

Reactive Docking: a computational method for high-throughput virtual screenings of reactive species

Giulia Bianco*, Matthew Holcomb, Diogo Santos-Martins, Andreas Tillack, Althea Hansel-Harris, Stefano Forli*

Department of Integrative Structural and Computational Biology, Scripps Research Institute,
10550 N. Torrey Pines, La Jolla, CA, USA

* corresponding authors: gbianco@scripps.edu, forli@scripps.edu

Abstract

We describe the formalization of the reactive docking protocol, a method developed to model and predict reactions between small molecules and biological macromolecules. The method has been successfully used in a number of applications already, including recapitulating large proteomics datasets, performing structure-reactivity target optimizations and prospective virtual screenings. By modeling a near-attack conformation-like state, no QM calculations are required to model ligand and receptor geometries. Here, we present its generalization using a large dataset containing more than 400 ligand-target complexes, 8 nucleophilic modifiable residue types, and more than 30 warheads. The method correctly predicts the modified residue in ~85% of complexes and shows enrichments comparable to standard focused virtual screenings in ranking ligands. This performance supports this approach for the docking and screening of reactive ligands in virtual chemoproteomics and drug design campaigns.

Main

Small molecules containing electrophilic reactive warheads have emerged as an important ligand class¹, especially in drug design and chemical proteomics. Covalent drugs have been employed against a variety of targets as well as disease classes and their rational design has been a topic of increasing study². Relative to conventional drugs, they can exhibit improved pharmacodynamics, higher potency, and improved selectivity. However, concerns have been raised due to their potential toxicity and promiscuous binding³.

Taking into account the issues that a reactive molecule could potentially cause, research in the past decade focused on merging the strengths of covalent and non-covalent binders (i.e.: high selectivity and high affinity) giving rise to the "targeted covalent inhibitors" (TCIs)¹. These molecules install low reactivity warheads (i.e.: the electrophile moiety of the ligand) on a molecular scaffold with high binding affinity with the target structure, resulting in less promiscuity^{4,5}. This increases the importance of having a structure or high-quality model on which to perform structure-based optimizations⁶.

Conversely, in chemical proteomics the promiscuity of electrophilic fragments is a feature that allows the identification of potentially modifiable residues across a proteome. Nonetheless, structure-based modeling can be critical here, for example in disambiguating which residue is labeled in a peptide, or which enantiomer of a racemic probe may be more active for further development. Modeling may also provide rational bases for the selectivity across probes for a given site, which can improve the SAR information extracted from the original proteomic experiment.

This interest in TCIs has been accompanied by a resurgence of interest in computational methods that can model such inhibitors, which come with unique modeling challenges⁷. Existing methods tend to focus on the evaluation of the end-point of the reaction, modeling the ligand in the bound state, such as the tethered^{8,9} or biased (or constrained) approaches^{10,11}. More computationally expensive methods supplement these models with free energy calculations^{6,12,13}.

These methods are suitable for the analysis of well-defined sets of molecules into well-characterized binding sites⁴ but are not ideal for binding sites prediction across nucleophilic residues in a protein, let alone proteomes. For that, an ideal method should be capable of predicting both the correct residue for modification and the optimal reactive ligand(s).

In this work, we present the formalization of the reactive docking method, which was specifically designed to address both challenges. This method uses a modified version of the AutoDock4 standard force field¹⁵ which modifies the near attack conformation¹⁶ (NAC, **Fig. 1a**), the last ground state geometry preceding the transition state geometries, with a decreased equilibrium distance for the reactive atoms. This species was chosen to model the ligand-target interaction because it represents an essential thermodynamic checkpoint for the reaction while containing minimal geometric perturbations to the ligand or target structures. With this model, the ligand is docked in the unmodified state prior to the reaction, with the reactive warhead still intact. Another advantage of modeling a near-attack conformation-like a NAC-like state is that, as a ground state model, no expensive QM calculations are required to determine the ligand and receptor geometries, minimizing the domain knowledge required to parameterize a new reaction class. Despite the geometric simplicity of this model, it serves as an essential step in the reaction pathway. The AutoDock standard force field estimates free energy of binding using the following equation for any pairwise interaction (i,j):

$$\begin{aligned}
 \Delta G_{(lig)} = & \Delta H_{vdW} && \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) \\
 & + \Delta H_{hbond} && \sum_{i,j} E(t) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) \\
 & + \Delta H_{elec} && \sum_{i,j} \left(\frac{q_i q_j}{\epsilon(r_{ij}) r_{ij}} \right) \\
 & + \Delta S_{tor} && N_{tor} \\
 & + \Delta G_{desolv} && \sum_{i,j} (S_i V_j + S_j V_i) e^{(-r_{ij}^2/2\sigma^2)}
 \end{aligned}$$

