DINC-Ensemble: A Web Server for Docking Large Ligands Incrementally to an Ensemble of Receptor Conformations

Anja Conev[a], Jing Chen[b], Lydia E. Kavraki[a]

[a] *Computer Science Department, Rice University, 6100 Main Street, Houston, 77005, Texas, USA*
[b] *Molecular Sciences Software Institute, 1880 Pratt Drive, Suite 1100, Blacksburg, 24060, Virginia, USA*

Abstract

Protein-ligand docking aids structure-based drug discovery by computationally modelling protein-ligand interactions. DINC (Docking INCrementally) is one approach to molecular docking that improved the docking of large ligands using a parallelized incremental meta-docking. Traditional docking tools, including DINC, explore the flexibility of the ligand in a single receptor binding pocket assuming limited flexibility of the receptor backbone. This simplifying assumption narrows down the docking search space but hinders successful docking for flexible receptors. DINC-Ensemble implicitly considers receptor backbone flexibility by running DINC docking in parallel on different receptor conformations. Inputs to DINC-Ensemble include (1) a ligand and (2) a list of different receptor conformations. For each ligand-receptor pair DINC-Ensemble performs incremental meta-docking in parallel. As a result, multiple ligand poses are generated in the binding pockets of different receptor conformations. These poses are then ranked, and the lowest scoring pose is selected. Two main outputs provided by a successful run of DINC-Ensemble are (1) the best scoring ligand poses and (2) a ranked list of selected receptor conformations. The best scoring ligand pose can be used to understand the interactions between the receptor and the ligand that influence the binding. The ranked list of receptor conformations shows the best receptor conformation fit for a given ligand and can provide insight into ligand-induced conformational selection. We provide DINC-Ensemble as a Python package and a free web server at `https://dinc-ensemble.kavrakilab.rice.edu/`.

Keywords: protein-ligand docking, ensemble docking, conformational selection, molecular docking web server
PACS: 0000, 1111
2000 MSC: 0000, 1111

1. Introduction

Molecular docking is a common computational approach for modeling protein-ligand interactions and is used to assist structure-based drug discovery [1]. Some of the early attempts at molecular docking date back to 1990s [2, 3] and have since been consolidated with dozens of widely used implementations such as AutoDock4 [4], Vina [5], Glide [6], GOLD [7]. Outlined traditional docking methods model the ligand binding in two stages (1) sampling and (2) scoring. In the sampling stage, ligand pose is generated in the protein binding pocket using search algorithms that explore the search space of the ligand and its degrees of freedom (i.e., translation, rotation, and angles of rotatable bond). In the scoring stage, the energy of the protein-ligand complex is approximated using scoring functions. The best-scoring ligand pose is selected, and the score is presented as an approximated binding energy. To deal with the exponential growth of the search space for large ligands, DINC [8] (Docking Incrementally) has been developed as a meta-docking incremental approach. Molecular docking results can be used for structure-based drug design in the geometry prediction setting or high-throughput virtual screening [9]. The ultimate goal of molecular docking is to speed up drug discovery by allowing for computational analysis of the protein-ligand complex, therefore avoiding the costly and time-consuming experimental resolution of the complex. While traditional docking methods have had considerable success in the past they encode simplifying assumptions that hinder the accuracy of the tools.

One major assumption of the traditional docking tools is that the backbone of the receptor is rigid. This is in line with the early 'key-lock' [10] theory of molecular recognition. However, other theories including 'induced-fit' [11] and 'conformational selection' [12] suggest that receptors are flexible during recognition. While the rigid backbone receptor assumption narrows down the search space and speeds up the sampling, it hinders the performance of the methods for flexible protein receptors. Some methods, including AutoDock Vina [5] provide an option for side-chain flexibility by sampling

several binding-site side-chains. This approach is useful for modeling localized changes in the binding pocket but does not model larger changes in the protein backbone. Accounting for receptor backbone flexibility has been recognized as a major challenge in molecular docking and one of the proposed approaches to tackle it is ensemble docking [9].

Ensemble docking can be achieved by docking a ligand to multiple conformations of the protein receptor[13]. Receptor backbone flexibility is not explicitly sampled during docking but is taken into account implicitly. A few works have recently applied this approach showing promising results on a range of protein targets including G protein-coupled receptors [14], kinases (ALK, CDK2, VEGFR2, wee1) [15], Estrogen Receptor α [15] and SARS-Cov2-related targets (N-protein, S-protein) [16, 17]. Besides providing the ligand pose and predicted binding energy of the complex, ensemble docking allows for a broader analysis of conformational selection with functional implications. Despite its long history and recent promising applications, the software infrastructure for ensemble docking is not widely accessible. In contrast to dozens of different molecular docking servers and tools, the only free web servers providing ensemble docking functionalities at the moment are EDock-ML [18] and DockThor-VS [19]. EDock-ML implements a machine-learning approach to virtual screening with ensemble docking. DockThor-VS is a platform for large scale virtual screening of small ligands and contains an ensemble docking setup.

In addition to employing an ensemble docking strategy, DINC-Ensemble also allows ensemble docking of large ligands incrementally to the receptor ensemble. Ligand size represents another big challenge for molecular docking, as the search space for identifying the optimal ligand pose increases with the increase in the number of rotatable bonds (i.e. degrees of freedom). Many advanced docking tools exhibit lower performance when docking large ligands compared to their performance with small ligands [20].

In this work, we propose a novel ensemble parallel docking program DINC-Ensemble. DINC-Ensemble is well suited for docking large ligands to flexible receptors and it builds on our previous tool DINC [8] for docking large ligands incrementally. DINC is a meta-docking approach that incrementally performs docking using an underlying docking engine. Main improvements over the previous approach include: (1) providing receptor ensemble inputs and an ensemble analysis of docking results, (2) upgrading the docking engine to the most recent version of AutoDock Vina[21], and (3) providing a Python package and a web server. DINC-Ensemble will allow researchers to

explore the ensemble docking strategy for large ligands.

