this case). Indeed, out of a set of 78 ligands with 3 actives, only Gnina CNN pose score consistently ranks two of the actives in the top 10% (see Figure 3). Almost all other metrics—Gnina affinity, MM-GBSA, MM-PBSA, and ligand RMSD—would likely fail to place even one active in the top 10% even in such a small set where the ratio of actives to inactives is several orders of magnitude greater than what is likely to be found in real screening decks. While it could be argued that there are many more tools, techniques, and modifications to the workflow that could be attempted, these results certainly warrant a healthy amount of caution and skepticism. Given that some of the most common tools and approaches do not place even 1 active in the top 1% out of 75 decoys that are utterly structurally different from the actives, how likely is it that another tool or method would be able to find the same active among 1 million or 1 billion decoys, some of which will be very similar to the active, some will be very different, with many more anywhere in between? And even if the active is within the top 1% out of millions to billions of decoys, it could still be buried under thousands to millions, potentially tens of millions of decoys. Clearly, finding 1 active out of 1M+ decoys is a very tall order, and what hits VHTS is currently furnishing are likely thanks to the random hit rate (probably within the 0.1%–1% range in most cases), with whatever enrichment it provides added on top of that. Accordingly, in the CACHE challenges, hit rates in a truly prospective setting have thus far tended to be between 0% and 1% (Herasymenko et al., 2025) with teams rarely exceeding success rates of 5% (Gutkin et al., 2024; Mettu et al., 2025; Eguida et al., 2024). Notably, in the first CACHE challenge, a docking-based search (Dunn et al., 2024) was tied for first place with pipelines based on much more sophisticated algorithms involving molecular dynamics, ABFEs, machine learning or deep learning. This is a clear indication that these methods and algorithms often do not work as well as their developers and users would hope and that much testing and validation remain ahead.

We also note that in its current implementation, MAYGEN produces a lot of unphysical structures that are allowed by sheer element valences but disallowed by other constraints. For example, it tends to produce a lot of structures with nonlinear consecutive double bonds, especially in rings, which are unrealistic. This creates the need to manually filter out such structures. We are hopeful that the MAYGEN development team will address these shortcomings and further enhance its utility and practicality.

Lastly, we reiterate that any criticism levied herein is so levied with only positivism and constructivism as its ultimate end. Herein, we build upon the ideas, tools, and work of others in a collaborative effort to make a meaningful improvement and move the field toward maturity and delivering the new medicines so very much needed in the 21st century.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Author contributions

SMI: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review and editing.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This study is financed by the European Union-NextGenerationEU, through the National Recovery and Resilience Plan of the Republic of Bulgaria, project BG-RRP-2.004-0004-C01 “Strategic research and innovation program for the development of Medical university - Sofia.” The *in silico* calculations were performed at the Centre of Excellence for Informatics and ICT, supported by the Science and Education for Smart Growth Operational Program and co-financed by the European Union through the European Structural and Investment Funds (Grant No. BG16RFPR002-1.014-0018-C01/2025). The funder had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements

This work has been published as a preprint at (Ivanov, 2024b).

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of

artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the

reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fbinf.2026.1756279/full#supplementary-material>

References

- Abagyan, R., Kuznetsov, D., and Totrov, M. (1994). ICM - New method for protein modeling and design: applications to docking and structure prediction from the distorted native conformation. *J. Comput. Chem.* 15 (5), 488–506. doi:10.1002/jcc.540150503
- Ackloo, S., Al-awar, R., Amaro, R. E., Arrowsmith, C. H., Azevedo, H., Batey, R. A., et al. (2022). CACHE (critical assessment of computational Hit-finding experiments): a public-private partnership benchmarking initiative to enable the development of computational methods for hit-finding. *Nat. Rev. Chem.* 6 (4), 287–295. doi:10.1038/s41570-022-00363-z
- Ahmed, M., Maldonado, A. M., and Durrant, J. D. (2023). From byte to bench to bedside: molecular dynamics simulations and drug discovery. *BMC Biol.* 21 (1), 1–4. doi:10.1186/s12915-023-01791-z
- Aldeghi, M., Heifetz, A., Bodkin, M. J., Knapp, S., and Biggin, P. C. (2016). Predictions of ligand selectivity from absolute binding free energy calculations. *J. Am. Chem. Soc.* 139 (2), 946–957. doi:10.1021/jacs.6b11467