In brief, ΔH_{vdW} , ΔH_{hbond} , and ΔH_{elec} are van der Waals, directional hydrogen bond, and electrostatic enthalpies, ΔS_{tor} is the torsional entropy, and ΔG_{desolv} is the desolvation free energy (for a full description of all the terms see reference¹⁷). The standard force field is modified by adding a 13-7 potential ΔH_{react} between the ligand reactive atom i^* and the receptor reactive atom j^* to represent incipient bond formation:

$$\Delta H_{\text{react}} = \frac{E_{i^*j^*}}{r_{i^*j^*}^{13}} - \frac{F_{i^*j^*}}{r_{i^*j^*}^7}$$

where

$$E_{i^*j^*} = 7 \left(\frac{\varepsilon r_{eq}^{13}}{13 - 7} \right)$$

$$F_{i^*j^*} = 13 \left(\frac{\varepsilon r_{eq}^7}{13 - 7} \right)$$

r_{eq} represents the equilibrium distance (Å) and ε represents the equilibrium energy (kcal/mol). Like van der Waals and hydrogen bond potentials, the reactive potential is also softened according to the standard AutoDock force field description¹⁸. As a result of this newly created pseudointeraction, atoms within two bonds of the two reactive atoms will be placed at distances shorter than their respective van der Waals equilibrium distances (**Fig. 1a**). To compensate for the tight proximity that this geometry would induce, the r_{eq} distances for pseudo 1-3 and 1-4 pairwise interactions around reactive atoms are scaled by factors $W_{1,3}$ and $W_{1,4}$, respectively, while any hydrogens bound to these atoms are ignored entirely from the pairwise calculation. Any atom beyond two bonds from the reactive atom pair is treated with default parameters. During docking, ligands are free to explore multiple binding modes including those not compatible with the formation of a covalent bond with the target residue, which is modeled as flexible¹⁵. Once dockings are completed, the result with the best docking score is analyzed to measure the distance between ligand and residue reactive atoms, determining the outcome of the reaction (i.e.: covalent or not, **Fig. 1b**).

This method was initially developed to model acrylamides and chloroacetamides reacting with cysteine thiols¹⁹, then it was extended to model more warheads and residues and successfully applied to a variety of chemical proteomic tasks. In particular, it was used successfully to resolve ambiguity in reactive residues in ABPP experiments on human T cells^{20,21}, identifying the residue most likely to be modified based on structural data. The method has been extended and applied to other reactions, such as those involving SuFEx warheads²², and used in an “*inverse drug discovery*” campaign targeting lysines and tyrosines in the human proteome²³. In addition, it has been used in focused drug design campaigns to model the reaction between SuFEx and serine residues in specific targets^{24,25}. In this work, the method is optimized and validated on a diverse dataset of targets (**Fig. 2**) and warheads and reactions (**Fig. 3**).

Results

Calibration. The identification of optimal parameters for the reactive docking forcefield was done by performing an exhaustive parameter sweep for ϵ , $W_{1,3}$, and $W_{1,4}$ values, optimizing the docking performance on the 80 complexes in the training set. The calibration showed multiple sets of values could achieve comparable success rates, with the ϵ being the dominant component. In fact, the analysis of the results shows accuracy starts dropping significantly with ϵ values lower than 2.0 Kcal/mol, below which, scaling factors have limited to no effect. Therefore, the optimal values maximizing success rates and minimizing false positive rates were $\epsilon = 3.5$ kcal/mol, $W_{1,3} = 0.8$, $W_{1,4} = 0.4$. These parameters achieved an overall success rate across the training set of 95% in identifying the correct residues as the top scored result, and >95% in the top 3 results (**Fig. 4a**). Predictions of reactions for cysteines with acrylamides and chloroacetamides, as well as serines with boronic acids, showed excellent performance (95% top score, in all cases), while predictions of reactions between serines and β -lactams performed slightly worse (80%) (**Fig. 4a**).

Due to the higher relative abundance of solvent-accessible serines (693, approximate reactive hit ratio 1:17) over cysteines (186, approximate reactive hit ratio 1:4, **Figs. 2b-c**) in the proteins of the *training set*, we anticipated that predicting the former would be a more challenging task, but overall success rates remained high for both residues.

