JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

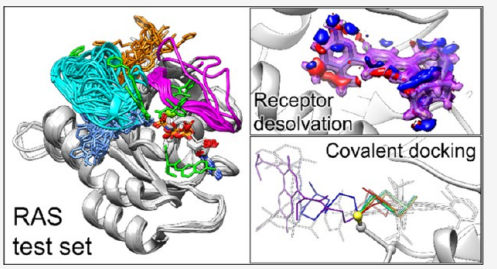
pubs.acs.org/jcim Article

Development of Receptor Desolvation Scoring and Covalent Sampling in DOCK 6: Methods Evaluated on a RAS Test Set

Y. Stanley Tan, Mayukh Chakrabarti, Reed M. Stein, Lauren E. Prentis, Robert C. Rizzo, Tom Kurtzman, Marcus Fischer, and Trent E. Balius*

Cite This: J. Chem. Inf. Model. 2025, 65, 722−748 Read Online

ABSTRACT: Molecular docking methods are widely used in drug discovery efforts. RAS proteins are important cancer drug targets, and are useful systems for evaluating docking methods, including accounting for solvation effects and covalent small molecule binding. Water often plays a key role in small molecule binding to RAS proteins, and many inhibitors—including FDA-approved drugs—covalently bind to oncogenic RAS proteins. We assembled a RAS test set, consisting of 138 RAS protein structures and 2 structures of KRAS DNA in complex with ligands. In DOCK 6, we have implemented a receptor desolvation scoring function and a covalent docking algorithm. These new features were evaluated using the test set, with pose reproduction, cross-docking, and enrichment calculations. We tested two solvation methods for generating receptor desolvation scoring grids: GIST and 3D-RISM. Using grids from GIST or 3D-RISM, water displacements are precomputed with Gaussian-weighting, and trilinear interpolation is used to speed up this scoring calculation. To test receptor desolvation scoring, we prepared GIST and 3D-RISM grids for all KRAS systems in the test set, and we compare enrichment performance with and without receptor desolvation. Accounting for receptor desolvation using GIST improves enrichment for 51% of systems and worsens enrichment for 35% of systems, while using 3D-RISM improves enrichment for 44% of systems and worsens enrichment for 30% of systems. To more rigorously test accounting for receptor desolvation using 3D-RISM, we compare pose reproduction with and without 3D-RISM receptor desolvation. Pose reproduction docking with 3D-RISM yields a 1.8 ± 2.41% increase in success rate compared to docking without 3D-RISM. Accounting for receptor desolvation provides a small, but significant, improvement in both enrichment and pose reproduction for this set. We tested the covalent attach-and-grow algorithm on 70 KRAS systems containing covalent ligands, obtaining similar pose reproduction success rates between covalent and noncovalent docking. Comparing covalent docking to noncovalent docking, there is a 2.4 ± 3.29% increase and a 1.27 ± 3.33% decline in the success rate when docking with experimental and SMILES-generated ligand conformations, respectively. As a proof-of-concept, we performed covalent virtual screens with and without receptor desolvation scoring, targeting the switch II pocket of KRAS, using 3.4 million make-on-demand acrylamide compounds from the Enamine REAL database. On average, the attach-and-grow algorithm spends approximately 17.61 s per molecule across the screen. The test set is available at https://github.com/tbalius/teb_docking_test_sets.



1. INTRODUCTION

RAS proteins play an important role in cellular signaling, particularly in modulating pathways governing cell growth and proliferation. RAS genes (*KRAS*, *HRAS*, and *NRAS*) are among the most prevalent oncogenes, with mutations occurring in ∼20% of human cancers.[1−3] Thus, RAS is an important and long sought-after cancer drug target. RAS proteins are small GTPases that bind a nucleotide, either GTP or GDP, and function as binary switches.[4] GTP-bound RAS is in an "active" state, and GDP-bound RAS is in an "inactive state". In recent years, there have been an influx of structures of RAS proteins in complex with inhibitors, including approved drugs.[5,6] This structural compendium[7] provides an excellent opportunity to develop benchmarks for evaluating the effectiveness of new docking software features that can contribute to virtual screening campaigns against this challenging target.

To assess therapeutically useful areas of chemical space, molecular docking methods are commonly used in drug discovery campaigns.[8] Numerous docking software are currently available, including AutoDock Vina,[9] Glide,[10] FRED,[11] GNINA,[12] FlexX,[13] rDOCK,[14] SEED,[15] PLANTS,[16] Gold,[17] DiffDock,[18] and many others.[19,20] DOCK is the first docking software,[21] and has undergone extensive development over the past four decades. Two versions of DOCK are being actively developed: DOCK 3 and DOCK 6. Both versions have benefits

Received: September 9, 2024
Revised: December 4, 2024
Accepted: December 17, 2024
Published: January 6, 2025

and drawbacks; DOCK 3 is faster and is well tested, and DOCK 6 is modular and amenable to the development of new features. There are many potential areas of improvement for docking methods, including better accounting for water, and improving covalent docking methods. In recognition of these limitations, we are porting desirable features from DOCK 3 into DOCK 6 and augmenting them, e.g., the receptor desolvation scoring function, and implementing new features that have been inspired by them, e.g., the covalent docking method introduced herein. We discuss these methods further in the following sections.

Water has long been recognized as an important factor and driver in molecular binding,[22,23] and appears to play a role in small molecule binding to RAS.[24−26] There are many methods to account for the contribution of water to binding in biological systems.[27] Two methods grounded in statistical mechanics are inhomogeneous solvation theory (IST)[28−30] and 3D reference interaction site model (3D-RISM) theory.[31,32] IST has been used by us and others, and it has been applied to ligand discovery efforts.[33−37] Grid-based IST (GIST) has been useful in understanding the thermodynamics and energetics of waters on surfaces of proteins (as well as other molecules), and in their cavities and openings. These grids provide a visualization of the hydration patterns about the protein binding site.[38] Accounting for receptor desolvation during docking with GIST has been implemented in DOCK 3.[31,32] Many other methods have been used to account for solvation effects in molecular docking.[39−41] As an alternative to GIST, 3D-RISM can be used to generate solvation energy grids that can be readily incorporated into receptor desolvation scoring.

In addition to noncovalent drugs, there are many examples of covalent drugs,[42,43] including the first approved drug targeting KRAS.[44] Better covalent docking methods would aid the discovery of new covalent drugs. Many covalent docking approaches have been reported, including DOCKovalent,[45] CovDOCK,[46,47] covalent docking using AutoDock4,[48] and many others.[49] Most of these methods target cysteine residues. However, other residues may also be attacked, including serines, aspartates, and arginines, all of which have been targeted for RAS,[50−52] as well as lysines, tyrosines, threonines, and histidines.[53,54] There are many chemical moieties that may be included in compounds as attacking groups, including acrylamides, chloroacetamides, vinyl sulfones, disulfides, phenol esters, and many more.[55] DOCKovalent was implemented within DOCK 3, and it has been used in ligand discovery campaigns.[45,56]

In this work, we present three interconnected topics: (1) We assembled a RAS test set for evaluating docking performance on RAS to aid drug discovery and method development. We evaluated new docking methods on the RAS test set, using both enrichment calculations and pose reproduction experiments. (2) We implemented receptor desolvation scoring functions within DOCK 6, and we tested two solvation methods that account for receptor desolvation within these scoring functions: GIST and 3D-RISM. GIST and 3D-RISM both calculate similar solvation energetic quantities that are then stored on a grid; however, the two methods differ in the underlying theory. GIST is quantified by analyzing an MD simulation, while 3D-RISM numerically solves the Ornstein−Zernike equation with a closure approximation without simulation (see Section 2 for more information). We evaluated these two methods for accounting for receptor desolvation by comparing enrichment and pose reproduction performance on KRAS systems in the test set (and, for enrichment, on an alternative test set of 43 diverse targets). (3) We implemented a covalent docking algorithm in DOCK 6, called *attach-and-grow*. In this method, the molecule is first attached to the receptor at the specified covalent residue, and then the molecule is grown out to generate the pose. That is, user-specified covalent bond information is used to position the attachment point, then the molecule is revealed one segment at a time, and degrees of freedom are sampled, until the molecule is fully displayed. We evaluated the *attach-and-grow* algorithm across 70 covalent KRAS systems (and on an alternative test set of 207 diverse targets), and we compared its performance to noncovalent docking using the *anchor-and-grow* algorithm. Finally, we combined receptor desolvation scoring and covalent docking by performing proof-of-concept virtual screens with a purchasable library of over 3 million covalent compounds to demonstrate applicability of these methods to ligand discovery campaigns.

2. METHODS

2.1. Test Set Generation. The RAS test set is made up of KRAS, HRAS, and NRAS structures in complex with ligands (SI Section 1 and SI Tables S1−S3). For a full list of ligands and their corresponding PDB IDs, see Table S1. These structures differ in nucleotide state and in sequence (Tables S2, S3, and Figure S1). In all docking experiments, the nucleotide and, when present, the ion cofactors, are treated as part of the receptor.

To add systems to the RAS test set, structures from the Protein Data Bank[57] (PDB) are prepared for docking in a semiautomated manner. The preparation procedure is organized into two stages: processing the PDB files and preparing the docking setups. In addition to the RAS test set, we evaluated our methods on test sets of diverse targets consisting of 207 covalent systems from Scarpino et al.[49] and 43 systems from the DUDE-Z test set,[58] prepared in a similar manner.

In the first stage, the PDB files are processed. The files are separated into ligand, cofactor, and protein files, and each piece is prepared by adding hydrogens and partial charges.

(1) PDB files are separated into ligand, cofactor, and protein. A ligand−receptor pair is flagged as covalent if there is a link definition in the PDB file between the ligand and the receptor. Ligands flagged as covalent are prepared for both covalent and noncovalent docking (to exploit covalent systems for noncovalent docking evaluation, and to directly compare covalent and noncovalent docking). Covalently prepared ligands are docked using the covalent *attach-and-grow* method. Noncovalently prepared ligands are docked using two noncovalent docking methods: *anchor-and-grow* and hierarchical database search (see Section 2.4 below). For covalent systems, the residue that forms a covalent bond with the ligand is mutated to an alanine to avoid steric clashes.

(2) Cofactor files are protonated using UCSF Chimera.[59]

(3) Proteins are passed through Chimera's *Dock Prep*, where solvent (including all structural waters) and alternative side chains are removed, and missing rotamers of side chains are added. Hydrogens are added using the program *Reduce*.[60]

(4) As with the cofactors, experimental conformations of the ligands (*xtal*) are protonated using Chimera, and partial charges are assigned using AM1-BCC charges[61] calculated in Chimera using Amber's *sqm* program.[62]

https://doi.org/10.1021/acs.jcim.4c01623
J. Chem. Inf. Model. 2025, 65, 722−748

(5) The processed ligand, cofactor, and protein files are aligned to the same frame (aligned to PDB ID 6GJ8) using Chimera's *MatchMaker*.

If needed, the processed ligand, cofactor, and protein files are modified by hand, allowing us to salvage systems that fail to run through our automated pipeline. The files are then placed in a staging directory for convenient access.

In the second stage, the files for docking produced in the first stage are used to prepare the docking setups.

(6) Partial charges of the protonated nucleotides are generated with Amber's *antechamber* using the AM1-BCC model. The protonated receptor is opened in Chimera, and partial charges are added using the *Dock Prep* module. The nucleotide is added to the receptor *mol2* file using the *combine* command in Chimera. The charged nucleotide *mol2* files are converted to parameter files to be read into *blastermaster.py*; charges are taken from the *mol2* file, and Sybyl *mol2* atom types are converted into DOCK atom types (used to define the van der Waals parameters).

(7) The processed ligand *mol2* files containing the experimental structure from step 1 are sufficient for docking with *anchor-and-grow* using Grid Score. *AMSOL 7.1*[63] is also used to recalculate partial charges and the per-atom desolvation parameters for use with Chemgrid Score. To prepare covalent ligands noncovalently, the atom involved in the covalent bond is removed to ensure that the experimental pose can be reproduced without clashing. The ligand is then prepared in the same way as other noncovalent ligands. To prepare ligands covalently, the ligand with the covalent adduct (i.e., the γ sulfur, Sγ, and β carbon, Cβ) is protonated, and partial charges are added using Chimera and *AMSOL 7.1*. *AMSOL* is also used to calculate the per-atom desolvation parameters. Hydrogens associated with the covalent adduct are removed, and the partial charges of the two remaining heavy atoms are set to zero. The sum of the removed partial charges is distributed to the remaining atoms. The two covalent adduct heavy atoms are then converted to dummy atoms. The Sγ and Cβ atoms are renamed D1 and D2, respectively, and the atom type is changed to Du.

(8) Ligands are prepared from SMILES. SMILES of each ligand are obtained from the PDB (see SI Table S1). Using the covalent linkage information from the PDB files, the SMILES of covalent ligands are modified for noncovalent docking (SI Table S1). Ligands are converted from 2D SMILES format to a 3D file format for docking in a multistep process (discussed also in Section 2.6). The scripts to prepare ligands from SMILES are now distributed with DOCK 6.12. Briefly, protonation states of the ligands are determined using ChemAxon's *cxcalc*[64] (https://www.chemaxon.com). The protonated SMILES strings are converted to a single 3D conformation in the *mol2* file format using *Corina*.[65] *AMSOL 7.1* is used to calculate partial charges and per-atom desolvation parameters. The partial charges and desolvation parameters are added to the *mol2* file. This *mol2* file can be used for docking with *anchor-and-grow* using Chemgrid Score. Covalent ligands are prepared noncovalently using the same procedure, but the covalent adduct atom is removed from the SMILES. The procedure to prepare ligands covalently from SMILES is described below in Section 2.3.1.

(9) DOCK 6 grids and spheres are generated. Grids are used for scoring, and spheres are used for orienting the ligand into the binding site. The program *DMS*[66] is used to generate the molecular surface of the protein. The program *sphgen*[67] is used to generate surface spheres representing the inverse image of the protein. The program *sphere_selector* is used to pick the spheres within 8 Å of the ligand, and a box is generated encompassing the spheres. DOCK 6's *grid* program[68] is used to calculate the van der Waals grids (using an all-atom force field and 6−9 exponents) and Coulombic grids (using a distance dependent dielectric).[69]

(10) DOCK 3 grids and spheres are generated. We use the *blastermaster.py* script distributed with DOCK 3.7. DOCK 3 spheres, called *matching spheres*, are generated by converting the crystallographic ligand atom to spheres and supplementing with surface spheres from *sphgen*. The DOCK 3 grids are converted to a DOCK 6 compatible format using the *gconv* program distributed with DOCK 6.12 (updated to work with grids produced with *blastermaster*). *Blastermaster* is first run without the nucleotide. The nucleotide is then added to the protonated receptor file. *Blastermaster* is run a second time without the hydrogen generation step (*addNOhydrogensflag*) and using the van der Waals and partial charge parameters supplemented with the nucleotide parameters generated in step 6.

(11) Receptor desolvation scoring grids are generated using GIST and 3D-RISM, as described below in Sections 2.2.2−2.2.5.

Scripts for the semiautomated procedure used to prepare the test set, as well as the processed ligand, cofactor, and protein files for docking, have been made available on GitHub.

2.2. DOCK Scoring Functions. In this study, we evaluate three DOCK 6 scoring functions: (1) Grid Score, (2) Chemgrid Score, and (3) Chemgrid + RecDesolv Score. RecDesolv Score refers to the receptor desolvation scoring function, which can be used with grids produced by either the GIST or 3D-RISM solvation methods.

Grid Score is a two-component scoring function

$$E_{\text{Grid\_Score}} = E_{\text{vdw}} + E_{\text{coul}}$$

It has a van der Waals component (E_{vdw}) (all-atom, 6−9 attractive and repulsive exponents), and an electrostatic component using Coulomb's law (E_{coul}) with a distance dependent dielectric.[69]

Chemgrid Score is a three-component scoring function

$$E_{\text{Chemgrid\_Score}} = E_{\text{vdw}} + E_{\text{ele}} + E_{\text{lig\_desolv}}$$

It has a van der Waals component (E_{vdw}) (united-atom, 6−12 attractive and repulsive exponents), an electrostatic component calculated using a Poisson−Boltzmann potential grid (E_{ele}), and a ligand desolvation ($E_{\text{lig\_desolv}}$) component.[68,70,71]

RecDesolv Score can be called by itself for rescoring, although it is more useful when combined with the above functions. The receptor desolvation term is calculated as follows

$$E_{\text{rec\_desolv}} = C^{*}E_{\text{water\_displacement}}$$

Here, the water displacement term ($E_{\text{water\_displacement}}$) is scaled by a scaling factor (C) (see SI Section 8). The water

https://doi.org/10.1021/acs.jcim.4c01623
J. Chem. Inf. Model. 2025, 65, 722−748

displacement term is calculated using the receptor desolvation grids, discussed further below.

To incorporate receptor desolvation scoring, we call the Descriptor Score scoring function, which is used to combine Chemgrid Score with RecDesolv Score

$$E_{\text{Chemgrid+RecDesolv\_Score}} = E_{\text{vdw}} + E_{\text{ele}} + E_{\text{lig\_desolv}} + E_{\text{rec\_desolv}}$$

2.2.1. Receptor Desolvation. We implemented a receptor desolvation scoring function into DOCK 6. There are three types of calculations: (1) *GIST displacement*, which is the same as that presented in Balius et al.,[31] (2) *GIST blurry displacement*, in which Gaussian weighting is applied to the voxels based on proximity to the center of an atom to avoid issues with double counting, and (3) *GIST blurry trilinear displacement*, in which atomic displacements are precomputed to make the calculation fast. *GIST blurry trilinear displacement* approximates *GIST blurry displacement*, and was introduced previously.[32] See Section 2.2.6. for more information on these types of receptor desolvation calculations. We first discuss how the receptor desolvation scoring grids are produced. We then discuss how to calculate the displacements. We also substitute the grids generated using GIST with grids generated using 3D-RISM, which, in general, is faster and easier to run (SI Figure S2).

2.2.2. GIST. To generate receptor desolvation grids using GIST, we perform an all-atom molecular dynamics (MD) simulation of the receptor in a box of water, where the receptor heavy atoms are restrained and the water molecules sample about the receptor. This simulation is then postprocessed. A grid is laid over the region of interest, and for each grid point, or voxel, the following thermodynamic properties are calculated and stored as dx files: the receptor (solute)-water enthalpy (E_{sw}), the water−water enthalpy (E_{ww}), the water−oxygen occupancy (g_o), the translational entropy (S_{trans}), and the orientational entropy (S_{orient}). These quantities can be combined in different ways. The g_o grid is used to reference the water−water term to bulk water. The solute (receptor)-water enthalpy and water−water enthalpy (referenced to bulk) is combined to give the total enthalpy grid. The entropy grids are combined with the total enthalpy grid to give the total free energy grid. Here, we focus on the total enthalpy grid.

2.2.3. GIST Grid Generation. We generate GIST grids using the procedure described in Balius et al.,[31] with a few key differences. Briefly, the GIST grid generation procedure involves running an MD simulation, followed by running the GIST calculation to generate GIST grids. The GIST grids are then processed and combined to generate the GIST docking grids, and then blurred using the procedure described in Stein et al.[32] to generate precomputed grids.

As discussed in Section 2.1, *antechamber* was used to parametrize the cofactor, AM1-BCC partial charges were calculated, and *frcmod* and *prep* files were produced. Amber's *tleap* is used to add the force field, and to place the protein, magnesium, and nucleotide in a box of water. The water box has a margin of 10 Å in all directions from the complex. The ff14SB force field[72] was applied to the protein, the TIP3P water model[73] was applied to the solvent, 12-6-4 Li/Merz ion parameters[74] were used for the magnesium ion, and the GAFF force field,[75] supplemented with the *frcmod* file, was used to parametrize the cofactor. We use Amber 22 *pmemd.cuda* to perform minimization and equilibration, and to run the MD simulation. As an initial test, for 20 KRAS systems, we run a 100 ns simulation, in ten chunks of 10 ns each. To reduce the computational demands of running GIST, for all of the KRAS systems (N = 123), we run a shorter 10 ns simulation (see Section 3 for more information). To run MD simulations on systems with missing backbone regions, we capped the missing regions with ACE and NME groups. See SI Table S1.1 for systems with missing regions that were capped for MD simulations. DUDE-Z systems are prepared similarly, and GIST grids are generated using MD simulations of 10 ns with all termini capped (see SI Section 5).

We then run the GIST calculation, using *cpptraj* version 6.22.0 of Amber, to process the simulation data and generate GIST grids. We used an MPI-parallelization of GIST,[76] which provides a significant speed up over the nonparallelized calculation (see SI Section 2 and Figure S2). At this stage, the dimensions of the GIST grid box are determined. We ensured the GIST grid box completely contains the other grid boxes (e.g., the van der Waals grid box) to not impact sampling coverage. If the GIST grids are smaller, poses that extend outside the GIST grid are discarded, and many poses that would normally be scored will not be considered. Note that changes to the box size will have an impact on enrichment and pose reproduction.

