

UCSF

UC San Francisco Previously Published Works

Title

Docking for Molecules That Bind in a Symmetric Stack with SymDOCK.

Permalink

<https://escholarship.org/uc/item/6298549t>

Journal

Journal of Chemical Information and Modeling, 64(2)

Authors

Smith, Matthew

Knight, Ian

Kormos, Rian

et al.

Publication Date

2024-01-22

DOI

10.1021/acs.jcim.3c01749

Peer reviewed

Docking for Molecules That Bind in a Symmetric Stack with SymDOCK

Matthew S. Smith, Ian S. Knight, Rian C. Kormos, Joseph G. Pepe, Peter Kunach, Marc I. Diamond, Sarah H. Shahmoradian, John J. Irwin, William F. DeGrado, and Brian K. Shoichet*



Cite This: *J. Chem. Inf. Model.* 2024, 64, 425–434



Read Online

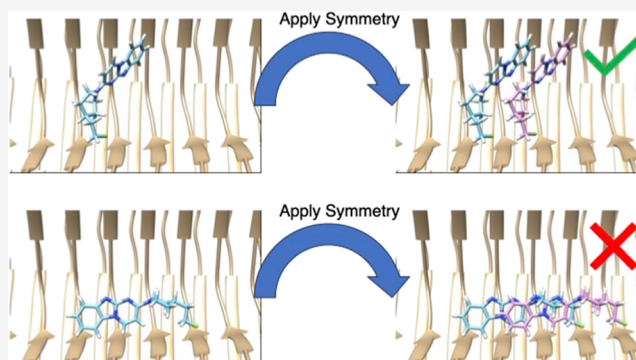
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: Discovering ligands for amyloid fibrils, such as those formed by the tau protein, is an area of great current interest. In recent structures, ligands bind in stacks in the tau fibrils to reflect the rotational and translational symmetry of the fibril itself; in these structures, the ligands make few interactions with the protein but interact extensively with each other. To exploit this symmetry and stacking, we developed SymDOCK, a method to dock molecules that follow the protein's symmetry. For each prospective ligand pose, we apply the symmetry operation of the fibril to generate a self-interacting and fibril-interacting stack, checking that doing so will not cause a clash between the original molecule and its image. Absent a clash, we retain that pose and add the ligand–ligand van der Waals energy to the ligand's docking score (here using DOCK3.8). We can check these geometries and energies using an implementation of ANI, a neural-network-based quantum-mechanical evaluation of the ligand stacking energies. In retrospective calculations, symmetry docking can reproduce the poses of three tau PET tracers whose structures have been determined. More convincingly, in a *prospective* study, SymDOCK predicted the structure of the PET tracer MK-6240 bound in a symmetrical stack to AD PHF tau before that structure was determined; the docked pose was used to determine how MK-6240 fit the cryo-EM density. In proof-of-concept studies, SymDOCK enriched known ligands over property-matched decoys in retrospective screens without sacrificing docking speed and can address large library screens that seek new symmetrical stackers. Future applications of this approach will be considered.



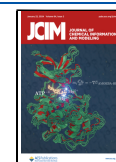
INTRODUCTION

An open problem in ligand discovery is understanding and exploiting the ability of high-affinity ligands to bind to protein amyloids. Tau (tubulin associated unit), an intrinsically disordered protein that can wrap around microtubules, is one such target for diagnostic tools.^{1–3} The accumulation of tau fibrils into toxic neurofibrillary tangles (NFTs) is characteristic of Alzheimer's disease (AD), chronic traumatic encephalopathy (CTE), and other neurodegenerative tauopathies.^{4–10} In cryogenic electron microscopy (cryo-EM) studies, tau adopts different polymorphs characteristic of different diseases: the paired helical filament (PHF) and the straight filament (SF) for AD and type I or type II for CTE.^{11–17} Recently, cryo-EM structures of AD tau and CTE tau with bound ligands have been determined. A striking feature of these structures is that the ligands bind in stacks to the tau filaments to reflect the rotational and translational symmetry of the fibril. Epigallocatechin gallate (EGCG), a tau disaggregator found in green tea, binds at the inter-protofilament cleft of AD PHF tau with a stoichiometry of one ligand molecule per protein monomer.¹⁸ The positron emission tomography (PET) radiotracers GTP-1 and MK-6240 bind within the

“C” shape of each protofilament with the same stoichiometry, as does flortaucipir (Tauvid) to CTE type I tau.^{19–22}

In each of these examples, the ligands within the stacks make substantial van der Waals and π – π interactions. Intriguingly, for the PET tracers, these ligand–ligand interactions seem more dominant than their direct contacts with the fibrils. For example, GTP-1 forms only a single hydrogen bond with lysine 353 on AD PHF tau.¹⁹ Flortaucipir's pyridine nitrogen in its benzo-pyrrolo-pyridine moiety makes a hydrogen bond to aspartate 358 in CTE type I tau; however, apart from this, the structure primarily extends into the solvent.²¹ EGCG establishes more hydrophilic contacts with the inter-protofilament cleft of AD PHF tau, including hydrogen bonds to asparagine 327, glutamate 338, and lysine 340, and polar contact with histidine 329. Yet, its interactions mainly involve

Received: October 29, 2023
Revised: December 21, 2023
Accepted: December 22, 2023
Published: January 8, 2024



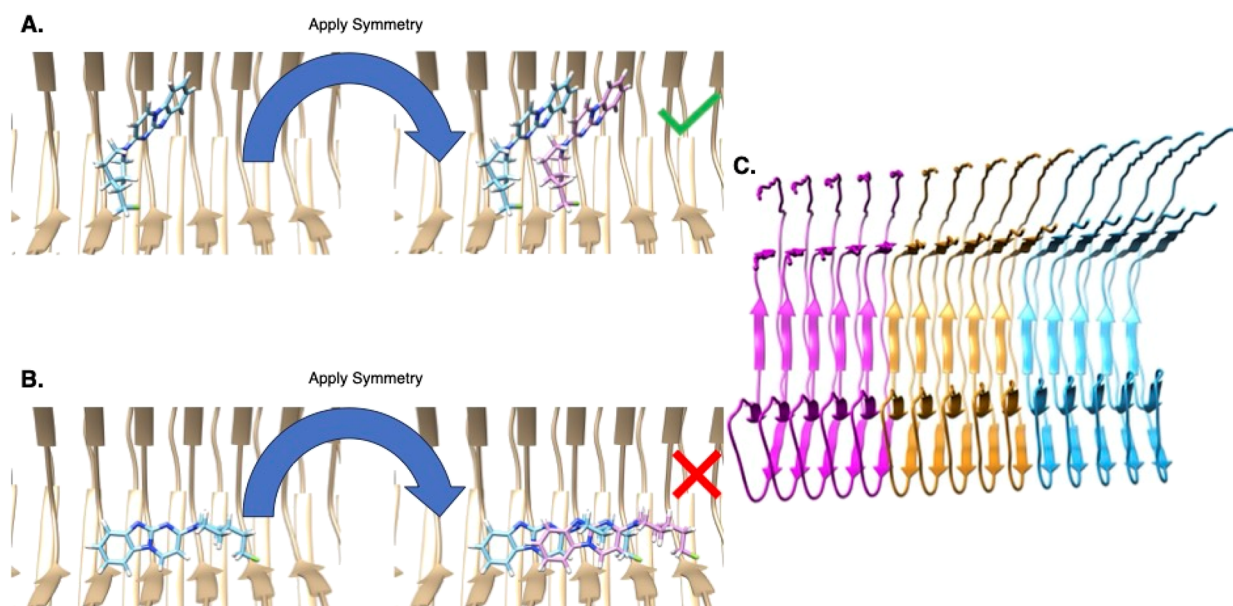


Figure 1. A demonstration of SymDOCK by docking GTP-1 into the structure of it bound to AD PHF tau (PDB ID: 8FUG). (a, b) SymDOCK generates poses for molecules as DOCK 3.8 normally does and then applies a symmetry check: if the molecule (blue) and its translated/rotated copy (purple) have no atoms within 2 Å of each other, then the pose passes the filter (a: with the dummy example of the experimental pose). If the molecule and its copy have atoms within this distance, then the pose fails the filter (b: showing the top pose of docking without the symmetry requirement). SymDOCK only evaluates ligand–ligand van der Waals energies of poses that pass the filter, and only does so for the first pair's energies because ligand–protein interactions will be the same at every site by symmetry (ligand–protein interaction terms are evaluated as in DOCK3.8 for the first ligand pose in the stack). (c) To avoid edge effects from modeling a small fraction of a fibril that should be nanometers in length, we artificially extend the input protein structure using the symmetry operation. This cryo-EM structure only contains five monomers of tau in each protofibril, but by applying the affine transformation and its inverse on the monomer atom positions, we can generate new atom positions for a fibril of 15 monomers (bronze for the original structure, blue for applying symmetry operation, and magenta for applying the inverse symmetry operation).