Testing of selected parameters. The performance of the optimal values obtained in calibration was tested on the 351 complexes in the test set, which is larger and significantly more diverse than the training set. This set includes all residue types and warheads in the training set and extends it with more well-represented warhead/residue combinations (e.g., α,β unsaturated carbonyls like cyano acrylamides, vinyl carbonyl, phosphates for cysteines) as well as residues and warheads for which it was not possible to obtain a representative number of complexes. While the overall performance in identifying the correct residue as the top result dropped from 95% of the training set to 86.6% (**Fig. 4b**) for the test set, all results showed a consistent performance of at least 75% success rate or better, except for carbonyl warheads targeting lysines (62%, **Fig. 4b**). The more marked drop for this residue is likely due to a combination of its high abundance in the proteins considered (14 lys/system, **Fig. 2c**) and their preferred localization in highly solvent-accessible regions on the surface of the protein. Both factors concurred in making docking predictions more challenging due to the potentially higher rate of false positives and the lack of well-defined pockets.

Virtual screening. The performance of the reactive docking was then validated against experimental results reported on a small covalent ligand discovery effort in a virtual screening (VS) setting. For that, a virtual screening set was built based on the experimental data published by Resnik et al.²⁶ which used a library of 993 reactive fragments functionalized with mild electrophiles including chloroacetamides ($n = 752$, 76%) and acrylamides ($n = 241$, 24%) **Fig. 4c**. The library was used to screen for binders of 10 cysteine-containing protein targets, using intact-protein MS and high-throughput crystallography to identify and characterize hits. The experimental characterization identified two classes of hits: strong hits (>50% labeling) and weak binders (<50% labeling) (**Fig. 4c**). The virtual screening set used here contains 3 of the 10 targets from the paper for which structures were available and potent and selective fragments were identified: two enzymes (deubiquitinase OTUB2; pyrophosphatase NUDT7), and K-

Ras^{G12C}. For the validation, the set of parameters obtained from the calibration step was used to dock the entire fragment library against the targets in a blind docking fashion, considering all solvent-accessible cysteines for each target and predicting both the most likely residue to be modified and the most likely ligands to react (Table S1).

For the first task, the method was able to detect the correct cysteines in all three targets, confirming its residue prediction capabilities. For the VS task, we then analyzed the docking results considering the top 0.5%, 1.0%, and 10% of the results ranked by docking score for the different warheads, as summarized in **Fig. 4d**. The method achieved true positive hit rates for top 0.5% of ranked results of 100% on NUDT7 chloroacetamide binders (50% strong and 50% weak binders) and 50% on OTUB2 chloroacetamide binders (25% strong and 25% weak binders). For KRAS, the true positive rate at the top 0.5% was only 25%. Acrylamide fragments were more difficult to predict correctly given the lower number of representatives of this warhead in the library, as well as the smaller number of strong binders. Combined success rates at identifying binders of both chemical classes in the top 0.5% of ranked results were 40% (20% strong and 20% weak binders) for OTUB2, 40% (weak binders) for NUDT7 and 40% (weak binders) for KRAS.

Discussion

Here, we present the reactive docking protocol, a predictive method for irreversible ligand binding events based on the analysis of a modified NAC, which combines descriptions of the ground state ligand and receptor with a bias toward the incipient bond formation. This relatively simple model is advantageous because it requires neither prior modification of the ligand structure (a non-trivial effort for large chemical collections) nor distinct parameters for different reaction classes while still describing the energetics of a key thermodynamic step in the covalent modification. Additionally, because of the large contribution of the ground state character to this model, the modified forcefield is still responsive to the structural features of the ligand-protein complex that dominates non-covalent interactions. The reactive parameters described here, covering the reactive potential and scalings on pseudo- 1-3 and 1-4 interactions across the incipient bond, were calibrated on an inverse docking task with a diverse set of warheads (β -lactams, boronic acids, chloroacetamides, and acrylamides), and different reaction classes (β -lactam addition, borylation, nucleophilic substitution, and Michael addition) targeting serine and cysteine residues. These parameters were then tested against a more diverse set containing both withheld examples from the training warhead classes, as well as out-of-domain reactions, differing in the warhead, mechanism, and labeled residue, performing excellently at predicting the labeled residue. This highlights how the relative simplicity of this model avoids overfitting and affords a generalizable forcefield that does not require extensive prior knowledge of either warhead or reaction mechanism.