The grids produced by the GIST calculation are combined to produce the docking GIST grid. We only use the enthalpy terms, and we scale the water−water term by two.[31,36] We apply a cap of ±3 kcal/mol/Å^3 to the values of the voxels, i.e., any voxel that exceeds a magnitude of 3 kcal/mol/Å^3 is set to the cap value. This energy capping step is used to remove extremely high magnitude energies, and the value of 3 kcal/mol/Å^3 used follows the procedure previously.[31] The grids produced at this stage are sufficient for docking with the GIST (full) displacement and GIST blurry displacement methods. We then apply Gaussian weighting and precompute displacements, as discussed above, to obtain the blurred GIST grids. The blurred grids are used for docking with the GIST blurry trilinear displacement method.

2.2.4. 3D-RISM. 3D-RISM is an alternative solvation model to GIST. Unlike GIST, 3D-RISM does not postprocess an MD simulation. Instead, 3D-RISM is an approximation of the molecular Ornstein−Zernike (OZ) equation,[77] derived from statistical mechanics. To solve the 3D-RISM equation, a closure relation is needed. Different approximate closures may be used, including Kovalenko−Hirata (KH),[78] partial series expansion of order n (PSEn),[79] and hyper-netted chain (HNC),[80,81] which are implemented in Amber.[82] The 3D-RISM and closure equations are solved on a 3D grid by iteratively converging on a solution,[83] accelerated using the modified direct inversion of the iterative subspace (MDIIS) method.[84] Convolution integrals in the 3D-RISM equation are evaluated in reciprocal space using the fast Fourier transform.[85,86] The 3D-RISM solution is then used to calculate the thermodynamic properties at each grid point, including the excess chemical potential (solvation free energy, labeled exchem), solvation energy (labeled solvene), and solute−solvent potential energy (labeled potUV).[82] The solvation energy (solvene) and solute−solvent potential energy (potUV) roughly correspond to the GIST total enthalpy (E_{sw} + E_{ww}) and the GIST solute-water enthalpy (E_{sw}), respectively.

3D-RISM models solvent molecules as being composed of interaction sites, e.g., water has oxygen and hydrogen sites; and 3D-RISM produces separate distribution grids for each solvent site. Because separate grids for oxygen and hydrogen solvent sites are produced, the grids must be combined to obtain molecular distribution grids like GIST. There are two approaches implemented in Amber: (1) combining the grids

https://doi.org/10.1021/acs.jcim.4c01623
J. Chem. Inf. Model. 2025, 65, 722−748

additively or (2) using the molecular reconstruction method described by Nguyen et al.[87] We generated grids using both methods, and we compare docking performance without receptor desolvation to docking with receptor desolvation calculated with additively combined 3D-RISM (see Section 3), or with molecular reconstruction 3D-RISM (see SI Section 6). Because additively combined 3D-RISM grids outperform molecular reconstruction 3D-RISM grids for docking using our scoring method, we focus only on additively combined 3D-RISM in the main text.

2.2.5. 3D-RISM Grid Generation. The 3D-RISM grid generation procedure closely follows the GIST procedure. Because 3D-RISM does not postprocess a simulation, it also does not require placing the protein in a water box, and 3D-RISM grids are generated without changing the frame.

We perform a minimization with *pmemd.cuda* on the receptor before running the 3D-RISM calculation using *rism3d.snglpnt* in Amber. The 3D-RISM equation is solved using the KH, PSE2, and PSE3 closures, in that order. The cTIP3P water model is used, and 1D-RISM results for cTIP3P water using PSE3 closure are provided through the *cTIP3P_pse3.xvv* input file. For systems that fail to converge, we adjust the MDIIS settings until the calculation is successful (SI Section 2 and Table S4). We tested the excess chemical potential, solvation energy, and solute−solvent potential energy grids summed over all solvent sites, labeled *exchem.1.dx*, *solvene.1.dx*, and *potUV.1.dx*, respectively, for our receptor desolvation energy calculations. These 3D-RISM grids are treated as a direct substitute for the total enthalpy GIST grids used for docking with the (full) displacement and blurry displacement methods.

The 3D-RISM grids encompass the entire protein, and we trim the box dimensions to include only the area around the binding pocket, matching the size of the GIST grid box. Note that the trimmed 3D-RISM grid box should completely contain the other grid boxes to not impact sampling coverage, like the GIST grids. After trimming the box dimensions, we apply a cap and generate precomputed grids, following the same procedure as for GIST. Because the magnitudes of the voxels for the 3D-RISM grids are different from GIST, we truncate the maximum values of the voxels in proportion to GIST by applying a cap of ±100 kcal/mol/Å^3 to the values of the voxels (i.e., any voxel that exceeds a magnitude of 100 kcal/mol/Å^3 is capped). Having tested the three different scoring methods with GIST, and having determined the behavior to be correlated (see SI Section 7), we only test receptor desolvation scoring with 3D-RISM by docking to the blurred grids with the blurry trilinear displacement method.

2.2.6. Using GIST and 3D-RISM Grids to Account for Desolvation in Docking. We calculate the receptor desolvation energy for ligand binding by summing up the energies of the voxels displaced by the ligand

$$E_{\text{water\_displacement}} = \sum_{v \in V_{\text{disp,l}}} E_v$$

The total displacement region ($V_{\text{disp,l}}$) is made up of the voxels proximal to the atoms of the ligand. Here, E_v has been multiplied by the volume of the voxel (0.125 Å^3), and is the energy in kcal/mol. If we assume no double counting, meaning no overlap between atoms, then

$$E_{\text{water\_displacement}} = \sum_{a \in \text{lig}} \sum_{v \in V_{\text{disp},a}} E_v$$

where a is an atom of the ligand (lig) and the atom-specific displacement region ($V_{\text{disp},a}$) is made up of the voxels proximal to that atom. However, some atoms in molecules will have overlapping displacement regions, so we must avoid double counting

$$E_{\text{water\_displacement}} = \sum_{i=1,N} \sum_{\substack{v \in V_{\text{disp},a(i)} \\ \wedge v \notin V_{\text{disp},a(j)}, \\ \text{for } j<i}} E_v$$

where N is the number of atoms in the ligand, and we only add the energy associated with a voxel that has not already been included. We described this method previously in Balius et al.[31] The bookkeeping associated with avoiding double counting is substantial, and partially results in a 6-times slower runtime in docking when evaluating displacement on-the-fly.[31] To obviate the double counting issue, we precompute receptor desolvation displacement, which allows us to evaluate receptor desolvation in the same manner as the other scoring components (electrostatics and van der Waals). Specifically, the closest points are looked up and trilinear interpolation is used to infer the energy at the atomic position.

To avoid the issue of double counting, we weight energy stored at voxels by their proximity to the center of the atom. We use the Gaussian equation for this purpose

$$p_a(\vec{x}) = \frac{1}{\sigma_a\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\|\vec{x} - \overrightarrow{\mu_a}\|}{\sigma_a}\right)^2\right)$$

where the mean, $\overrightarrow{\mu_a}$, is the atomic center of atom a, and the standard deviation, σ_a, is a function of the radius of atom a: $\sigma_a = r_a/D$. Modifying the divisor, D, changes the standard deviation, which determines the sharpness of the peak of the Gaussian. We refer to applying Gaussian weighting to the voxels as "blurring the grids". Figure 1 shows a visualization of how the divisor influences the sharpness of weight centered at the atomic positions. For a smaller divisor of 2, the weight is less sharp, and we cannot distinguish the shape of the aromatic rings of the molecule, while for a larger divisor of 3, the weight is sharper and clearly contours the molecule.

We write the water displacement as the sum of energies stored on all voxels, weighted by the Gaussian distribution. This Gaussian term will weight voxels closer to the atomic center more than atoms at the periphery

$$E_{\text{water\_displacement}} = \sum_{a \in \text{lig}} \sum_{v \in \text{grid}} p_a(v)^* E_v$$

Due to the Gaussian weighting, we no longer need to determine if a voxel is proximal to any atom, since the weight will approach zero for voxels further away from the ligand atoms. For voxels with overlapping regions, we see that the weight on that voxel is a sum of all weights from all atoms

$$E_{\text{water\_displacement}} = \sum_{v \in \text{grid}} \sum_{a \in \text{lig}} p_a(v)^* E_v = \sum_{v \in \text{grid}} E_v \sum_{a \in \text{lig}} p_a(v)$$

This allows us to visualize the weights (Figure 1).

The farther away a voxel, the less it contributes. Thus, for speed we only calculate the sum of the weighted energies over the displaced voxels, that is, the voxels proximal to a ligand atom

$$E_{\text{water\_displacement}} \approx \sum_{a \in \text{lig}} \sum_{v \in V_{\text{disp},a}} p_a(v)^* E_v$$

https://doi.org/10.1021/acs.jcim.4c01623
J. Chem. Inf. Model. 2025, 65, 722−748

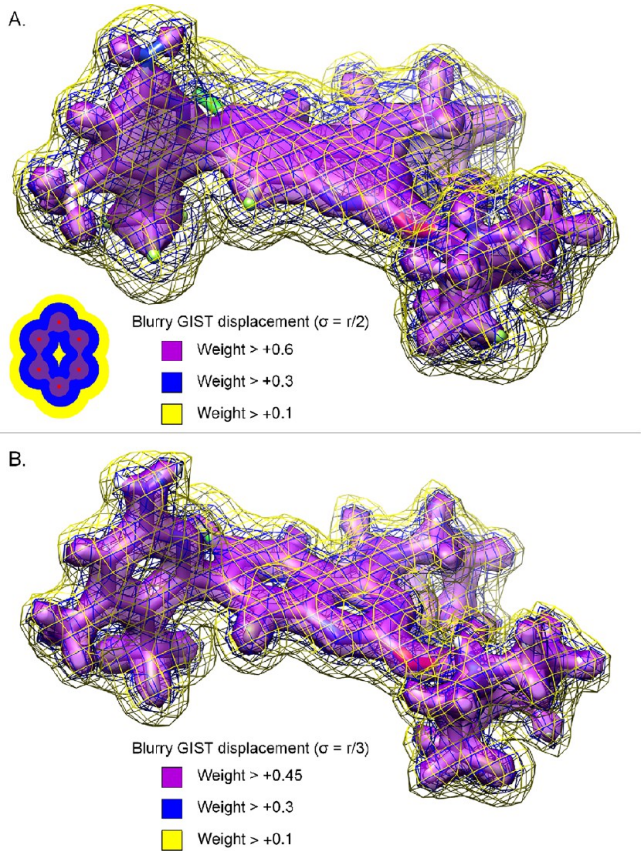


Figure 1. Using the Gaussian function centered at each atom to define the weights of grid voxels for the blurry-displacement version of the GIST scoring function. (A) The weights determined using a Gaussian function with a divisor of 2: $\sigma_a = r_a/2$. (B) The weights determined using a Gaussian function with a divisor of 3: $\sigma_a = r_a/3$. For both panels, three contour levels are shown: purple surface is weighted the highest (weight > (A) 0.6 or (B) 0.45), blue mesh the next highest (weight > 0.3), and finally, yellow mesh the lowest (weight > 0.1).

We precompute atomic displacement by placing a probe atom at each grid point. All voxels displaced by the atom are weighted by the Gaussian, and then summed to give the new precomputed value. We generate two precomputed (blurred) grids: one with a probe radius of 1.0 Å for hydrogens, and the other with the probe radius of 1.8 Å for heavy atoms. To ensure that the probe atom is contained within the original grid, we subtract a margin value in all directions. The dimensions of both blurred grids are reduced by 1.8 Å, so that these new grids have the same dimensions.

The precomputed grids are read into DOCK, and the trilinear interpolation procedure is used to calculate the ligand displacement. For each atom, the eight closest grid points are looked up and combined by linear interpolation to give the value at the atomic position. Table 1 shows the receptor desolvation parameters as implemented in DOCK 6.12.

Although the receptor desolvation scoring method was developed for GIST, the same procedures, i.e., blurring with the Gaussian and precomputing displacements, may be applied using 3D-RISM grids instead. When comparing GIST and 3D-RISM, we replace the processed GIST grids with 3D-RISM grids processed in the same way.

2.3. Covalent Docking with *Attach-and-Grow*. We have implemented a covalent docking method into DOCK 6, called *attach-and-grow*. The *attach-and-grow* algorithm aligns the modified covalent molecule to the covalent attachment point residue. The covalent bond environment is defined by four parameters that are sampled during docking: a dihedral angle (φ), a bond angle (θ), and two bond distances (d1 and d2). A cartoon shows these parameters (Figure 2). The dihedral angle is sampled starting at 0°, and steps to 360° with a user-specified step size. The dihedral angle sampling is analogous to the orients in *anchor-and-grow*. Values for the bond angle and two distances are user-specified. The different bond environments are analogous to different anchors (e.g., rings) in *anchor-and-grow*. Also analogous to *anchor-and-grow*, *attach-and-grow* minimizes torsions at every stage of growth using the *premin* option in DOCK 6. Including the verbose flag at runtime for DOCK 6 displays the progress of growth in the dock statistics, similarly to *anchor-and-grow*. The growth statistics may be useful for

Table 1. Receptor Desolvation Scoring Function Parameters in DOCK 6.12[a]

parameter	input used	description
gist_score_primary	yes	parameter to select GIST score as the scoring function.
gist_score_att_exp	6	attractive exponent. used for calculating the atomic radius.
gist_score_rep_exp	12	repulsive exponent. used for calculating the atomic radius.
gist_score_gist_scale	−1.0	scaling factor.
	parameters for full displacement receptor desolvation	
gist_score_gist_type	displace	type of GIST calculation; use displacement and avoid double counting (for rescoring).
gist_score_grid_file	gist-EswPlus2Eww_ref_cap.dx	
parameters for *blurry displacement* receptor desolvation		
gist_score_gist_type	blurry_displace	type of GIST calculation; use displacement and proximity weighting (for rescoring).
gist_score_grid_file	gist-EswPlus2Eww_ref_cap.dx	
gist_score_blurry_gist_div	2	sigma determines the sharpness of the peak of the Gaussian; equal to the radius divided by a divisor: $\sigma_a = r_a/D$. This parameter is only used for blurry displacement.
	parameters for blurry trilinear displacement receptor desolvation	
gist_score_gist_type	trilinear	type of GIST calculation; use trilinear (for docking).
gist_score_grid_file	gist-EswPlus2Eww_ref_cap_blurr_1pt8.dx	run calculation on precomputed displacement grid using blurry GIST. Generated using 1.8 Å probe radius for heavy atoms.
gist_score_Hydrogen_grid_file	gist-EswPlus2Eww_ref_cap_blurr_1pt0.dx	generated using 1.0 Å probe radius for hydrogens.

[a]GIST Score may be used as a stand-alone scoring function or combined with other scoring functions as part of Descriptor Score.

https://doi.org/10.1021/acs.jcim.4c01623
J. Chem. Inf. Model. 2025, 65, 722−748

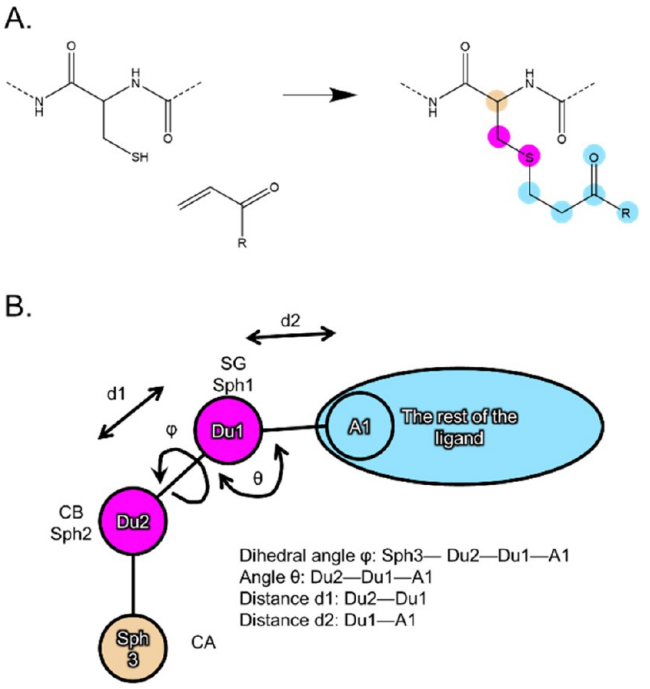


Figure 2. Covalent bond formation and environment. (A) The thiol−ene reaction to attach an acrylamide warhead to a cysteine residue is shown. (B) The covalent bond environment is defined by a bond angle (θ), and two distances (d1 and d2). The dihedral angle (φ) is sampled.

debugging if a molecule fails to produce a pose. This method has been discussed previously.[88] Table 2 shows the sampling parameters for the *attach-and-grow* algorithm as implemented in the DOCK 6.12 release.

Table 2. *Attach-and-grow* Sampling Parameters in DOCK 6.12

parameter	input used[a]	description[b]
covalent_bondlength	1.8, 1.7:0.1:1.91	bond length between Du2 and Du1. A value or a range[c] is given.
covalent_bondlength2	1.8, 1.7:0.1:1.91	bond length between Du1 and A1. A value or a range[c] is given.
covalent_angle	104.0, 103.0:1.0:105.0	angle between Du2, Du1, and A1. A value or a range[c] is given.
covalent_dihedral_step	10.0	step size for exploring the dihedral angle. We start at 0° and go to 360°. The dihedral is defined by Sph3, Du2, Du1, and A1.

[a]First input is a single value used for virtual screens; second input is a range of values used for test set pose reproduction. [b]See Figure 2B for reference. [c]A range is provided using the form: start:step:stop; the form: start:stop may also be used.

2.3.1. Covalent Database Generation. For the example virtual screens, we first prepare a database of molecules containing acrylamide warheads for docking. We modify the acrylamide warheads by attaching two silicon dummy atoms to the adduct in the place of the γ-sulfur and β-carbon. This is done using a Python script (*simple_reaction_file.py*, available on GitHub in the *teb_scripts_programs* repository) that takes as input a SMARTS string specifying a reaction and a file containing SMILES, and uses *RDKit*[89] to apply the chemical transformation.

Here is the SMARTS string that we use for acrylamides:

[SiH3:1][SiH2:2].[C:3] = [C:4][C:5](=[O:6])[N:7] ≫ [SiH3:1][SiH2:2][C:3][C:4][C:5](=[O:6])[N:7]

We then process the transformed molecules using a script now distributed with DOCK 6.12:

$DOCK6BASE/template_pipeline/hdb_lig_gen/generate/enumerate_stereoisomers.csh

This script enumerates all unspecified stereoisomers using ChemAxon's *cxcalc*[64] (https://www.chemaxon.com).

To generate *mol2* files for covalent docking, we used another script now distributed with DOCK 6.12:

$DOCK6BASE/template_pipeline/hdb_lig_gen/generate/build_ligand_simple_dock6_covalent_solv.csh

ChemAxon's *cxcalc* was used to generate protomers (i.e., the protonation states of the ligand) and isomers. *Corina* was used to convert the ligand chemical structure into a set of 3D atomic coordinates. *AMSOL 7.1* was used to calculate partial charges and desolvation parameters. Finally, a Python script (*mol_covalent_Si_to_Du_solv.py*, available on GitHub in the *teb_scripts_programs* repository) was called to finish preparing the *mol2* for covalent docking. This script converts the silicon atoms to dummy atoms, removes hydrogens that were connected to the silicon atoms, changes dummy atom charges to zero, and, to ensure the formal charge is an integer, distributes the sum of the removed partial charges to the remaining atoms equally. It also adds the per-atom solvation to the *mol2* file.

Because the *attach-and-grow* algorithm docks from a *mol2* file, and we do not need to conformationally expand our molecules to generate DB2 files, database generation is faster and requires less disk space in comparison to DOCK 3.7 (DOCKovalent).

2.4. Noncovalent Docking. We perform noncovalent docking in pose reproduction experiments and enrichment calculations with both noncovalent and covalent ligands. For noncovalent docking of covalent ligands, the covalent attachment point is not restrained, and the full degrees of freedom are sampled. We prepare the covalent ligands by removing an atom so that it is possible for the experimental pose to be reproduced without a clash between the ligand and the protein (see SI Table S1). This is done for ligands prepared from the experimental structure, and ligands prepared from SMILES. There are two methods that we use for noncovalent docking: *Anchor-and-grow* and Hierarchical Database Search.[71,90] *Anchor-and-grow* is used for pose reproduction experiments for more thorough sampling, and Hierarchical Database Search is used for enrichment calculations for speed.

2.4.1. Anchor-and-Grow. *Anchor-and-grow* is a breadth-first search method where ligand conformational degrees of freedom are explored on-the-fly.[91] The molecule is broken into rigid segments at rotatable bonds. Segments with the most atoms and attachment points become anchors, and, for each anchor, the remaining segments are arranged in layers about the anchor. The anchor is positioned in the binding site in multiple orientations using the orienting spheres. For each partially grown pose (e.g., anchor placement), the next segment is revealed, and dihedral angles are sampled and minimized. Conformations that exceed an energy threshold are pruned before proceeding to the next step of growth. The growth procedure continues until a completed conformation of the ligand is generated. We use the *anchor-and-grow* search method for noncovalent pose reproduction experiments.