other ligands in the stack and the surrounding solvent.¹⁸ Despite these minimal fibril contacts, GTP-1 exhibits an 11 nM affinity to tau, whereas flortaucipir binds to brain-derived tau in the low nM to high pM concentration range.^{23–30}

To use this new mode of binding to find other ligands that might bind to different tau polymorphs and to other protein fibrils, we developed the symmetric extension to DOCK3.8,^{31–33} “SymDOCK”. This method uses the same pose sampling and scoring as in DOCK3.8 but retains only those poses that can form a symmetrical stack in the protein. We then evaluate the ligand–ligand van der Waals energy as a crude measure of favorable self-interaction and add that to the DOCK score. Given the imposed symmetry, we assume that the interactions with the protein are copied from the ligand monomer to monomer. We found that in both retrospective and, more convincingly, prospective predictions, SymDOCK succeeded in recapitulating and in predicting the poses of different tau ligands. In a test-of-concept screen of 22 million molecules, SymDOCK was not much slower than base DOCK3.8. This suggests that the extra cost of the new symmetry operations and of calculating inter-ligand van der Waals energies is compensated by the lower sampling imposed by the symmetry constraints. To further test SymDOCK’s generated geometries, we computed their energies using ANI, a neural network force field trained on density functional theory (DFT) energy data.^{34,35} We found that their ANI energies and exact geometries differed only slightly from those of ANI-optimized stacks. We consider further generalizations of this approach.

METHODS

Symmetric Pose Sampling. The SymDOCK extension uses the same ligand building and initial pose generation as base DOCK3.8, allowing a user to dock the built section of the ZINC22 database without any modifications.^{31,36} To keep SymDOCK from decreasing the speed of docking each molecule, we do not generate or attempt to dock a stack of molecules. Rather, we generate poses (conformations and orientations) for each ligand as always in DOCK but only keep those that can self-stack.^{31,37,38} Because of the symmetry of the system, we need to evaluate only the ligand–protein interaction (checks for clashes and energetic favorability) for a single site; the ligand–protein interactions at all other sites will be identical.

Once we generate a potential pose, we apply the symmetry operation of the fibril and check the ligand–ligand interactions (Figure 1a,b). The symmetry operation of the fibril is the four-dimensional affine transformation matrix that acts on a vector in three dimensions, applying a rotation followed by a translation. For the tau fibril, this transformation would be a 4.7 Å translation down the long axis of the fibril along with a $\sim -1^\circ$ rotation in the plane perpendicular to that axis.^{11,18} We describe a procedure for generating this matrix in the discussion of Figure S1. Once we have a rotated/translated pose of the molecule, we check for a van der Waals clash with the initial pose, defined as any atom from the initial pose being within 2 Å of any atom in the transformed pose. We choose this conservative definition of what constitutes a clash based on the experimental structure of GTP-1 bound to AD PHF tau (Figure S2).¹⁹ For each pose that passes the symmetry check,

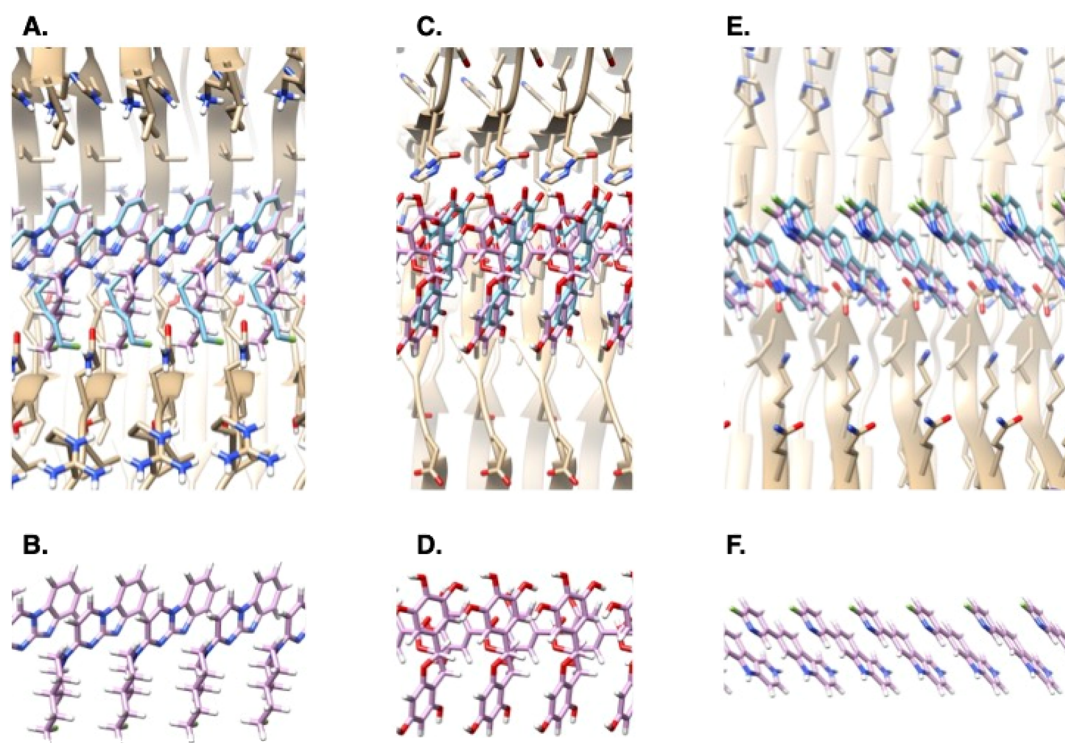


Figure 2. Retrospective pose reproduction, with the protein in beige, the cryo-EM pose in blue, and the SymDOCK-predicted pose in purple, with (a, c, e) and without (b, d, f; emphasizing the docked ligand stacking) the experimental structure shown. (a, b) GTP-1 docked into in the PET-tracer-site of AD PHF tau; RMSD to the experimental structure 1.2 Å. (c, d) EGCG docked into the inter-protofilament cleft of AD PHF tau; RMSD to the experimental structure 2.4 Å. (e, f) Flortaucipir docked into CTE type I tau; RMSD to the experimental structure 0.48 Å.

we add the van der Waals interaction of the original molecule and the transformed version to the total DOCK score function with the same AMBER 4.0 united-atom force field parameters as the ligand–protein van der Waals energy.^{31,39–41}

Changes to the DOCK Score Grids. In addition to the explicit ligand–ligand interactions, we slightly changed how we calculate the grids for the ligand–protein interactions. Because the protein fibril is hundreds of nanometers long, its interactions with any small molecule ligand should be periodic.¹¹ We simulate this by artificially extending the fibril structure to more monomers than what are in the deposited structure (Figure 1c). For example, the cryo-EM structure of GTP-1 bound to AD PHF tau (PDB ID: 8FUG; we used an earlier version of the structure) only contains five monomers of tau in each protofilament. By applying the affine transformation and its inverse on the monomer atom positions, we can generate new atom positions for a fibril of 15 monomers. We then feed the longer fibril into the Pydock3 procedure⁴² for generating the van der Waals, electrostatic, and desolvation grids and see that these grids do not have edge effects in the region of interest to docking (Figure S3).

We also change the electrostatic and desolvation grids to account for the changed dielectric environment of a stack of organic ligand molecules (Figure S4). Part of the normal DOCK3.8 optimization procedure includes adding a low dielectric layer that extends the boundary of protein atoms into the bulk solvent.^{31,39,42} This increases the magnitude of electrostatic interactions, attempting to balance the nonpolar terms in docking, which often dominate. In SymDOCK, we have two sources of large-magnitude van der Waals energies: ligand–protein interactions and ligand–ligand interactions. Extending the low dielectric boundary for both electrostatic

and desolvation grids to the positions of the heavy atoms in a stack of ligands in the experimental structure helps to overcome this bias toward nonpolar terms and mimics the ligand binding's effect on the dielectric boundary (Figure S5, made using the R Programming Language and the ggplot2 package).^{43,44} Such a ligand perturbation of the dielectric boundary becomes even more important with a stack of ligands bound to the fibril.

