State-of-the-art covalent virtual screening with AlphaFold3

Yoav Shamir[1], Nir London[1,#]

[1] Department of Chemical and Structural Biology, The Weizmann Institute of Science, Rehovot 7610001, Israel

[#] For correspondence: nir.london@weizmann.ac.il

Abstract

Recent years have seen an explosion in the prominence of covalent inhibitors as research and therapeutic tools. However, a lag in application of computational methods for covalent docking slows progress in this field. AI models such as AlphaFold3 have shown accuracy in ligand pose prediction but were never assessed for virtual screening. We show that AlphaFold3 reaches near-perfect classification (average AUC=98.3%) of covalent active binders over property-matched decoys, dramatically outperforming classical covalent docking tools. We identify a predicted metric that allows to reliably assign a probability of binding and demonstrate it also improves non-covalent virtual screening.

Covalent acting small molecules represent new opportunities in chemical biology and drug discovery (*1–4*). Such ligands typically display increased potency, selectivity and prolonged target engagement compared with their non-covalent counterparts. Advances in chemical biology have mitigated early concerns regarding off-target reactivity and promiscuous irreversible binding and enabled rational and successful design processes (*2*). This has led to the discovery of drugs for decades-long challenging targets such as K-Ras (*5*). Currently more than 50 covalent drugs have been approved by the FDA (*4*). Recent examples include Nirmatrelvir for COVID-19 (*6*), Ritlectinib for alopecia (*7*), and Adagrasib for cancer (*8*), and many others are undergoing Phase III clinical trials.

Several computational tools were developed for virtual screening of covalent libraries (*9–16*). Given a protein structure or model as input, these docking tools employ physics-based scoring functions to predict and energetically score the protein-bound pose of each ligand. Multiple covalent binders targeting a range of proteins have been developed and experimentally validated based on the results of such *in silico* screening (*9*, *17–20*). Several covalent datasets of experimental structures have been curated, enabling the evaluation of binding-pose prediction accuracy of docking tools (*21–25*). However, for virtual screening and practical applications, the ability of docking tools to rank a compound library such that active compounds are at the top is arguably more important than accurate pose recapitulation. While datasets designed to evaluate such enrichment are prevalent for non-covalent docking(*26–30*), we are unaware of any for the covalent domain.

AlphaFold3 (AF3), the latest AI-based all-atom structure prediction model from Google DeepMind, was released for free use in November 2024 (*31*). AF3 facilitates high-accuracy prediction of biomolecular complexes, including the prediction of covalent protein-ligand complexes. To probe the enrichment performance of AF3 in the covalent domain, we constructed COValid, a first of its kind test set for the enrichment analysis of covalent virtual screening. We show that ranking the AF3-predicted compounds using a physics-based scoring function significantly outperforms the ranking produced by classical covalent docking tools. Strikingly, we find an AF3 predicted confidence metric

that significantly outperforms all physics-based methods tested across all protein targets in COValid, yielding exceptionally high success rates in nominating active compounds.

Results

COValid - a novel benchmark for covalent virtual screening

Because cysteines are the most commonly targeted nucleophile by covalent binders, we focused our analysis on the most common electrophile that targets cysteines – acrylamide (Fig. S1) (*32*). We curated compounds with experimental activity annotations from two databases – BindingDB (*33*) and ChEMBL (*34*). The majority of targets we identified were protein kinases, likely due to the intense interest in covalent kinase inhibitors (*32*, *35*, *36*), as they provide superior selectivity and potency for this wide-spread protein family. We included the eight most populated kinases in COValid (Fig. 1A) as well as K-Ras^{G12C}, a prime target for covalent inhibitor design (*5*, *37*). For each target/ligand pair we also annotated the target cysteine position. We made sure that for each target the benchmark included a structure with both a covalent and noncovalent binder (selected from the PDB; Dataset S1).

Following the curation of active ligands, we generated 50 property-matched, acrylamide-bearing decoys for each ligand (see Supplementary Methods), to represent chemically feasible but likely inactive compounds. This computational design of decoys is necessary since annotations of ‘true’ inactive compounds are scarce. The decoy generation protocol (Fig. 1B) followed two design principles borrowed from the popular DUD-E benchmark set for non-covalent enrichment analysis (*27*). First - the physicochemical properties of the designed decoys should match those of the active. This mitigates the risk that different property distributions for the actives and decoys yield artificially high enrichment owing to property mismatch (Fig. 1C, Dataset S2) (*38*). The second principle - the chemical topology of the decoys should be dissimilar to that of the actives. We therefore scrambled the connectivity of the compounds to generate decoys that are compositionally matched to actives but unlikely to conserve favorable interactions with the target. Finally, the decoy scaffolds are selected from ZINC20 (*39*) representing commercially available compounds to circumvent the risk of generating non-realistic

molecules. In total, COValid includes 1,845 active protomers and 92,250 decoy protomers, across ten cysteine attachment sites from nine protein targets (Datasets S3,4).