2. Materials and methods

2.1. Algorithm

DINC-Ensemble can tackle multiple protein receptor inputs in parallel and relies on an incremental meta-docking approach for docking the ligand to each of the binding pockets of the receptor. As input, DINC-Ensemble takes (1) a ligand structure file (.mol2 format), (2) $\#N$ receptor conformations (.pdb, .pdbqt formats), and (3) a set of docking parameters (see Supplementary Table 1). Ligand and receptor inputs are first prepared for docking. Next, the incremental docking procedure is started. As output, DINC-Ensemble provides (1) structures of a few best scoring ligand poses (.pdb files), (2) a list of energies and RMSD values for the ligand poses generated across receptor conformations, and (3) a list of receptor conformations ranked by the best score achieved for the ligand. Improvements of DINC-Ensemble in comparison to the previous iteration of the DINC algorithm are highlighted in the Supplementary Table 6.

2.1.1. Preparing the input ligand and receptor

First, we prepare the ligand for docking with DINC-Ensemble. Hydrogens and gasteiger charges are added to the ligand and types and unique names are assigned to ligand atoms using AutoDock Vina prepare_ligand protocol. With this, an initial torsion tree for the ligand is created and saved in a .pdbqt format. Next, we fragment the ligand for incremental docking. We first select the root atom and the root node of the torsion tree and then split the tree into $\#K$ incremental overlapping subtrees (fragments). Each fragment has a subset of the rotatable bonds active during docking limiting the search space at each step to a fixed size.

Next, we prepare receptors with AutoDock Vina prepare_receptor script. Missing hydrogens are added and gasteiger charges are calculated. We collect the binding box parameters for the reference receptor from user input (see Appendix Parameters). We align additional receptors in the ensemble to the reference receptor using sequence alignment followed by structural superposition (using PyMol's align implementation). We identify the binding site for all receptors using the input binding box.

With these steps, we prepare the input ligand, receptors, and binding sites for docking with DINC-Ensemble. Note that we recommend for users to provide protonated inputs, as we do not assign protonation states internally.

2.1.2. Incremental ensemble meta-docking

The incremental ensemble meta-docking approach used by DINC-Ensemble is described in Figure 1. The x-axis follows the incremental component of the algorithm, the y-axis follows the ensemble component, and the z-axis shows the replicas of parallelized meta-docking. Each orange cell in the figure corresponds to a single docking thread that runs a docking engine. Under the main schema in the Figure 1 is the outline of the algorithm is described with the three main components of (1) initializing docking threads, (2) running the docking threads, and (3) aggregating the results of the docking threads. These steps are repeated as the fragments grow and until the full ligand is reconstructed in each of the receptor conformations.

A docking thread is initialized by assigning a single receptor-ligand pair to the thread. The ligand assigned to a thread at incremental step i corresponds to the $i-th$ fragment generated in the preparation stage. At each incremental step, $\#K$ replicas are initialized for each of the $\#N$ receptor conformations resulting in $K * N$ active threads at a time. Each of the docking threads performs docking with the docking engine. The docking engine used by default is Vina 1.2.0 [21]. In each run, a docking thread generates several output ligand poses with corresponding energy scores. From one incremental time step to the next the results are aggregated by choosing a few unique minimum energy poses (clustering of poses is used to reduce redundancy). These "best" poses are used to initialize the next round of replicas.

Finally, full final ligand poses across multiple receptors in the ensemble are clustered and aggregated across different receptors. Ligand poses are ranked using the predicted energy scores as well as the cluster size. Besides the ranked ligand poses, input receptors are also ranked based on how well the ligand is docked to each of the receptors.

2.1.3. Parallel implementation

Docking across multiple replicas and receptors is parallelized for each iteration (column of yellow rectangles in Figure 1). Parallel processing is implemented using python's multiprocess library and single docking rounds (one fragment, one receptor and one replica) are implemented as a subprocess. Each subprocess in an iteration waits for others to complete (joining).

When the subprocess is joined with others from the same iteration, docking results are aggregated and clustered. Next iteration is initialized and new subprocesses are started. This parallelization allows users to perform ensemble docking without significant time costs. For example, if a user has access to a machine with 20 cores, they can perform docking across 5 receptors with 4 replicas each (using all $5 \times 4 = 20$ cores) in the same time frame as they would need to dock a ligand to a single receptor.

2.2. Benchmarking and case study

2.2.1. Incremental meta-docking benchmark

We first test the ability of DINC-Ensemble to dock large ligands to a single receptor. To that end, we construct Dataset 1 (Supplementary Tables 2,3,4) as a subset of the dataset compiled by Devaurs et al [20] to benchmark DINC-Ensemble performance in docking large ligands.

Dataset 1 is based on five datasets from the literature (Dhanik [8], Renard [22], LEADS [23], Hou [24], and PPDBench[25]). These datasets were used to evaluate large ligand docking performance in different settings. Devaurs et al [20] further filtered these datasets to identify challenging docking targets.

Dhanik dataset extracted large ligands from an older version of the PDBBind dataset [26]. Devaurs et al. further filtered the Dhanik dataset and identified challenging large ligands within it that have between 7 and 30 rotatable bonds. The Renard dataset was used to benchmark docking methods on small peptides. Devaurs et al. extracted a subset of peptides from this dataset having between 10 and 22 rotatable bonds that were challenging for docking with classical approaches. The LEADS dataset was constructed to benchmark docking performance on peptides of length 3 to 12. Devaurs et al filtered out the small easy peptides from that dataset. The Hou dataset contains ligands from a more recent version of PDBBind dataset [26] that are particulary challenging for traditional docking methods. The PPDBench dataset contains protein-peptide crystal structures composed of 9 to 15 amino acids.

Dataset 1 contains 92 experimentally resolved crystal structures of large ligands with degrees of freedom ranging from 7 to 30. We use this dataset in a redocking setting to test docking performance of large ligands to their native receptor conformations. Note that we do not test the ensemble docking paradigm with this dataset, but just validate our implementation of the parallelized incremental docking paradigm.

We run DINC-Ensemble with the following settings inspired by a previous benchmerk of DINC [20]. We use a single receptor input per ligand (native conformation from the crystal structure). We randomize each ligand prior to docking. We use 24 replicas per ligand and split ligands into 3 fragments adding 3 new active bonds per iteration. We define the center and dimensions of the binding box based on the input ligand with 5Å padding. We use the Vina docking engine with exhaustiveness parameter set to 4, using 1 CPU per Vina run.