- Aldeghi, M., Bodkin, M. J., Knapp, S., and Biggin, P. C. (2017). Statistical analysis on the performance of molecular mechanics poisson-Boltzmann surface area versus absolute binding free energy calculations: bromodomains as a case Study. *J. Chem. Inf. Model* 57 (9), 2203–2221. doi:10.1021/acs.jcim.7b00347
- Aldeghi, M., Bluck, J. P., and Biggin, P. C. (2018). Absolute alchemical free energy calculations for ligand binding: a beginner's guide. *Methods Mol. Biol.* 1762, 199–232. doi:10.1007/978-1-4939-7756-7_11
- Alturki, N. A., Mashraqi, M. M., Alzamami, A., Alghamdi, Y. S., Alharthi, A. A., Asiri, S. A., et al. (2022). *In-Silico* screening and molecular dynamics simulation of drug bank experimental compounds against SARS-CoV-2. *Molecules* 27 (14), 1–12. doi:10.3390/molecules27144391
- Amaral, P., Carbonell-Sala, S., De La Vega, F. M., Faial, T., Frankish, A., Gingeras, T., et al. (2023). The status of the human gene catalogue. *Nature* 622 (7981), 41–47. doi:10.1038/s41586-023-06490-x
- Apweiler, R., Bairoch, A., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., et al. (2004). UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* 32 (Database issue), D115–D119. doi:10.1093/nar/gkh131
- Arús-Pous, J., Patronov, A., Bjerrum, E. J., Tyrchan, C., Reymond, J. L., Chen, H., et al. (2020). SMILES-based deep generative scaffold decorator for *de-novo* drug design. *J. Cheminform* 12 (1), 1–18. doi:10.1186/s13321-020-00441-8
- Atanasova, M. (2025). Small-molecule inhibitors of amyloid beta: insights from molecular dynamics—part A: endogenous compounds and repurposed drugs. *Pharmaceuticals* 18 (3), 1–62. doi:10.3390/ph18030306
- Atanasova, M., Dimitrov, I., Ivanov, S., Georgiev, B., Berkov, S., Zheleva-Dimitrova, D., et al. (2022). Virtual screening and hit selection of natural compounds as acetylcholinesterase inhibitors. *Molecules* 27 (10), 1–19. doi:10.3390/molecules27103139
- Atanasova, M., Stavrakov, G., Philipova, I., Georgiev, B., Bastida, J., Doytchinova, I., et al. (2024). AChE inhibitory activity of N-substituted natural galanthamine derivatives. *Bioorg. Med. Chem. Lett.* 112 (July), 129937. doi:10.1016/j.bmcl.2024.129937
- Austin, N. D., Sahinidis, N. V., and Trahan, D. W. (2016). Computer-aided molecular design: an introduction and review of tools, applications, and solution techniques. *Chem. Eng. Res. Des.* 116, 2–26. doi:10.1016/j.cherd.2016.10.014
- Badertscher, M., Korytko, A., Schulz, K., Madison, M., Munk, M. E., Portmann, P., et al. (2000). Assemble 2.0: a structure generator. *Chemom. Intell. Lab. Syst.* 51 (1), 73–79. doi:10.1016/s0169-7439(00)00056-3
- Baell, J. B., and Holloway, G. A. (2010). New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J. Med. Chem.* 53 (7), 2719–2740. doi:10.1021/jm901137j
- Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81, 3684–3690. doi:10.1063/1.448118
- Bertrand, H., and Triballeau, N. (2005). Virtual screening workflow development guided by the ROC curve approach. Application to high-throughput docking on metabotropic glutamate receptor subtype 4. *J. Med. Chem.* 48 (7), 2534–2547. doi:10.1021/jm049092j
- Bhakat, S., and Söderhjelm, P. (2017). Resolving the problem of trapped water in binding cavities: prediction of host-guest binding free energies in the SAMPL5 challenge by funnel metadynamics. *J. Comput. Aided Mol. Des.* 31 (1), 119–132. doi:10.1007/s10822-016-9948-6
- Blaschke, T., Arús-Pous, J., Chen, H., Margreitter, C., Tyrchan, C., Engkvist, O., et al. (2020). REINVENT 2.0 – an AI Tool for *de novo* Drug Design. *J. Chem. Inf. Model* 60 (12), 5918–5922. doi:10.1021/acs.jcim.0c00915
- Boger, D. L., Goldberg, J., and Andersson, C. M. (1999). Solution phase combinatorial synthesis of biaryl libraries employing heterogeneous conditions for catalysis and isolation with size exclusion chromatography for purification. *J. Org. Chem.* 64 (7), 2422–2427. doi:10.1021/jo982234v
- Bowers, K. J., Sacerdoti, F. D., Salmon, J. K., Shan, Y., Shaw, D. E., Chow, E., et al. (2006). Molecular dynamics—Scalable algorithms for molecular dynamics simulations on commodity clusters. *Proc. 2006 ACM/IEEE Conf. Supercomput - S. C. '06* 43. doi:10.1109/SC.2006.54
- Brandsdal, B. O., Österberg, F., Almlöf, M., Freierberg, I., Luzhkov, V. B., and Åqvist, J. (2003). Free energy calculations and ligand binding. *Adv. Protein Chem.* 66, 123–158. doi:10.1016/s0065-3233(03)66004-3
- Brooks, W. H. (2024). Polyamine dysregulation and nucleolar disruption in Alzheimer's disease. *J. Alzheimer's Dis.* 98 (3), 837–857. doi:10.3233/JAD-231184
- Carlsson, J., and Lutten, A. (2024). Structure-based virtual screening of vast chemical space as a starting point for drug discovery. *Curr. Opin. Struct. Biol.* 87, 102829. doi:10.1016/j.sbi.2024.102829
- Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., et al. (2005). The amber biomolecular simulation programs. *J. Comput. Chem.* 26 (16), 1668–1688. doi:10.1002/jcc.20290
- Chaikuad, A., Petros, A. M., Fedorov, O., Xu, J., and Knapp, S. (2014). Structure-based approaches towards identification of fragments for the low-druggability ATAD2 bromodomain. *Medchemcomm* 5 (12), 1843–1848. doi:10.1039/c4md00237g