ANI-Based Stack Filter. The energetic favorability of the molecular poses returned by SymDOCK may be screened using the ANI-2x molecular force field, as implemented in TorchANI.^{34,35} ANI-2x is a neural network trained to predict the ground state energies of small organic molecules as calculated by density functional theory at the ω B97X/6-31G* level. We reasoned that the ANI-2x force field provided the best compromise between throughput and accuracy for the goal of determining the per-monomer ground state energies of the stacks of small molecules generated by SymDOCK.

We first estimate the SymDOCK-predicted geometry's per-monomer energy, and then use that as a starting geometry for a Monte Carlo (MC) optimization via the Metropolis–Hastings algorithm.⁴⁵ We do not perform any optimization prior to the ANI calculations: the MC procedure works only using the ANI energies. We sample each candidate move by taking translations from a multivariate normal distribution and rotations from an isotropic normal distribution on SO(3) (the group or rotations in 3D).⁴⁶ For each candidate move, we generate the two nearest symmetry mates of the molecule, evaluate the ANI energy, and determine the per-monomer energy (Supplemental Methods). After 1000 Monte Carlo steps, we compare the minimum energy configuration achieved to SymDOCK's output in terms of both ANI-generated energy difference and

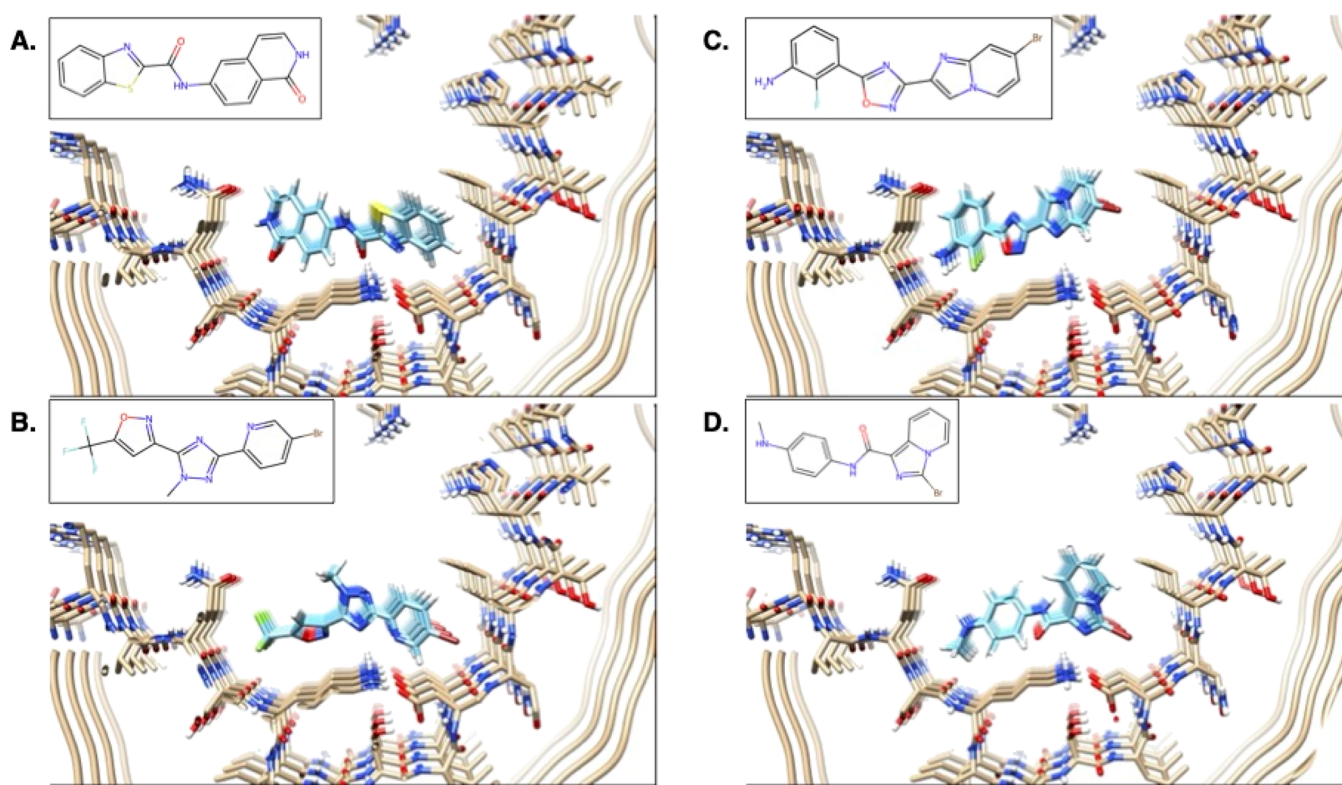


Figure 3. Stacking and polar interactions to the AD PHF tau (beige) among docked poses of four representative high-ranking molecules from a 22-million-molecule library screen (blue). (a) The benzothiophene-quinoline ZINCnD000001Ahba. (b and c) The triaryls ZINCmD000004I6kF and ZINCnF000004Bxys. (d) The amide-linked phenyl-benzimidazole ZINCID000002Tbny.

RMSD to the original docking pose. At this point in the development of the method, this is mostly done to test whether the docking-predicted ligand stacking is sensible; we do not alter the conformation of the molecule from what SymDOCK produces in the docking predictions and do not currently use the MC protocol to refine best-scoring poses, although doing so may be explored in future use of this approach.

RESULTS

Retrospective Pose Reproduction. Reproducing experimental poses of known binders was a first metric we used to evaluate SymDOCK. Typically, reproducing the ligand pose by docking it back into its native complex is considered necessary in the field, but here, there is the complication of regenerating the full ligand stack. We began with the PET tracer GTP-1, docking it back into its complex with the AD PHF tau fibril. As is typical, we used pseudo atoms (“spheres”) nucleated around the known ligand coordinates to define the sampling and calculated van der Waals, Poisson–Boltzmann electrostatic, and desolvation energy potential grids to evaluate ligand protein complementarity.^{31,39,47} SymDOCK sampled GTP-1 in 1507 orientations in the fibril site and multiple conformations within each orientation.⁴⁸ Each pose was checked for the ability to generate a symmetry partner that does not clash with itself using the fibril symmetry operator (Figure 1a). If a pose passed this check, then SymDOCK calculated the van der Waals, electrostatic, desolvation and, new to SymDOCK, ligand–ligand van der Waals energies. We then ranked allowed poses by the total DOCK score. When we compared the highest-scoring docked pose of GTP-1 to its experimental geometry, the root-mean-squared difference (RMSD) was

1.165 Å, with the three-ring system of the ligand almost exactly superposed and the differences coming from the flexible tail (Figure 2a,b, Figure S6). Similarly, docking EGCG back into its complex in the inter-protofilament cleft of AD PHF tau led to an RMSD of 2.37 Å to the experimental structure for the top-scoring pose (Figure 2c,d, Figure S7). Even with this RMSD, SymDOCK recapitulated most of the key polar interactions observed in the experimental structure. Just as importantly, all of the predicted poses captured the ligand–ligand aromatic stacking. Finally, docking flortaucipir back into its complex with CTE type I tau led to an RMSD of 0.477 Å for the best docking pose, 7.78 Å for the top scoring one, and RMSD values between of 7.49 and 7.74 Å for the poses scoring between those two (Figure 2e,f, Figure S8). When we compared GTP-1’s, EGCG’s, and flortaucipir’s SymDOCK poses to their ANI-optimized geometries, the ligand–ligand energies improved, but the structures changed only between 0.6 and 1.1 Å in RMSD (Table S1). Crucially, in all three of the docked complexes, the ligand–ligand packing and quadrupole stacking closely resembled those in the experimental structures.

Retrospective Enrichment of Ligands over Decoys.