Three docking methods DINC, DINC-Ensemble and Vina are compared in this experiment. The major difference between DINC-Ensemble and DINC in this setting is that DINC relies on an older version of Vina in its incremental approach. We run DINC [27] using the same parameters as DINC-Ensemble. The major difference between the incremental approaches (DINC and DINC-Ensemble) and Vina is that Vina docks the ligands with all degrees of freedom active and without incremental steps. To make the comparison fair between Vina and the incremental approaches, we run 24 replicas of Vina and use the exhaustiveness 12 (while in the incremental runs we had 3 iterations with exhaustiveness 4). This way all methods are given equal number of docking cycles to find a good ligand pose.

2.2.2. Ensemble docking case study

To showcase the utility of ensemble docking, we devise a case study of ensemble docking for cyclin dependent kinase CDK2. CDK2 performs a vital signaling function in the cell cycle controlling the G1/S transition. Hyperactivity of CDK2 can lead to disregulation of the cell devision which is one of the hallmarks of cancer. Hence, CDK2 inhibitors are an active target for cancer therapy [28].

CDK2 activity is characterized by a large conformational change induced by cyclin binding [29]. This change is achieved by the motion of the CDK2 activation loop (A-loop) from an A-loop OUT state to an A-loop IN state. Recent work revealed allosteric effects that some inhibitors have on cyclin binding [30] and identified ligands that stabilize the conformational ensemble in one of the two states. Another important structural component of the CDK2 is the DFG motif located in the binding site. The displacement of the DFG domain from the DFG-in to DFG-out state opens up more space within the binding pocket and allows for larger inhibitors to bind (type II inhibitors) [31]. The binding of Type II inhibitors triggers an allosteric effect blocking cyclin binding and leading to less associated toxicity. Recent work

[31] revealed a first available type II inhibitor CDK2 cocrystal structure.

This intricate interplay between the ligand binding and CDK2 flexibility makes it an interesting target for an ensemble docking paradigm. We queried the Protein Databank (PDB) and found crystal structures of ligands mentioned in the above works [31, 30]. We collected 10 crystal structures of the five ligands investigated by Knapp et al and Levinson et al [31, 30]. Supplementary Table 5 lists those crystal structures and the degrees of freedom of the respective ligands. In addition, the conformational features of the crystalized CDK2 are also labeled.

With this data we evaluate the benefits of ensemble docking. We first perform a re-docking experiment (docking each of the ligands to the native receptor). Next, we perform the ensemble docking (cross-docking to multiple receptors), where we dock the ligand to the full ensemble of the 10 crystal structures.

2.3. Evaluation setup

2.3.1. RMSD Evaluation Metric

To quantify the quality of the ligand poses generated by docking methods, we utilize the widely used metric, Root Mean Square Deviation (RMSD). The input ligand conformation serves as the reference or ground truth ligand, against which the RMSD of the docked pose is calculated. In all of our experiments, we utilize the all-atom RMSD, incorporating all ligand atoms in the calculation.

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=0}^{N} d_i^2},$$

where N is the number of atoms and d_i is the distance of the ith atom of the reference ligand and the docked pose.

In ensemble docking scenarios, where the ligand is cross-docked to a non-native receptor, the native and non-native complexes are first aligned, and the RMSD is subsequently calculated between the native pose and the predicted docked pose.

Ligand poses with RMSDs under 2Å are typically regarded as high-quality poses. In this study, we focus on docking particularly challenging and large ligands. Thus, we establish several thresholds to better interpret the results. Poses with RMSDs exceeding 6Å are deemed failed poses. While this threshold is generous, it effectively filters out significantly inaccurate poses and

allows for a thorough analysis of the rest of the results, revealing the power of our approach. We determine the failure rate of the docking tools by identifying the number of docking runs that result in poses with RMSDs greater than 6Å. For large ligands, we use additional thresholds of 2Å, 3Å, and 4Å to distinguish between satisfactory and high-quality poses.

2.3.2. Evaluation Schemes

Docking tools generate multiple ligand poses as output, which are then scored and ranked using a scoring function. Identifying the most representative docking output for RMSD evaluation can be challenging due to the limitations of scoring functions, which may not always rank high-quality poses favorably. To disentangle the effects of the scoring function from the sampling process, we evaluate RMSD using several different schemas."Top X RMSD" refers to the RMSD of the best pose within the top X ranked poses. For instance, "Top 1 RMSD" denotes the RMSD of the single top-scoring pose produced by the docking method, while "Top 5 RMSD" represents the best RMSD among the top 5 scored poses. Lastly, "Best RMSD" is the RMSD of the highest-quality pose across all outputs, irrespective of their scored rank.

3. Results

3.1. Incremental docking benchmark

We compare the performance of DINC-Ensemble to the state-of-the-art methods, Vina and DINC, in docking large ligands. Figure 2 illustrates the performance of all three methods on a large ligand dataset, with degrees of freedom ranging from 7 to 33.

Figure 2A examines the failure rates for all three docking methods as the ligand's degrees of freedom increase. A run is considered failed if the RMSD of the resulting docked pose exceeds 6Å. We evaluate the methods using four different schemes: Top 1, Top 5, Top 10, and the Best pose uncovered by docking (displayed in different facets of the plot). The cumulative failure rate, shown on the y-axis, represents the number of failed poses in a subset of the dataset with less than x degrees of freedom. The x-axis shows the degree of freedom thresholds. Note that the total number of ligands in this dataset is 92, and the y-axis limit is set to 92 to facilitate interpretation within the context of the dataset size.

All docking methods exhibit increasing failure rates for the Top 1 scoring pose as the degrees of freedom increase. However, DINC-Ensemble shows improved performance, reducing the number of failed ligands from approximately 60/92 (for Vina and DINC) to 40/92. When considering the Top 5 and Top 10 scoring poses, the failure rates decrease, with a noticeable performance gap favoring DINC-Ensemble. For the Top 10 evaluation scheme, DINC-Ensemble fails on only 10/92 ligands, whereas Vina and DINC fail on more than 30/92. Furthermore, when examining the Best quality pose generated by the docking methods, DINC-Ensemble and DINC fail on very few ligands (less than 5), while Vina continues to have high failure rates.