A

Nucleophilic attachment site	Active protomers	Decoy protomers
BMX C496	58	2900
BTK C481	505	25250
EGFR C797	237	11850
FGFR1 C488	93	4650
FGFR4 C477	40	2000
FGFR4 C552	82	4100
ITK C442	176	8800
JAK3 C909	438	21900
KRAS G12C C12	198	9900
MAP3K7 C174	18	900
Total	**1845**	**92250**

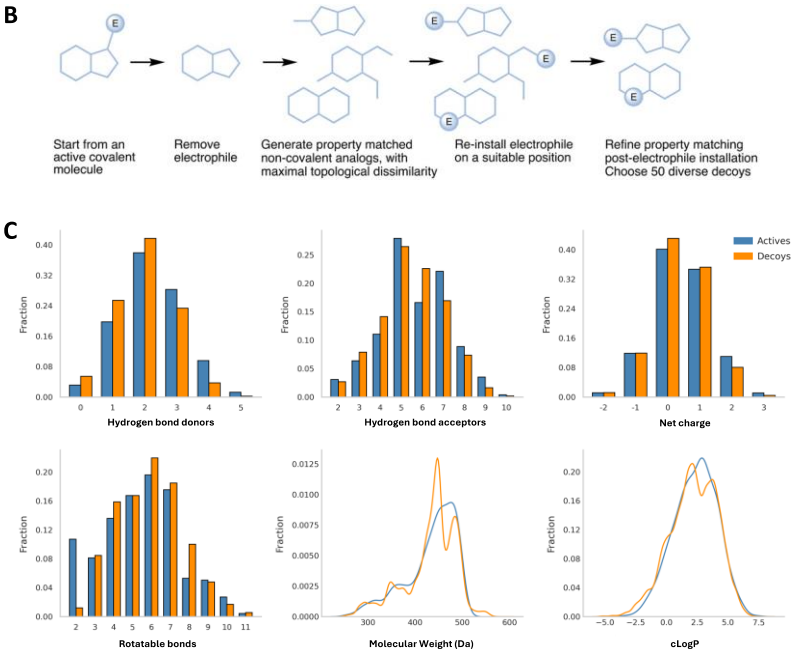

Figure 1. Constructing a balanced benchmark for covalent virtual screening. **A.** Number of active and decoy protomers per Cys attachment site. **B.** Decoy design scheme - the electrophile (E) is removed from the covalent 'active', next, property-matching, topologically dissimilar non-covalent decoys are collected for all actives. Finally, 50 decoy protomers are matched to each active protomer by attaching an acrylamide to free amines of the non-covalent decoys and ensuring property-matching to the active. **C.** Property distributions of the active and decoy protomers (colored blue and orange, respectively) for the six properties matched during decoy design.

COValid enables comparison and optimization of physics-based covalent docking tools

We evaluated three physics-based covalent docking tools using COValid: the DOCKovalent method (*9*) based on DOCK3.7, the attach-and-grow method based on DOCK6.12 (*15*), and the flexible sidechain method of AutoDock (*10*). To quantify the enrichment performance, we used the adjusted LogAUC metric (*40*), which emphasizes early enrichment in ranking hit lists (*41*). In a real-life scenario it is more important to have active molecules at the top fraction of the hit list than overall better ranking across the entire library. An adjusted LogAUC of 0% equals random performance, whereas 85.5% is perfect performance (*42*).

Each docking tool offers many adjustable parameters and configurations that can be non-trivial to optimize and can result in significant increases in run-time. Since evaluating all different combinations of parameters is an intractable combinatorial problem, we chose several representative features for each software and assessed their effect on accuracy and runtime (See Fig. 2A for one example, and the Supplementary Results for the full analysis).

The performance of leading configurations of DOCK6 and DOCKovalent were similar (Fig. 2B; Table 1; see docking parameters in Table S1). AutoDock, with a similar run-time showed an average enrichment twofold lower than DOCKovalent and DOCK6. For the three docking tools, the differences in the enrichment across the covalent-bound structures, or the non-covalent structures, were insignificant (*p-values* > *0.1*), indicating that non-covalent *holo* structures could be useful for covalent virtual screening when a covalent complex structure is not available.