We were curious if symmetry docking could identify plausible fibril ligands among a large molecular library. We first asked whether SymDOCK could prioritize known ligands versus a much larger set of property-matched decoys. Docking against AD PHF tau, we docked the PET ligands GTP-1 and MK-6240 and a set of 14 MK-6240 analogs drawn from a larger group of 94 to maximize diversity (Table S2).^{20,49} We generated 50 property matched decoys for each ligand, for 800 total.^{50,51} Docking the 16 ligands and 800 decoys against the GTP-1-bound structure of AD PHF tau resulted in an

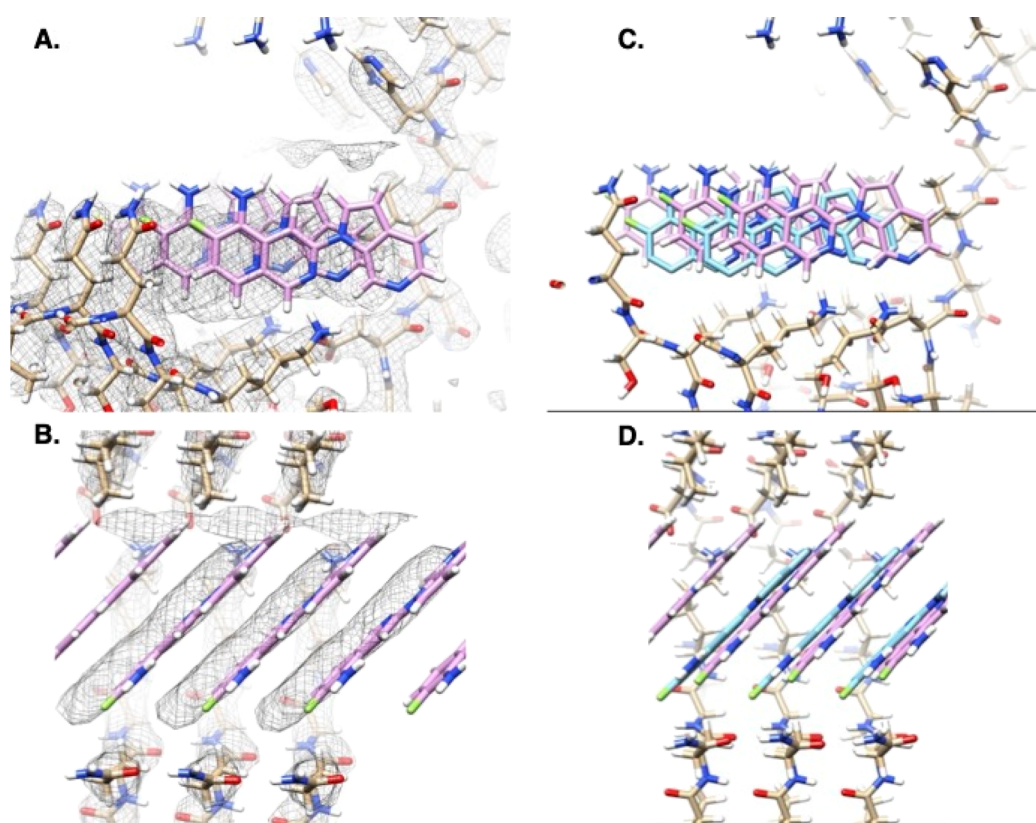


Figure 4. Prospective pose prediction for MK-6240 in AD PHF tau. (a, b) The electron density of the MK-6240-tau structure (black mesh showing the 0.0117 level) with the modeled protein in beige and SymDOCK-predicted pose in purple. (a) Looking down the axis of the fibril. (b) Looking into the open trough of the fibril. (c, d) The final refinement of the MK-6240 pose into the electron density (blue) using the SymDOCK-predicted pose as input.

adjusted logAUC enrichment of 32.7 and an enrichment factor at 1% ($EF_{1\%}$) of 6.4 (Figure S9).^{39,52} The better ranking of the known ligands versus the decoys reflects their better ability to form stacks that recapitulate fibril symmetry and interact with it. For instance, ZINC000000035542 is a 1.2.2-bicyclo that can dock to the receptor grids using the nonsymmetric default DOCK3.8 procedure, but its three-dimensional ring structure prevents it from docking symmetrically (structure in Figure S10). Even though this decoy has properties matched with the known ligands, the symmetry check prevents it from scoring and improves enrichment.

Docking 22 Million Molecules from ZINC22. Some of the decoys from the enrichment calculation scored well and looked plausible, stacking well with themselves and packing in the fibril to mimic its symmetry. We wondered whether there might be many molecules in a general library that might be suitable to form symmetrical stacks in a fibril. To test how symmetry docking would work in a large library screen, we docked 22 million molecules from the ZINC22 database³⁶ against the GTP-1-bound form of AD PHF tau. We selected molecules that were similar in gross properties to GTP-1: neutral, with heavy atom count between 21 and 23 and clogP values between 3.30 and 3.50.^{53,54} The 22 million molecules docked in 17,800 core hours (less than a day on a typical 1000-core cluster), 2.90 s per molecule per core, two to three times slower than typical for standard docking with DOCK3.8.³² When passed through the ANI-based filter, the configurations of the stacked ligands typically changed only modestly (Figure S11, made using the R Programming Language): 74% of the 5000 molecules changed by less than 1.5 Å RMSD, and 82% of

them changed by less than 2.0 Å RMSD.⁴³ Practically, these observations suggest that most of the stacking interactions in the docking poses are energetically sensible when they are evaluated at a higher level of theory.

Visually, many of the top 5000 molecules from the large library docking stacked in plausible ways, typically making one or two polar interactions with the fibril, akin to what is observed in the experimental ligand–fibril structures. For instance, ZINCnD000001AHba appears to make a single hydrogen bond to Lys353 from its amide carbonyl (Figure 3a). The molecule stacks with symmetry reflecting that of the overall fibril, including 4.7 Å translation. The orientation of the benzothiophenes and quinoline rings, however, creates interplanar distances of 3.4 and 3.5 Å, respectively. These values agree with the 3.3 Å planar separation between the ring systems in the bound structure of GTP-1 to AD PHF tau and ensure that putative quadrupole–quadrupole interactions are strong.^{19,55–57} Meanwhile, ZINCmD000004I6kF and ZINCnF000004BXys are not posed to form explicit hydrogen bonds with the fibril, instead making van der Waals contacts with it and perhaps more importantly making extensive ligand–ligand stacking interactions (Figure 3b,c). Self-stacking also seems to dominate in the pose of the amide-linked phenyl-benzimidazole ZINCID000002Tbny, which again makes no formal hydrogen bond to the fibril (Figure 3d). When docked to the same site, the known PET ligands GTP-1 and MK-6240 would have placed among the top 5 to 7% best-scoring molecules in this screen of 22 million (the ranking of the other 14 known ligands used in the retrospective study is in Table S3). The poses of the molecules taken from ZINC22 lack

experimental testing should not carry too much weight, but they do support the plausibility of large library screens using a symmetry-based ligand stacking approach.

Prospective Pose Prediction. Encouraged by the retrospective results and the ability to generate plausible geometries from large library screens, we tried the method for a genuine prospective pose prediction. Aware that the structure of the PET ligand MK-6240 was being determined against AD PHF tau fibrils, we docked that molecule against the structure of AD PHF tau in its GTP-1 complex using SymDOCK. When we compared the predicted docked poses to the subsequently determined cryo-EM electron density, we found that the ninth-best scoring pose came close to fitting the electron density (Figure 4a,b) and captured the symmetry and the overall placement of the ligand stack. The one discrepancy was that the ligand stack extended out of the cryo-EM density in its pyrrolo-pyridine ring and did not completely fill it on the other side of the molecule, indicating that a simple rigid-body translation of the SymDOCK pose could better fit the density. Comparing the final refined position for the ligand stack to that predicted by the docking, the RMSD of 2.13 Å can be mostly attributed to a rotation about the axis of the fibril (Figure 4c,d). Indeed, the structure was close enough that the docked pose was used as the input for refinement of the final experimental structure of the MK-6240/AD PHF tau complex.⁵⁸ We note that although this was the ninth-best scoring structure, the best pose by DOCK score had an RMSD of 3.01 Å from the experimental structure, and what is effectively the second-best scoring pose had an RMSD of 2.27 Å to that structure. From a virtual screening standpoint, either of these top-ranking poses would have been sufficient to identify MK-6240 as a plausible candidate even though it was the ninth best pose that we used to prospectively predict the experimental structure and indeed help in its experimental determination.

DISCUSSION

An arresting feature of the recent tau fibril structures is the adoption of symmetrical stacks by their ligands (Figures 2 and 4). Ligand–receptor interactions are few in these structures, with the most extensive contacts coming between the ligands themselves. For a receptor with multiple symmetrical sites, the adoption of such ligand stacks may be the favorable mode. Considerations of entropy coming from the number of sites and cooperativity from the ligand–ligand interactions may explain how molecules with so few fibril interactions may nevertheless bind in the nanomolar range, as many of the PET ligands do. A challenge for ligand discovery is exploiting the symmetrical fibril sites and ligand–ligand interactions, as most docking methods anticipate a 1:1 ligand-to-receptor stoichiometry and are designed to explore multiple ligand poses unencumbered by symmetry or stacking.