Overall, this analysis provides several key insights. First, the incremental approach of both DINC and DINC-Ensemble reduces failure rates, identifying a ligand pose of less than 6Å for the majority of ligands. Second, these poses are not always highly ranked by the scoring function, resulting in high failure rates for the top-scoring poses in DINC. Third, DINC-Ensemble reduces the failure rates among the top-scoring poses, indicating that the scoring function ranks the DINC-Ensemble poses better than those of Vina and DINC. With advancements in scoring functions and their increasing reliability, the performance of DINC-Ensemble is expected to improve.

Figure 2B provides a detailed analysis of the quality of poses successfully docked under 6Å. The results presented are for the best docked pose, with additional evaluation schemes outlined in Figure 1. The x-axis represents the different docking methods, while the y-axis shows the number of ligands docked below specific thresholds. Thresholds of 2Å, 3Å, and 4Å are presented across different facets. While Vina identifies slightly more poses under 2Å, both DINC and DINC-Ensemble achieve better performance for poses under 3Å and 4Å. Notably, DINC-Ensemble successfully docks over 60 out of 92 poses under 3Å.

Figure 2C displays an example of ligand poses docked by all three methods. Docked ligands are represented in red (DINC-Ensemble), yellow (DINC) and green (Vina). In addition, the native crystal structure is overlayed with the docking results and displayed in light blue.

It is important to note that throughout this analysis, DINC-Ensemble was tested in a redocking setting, where each ligand was docked to a single native receptor conformation. These experiments were designed to validate the performance of the methods on large ligands. The subsequent sections further explore the advantages of the ensemble approach.

3.2. Ensemble docking case study

In this analysis, we compare the performance of DINC-Ensemble in a redocking setting (docking ligands to a single native conformation) to its performance in an ensemble setting, where ligands are docked to multiple receptor conformations. The dataset used here is a diverse CDK2 ensemble dataset of 6 ligands and 10 receptor conformations described in Supplementary Table 5. For each of the 10 crystal structures in the dataset, a single redocking experiment and a single ensemble docking experiment were conducted. Note that ensemble docking experiments described here include the native receptor in the ensemble.

Supplementary Figure 1A presents RMSD values achieved in the redocking setting on the x-axis and those achieved in the ensemble docking setting on the y-axis. Each point corresponds to a docking run for one of the 10 data points from the CDK2 dataset (Supplementary Table 5). Values below the diagonal indicate instances where ensemble docking yields a better pose compared to redocking, while values on the diagonal represent cases where the same quality pose is retrieved in both settings. Values above the diagonal show cases where ensemble docking results in a worse pose than redocking. Pink rectangles highlight regions with high-quality poses (RMSDs under 2Å). Performance is evaluated using the Best RMSD pose (regardless of the scoring function rank).

Ensemble docking improves the quality of the docked pose for 5 out of 10 experiments when evaluated. In most of the remaining cases, ensemble docking yields the results of equal quality as redocking, meaning that the native receptor inside the ensemble had the best RMSD pose. We performed additional experiments, excluding the native conformation from the ensemble and those results are presented in the Supplementary Figure 2. Still, ensemble docking is able to uncover lower RMSD poses for 5 out of 10 cases, while for the other 5, the RMSD is higher than that achieved in the native receptor.

Supplementary Figure 1B zooms into a particular case of ensemble docking for the Type II inhibitor K03861 (PDB: 5A14). This plot shows all poses generated by the ensemble docking run for K03861, with the RMSD values on the x-axis and the energy scores on the y-axis. The receptor conformation to which each pose is bound is indicated by the marker's color, while the receptor conformation type is indicated by the marker's shape. An optimal docking run should produce poses in the bottom-left corner of the plot, identifying low scoring, low RMSD poses.

For K03861, the only receptor conformation that achieves both low RMSD and low energy is the native 5A14 receptor. Supplementary Figure 1C provides insight into this result. The 5A14 is the only receptor conformation in the dataset in the DFG-out state (left, yellow pocket in Supplementary Figure 1C). The DFG-out state features a larger binding site volume, accommodating Type II inhibitors like K03861 (yellow ligand on the left of Supplementary Figure 1C). In contrast, the binding pocket of the DFG-in state has a smaller volume and is "blocked" by the DFG loop (right, purple pocket in Supplementary Figure 1C). It is known that the Type II inhibitors do not bind to the receptors in DFG-in state [31]. The K03861 ligand is large and would clash with the protein if placed in the DFG-in binding pocket (yellow ligand on the right of Supplementary Figure 1C). This example illustrates importance of having a representative ensemble. Docking to an ensemble without a DFG-out representative would likely fail to identify a correct pose for K03861 and possibly other type II inhibitor ligands. To highlight this point, we show results for ensemble docking without the native receptor in the ensemble in the Supplementary Figure 3.

Additionally, the energy scores of generated poses within different receptor conformations reveal that the energies are extremely high for the cyclin-bound conformations (triangles). For some cyclin-bound conformations (crosses), negative energy scores are observed. Type II inhibitors are known to stabilize the receptor in a state that prevents cyclin binding [31], and it is interesting to see this trend reflected in the ensemble docking results.

3.3. Web Server and the Python package

DINC-Ensemble web server is freely available at: `https://dinc-ensemble.kavrakilab.rice.edu/`. The input page prompts the user to provide a ligand structure, a reference receptor, a receptor ensemble, and an email (Figure 3). Additional parameters can be selected by the user, but are not required as the defaults are provided. Upon submission, the user receives an email that the job has been submitted along with a pointer to a progress page where users can track the progress of submitted jobs. Once the ensemble docking round has been processed the user receives an email with the link to the results page. There, the user can see the list of ranked docking results as well at visualize the top few predictions in the MolStar 3D viewer (PDBe version). The Python package provides more flexibility for advanced users. It is accessible at `https://github.com/KavrakiLab/dinc-ensemble`.

4. Discussion

We developed the DINC-Ensemble parallel docking program to extend the DINC incremental docking approach to the ensemble docking paradigm, facilitating exploration of both receptor backbone flexibility and large ligand flexibility during docking. We evaluate the DINC-Ensemble program through two case studies: first, assessing its incremental approach for large ligands using a challenging dataset; second, in an ensemble docking scenario of moderate-sized ligands docked to a diverse set of CDK2 receptor conformations.