- Chang, C. E. A., Chen, W., and Gilson, M. K. (2007). Ligand configurational entropy and protein binding. *Proc. Natl. Acad. Sci. U. S. A.* 104 (5), 1534–1539. doi:10.1073/pnas.0610494104
- Chaput, L., Martinez-Sanz, J., Saettel, N., and Mouawad, L. (2016). Benchmark of four popular virtual screening programs: construction of the active/decoy dataset remains a major determinant of measured performance. *J. Cheminform* 8 (1), 1–17. doi:10.1186/s13321-016-0167-x
- Childers, M. C., Daggett, V., Beck, D. A. C., Benson, N. C., Jonsson, A. L., Merkley, E. D., et al. (2017). Insights from molecular dynamics simulations for computational protein design. *Mol. Syst. Des. Eng.* 2 (1), 9–33. doi:10.1039/C6ME00083E
- Chipot, C. (2023). Free energy methods for the description of molecular processes. *Annu. Rev. Biophys.* 52, 113–138. doi:10.1146/annurev-biophys-062722-093258
- Ciccotti, G., and Ryckaert, J. P. (1986). Molecular dynamics simulation of rigid molecules. *Comput. Phys. Rep.* 4 (6), 346–392. doi:10.1016/0167-7977(86)90022-5
- Clark, J. A., Kusy, D., Vakuliuk, O., Krzeszewski, M., Kochanowski, K. J., Koszarna, B., et al. (2023). The magic of biaryl linkers: the electronic coupling through them defines the propensity for excited-state symmetry breaking in quadrupolar acceptor-donor-acceptor fluorophores. *Chem. Sci.* 14 (46), 13537–13550. doi:10.1039/d3sc03812b
- Cournia, Z., Allen, B. K., Beuming, T., Pearlman, D. A., Radak, B. K., and Sherman, W. (2020). Rigorous free energy simulations in virtual screening. *J. Chem. Inf. Model* 60 (9), 4153–4169. doi:10.1021/acs.jcim.0c00116
- Dahlgren, M. K., Schyman, P., Tirado-Rives, J., and Jorgensen, W. L. (2013). Characterization of biaryl torsional energetics and its treatment in OPLS all-atom force fields. *J. Chem. Inf. Model* 53 (5), 1191–1199. doi:10.1021/ci4001597

- Darden, T., York, D., and Pedersen, L. (1993). Particle mesh ewald: an N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98 (1993), 10089–10092. doi:10.1063/1.464397
- Demont, E. H., Chung, C. W., Furze, R. C., Grandi, P., Michon, A. M., Wellaway, C., et al. (2015). Fragment-based discovery of low-micromolar ATAD2 bromodomain inhibitors. *J. Med. Chem.* 58 (14), 5649–5673. doi:10.1021/acs.jmedchem.5b00772
- Dodd, T., Botto, M., Paul, F., Fernandez-Leiro, R., Lamers, M. H., and Ivanov, I. (2020). Polymerization and editing modes of a high-fidelity DNA polymerase are linked by a well-defined path. *Nat. Commun.* 11 (1), 1–11. doi:10.1038/s41467-020-19165-2
- Doig, A. J., and Sternberg, M. J. (1995). Side-chain conformational entropy in protein folding. *Protein Sci.* 4 (11), 2247–2251. doi:10.1002/pro.5560041101
- Dolbois, A., Batiste, L., Wiedmer, L., Dong, J., Brüttsch, M., Huang, D., et al. (2020). Hitting a moving target: simulation and crystallography study of ATAD2 bromodomain blockers. *ACS Med. Chem. Lett.* 11 (8), 1573–1580. doi:10.1021/acmedchemlett.0c00800
- Draper, S. J., Sack, B. K., King, C. R., Nielsen, C. M., Rayner, J. C., Higgins, M. K., et al. (2018). Malaria vaccines: recent advances and new Horizons. *Cell Host Microbe* 24 (1), 43–56. doi:10.1016/j.chom.2018.06.008
- Dunn, I., Pirhadi, S., Wang, Y., Ravindran, S., Concepcion, C., and Koes, D. R. (2024). CACHE challenge #1: docking with GNINA is all you need. *J. Chem. Inf. Model* 64 (24), 9388–9396. doi:10.1021/acs.jcim.4c01429
- Eguida, M., Bret, G., Sindt, F., Li, F., Chau, I., Ackloo, S., et al. (2024). Subpocket similarity-based hit identification for challenging targets: application to the WDR domain of LRRK2. *J. Chem. Inf. Model* 64 (13), 5344–5355. doi:10.1021/acs.jcim.4c00601
- Fialková, V., Zhao, J., Papadopoulos, K., Engkvist, O., Bjerrum, E. J., Kogej, T., et al. (2021). LibINVENT: reaction-based generative scaffold decoration for *in silico* library design. *J. Chem. Inf. Model* 62 (9), 2046–2063. doi:10.1021/acs.jcim.1c00469
- Fogolari, F., Maloku, O., Dongmo Fomthum, C. J., Corazza, A., and Esposito, G. (2018). PDB2ENTROPY and PDB2TRENT: conformational and translational-rotational entropy from molecular ensembles. *J. Chem. Inf. Model* 58 (7), 1319–1324. doi:10.1021/acs.jcim.8b00143
- Fratev, F., Steinbrecher, T., and S6, J. (2018). Prediction of accurate binding modes using combination of classical and accelerated molecular dynamics and free-energy perturbation calculations: an application to toxicity studies. *ACS Omega* 3 (4), 4357–4371. doi:10.1021/acsomega.8b00123
- Genheden, S., and Ryde, U. (2015). The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov.* 10 (5), 1–13. doi:10.1517/17460441.2015.1032936
- Kim, S., and Vincent, W. R. (2013). Contemporary approaches for evidence-based pharmacotherapy. *J. Pharm. Pract.* 26 (2), 95–102. doi:10.1177/0897190012474232
- Good, A. C., and Oprea, T. I. (2008). Optimization of CAMD techniques 3. Virtual screening enrichment studies: a help or hindrance in tool selection? *J. Comput. Aided Mol. Des.* 22 (3–4), 169–178. doi:10.1007/s10822-007-9167-2
- Gugisch, R., Kerber, A., Kohnert, A., Laue, R., Meringer, M., and Rücker, C. (2015). Molgen 5.0, A molecular structure generator. *Adv. Math. Chem. Appl.* 1, 113–138.