The SymDOCK approach adapts docking by imposing the symmetry of the fibril on any ligand geometry, insisting that there are no ligand–ligand van der Waals clashes. The method succeeds retrospectively in finding the experimental poses of four topologically unrelated PET ligands to tau (Figures 2 and 4). Although the symmetry calculation and the internal clash checks add operations to docking, these are compensated by the constraint imposed on pose selection by insisting on symmetry, and so the method loses little speed. A user can thus perform large library screens seeking new ligands. Although these predictions remain to be tested experimentally, known

fibril ligands are enriched both against property matched decoys (Figure S9) and in screens of 22 million diverse molecules (Table S3). The new ligands' poses seem plausible, forming stacks around planar aromatic and heteroaromatic systems while preserving the symmetry of the fibril (Figure 3). In a genuine prospective prediction, the method correctly predicts the binding pose of MK-6240 to AD PHF tau (Figure 4), and that docking pose was in fact used in the solution of the ligand-complex structure as determined by cryo-EM.⁵⁸

Several limitations of SymDOCK merit airing. For speed of calculation, we have used a 2 Å cutoff to define ligand–ligand clashes, which misses those owing to large radii atoms like bromine or iodine. Only evaluating the favorability of a stacking geometry using the AMBER 4.0 united-atom force field's van der Waals parameters ignores higher-order effects from π – π stacking and electrostatics.^{40,41} We have explored compensating for this with more detailed evaluation using the DFT-derived ANI,^{34,35} although this is slow enough to limit application to rescoring post-docking. We certainly have not optimized to balance the self-stacking versus the ligand–protein interactions other than relying on the same scoring function for both—if, for some ligands, the balance is off, this could lead to false negatives and false positives. We currently only use the symmetry operation (affine transformation) from the protein, but it is possible that other symmetries of a ligand stack could fit the pocket. For example, one could imagine ligands that bind in a 2:1 stoichiometry by forming stacks of molecules alternating in orientation with respect to the protein by 180° rotations, and more complex repeat patterns among the ligands are also conceivable. Finally, we have shown that prospective large library screens are mechanically possible using SymDOCK and produce plausible symmetrical stacks, but prospective hits and binding modes will require experimental testing. For now, these predictions merely suggest that it should be possible to test this method for ligand discovery.

Notwithstanding these caveats, the key observations from this study should be clear. Imposing symmetry and excluding ligand–ligand clashes are effective approaches for docking ligands that form extensive, symmetric stacks against fibril proteins and potentially other proteins with symmetrical sites. Because we used a relatively simple approach to this problem, the method is fast enough for use in large screens. Encouragingly, by insisting on symmetry and excluding clashes, the method seems to capture the key features of the symmetrical ligand stacks despite ignoring higher-order energetic terms. We have implemented the symmetry docking approach in the program SymDOCK (freely available for academic research at <https://dock.compbio.ucsf.edu/>); we suspect that this approach may be readily adapted to most docking methods.^{59–75}

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jcim.3c01749>.

Additional methods describing the calculation of ANI energies; Tables S1 to S3 show energies and RMSD values before and after ANI refinement, known ligands of PHF tau used for enrichment calculations, and their ranks, respectively; and Figures S1 to S11 display further details on the symmetry calculations, calculation of

internal clashes among symmetry-docked molecules, ligand–protein docked complexes, effects of dielectric boundaries on docking scores, enrichment calculations, and how ANI optimization affects docked poses (PDF)

AUTHOR INFORMATION

Corresponding Author

Brian K. Shoichet – Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94158, United States; orcid.org/0000-0002-6098-7367; Email: bshoichet@gmail.com

Authors

Matthew S. Smith – Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94158, United States; Program in Biophysics, University of California, San Francisco, California 94143, United States

Ian S. Knight – Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94158, United States

Rian C. Kormos – Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94158, United States; Program in Biophysics, University of California, San Francisco, California 94143, United States

Joseph G. Pepe – Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94158, United States; Program in Biophysics, University of California, San Francisco, California 94143, United States

Peter Kunach – McGill Research Centre for Studies in Aging, McGill University, Montreal, Quebec H4H 1R3, Canada; Department of Neurology and Neurosurgery, McGill University, Montreal, Quebec H3A 1A1, Canada; Center for Alzheimer's and Neurodegenerative Diseases, Peter O'Donnell Jr. Brain Institute, University of Texas Southwestern Medical Center, Dallas, Texas 75390, United States; Department of Neurology, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9368, United States; Department of Neuroscience, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9111, United States

Marc I. Diamond – Center for Alzheimer's and Neurodegenerative Diseases, Peter O'Donnell Jr. Brain Institute, University of Texas Southwestern Medical Center, Dallas, Texas 75390, United States; Department of Neurology, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9368, United States; Department of Neuroscience, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9111, United States

Sarah H. Shahmoradian – Center for Alzheimer's and Neurodegenerative Diseases, Peter O'Donnell Jr. Brain Institute, University of Texas Southwestern Medical Center, Dallas, Texas 75390, United States; Department of Biophysics, University of Texas Southwestern Medical Center, Dallas, Texas 75390-8816, United States

John J. Irwin – Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94158, United States

William F. DeGrado – Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94158, United States; Cardiovascular Research Institute, University of California, San Francisco, California 94158-9001, United States

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.jcim.3c01749>

Author Contributions

The symmetry docking method was conceived of by M.S.S. and B.K.S. and implemented in DOCK3.8 by I.S.K. under the supervision of J.J.I. M.S.S. performed all docking calculations reported here, with assistance from J.G.P. for flortaucipir. R.C.K. developed the ANI-based stack filter under the supervision of W.F.D., and M.S.S. used it to analyze the SymDOCK-derived poses. P.K. ran the experiments and computational analysis to provide the electron density and protein model for the structure of MK-6240 bound to AD PHF tau, under the supervision of M.I.D. and S.H.S., as described in an accompanying manuscript.⁵⁸ All authors contributed to writing and editing the paper.

Funding

Supported by the US NIH grants R35GM122481 (to B.K.S.) and RF1AG065407 (to M.I.D.). R.C.K. was supported by a DoD NDSEG Fellowship.

Notes

The authors declare the following competing financial interest(s): B.K.S. serves on the SAB of Schrodinger, Vilya Therapeutics, and of Hyku Therapeutics, is a founder of Epiodyne, and with J.J.I. of Deep Apple Therapeutics and BlueDolphin Leads LLC, and consults for Great Point Ventures and for Levator Therapeutics.

[¶]M.S.S. and I.S.K. are co-first authors.

SymDOCK is available without charge for academic research as part of the DOCK3.8 suite of programs at <https://dock.compbio.ucsf.edu/>.

REFERENCES

- (1) Cleveland, D. W.; Hwo, S.-Y.; Kirschner, M. W. Purification of Tau, a Microtubule-Associated Protein That Induces Assembly of Microtubules from Purified Tubulin. *J. Mol. Biol.* **1977**, *116* (2), 207–225.
- (2) Cleveland, D. W.; Hwo, S.-Y.; Kirschner, M. W. Physical and Chemical Properties of Purified Tau Factor and the Role of Tau in Microtubule Assembly. *J. Mol. Biol.* **1977**, *116* (2), 227–247.
- (3) Kadavath, H.; Jaremko, M.; Jaremko, L.; Biernat, J.; Mandelkow, E.; Zweckstetter, M. Folding of the Tau Protein on Microtubules. *Angew. Chem., Int. Ed.* **2015**, *54* (35), 10347–10351.
- (4) Kidd, M. Paired Helical Filaments in Electron Microscopy of Alzheimer's Disease. *Nature* **1963**, *197* (4863), 192–193.
- (5) Wischik, C. M.; Novak, M.; Edwards, P. C.; Klug, A.; Tichelaar, W.; Crowther, R. A. Structural Characterization of the Core of the Paired Helical Filament of Alzheimer Disease. *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85* (13), 4884–4888.
- (6) Coedert, M.; Spillantini, M. G.; Jakes, D.; Rutherford, D.; Crowther, R. A. Multiple Isoforms of Human Microtubule-Associated Protein Tau: Sequences and Localization in Neurofibrillary Tangles of Alzheimer's Disease. *Neuron* **1989**, *3*, 519 DOI: [10.1016/0896-6273\(89\)90210-9](https://doi.org/10.1016/0896-6273(89)90210-9).
- (7) Crowther, R. A. Straight and Paired Helical Filaments in Alzheimer Disease Have a Common Structural Unit. *Proc. Natl. Acad. Sci. U. S. A.* **1991**, *88* (6), 2288–2292.
- (8) Tokuda, T.; Ikeda, S.; Yanagisawa, N.; Ihara, Y.; Glenner, G. G. Re-Examination of Ex-Boxers' Brains Using Immunohistochemistry with Antibodies to Amyloid β -Protein and Tau Protein. *Acta Neuropathol. (Berl.)* **1991**, *82* (4), 280–285.
- (9) Goedert, M.; Spillantini, M. G.; Cairns, N. J.; Crowther, R. A. Tau Proteins of Alzheimer Paired Helical Filaments: Abnormal Phosphorylation of All Six Brain Isoforms I. *Neuron* **1992**, *8*, 159–168.