Docking large ligands with significant flexibility remains challenging for traditional methods due to the expansive search space. The previous incremental approach, DINC [27], demonstrated improved performance for large ligands. In DINC-Ensemble, we update DINC with the newer docking engine Vina1.2.0 [21], leading to improved performance on a large ligand dataset. DINC-Ensemble achieves lower failure rates and better quality ligand poses compared to both DINC and Vina1.2.0, particularly improving the quality of the Top 5 and Top 10 scoring poses. This suggests that combining DINC's incremental approach with Vina1.2.0's advanced search heuristics improves pose quality for top-scoring results. As all traditional docking programs, DINC-Ensemble is limited by the power of scoring functions and its performance is expected to improve as better scoring functions emerge. Large ligand docking remains challenging and to assess pose quality we utilized the RMSD thresholds of 2Å, 3Å, and 4Å, while for docking smaller ligands a stricter thresholds of 1Å and 2Å are used. There is still potential for accessing those lower RMSD pose values by increasing the number of replicas and run time of docking and carefully tuning incremental approach parameters, as showcased by Devaurs et al. [20].

The unique contribution of DINC-Ensemble is in combining the incremental approach of DINC with the ensemble docking paradigm. We showcase the benefits that ensemble docking exhibits in a case study of medium size ligands and the CDK2 ensemble. This case study is interesting because of the scale of the receptor conformational change induced by cyclin binding as well as a diverse set of ligands that induce allosteric effects [30, 31]. Ensemble docking strategy can in some cases provide us with a pose of lower RMSD that that of a redocking experiment (green markers in Figure 2A and Supplementary Figure 2). Another interesting trend is observed for the ligand K03861 labeled as a type II inhibitor. It is clear that docking this ligand

would have failed if the ensemble did not include a unique conformation that facilitates the binding of this inhibitor (5A14). This example hints on how using a single receptor conformation could lead to missing a whole range of ligands of a certain type in a virtual screening setting. It further emphasizes the need for access to a representative ensemble for docking.

There are many challenges still to the ensemble docking paradigm and potential for future advancements. Across our experiments we evaluated the results using different evaluation schemes (i.e., Top 5, Top 10, Best) to address the limitations of current scoring functions, highlighting the need for improved scoring methods to ensure that high-quality poses are accurately ranked. Better parameter tuning of the DINC-Ensemble and longer running times could yield improved results for large ligand docking. Furthermore, the performance of ensemble docking is highly dependent on the quality of the receptor ensemble. Our method, EnGens [32], addresses this by applying unsupervised learning techniques to identify a representative ensemble, underscoring the importance of ensemble diversity. Finally, there is a need for a systematic benchmark of ensemble docking tools across a wider range of targets.

The DINC-Ensemble web server (`https://dinc-ensemble.kavrakilab.rice.edu/`) represents a novel and accessible platform for leveraging the DINC-Ensemble program, catering to users across diverse computational expertise levels. It offers the capability to download and manually analyze detailed results from all docking rounds, complemented by an intuitive overview available directly on the web interface. A notable strength of the DINC-Ensemble implementation is its ability to perform parallelized computations across multiple CPUs, enabling efficient processing of receptor ensembles and replicas. For users requiring tailored workflows or high-throughput docking, the associated Python package is available at `https://github.com/KavrakiLab/dinc-ensemble`, supporting deployment on custom computational resources. We anticipate that this resource will play an important role in advancing efforts to integrate receptor flexibility into molecular docking, addressing a critical need in the field.

5. Funding and Acknowledgments

The authors would like to acknowledge the work of Nicole Tan, Aravind Kuchibhatla and Felix Quintana in supporting the development of the web server. In addition, the authors would like to acknowledge the conversations

and branstorming sessions with other members of KavrakiLab that helped shape up this work.

Work on this project by L.E.K. has been supported in part by the National Institutes of Health NIH [U01CA258512]. Work on this project by A.C. has been supported by in part by the National Institutes of Health NIH [U01CA258512] and the MolSSI Institute through the NSF grant CHE-2136142.

References

[1] B. J. Bender, S. Gahbauer, A. Luttens, J. Lyu, C. M. Webb, R. M. Stein, E. A. Fink, T. E. Balius, J. Carlsson, J. J. Irwin, B. K. Shoichet, A practical guide to large-scale docking, Nature Protocols 2021 16:10 16 (10) (2021) 4799–4832. doi:10.1038/s41596-021-00597-z.
URL https://www.nature.com/articles/s41596-021-00597-z

[2] S. Brenner, R. A. Lerner, Encoded combinatorial chemistry, Proceedings of the National Academy of Sciences of the United States of America 89 (12) (1992) 5381–5383. doi:10.1073/PNAS.89.12.5381.
URL https://pubmed.ncbi.nlm.nih.gov/1608946/

[3] R. S. Bohacek, C. Mcmartin, W. C. Guida, The Art and Practice of Structure-Based Drug Design: A Molecular Modeling Perspectivedoi:10.1002/(SICI)1098-1128(199601)16:1.
URL https://onlinelibrary.wiley.com/terms-and-conditions

[4] G. M. Morris, H. Ruth, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell, A. J. Olson, AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility, Journal of computational chemistry 30 (16) (2009) 2785–2791. doi:10.1002/JCC.21256.
URL https://pubmed.ncbi.nlm.nih.gov/19399780/

[5] O. Trott, A. J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading, Journal of computational chemistry 31 (2) (2010) 455. doi:10.1002/JCC.21334.
URL /pmc/articles/PMC3041641//pmc/articles/PMC3041641/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC3041641/

[6] R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis, P. S. Shenkin, Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy, Journal of Medicinal Chemistry 47 (7) (2004) 1739–1749. doi:10.1021/JM0306430/SUPPL{_}FILE/JM0306430{_}S.PDF.
URL https://pubs.acs.org/doi/full/10.1021/jm0306430

[7] G. Jones, P. Willett, R. C. Glen, A. R. Leach, R. Taylor, Development and validation of a genetic algorithm for flexible docking, Journal of Molecular Biology 267 (3) (1997) 727–748. doi:10.1006/JMBI.1996.0897.