2.4.2. Hierarchical DataBase (HDB) Search. HDB search is a depth-first search method where we search through precomputed ligand conformations stored in a hierarchical database, formatted as a *DB2* file.[71,90,92] The *DB2* file format stores ligand

https://doi.org/10.1021/acs.jcim.4c01623
J. Chem. Inf. Model. 2025, 65, 722−748

conformations as a tree, with each node representing a part or segment of the molecule in a specific conformation. A branch of the tree represents a conformation of the complete molecule. Each node appears only once, e.g., the root node (the rigid segment), which is shared among all branches, and thus the *DB2* file format is a highly compressed way of storing conformations. Like *anchor-and-grow*, the rigid segment of the molecule is placed in the binding site and scored. Then, for each branch (conformation of the molecule), each node (segment) of the branch is scored. The score is stored, and the node is flagged. If the score of the node exceeds a threshold, the branch is terminated early. If a node that has already been flagged is re-encountered, the score value is looked up. To make the algorithm efficient, the node is only oriented in the pocket before scoring. HDB search is 16-times faster than *anchor-and-grow*.[90] We use the HDB search method for enrichment calculations.

2.5. Pose Reproduction Outcomes. In a completed docking experiment, we classify the result into one of three outcomes: docking success, scoring failure, and sampling failure.[93] Docking success is defined when the top scoring pose reproduces the experimental pose. Scoring failure is defined when there exists a pose that reproduces the experimental pose, but it is not the top scoring pose. Sampling failure is defined when none of the ten poses generated reproduce the experimental pose. A docked pose is considered to reproduce the experimental pose if it is within 2 Å heavy atom root-mean-square deviation (RMSD). We use a symmetry-corrected RMSD (Hungarian algorithm[94]) implemented in DOCK 6. Docking experiments might not run to completion and no poses may be produced; we classified these as *did-not-dock*.

2.6. Cross-Docking. The structures were aligned to a single PDB (PDB ID 6GJ8) prior to being prepared for docking. Aligning all the structures to the same frame allows us to evaluate cross-docked poses using the aligned ligand experimental pose as a reference. Using 127 KRAS ligand−receptor pairs from the test set, we performed cross-docking of each ligand into each receptor, for a total of 16 129 docking calculations.

For cross-docking pose reproduction, RMSD values for a cross-docked pose of a ligand were calculated using the aligned experimental pose of the same ligand as the reference. Because the receptor conformation is different between systems and is kept rigid during docking, the experimental ligand pose may not be possible to reproduce for some cross-docking experiments. If the experimental pose clashes with the receptor, we classify the ligand−receptor pair as incompatible.[93] To determine which pairs are incompatible, we perform a constrained minimization of the ligand. If, after the constrained minimization, the Chemgrid Score is high (above 1000.0 kcal/mol), or if the molecule moves significantly (more than 2 Å RMSD), the pair is classified as incompatible (see SI Figure S3). Docking experiments with ligand−receptor pairs classified as incompatible usually result in a sampling failure or *did-not-dock* outcome, with a few exceptions (see SI Section 3 and Table S5). Note that an incompatible pair (i, j) is determined by comparing the experimental ligand pose (from system i) with the receptor (from system j) without any docking, while the sampling failure and *did-not-dock* outcomes are determined by the docking experiments.

2.7. Enrichments. DUDE-Z-like decoy backgrounds were generated for each of the 110 KRAS ligands for enrichment calculations on KRAS systems. The decoy generation procedure is adapted from the procedure for generating DUDE-Z decoys described in Stein et al.[58] The ligands' SMILES were protonated at pH 7.2 using ChemAxon's *cxcalc*. For each ligand protomer, the ZINC20 database[95] was searched for potential decoy molecules that matched the ligand's molecular weight, calculated water-octanol partition coefficient, number of rotatable bonds, number of hydrogen-bond donors, and number of hydrogen-bond acceptors. The potential decoy molecules were protonated, and molecules that did not match the formal charge of the ligand were discarded. Morgan fingerprints[96] (using a radius of 4) for the ligands and decoys were generated using *RDKit*, and Tanimoto coefficients (T_c) were used to determine similarity between molecules. Decoys too similar to any ligand ($T_c > 0.35$) were discarded, and the remaining decoys were clustered using a T_c threshold of 0.5. After clustering, the decoys were assigned to the ligands, with a goal of 50 property-matched decoys per ligand protomer. See SI Section 15 for discussion of parameter choice. The procedure to generate charge extrema decoys is the same, except, instead of matching the formal charge of the ligand, decoys were found for each of the five formal charges between −2 and +2. A goal of 50 extrema decoys per charge per ligand protomer was set, for a total of 250 extrema decoys per ligand. Charge extrema decoys were used to check if the docking scoring function was overoptimizing for charge interactions.

Enrichment benchmark databases are generated in the *DB2* file format. ChemAxon's *cxcalc* was used to protonate the SMILES at pH 7.2. *Corina* was used to generate 3D conformations of the molecules from the 2D SMILES representations. *RDKit*, *AMSOL 7.1*, and DOCK 6 were used to generate multiple conformations of each molecule. Conformations were expanded about a rigid segment of the molecule; any rigid segment with more than 4 heavy atoms was used. The multiconformation *mol2*, partial charges, and per-atom desolvation (calculated from the starting conformation) are converted to the *DB2* format using the *mol2db2.py* script distributed with DOCK 6.12 (in the *$DOCK6BASE/template_-pipeline/hdb_lig_gen/mol2db2/* directory). All molecules are generated as noncovalent databases. For covalent ligands, one atom was removed from the SMILES prior to database generation to ensure that the molecule could reproduce the experimental pose without clashing. To evaluate the docking performance on each pocket, the enrichment benchmark sets were partitioned based on the binding pocket of the ligands. Because a significant proportion of switch II pocket-binding ligands were covalent, the switch II pocket enrichment set was further partitioned based on whether the ligand binds covalently or noncovalently. This resulted in three enrichment sets: switch I/II pocket, switch II pocket noncovalent, and switch II pocket covalent.

To quantify enrichment, we use the area under the curve (AUC) of the receiver operator characteristic (ROC) curve, which plots the true positive rate (% of ligands found) versus the false positive rate (% of decoys found). To produce the curve, first, a sorted list of the best scores for each unique ligand and decoy is obtained. We iterate through the list, stepping up when a ligand is encountered and stepping right when a decoy is encountered. Perfect enrichment of the ligands from the decoys results in an AUC of 100%. Random enrichment where the ligands are uniformly distributed among the decoys results in an AUC of 50%. Any AUC value above 50% is better than random, and any value below 50% is subrandom. To weight early enrichment more than later enrichment, we calculate at area under the log adjusted ROC curve, where we take the logarithm of the false positive rate, and subtract the AUC of the random log

https://doi.org/10.1021/acs.jcim.4c01623
J. Chem. Inf. Model. 2025, 65, 722−748

adjusted ROC curve to obtain the logAUC.[70] As a result, the logAUC will be positive for better than random enrichment and negative for subrandom enrichment. Because the logarithm function is undefined at 0, we define a lower bound of 0.001. That is, the logAUC is calculated on the range from 0.1 to 100% for the % of decoys found.

2.7.1. Error in Enrichment Calculations Using Bootstrapping. To quantify error in our enrichment calculations, we use bootstrapping[97] to resample. We perform random sampling with replacement on the list of scored molecules to obtain a resampled list of the same length as the original. The ligands and decoys are sampled separately, so that the number of ligands and decoys remains consistent. We perform 1000 bootstrap runs and calculate enrichment metrics for each, generating a distribution of logAUC values. To compare enrichment between different docking setups, we perform paired bootstrapping. We dock the same ligands and decoys using two different docking setups and perform the same resampling of the docked molecules for both enrichment runs. This results in 1000 paired logAUC values, from which we calculate the difference of each pair.

We calculate *p*-values using two *t*-tests from SciPy:[98] two-sample *t*-test (using *stats.ttest_ind*) and one-sample *t*-test (using *stats.ttest_1samp*). The two-sample *t*-test compares two distributions, testing the null hypothesis that the two distributions have the same mean. The one-sample *t*-test tests the null hypothesis that the mean of the distribution of logAUC differences is equal to a given population mean, set to zero. In both cases, we are testing whether the mean logAUC difference between the two bootstrapped distributions is significant. We reject the null hypothesis with a level of significance of 0.01 (p-value $\leq$ 0.01) and classify the enrichment calculation as significant or not significant (no significant difference) accordingly. Rejecting the null hypothesis indicates that the change in the docking setup (e.g., incorporating receptor desolvation) had a significant impact on the enrichment results. We classify changes to the docking setup as better or worse based on the results of the two-sample *t*-test and the change in logAUC. For example, when comparing enrichment with and without 3D-RISM receptor desolvation for one system, if the *p*-value is less than 0.01, and the logAUC with 3D-RISM is greater than the logAUC without 3D-RISM, or vice versa, we classify the system as better or worse, respectively.

2.8. Error in Pose Reproduction Experiments. To quantify error in pose reproduction experiments, we ran the docking with different random seeds for the simplex minimization procedure.[91] Because we use the simplex minimization from the start of the sampling procedure for both *anchor-and-grow* and *attach-and-grow*, modifying the random seed of the simplex will perturb the sampling from the start, ensuring sampling divergence.

To compare pose reproduction between different docking setups, we ran the pose reproduction experiments with each docking setup 100 times with different random seeds. For each of the 100 random seed runs (or replicas), we classify the completed docking calculation for each system into one of the three docking outcomes: docking success, scoring failure, and sampling failure. Docking calculations that do not run to completion were classified as *did-not-dock*. This results in 100 replicas from which we determine the number of systems with docking success, scoring failure, sampling failure, and *did-not-dock* outcomes. For each outcome, we calculate the mean and standard deviation. We calculate *p*-values using the same *t*-tests from SciPy used for the enrichment error calculations. The *p*-values are used to compare the mean number of each docking outcome between the two docking setups. A smaller *p*-value indicates a greater likelihood that a change in the mean number of each docking outcome is significant, and results from a change in the docking setup (e.g., incorporating 3D-RISM receptor desolvation or covalent docking). For example, if the number of docking success outcomes is greater with covalent docking than with noncovalent docking, and the *p*-value is small ($\leq$0.01), then covalent docking statistically significantly increases docking success rate for that set of systems.

3. RESULTS

To evaluate docking performance on the RAS test set, we performed pose reproduction experiments, for both cognate docking and cross-docking cases, and enrichment calculations.

3.1. Description of the Test Set. We assembled a RAS test set consisting of 138 RAS protein structures (125 KRAS4B, 11 HRAS, and 2 NRAS), and 2 structures of KRAS DNA G-quadruplexes in complex with ligands. To construct this test set, we searched the PDB for structures of RAS proteins in complex with small molecule inhibitors. We filtered and discarded structures containing GTP-competitive inhibitors. From the remaining structures, we extracted the PDB ligand chemical IDs, resulting in 110, 5, and 1 unique ligand(s) for the 125 KRAS, 11 HRAS, and 2 NRAS protein structures, respectively. See SI Table S1 for a full list of ligands and corresponding PDB IDs. This test set is prepared for docking as described in Section 2. Shown in Table 3 is a breakdown of the number of systems and unique ligands for KRAS.

Table 3. Breakdown of the KRAS Systems and Ligands

	# systems	# unique ligands
switch I/II pocket	40	32
switch II pocket	83	74
switch II pocket, noncovalent	24	18
switch II pocket, covalent	59	56
miscellaneous	4	4
total	127[a]	110

[a]There are 125 PDB structures; two structures have two ligands for a total of 127 ligand−receptor pairs.

The majority of the ligands bind to one of two main pockets: the switch I/II pocket (SI−SII), and the switch II pocket (SII). Two KRAS structures (PDB IDs 7U8H and 8AFD) contain ligands in both the SI−SII and SII pockets, resulting in 127 ligand−receptor pairs from the 125 KRAS protein structures. We classify ligands that do not bind to either the SI−SII or SII pockets as miscellaneous binders. The test set contains ligands that bind noncovalently and ligands that bind covalently. We prepared covalent ligands for standard noncovalent docking, and covalent docking with *attach-and-grow* (see Section 2).

The locations of the two main pockets, and ligands that bind to these sites, are seen through an overlay of 48 representative KRAS structures (Figure 3). Some miscellaneous binders are shown as well. The bound nucleotides are shown and may be GDP, or one of three GTP analogs: GPP[NH]P (GNP), GPP[CH2]P (GCP), or GTPγS (GSP). The fluctuations of the Switch I and Switch II regions are also seen in the overlay of structures in Figure 3.

3.2. Pose Reproduction. Using the 127 KRAS ligand−receptor pairs, we docked all ligands to all receptors and

https://doi.org/10.1021/acs.jcim.4c01623
J. Chem. Inf. Model. 2025, 65, 722−748

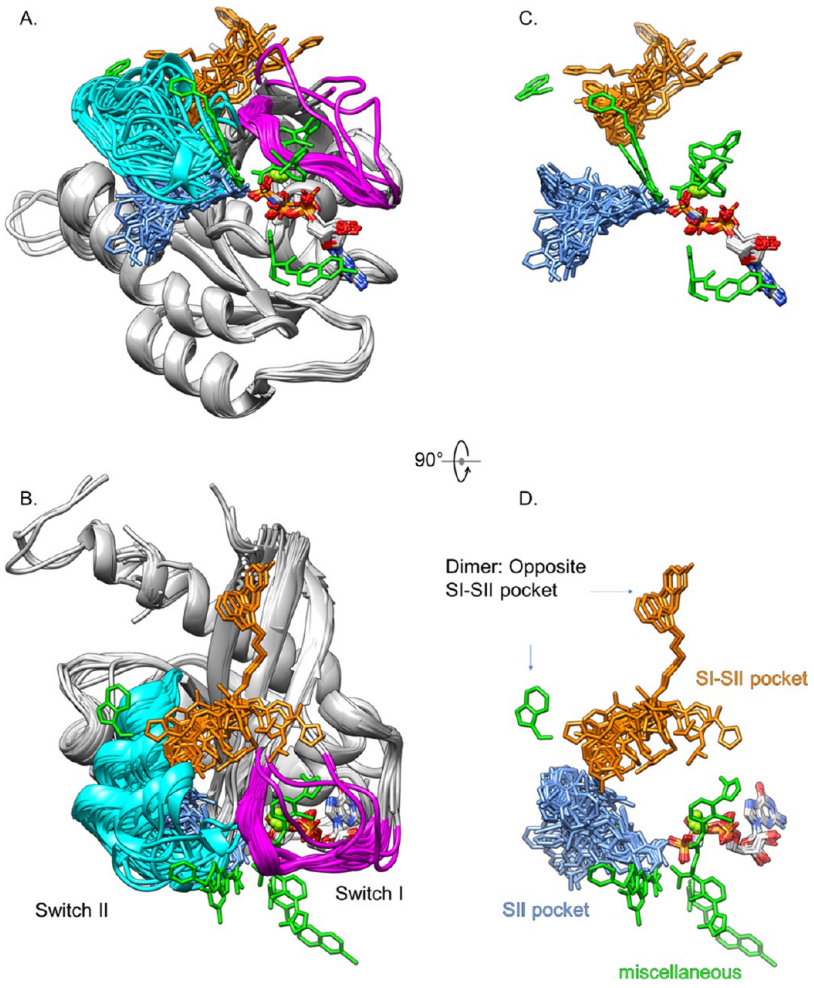


Figure 3. Representative KRAS structures (N = 48) from the test set. Switch I (residues 30−38, magenta) and Switch II (residues 60−76, cyan) regions are highlighted. Ligands that occupy the SI−SII pocket are shown in orange. Ligands that occupy the SII pocket are shown in blue. Ligands that do not bind to either of these two pockets are classified as miscellaneous and shown in green. (A, B) Overlays of 48 representative structurally distant KRAS protein−ligand complexes. (C, D) Ligands are shown without the protein. (B) is rotated 90° of (A), and, likewise, (D) is rotated 90° of (C). Cluster representatives are reported in SI Table S1.1.

generated cross-docking matrices using the results. Docking calculations were performed using the noncovalent *anchor-and-grow* sampling method with Chemgrid Score. We organize the cross-docking results by ligand binding pocket, i.e., SI−SII pocket systems (Figure 4A) and SII pocket systems (Figure 4B); or view the results for all systems (SI Figure S4A). For each completed docking experiment (one entry of the matrix), we define three possible outcomes: docking success, scoring failure, and sampling failure shown in blue, yellow and red, respectively (Figure 4), defined above in Section 2. Sometimes a docking calculation does not run to completion and no pose is generated; these we classify as a *did-not-dock* outcome and are shown in black (Figure 4). There are also often cases of incompatibility between a ligand and a receptor (see SI Figure S4). RAS is a flexible protein that adopts many distinct conformations,[7,99] and these ligand−receptor incompatibilities are likely due to conformational differences in RAS, particularly within the SII pocket.[6] In the SI−SII pocket systems we observe a swath of 8 systems where cross-docking experiments with the associated ligands results in sampling failures for all receptors (Figure 4A), see the Section 4 for more about these 8 systems. A fractional breakdown of the pose reproduction outcomes for the cross-docking matrix diagonal (cognate docking), matrix off-diagonal (cross-docking), and full matrix is shown in Figure 4C. The cognate docking outcomes (Table 4) and cross-docking outcomes (Table 5) are also reported below. We use ligand conformations generated using the experimental conformation as a starting point, for results using ligand conformations generated from SMILES see SI Figure S4B,C.

Tables 4 and 5 show rates of each docking outcome for cognate docking (diagonal) and cross-docking (off-diagonal). As expected, we see that cognate docking has a higher docking success rate than cross-docking. For the SI−SII pocket systems, we see a 25.9% higher docking success rate for cognate docking compared to cross-docking. We see a similar trend for the SII pocket systems, 62.8% higher, and for all systems, 59.4% higher. The SII pocket has a higher cognate docking success rate compared to the SI−SII pocket, 81.9% compared to 52.5%, respectively, but a lower rate for cross-docking, 19.1% compared to 26.6%, respectively.

The SI−SII pocket has a lower sampling failure rate than the SII pocket, 62.3% compared to 73.3%, respectively. This corresponds to the incompatibility rate: 20.2% for the SI−SII pocket compared to 32.2% for the SII pocket. For cross-docking, the docking success rate for the entire set falls to 10.7%, lower than either the SI−SII or SII pocket alone. This decline in

https://doi.org/10.1021/acs.jcim.4c01623
J. Chem. Inf. Model. 2025, 65, 722−748

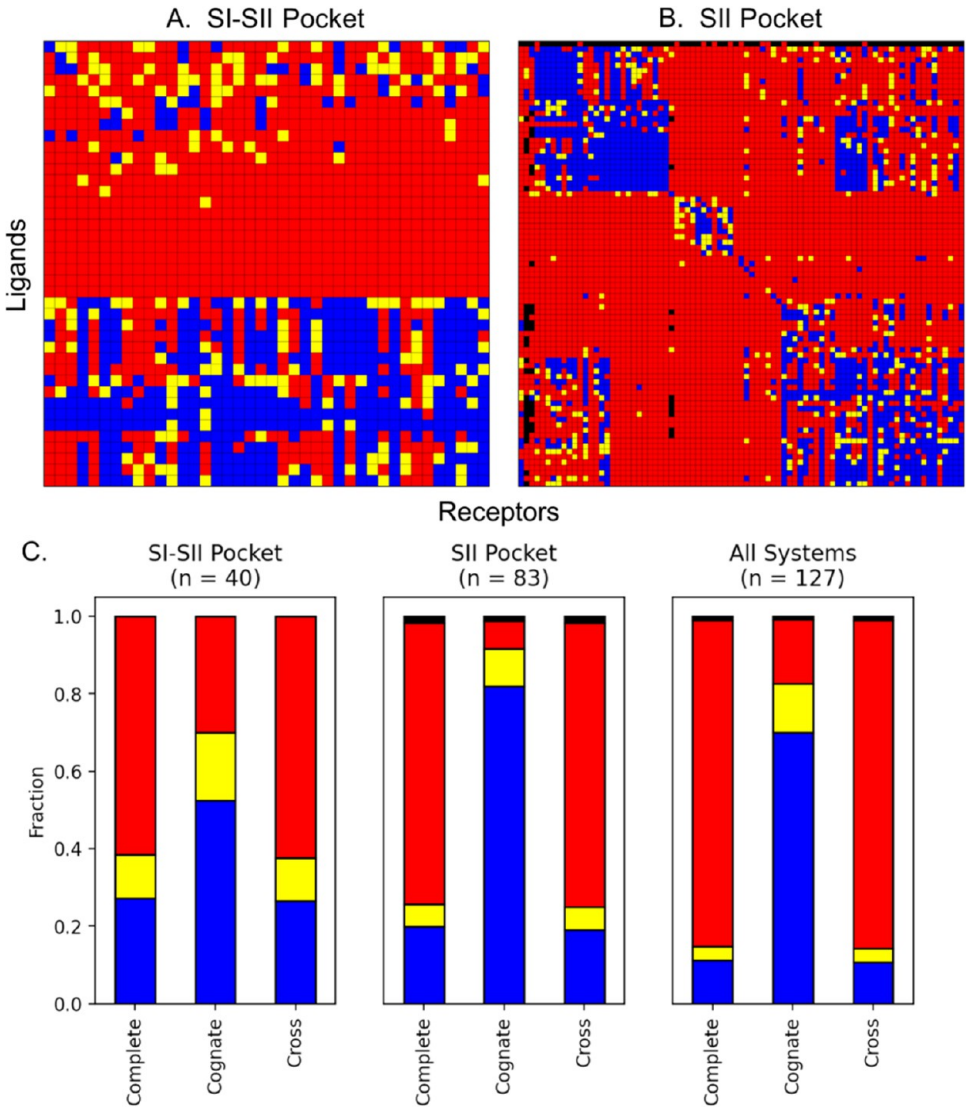


Figure 4. Cross-docking and cognate docking results using Chemgrid Score. Blue, yellow, red, and black indicate docking success, scoring failure, sampling failure, and *did-not-dock* outcomes, respectively. Cross-docking matrix showing docking outcomes for (A). 40 SI−SII pocket systems and (B). 83 SII pocket systems. Arrangement of systems determined using hierarchical clustering by similarity between ligand outcomes. (C) Fractional breakdown of docking outcomes for the complete matrix, cognate docking (diagonal), and cross-docking (off-diagonal).