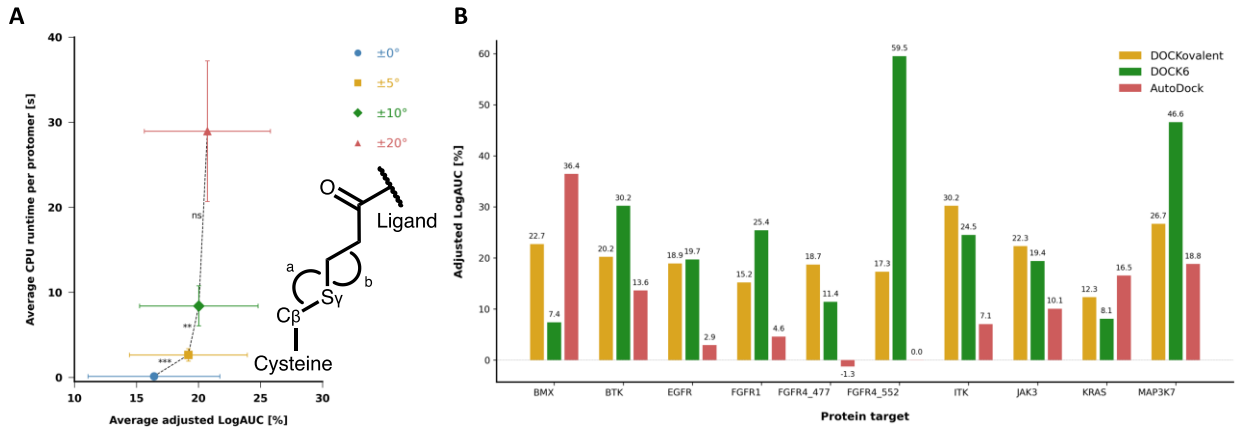

Figure 2. optimizing covalent docking algorithms based on performance in COValid. A. Example of using COValid for parameter optimization. The angle range parameter of DOCovalent defines a range Δ around a user-defined angle (109.5°), such that two bonds angles (angles a and b, defined by Cysteine Cβ, Sγ, and the acrylamide Cβ atoms, and by the Cysteine Sγ, acrylamide Cβ and Cα atoms) are sampled ±Δ during docking. Sampling with ±5º and further with ±10º leads to higher averaged adjusted LogAUC with statistical significance (*p-values* are 0.0002 and 0.001, respectively), but further increasing the exhaustiveness to Δ=20° yields no significant improvement (*p-value*: 0.1294), while yielding significantly longer runtime. **B.** Comparison of enrichment for the best identified configuration of the three covalent docking tools, docked to the ten covalent PDB structures.

Table 1. Performance over the COValid set

	DOCKovalent[a]	DOCK6	AutoDock	AF3 (Rosetta)	AF3 (mPEA)
Average Adj. LocAUC (%)	20.0±4.8	21.1±10.0	11.1±8.8	28.1±15.4	75.35±7.8
Average run-time per ligand (Sec.)	8.4±2.4	7.0±2.0	6.7±0.4	266±114[b]	

[a] The exact docking configurations corresponding to these runs are in Table S1.
[b] Run-time does not include Rosetta rescoring – only AF3 modeling per ligand, run-time varied based on GPU used. The initial step (multiple sequence alignment generation and curation of template structures, using 20 allocated CPUs) took on average 22±8 min. per protein.

Physics-based scoring of AlphaFold3 models outperforms docking tools for multiple targets

AF3 is the latest AI model from DeepMind for all-atom prediction of biomolecular structures (*31*). It enables prediction of non-covalent as well as covalent protein-ligand complexes. On a pose-prediction benchmark set of covalent complexes curated by the AF3 authors, the model achieved a success rate (defined as RMSD<2Å across the pocket atoms) of 78.5% (*31*). To our knowledge AF3 performance in the context of covalent virtual screening (or in fact for virtual screening in general) has not been reported. We set out to use COValid to assess its capacity for screening. We used the prediction pipeline of AF3 to predict covalent complexes for all compounds in COValid using the protein sequence and geometrically optimized 3D conformer as input. The input also includes specification of the ligand atom and the residue atom to be covalently bonded.

We sought to rank the predicted complex structures using a physics-based docking score. We used Rosetta (*43*) to minimize and energetically score all the predicted AF3 models and used this score to calculate the enrichment performance across COValid. Re-scoring of AF3 models yielded significantly better results for five out of the 10 sites (Fig. S2) and an overall better average adjusted LogAUC across the ten sites (Table 1). The fact this enrichment was achieved just based on the models produced by AF3 (and not by any AI-based metric) suggests the models are sufficiently realistic to differentiate true binding modes from artificial ones. Indeed, in the few cases where a crystal structure was available for a target-ligand pair from the benchmark, the AF3 models proved accurate (SI Fig. S3). While most of these examples were likely included in the AF3 training set, one structure was determined after the training, and in this case as well AF3 predicted it accurately (0.45Å pocket-aligned RMSD; PDB:7O70, Fig. 3A).

AF3 predicted alignment error is a superior metric for virtual screening enrichment

Along with the 3D coordinates of the predicted molecular structure, the AF3 model outputs predicted confidence metrics for the accuracy of the pose reproduction. We were interested in the ability of such metrics to rank compound libraries and enrich actives, though these metrics were not trained to predict activity or affinity. We evaluated the

adjusted LogAUC values over the COValid set for various AF3-predicted confidence metrics (See definitions in Supplementary Methods and results in Fig. S4). When attempting to rank COValid AF3 models by these metrics, we noticed that most metrics lack the resolution required for a decisive ranking of the lists. In several cases, thousands of protomers were ranked with the same exact predicted value (Fig. S5). Nevertheless, to probe their enrichment performance, we evaluated the adjusted LogAUC value for each cysteine site as a range spanning two extremes: We assessed a best-case scenario, where for any given value we rank actives higher than decoys for the same metric value, and a worst-case scenario, where decoys are ranked higher than actives of the same value.