- Guo, J., Janet, J. P., Bauer, M. R., Nittinger, E., Giblin, K. A., Papadopoulos, K., et al. (2021). DockStream: a docking wrapper to enhance *de novo* molecular design. *J. Cheminform* 13 (1), 1–21. doi:10.1186/s13321-021-00563-7
- Guo, J., Knuth, F., Margreitter, C., Janet, J. P., Papadopoulos, K., Engkvist, O., et al. (2023). Link-INVENT: generative linker design with reinforcement learning. *Digit. Discov.* 2 (2), 392–408. doi:10.1039/d2dd00115b
- Gutkin, E., Gusev, F., Gentile, F., Ban, F., Koby, S. B., Narangoda, C., et al. (2024). *In silico* screening of LRRK2 WDR domain inhibitors using deep docking and free energy simulations. *Chem. Sci.* 15 (23), 8800–8812. doi:10.1039/d3sc06880c
- Hahn, D. F., Bayly, C. I., Boby, M. L., Bruce Macdonald, H. E., Chodera, J. D., Gapsys, V., et al. (2022). Best practices for constructing, preparing, and evaluating protein-ligand binding affinity benchmarks. *Living J. Comput. Mol. Sci.* 4 (1), doi:10.33011/livecoms.4.1.1497
- Hande, A. S., and Tamboli, F. A. (2024). The role of cloud computing in modern IT infrastructure: opportunities and risks. *Int. J. Adv. Res. Sci. Commun. Technol.* 4 (1), 152–155. doi:10.48175/IJARSC19527
- Hawkins, P. C. D., Warren, G. L., Skillman, A. G., and Nicholls, A. (2008). How to do an evaluation: pitfalls and traps. *J. Comput. Aided Mol. Des.* 22 (3–4), 179–190. doi:10.1007/s10822-007-9166-3
- Hawkins, P. C. D., Skillman, A. G., Warren, G. L., Ellingson, B. A., and Stahl, M. T. (2010). Conformer generation with OMEGA: algorithm and validation using high quality structures from the protein databank and Cambridge structural database. *J. Chem. Inf. Model* 50 (4), 572–584. doi:10.1021/ci100031x
- Heinzelmann, G., and Gilson, M. K. (2021). Automation of absolute protein-ligand binding free energy calculations for docking refinement and compound evaluation. *Sci. Rep.* 11 (1), 1–18. doi:10.1038/s41598-020-80769-1
- Heinzelmann, G., Huggins, D. J., and Gilson, M. K. (2024). BAT2: an open-source tool for flexible, automated, and low cost absolute binding free energy calculations. *J. Chem. Theory Comput.* 20 (15), 6518–6530. doi:10.1021/acs.jctc.4c00205
- Heinzlmeier, S., Kudlinzki, D., Sreeramulu, S., Klaeger, S., Gande, S. L., Linhard, V., et al. (2016). Chemical proteomics and structural biology define EPHA2 inhibition by clinical kinase drugs. *ACS Chem. Biol.* 11 (12), 3400–3411. doi:10.1021/acscchembio.6b00709
- Heinzlmeier, S., Lohse, J., Treiber, T., Kudlinzki, D., Linhard, V., Gande, S. L., et al. (2017). Chemoproteomics-aided medicinal chemistry for the discovery of EPHA2 inhibitors. *ChemMedChem* 12 (12), 999–1011. doi:10.1002/cmdc.201700217
- Herasymenko, O., Silva, M., Abu-Saleh, A. A. A., Alvarado-huayhuaz, J., Arce, O. E. A., Armstrong, R. J., et al. (2025). CACHE challenge #2: targeting the RNA site of the SARS-CoV-2 helicase Nsp13. *J. Chem. Inf. Model* 65 (13), 6884–6898. doi:10.1021/acs.jcim.5c00535
- Hopkins, C. W., Le Grand, S., Walker, R. C., and Roitberg, A. E. (2015). Long-time-step molecular dynamics through hydrogen mass repartitioning. *J. Chem. Theory Comput.* 11 (4), 1864–1874. doi:10.1021/ct5010406
- Huang, N., Shoichet, B. K., and Irwin, J. J. (2006). Benchmarking sets for molecular docking. *J. Med. Chem.* 49 (23), 6789–6801. doi:10.1021/jm0608356
- Hussain, I., Capricho, J., and Yawer, M. A. (2016). Synthesis of biaryls via ligand-free Suzuki–Miyaura cross-coupling reactions: a review of homogeneous and heterogeneous catalytic developments. *Adv. Synth. Catal.* 358 (21), 3320–3349. doi:10.1002/adsc.201600354
- Irudayam, S. J., and Henschman, R. H. (2009). Entropic cost of protein-ligand binding and its dependence on the entropy in solution. *J. Phys. Chem. B* 113, 5871–5884. doi:10.1021/jp809968p
- Ivanov, S. M. (2024a). Calculated hydration free energies become less accurate with increases in molecular weight. *PLoS One* 19 (9), e0309996. doi:10.1371/journal.pone.0309996
- Ivanov, S. M. (2024b). An efficient computational chemistry approach to generating negative data for drug discovery pipeline validation. *ChemRxiv*. [preprint]. doi:10.26434/chemrxiv-2024-54m5x-v3
- Ivanov, S. M. (2026). *In silico* isomerization produces apt negative data for VHTS validation. *ChemRxiv Prepr.* doi:10.26434/chemrxiv-2026-bgc3j-v2