This method was additionally validated on a virtual screening task on reactive fragments against reactive cysteines in three separate proteins. Here we demonstrate that in a fully predictive setting (i.e.: neither the site nor the ligand are known in advance) the method can provide an enrichment of binders (particularly strong binders) comparable to focused conventional virtual screenings. While the use of a fragment library for assessing the VS performance is intrinsically more challenging compared to drug-like ligands²⁷, the overall

results in simultaneous prediction of both residue and ligands suggest that this protocol is very suitable for prospective VS campaigns.

This method and the calibration protocol also serve as a baseline for future development. Force field parameters described here can be easily modified, allowing interested researchers to optimize the parameters for an individual target or warhead class if sufficient data is available. Similarly, with the availability of more diverse and structurally rich experimental datasets, we anticipate the performance of the model on ligand ranking can be improved significantly.

Collectively, the results presented here confirm and expand the successful applications previously reported, suggesting this method can be readily applied to a diverse range of docking tasks and warheads, requiring only the reassignment of atom types and inclusion of the modified forcefield parameters. To the best of our knowledge, there is no other computational method that is capable of addressing both predictive aspects of proteomics experiments, that is predicting both residue to be modified and ligands capable of doing so. The very limited computational overhead and the predictive power make this method ideally suited for devising large virtual screenings campaigns to screen the growing commercially available chemical collections of reactive molecules.

Moreover, by leveraging the structural data generated by AlphaFold^{28,29}, it makes possible proteome-wide labeling predictions, providing *in silico* support for large chemical proteomics applications.

Methods

Ligand preparation. Ligand structures identified from PDB were built in their native form restoring the active warheads using the primary literature citations, (e.g.: reconstruction of the β -lactam ring, addition of leaving groups, restoring unsaturated bonds, etc.). Initial 3D coordinates were generated using OpenBabel³⁰ modeling protonation states at pH 7.4, then minimized (MMFF94s; 300 steps Steepest Descent; 300 Conjugated Gradient). Partial charges, torsions, standard and reactive atom type parameters were assigned according to the AutoDock protocol³¹ using Meeko (<https://github.com/forlilab/Meeko.git>) with SMARTS patterns to define the warhead atoms and assign the reactive docking force field parameters (**Fig. 1a**). If present, macrocyclic structures were modeled as flexible¹⁵ by default during docking³².

Target preparation. Target structures were retrieved from the Protein Data Bank and hydrogens were added with Reduce³³. For oligomeric structures, dockings were performed only on the first chain. Partial charges, torsions, and standard and reactive atom type parameters were assigned using Meeko according to the AutoDock protocol³¹. For each site, a cubic docking box of 30 Å side (80 points in the AutoGrid parameter file¹⁵) was defined and centered on the C α of the target residue.

Reactive docking. By default, the residues to be evaluated by reactive docking are automatically calculated on each target structure using MSMS³⁴ to identify all solvent-accessible residues (default probe radius: 1.5 Å) of a given type (i.e.: cysteines) including buried cavities (**Fig. 1b**). Optionally, a user-defined list can be provided. Then, each ligand is docked against the individual residues to be evaluated. During docking, ligands are modeled in their unmodified form with reactive warheads in place using a conventional (untethered) docking method, while the side chain of the target residue is modeled as

flexible¹⁵. Dockings are performed using AutoDock-GPU¹⁸, generating 50 poses for each ligand using the default Lamarckian Genetic Algorithm (LGA) parameters³¹. A ligand and a residue are considered reacting if the distance between their reactive atoms in the lowest energy pose is ≤ 2.0 Å. Residues are then ranked by their likelihood of reacting based on the best energy of the ligand(s) predicted to react with it. Reactive docking parameters do not affect the length nor the complexity of the calculation, resulting in nearly identical docking times as conventional dockings.

Datasets. In order to calibrate and validate the reactive docking parameters, covalent complexes were collected from the PDB³⁵, containing a very diverse pool of residues (**Fig. 2**) and warheads (**Fig. 3**). Complexes were visually inspected to discard highly distorted or problematic structures selecting a dataset of 431 structurally diverse ligands and chemical reactions. The dataset was then subdivided into specialized sets, with no overlap between them. The list of all PDBs and warheads in the dataset is available in Supplementary Material.

Training and test sets. First, we selected a training set of 80 complexes using the most represented residues (cysteine, serine) for which the largest number of complexes with the most diverse ligands and protein families is available. The training set was built by randomly selecting 20 complexes representative of each of the four most abundant warheads (**Fig. 3a**): acrylamides and chloroacetamides for cysteine; boronic acids and β -lactams for serine. The training set was used to calibrate the reactive docking force field parameters.