Table 4. Cognate Docking (Diagonal) Outcomes Using Chemgrid Score

	docking success	scoring failure	sampling failure	did not dock
SI−SII pocket[a]	21 (52.5%)	7 (17.5%)	12 (30.0%)	0 (0.0%)
SII pocket[b]	68 (81.9%)	8 (9.6%)	6 (7.2%)	1 (1.2%)
all systems[c]	89 (70.1%)	16 (12.6%)	21 (16.5%)	1 (0.8%)

[a]Number of cognate docking (diagonal) experiments: 40. [b]83. [c]127.

docking success rate corresponds with an increase in sampling failure rate, which is higher for the full set, 84.7%, than for either pocket alone, 62.3 or 73.3%. Unsurprisingly, the incompatibility rate is also higher for the full set, 61.1%, than for either pocket, 20.2 or 32.2%. (See SI Section 3 and Table S5 for systems classified as incompatible, but where cross-docking reproduces the experimental pose).

For cognate docking, we generally see that the rates of each docking outcome for all systems falls between the corresponding rates for the two pockets (Table 4). For instance, the scoring failure rate for all systems, 12.6%, falls between the scoring

Table 5. Cross-Docking (Off-Diagonal) Outcomes Using Chemgrid Score

	docking success	scoring failure	sampling failure	did not dock	incompatible
SI−SII pocket[a]	415 (26.6%)	173 (11.1%)	972 (62.3%)	0 (0.0%)	315 (20.2%)
SII pocket[b]	1301 (19.1%)	396 (5.8%)	4988 (73.3%)	121 (1.8%)	2190 (32.2%)
All systems[c]	1716 (10.7%)	569 (3.6%)	13554 (84.7%)	163 (1.0%)	9782 (61.1%)

[a]Number of cross-docking (off-diagonal) experiments: 1560. [b]6806. [c]16 002.

https://doi.org/10.1021/acs.jcim.4c01623
J. Chem. Inf. Model. 2025, 65, 722−748

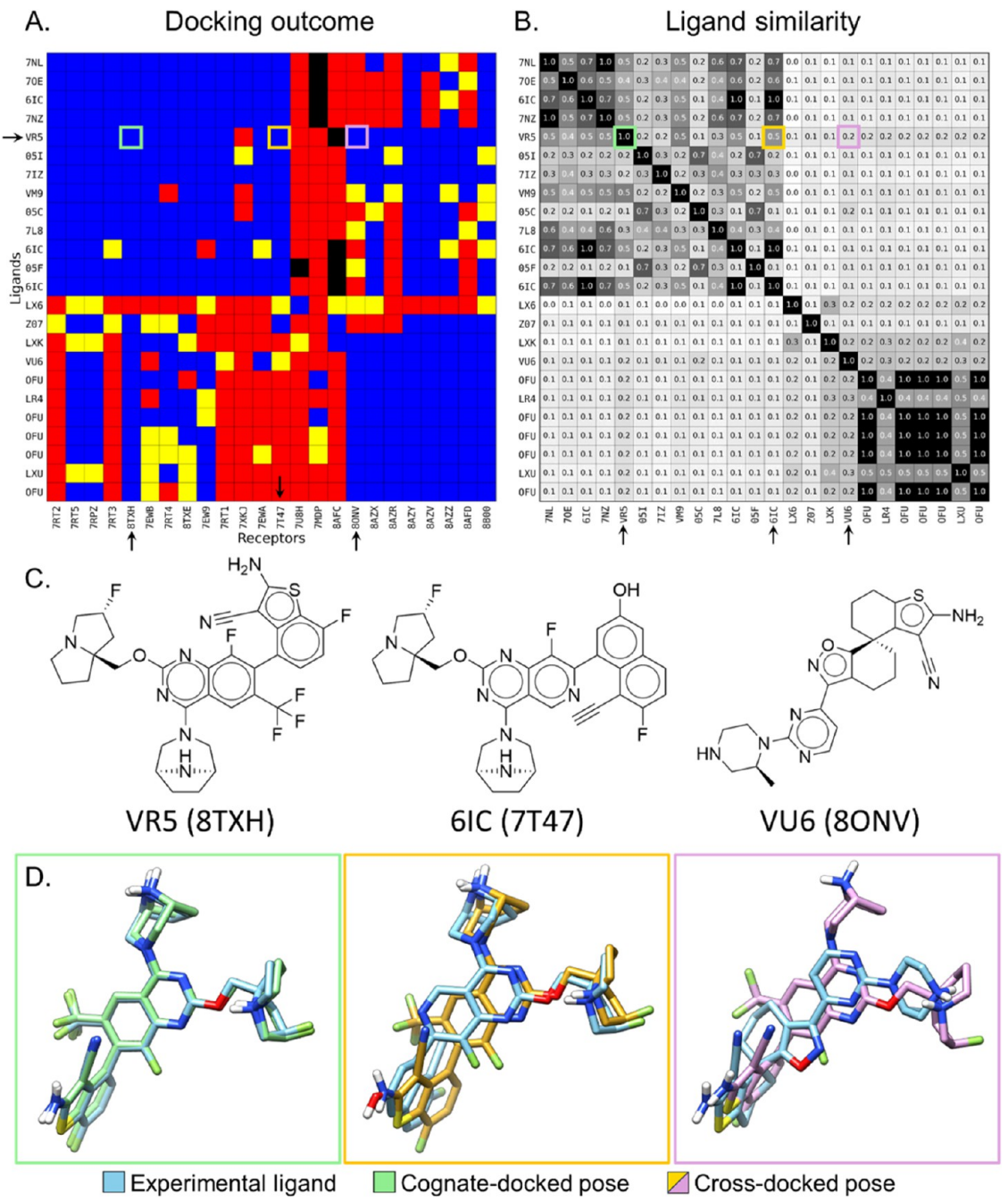


Figure 5. Pose reproduction for SII pocket noncovalent systems. (A) Docking outcomes for 24 SII pocket systems containing noncovalent ligands. Blue, yellow, red, and black indicate docking success, scoring failure, sampling failure, and did-not-dock outcomes, respectively. (B) Ligand chemical similarity determined using Tanimoto coefficients (Tc) between Morgan fingerprints generated using *RDKit*, where a higher Tc indicates greater similarity. Black indicates the identical or high similarity, gray indicates some similarity, and white indicates poor similarity. (C) Chemical structures of three selected ligands (PDB ligand chemical IDs VR5, 6IC, and VU6). (D) Cognate docking, and two cross-docking poses for the ligand VR5 to receptors 8TXH (green), 7T47 (orange), and 8ONV (purple). The receptor is hidden for clarity.

failure rates for the two pockets, 17.5 and 9.6%. Since most systems contain ligands that bind to the SI−SII or SII pockets (i.e., there are few miscellaneous systems), cognate docking for all systems only differs from the union of results for the two individual pockets by a few systems. However, this is not the case for cross-docking (Table 5). This is due to the additional cases added to cross-docking across all systems, namely, when ligands that bind to the SI−SII pocket are docked to SII pocket receptors, or vice versa. For instance, the cross-docking scoring failure rate for all systems, 3.6%, is lower than the rates for either of the two pockets, 11.1 or 5.8%.

Focusing on the subset of SII pocket systems containing noncovalent ligands, we obtain a much higher docking success rate (Figure 5A). There are 24 systems in this subset, with some systems sharing the same ligand (Table 3). We generated a ligand chemical similarity matrix (Figure 5B) to compare to the cross-docking matrix (Figure 5A). Regions of high ligand similarity seem to correspond to regions of high docking success rate. We select three systems from the matrix based on chemical

https://doi.org/10.1021/acs.jcim.4c01623
J. Chem. Inf. Model. 2025, 65, 722−748

similarity of the three ligands (Figure 5B,C). The ligands 6IC and VU6 serve as representatives of docking success and ligand similarity from the two clusters. The ligand VR5 was selected for having high docking success rate for both clusters. The ligands 6IC and VR5 have high chemical similarity, and cross-docking VR5 to receptor 7T47 (associated with 6IC) results in a docking success (Figure 5D, outlined in orange). The ligands VU6 and VR5 have a lower Tc, indicating less similarity. However, both molecules contain benzothiophene moieties that orient deep in the pocket. Cross-docking VR5 to receptor 8ONV (associated with VU6) also results in a docking success (Figure 5D, outlined in purple). Chemical similarity between ligands seems to result in increased compatibility between systems, due to similar receptor conformations. Similar ligands may also be easier to orient to the matching spheres, which are generated using ligand atoms. As a result of sharing chemical similarity with ligands from both clusters, the ligand VR5 has one of the highest rates of docking success (83.3%) across the 24 noncovalent, SII pocket systems.

3.3. Receptor Desolvation. We used systems from the test set to test the newly implemented scoring functions that account for receptor desolvation. Docking calculations here in Section 3.3 are performed using the noncovalent sampling methods: hierarchical database search for enrichment calculations, and *anchor-and-grow* for pose reproduction.

3.3.1. Enrichment Calculations. Using the decoy generation procedure described in Section 2, we generated DUDE-Z-like benchmark sets for enrichment quantification. These sets consist of the SI−SII and SII pocket-binding ligands from the test set, and corresponding property-matched and charge extrema decoys sets. The ligands were first partitioned by ligand-binding pocket, resulting in 32 SI−SII pocket ligands and 74 SII pocket ligands. A large portion of ligands that bind to the SII pocket are covalent binders, which we modify by removing an atom from the covalent warhead to prevent clashes during noncovalent docking. The 74 SII pocket ligands were further partitioned based on whether the ligands bind covalently or noncovalently, resulting in 18 noncovalent ligands and 56 covalent ligands. The result is three sets of ligands: 32 SI−SII pocket ligands, 18 noncovalent SII pocket ligands, and 56 covalent SII pocket ligands. Each set of ligands has two corresponding sets of decoys: property-matched decoys and charge extrema decoys (Table 6).

We performed benchmark enrichment calculations on 123 KRAS systems containing ligands that bind to either the SI−SII or SII pockets with Chemgrid Score, omitting 4 miscellaneous systems containing ligands that do not bind to either the SI−SII or SII pockets from enrichment calculations. Due to the computational cost of running GIST using 100 ns simulations, we first selected the 20 systems with the best enrichment to run GIST on: 10 SI−SII pocket systems, and 10 SII pocket systems. Receptors with missing residues were omitted from our selection. For more discussion on enrichment results for all KRAS systems, see Section 3.3.2 and SI Section 5. We further run GIST, using 10 ns simulations to reduce computational cost, and 3D-RISM on all KRAS systems in the test set.

Table 6. Number of Ligands, Property-Matched Decoys, and Charge Extrema Decoys for the DUDE-Z-like Enrichment Benchmark Sets[a]

	# ligands	# property-matched	# extrema
SI−SII pocket	32	3100 (2922)	7723 (7239)
SII-pocket-noncovalent	18	1147 (1107)	5431 (5032)
SII-pocket-covalent	56	1800 (1723)	14 504 (13 930)

[a]Outside the parentheses is the number of decoys matched, inside the parentheses is the number of successfully built molecules. SMILES that fail to build may fail to run through one of the programs used to prepare molecules in the DB2 file format, e.g., ChemAxon, *Corina*, *RDKit*, *Amsol 7.1*, or DOCK 6.

3.3.2. Comparison of GIST and 3D-RISM Scoring. We tested incorporating receptor desolvation scoring by comparing docking performance with and without the receptor desolvation term, using enrichment calculations. Receptor desolvation grids were generated using either GIST or 3D-RISM, and they were incorporated into the scoring function for docking using the trilinear interpolation method (i.e., the *gist_score_gist_type* parameter was set to trilinear). Thus, we compare three scoring functions: Chemgrid Score (baseline, no receptor desolvation), Chemgrid Score with GIST, and Chemgrid Score with 3D-RISM.

For the GIST calculations, the protein is translated and rotated into a water box for the MD simulation; this changes the frame, which impacts the docking. To accurately compare all three scoring functions, we use receptors (and therefore grids) aligned to the new GIST MD simulation frame. We regenerate docking grids for the aligned receptors. We also prepare three blurred 3D-RISM grids for docking: the excess chemical potential (*exchem*), total solvation energy (*solvene*), and solute−solvent potential energy (*potUV*) grids. We visualize and compare the GIST and *potUV* 3D-RISM grids (Figure 6). Using these grids, we quantify water displacement energies: waters displaced from red regions are unfavorable, resulting in an energy boost, while waters displaced from blue regions are favorable, resulting in an energy penalty. There are visual similarities between GIST grids (Figure 6A) and 3D-RISM grids (Figure 6B), but also substantial differences. To use the trilinear scoring method, we precompute displacement using a Gaussian weight, as discussed in Section 2. In Figure 6C−F, precomputed displacement grids using two different probe radii are shown. We refer to the process of precomputing displacement using a Gaussian weight as blurring, and the resulting grids as blurred. GIST and 3D-RISM grids blurred with a probe radius of 1.8 Å (Figure 6E,F) are the most visually similar. The magnitude of the voxels also differs between GIST and 3D-RISM, and we adjust the thresholds used for visual comparison of the grids: ±0.25 kcal/mol/Å^3 for GIST, and ±20 kcal/mol/Å^3 for 3D-RISM. This is also the case for the blurred grids, and we use thresholds of ±2.5 kcal/mol for blurred GIST and ±50 kcal/mol for blurred 3D-RISM.

To compare the trilinear, blurry displacement, and (full) displacement scoring methods, we rescored enrichment poses and compared the receptor desolvation energies using correlation plots. For further discussion of the correlation between the scoring methods, and between GIST and 3D-RISM, see SI Section 7. The magnitude of the GIST and 3D-RISM receptor desolvation energies are also different. We apply different weighting factors to GIST and 3D-RISM to balance the receptor desolvation components with the other scoring function terms. See SI Section 8 for exploration of different weighting factors.

As mentioned above, docking using GIST requires changing the frame to orient the protein into a water box for the MD simulation. Changing the frame perturbs the docking calculation. To accurately compare GIST and 3D-RISM, we generate

https://doi.org/10.1021/acs.jcim.4c01623
J. Chem. Inf. Model. 2025, 65, 722−748

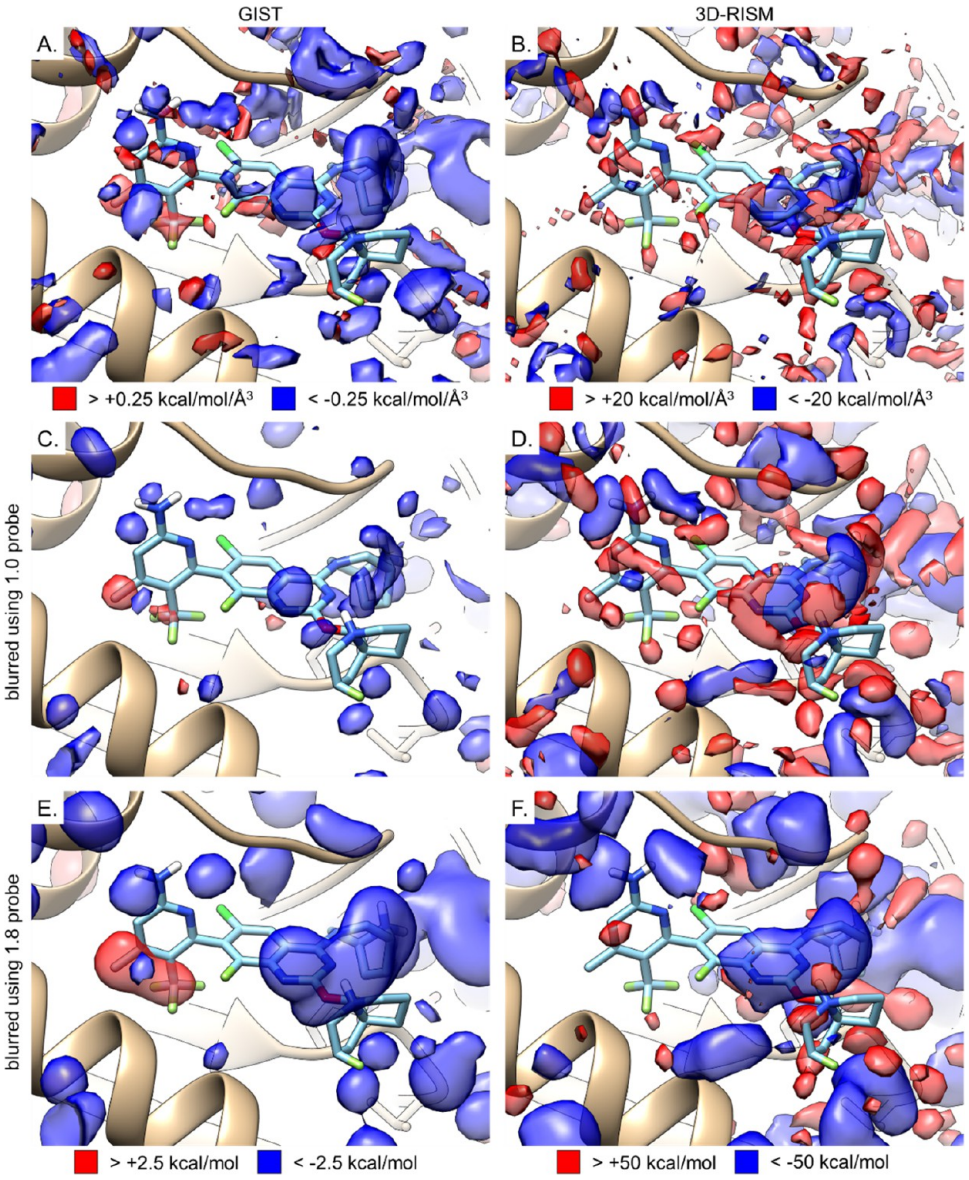


Figure 6. Comparison of solvation energy density distributions with GIST and 3D-RISM grids. Representative SII pocket system shown (PDB ID 8TXE, PDB ligand chemical ID VM9). (A, B) Standard grids. (C, D) Grids blurred using a probe with radius of 1.0. (E, F) Grids blurred using a probe with radius of 1.8. (A, C, E) Grids produced using GIST. (B, D, F) Grids produced using 3D-RISM solute−solvent potential energy. Red surfaces indicate unfavorable water regions (favorable if displaced), and blue surfaces indicate favorable water regions (penalty if displaced).

3D-RISM grids in the new frame, even though 3D-RISM does not require changing the frame to run. Shown in Table S6 and Figure 7A,B is a comparison of enrichment performance between GIST and 3D-RISM in the new frame. Enrichment performance for 3D-RISM in the original test set frame is also shown in Table S6 and Figure 7B. In Figure 7A,B and Table S6, the no receptor desolvation, GIST, and 3D-RISM labels refer to using only Chemgrid Score, Chemgrid Score with GIST, and Chemgrid Score with 3D-RISM, respectively (see Section 2). Thus, we are comparing enrichment performance using Chemgrid Score with and without GIST receptor desolvation, and Chemgrid Score with and without 3D-RISM receptor desolvation.

Incorporating GIST receptor desolvation results in positive ΔlogAUC for all systems except for one, 8AZR, where it is slightly negative (ΔlogAUC = −0.1). Two-sampled *t*-tests of the bootstrapped logAUC distributions (see Section 2) indicate that including GIST results in better logAUC values for 16 out of 20 systems, worse for 0 systems, and no significant difference for 4 systems, compared to enrichment without receptor desolvation. Mean ΔlogAUC is positive with GIST receptor desolvation for both binding pockets, +1.40 for SI−SII pocket systems, and +2.60 for SII pocket systems, indicating GIST improves enrichment performance on average.