- Ivanov, S. M., Huber, R. G., Warwicker, J., and Bond, P. J. (2016). Energetics and dynamics across the Bcl-2-Regulated apoptotic pathway reveal distinct evolutionary determinants of specificity and affinity. *Structure* 24 (11), 2024–2033. doi:10.1016/j.str.2016.09.006
- Ivanov, S. M., Dimitrov, I., and Doytchinova, I. A. (2019a). Bridging solvent molecules mediate RNase A – ligand binding. *PLoS One* 14 (10), e0224271. doi:10.1371/journal.pone.0224271
- Ivanov, S. M., Huber, R. G., Alibay, I., Warwicker, J., and Bond, P. J. (2019b). Energetic fingerprinting of ligand binding to paralogous proteins: the case of the apoptotic pathway. *J. Chem. Inf. Model* 59 (1), 245–261. doi:10.1021/acs.jcim.8b00765
- Ivanov, S. M., Atanasova, M., Dimitrov, I., and Doytchinova, I. A. (2020). Cellular polyamines condense hyperphosphorylated Tau, triggering Alzheimer's disease. *Sci. Rep.* 10 (1), 1–13. doi:10.1038/s41598-020-67119-x
- Izadi, S., and Onufriev, A. V. (2016). Accuracy limit of rigid 3-point water models. *J. Chem. Phys.* 145 (7), 074501. doi:10.1063/1.4960175
- Izadi, S., Anandakrishnan, R., and Onufriev, A. V. (2014). Building water models: a different approach. *J. Phys. Chem. Lett.* 5 (21), 3863–3871. doi:10.1021/jz501780a
- Jakalian, A., Bush, B. L., Jack, D. B., and Bayly, C. I. (2000). Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. *J. Comput. Chem.* 21 (2), 132–146. doi:10.1002/(sici)1096-987x(20000130)21:2<132::aid-jcc5>3.3.co;2-g
- Jiménez, J., Škalič, M., Martínez-Rosell, G., and De Fabritiis, G. (2018). KDEEP: protein-ligand absolute binding affinity prediction via 3D-Convolutional neural networks. *J. Chem. Inf. Model* 58 (2), 287–296. doi:10.1021/acs.jcim.7b00650
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79, 926–935. doi:10.1063/1.445869
- Joung, I. S., and Cheatham, T. E. (2008). Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B* 112 (30), 9020–9041. doi:10.1021/jp8001614
- Karplus, M., and McCammon, J. A. (2002). Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* 9 (9), 646–652. doi:10.1038/nsb0902-646
- Kim, J., Jin, H., Kim, J., Cho, S. Y., Moon, S., Wang, J., et al. (2024). Leveraging the fragment molecular orbital and MM-GBSA methods in virtual screening for the discovery of novel non-covalent inhibitors targeting the TEAD lipid binding pocket. *Int. J. Mol. Sci.* 25 (10), 5358. doi:10.3390/ijms25105358
- Kumalo, H. M., Bhakat, S., and Soliman, M. E. (2016). Investigation of flap flexibility of β -secretase using molecular dynamic simulations. *J. Biomol. Struct. Dyn.* 34 (5), 1008–1019. doi:10.1080/07391102.2015.1064831
- Kundra, R., Zhang, H., Sheridan, R., Sirintrapun, S. J., Wang, A., Ochoa, A., et al. (2021). OncoTree: a cancer classification system for precision oncology. *JCO Clin. Cancer Inf.* 5 (5), 221–230. doi:10.1200/CCI.20.00108
- Lawrenz, M., Shukla, D., and Pande, V. S. (2015). Cloud computing approaches for prediction of ligand binding poses and pathways. *Sci. Rep.* 5 (1), 7918. doi:10.1038/srep07918

- Leckband, D., and Sivasankar, S. (1999). Forces controlling protein interactions: theory and experiment. *Colloids Surfaces B Biointerfaces* 14 (1–4), 83–97. doi:10.1016/s0927-7765(99)00027-2
- Lee, T.-S., Allen, B. K., Giese, T. J., Guo, Z., Li, P., Lin, C., et al. (2020). Alchemical binding free energy calculations in AMBER20: advances and best practices for drug discovery. *J. Chem. Inf. Model* 60 (11), 5595–5623. doi:10.1021/acs.jcim.0c00613
- Lionta, E., Spyrou, G., Vassiliadis, D., and Cournia, Z. (2014). Structure-based virtual screening for drug discovery: principles, applications and recent advances. *Curr. Top. Med. Chem.* 14 (16), 1923–1938. doi:10.2174/1568026614666140929124445
- Macchiagodena, M., Pagliai, M., Karrenbrock, M., Guarnieri, G., Iannone, F., and Procacci, P. (2020). Virtual double-system single-box: a nonequilibrium alchemical technique for absolute binding free energy calculations: application to ligands of the SARS-CoV-2 main protease. *J. Chem. Theory Comput.* 16 (11), 7160–7172. doi:10.1021/acs.jctc.0c00634
- Maffucci, I., and Contini, A. (2013). Explicit ligand hydration shells improve the correlation between MM-PB/GBSA binding energies and experimental activities. *J. Chem. Theory Comput.* 9 (6), 2706–2717. doi:10.1021/ct400045d
- Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, T. B. (2015). ff14SB: improving the accuracy of protein side chain and backbone parameters for ff99SB. *J. Chem. Theory Comput.* 11 (8), 3696–3713. doi:10.1021/acs.jctc.5b00255