The remaining set of complexes was used to build the test set of 351 complexes, which includes 8 residue types (cystine, serine, threonine, lysine, tyrosine, glutamic acid, histidine, and methionine, **Fig. 2a**) and 37 warheads (**Fig. 3a**). The test set was used to assess the performance and transferability of the parameters obtained with the calibration process on the training set.

Virtual screening set. A virtual screening set was built based on the experimental data published by Resnik et al.²⁶ which used a library of 993 reactive fragments functionalized with mild electrophiles including chloroacetamides ($n = 752$, 76%) and acrylamides ($n=241$, 24%) targeting cysteines. This virtual screening set contains three targets for which structure coordinates were available and potent and selective fragments were identified: two enzymes, deubiquitinase OTUB2³⁶ and the pyrophosphatase NUDT7³⁷, and K-Ras^{G12C}³⁸.

Parameter selection. The training set was used to calibrate the reactive docking parameters to obtain a single set of weights to be used to maximize the predictive accuracy. The r_{eq} value of 1.7 Å is fixed for all reactions and was chosen for being between the lengths of the shortest bond in our dataset (C-C bond, ~ 1.5 Å) and the longest (C-S bond, ~ 1.8 Å). After preliminary tests, values were obtained by performing an exhaustive search with respect to ϵ value (from 2.0 to 10.0 kcal/mol, 0.5 intervals), and the scaling factors $W_{1,3}$, and $W_{1,4}$ (0.0 to 0.9, 0.1 interval). During the calibration, each parameter set was tested with the entire training set by docking each ligand on its cognate target including all the solvent-accessible residues that would be modified by its warhead (i.e.: all solvent-accessible cysteines for acrylamides).

Virtual screening protocol. Firstly, we assessed the ability of the reactive docking to find the correct modified cysteine from the pool of all solvent-accessible cysteines in each structure. We calculated the average, median, and standard deviation of the docking

score of the whole library for each residue (Table S1). Then the average docking score of the whole library docking against each target calculation was used to rank the most likely cysteine to be alkylated by the fragments. True positive success rates for each target were then calculated by ranking ligands predicted to react with the target residue by their best docking score, and true hit rates of known binders were calculated for the top 0.5%, 1%, and 10% of the docking results.

Acknowledgments

Code availability. All the code is distributed under open source licenses and is available at <https://github.com/forlilab/Meeko> and <https://github.com/ccsb-scripps/AutoDock-GPU>

Acknowledgements

This work was supported by the National Institutes of Health grants R01GM069832 and U54AI150472.

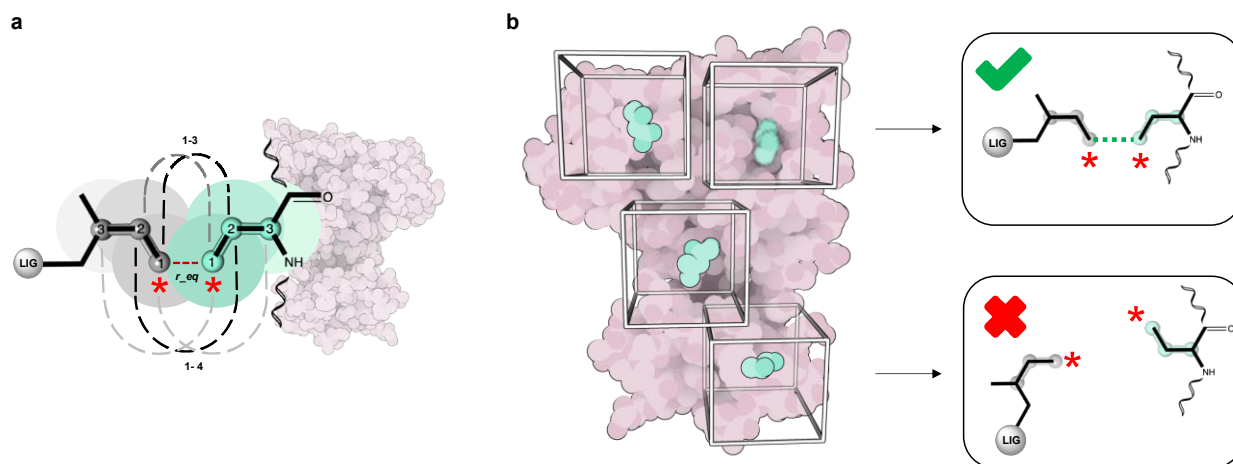


Figure 1. a) Schematic representation of the reactive docking model and the NAC-like state, showing the pseudo 1-3 and 1-4 interactions and their van der Waals radii. Reactive atoms are highlighted with *red* asterisks, pseudo 1-3 and 1-4 interactions are shown with dashed lines. b) Representation of the independent docking on selected solvent-exposed residue (cysteine, teal spheres) on the target protein and the possible outcomes of the simulations.