[8] A. Dhanik, J. S. McMurray, L. E. Kavraki, DINC: A new AutoDock-based protocol for docking large ligands, BMC Structural Biology 13 (SUPPL.1) (2013) 1–14. doi:10.1186/1472-6807-13-S1-S11/FIGURES/7.
URL https://bmcstructbiol.biomedcentral.com/articles/10.1186/1472-6807-13-S1-S11

[9] D. A. Antunes, D. Devaurs, L. E. Kavraki, Understanding the challenges of protein flexibility in drug design, Expert Opinion on Drug Discovery 10 (12) (2015) 1301–1313. doi:10.1517/17460441.2015.1094458.
URL https://www.tandfonline.com/doi/abs/10.1517/17460441.2015.1094458

[10] E. Fischer, Einfluss der Configuration auf die Wirkung der Enzyme, Berichte der deutschen chemischen Gesellschaft 27 (3) (1894) 2985–2993. doi:10.1002/CBER.18940270364.
URL https://onlinelibrary.wiley.com/doi/full/10.1002/cber.18940270364https://onlinelibrary.wiley.com/doi/abs/10.1002/cber.18940270364https://chemistry-europe.onlinelibrary.wiley.com/doi/abs/10.1002/cber.18940270364

[11] D. E. Koshland, Application of a Theory of Enzyme Specificity to Protein Synthesis, Proceedings of the National Academy of Sciences 44 (2) (1958) 98–104. doi:10.1073/PNAS.44.2.98/ASSET/33D43A5D-1879-4B8E-8263-5FE6FB86AEE1/ASSETS/PNAS.44.2.98.FP.PNG.
URL https://www.pnas.org/doi/abs/10.1073/pnas.44.2.98

[12] B. Ma, S. Kumar, C.-J. Tsai, R. Nussinov, Folding funnels and binding mechanisms, Protein Engineering vol 12 (9) (1999) 713–720.

[13] R. M. Knegtel, I. D. Kuntz, C. M. Oshiro, Molecular docking to ensembles of protein structures, Journal of Molecular Biology 266 (2) (1997) 424–440. doi:10.1006/JMBI.1996.0776.

[14] S. Vilar, S. Costanzi, Application of Monte Carlo-Based Receptor Ensemble Docking to Virtual Screening for GPCR Ligands, Methods in Enzymology 522 (2013) 263–278. doi:10.1016/B978-0-12-407865-9.00014-5.

[15] Y. Li, D. M. Wu, L. M. Kong, S. Zhang, H. Du, W. Sun, L. Zhang, Y. Li, Z. Zuo, Ensemble Docking-Based Virtual Screening Toward Identifying Inhibitors against Wee1 Kinase, Future Medicinal Chemistry 11 (16) (2019) 1889–1906. doi:10.4155/FMC-2019-0022.
URL https://www.tandfonline.com/doi/abs/10.4155/fmc-2019-0022

[16] S. Hall-Swan, D. Devaurs, M. M. Rigo, D. A. Antunes, L. E. Kavraki, G. Zanatta, DINC-COVID: A webserver for ensemble docking with flexible SARS-CoV-2 proteins, Computers in Biology and Medicine 139 (2021) 104943. doi:10.1016/J.COMPBIOMED.2021.104943.

[17] A. Acharya, R. Agarwal, M. B. Baker, J. Baudry, D. Bhowmik, S. Boehm, K. G. Byler, S. Y. Chen, L. Coates, C. J. Cooper, O. Demerdash, I. Daidone, J. D. Eblen, S. Ellingson, S. Forli, J. Glaser, J. C. Gumbart, J. Gunnels, O. Hernandez, S. Irle, D. W. Kneller, A. Kovalevsky, J. Larkin, T. J. Lawrence, S. Legrand, S. H. Liu, J. C. Mitchell, G. Park, J. M. Parks, A. Pavlova, L. Petridis, D. Poole, L. Pouchard, A. Ramanathan, D. M. Rogers, D. Santos-Martins, A. Scheinberg, A. Sedova, Y. Shen, J. C. Smith, M. D. Smith, C. Soto, A. Tsaris, M. Thavappiragasam, A. F. Tillack, J. V. Vermaas, V. Q. Vuong, J. Yin, S. Yoo, M. Zahran, L. Zanetti-Polzi, Supercomputer-Based Ensemble Docking Drug Discovery Pipeline with Application to Covid-19, Journal of Chemical Information and Modeling 60 (12) (2020) 5832–5852. doi:10.1021/ACS.JCIM.0C01010/ASSET/IMAGES/LARGE/CI0C01010{_}0009.JPEG.
URL https://pubs.acs.org/doi/full/10.1021/acs.jcim.0c01010

[18] T. Chandak, C. F. Wong, EDock-ML: A web server for using ensemble docking with machine learning to aid drug discovery, Protein Science 30 (5) (2021) 1087–1097. doi:10.1002/PRO.4065.
URL https://onlinelibrary.wiley.com/doi/full/10.1002/pro.4065https://onlinelibrary.wiley.com/doi/abs/10.1002/pro.4065https://onlinelibrary.wiley.com/doi/10.1002/pro.4065

[19] I. A. Guedes, M. M. Pereira da Silva, M. Galheigo, E. Krempser, C. S. de Magalhães, H. J. Correa Barbosa, L. E. Dardenne, Dockthor-vs: A free platform for receptor-ligand virtual screening, Journal of Molecular Biology 436 (17) (2024) 168548. doi:10.1016/j.jmb.2024.168548.
URL http://dx.doi.org/10.1016/j.jmb.2024.168548

[20] D. Devaurs, D. A. Antunes, S. Hall-Swan, N. Mitchell, M. Moll, G. Lizée, L. E. Kavraki, Using parallelized incremental meta-docking can solve the conformational sampling issue when docking large ligands to proteins, BMC Molecular and Cell Biology 20 (1) (Sep. 2019). doi:10.1186/s12860-019-0218-z.
URL https://doi.org/10.1186/s12860-019-0218-z

[21] J. Eberhardt, D. Santos-Martins, A. F. Tillack, S. Forli, AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings, Cite This: J. Chem. Inf. Model 61 (2021) 3891–3898. doi:10.1021/acs.jcim.1c00203.
URL https://doi.org/10.1021/acs.jcim.1c00203