- Maraf, M. B., Mountessou, B. Y. G., Hans Merlin, T. F., Ariane, P., Fekoua, J. N. N., Jean Yves, T. B., et al. (2024). Virtual screening, MMGBSA, and molecular dynamics approaches for identification of natural products from South African biodiversity as potential *Onchocerca volvulus* pi-class glutathione S-transferase inhibitors. *Heliyon* 10 (9), e29560. Contents. doi:10.1016/j.heliyon.2024.e29560
- Marković, A., Atanasova, M., Buyukliev, R., Bakalova, A., Šmelcerović, A., and Cherneva, E. (2024). 3'-Methyl-4-thio-1H-tetrahydropyranspiro-5'-hydantoin platinum complex as a novel deoxyribonuclease I inhibitor. *Pharmacia* 71, 1–7. doi:10.3897/pharmacia.71.e126246
- McGann, M. (2011). FRED pose prediction and virtual screening accuracy. *J. Chem. Inf. Model* 51 (3), 578–596. doi:10.1021/ci100436p
- McNutt, A. T., Francoeur, P., Aggarwal, R., Masuda, T., Meli, R., Ragoza, M., et al. (2021). GNINA 1.0: molecular docking with deep learning. *J. Cheminform* 13 (1), 1–20. doi:10.1186/s13321-021-00522-2
- Menchon, G., Maveyraud, L., and Czaplicki, G. (2018). Molecular dynamics as a tool for virtual ligand screening. *Methods Mol. Biol.* 1762, 145–178. doi:10.1007/978-1-4939-7756-7_9
- Mettu, A., Glavatskikh, M., Wang, X., Lara Ordóñez, A. J., Li, F., Chau, I., et al. (2025). Functionally active modulators targeting the LRRK2 WD40 repeat domain identified by FRASE-bot in CACHE challenge #1. *Chem. Sci.* 16 (8), 3430–3439. doi:10.1039/d4sc07532c
- Moore, J. H., Bauer, M. R., Guo, J., Patronov, A., Engkvist, O., and Margreitter, C. (2022). Icolos: a workflow manager for structure-based post-processing of *de novo* generated small molecules. *Bioinformatics* 38 (21), 4951–4952. doi:10.1093/bioinformatics/btac614
- Moore, J. H., Margreitter, C., Janet, J. P., Engkvist, O., de Groot, B. L., and Gapsys, V. (2023). Automated relative binding free energy calculations from SMILES to $\Delta\Delta G$. *Commun. Chem.* 6 (1), 1–12. doi:10.1038/s42004-023-00859-9
- Mysinger, M. M., Carchia, M., Irwin, J. J., and Shoichet, B. K. (2012). Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *J. Med. Chem.* 55 (14), 6582–6594. doi:10.1021/jm300687e
- Nasir, I. A., Yakubu, S., and Mustapha, J. O. (2017). Epidemiology and synergistic hepatopathology of malaria and hepatitis C virus coinfection. *Virol. Res. Treat.* 8, 1–4. doi:10.1177/1178122X17724411
- Niazi, S. K., and Mariam, Z. (2024). Computer-aided drug design and drug discovery: a prospective analysis. *Pharmaceuticals* 17 (1), 1–22. doi:10.3390/ph17010022
- Nuzillard, J.-M., and Massiot, G. (1991). Logic for structure determination. *Tetrahedron* 47 (22), 3655–3664. doi:10.1016/s0040-4020(01)80878-4
- Okimoto, N., Futatsugi, N., Fuji, H., Suenaga, A., Morimoto, G., Yanai, R., et al. (2009). High-performance drug discovery: computational screening by combining docking and molecular dynamics simulations. *PLoS Comput. Biol.* 5 (10), e1000528. doi:10.1371/journal.pcbi.1000528
- O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., and Hutchison, G. R. (2011). Open babel. *J. Cheminform* 3 (33), 1–14. doi:10.1186/1758-2946-3-33
- Paggi, J. M., Pandit, A., and Dror, R. O. (2024). The art and science of molecular docking. *Annu. Rev. Biochem.* 93 (1), 389–410. doi:10.1146/annurev-biochem-030222-120000
- Pereira, J. C., Caffarena, E. R., and Dos Santos, C. N. (2016). Boosting docking-based virtual screening with deep learning. *J. Chem. Inf. Model* 56 (12), 2495–2506. doi:10.1021/acs.jcim.6b00355
- Philipova, I., Atanasova, M., Mihaylova, R., Dailova-Barzeva, A., Ivanov, S. M., Simeonova, R. L., et al. (2025). Design, docking analysis, and structure–activity relationship of ferrocene-modified tyrosine kinase inhibitors: insights into BCR-ABL interactions. *Molecules* 30 (15), 3101. doi:10.3390/molecules30153101
- Poli, G., Granchi, C., Rizzolio, F., and Tuccinardi, T. (2020). Application of MM-PBSA methods in virtual screening. *Molecules* 25 (8), 1–19. doi:10.3390/molecules25081971
- Poncet-Montange, G., Zhan, Y., Bardenhagen, J. P., Petrocchi, A., Leo, E., Shi, X., et al. (2015). Observed bromodomain flexibility reveals histone peptide- and small molecule ligand-compatible forms of ATAD2. *Biochem. J.* 466, 337–346. doi:10.1042/BJ20140933