- (10) Hof, P. R.; Bouras, C.; Buée, L.; Delacourte, A.; Perl, D. P.; Morrison, J. H. Differential Distribution of Neurofibrillary Tangles in the Cerebral Cortex of Dementia Pugilistica and Alzheimer's Disease Cases. *Acta Neuropathol.* **1992**, *85* (1), 23–30.
- (11) Fitzpatrick, A. W. P.; Falcon, B.; He, S.; Murzin, A. G.; Murshudov, G.; Garringer, H. J.; Crowther, R. A.; Ghetti, B.; Goedert, M.; Scheres, S. H. W. Cryo-EM Structures of Tau Filaments from Alzheimer's Disease. *Nature* **2017**, *547* (7662), 185–190.
- (12) Falcon, B.; Zhang, W.; Murzin, A. G.; Murshudov, G.; Garringer, H. J.; Vidal, R.; Crowther, R. A.; Ghetti, B.; Scheres, S. H. W.; Goedert, M. Structures of Filaments from Pick's Disease Reveal a Novel Tau Protein Fold. *Nature* **2018**, *561* (7721), 137–140.
- (13) Falcon, B.; Zhang, W.; Schweighauser, M.; Murzin, A. G.; Vidal, R.; Garringer, H. J.; Ghetti, B.; Scheres, S. H. W.; Goedert, M. Tau Filaments from Multiple Cases of Sporadic and Inherited Alzheimer's Disease Adopt a Common Fold. *Acta Neuropathol. (Berl.)* **2018**, *136* (5), 699–708.
- (14) Falcon, B.; Zivanov, J.; Zhang, W.; Murzin, A. G.; Garringer, H. J.; Vidal, R.; Crowther, R. A.; Newell, K. L.; Ghetti, B.; Goedert, M.; Scheres, S. H. W. Novel Tau Filament Fold in Chronic Traumatic Encephalopathy Encloses Hydrophobic Molecules. *Nature* **2019**, *568* (7752), 420–423.
- (15) Zhang, W.; Tarutani, A.; Newell, K. L.; Murzin, A. G.; Matsubara, T.; Falcon, B.; Vidal, R.; Garringer, H. J.; Shi, Y.; Ikeuchi, T.; Murayama, S.; Ghetti, B.; Hasegawa, M.; Goedert, M.; Scheres, S. H. W. Novel Tau Filament Fold in Corticobasal Degeneration. *Nature* **2020**, *580* (7802), 283–287.
- (16) Shi, Y.; Zhang, W.; Yang, Y.; Murzin, A. G.; Falcon, B.; Kotecha, A.; Van Beers, M.; Tarutani, A.; Kametani, F.; Garringer, H. J.; Vidal, R.; Hallinan, G. I.; Lashley, T.; Saito, Y.; Murayama, S.; Yoshida, M.; Tanaka, H.; Kakita, A.; Ikeuchi, T.; Robinson, A. C.; Mann, D. M. A.; Kovacs, G. G.; Revesz, T.; Ghetti, B.; Hasegawa, M.; Goedert, M.; Scheres, S. H. W. Structure-Based Classification of Tauopathies. *Nature* **2021**, *598* (7880), 359–363.
- (17) Zhang, W.; Falcon, B.; Murzin, A. G.; Fan, J.; Crowther, R. A.; Goedert, M.; Scheres, S. H. Heparin-Induced Tau Filaments Are Polymorphic and Differ from Those in Alzheimer's and Pick's Diseases. *eLife* **2019**, *8*, No. e43584.
- (18) Seidler, P. M.; Murray, K. A.; Boyer, D. R.; Ge, P.; Sawaya, M. R.; Hu, C. J.; Cheng, X.; Abskharon, R.; Pan, H.; DeTure, M. A.; Williams, C. K.; Dickson, D. W.; Vinters, H. V.; Eisenberg, D. S. Structure-Based Discovery of Small Molecules That Disaggregate Alzheimer's Disease Tissue Derived Tau Fibrils in Vitro. *Nat. Commun.* **2022**, *13* (1), 5451.
- (19) Merz, G. E.; Chalkley, M. J.; Tan, S. K.; Tse, E.; Lee, J.; Prusiner, S. B.; Paras, N. A.; DeGrado, W. F.; Southworth, D. R. Stacked Binding of a PET Ligand to Alzheimer's Tau Paired Helical Filaments. *Nat. Commun.* **2023**, *14* (1), 3048.
- (20) Walji, A. M.; Hostetler, E. D.; Selnick, H.; Zeng, Z.; Miller, P.; Bennacef, I.; Salinas, C.; Connolly, B.; Gantert, L.; Holahan, M.; O'Malley, S.; Purcell, M.; Riffel, K.; Li, J.; Balsells, J.; O'Brien, J. A.; Melquist, S.; Soriano, A.; Zhang, X.; Ogawa, A.; Xu, S.; Joshi, E.; Della Rocca, J.; Hess, F. J.; Schachter, J.; Hesk, D.; Schenk, D.; Struyk, A.; Babaoglu, K.; Lohith, T. G.; Wang, Y.; Yang, K.; Fu, J.; Evelhoch, J. L.; Coleman, P. J. Discovery of 6-(Fluoro-¹⁸F)-3-(1H-Pyrrolo[2,3-c]Pyridin-1-Yl)isoquinolin-5-Amine ([¹⁸F]-MK-6240): A Positron Emission Tomography (PET) Imaging Agent for Quantification of Neurofibrillary Tangles (NFTs). *J. Med. Chem.* **2016**, *59* (10), 4778–4789.
- (21) Shi, Y.; Ghetti, B.; Goedert, M.; Scheres, S. H. W. Cryo-EM Structures of Chronic Traumatic Encephalopathy Tau Filaments with PET Ligand Flortaucipir. *J. Mol. Biol.* **2023**, *435* (11), No. 168025.
- (22) Shi, Y.; Murzin, A. G.; Falcon, B.; Epstein, A.; Machin, J.; Tempest, P.; Newell, K. L.; Vidal, R.; Garringer, H. J.; Sahara, N.; Higuchi, M.; Ghetti, B.; Jang, M.-K.; Scheres, S. H. W.; Goedert, M. Cryo-EM Structures of Tau Filaments from Alzheimer's Disease with PET Ligand APN-1607. *Acta Neuropathol. (Berl.)* **2021**, *141* (5), 697–708.
- (23) Sanabria Bohórquez, S.; Marik, J.; Ogasawara, A.; Tinianow, J. N.; Gill, H. S.; Barret, O.; Tamagnan, G.; Allagille, D.; Ayalon, G.; Manser, P.; Bengtsson, T.; Ward, M.; Williams, S.-P.; Kerchner, G. A.; Seibyl, J. P.; Marek, K.; Weimer, R. M. [¹⁸F]GTP1 (Genentech Tau Probe 1), a Radioligand for Detecting Neurofibrillary Tangle Tau Pathology in Alzheimer's Disease. *Eur. J. Nucl. Med. Mol. Imaging* **2019**, *46* (10), 2077–2089.
- (24) Xia, C.; Arteaga, J.; Chen, G.; Gangadharmath, U.; Gomez, L. F.; Kasi, D.; Lam, C.; Liang, Q.; Liu, C.; Mocharla, V. P.; Mu, F.; Sinha, A.; Su, H.; Szardenings, A. K.; Walsh, J. C.; Wang, E.; Yu, C.; Zhang, W.; Zhao, T.; Kolb, H. C. [¹⁸F]T807, a Novel Tau Positron Emission Tomography Imaging Agent for Alzheimer's Disease. *Alzheimers Dement.* **2013**, *9* (6), 666–676.
- (25) Ono, M.; Sahara, N.; Kumata, K.; Ji, B.; Ni, R.; Koga, S.; Dickson, D. W.; Trojanowski, J. Q.; Lee, V. M.-Y.; Yoshida, M.; Hozumi, I.; Yoshiyama, Y.; Van Swieten, J. C.; Nordberg, A.; Sahara, T.; Zhang, M.-R.; Higuchi, M. Distinct Binding of PET Ligands PBB3 and AV-1451 to Tau Fibril Strains in Neurodegenerative Tauopathies. *Brain* **2017**, No. aww339.
- (26) Fleisher, A. S.; Pontecorvo, M. J.; Devous, M. D.; Lu, M.; Arora, A. K.; Truocchio, S. P.; Aldea, P.; Flitter, M.; Locascio, T.; Devine, M.; Siderowf, A.; Beach, T. G.; Montine, T. J.; Serrano, G. E.; Curtis, C.; Perrin, A.; Salloway, S.; Daniel, M.; Wellman, C.; Joshi, A. D.; Irwin, D. J.; Lowe, V. J.; Seeley, W. W.; Ikonovic, M. D.; Masdeu, J. C.; Kennedy, I.; Harris, T.; Navitsky, M.; Southekal, S.; Mintun, M. A. for the A16 Study Investigators. Positron Emission Tomography Imaging With [¹⁸F]Flortaucipir and Postmortem Assessment of Alzheimer Disease Neuropathologic Changes. *JAMA Neurol.* **2020**, *77* (7), 829.
- (27) Gogola, A.; Minhas, D. S.; Villemagne, V. L.; Cohen, A. D.; Mountz, J. M.; Pascoal, T. A.; Laymon, C. M.; Mason, N. S.; Ikonovic, M. D.; Mathis, C. A.; Snitz, B. E.; Lopez, O. L.; Klunk, W. E.; Lopresti, B. J. Direct Comparison of the Tau PET Tracers ¹⁸F-Flortaucipir and ¹⁸F-MK-6240 in Human Subjects. *J. Nucl. Med.* **2022**, *63* (1), 108–116.
- (28) Mathis, C. A.; Wang, Y.; Holt, D. P.; Huang, G.-F.; Debnath, M. L.; Klunk, W. E. Synthesis and Evaluation of ¹¹C-Labeled 6-Substituted 2-Arylbenzothiazoles as Amyloid Imaging Agents. *J. Med. Chem.* **2003**, *46* (13), 2740–2754.
- (29) Klunk, W. E.; Engler, H.; Nordberg, A.; Wang, Y.; Blomqvist, G.; Holt, D. P.; Bergström, M.; Savitcheva, I.; Huang, G.; Estrada, S.; Ausén, B.; Debnath, M. L.; Barletta, J.; Price, J. C.; Sandell, J.; Lopresti, B. J.; Wall, A.; Koivisto, P.; Antoni, G.; Mathis, C. A.; Långström, B. Imaging Brain Amyloid in Alzheimer's Disease with Pittsburgh Compound-B. *Ann. Neurol.* **2004**, *55* (3), 306–319.
- (30) Fagan, A. M.; Mintun, M. A.; Mach, R. H.; Lee, S.; Dence, C. S.; Shah, A. R.; LaRossa, G. N.; Spinner, M. L.; Klunk, W. E.; Mathis, C. A.; DeKosky, S. T.; Morris, J. C.; Holtzman, D. M. Inverse Relation between in Vivo Amyloid Imaging Load and Cerebrospinal Fluid A β 42 in Humans. *Ann. Neurol.* **2006**, *59* (3), 512–519.
- (31) Coleman, R. G.; Carchia, M.; Sterling, T.; Irwin, J. J.; Shoichet, B. K. Ligand Pose and Orientational Sampling in Molecular Docking. *PLoS One* **2013**, *8* (10), No. e75992.
- (32) Lyu, J.; Wang, S.; Balias, T. E.; Singh, I.; Levit, A.; Moroz, Y. S.; O'Meara, M. J.; Che, T.; Alga, E.; Tolmacheva, K.; Tolmachev, A. A.; Shoichet, B. K.; Roth, B. L.; Irwin, J. J. Ultra-Large Library Docking for Discovering New Chemotypes. *Nature* **2019**, *566* (7743), 224–229.
- (33) Shuo, Gu; Smith, M. S.; Yang, Y.; Irwin, J. J.; Shoichet, B. K. Ligand Strain Energy in Large Library Docking. *J. Chem. Inf. Model.* **2021**, *61* (9), 4331–4341.
- (34) Devereux, C.; Smith, J. S.; Huddleston, K. K.; Barros, K.; Zubatyuk, R.; Isayev, O.; Roitberg, A. E. Extending the Applicability of the ANI Deep Learning Molecular Potential to Sulfur and Halogens. *J. Chem. Theory Comput.* **2020**, *16* (7), 4192–4202.
- (35) Gao, X.; Ramezanghorbani, F.; Isayev, O.; Smith, J. S.; Roitberg, A. E. TorchANI: A Free and Open Source PyTorch-Based Deep Learning Implementation of the ANI Neural Network Potentials. *J. Chem. Inf. Model.* **2020**, *60* (7), 3408–3415.