[22] R. Rentzsch, B. Y. Renard, Docking small peptides remains a great challenge: an assessment using autodock vina, Briefings in Bioinformatics 16 (6) (2015) 1045–1056. doi:10.1093/bib/bbv008.
URL http://dx.doi.org/10.1093/bib/bbv008

[23] A. S. Hauser, B. Windshügel, Leads-pep: A benchmark data set for assessment of peptide docking performance, Journal of Chemical Information and Modeling 56 (1) (2016) 188–200. doi:10.1021/acs.jcim.5b00234.
URL http://dx.doi.org/10.1021/acs.jcim.5b00234

[24] Z. Wang, H. Sun, X. Yao, D. Li, L. Xu, Y. Li, S. Tian, T. Hou, Comprehensive evaluation of ten docking programs on a diverse set of protein–ligand complexes: the prediction accuracy of sampling power

and scoring power, Physical Chemistry Chemical Physics 18 (18) (2016) 12964–12975. doi:10.1039/c6cp01555g.
URL http://dx.doi.org/10.1039/C6CP01555G

[25] P. Agrawal, H. Singh, H. K. Srivastava, S. Singh, G. Kishore, G. P. S. Raghava, Benchmarking of different molecular docking methods for protein-peptide docking, BMC Bioinformatics 19 (S13) (Feb. 2019). doi:10.1186/s12859-018-2449-y.
URL http://dx.doi.org/10.1186/s12859-018-2449-y

[26] R. Wang, X. Fang, Y. Lu, C.-Y. Yang, S. Wang, The pdbbind database: methodologies and updates, Journal of Medicinal Chemistry 48 (12) (2005) 4111–4119. doi:10.1021/jm048957q.
URL http://dx.doi.org/10.1021/jm048957q

[27] D. A. Antunes, M. Moll, D. Devaurs, K. R. Jackson, G. Lizée, L. E. Kavraki, DINC 2.0: A new protein-peptide docking webserver using an incremental approach, Cancer Research 77 (21) (2017) e55–e57. doi:10.1158/0008-5472.CAN-17-0511/SUPPLEMENTARY-VIDEO-S2.
URL /cancerres/article/77/21/e55/662392/DINC-2-0-A-New-Protein-Peptide-Docking-Webserverhttps://dx.doi.org/10.1158/0008-5472.CAN-17-0511

[28] S. Tadesse, A. T. Anshabo, N. Portman, E. Lim, W. Tilley, C. E. Caldon, S. Wang, Targeting cdk2 in cancer: challenges and opportunities for therapy, Drug Discovery Today 25 (2) (2020) 406–413. doi:10.1016/j.drudis.2019.12.001.
URL http://dx.doi.org/10.1016/j.drudis.2019.12.001

[29] M. Pellerano, S. Tcherniuk, C. Perals, T. N. Ngoc Van, E. Garcin, F. Mahuteau-Betzer, M. Teulade-Fichou, M. C. Morris, Targeting conformational activation of cdk2 kinase, Biotechnology Journal 12 (8) (May 2017). doi:10.1002/biot.201600531.
URL http://dx.doi.org/10.1002/biot.201600531

[30] A. Majumdar, D. J. Burban, J. M. Muretta, A. R. Thompson, T. A. Engel, D. M. Rasmussen, M. V. Subrahmanian, G. Veglia, D. D. Thomas, N. M. Levinson, Allostery governs cdk2 activation and differential recognition of cdk inhibitors, Nature Chemical Biology 17 (4) (2021) 456–464.

doi:10.1038/s41589-020-00725-y.
URL http://dx.doi.org/10.1038/s41589-020-00725-y

[31] L. T. Alexander, H. Möbitz, P. Drueckes, P. Savitsky, O. Fedorov, J. M. Elkins, C. M. Deane, S. W. Cowan-Jacob, S. Knapp, Type ii inhibitors targeting cdk2, ACS Chemical Biology 10 (9) (2015) 2116–2125. doi:10.1021/acschembio.5b00398.
URL http://dx.doi.org/10.1021/acschembio.5b00398

[32] A. Conev, M. M. Rigo, D. Devaurs, A. F. Fonseca, H. Kalavadwala, M. V. de Freitas, C. Clementi, G. Zanatta, D. A. Antunes, L. E. Kavraki, EnGens: a computational framework for generation and analysis of representative protein conformational ensembles, Briefings in Bioinformatics 24 (4) (2023) 1–11. doi:10.1093/BIB/BBAD242.
URL https://dx.doi.org/10.1093/bib/bbad242