- Qiu, C., Kivipelto, M., and Von Strauss, E. (2009). Epidemiology of Alzheimer's disease: occurrence, determinants, and strategies toward intervention. *Dialogues Clin. Neurosci.* 11 (2), 111–128. doi:10.31887/DCNS.2009.11.2/cqiu
- Ries, B., Alibay, I., Anand, N. M., Biggin, P. C., and Magarkar, A. (2024). Automated absolute binding free energy calculation workflow for drug discovery. *J. Chem. Inf. Model* 64 (14), 5357–5364. doi:10.1021/acs.jcim.4c00343
- Rohrer, S. G., and Baumann, K. (2009). Maximum unbiased validation (MUV) data sets for virtual screening based on PubChem bioactivity data. *J. Chem. Inf. Model* 49 (2), 169–184. doi:10.1021/ci8002649
- Röhrig, U. F., Goullieux, M., Bugnon, M., and Zoete, V. (2023). Attracting cavities 2.0: improving the flexibility and robustness for small-molecule docking. *J. Chem. Inf. Model* 63 (12), 3925–3940. doi:10.1021/acs.jcim.3c00054
- Ross, G. A., Lu, C., Scarabelli, G., Albanese, S. K., Houang, E., Abel, R., et al. (2023). The maximal and current accuracy of rigorous protein-ligand binding free energy calculations. *Commun. Chem.* 6 (1), 1–12. doi:10.1038/s42004-023-01019-9
- Sahakyan, H. (2021). Improving virtual screening results with MM/GBSA and MM/PBSA rescoring. *J. Comput. Aided Mol. Des.* 35 (6), 731–736. doi:10.1007/s10822-021-00389-3
- Schlender, M., Hernandez-Villafuerte, K., Cheng, C. Y., Mestre-Ferrandiz, J., and Baumann, M. (2021). How much does it cost to research and develop a new drug? A systematic review and assessment. *Pharmacoeconomics* 39 (11), 1243–1269. doi:10.1007/s40273-021-01065-y
- Sertkaya, A., Beleche, T., Jessup, A., and Sommers, B. D. (2024). Costs of drug development and research and development intensity in the US, 2000–2018. *JAMA Netw. Open* 7 (6), e2415445. doi:10.1001/jamanetworkopen.2024.15445
- Spassov, D. S. (2024). Binding affinity determination in drug design: insights from lock and key, induced fit, conformational selection, and inhibitor trapping models. *Int. J. Mol. Sci.* 25 (13), 7124. doi:10.3390/ijms25137124
- Spassov, D. S., Atanasova, M., and Doytchinova, I. (2024). Inhibitor trapping in kinases. *Int. J. Mol. Sci.* 25 (6), 3249. doi:10.3390/ijms25063249
- Stammers, E., Parsons, C. D., Clayden, J., and Lennox, A. J. J. (2023). Electrochemical synthesis of biaryls by reductive extrusion from N,N'-diaryliureas. *Nat. Commun.* 14 (1), 2–7. doi:10.1038/s41467-023-40237-6
- Stavrakou, G., Philipova, I., Lukarski, A., Atanasova, M., Zheleva, D., Zhivkova, Z. D., et al. (2020). Galantamine-curcumin hybrids as dual-site binding acetylcholinesterase inhibitors. *Molecules* 25 (15), 1–18. doi:10.3390/molecules25153341
- Stein, R. M., Yang, Y., Balius, T. E., O'Meara, M. J., Lyu, J. K., Young, J., et al. (2021). Property-unmatched decoys in docking benchmarks. *J. Chem. Inf. Model* 61 (2), 699–714. doi:10.1021/acs.jcim.0c00598
- Steinbeck, C. (2001). SENECA: a platform-independent, distributed, and parallel system for computer-assisted structure elucidation in organic chemistry. *J. Chem. Inf. Comput. Sci.* 41 (6), 1500–1507. doi:10.1021/ci000407n
- Su, Y., Gallicchio, E., Das, K., Arnold, E., and Levy, R. M. (2007). Linear interaction energy (LIE) models for ligand binding in implicit solvent: theory and application to the binding of NNRTIs to HIV-1 reverse transcriptase. *J. Chem. Theory Comput.* 3 (1), 256–277. doi:10.1021/ct600258e
- Sundin, I., Voronov, A., Xiao, H., Papadopoulos, K., Bjerrum, E. J., Heinonen, M., et al. (2022). Human-in-the-loop assisted *de novo* molecular design. *J. Cheminform* 14 (1), 1–16. doi:10.1186/s13321-022-00667-8
- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 71 (3), 209–249. doi:10.3322/caac.21660
- The Lancet Gastroenterology & Hepatology (2021). The hunt for a vaccine for hepatitis C virus continues. *Lancet Gastroenterol. Hepatol.* 6 (4), 253. doi:10.1016/S2468-1253(21)00073-X
- Tian, C., Kasavajhala, K., Belfon, K. A. A., Raguette, L., Huang, H., Miguez, A. N., et al. (2020). Ff19SB: amino-acid-specific protein backbone parameters trained against Quantum mechanics energy surfaces in solution. *J. Chem. Theory Comput.* 16 (1), 528–552. doi:10.1021/acs.jctc.9b00591
- Tong, J., and Zhao, S. (2021). Large-scale analysis of bioactive ligand conformational strain energy by *ab initio* calculation. *J. Chem. Inf. Model* 61 (3), 1180–1192. doi:10.1021/acs.jcim.0c01197