However, the impact of incorporating 3D-RISM receptor desolvation is not as clear (Figure 7). We calculated enrichment using the blurred exchem, solvene, and potUV 3D-RISM grids, all of which perform similarly. Across the 20 selected KRAS systems, incorporating receptor desolvation using GIST performs better than with any of the three 3D-RISM grids. Docking with GIST does not result in significantly worse logAUC for any of the 20 systems, and results in improved enrichment on average (positive mean ΔlogAUC) for both pockets. In contrast, docking with 3D-RISM sometimes results

https://doi.org/10.1021/acs.jcim.4c01623
J. Chem. Inf. Model. 2025, 65, 722−748

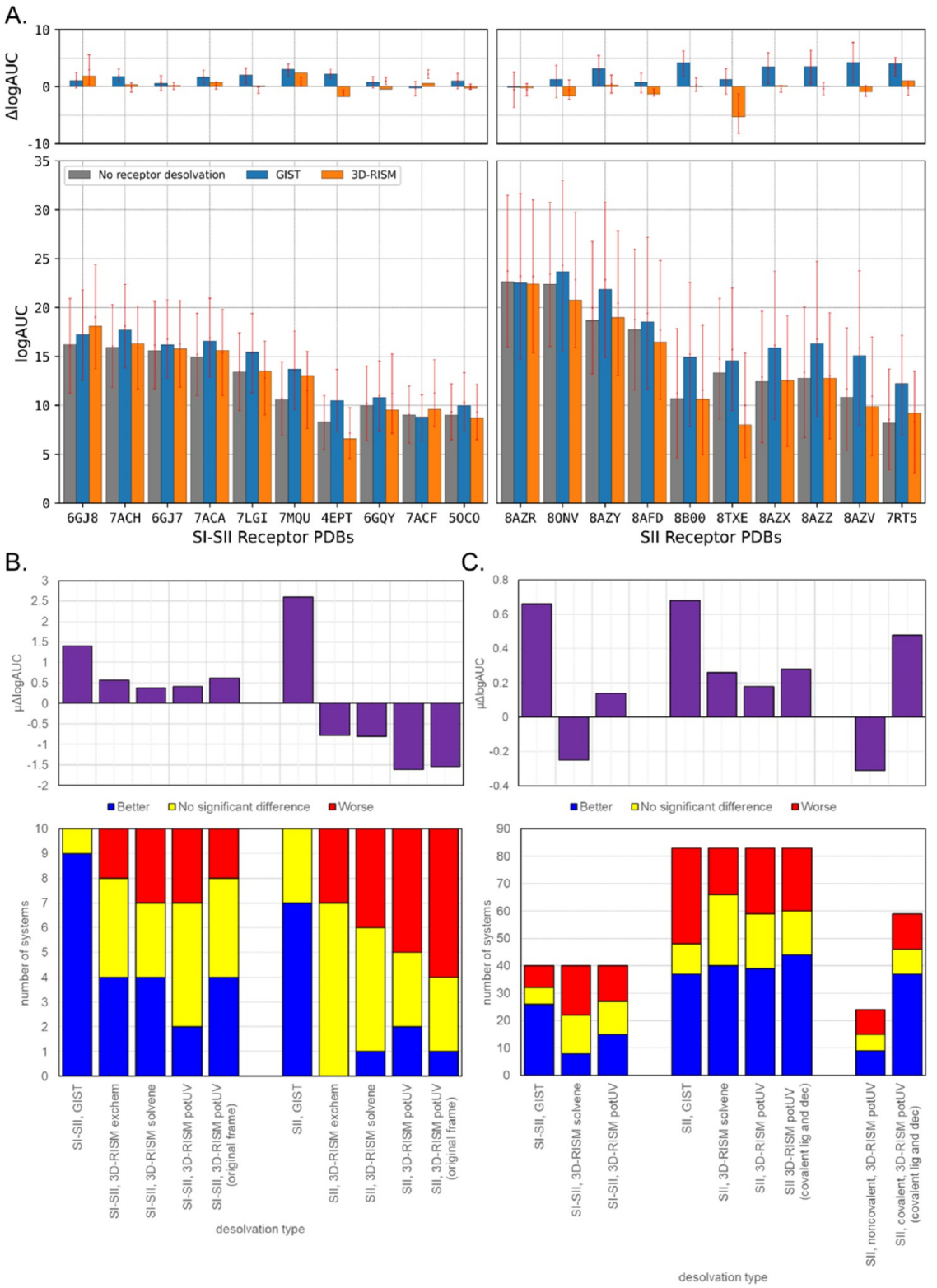


Figure 7. Comparison of enrichment with and without receptor desolvation. (A) Enrichment with and without receptor desolvation over 10 SI−SII systems and 10 SII systems. Bottom panels show logAUC values for no receptor desolvation (gray), GIST (blue) or 3D-RISM (orange). Top panels show differences in logAUC values (ΔlogAUC) between GIST (blue) or 3D-RISM (orange) and no receptor desolvation. 3D-RISM receptor desolvation calculated using blurred total solvation energy (solvene) grids. (B) Comparison of enrichment with and without receptor desolvation over 10 SI−SII systems and 10 SII systems. (C) Comparison of enrichment with and without receptor desolvation over 40 SI−SII systems and 83 SII systems. (B, C) Bottom panels show breakdown of systems by whether receptor desolvation results in better (blue), no significant difference (yellow), or worse (red) enrichment. Top panels show the average change logAUC.

in significantly worse enrichment, especially for the SII pocket systems. Regardless of which 3D-RISM grid is used, enrichment with 3D-RISM improves on average for the SI−SII pocket, but it worsens on average for the SII pocket (negative mean

https://doi.org/10.1021/acs.jcim.4c01623
J. Chem. Inf. Model. 2025, 65, 722−748

ΔlogAUC). The system with the largest disparity in performance between GIST and 3D-RISM is the SII pocket system 8TXE. Compared to enrichment without receptor desolvation, logAUC increases by 1.24 points with GIST, but decreases by 4.38, 5.31, and 5.91 points with *exchem*, *solvene*, and *potUV* 3D-RISM, respectively (Table S6).

We compared enrichments with 3D-RISM receptor desolvation in the new frame and in the original test set frame to confirm that differences between 3D-RISM and GIST are mostly not due to the frame shift (Figure 7B). Docking with the *potUV* 3D-RISM grids results in better logAUC for only 4 out of 20 systems, a worse logAUC for 8 systems, and no significant difference for 8 systems, compared to enrichment without receptor desolvation. For example, comparing the original frame to the new frame for the 10 SI−SII pocket systems, 4 systems in the original frame compared to 2 systems in the new frame have better logAUC, 2 compared to 3 are worse, and 4 compared to 5 have no significant difference. However, the general trends remain the same for the mean ΔlogAUC. For both the new frame and original frame, the mean ΔlogAUC for the SI−SII pocket is positive, +0.41 and +0.62, and the mean ΔlogAUC for the SII pocket is negative, −1.61 and −1.54 (Table S6).

After testing GIST and 3D-RISM on the smaller subset of 20 systems, we further test GIST and 3D-RISM scoring on a larger set of systems with some changes to the grid generation procedures. We generate GIST and 3D-RISM grids and perform enrichment calculations on 123 KRAS systems in the test set. Because 3D-RISM is faster and easier to run than GIST, and does not require an MD simulation, we generated 3D-RISM grids aligned to the original test set frame for comparison to our original enrichment calculations discussed in Section 3.3.1. To run GIST on the larger set of systems, we reduce the computational demands of the GIST calculations by shortening the simulation length to 10 ns. We identify systems that have significantly different enrichment with and without receptor desolvation from GIST or 3D-RISM using the same criteria as discussed above (see Section 2). Because the total enthalpy GIST grids best correspond to the 3D-RISM *solvene* grids, we compare enrichment with GIST using the 10 ns simulations to enrichment with 3D-RISM *solvene*, shown in Figure 7C and Table S7. Note that the GIST and 3D-RISM results are in different frames. For the 3D-RISM *solvene* grids, we see that 3D-RISM performs worse on the SI−SII pocket systems: 8/40 (20%) systems have better enrichment with 3D-RISM, while 18/40 (45%) systems have worse enrichment. For the SII pocket, 40/83 (48%) systems have better enrichment, and 17/83 (20%) systems have worse. These are consistent with the mean ΔlogAUC ($\mu_{\Delta\text{logAUC}}$) of −0.25 compared to 0.26, for SI−SII and SII pockets, respectively. When GIST grids are used for receptor desolvation scoring, we see an improvement in enrichment for the SI−SII pocket systems: 26/40 (65%) have better enrichment and 8/40 (20%) have worse. There is not as much of an improvement for the SII pocket: 37/83 (45%) have better enrichment, and 35/83 (42%) have worse. However, for GIST, both pockets have a positive mean ΔlogAUC ($\mu_{\Delta\text{logAUC}}$), 0.66 and 0.68. Both GIST and 3D-RISM seem to help with enrichment. Combining the two pockets, we see 63/123 (51%) improve with GIST, and 54/123 (44%) improve with 3D-RISM (*potUV*); 43/123 (35%) and 37/123 (30%) perform worse with GIST and 3D-RISM (*potUV*), respectively. The remaining systems have no significant difference (Figure 7C, yellow). Thus, more systems perform better than worse for both GIST and 3D-RISM. GIST receptor desolvation does not seem to help as much as we expected based on enrichment to the subset of 20 systems, but still improves enrichment overall. For these enrichment tests, 3D-RISM does not perform as well on the SI−SII pocket as GIST, but it outperforms GIST on the SII pocket in terms of percent of systems, with more systems better and fewer systems worse. GIST outperforms 3D-RISM on both pockets in $\mu_{\Delta\text{logAUC}}$.

We also partition the SII pocket receptors based on whether the experimental structure contains covalent or noncovalent ligand bound. Since we already partition the enrichment ligand and decoy sets, we examine the enrichment performance of the noncovalent set on the noncovalent systems, and the covalent set on the covalent systems (Figure 7C, last two stacked bars, and Table S7). For the 24 SII pocket noncovalent systems, 9 systems are better, 9 systems are worse, 6 systems show no significant change, and the mean ΔlogAUC is negative (−0.31). For the 59 SII pocket covalent systems, 37 systems are better, 13 systems are worse, 9 systems show no significant change, and the mean ΔlogAUC is positive (0.48). See SI Section 5 for more discussion on enrichment across the 123 KRAS systems. Results shown in Figure S6 are summarized in Table S7. We also compare enrichment with property-matched decoys to enrichment with charge extrema decoys (Figure S7).

We also compare enrichment with and without GIST receptor desolvation for the 43 systems in the DUDE-Z set (SI Section 5, Table S8, Figure S8). When comparing blurry GIST, using a weighting factor of 0.5, to no receptor desolvation, we see that enrichment improves for 19 systems, is the same (not significantly different) for 14 systems, and worsens for 10 systems. However, when including blurry GIST with weights of 1 and 2, more systems worsen than improve.

3.3.3. Additional Testing of 3D-RISM Scoring. To more rigorously evaluate the effect of including 3D-RISM receptor desolvation scoring in docking, we performed pose reproduction experiments using the 3D-RISM grids generated in the test set frame. We compare pose reproduction results generated with 3D-RISM receptor desolvation scoring to results generated without receptor desolvation discussed in Section 3.2. The blurred solute−solvent potential energy (*potUV*) 3D-RISM grids and the blurry trilinear displacement receptor desolvation scoring method are used.

To evaluate the impact of 3D-RISM receptor desolvation scoring on pose reproduction, we performed cognate docking using the 127 KRAS systems with and without 3D-RISM. We classified each of the completed calculations as a docking success, scoring failure, or sampling failure, and compared systems that swap docking outcomes when docked with and without 3D-RISM. Table 7 shows the number of systems that swap between each docking outcome category. The diagonal shows the number of systems that do not change docking outcome, while the off-diagonal shows the number of systems that do. For instance, 5 systems that are scoring failures when docked without 3D-RISM are docking successes with 3D-RISM, and 8 systems are scoring failures both with 3D-RISM and without 3D-RISM. Most systems do not change docking outcomes (113 systems), while 13 systems swap between classifications. Overall, there are 7 more docking successes with 3D-RISM compared to without, 6 fewer scoring failures, and 1 less sampling failure. Including 3D-RISM in pose reproduction across all KRAS systems yields a 75.6% docking success rate, while pose reproduction without 3D-RISM yields a 70.1% docking success rate, showing a 5.5% improvement with 3D-

https://doi.org/10.1021/acs.jcim.4c01623
J. Chem. Inf. Model. 2025, 65, 722−748

Table 7. Number of Systems That Swap Pose Reproduction Docking Outcomes with and without 3D-RISM Receptor Desolvation[a]

	docking success (w/o 3D-RISM)	scoring failure (w/o 3D-RISM)	sampling failure (w/o 3D-RISM)	total (w/o 3D-RISM)
docking success (w/ 3D-RISM)	88	1	0	89
scoring failure (w/3D-RISM)	5	8	3	16
sampling failure (w/ 3D-RISM)	3	1	17	21
total (w/3D-RISM)	96	10	20	126[b]

[a]A system is said to swap outcomes when it has different outcomes between the two runs. [b]N = 127, one system fails to dock.

RISM. Although it is a small effect, including 3D-RISM does seem to improve sampling.

We investigate systems that swap between docking outcome classifications (Table 7 and SI Figure S5). In Figure 8, we show two examples where including 3D-RISM has a dramatic impact on sampling. An example of a system that swaps from a sampling failure without 3D-RISM to a docking success with 3D-RISM is 6GQY (Figure 8A–C). All poses generated with Chemgrid Score (blue points, Figure 8A) have a high RMSD to the experimental pose (all RMSDs > 6 Å), while the poses generated with Chemgrid Score with 3D-RISM (orange points, Figure 8A) have better RMSDs, including the best scoring pose, which has an RMSD < 1 Å. By visually examining the best scoring pose from both calculations, we see that the docked pose generated with Chemgrid Score alone, although partially overlapping spatially with the experimental pose, is flipped (Figure 8B). While for 3D-RISM, the red unfavorable water regions (favorable if displaced) help position the phenyl ether end of the ligand into the pocket and the pose is well overlaid (Figure 8C).

Another example of a system that swaps is 4M1W, which swaps from a docking success without 3D-RISM to a scoring failure with 3D-RISM. In this case, we see that both scoring methods produce poses with good RMSDs. However, when docking with 3D-RISM, the best scoring molecule has a large RMSD (~12 Å) (Figure 8D). Examining the poses, we see that the pose docked without 3D-RISM is well overlaid (Figure 8E), while the pose docked with 3D-RISM is positioned outside the pocket, and it does not overlap with the experimental pose at all (Figure 8D). See SI Figure S5 for DOCK Score vs RMSD plots for the other systems that swap docking outcomes.

We compare 3D-RISM docking grids produced with the molecular reconstruction approach (SI Section 6, Figure S9, and Tables S9–S12) to grids produced with the additive approach. Overall, docking with the additively combined 3D-RISM grids performs better than molecular reconstruction 3D-RISM grids,

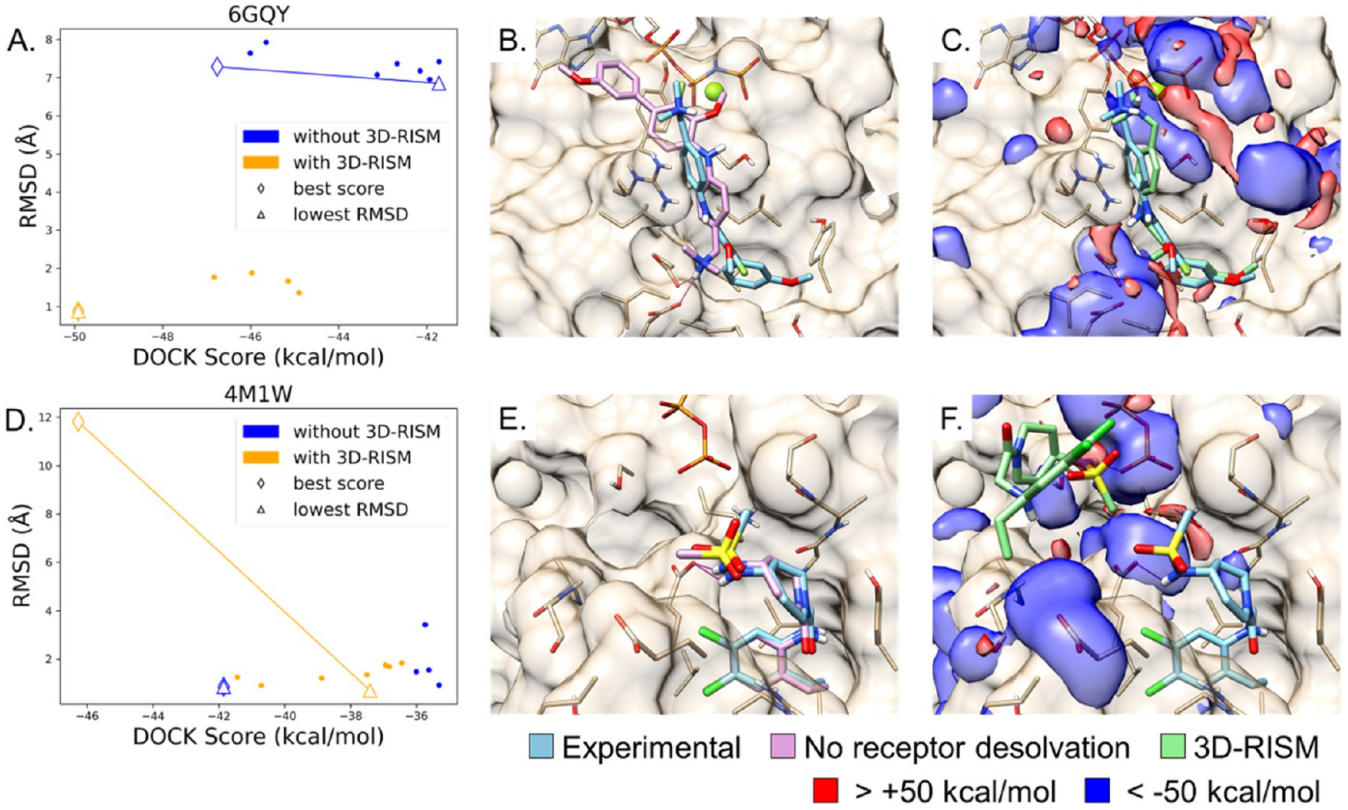


Figure 8. Two examples where incorporating receptor desolvation results in a different docking outcome and a dramatic pose change. (A–C) System that swaps from a sampling failure to a docking success after incorporating 3D-RISM receptor desolvation (PDB ID 6GQY). (D–F) System that swaps from a docking success to a scoring failure (PDB ID 4M1W). (A, D) Comparison of RMSD and DOCK score for poses sampled without (blue) and with (orange) 3D-RISM receptor desolvation. (B, E) Best scoring docked pose sampled without 3D-RISM receptor desolvation (purple). (C, F) Best scoring docked pose sampled with 3D-RISM receptor desolvation (green). Experimental poses shown in blue. 3D-RISM potUV grids shown in blue (favorable) and red (unfavorable) surfaces.

https://doi.org/10.1021/acs.jcim.4c01623
J. Chem. Inf. Model. 2025, 65, 722–748

and additively combined 3D-RISM is used in the main text unless otherwise stated. The different receptor desolvation scoring methods are compared using rescoring in SI Section 7. We analyze the impact of scoring function choice on sampling (Figures S10−S12), and look at the correlation between receptor desolvation terms for the different scoring methods (Figure S13). We also examined different weighting factors for GIST and 3D-RISM receptor desolvation terms, using the Gaussian-blurred grids and the trilinear scoring method (SI Section 8 and Figures S14 and S15). Weighting factors of 1 for GIST and 0.01 for 3D-RISM were determined to be reasonable and were used in the analyses of GIST and 3D-RISM scoring, unless otherwise stated.

3.4. Covalent Docking. We used systems from the test set to test the newly implemented covalent docking method, attach-and-grow. Docking calculations in Section 3.4.1 do not use receptor desolvation scoring. Covalent pose reproduction results with *attach-and-grow* were compared to noncovalent pose reproduction results, discussed in Section 3.2. In the proof-of-concept virtual screens discussed in Section 3.4.2, we combine the covalent *attach-and-grow* algorithm with receptor desolvation scoring.

3.4.1. Covalent Pose Reproduction. We performed pose reproduction using the *attach-and-grow* algorithm on 70 KRAS systems containing covalent ligands. Ligand conformations were prepared from the experimental structures (*xtal*) and from the SMILES strings (*smi*). Ligands prepared from SMILES may not be docked in the same conformation as ligands prepared from the experimental structure due to under-sampling, e.g., only docking chair ring conformations, or failing to sample the native geometry.

For *xtal* ligands, covalent docking results in a *docking success* rate of 78.6%, scoring failure rate of 2.9%, and sampling failure rate of 15.7% (Table 8). Noncovalent docking results in a docking success rate of 71.1%, scoring failure rate of 14.3% and sampling failure rate of 12.9%. Compared to noncovalent docking, covalent docking results in an increase in docking successes by 5 systems (+7.2%), a decrease in scoring failures by 8 systems (−11.4%), and an increase in sampling failures by 2 systems (+3.8%). For *smi* ligands, we see a substantial decrease in docking success rate: 47.1% for *smi* compared to 78.6% for *xtal* (−31.5%). The docking success rate is close to the sampling failure rate (47.1% and 42.9%), a substantial increase in sampling failures by 19 systems (27.2%) when compared to xtal covalent docking. The covalent and noncovalent docking success rates for *smi* ligands are quite similar (47.1% and 50.0%), only differing by 2 systems.