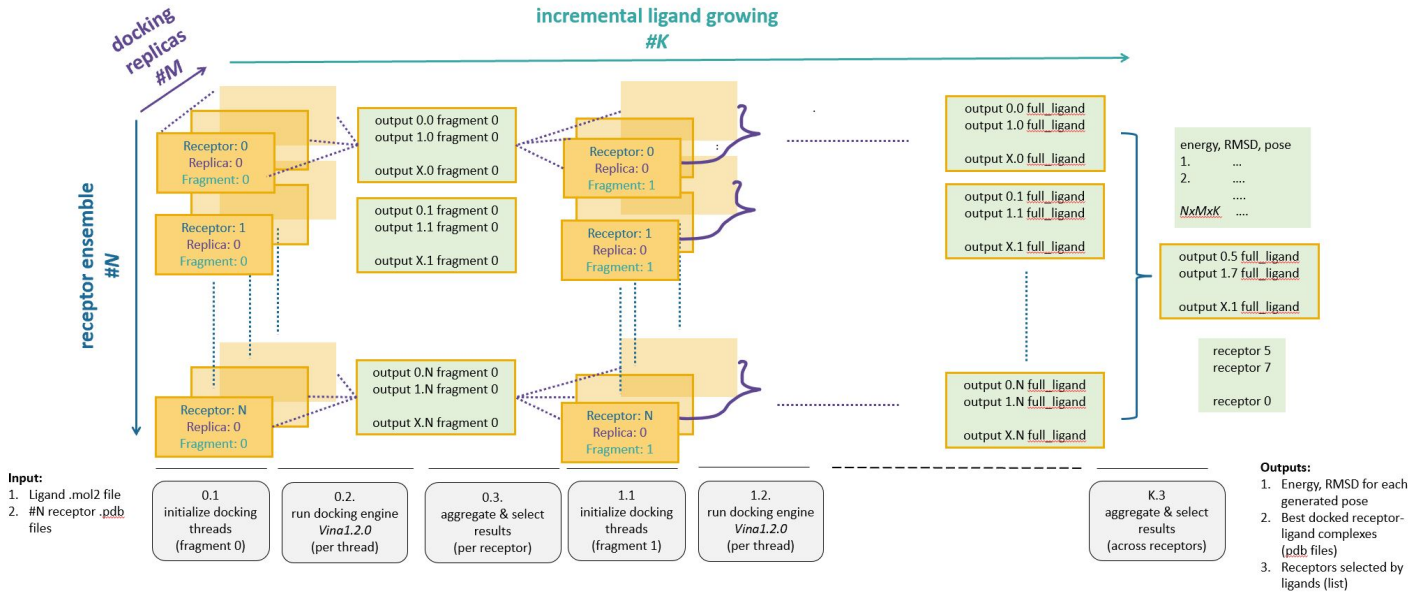

Figure 1: DINC-Ensemble Incremental Meta-Docking Program: The x-axis follows the incremental ligand growing pipeline. The y-axis outlines multiple receptor conformations and follows the ensemble docking stream. The z-axis shows the multiple replica docking in a parallelized approach. Yellow rectangles represent the docking engine that performs the search for single fragments. Green rectangles denote outputs of docking for each iteration as well as the final DINC-Ensemble output. The gray rectangles in the bottom of the plot outline the consecutive steps of DINC-Ensemble's incremental ensemble docking program.

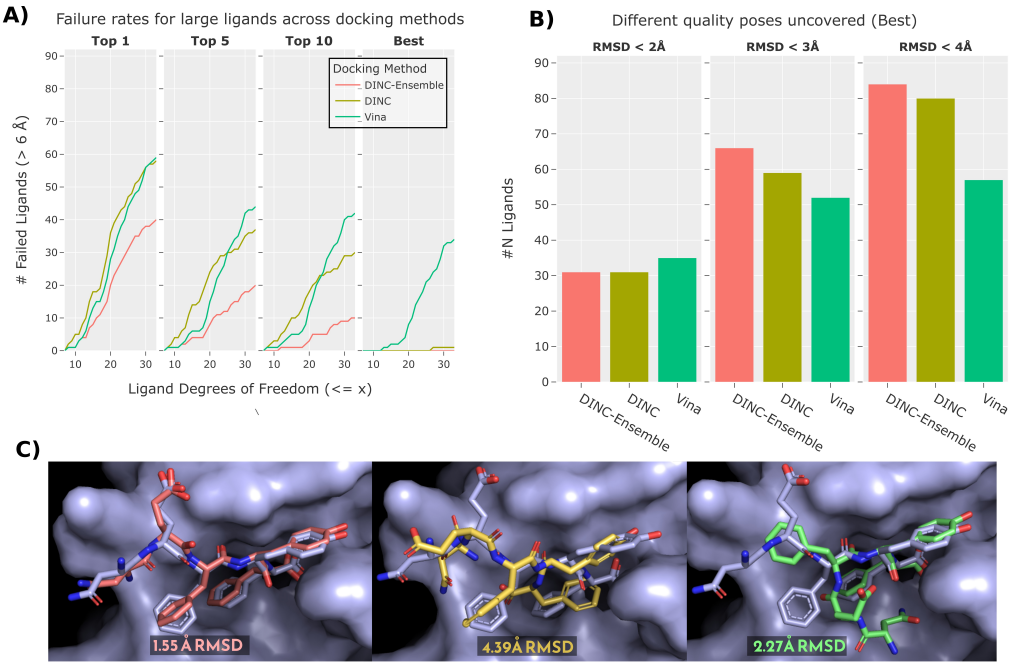

Figure 2: Docking performance on the large ligand case study. A) Cumulative failure rates (the lower the better) for docking methods with increased ligand size. X-axis represents the ligand size threshold. Y-axis shows cummulative failure rate. Facets correspond to distinct RMSD evaluation schemas. Docking methods are color coded. B) Quality of the generated poses (RMSD below 6Å) for the docking methods using the Best evaluation schema. Y-axis shows the number of ligands that achieved RMSD below the given threshold. Facets represent different RMSD thresholds. C) Docking result poses for a large ligand with 20 degrees of freedom (PDB: 1W92) are visualized across different models. The crystal structure complex, displayed in light blue, is overlayed with each of the docking results. The optimal pose from DINC-Ensemble is shown in red on the left (1.55Å). The pose from DINC is shown in yellow in the center(4.39Å). The Vina result is shown in green on the right (2.27Å).

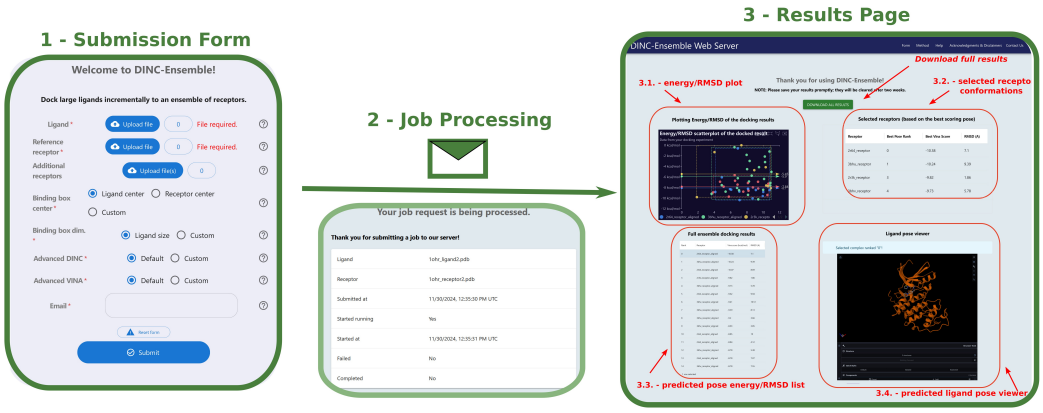

Figure 3: DINC-Ensemble Webserver: Submission Form (1), Job Processing (2) and Results Page (3). Different views are provided on the results page.