- Tran-Nguyen, V. K., Jacquemard, C., and Rognan, D. (2020). LIT-PCBA: an unbiased data set for machine learning and virtual screening. *J. Chem. Inf. Model* 60 (9), 4263–4273. doi:10.1021/acs.jcim.0c00155
- Vázquez, J., García, R., Llinares, P., Luque, F. J., and Herrero, E. (2024). On the relevance of query definition in the performance of 3D ligand-based virtual screening. *J. Comput. Aided Mol. Des.* 38 (1), 18. doi:10.1007/s10822-024-00561-5

- Wade, A. D., Bhati, A. P., Wan, S., and Coveney, P. V. (2022). Alchemical free energy estimators and molecular dynamics engines: accuracy, precision, and reproducibility. *J. Chem. Theory Comput.* 18 (6), 3972–3987. doi:10.1021/acs.jctc.2c00114
- Wallach, I., and Heifets, A. (2018). Most ligand-based classification benchmarks reward memorization rather than generalization. *J. Chem. Inf. Model* 58 (5), 916–932. doi:10.1021/acs.jcim.7b00403
- Wallach, I., Dzamba, M., and Heifets, A. (2015). AtomNet: a deep convolutional neural network for bioactivity prediction in. *Structure-based Drug Discov.*, 1–11. doi:10.48550/arXiv.1510.02855
- Wang, R., Fang, X., Lu, Y., and Wang, S. (2004). The PDBbind database: collection of binding affinities for protein–ligand complexes with known three-dimensional structures. *J. Med. Chem.* 47 (12), 2977–2980. doi:10.1021/jm030580l
- Wang, J. M., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004). Development and testing of a general amber force field. *J. Comput. Chem.* 25 (9), 1157–1174. doi:10.1002/jcc.20035
- Wang, Y., Bryant, S. H., Cheng, T., Wang, J., Gindulyte, A., Shoemaker, B. A., et al. (2017). PubChem BioAssay: 2017 update. *Nucleic Acids Res.* 45 (D1), D955–D963. doi:10.1093/nar/gkw1118
- Wang, P., Zheng, S., Jiang, Y., Li, C., Liu, J., Wen, C., et al. (2022). Structure-aware multimodal deep learning for drug-protein interaction prediction. *J. Chem. Inf. Model* 62 (5), 1308–1317. doi:10.1021/acs.jcim.2c00060
- Warshel, A., and Lifson, S. (1970). Consistent force field calculations. II. Crystal structures, sublimation energies, molecular and lattice vibrations, molecular conformations, and enthalpies of alkanes. *J. Chem. Phys.* 53 (2), 582–594. doi:10.1063/1.1674031
- Wicker, J. G. P., and Cooper, R. I. (2016). Beyond rotatable bond counts: capturing 3D conformational flexibility in a single descriptor. *J. Chem. Inf. Model* 56 (12), 2347–2352. doi:10.1021/acs.jcim.6b00565
- Wu, C., Lu, X., Lu, S., Wang, H., Li, D., Zhao, J., et al. (2022). Efficient detection of the alternative spliced human proteome using translome sequencing. *Front. Mol. Biosci.* 9 (June), 1–9. doi:10.3389/fmolb.2022.895746
- Xia, J., Tilahun, E. L., Reid, T. E., Zhang, L., and Wang, X. S. (2015). Benchmarking methods and data sets for ligand enrichment assessment in virtual screening. *Methods* 71, 146–157. doi:10.1016/j.ymeth.2014.11.015
- Xiong, Y., Shabane, P. S., and Onufriev, A. V. (2020). Melting points of OPC and OPC3 water models. *ACS Omega* 5 (39), 25087–25094. doi:10.1021/acsomega.0c02638
- Yazdani, B., Sirous, H., Brogi, S., and Calderone, V. (2023). Structure-based high-throughput virtual screening and molecular dynamics simulation for the discovery of novel SARS-CoV-2 NSP3 Mac1 domain inhibitors. *Viruses* 15 (12), 2291. doi:10.3390/v15122291
- Yirik, M. A., Sorokina, M., and Steinbeck, C. (2021). MAYGEN: an open-source chemical structure generator for constitutional isomers based on the orderly generation principle. *J. Cheminform* 13 (1), 1–14. doi:10.1186/s13321-021-00529-9
- Yu, J., Yan, C., Dodd, T., Tsai, C. L., Tainer, J. A., Tsutakawa, S. E., et al. (2023). Dynamic conformational switching underlies TFIIH function in transcription and DNA repair and impacts genetic diseases. *Nat. Commun.* 14 (1), 2758. doi:10.1038/s41467-023-38416-6
- Zhang, S., Hahn, D. F., Shirts, M. R., and Voelz, V. A. (2021). Expanded ensemble methods can be used to accurately predict protein-ligand relative binding free energies. *J. Chem. Theory Comput.* 17 (10), 6536–6547. doi:10.1021/acs.jctc.1c00513
- Zhang, B., Li, H., Yu, K., and Jin, Z. (2022). Molecular docking-based computational platform for high-throughput virtual screening. *CCF Trans. High. Perform. Comput.* 4 (1), 63–74. doi:10.1007/s42514-021-00086-5
- Zhang, Y., Vass, M., Shi, D., Abualrous, E., Chambers, J. M., Chopra, N., et al. (2023). Benchmarking refined and unrefined AlphaFold2 structures for hit discovery. *J. Chem. Inf. Model* 63 (6), 1656–1667. doi:10.1021/acs.jcim.2c01219