Table 8. Number of Systems (and Percentages) for Cognate Docking Pose Reproduction Outcomes Using Chemgrid Score for Both Covalent and Standard Docking

	docking success	scoring failure	sampling failure	did not dock
covalent (*xtal*)	55 (78.6%)	2 (2.9%)	11 (15.7%)	2 (2.9%)
noncovalent (*xtal*)	50 (71.4%)	10 (14.3%)	9 (12.9%)	1 (1.4%)
covalent (*smi*)	33 (47.1%)	7 (10.0%)	30 (42.9%)	0 (0.0%)
noncovalent (*smi*)	35 (50.0%)	7 (10.0%)	25 (35.7%)	3 (4.3%)

We compare Chemgrid Score and Grid Score, exploring different combination of terms and parameters, with both covalent and noncovalent docking (see SI Section 9). For both covalent and noncovalent *xtal* docking, more systems have a docking success outcome when Grid Score is used (SI Table S13), compared with Chemgrid Score (Table 8). This difference in behavior is not due to the ligand desolvation component—in fact, turning off desolvation decreases docking success outcomes by one system, and increases sampling failure outcomes by three systems for covalent *xtal* docking (SI Table S14). A more likely explanation is the difference in van der Waals exponents: we use a softer surface (6−9 exponents) for Grid Score, while a harder surface (6−12 exponents) is used for Chemgrid score. This harder surface would explain why there are fewer sampling failure outcomes for Grid Score compared to Chemgrid score without desolvation, particularly for *smi* docking (SI Tables S13 and S14). Softening the surface is known to help pose generation, but it hurts enrichment.[100] We reran the grid generation with the harder van der Waals exponents and, as expected, see an increase in sampling failure outcomes (SI Table S15). For *xtal* docking, the number of docking success outcomes is unchanged for covalent docking, and it increases for noncovalent docking with harder van der Waals exponents. This contrasts with *smi* docking, where docking success outcomes decrease for both covalent and noncovalent docking. Another factor that may explain the differences in the number of docking success outcomes is the difference in the treatment of electrostatics between the two scoring methods. Chemgrid Score uses Poisson−Boltzmann-generated grids with the binding site evacuated, while Grid Score uses grids generated using Coulomb's law with a distance-dependent dielectric. For *xtal* docking, more systems have a docking success outcome when covalent docking is performed, compared to noncovalent docking. For both Chemgrid Score and Grid Score, *xtal* covalent docking results in more *docking success* outcomes than *xtal* noncovalent docking (Tables 8 and S13). The opposite is true for *smi* docking, where noncovalent docking results in slightly more docking success outcomes for both scoring functions. Additionally, we ran Grid Score with the Chemgrid Score ligand desolvation component (Table S16). Including ligand desolvation (comparing Tables S13 and S16) in *xtal* docking increases *docking successes* by one and two systems for covalent and noncovalent docking, respectively. For *smi* docking, the results are mixed, increasing docking successes by one system for covalent docking, but decreasing by two systems for noncovalent docking. This is due to the increase in *did-not-dock* and sampling failure by one system for both. Overall, although a small effect, including ligand desolvation seems to help.

An example of a system that swaps from a sampling failure with noncovalent docking to a docking success with covalent docking (using *attach-and-grow*) is 6P8X (Figure 9A−C). We see that all covalently docked poses have RMSDs less than 2 Å, while all the noncovalently docked poses have RMSDs greater than 5 Å (Figure 9A). The covalent poses also have lower energies compared to the noncovalent poses. For both covalent and noncovalent docking, the best scoring pose is different from the lowest RMSD pose. For covalent docking, the two poses are both close in RMSD to the experimental pose, and by extension, to each other, despite the difference in score. We see that the best scoring pose for noncovalent docking partially occupies the same region as the experimental pose, but it is flipped and does not reproduce the experimental pose (Figure 9B). For covalent docking, we see that the best scoring pose is well overlaid with the experimental pose (Figure 9C).

An example of a system that swaps from a *docking success* with noncovalent docking to a *sampling failure* with covalent docking

https://doi.org/10.1021/acs.jcim.4c01623
J. Chem. Inf. Model. 2025, 65, 722−748

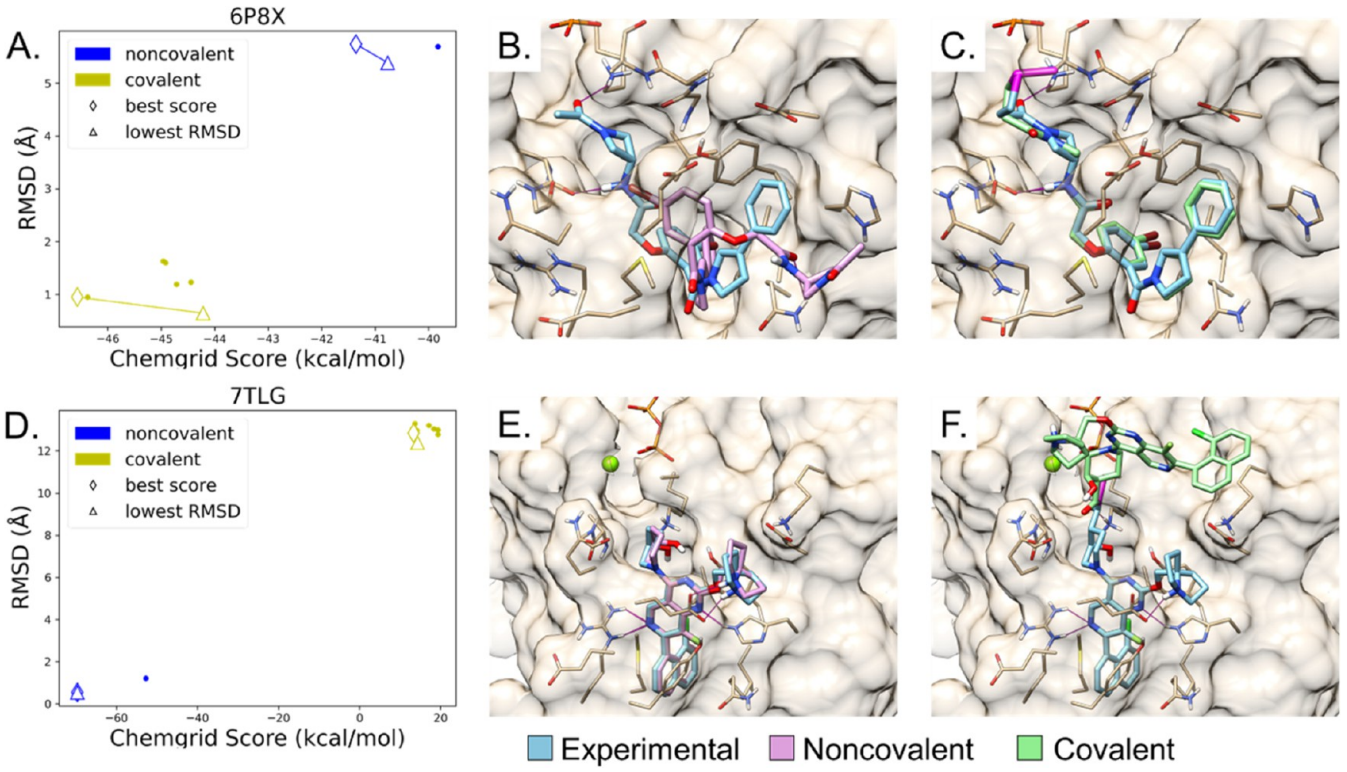


Figure 9. Two examples where covalent docking results in a different docking outcome and a dramatic pose change. (A−C) Example system that swaps from a sampling failure to a docking success after covalent docking with *attach-and-grow* (PDB ID 6P8X). (D−F) Example system that swaps from a docking success to a sampling failure (PDB ID 7TLG). (A, D) Comparison of RMSD and DOCK score for poses from standard noncovalent docking (blue) and poses from covalent docking (yellow). B, E. Best scoring docked pose from standard noncovalent docking (purple). (C, F) Best scoring docked pose from covalent docking (green). Experimental poses shown in blue.

(using *attach-and-grow*) is 7TLG (Figure 9D−F). For this system, the covalently docked poses all have extremely high RMSDs greater than 10 Å, while the noncovalently docked poses have RMSDs less than 2 Å (Figure 9D). The noncovalently docked poses also all have lower energies, while the covalently docked poses all have high energies above 0 kcal/mol. For noncovalent docking, the best scoring pose is also the lowest RMSD pose, and it is well overlaid with the experimental pose (Figure 9E). For covalent docking, the best scoring pose is completely outside of the binding pocket, and it does not overlay with the experimental pose at all (Figure 9F). For these two examples, we picked systems with a strong contrast between best-scoring and lowest RMSD poses. For RMSD-Score plots for all systems that swap classifications, see SI Section 10 and Figure S16.

3.4.2. Example Covalent Virtual Screen. We performed covalent virtual screens using databases of over 27 thousand Enamine in-stock and 3.4 million Enamine REAL make-on-demand acrylamide compounds,[101] with and without 3D-RISM receptor desolvation scoring. We focus on 3D-RISM to more rigorously evaluate this alternative solvation method in virtual screening, since GIST has already been tested in virtual screening campaigns.[31] We selected a covalent SII pocket system to dock to (PDB ID 7O83), based on a combination of overall enrichment performance and enrichment improvement with 3D-RISM. We first prepared the 27 thousand Enamine in-stock database molecules for covalent docking (see Section 2), and we docked the molecules with and without 3D-RISM. Specifically, we docked 46 955 molecules (stereoisomers/tautomers/protomers), which map to 27 921 unique Enamine codes, implying that on average there are 1.68 stereoisomers/tautomers/protomers per code. For Chemgrid score, DOCK 6 spends on average 9.17 s per molecule. With 3D-RISM, DOCK 6 spends on average 14.56 s per molecule, approximately 1.59 times longer per molecule than Chemgrid. Note these numbers are approximate because the runtime information was calculated over a cluster and not on one machine.

We also screened a database of approximately 3.4 million Enamine REAL make-on-demand acrylamide compounds. The make-on-demand database includes heavier and greasier compounds, and better matches the properties of the ligands compared to the in-stock database (see SI Section 11 and SI Figure S17). We docked 5 920 811 molecules (stereoisomers/tautomers/protomers) from 3 219 271 unique Enamine codes (1.84 stereoisomers/tautomers/protomers per code). For Chemgrid score, DOCK 6 spends on average 17.61 s per molecule. With 3D-RISM, DOCK 6 spends on average 26.52 s per molecule, approximately 1.51 times longer per molecule than Chemgrid.

We also docked 46 acrylamide ligands from the test set, to compare docking performance between the ligands and the Enamine compounds (Figure 10A,B). Using the Enamine molecules as our decoy background, we calculate AUC and logAUC to quantify how well the ligands from the test set rank among the screened molecules. When the Enamine in-stock acrylamide compounds are used as the "decoy" background, we obtain an AUC of 59.34 and logAUC of 12.77 for 3D-RISM, and an AUC of 62.39 and logAUC of 16.60 for no receptor desolvation, a ΔlogAUC of 3.83. When the Enamine REAL make-on-demand acrylamide compounds are used as the

https://doi.org/10.1021/acs.jcim.4c01623
J. Chem. Inf. Model. 2025, 65, 722−748

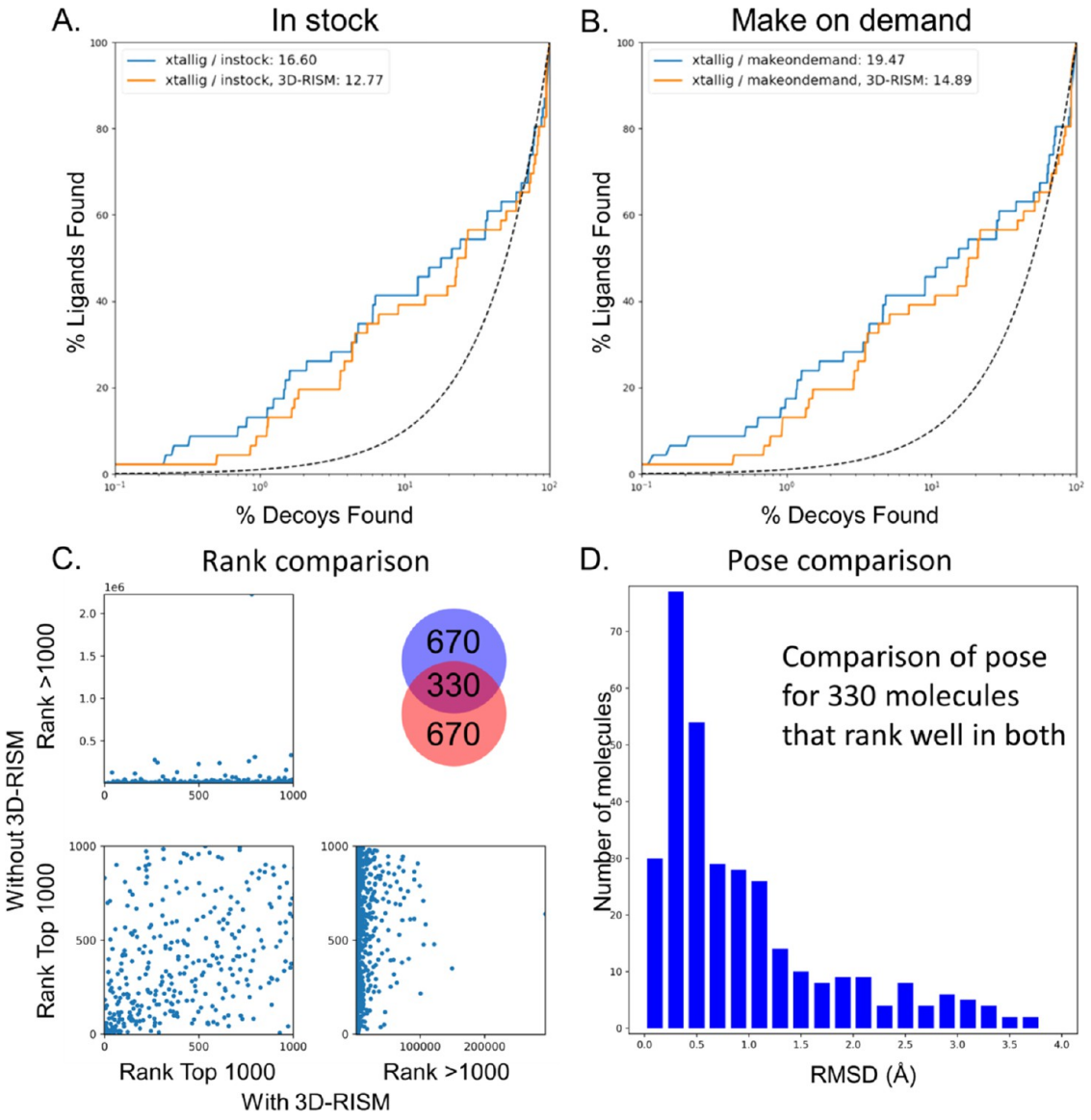


Figure 10. Receiver operator characteristic curves for proof-of-concept covalent virtual screens. *X*-axis is plotted on a logarithmic scale. Covalent acrylamide binders from the test set used as ligands. (A) Library of 27K Enamine in-stock acrylamide compounds was used as the decoy background. (B) Library of 3.4 M Enamine REAL make-on-demand acrylamide compounds was used as the decoy background. Screens without receptor desolvation plotted in blue, screens with 3D-RISM receptor desolvation plotted in orange. (C) For make-on-demand screens, comparison of the ranks for the top 1000 poses with and without 3D-RISM. Bottom-left quadrant shows the rank comparison for the 330 molecules the rank well in both. Bottom-right quadrant shows the rank comparison of the 670 molecules that rank well without 3D-RISM, but poorly with 3D-RISM. Top-left quadrant shows the rank comparison of the 670 molecules that rank well with 3D-RISM, but poorly without 3D-RISM. The Venn diagram comparing the 1000 top-ranked molecules that are in common and different between the two screens is shown. (D) For make-on-demand screens, for the 330 molecules that rank well in both, we compare RMSD to quantify pose similarity between the two methods in a histogram.

"decoy" background, we obtain an AUC of 62.08 and logAUC of 14.89 for 3D-RISM, and an AUC of 65.45 and logAUC of 19.47 for no receptor desolvation, a ΔlogAUC of 4.58. The logAUC using the Enamine make-on-demand compounds as the decoy background was slightly higher compared to the in-stock compounds, i.e., the ligands were proportionally ranked higher with the make-on-demand molecules as the background. Despite this system having significantly better enrichment with 3D-RISM, when using noncovalent HDB docking on the SII pocket enrichment set, for both the in-stock and make-on-demand screens, enrichment with 3D-RISM was worse than enrichment with only Chemgrid. It is worth noting that the logAUC calculated on the screening results is not necessarily indicative of better enrichment performance, as top scoring molecules from the screen might not be decoys.

To analyze the impact of including 3D-RISM receptor desolvation scoring compared to using Chemgrid Score alone on our docking behavior, we compare the changes in rank (Figure 10C) and poses (Figure 10D) for the virtual screens using the Enamine make-on-demand compounds with and without receptor desolvation scoring. The union of the top scoring 1000 molecules for both screens yields 1670 molecules. The intersection of the top scoring 1000 molecules for both screens yields 330 molecules, i.e., there are 330 molecules in common in the top scoring 1000 molecules between the two screens. For both screens, 670 molecules rank well in one screen and not in the other. Some of the molecules dramatically change in rank.

https://doi.org/10.1021/acs.jcim.4c01623
J. Chem. Inf. Model. 2025, 65, 722−748

Table 9. Number of Systems Categorized by Docking Outcome and Pose Reproduction Error Analysis over 100 Simplex Random Seeds for the Minimizer[a], Comparing Covalent and Noncovalent Docking, and Docking with and without 3D-RISM

	docking success	scoring failure	sampling failure	did not dock
covalent (*xtal*)[b]	54.60 ± 1.21	2.54 ± 1.24	10.89 ± 0.80	1.96 ± 0.20
noncovalent (*xtal*)	52.91 ± 1.92	6.02 ± 1.95	10.19 ± 1.47	0.88 ± 0.33
p-value (2-sample)	2.19e-12	6.99e-35	3.81e-05	1.63e-72
difference	1.69 ± 2.30 (2.40 ± 3.29%)	−3.48 ± 2.37	0.70 ± 1.59	1.08 ± 0.39
p-value (1-sample)	4.44e-11	9.05e-27	2.32e-05	1.16e-48
covalent (smi)[b]	33.94 ± 1.22	4.34 ± 1.70	31.72 ± 1.30	0.00 ± 0.00
noncovalent (*smi*)	34.83 ± 2.04	7.26 ± 2.10	24.91 ± 2.06	3.00 ± 0.00
p-value (2-sample)	2.19e-04	6.54e-22	2.56e-71	N/A
difference	−0.89 ± 2.33 (−1.27 ± 3.33%)	2.92 ± 2.58	6.81 ± 2.39	−3.00 ± 0.00
p-value (1-sample)	2.10e-04	8.83e-20	4.62e-50	N/A
3D-RISM[c]	95.53 ± 2.11	11.38 ± 2.31	19.25 ± 1.65	0.84 ± 0.37
no receptor desolvation	93.23 ± 2.28	11.49 ± 2.46	21.41 ± 1.88	0.88 ± 0.33
p-value (2-sample)	2.53e-12	7.46e-01	1.43e-15	4.18e-01
difference	2.31 ± 3.06 (1.80 ± 2.41%)	−0.11 ± 3.23	−2.16 ± 2.64	−0.04 ± 0.45
p-value (1-sample)	1.73e-11	7.35e-01	7.42e-13	3.74e-01

[a]Statistics based on 100 runs. [b]Covalent and noncovalent comparison is on 70 systems. [c]3D-RISM and no receptor desolvation comparison is on 127 systems.

For example, the most extreme difference in rank, a molecule that has a rank of 798 without 3D-RISM and changes rank to 311 340 with 3D-RISM. Another molecule has a rank of 639 with 3D-RISM and changes rank to 293 879 without 3D-RISM. For the 330 molecules the rank well in both screens, we compare the RMSDs to quantify pose similarity. Most of the time the poses are similar (218 with RMSD < 1 Å) and (285 with RMSD < 2 Å). A small number of the poses are different (14 with RMSD > 3 Å).

3.5. Confidence in Pose Reproduction Experiments. To obtain confidence in our pose reproduction calculations, we perturbed sampling by modifying the simplex random seed for minimization (see Section 2).

For both covalent docking and docking with 3D-RISM, we see a slight average increase in the number of docking success outcomes when compared to standard noncovalent docking (Table 9). The improvement is small, an increase of 1.69 and 2.31 systems, respectively, albeit significant (p-value $\leq$ 0.001). For covalent docking, there is a decrease in scoring failure outcomes and an increase in sampling failure outcomes. This trend also follows the covalent and noncovalent pose reproduction results for a single random seed, shown in Table 8, where the number of docking success and sampling failure outcomes increase, compensated by scoring failure outcomes decreasing. For 3D-RISM docking, there is a decrease in sampling failure outcomes, and a slight decrease or no change in scoring failure outcomes. This differs slightly from the pose reproduction results with a single random seed, shown in Table 7. For both (Tables 7 and 9), the docking success outcomes increase with 3D-RISM (7 and 2.31 systems). Most of the change comes from a decrease in scoring failures with 3D-RISM, rather than a decrease in sampling failures.