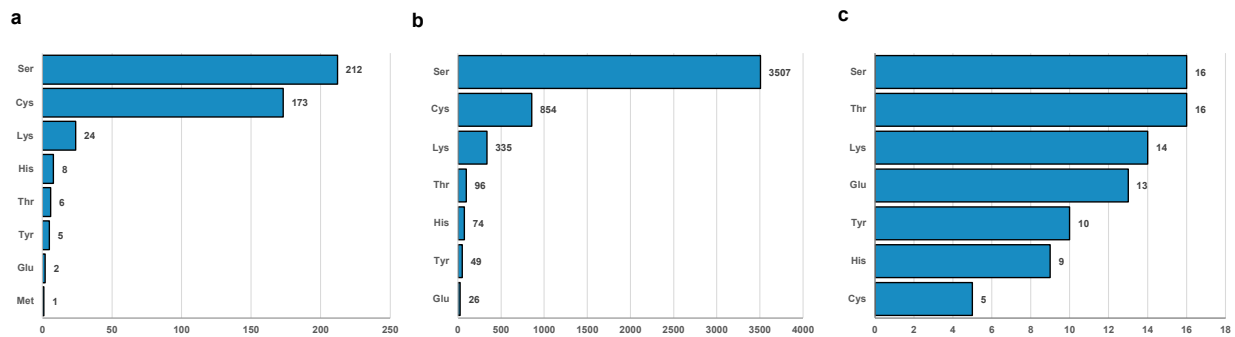


Figure 2. Dataset statistics. *a*) Count of complexes containing a given modified residue type; *b*) Total number of solvent-accessible residues by type. *c*) Average number of solvent-accessible residue types per complex.