Ranking by a global metric (pTM) or protein-based metric (protein chain pTM) yielded large ranges between the best- and worst-case enrichment values (Fig. S4). This lack of resolution is perhaps in line with the fact that the protein is constant for any specific target. Taking into account the protein-ligand interface (ipTM and 'ranking score'), yielded smaller, more informative enrichment ranges for some targets, while other targets still displayed a wide range of enrichment values. The metric that yielded the most decisive adjusted LogAUC ranges of this group was the ligand chain pTM. Even when using the worst-case ranking, AF3 predictions sorted by the ligand chain pTM metrics significantly outperformed the physics-based docking tools across nine of the ten COValid Cys sites (Fig. S6).

Finally, we studied the relevance of the predicted aligned error matrix (PAE), a structural confidence metric provided by AF3. In the internal representation used by the AF3 model, each protein residue and each ligand atom are represented by a single token. Each (*i,j*) element of the PAE matrix contains a prediction of the error in the positions of the AF3 token *j* when aligned to the ground-truth based on token *i*. (*31*). We define mPAE as the minimal predicted error in the positions of the protein residues, when aligned to the ground-truth complex with respect to the ligand (See details in Supplementary Methods; this value is provided as output by AF3).

Compared with other AF3-predicted confidence metrics, mPAE yields much better resolution between compounds, enabling conclusive ranking of all ten COValid compound lists. Remarkably, AF3 modeling followed by ranking by mPAE, significantly outperforms other metrics including all covalent docking tools across all ten cases (Fig. 3B; Fig. S4). The worst-case adjusted LogAUC values ranged from 58.6% for BMX to 85.1% for Cys477 of FGFR4, approaching perfect classification. Comparing the enrichment performance to that of the physics-based covalent docking tools, the differences are striking (Fig. 3B). If we disregard early enrichment and report the more typical ‘AUC’ (area under the receiver operating curve), the values range from 0.9690 to 0.9998 (SI Table 2).