We also test our covalent docking method on an existing test set (see SI Section 12 and Table S17) comprised of 207 systems of diverse targets.[49] We see that DOCK 6 with the *attach-and-grow* algorithm has a docking success rate of 53.11 ± 1.31%, which is in the middle of the pack of the different covalent docking methods discussed in Scarpino et al.[49] The outcome for this more diverse set (53.11%) is somewhat stronger than 48.49% (33.94/70) for the RAS-specific set, but is comparable.

3.6. Future Directions. In addition to GIST and 3D-RISM, we are exploring the Gaussian weighting method for docking to other types of grids. As an alternative method to *attach-and-grow*, we have implemented the DOCKovalent algorithm in DOCK 6, which we are evaluating. A scoring function is being evaluated for covalent pose minimization to help refine covalent poses generated with the DOCKovalent approach. These methods will be released in a future version of DOCK.

The RAS test set is available on GitHub at https://github.com/tbalius/teb_docking_test_sets. We intend to update and maintain this test set as new structures are released. In addition to KRAS, we have also docked to NRAS and HRAS, as well as KRAS DNA (SI Section 13 and Figure S18). A few improvements to the test set have been made in an update (see SI Section 14, Figures S19 and S20). The GitHub repository also includes several tutorials for docking to RAS using different methods, which we plan on maintaining, updating, and expanding.

4. DISCUSSION

Five key findings have emerged in this work:

1. Through the construction of a RAS test set, we evaluated docking performance on this enigmatic target. In this work, we introduce a focused test set of RAS systems. Although there are other test sets that are more extensive,[93] and feature diverse targets,[102,103] by focusing on RAS, we can better examine the behavior of our docking algorithms in depth on a few systems. Even when focusing on only RAS proteins, we see a large variability in docking behavior. We explore the behavior of different sampling methods, e.g., *attach-and-grow* and *anchor-and-grow*, and different scoring methods, e.g., Grid Score, Chemgrid Score, and Chemgrid Score with RecDesolv Score, as well as the interplay between sampling and scoring.
2. Through cross-docking experiments, we see that within and among pockets, there are many combinations of ligands and receptors that fail to reproduce the experimental pose, contrasting with the high cognate docking success rate. In Figure 4 and Tables 4 and 5, we see that the cross-docking success rate is always lower

https://doi.org/10.1021/acs.jcim.4c01623
J. Chem. Inf. Model. 2025, 65, 722−748

than cognate docking. There are 8 systems that result in all sampling failures from the SI−SII pocket cross-docking experiment (Figure 4A). PDB ID 6GJ8 features the small molecule BI-2852 (PDB chemical ID F0K) which binds at a dimer interface;[104] docking to the monomer results in a sampling failure, however, when we dock to the dimer we reproduce the experimental pose.[105] Similarly, PDB IDs 7ACA, 7ACF, 7ACH, and 7ACQ all feature the same ligand, PDB chemical ID R6W, a large symmetric analog of BI-2852 designed to bind at the dimer interface.[106] PDB ID 4PZY features a ligand that binds to a symmetric pocket on a RAS molecule in the crystal structure (addressed in SI Section 14, Figure S19). For these cases, the cross-docking sampling failures result from not considering the additional interactions from a neighboring RAS protein, making it difficult to sample the experimental pose. For PDB ID 6GQT, the portion of the ligand that engages with the receptor is well overlaid while the solvent-exposed portion of the ligand is not. The docked pose also clashes with a neighboring RAS protein in the crystal, which is not included in our docking. For PDB ID 4EPX, the ligand binds with a polar group buried in the binding pocket, however, the Chemgrid Score ligand desolvation component penalizes this and prevents the experimental pose from being adopted. When we perform cognate docking with Grid Score—which does not include ligand desolvation—the experimental pose is reproduced (lowest RMSD = 1.4 Å). We see that the SI−SII pocket has a lower cognate docking success rate, in part due to these 8 systems, but a higher cross-docking success rate, when compared to the SII pocket. The poor cross-docking success rate is likely due to incompatibilities between ligand−receptor pairs, and within pockets, there are fewer, although still substantial, incompatibilities than between pockets. Since RAS proteins are intrinsically dynamic, incompatibilities may result from conformational differences in the backbone or side chains. In addition, incompatibilities may result from differences in sequence, or missing regions of the protein. Conformational changes of the backbone may dramatically change the shape and nature of the ligand binding pockets. As seen in Figure 3, the Switch I and Switch II regions can have dramatic changes in conformation, and the state of these Switch regions can determine if a binding pocket is open or closed to ligands. Conformational changes of the side chains can also alter the shape and nature of the pocket; however, backbone conformational changes are often more substantial than side chain movements. Differences in sequence are another factor that can change the nature of a site, e.g., a mutation from a hydrophilic to hydrophobic side chain. Among the protein structures of the test set, there are several sequence differences (see SI Figure S1 and Table S2). There are some structures that are missing stretches of residues (see SI Table S1.1). If these missing regions are proximal to the binding site, there may be an impact on the docking results. Differences in the matching spheres can impact the docking. Because ligand atoms are converted to spheres, even for similar protein structures, different ligands will lead to different matching spheres used for orienting. Preparation of the proteins, particularly for histidine residues, may differ among systems. The proteins in the test set are complexed with different nucleotides: GDP, GPP[NH]P, GPP[CH2]P, or GTPγS (SI Table S3), which may impact the docking. When defining incompatibility, we focus on steric incompatibilities, mostly driven by conformational changes in the backbone and side chains. To overcome these steric issues, there are many methods that attempt to account for protein flexibility in docking.[107−109] Here, we use the rigid receptor approximation, and do not explore receptor flexibility during docking.

3. Including receptor desolvation provides a small, but significant, improvement to docking performance. This is in agreement with earlier work[31] on receptor desolvation scoring in DOCK 3. Previously,[32] we tested the blurry trilinear displacement GIST implementation in DOCK 3.7/3.8 on the DUDEZ benchmark, and on prospective screens against AmpC β-lactamase. We further examined the importance of water rearrangement through molecular dynamics simulations. In comparing two different scoring functions, we surprisingly observe that sometimes sampling with one scoring function results in a better score with the other. We found this to be so in both previous work[32] and herein (see SI Section 7). To test receptor desolvation scoring in DOCK 6, we prepared both GIST (100 ns) and 3D-RISM grids for 20 KRAS systems, and we compared enrichment performance with and without receptor desolvation. In addition, we prepared GIST (10 ns) and 3D-RISM grids for all KRAS protein systems in the test set, and we compared pose reproduction and enrichment performance with and without receptor desolvation. Over the set of 20 KRAS systems, we see that accounting for desolvation with GIST (Chemgrid + GIST Score) consistently improved enrichment when compared to using Chemgrid Score alone. For the same set, including 3D-RISM did not improve enrichment. Using GIST to account for receptor desolvation improved enrichment for 80% of the systems, while using 3D-RISM only improved enrichment for 20−25% of systems. However, extending enrichment calculations across the whole test set (N = 123), we see that both GIST and 3D-RISM improve enrichment more often than not. When we incorporate either GIST or 3D-RISM, 51 or 44% of systems have improved logAUC, while 35 or 30% of systems have worse logAUC, respectively. When we expand our receptor desolvation calculations to the whole test set, we saw a strong performance for the SI−SII pocket systems but a poorer performance on the SII pocket systems. We speculate that the worse performance with GIST over the full set compared to the subset of 20 systems might be due to similarities in ligands and receptor conformation in the systems selected, as well as conformational variability in the full set of SII pocket systems. Also, some systems are missing loop regions. These improvements in enrichment logAUC are small, but statistically significant. In pose reproduction, docking with 3D-RISM results in a 5.5% increase in success rate, compared to docking without 3D-RISM. Accounting for receptor desolvation provides a small, but significant, improvement in both enrichment and pose reproduction on KRAS. Resampling and multiple replicas allowed us to estimate uncertainty in our docking calculations. We resampled our enrichment results using bootstrapping. We calculated multiple replicas by docking with different random seeds for

https://doi.org/10.1021/acs.jcim.4c01623
J. Chem. Inf. Model. 2025, 65, 722−748

pose reproduction. For both resampling with bootstrapping and multiple replicas, we generated distributions that we used to calculate *p*-values to quantify confidence in the result. For example, even though the change in docking success is small in Table 9, the p-values are less than 10^{-11}, implying strong confidence in the difference.

4. Covalent docking with *attach-and-grow* gives comparable pose reproduction results to noncovalent docking using *anchor-and-grow*. Covalent docking also provides an inherent speed up over noncovalent docking, due to sampling less degrees of freedom. We tested the covalent *attach-and-grow* algorithm on 70 KRAS systems containing covalent ligands. For pose reproduction with the experimental ligand conformation, in which the starting conformation preserves the experimental bonds and angles but forgets the dihedral angles, the docking success rate was 71.4% using noncovalent *anchor-and-grow*, and 78.6% using covalent *attach-and-grow*. We see an increase in the number of docking successes by 5 systems with covalent docking (Table 8), and for the error analysis experiment, we see an average increase in docking success by 1.69 ± 2.30 systems (Table 8). However, this is not the case when docking from SMILES, where *anchor-and-grow* outperforms *attach-and-grow* by 2 systems (Table 8). One reason for the discrepancy is because including the covalent bond in covalent docking adds additional constraints to the sampling, while noncovalent docking allows more degrees of freedom that can resolve issues arising from using a generated ligand conformation, which differs from the experimental conformation. The three translational and three rotational degrees of freedom give the molecule more opportunity to be placed in the right location. Previously,[88] a developmental version of the *attach-and-grow* covalent docking method was tested on a different set of proteins. The *attach-and-grow* algorithm was inspired by both DOCKovalent (which is a modification of the HDB-search routine implemented within DOCK 3.7) and *anchor-and-grow*. In contrast to *attach-and-grow*, one weakness of DOCKovalent is not minimizing the ligand conformation. In *attach-and-grow*, the torsions are minimized at every stage of growth (using the *premin* option in DOCK 6). Likely due to minimization, the *attach-and-grow* algorithm seems to perform better in pose reproduction.
5. Through our proof-of-concept virtual screen, we demonstrate our ability to perform a ligand discovery campaign using the covalent *attach-and-grow* algorithm. One of the main applications of docking is as a ligand discovery tool. As a proof-of-concept, we performed covalent virtual screens with and without receptor desolvation scoring, targeting the switch II pocket of KRAS using 3.4 million make-on-demand acrylamide compounds from the Enamine REAL database. On average, the *attach-and-grow* algorithm spends approximately 17.61 s per molecule across the screened molecules. The covalent database of SMILES is processed using *RDKit*, and the covalent adduct is attached to prepare the molecules for docking, as discussed in Section 2. Because we process the SMILES in this way, we do not mix covalent warheads when performing a virtual screen, due to differences in the covalent bond environment. Molecules with different warheads should be prepared and docked separately due to differences in reactivity and other issues with docking the adduct. Herein, we dock a database containing only acrylamide compounds. The modular nature of DOCK 6 allows swapping scoring and sampling methods, and we can use Grid Score, Chemgrid Score, or Chemgrid + 3D-RISM Score with *anchor-and-grow*, *attach-and-grow*, or HDB search. In this way, new features may be combined with existing features, and tested. We highlight this in the virtual screen, where we combine covalent docking with receptor desolvation scoring.

There are some caveats that merit discussion. (1) There are many parameters to test, of which we are limited to only explore a fraction (SI Section 15). In the same vein, we only explore these methods on a small test set comprised of a small number of systems relative to the extensive body of structural data available. If we expand our study to more and more parameters and more and more systems, we might find results contradictory to some of our conclusions. However, these are control calculations, and they are used to demonstrate the utility of the methods. (2) To be fast, docking scoring functions ignore many important contributions to binding. Neglected terms include entropic contributions, receptor flexibility, and bridging waters, to name a few. To more accurately model water's role in binding, we should not only account for desolvation, but we should also model waters that are not displaced. Furthermore, these waters are not static, and we should also model how they rearrange. (3) We neglect accounting for the covalent bond formation in scoring. One of the most important drivers for covalent ligands binding to the receptor is the covalent bond formation, in which the covalent warhead reacts with a residue. For example, an acrylamide warhead reacts with a cysteine residue through a nucleophilic attack.[42] Different warheads, their substituents, different reacting residues, and the protein environment around the residue all play a role in the strength of the reaction. Sampling with the *attach-and-grow* algorithm is performed with the reaction having already occurred, and we neglect the strength of the reaction in our scoring. (4) All results here are retrospective, and one of the most important applications of docking is prospective predictions. While our control calculations demonstrate the utility of including receptor desolvation scoring and of the covalent *attach-and-grow* algorithm, the most convincing tests are prospective tests.

These caveats should not detract from the key findings. In this work, we present three interwoven topics: the RAS test set, and the evaluation of two docking improvements—receptor desolvation scoring and covalent sampling. To test docking methods, we have assembled a RAS test set consisting of 140 structures (138 protein and 2 DNA), in complex with ligands. We have introduced two new features in DOCK 6.12: a covalent docking algorithm, coined *attach-and-grow*, and a scoring function that accounts for receptor desolvation using grids from GIST or 3D-RISM. We precompute displacements with a Gaussian-weighting, so that trilinear interpolation may be used to speed up the receptor desolvation calculation compared to the earlier displacement implementation. We use this RAS test set to evaluate UCSF DOCK 6 performance. We tested DOCK 6 and these new features, with pose reproduction experiments, through cognate and cross-docking, and enrichment calculations. Within and among pockets, there are many combinations of ligands and receptors that are incompatible. Our results suggest that accounting for receptor desolvation provides a small, but significant, improvement in both enrichment and pose

https://doi.org/10.1021/acs.jcim.4c01623
J. Chem. Inf. Model. 2025, 65, 722−748

reproduction. Covalent docking with *attach-and-grow* gives comparable pose reproduction results to noncovalent docking with *anchor-and-grow*. As a proof-of-concept, we performed covalent virtual screens targeting the switch II pocket of KRAS with and without receptor desolvation scoring. The test set is available at https://github.com/tbalius/teb_docking_test_sets.

■ ASSOCIATED CONTENT

Data Availability Statement

The docking methods are available through UCSF DOCK 6.12 at https://dock.compbio.ucsf.edu/DOCK_6/index.htm. The test set and associated scripts are available on GitHub at https://github.com/tbalius/teb_docking_test_sets. Additional scripts are made available on GitHub at https://github.com/tbalius/teb_scripts_programs. Other data is available upon request.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.4c01623.

In this Supporting Information, there are 15 sections, including sections featuring additional test set information, enrichment and pose reproduction experiments, updates to the test set, and parameter selection. These supporting sections include supplemental text, tables, and figures. (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Trent E. Balius − *NCI RAS Initiative, Cancer Research Technology Program, Frederick National Laboratory for Cancer Research, Leidos Biomedical Research, Inc., Frederick 21702 Maryland, United States;* orcid.org/0000-0002-6811-4667; Email: trent.balius@nih.gov

Authors

Y. Stanley Tan − *NCI RAS Initiative, Cancer Research Technology Program, Frederick National Laboratory for Cancer Research, Leidos Biomedical Research, Inc., Frederick 21702 Maryland, United States*

Mayukh Chakrabarti − *NCI RAS Initiative, Cancer Research Technology Program, Frederick National Laboratory for Cancer Research, Leidos Biomedical Research, Inc., Frederick 21702 Maryland, United States;* orcid.org/0000-0003-2653-9728

Reed M. Stein − *Department of Pharmaceutical Chemistry, University of California—San Francisco, San Francisco 94158 California, United States*

Lauren E. Prentis − *Department of Biochemistry and Structural Biology, Stony Brook University, Stony Brook 11794 New York, United States; Institute of Chemical Biology and Drug Discovery, Stony Brook University, Stony Brook 11794 New York, United States*

Robert C. Rizzo − *Institute of Chemical Biology and Drug Discovery, Stony Brook University, Stony Brook 11794 New York, United States; Department of Applied Mathematics and Statistics and Laufer Center for Physical and Quantitative Biology, Stony Brook University, Stony Brook 11794 New York, United States;* orcid.org/0000-0003-0525-6147

Tom Kurtzman − *PhD Programs in Chemistry, Biochemistry, and Biology, The Graduate Center of the City University of New York, New York 10016, New York, United States; Department of Chemistry, Lehman College, The City University of New York, Bronx 10468 New York, United States;* orcid.org/0000-0003-0900-772X

Marcus Fischer − *Department of Chemical Biology and Therapeutics, St. Jude Children's Research Hospital, Memphis 38105 Tennessee, United States;* orcid.org/0000-0002-7179-2581

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.4c01623

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank the DOCK Community, including Drs. Brian K. Shoichet, Nir London, John J. Irwin, and Jiankun Lyu, for helpful discussions, the NCI RAS Initiative, including Drs. Dwight Nissley and Frank McCormick, for support and the Frederick Research Computing Environment (FRCE). We thank Drs. Chao Yang for beta testing the covalent code and reporting bugs, Scott Brozell for work on the DOCK 6.12 release, and Tyler Luchko for answering our questions on 3D-RISM. This project was funded in whole or in part with federal funds from the National Cancer Institute, NIH Contract 75N91019D00024, NIH grants R35GM126906 (to R.C.R.), R35GM144089 (to T.K.), R35GM142772 (to M.F.), and ALSAC (to M.F.). The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, and the mention of trade names, commercial products, or organizations does not imply endorsement by the US Government.

■ REFERENCES

(1) Prior, I. A.; Hood, F. E.; Hartley, J. L. The Frequency of Ras Mutations in Cancer. *Cancer Res.* **2020**, *80* (14), 2969−2974. (acccessed 6/27/2024)

(2) Hood, F. E.; Sahraoui, Y. M.; Jenkins, R. E.; Prior, I. A. Ras protein abundance correlates with Ras isoform mutation patterns in cancer. *Oncogene* **2023**, *42* (15), 1224−1232.

(3) Liu, Y.; Xie, B.; Chen, Q. RAS signaling and immune cells: a sinister crosstalk in the tumor microenvironment. *J. Transl. Med.* **2023**, *21* (1), 595.

(4) Simanshu, D. K.; Nissley, D. V.; McCormick, F. RAS Proteins and Their Regulators in Human Disease. *Cell* **2017**, *170* (1), 17−33.

(5) Moore, A. R.; Rosenberg, S. C.; McCormick, F.; Malek, S. RAS-targeted therapies: is the undruggable drugged? *Nat. Rev. Drug Discovery* **2020**, *19* (8), 533−552.

(6) Zhao, Z.; Bohidar, N.; Bourne, P. E. Analysis of KRAS−Ligand Interaction Modes and Flexibilities Reveals the Binding Characteristics. *J. Chem. Inf. Model.* **2023**, *63* (4), 1362−1370.

(7) Parker, M. I.; Meyer, J. E.; Golemis, E. A.; Dunbrack, R. L., Jr. Delineating the RAS Conformational Landscape. *Cancer Res.* **2022**, *82* (13), 2485−2498.

(8) Irwin, J. J.; Shoichet, B. K. Docking Screens for Novel Ligands Conferring New Biology. *J. Med. Chem.* **2016**, *59* (9), 4103−4120.

(9) Trott, O.; Olson, A. J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, *31* (2), 455−461. https://doi.org/10.1002/jcc.21334

(10) 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.; et al. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47* (7), 1739−1749.

(11) McGann, M. R.; Almond, H. R.; Nicholls, A.; Grant, J. A.; Brown, F. K. Gaussian docking functions. *Biopolymers* **2003**, *68* (1), 76−90.