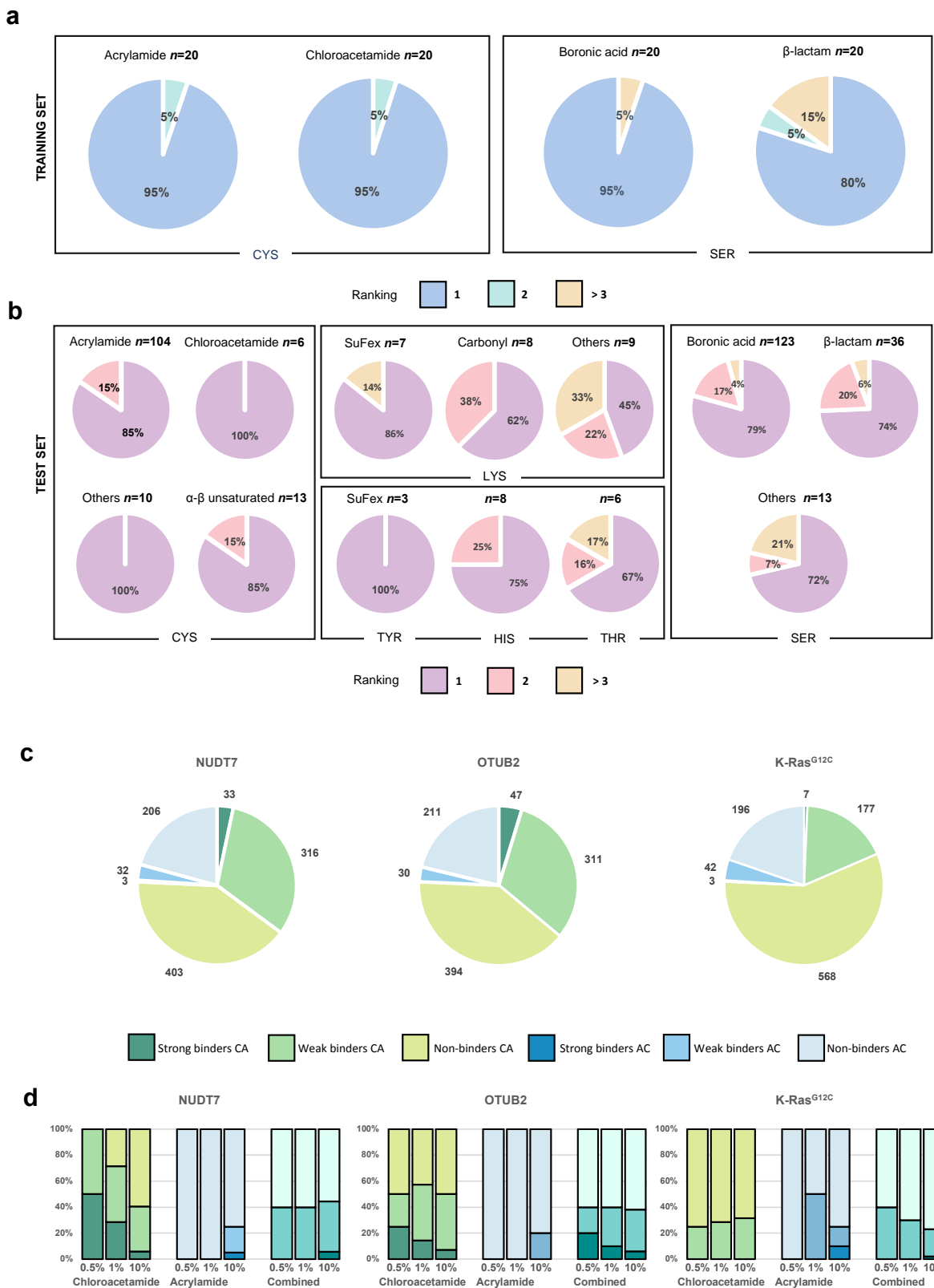


Figure 4. a) Training set docking performance in reactive residues predictions (1:best docking score; 2:second best-docking score; 3:third best or higher docking score). b) Test

set performance in reactive residues predictions (1:best docking score; 2:second best-docking score; 3:third best or higher docking score). c) Distribution of experimental binder strengths by warhead (CA: chloroacetamide; AC: acrylamide) from Resnick et al.[REF] for the three targets considered in the virtual screening. d) Virtual screening success rates in binders recovery (0.5%, 1.0% and 10% fractions).

References

1. Singh, J. The Ascension of Targeted Covalent Inhibitors. *J. Med. Chem.* **65**, 5886–5901 (2022).
2. Boike, L., Henning, N. J. & Nomura, D. K. Advances in covalent drug discovery. *Nat. Rev. Drug Discov.* **21**, 881–898 (2022).
3. Bandyopadhyay, A. & Gao, J. Targeting Biomolecules with Reversible Covalent Chemistry. *Curr. Opin. Chem. Biol.* **34**, 110–116 (2016).
4. Singh, J., Petter, R. C., Baillie, T. A. & Whitty, A. The resurgence of covalent drugs. *Nat. Rev. Drug Discov.* **10**, 307–317 (2011).
5. Gehringer, M. & Laufer, S. A. Emerging and Re-Emerging Warheads for Targeted Covalent Inhibitors: Applications in Medicinal Chemistry and Chemical Biology. *J. Med. Chem.* **62**, 5673–5724 (2019).
6. Lonsdale, R. & Ward, R. A. Structure-based design of targeted covalent inhibitors. *Chem. Soc. Rev.* **47**, 3816–3830 (2018).
7. Bianco, G., Goodsell, D. S. & Forli, S. Selective and Effective: Current Progress in Computational Structure-Based Drug Discovery of Targeted Covalent Inhibitors. *Trends Pharmacol. Sci.* **41**, 1038–1049 (2020).
8. Bianco, G., Forli, S., Goodsell, D. S. & Olson, A. J. Covalent docking using autodock: Two-point attractor and flexible side chain methods. *Protein Sci. Publ. Protein Soc.* **25**, 295–301 (2016).
9. Schröder, J. *et al.* Docking-Based Virtual Screening of Covalently Binding Ligands: An Orthogonal Lead Discovery Approach. *J. Med. Chem.* **56**, 1478–1490 (2013).