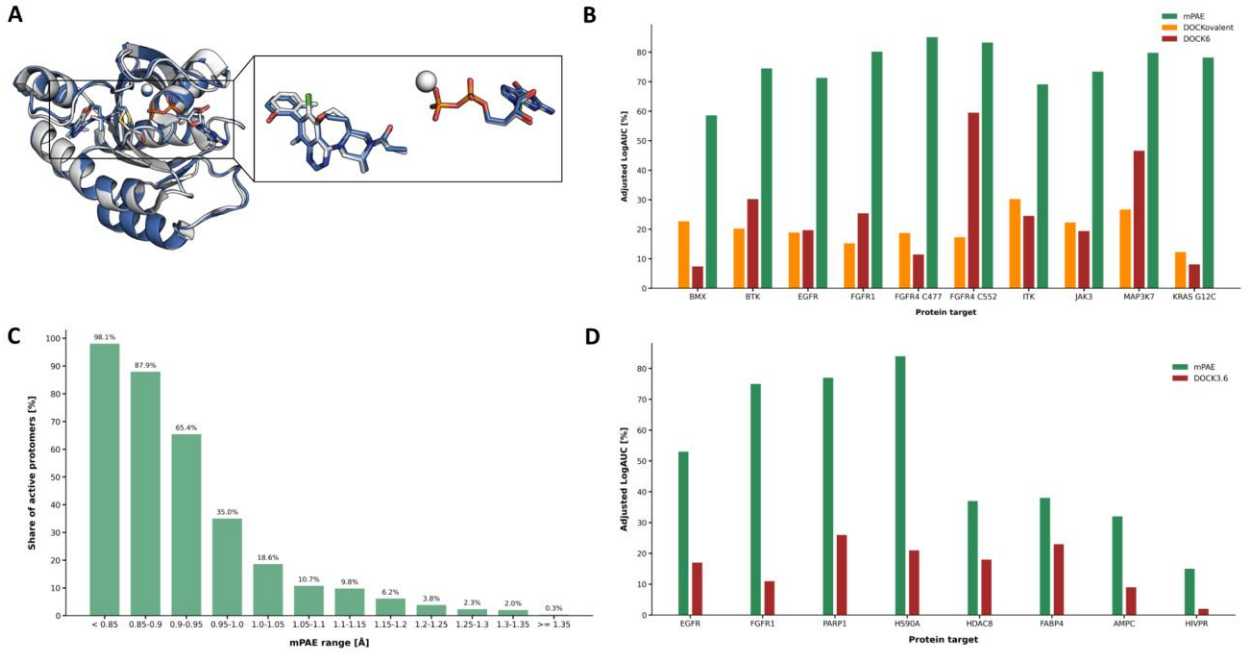

Figure 3. Near perfect enrichment of true inhibitors using AF3 mPAE A. Example of an accurate AF3 prediction of a covalent complex (KRAS G12C with Mg^{2+} and GDP, PDB 7O70, pocket-aligned RMSD 0.45Å) not included in its training set (the experimental structure and the predicted model are colored in gray and blue, respectively). **B.** Comparison of the two best physics-based covalent docking enrichment to AF3 structural predictions ranked by the mPAE element. **C.** The distribution of the proportion of active COValid protomers, according to their AF3 mPAE element. Each column indicates the share of active protomers out of all protomers within the indicated mPAE value range. This could be interpreted as the probability for being 'active' given a mPAE value. **D.** AF3 mPAE enrichment performance on DUDE-Z targets. Across a subset of eight DUDE-Z targets, we compare the adjusted LogAUC values reported in the DUDE-Z paper (using an optimized scoring function), and the values calculated from our AF3 predictions, ranked by mPAE (actives are sorted prior to enrichment calculations, using only the best-scoring active protomer of each compound). Protein sequence input for the data step of AF3 was extracted from the DUDE-Z PDB files, and the non-covalent ligands were provided as SMILES strings from DUDE-Z.

We also studied the relevance of the absolute mPAE values for prospective covalent screening. Within the applicability domain of COValid, mPAE enables assignment of confidence to the classification of a compound as active (Fig. 3C). Across all COVAlid protomers, 98.1% of the protomers with the lowest mPAE values (<0.85Å) correspond to actives. As mPAE values increase (indicating a decrease in structural confidence) the probability that the protomer corresponds to an active binder drops monotonically, decreasing to less than 10% at mPAE values higher than 1.05Å.

We considered various confounding factors, such as the possibility of overlap between the AF3 training set and the COValid compounds. If the model was trained on some COValid 'actives', it could potentially predict higher structural confidence for these compared to the designed decoys. We conducted ablation studies by removing active compounds with high topological similarity to any compound in the PDB (along with their corresponding decoys; Fig. S7) and re-calculated the adjusted LogAUC values in the absence of the removed protomers (Fig. S8). Even at strict similarity cutoffs that remove most of the active and decoy protomers from the ten ranked lists, enrichment remains high, with little effect due to ablation. For the two targets that exhibit the sharpest drop in enrichment for the strictest cutoff used (Tanimoto coefficient of 0.4), EGFR and ITK (45.2% and 47.7%, respectively), enrichment is still significantly higher than that of the leading docking tools. This suggests that the performance of mPAE is not due to overlap with the training data. In addition, analysis of the mPAE element of all the COValid 'actives' against their maximal Tanimoto similarity to any PDB yielded no correlation between the two metrics (Pearson correlation coefficient -0.047, *p-value* 0.29; Fig. S9).

Another potential confounding factor is that our decoys, despite matching in physicochemical properties and being based on commercially available scaffolds, differ in some rudimentary way from the 'actives'. To address this, we searched ChEMBL 33 (*34*) for experimental decoys - acrylamide-bearing structures that were tested against benchmark targets but showed worse than 10 μM activity (Fig. S10). We found 64 experimental decoys across six COValid kinases. We used AF3 to predict their covalent complexes and evaluated their mPAE values. The average mPAE was 1.5±0.5Å, with 85% of the decoys yielding a mPAE higher than 0.95Å. We found, however, that there is

no apparent correlation between mPAE and affinity (Fig. S11) suggesting it is useful primarily as a classification metric.

Finally, we tested the relevance of AF3 predictions with mPAE ranking for the non-covalent case. We selected eight diverse protein targets from the DUDE-Z benchmark set (*28*), and calculated enrichment with the non-covalent compound libraries. Across all eight targets, the AF3-based enrichment significantly outperformed the results reported for DUDE-Z with an optimized scoring function (Fig. 3D). The improvement in enrichment ranged from 13% to 64%. Across this diverse set, we observe a larger range of enrichment values than in COValid, with a minimal value of 15% and a maximal value of 84% (for HIV protease and HSP90A, respectively), suggesting some targets are more challenging than other.

Discussion

Virtual screening of large ligand libraries is a cornerstone of modern drug discovery(*44*). Despite the growing prominence of covalent drug discovery, prospective covalent virtual screening remains underdeveloped. We sought to test whether AF3 can address this gap. Inspired by the importance of benchmarks of noncovalent molecules for the development of virtual screening in that domain(*27*) we developed COValid. Our key finding was that AF3 vastly outperformed traditional methods. This is especially encouraging as it only requires the sequence of the protein and the target ligand rather than a crystallographic structure of the protein as do conventional virtual screening approaches.

Previous studies have shown that AI models trained on ligand binding data could not surpass AutoDock Vina on the DUD-E enrichment test set (49). Moreover, an analysis conducted prior to the development of AF3 revealed that deep learning–based methods lagged classical docking tools, particularly in generating physically valid poses (50). Most such analyses focused, however, on pose reproduction(*45*). In terms of screening, AI models were only evaluated for their capacity to generate structural models to support virtual screens using traditional docking tools(*46–48*). In contrast, we show for the first time that AF3 outperformed classical docking on both COValid and eight targets from DUDE-Z in a screening scenario (Fig. 3).