- (36) Tingle, B. I.; Tang, K. G.; Castanon, M.; Gutierrez, J. J.; Khurelbaatar, M.; Dandarchuluun, C.; Moroz, Y. S.; Irwin, J. J. ZINC-22—A Free Multi-Billion-Scale Database of Tangible Compounds for Ligand Discovery. *J. Chem. Inf. Model.* **2023**, *63* (4), 1166–1176.
- (37) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. A Geometric Approach to Macromolecule-Ligand Interactions. *J. Mol. Biol.* **1982**, *161* (2), 269–288.
- (38) Shoichet, B. K.; Kuntz, I. D.; Bodian, D. L. Molecular Docking Using Shape Descriptors. *J. Comput. Chem.* **1992**, *13* (3), 380–397.
- (39) Mysinger, M. M.; Shoichet, B. K. Rapid Context-Dependent Ligand Desolvation in Molecular Docking. *J. Chem. Inf. Model.* **2010**, *50* (9), 1561–1573.
- (40) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins. *J. Am. Chem. Soc.* **1984**, *106* (3), 765–784.
- (41) Case, D. A.; Aktulga, H. M.; Belfon, K.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E. I.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Giambasu, G.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Harris, R.; Izadi, S.; Izmailov, S. A.; Jin, C.; Kasavajhala, K.; Kaymak, M. C.; King, E.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Machado, M.; Man, V.; Manathunga, M.; Merz, K. M.; Miao, Y.; Mikhailovskii, O.; Monard, G.; Nguyen, H.; O’Hearn, K. A.; Onufriev, A.; Pan, F.; Pantano, S.; Qi, R.; Rahnamoun, A.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Skrynnikov, N. R.; Smith, J.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.; Wolf, R. M.; Wu, X.; Xue, Y.; York, D. M.; Zhao, S.; Kollman, P. A. *Amber*, 2021 University of California. <https://ambermd.org/index.php>.
- (42) Knight, I.; Tang, K.; Mailhot, O.; Irwin, J. DockOpt: A Tool for Automatic Optimization of Docking Models; *ChemRxiv* preprint; Chemistry, 2023.
- (43) R Core Team. *R: A Language and Environment for Statistical Computing* R Foundation for Statistical Computing, 2019, <https://www.R-project.org/>.
- (44) Wickham, H.; Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*, 2016, <https://ggplot2.tidyverse.org>.
- (45) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.* **1953**, *21* (6), 1087–1092.
- (46) Leach, A.; Schmon, S. M.; Degiacomi, M. T.; Willcocks, C. G. DENOISING DIFFUSION PROBABILISTIC MODELS ON SO(3) FOR ROTATIONAL ALIGNMENT. In *ICLR 2022 Workshop on Geometrical and Topological Representation Learning* 2022.
- (47) Irwin, J. J.; Shoichet, B. K.; Mysinger, M. M.; Huang, N.; Colizzi, F.; Wassam, P.; Cao, Y. Automated Docking Screens: A Feasibility Study. *J. Med. Chem.* **2009**, *52* (18), 5712–5720.
- (48) Lorber, D.; Shoichet, B. Hierarchical Docking of Databases of Multiple Ligand Conformations. *Curr. Top. Med. Chem.* **2005**, *5* (8), 739–749.
- (49) Abbas, W.; Hostetler, E.; Greshock, T.; Li, J.; Moore, K. P.; Bennacef, I.; Mulhearn, J.; Selnick, H.; Wang, Y.; Yang, K.; Fu, J. PYRROLO[2,3-C]PYRIDINES AS IMAGING AGENTS FOR NEUROFIBRILARY TANGLES. WO2015188368A1 <https://patentscope.wipo.int/search/en/detail.jsf?docId=WO2015191506> 2014.
- (50) Mysinger, M. M.; Carchia, M.; Irwin, J. J.; Shoichet, B. K. Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. *J. Med. Chem.* **2012**, *55* (14), 6582–6594.
- (51) Stein, R. M.; Yang, Y.; Balius, T. E.; O’Meara, M. J.; Lyu, J.; Young, J.; Tang, K.; Shoichet, B. K.; Irwin, J. J. Property-Unmatched Decoys in Docking Benchmarks. *J. Chem. Inf. Model.* **2021**, *61* (2), 699–714.
- (52) Clark, R. D.; Webster-Clark, D. J. Managing Bias in ROC Curves. *J. Comput. Aided Mol. Des.* **2008**, *22* (3–4), 141–146.
- (53) Wildman, S. A.; Crippen, G. M. Prediction of Physicochemical Parameters by Atomic Contributions. *J. Chem. Inf. Comput. Sci.* **1999**, *39* (5), 868–873.
- (54) Landrum, G. *RDKit: Open-Source Cheminformatics*, 2010. <https://www.rdkit.org/>.
- (55) Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P. *Mol. Biol. Cell*, 4th ed.; Garland Science: New York, NY, 2002.
- (56) Madoori, P. K.; Agustindari, H.; Driessen, A. J. M.; Thunnissen, A.-M. W. H. Structure of the Transcriptional Regulator LmrR and Its Mechanism of Multidrug Recognition. *EMBO J.* **2009**, *28* (2), 156–166.
- (57) Shao, J.; Kuiper, B. P.; Thunnissen, A.-M. W. H.; Cool, R. H.; Zhou, L.; Huang, C.; Dijkstra, B. W.; Broos, J. The Role of Tryptophan in π Interactions in Proteins: An Experimental Approach. *J. Am. Chem. Soc.* **2022**, *144* (30), 13815–13822.
- (58) Kunach, P.; Vaquer-Alicea, J.; Smith, M. S.; Hopewell, R.; Monistrol, J.; Moquin, L.; Theriault, J.; Tissot, C.; Rahmouni, N.; Massarweh, G.; Soucy, J.-P.; Guiot, M.-C.; Shoichet, B. K.; Rosa-Neto, P.; Diamond, M. I.; Shahmoradian, S. H. Cryo-EM Structure of Alzheimer’s Disease Tau Filaments with PET Ligand MK-6240; preprint; Neuroscience, *bioRxiv* **2023**. DOI: 10.1101/2023.09.22.558671.
- (59) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A New Approach for Rapid, Accurate Docking and Scoring. I. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47* (7), 1739–1749.
- (60) Sherman, W.; Day, T.; Jacobson, M. P.; Friesner, R. A.; Farid, R. Novel Procedure for Modeling Ligand/Receptor Induced Fit Effects. *J. Med. Chem.* **2006**, *49* (2), 534–553.
- (61) Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; Romero, D. L.; Masse, C.; Knight, J. L.; Steinbrecher, T.; Beuming, T.; Damm, W.; Harder, E.; Sherman, W.; Brewer, M.; Wester, R.; Murcko, M.; Frye, L.; Farid, R.; Lin, T.; Mobley, D. L.; Jorgensen, W. L.; Berne, B. J.; Friesner, R. A.; Abel, R. Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. *J. Am. Chem. Soc.* **2015**, *137* (7), 2695–2703.
- (62) Abel, R.; Wang, L.; Harder, E. D.; Berne, B. J.; Friesner, R. A. Advancing Drug Discovery through Enhanced Free Energy Calculations. *Acc. Chem. Res.* **2017**, *50* (7), 1625–1632.
- (63) Lu, C.; Wu, C.; Ghoreishi, D.; Chen, W.; Wang, L.; Damm, W.; Ross, G. A.; Dahlgren, M. K.; Russell, E.; Von Bargen, C. D.; Abel, R.; Friesner, R. A.; Harder, E. D. OPLS4: Improving Force Field Accuracy on Challenging Regimes of Chemical Space. *J. Chem. Theory Comput.* **2021**, *17* (7), 4291–4300.
- (64) Totrov, M.; Abagyan, R. Flexible Protein–Ligand Docking by Global Energy Optimization in Internal Coordinates. *Proteins Struct. Funct. Genet.* **1997**, *29* (S1), 215–220.
- (65) Katritch, V.; Totrov, M.; Abagyan, R. ICFF: A New Method to Incorporate Implicit Flexibility into an Internal Coordinate Force Field. *J. Comput. Chem.* **2003**, *24* (2), 254–265.
- (66) Bottegoni, G.; Kufareva, I.; Totrov, M.; Abagyan, R. Four-Dimensional Docking: A Fast and Accurate Account of Discrete Receptor Flexibility in Ligand Docking. *J. Med. Chem.* **2009**, *52* (2), 397–406.
- (67) Sadybekov, A. A.; Sadybekov, A. V.; Liu, Y.; Iliopoulos-Tsoutsouvas, C.; Huang, X.-P.; Pickett, J.; Houser, B.; Patel, N.; Tran, N. K.; Tong, F.; Zvonok, N.; Jain, M. K.; Savych, O.; Radchenko, D. S.; Nikas, S. P.; Petasis, N. A.; Moroz, Y. S.; Roth, B. L.; Makriyannis, A.; Katritch, V. Synthon-Based Ligand Discovery in Virtual Libraries of over 11 Billion Compounds. *Nature* **2022**, *601* (7893), 452–459.
- (68) Yang, Y.; Yao, K.; Repasky, M. P.; Leswing, K.; Abel, R.; Shoichet, B. K.; Jerome, S. V. Efficient Exploration of Chemical Space with Docking and Deep Learning. *J. Chem. Theory Comput.* **2021**, *17* (11), 7106–7119.
- (69) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and Validation of a Genetic Algorithm for Flexible Docking. *J. Mol. Biol.* **1997**, *267* (3), 727–748.