10. Ouyang, X. *et al.* CovalentDock: Automated covalent docking with parameterized covalent linkage energy estimation and molecular geometry constraints. *J. Comput. Chem.* **34**, 326–336 (2013).
11. Ai, Y., Yu, L., Tan, X., Chai, X. & Liu, S. Discovery of Covalent Ligands via Noncovalent Docking by Dissecting Covalent Docking Based on a ‘Steric-Clashes Alleviating Receptor (SCAR)’ Strategy. *J. Chem. Inf. Model.* **56**, 1563–1575 (2016).
12. Zhu, K. *et al.* Docking covalent inhibitors: a parameter free approach to pose prediction and scoring. *J. Chem. Inf. Model.* **54**, 1932–1940 (2014).
13. Friesner, R. A. *et al.* Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein–Ligand Complexes. *J. Med. Chem.* **49**, 6177–6196 (2006).
14. London, N. *et al.* Covalent docking of large libraries for the discovery of chemical probes. *Nat. Chem. Biol.* **10**, 1066–1072 (2014).
15. Morris, G. M. *et al.* AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J. Comput. Chem.* **30**, 2785–2791 (2009).
16. Hur, S. & Bruice, T. C. The near attack conformation approach to the study of the chorismate to prephenate reaction. *Proc. Natl. Acad. Sci.* **100**, 12015–12020 (2003).
17. Morris, G. M. *et al.* Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **19**, 1639–1662 (1998).
18. Santos-Martins, D. *et al.* Accelerating AutoDock4 with GPUs and Gradient-Based Local Search. *J. Chem. Theory Comput.* **17**, 1060–1073 (2021).
19. Backus, K. M. *et al.* Proteome-wide covalent ligand discovery in native biological systems. *Nature* **534**, 570–574 (2016).
20. Weerapana, E. *et al.* Quantitative reactivity profiling predicts functional cysteines in proteomes. *Nature* **468**, 790–795 (2010).
21. Vinogradova, E. V. *et al.* An Activity-Guided Map of Electrophile-Cysteine Interactions in Primary Human T Cells. *Cell* **182**, 1009-1026.e29 (2020).

22. Dong, J., Krasnova, L., Finn, M. G. & Sharpless, K. B. Sulfur(VI) Fluoride Exchange (SuFEx): Another Good Reaction for Click Chemistry. *Angew. Chem. Int. Ed.* **53**, 9430–9448 (2014).
23. Mortenson, D. E. *et al.* “Inverse Drug Discovery” Strategy To Identify Proteins That Are Targeted by Latent Electrophiles As Exemplified by Aryl Fluorosulfates. *J. Am. Chem. Soc.* **140**, 200–210 (2018).
24. Zheng, Q. *et al.* SuFEx-enabled, agnostic discovery of covalent inhibitors of human neutrophil elastase. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 18808–18814 (2019).
25. Cheng, Y. *et al.* Diversity oriented clicking delivers β -substituted alkenyl sulfonyl fluorides as covalent human neutrophil elastase inhibitors. *Proc. Natl. Acad. Sci.* **119**, e2208540119 (2022).
26. Resnick, E. *et al.* Rapid Covalent-Probe Discovery by Electrophile-Fragment Screening. *J. Am. Chem. Soc.* **141**, 8951–8968 (2019).
27. Hubbard, R. E., Chen, I. & Davis, B. Informatics and modeling challenges in fragment-based drug discovery. *Curr. Opin. Drug Discov. Devel.* **10**, 289–297 (2007).
28. Tunyasuvunakool, K. *et al.* Highly accurate protein structure prediction for the human proteome. *Nature* **596**, 590–596 (2021).
29. Holcomb, M., Chang, Y.-T., Goodsell, D. S. & Forli, S. Evaluation of AlphaFold2 structures as docking targets. *Protein Sci.* **32**, e4530 (2023).
30. O’Boyle, N. M. *et al.* Open Babel: An open chemical toolbox. *J. Cheminformatics* **3**, 33 (2011).
31. Forli, S. *et al.* Computational protein–ligand docking and virtual drug screening with the AutoDock suite. *Nat. Protoc.* **11**, 905–919 (2016).
32. Santos-Martins, D. *et al.* D3R Grand Challenge 4: prospective pose prediction of BACE1 ligands with AutoDock-GPU. *J. Comput. Aided Mol. Des.* **33**, 1071–1081 (2019).
33. 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 orientation. *J. Mol. Biol.* **285**, 1735–1747 (1999).

34. Sanner, M. F., Olson, A. J. & Spehner, J.-C. Reduced surface: An efficient way to compute molecular surfaces. *Biopolymers* **38**, 305–320 (1996).
35. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
36. Borodovsky, A. *et al.* Chemistry-based functional proteomics reveals novel members of the deubiquitinating enzyme family. *Chem. Biol.* **9**, 1149–1159 (2002).
37. Reilly, S.-J., Tillander, V., Ofman, R., Alexson, S. E. H. & Hunt, M. C. The nudix hydrolase 7 is an Acyl-CoA diphosphatase involved in regulating peroxisomal coenzyme A homeostasis. *J. Biochem. (Tokyo)* **144**, 655–663 (2008).
38. Janes, M. R. *et al.* Targeting KRAS Mutant Cancers with a Covalent G12C-Specific Inhibitor. *Cell* **172**, 578-589.e17 (2018).