Several factors may contribute to the exceptional performance. First, specifying the covalent attachment point may improve accuracy; prior work demonstrated that indicating the ligand binding pocket increased pose recapitulation accuracy from 81% to 93% in AF3(*31*). Second, by incorporating backbone flexibility, AF3 can resolve minor clashes that typically confound classical fixed-backbone docking methods or expose cryptic pockets.

A more cautious explanation, however, is that the structural data used to train AF3 might overlap with our test set. Kinases—common drug targets that share similar features—are heavily represented in both the PDB and in drug discovery generally(*49*, *50*). Our dataset is heavily biased toward kinases (nine out of ten targets), highlighting

the potential for data leakage. Yet, this bias also underscores the importance of kinases, given that only this class of proteins provided a critical mass of annotated covalent binders. It is possible that the impressive performance we reported is limited to kinases; even so, the impact on accelerating ligand discovery for this crucial class would still be immense. Encouragingly, two observations suggest broader generalization. First, AF3 achieved high enrichment against KRas—a GTPase unrelated to kinases—even when considering only actives dissimilar to any PDB ligand (Fig. S8). Second, AF3 demonstrated strong non-covalent enrichment for six non-kinase targets (Fig. 3D): while EGFR and FGFR1 performed exceptionally well, HSP90 and PARP1 also achieved near-perfect ranking, and four additional targets, despite lower enrichment, still outperformed classical docking.

AF3 confidence metrics were trained to quantify modeling accuracy. Most predicted metrics lack the resolution to rank large libraries of models (Fig. S4), particularly those focusing on the protein side. A recent adversarial experiment suggested that AF3 does not rely on specific physical interactions for ligand placement (51), and indeed, there is no apparent correlation between the mPAE score and the binding affinity of experimental actives (Fig. S11). Nevertheless, the fact that rescoring AF3 models with Rosetta—an established physics-based scoring method—significantly outperformed covalent docking tools in terms of enrichment for five of the ten sites suggests that the models generated by AF3 are physically sensible. In that respect, while AF3's covalent docking tends to produce malformed poses with respect to the bond length and angle geometry (Fig. S12), it compensates so that the overall ligand pose remains accurate. An advantage of mPAE is its apparent insensitivity to the protein target, allowing comparisons across complexes and enabling the assignment of binding probabilities based on mPAE values (Fig. 3C).

In practical terms, despite AF3's superior enrichment, its long run-time may render it impractical for screening the extremely large virtual libraries that have become standard (*51–53*). A feasible solution is to use a fast classical docking method to screen very large virtual libraries—potentially comprising billions of compounds—followed by re-docking the top million or so predicted compounds with AF3 to more effectively enrich true binders.

If AF3's enrichment would be maintained in prospective virtual screening it would open the door to large scale covalent ligand discovery against novel classes of targets for which structures have not yet been determined, vastly accelerating chemical biology and drug discovery. The surprisingly good performance also suggests AF3 could be utilized to bear on the open question of generative covalent ligand design.

Acknowledgements

We thank Sarel Fleishman for critical reading of the manuscript. The Irwin lab in UCSF for access to their cluster for DOCK6 and DOCKovalent ligand generation and in particular Dr. Khanh Tang for technical assistance. We thank Dr. Trent Balius for assistance with DOCK6.12, and Dr. Alexey Orlov for sharing his ligand generation pipeline. Y.S. is funded by the CHE fellowship for data sciences. This research was generously supported by the Institute for Artificial Intelligence, the Honey and Dr. Barry Sherman Lab, the Dr. Barry Sherman Institute for Medicinal Chemistry, the Abisch-Frenkel RNA Therapeutics Center, the Moross Integrated Cancer Center, the Goldhirsh-Yellin Foundation and Celia Zwillenberg-Fridman.

References

1. T. A. Baillie, Targeted covalent inhibitors for drug design. *Angew. Chem. Int. Ed Engl.* **55**, 13408–13421 (2016).

2. L. Boike, N. J. Henning, D. K. Nomura, Advances in covalent drug discovery. *Nat. Rev. Drug Discov.* **21**, 881–898 (2022).

3. N. London, Covalent proximity inducers. *Chem. Rev.* **125**, 326–368 (2025).

4. F. Sutanto, M. Konstantinidou, A. Dömling, Covalent inhibitors: a rational approach to drug discovery. *RSC Med. Chem.* **11**, 876–884 (2020).

5. J. M. L. Ostrem, K. M. Shokat, Targeting KRAS G12C with covalent inhibitors. *Annu. Rev. Cancer Biol.* **6**, 49–64 (2022).

6. J. Hammond, H. Leister-Tebbe, A. Gardner, P. Abreu, W. Bao, W. Wisemandle, M. Baniecki, V. M. Hendrick, B. Damle, A. Simón-Campos, R. Pypstra, J. M. Rusnak, EPIC-HR Investigators, Oral nirmatrelvir for high-risk, nonhospitalized adults with Covid-19. *N. Engl. J. Med.* **386**, 1397–1408 (2022).