(70) Forli, S.; Huey, R.; Pique, M. E.; Sanner, M. F.; Goodsell, D. S.; Olson, A. J. Computational Protein–Ligand Docking and Virtual Drug Screening with the AutoDock Suite. *Nat. Protoc.* **2016**, *11* (5), 905–919.

(71) Eberhardt, J.; Santos-Martins, D.; Tillack, A. F.; Forli, S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J. Chem. Inf. Model.* **2021**, *61* (8), 3891–3898.

(72) Gorgulla, C.; Boeszoeremnyi, A.; Wang, Z.-F.; Fischer, P. D.; Coote, P. W.; Padmanabha Das, K. M.; Malets, Y. S.; Radchenko, D. S.; Moroz, Y. S.; Scott, D. A.; Fackeldey, K.; Hoffmann, M.; Iavniuk, I.; Wagner, G.; Arthanari, H. An Open-Source Drug Discovery Platform Enables Ultra-Large Virtual Screens. *Nature* **2020**, *580* (7805), 663–668.

(73) Gorgulla, C.; Nigam, A.; Koop, M.; Selim Çınaroğlu, S.; Secker, C.; Haddadnia, M.; Kumar, A.; Malets, Y.; Hasson, A.; Li, M.; Tang, M.; Levin-Konigsberg, R.; Radchenko, D.; Kumar, A.; Gehev, M.; Aquilanti, P.-Y.; Gabb, H.; Alhossary, A.; Wagner, G.; Aspuru-Guzik, A.; Moroz, Y. S.; Fackeldey, K.; Arthanari, H. VirtualFlow 2.0 - The Next Generation Drug Discovery Platform Enabling Adaptive Screens of 69 Billion Molecules; preprint; Bioinformatics, *bioRxiv*, 2023. DOI: [10.1101/2023.04.25.537981](https://doi.org/10.1101/2023.04.25.537981).

(74) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A Fast Flexible Docking Method Using an Incremental Construction Algorithm. *J. Mol. Biol.* **1996**, *261* (3), 470–489.

(75) Schellhammer, I.; Rarey, M. FlexX-Scan: Fast, Structure-based Virtual Screening. *Proteins Struct. Funct. Bioinforma.* **2004**, *57* (3), 504–517.