(12) McNutt, A. T.; Francoeur, P.; Aggarwal, R.; Masuda, T.; Meli, R.; Ragoza, M.; Sunseri, J.; Koes, D. R. GNINA 1.0: molecular docking with deep learning. *J. Cheminf.* **2021**, *13* (1), 43.
(13) 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.
(14) Ruiz-Carmona, S.; Alvarez-Garcia, D.; Foloppe, N.; Garmendia-Doval, A. B.; Juhos, S.; Schmidtke, P.; Barril, X.; Hubbard, R. E.; Morley, S. D. rDock: A Fast, Versatile and Open Source Program for Docking Ligands to Proteins and Nucleic Acids. *PLOS Comput. Biol.* **2014**, *10* (4), No. e1003571.
(15) Majeux, N.; Scarsi, M.; Apostolakis, J.; Ehrhardt, C.; Caflisch, A. Exhaustive docking of molecular fragments with electrostatic solvation. *Proteins: Struct., Funct., Bioinf.* **1999**, *37* (1), 88−105.
(16) Korb, O.; Stützle, T.; Exner, T. E. PLANTS: Application of Ant Colony Optimization to Structure-Based Drug Design. In *Ant Colony Optimization and Swarm Intelligence*; Dorigo, M.; Gambardella, L. M.; Birattari, M.; Martinoli, A.; Poli, R.; Stützle, T., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2006; pp 247−258.
(17) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking11Edited by F. E. Cohen. *J. Mol. Biol.* **1997**, *267* (3), 727−748.
(18) Corso, G.; Stärk, H.; Jing, B.; Barzilay, R.; Jaakkola, T.DiffDock: Diffusion Steps, Twists, and Turns for Molecular Docking *arXiv preprint arXiv:2210.01776* 2023 DOI: 10.48550/arXiv.2210.01776.
(19) Pagadala, N. S.; Syed, K.; Tuszynski, J. Software for molecular docking: a review. *Biophys Rev.* **2017**, *9* (2), 91−102. From NLM
(20) Paggi, J. M.; Pandit, A.; Dror, R. O. The Art and Science of Molecular Docking. *Annu. Rev. Biochem.* **2024**, *93*, 389.
(21) 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.
(22) Darby, J. F.; Hopkins, A. P.; Shimizu, S.; Roberts, S. M.; Brannigan, J. A.; Turkenburg, J. P.; Thomas, G. H.; Hubbard, R. E.; Fischer, M. Water Networks Can Determine the Affinity of Ligand Binding to Proteins. *J. Am. Chem. Soc.* **2019**, *141* (40), 15818−15826.
(23) Stachowski, T. R.; Vanarotti, M.; Seetharaman, J.; Lopez, K.; Fischer, M. Water Networks Repopulate Protein−Ligand Interfaces with Temperature. *Angew. Chem.* **2022**, *134* (31), No. e202112919.
(24) Leini, R.; Pantsar, T. In Silico Evaluation of the Thr58-Associated Conserved Water with KRAS Switch-II Pocket Binders. *J. Chem. Inf. Model.* **2023**, *63* (5), 1490−1505.
(25) Kumar, S. U.; Priya Doss, C. G. Residue interaction networks of K-Ras protein with water molecules identifies the potential role of switch II and P-loop. *Comput. Biol. Med.* **2021**, *135*, No. 104597.
(26) Pantsar, T. KRAS(G12C)−AMG 510 interaction dynamics revealed by all-atom molecular dynamics simulations. *Sci. Rep.* **2020**, *10* (1), No. 11992.
(27) Hu, X.; Maffucci, I.; Contini, A. Advances in the Treatment of Explicit Water Molecules in Docking and Binding Free Energy Calculations. *Curr. Med. Chem.* **2020**, *26* (42), 7598−7622.
(28) Lazaridis, T. Inhomogeneous Fluid Approach to Solvation Thermodynamics. 1. Theory. *J. Phys. Chem. B* **1998**, *102* (18), 3531−3541.
(29) Lazaridis, T. Inhomogeneous Fluid Approach to Solvation Thermodynamics. 2. Applications to Simple Fluids. *J. Phys. Chem. B* **1998**, *102* (18), 3542−3550.
(30) Nguyen, C. N.; Kurtzman Young, T.; Gilson, M. K. Grid inhomogeneous solvation theory: Hydration structure and thermodynamics of the miniature receptor cucurbit[7]uril. *J. Chem. Phys.* **2012**, *137* (4), No. 044101.
(31) Balius, T. E.; Fischer, M.; Stein, R. M.; Adler, T. B.; Nguyen, C. N.; Cruz, A.; Gilson, M. K.; Kurtzman, T.; Shoichet, B. K. Testing inhomogeneous solvation theory in structure-based ligand discovery. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114* (33), E6839−E6846.
(32) Stein, R. M.; Singh, I.; Ghattas, M.; Kurtzman, T.; Balius, T. E.; Shoichet, B. K. Chapter 2: Testing a Faster Implementation of IST in Ligand Discovery. In *Understanding Virtual Solvent Through Large-scale Ligand Discovery*; University of California, 2020; pp 61−110.
(33) Abel, R.; Salam, N. K.; Shelley, J.; Farid, R.; Friesner, R. A.; Sherman, W. Contribution of Explicit Solvent Effects to the Binding Affinity of Small-Molecule Inhibitors in Blood Coagulation Factor Serine Proteases. *ChemMedChem* **2011**, *6* (6), 1049−1066.
(34) Abel, R.; Young, T.; Farid, R.; Berne, B. J.; Friesner, R. A. Role of the Active-Site Solvent in the Thermodynamics of Factor Xa Ligand Binding. *J. Am. Chem. Soc.* **2008**, *130* (9), 2817−2831.
(35) Li, Z.; Lazaridis, T. Computing the Thermodynamic Contributions of Interfacial Water. In *Computational Drug Discovery and Design*; Baron, R., Ed.; Springer New York, 2012; pp 393−404.
(36) Nguyen, C. N.; Cruz, A.; Gilson, M. K.; Kurtzman, T. Thermodynamics of Water in an Enzyme Active Site: Grid-Based Hydration Analysis of Coagulation Factor Xa. *J. Chem. Theory Comput.* **2014**, *10* (7), 2769−2780.
(37) Haider, K.; Cruz, A.; Ramsey, S.; Gilson, M. K.; Kurtzman, T. Solvation Structure and Thermodynamic Mapping (SSTMap): An Open-Source, Flexible Package for the Analysis of Water in Molecular Dynamics Trajectories. *J. Chem. Theory Comput.* **2018**, *14* (1), 418−425.
(38) Ramsey, S.; Nguyen, C.; Salomon-Ferrer, R.; Walker, R. C.; Gilson, M. K.; Kurtzman, T. Solvation thermodynamic mapping of molecular surfaces in AmberTools: GIST. *J. Comput. Chem.* **2016**, *37* (21), 2029−2037.
(39) Majeux, N.; Scarsi, M.; Caflisch, A. Efficient electrostatic solvation model for protein-fragment docking. *Proteins: Struct., Funct., Bioinf.* **2001**, *42* (2), 256−268.
(40) Forli, S.; Olson, A. J. A Force Field with Discrete Displaceable Waters and Desolvation Entropy for Hydrated Ligand Docking. *J. Med. Chem.* **2012**, *55* (2), 623−638.
(41) Murphy, R. B.; Repasky, M. P.; Greenwood, J. R.; Tubert-Brohman, I.; Jerome, S.; Annabhimoju, R.; Boyles, N. A.; Schmitz, C. D.; Abel, R.; Farid, R.; Friesner, R. A. WScore: A Flexible and Accurate Treatment of Explicit Water Molecules in Ligand−Receptor Docking. *J. Med. Chem.* **2016**, *59* (9), 4364−4384.
(42) Boike, L.; Henning, N. J.; Nomura, D. K. Advances in covalent drug discovery. *Nat. Rev. Drug Discovery* **2022**, *21* (12), 881−898.
(43) Schaefer, D.; Cheng, X. Recent Advances in Covalent Drug Discovery. *Pharmaceuticals* **2023**, *16* (5), 663 DOI: 10.3390/ph16050663.
(44) Lanman, B. A.; Allen, J. R.; Allen, J. G.; Amegadzie, A. K.; Ashton, K. S.; Booker, S. K.; Chen, J. J.; Chen, N.; Frohn, M. J.; Goodman, G.; et al. Discovery of a Covalent Inhibitor of KRASG12C (AMG 510) for the Treatment of Solid Tumors. *J. Med. Chem.* **2020**, *63* (1), 52−65.
(45) London, N.; Miller, R. M.; Krishnan, S.; Uchida, K.; Irwin, J. J.; Eidam, O.; Gibold, L.; Cimermančič, P.; Bonnet, R.; Shoichet, B. K.; Taunton, J. Covalent docking of large libraries for the discovery of chemical probes. *Nat. Chem. Biol.* **2014**, *10* (12), 1066−1072.
(46) Zhu, K.; Borrelli, K. W.; Greenwood, J. R.; Day, T.; Abel, R.; Farid, R. S.; Harder, E. Docking Covalent Inhibitors: A Parameter Free Approach To Pose Prediction and Scoring. *J. Chem. Inf. Model.* **2014**, *54* (7), 1932−1940.
(47) Toledo Warshaviak, D.; Golan, G.; Borrelli, K. W.; Zhu, K.; Kalid, O. Structure-Based Virtual Screening Approach for Discovery of Covalently Bound Ligands. *J. Chem. Inf. Model.* **2014**, *54* (7), 1941−1950.
(48) Bianco, G.; Forli, S.; Goodsell, D. S.; Olson, A. J. Covalent docking using autodock: Two-point attractor and flexible side chain methods. *Protein Sci.* **2016**, *25* (1), 295−301.
(49) Scarpino, A.; Ferenczy, G. G.; Keserű, G. M. Comparative Evaluation of Covalent Docking Tools. *J. Chem. Inf. Model.* **2018**, *58* (7), 1441−1458.
(50) Zhang, Z.; Guiley, K. Z.; Shokat, K. M. Chemical acylation of an acquired serine suppresses oncogenic signaling of K-Ras(G12S). *Nat. Chem. Biol.* **2022**, *18* (11), 1177−1183.
(51) Zheng, Q.; Zhang, Z.; Guiley, K. Z.; Shokat, K. M. Strain-release alkylation of Asp12 enables mutant selective targeting of K-Ras-G12D. *Nat. Chem. Biol.* **2024**, *20*, 1114.
(52) Zhang, Z.; Morstein, J.; Ecker, A. K.; Guiley, K. Z.; Shokat, K. M. Chemoselective Covalent Modification of K-Ras(G12R) with a Small

Molecule Electrophile. *J. Am. Chem. Soc.* **2022**, *144* (35), 15916−15921.
(53) Wan, X.; Yang, T.; Cuesta, A.; Pang, X.; Balius, T. E.; Irwin, J. J.; Shoichet, B. K.; Taunton, J. Discovery of Lysine-Targeted eIF4E Inhibitors through Covalent Docking. *J. Am. Chem. Soc.* **2020**, *142* (11), 4960−4964.
(54) Zheng, L.; Li, Y.; Wu, D.; Xiao, H.; Zheng, S.; Wang, G.; Sun, Q. Development of covalent inhibitors: Principle, design, and application in cancer. *MedComm-Oncology* **2023**, *2* (4), No. e56.
(55) Mehta, N. V.; Degani, M. S. The expanding repertoire of covalent warheads for drug discovery. *Drug Discovery Today* **2023**, *28* (12), No. 103799.
(56) Nnadi, C. I.; Jenkins, M. L.; Gentile, D. R.; Bateman, L. A.; Zaidman, D.; Balius, T. E.; Nomura, D. K.; Burke, J. E.; Shokat, K. M.; London, N. Novel K-Ras G12C Switch-II Covalent Binders Destabilize Ras and Accelerate Nucleotide Exchange. *J. Chem. Inf. Model.* **2018**, *58* (2), 464−471.
(57) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28* (1), 235−242.
(58) 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.
(59) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25* (13), 1605−1612. https://doi.org/10.1002/jcc.20084
(60) Word, J. M.; Lovell, S. C.; Richardson, J. S.; Richardson, D. C. Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation11Edited by J. Thornton. *J. Mol. Biol.* **1999**, *285* (4), 1735−1747.
(61) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **2002**, *23* (16), 1623−1641.
(62) *Amber 2022*; University of California: San Francisco, 2022. (accessed June 13, 2024).
(63) *AMSOL*; University of Minnesota: Minneapolis, 2004. (accessed June 13, 2024).
(64) ChemAxon. https://www.chemaxon.com. (accessed June 13, 2024).
(65) Gasteiger, J.; Rudolph, C.; Sadowski, J. Automatic generation of 3D-atomic coordinates for organic molecules. *Tetrahedron Comput. Methodol.* **1990**, *3* (6, Part C), 537−547.
(66) Bash, P. A.; Pattabiraman, N.; Huang, C.; Ferrin, T. E.; Langridge, R. Van der Waals Surfaces in Molecular Modeling: Implementation with Real-Time Computer Graphics. *Science* **1983**, *222* (4630), 1325−1327.
(67) DesJarlais, R. L.; Sheridan, R. P.; Seibel, G. L.; Dixon, J. S.; Kuntz, I. D.; Venkataraghavan, R. Using shape complementarity as an initial screen in designing ligands for a receptor binding site of known three-dimensional structure. *J. Med. Chem.* **1988**, *31* (4), 722−729.
(68) Meng, E. C.; Shoichet, B. K.; Kuntz, I. D. Automated docking with grid-based energy evaluation. *J. Comput. Chem.* **1992**, *13* (4), 505−524. DOI: https://doi.org/10.1002/jcc.540130412.
(69) Allen, W. J.; Balius, T. E.; Mukherjee, S.; Brozell, S. R.; Moustakas, D. T.; Lang, P. T.; Case, D. A.; Kuntz, I. D.; Rizzo, R. C. DOCK 6: Impact of new features and current docking performance. *J. Comput. Chem.* **2015**, *36* (15), 1132−1156. DOI: https://doi.org/10.1002/jcc.23905.
(70) Mysinger, M. M.; Shoichet, B. K. Rapid Context-Dependent Ligand Desolvation in Molecular Docking. *J. Chem. Inf. Model.* **2010**, *50* (9), 1561−1573.
(71) 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.
(72) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696−3713.
(73) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79* (2), 926−935.
(74) Li, P.; Song, L. F.; Merz, K. M., Jr. Parameterization of Highly Charged Metal Ions Using the 12−6-4 LJ-Type Nonbonded Model in Explicit Water. *J. Phys. Chem. B* **2015**, *119* (3), 883−895.
(75) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25* (9), 1157−1174.
(76) Roe, D. R.; Brooks, B. R. MPI-parallelization of the grid inhomogeneous solvation theory calculation. *J. Comput. Chem.* **2024**, *45* (10), 633−637.
(77) Ornstein, L. S. Accidental deviations of density and opalescence at the critical point of a single substance. *Proc. Akad. Sci.* **1914**, *17*, 793.
(78) Kovalenko, A.; Hirata, F. Potentials of mean force of simple ions in ambient aqueous solution. I. Three-dimensional reference interaction site model approach. *J. Chem. Phys.* **2000**, *112* (23), 10391−10402.
(79) Kast, S. M.; Kloss, T. Closed-form expressions of the chemical potential for integral equation closures with certain bridge functions. *J. Chem. Phys.* **2008**, *129* (23), No. 236101.
(80) Singer, S. J.; Chandler, D. Free energy functions in the extended RISM approximation. *Mol. Phys.* **1985**, *55* (3), 621−625.
(81) Morita, T.; Hiroike, K. A New Approach to the Theory of Classical Fluids. I. *Prog. Theor. Phys.* **1960**, *23* (6), 1003−1027.
(82) Luchko, T.; Gusarov, S.; Roe, D. R.; Simmerling, C.; Case, D. A.; Tuszynski, J.; Kovalenko, A. Three-Dimensional Molecular Theory of Solvation Coupled with Molecular Dynamics in Amber. *J. Chem. Theory Comput.* **2010**, *6* (3), 607−624.
(83) Kovalenko, A.; Hirata, F. Three-dimensional density profiles of water in contact with a solute of arbitrary shape: a RISM approach. *Chem. Phys. Lett.* **1998**, *290* (1), 237−244.
(84) Kovalenko, A.; Ten-no, S.; Hirata, F. Solution of three-dimensional reference interaction site model and hypernetted chain equations for simple point charge water by modified method of direct inversion in iterative subspace. *J. Comput. Chem.* **1999**, *20* (9), 928−936.
(85) Frigo, M.; Zorn, B. G.; Frigo, M.et al. A Fast Fourier Transform Compiler. In *Proceedings of the ACM SIGPLAN 1999 Cnference on Programming Language Design and Implementation* Atlanta, Georgia, USA, 1999.
(86) Frigo, M.A Fast Fourier Transform Compiler. In *Proceedings of the ACM SIGPLAN 1999 conference on Programming Language Design and Implementation* 1999; pp 169−180 .
(87) Nguyen, C.; Yamazaki, T.; Kovalenko, A.; Case, D. A.; Gilson, M. K.; Kurtzman, T.; Luchko, T. A molecular reconstruction approach to site-based 3D-RISM and comparison to GIST hydration thermodynamic maps in an enzyme active site. *PLoS One* **2019**, *14* (7), No. e0219473.
(88) Prentis, L. E. Chapter 5: Implementation and Validation of Covalent Attach and Grow Algorithm. In *Development and Application of Enhanced Search Algorithm for Computer-Aided Drug Design*; Stony Brook University, 2022; pp 129−147.
(89) RDKit: Open-source 2022 cheminformaticshttps://www.rdkit.org. (accessed June13, 2024).
(90) Balius, T. E.; Tan, Y. S.; Chakrabarti, M. DOCK 6: Incorporating hierarchical traversal through precomputed ligand conformations to enable large-scale docking. *J. Comput. Chem.* **2024**, *45* (1), 47−63.
(91) Ewing, T. J. A.; Makino, S.; Skillman, A. G.; Kuntz, I. D. DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases. *J. Comput.-Aided Mol. Des.* **2001**, *15* (5), 411−428.
(92) Lorber, D. M.; Shoichet, B. K. Flexible ligand docking using conformational ensembles. *Protein Sci.* **1998**, *7* (4), 938−950. https://doi.org/10.1002/pro.5560070411
(93) Mukherjee, S.; Balius, T. E.; Rizzo, R. C. Docking Validation Resources: Protein Family and Ligand Flexibility Experiments. *J. Chem. Inf. Model.* **2010**, *50* (11), 1986−2000.

(94) Allen, W. J.; Rizzo, R. C. Implementation of the Hungarian Algorithm to Account for Ligand Symmetry and Similarity in Structure-Based Design. *J. Chem. Inf. Model.* **2014**, *54* (2), 518−529.
(95) Irwin, J. J.; Tang, K. G.; Young, J.; Dandarchuluun, C.; Wong, B. R.; Khurelbaatar, M.; Moroz, Y. S.; Mayfield, J.; Sayle, R. A. ZINC20−A Free Ultralarge-Scale Chemical Database for Ligand Discovery. *J. Chem. Inf. Model.* **2020**, *60* (12), 6065−6073, DOI: 10.1021/acs.jcim.0c00675.
(96) Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* **2010**, *50* (5), 742−754.
(97) Efron, B.Bootstrap Methods: Another Look at the Jackknife. In *Breakthroughs in Statistics: Methodology and Distribution*, Springer Series in Statistics; Springer New York: New York, NY, 1992; Vol. 7, pp 569−593.
(98) Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **2020**, *17* (3), 261−272.
(99) Lu, S.; Jang, H.; Muratcioglu, S.; Gursoy, A.; Keskin, O.; Nussinov, R.; Zhang, J. Ras Conformational Ensembles, Allostery, and Signaling. *Chem. Rev.* **2016**, *116* (11), 6607−6665.
(100) Ferrari, A. M.; Wei, B. Q.; Costantino, L.; Shoichet, B. K. Soft Docking and Multiple Receptor Conformations in Virtual Screening. *J. Med. Chem.* **2004**, *47* (21), 5076−5084.
(101) https://enamine.net/compound-collections/covalent-compounds/acrylamides. (accessed Dec 11, 2023).
(102) Hartshorn, M. J.; Verdonk, M. L.; Chessari, G.; Brewerton, S. C.; Mooij, W. T. M.; Mortenson, P. N.; Murray, C. W. Diverse, High-Quality Test Set for the Validation of Protein−Ligand Docking Performance. *J. Med. Chem.* **2007**, *50* (4), 726−741.
(103) Flachsenberg, F.; Ehrt, C.; Gutermuth, T.; Rarey, M. Redocking the PDB. *J. Chem. Inf. Model.* **2024**, *64* (1), 219−237.
(104) Tran, T. H.; Alexander, P.; Dharmaiah, S.; Agamasu, C.; Nissley, D. V.; McCormick, F.; Esposito, D.; Simanshu, D. K.; Stephen, A. G.; Balius, T. E. The small molecule BI-2852 induces a nonfunctional dimer of KRAS. *Proc. Natl. Acad. Sci. U.S.A.* **2020**, *117* (7), 3363−3364.
(105) Chakrabarti, M.; Tan, Y. S.; Balius, T. E. Considerations Around Structure-Based Drug Discovery for KRAS Using DOCK. In *KRAS: Methods and Protocols*; Stephen, A. G.; Esposito, D., Eds.; Springer US, 2024; pp 67−90.
(106) Kessler, D.; Gollner, A.; Gmachl, M.; Mantoulidis, A.; Martin, L. J.; Zoephel, A.; Mayer, M.; Covini, D.; Fischer, S.; Gerstberger, T.; et al. Reply to Tran et al.: Dimeric KRAS protein−protein interaction stabilizers. *Proc. Natl. Acad. Sci. U.S.A.* **2020**, *117* (7), 3365−3367.
(107) Fischer, M.; Coleman, R. G.; Fraser, J. S.; Shoichet, B. K. Incorporation of protein flexibility and conformational energy penalties in docking screens to improve ligand discovery. *Nat. Chem.* **2014**, *6* (7), 575−583.
(108) Meiler, J.; Baker, D. ROSETTALIGAND: Protein−small molecule docking with full side-chain flexibility. *Proteins: Struct., Funct., Bioinf.* **2006**, *65* (3), 538−548.
(109) Abagyan, R.; Totrov, M.; Kuznetsov, D. ICM—A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation. *J. Comput. Chem.* **1994**, *15* (5), 488−506.

https://doi.org/10.1021/acs.jcim.4c01623
J. Chem. Inf. Model. 2025, 65, 722−748