7. H. A. Blair, Ritlecitinib: First approval. *Drugs* **83**, 1315–1321 (2023).

8. P. A. Jänne, G. J. Riely, S. M. Gadgeel, R. S. Heist, S.-H. I. Ou, J. M. Pacheco, M. L. Johnson, J. K. Sabari, K. Leventakos, E. Yau, L. Bazhenova, M. V. Negrao, N. A. Pennell, J. Zhang, K. Anderes, H. Der-Torossian, T. Kheoh, K. Velastegui, X. Yan, J. G. Christensen, R. C. Chao, A. I. Spira, Adagrasib in non-small-cell lung cancer harboring a KRASG12C mutation. *N. Engl. J. Med.* **387**, 120–131 (2022).

9. N. London, R. M. Miller, S. Krishnan, K. Uchida, J. J. Irwin, O. Eidam, L. Gibold, P. Cimermančič, R. Bonnet, B. K. Shoichet, J. Taunton, Covalent docking of large libraries for the discovery of chemical probes. *Nat. Chem. Biol.* **10**, 1066–1072 (2014).

10. G. Bianco, S. Forli, D. S. Goodsell, A. J. Olson, Covalent docking using autodock: Two-point attractor and flexible side chain methods: Covalent Docking with AutoDock. *Protein Sci.* **25**, 295–301 (2016).

11. Y. Wu, C. L. Brooks Iii, Covalent docking in CDOCKER. *J. Comput. Aided Mol. Des.* **36**, 563–574 (2022).

12. D. Toledo Warshaviak, G. Golan, K. W. Borrelli, K. Zhu, O. Kalid, Structure-based virtual screening approach for discovery of covalently bound ligands. *J. Chem. Inf. Model.* **54**, 1941–1950 (2014).

13. S. De Cesco, S. Deslandes, E. Therrien, D. Levan, M. Cueto, R. Schmidt, L.-D. Cantin, A. Mittermaier, L. Juillerat-Jeanneret, N. Moitessier, Virtual screening and computational optimization for the discovery of covalent prolyl oligopeptidase inhibitors with activity in human cells. *J. Med. Chem.* **55**, 6306–6315 (2012).

14. V. Katritch, C. M. Byrd, V. Tseitin, D. Dai, E. Raush, M. Totrov, R. Abagyan, R. Jordan, D. E. Hruby, Discovery of small molecule inhibitors of ubiquitin-like poxvirus proteinase I7L using homology modeling and covalent docking approaches. *J. Comput. Aided Mol. Des.*

21, 549–558 (2007).

15. Y. S. Tan, M. Chakrabarti, R. M. Stein, L. E. Prentis, R. C. Rizzo, T. Kurtzman, M. Fischer, T. E. Balius, Development of receptor desolvation scoring and covalent sampling in DOCK 6: Methods evaluated on a RAS test set. *J. Chem. Inf. Model.*, doi: 10.1021/acs.jcim.4c01623 (2025).

16. M. Rachman, A. Scarpino, D. Bajusz, G. Pálfy, I. Vida, A. Perczel, X. Barril, G. M. Keserű, DUckCov: A Dynamic Undocking-based virtual screening protocol for covalent binders. *ChemMedChem* **14**, 1011–1021 (2019).

17. S. Zhang, J. Tan, Z. Lai, Y. Li, J. Pang, J. Xiao, Z. Huang, Y. Zhang, H. Ji, Y. Lai, Effective virtual screening strategy toward covalent ligands: identification of novel NEDD8-activating enzyme inhibitors. *J. Chem. Inf. Model.* **54**, 1785–1797 (2014).

18. A. Shraga, E. Olshvang, N. Davidzohn, P. Khoshkenar, N. Germain, K. Shurrush, S. Carvalho, L. Avram, S. Albeck, T. Unger, B. Lefker, C. Subramanyam, R. L. Hudkins, A. Mitchell, Z. Shulman, T. Kinoshita, N. London, Covalent docking identifies a potent and selective MKK7 inhibitor. *Cell Chem. Biol.* **26**, 98–108.e5 (2019).

19. C. I. Nnadi, M. L. Jenkins, D. R. Gentile, L. A. Bateman, D. Zaidman, T. E. Balius, D. K. Nomura, J. E. Burke, K. M. Shokat, N. London, Novel K-Ras G12C switch-II covalent binders destabilize Ras and accelerate nucleotide exchange. *J. Chem. Inf. Model.* **58**, 464–471 (2018).

20. E. A. Fink, C. Bardine, S. Gahbauer, I. Singh, T. C. Detomasi, K. White, S. Gu, X. Wan, J. Chen, B. Ary, I. Glenn, J. O'Connell, H. O'Donnell, P. Fajtová, J. Lyu, S. Vigneron, N. J. Young, I. S. Kondratov, A. Alisoltani, L. M. Simons, R. Lorenzo-Redondo, E. A. Ozer, J. F. Hultquist, A. J. O'Donoghue, Y. S. Moroz, J. Taunton, A. R. Renslo, J. J. Irwin, A. García-Sastre, B. K. Shoichet, C. S. Craik, Large library docking for novel SARS-CoV-2 main protease non-covalent and covalent inhibitors. *Protein Sci.* **32**, e4712 (2023).

21. A. Scarpino, G. G. Ferenczy, G. M. Keserű, Comparative evaluation of covalent docking tools. *J. Chem. Inf. Model.* **58**, 1441–1458 (2018).

22. M. Gao, A. F. A. Moumbock, A. Qaseem, Q. Xu, S. Günther, CovPDB: a high-resolution coverage of the covalent protein-ligand interactome. *Nucleic Acids Res.* **50**, D445–D450 (2022).

23. X.-K. Guo, Y. Zhang, CovBinderInPDB: A structure-based covalent binder database. *J. Chem. Inf. Model.* **62**, 6057–6068 (2022).

24. H. Du, X. Zhang, Z. Wu, O. Zhang, S. Gu, M. Wang, F. Zhu, D. Li, T. Hou, P. Pan, CovalentInDB 2.0: an updated comprehensive database for structure-based and ligand-based covalent inhibitor design and screening. *Nucleic Acids Res.* **53**, D1322–D1327 (2024).

25. C. Wen, X. Yan, Q. Gu, J. Du, D. Wu, Y. Lu, H. Zhou, J. Xu, Systematic studies on the protocol and criteria for selecting a covalent docking tool. *Molecules* **24**, 2183 (2019).

26. N. Huang, B. K. Shoichet, J. J. Irwin, Benchmarking sets for molecular docking. *J. Med. Chem.* **49**, 6789–6801 (2006).

27. M. M. Mysinger, M. Carchia, J. J. Irwin, B. K. Shoichet, Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *J. Med. Chem.* **55**, 6582–6594 (2012).

28. R. M. Stein, Y. Yang, T. E. Balius, M. J. O'Meara, J. Lyu, J. Young, K. Tang, B. K. Shoichet, J. J. Irwin, Property-unmatched decoys in docking benchmarks. *J. Chem. Inf. Model.* **61**, 699–714 (2021).

29. M. R. Bauer, T. M. Ibrahim, S. M. Vogel, F. M. Boeckler, Evaluation and optimization of virtual screening workflows with DEKOIS 2.0--a public library of challenging docking benchmark sets. *J. Chem. Inf. Model.* **53**, 1447–1462 (2013).

30. S. G. Rohrer, K. Baumann, Maximum unbiased validation (MUV) data sets for virtual screening based on PubChem bioactivity data. *J. Chem. Inf. Model.* **49**, 169–184 (2009).

31. J. Abramson, J. Adler, J. Dunger, R. Evans, T. Green, A. Pritzel, O. Ronneberger, L. Willmore, A. J. Ballard, J. Bambrick, S. W. Bodenstein, D. A. Evans, C.-C. Hung, M. O'Neill, D. Reiman, K. Tunyasuvunakool, Z. Wu, A. Žemgulytė, E. Arvaniti, C. Beattie, O. Bertolli, A. Bridgland, A. Cherepanov, M. Congreve, A. I. Cowen-Rivers, A. Cowie, M. Figurnov, F. B. Fuchs, H. Gladman, R. Jain, Y. A. Khan, C. M. R. Low, K. Perlin, A. Potapenko, P. Savy, S. Singh, A. Stecula, A. Thillaisundaram, C. Tong, S. Yakneen, E. D. Zhong, M. Zielinski, A. Žídek, V. Bapst, P. Kohli, M. Jaderberg, D. Hassabis, J. M. Jumper, Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).

32. A. Abdeldayem, Y. S. Raouf, S. N. Constantinescu, R. Moriggl, P. T. Gunning, Advances in covalent kinase inhibitors. *Chem. Soc. Rev.* **49**, 2617–2687 (2020).

33. M. K. Gilson, T. Liu, M. Baitaluk, G. Nicola, L. Hwang, J. Chong, BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Res.* **44**, D1045–53 (2016).

34. B. Zdrazil, E. Felix, F. Hunter, E. J. Manners, J. Blackshaw, S. Corbett, M. de Veij, H. Ioannidis, D. M. Lopez, J. F. Mosquera, M. P. Magarinos, N. Bosc, R. Arcila, T. Kizilören, A. Gaulton, A. P. Bento, M. F. Adasme, P. Monecke, G. A. Landrum, A. R. Leach, The ChEMBL Database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. *Nucleic Acids Res.* **52**, D1180–D1192 (2024).

35. Q. Liu, Y. Sabnis, Z. Zhao, T. Zhang, S. J. Buhrlage, L. H. Jones, N. S. Gray, Developing irreversible inhibitors of the protein kinase cysteinome. *Chem. Biol.* **20**, 146–159 (2013).

36. Z. Zhao, P. E. Bourne, Progress with covalent small-molecule kinase inhibitors. *Drug Discov. Today* **23**, 727–735 (2018).

37. L. S. Rathod, P. S. Dabhade, S. N. Mokale, Recent progress in targeting KRAS mutant cancers with covalent G12C-specific inhibitors. *Drug Discov. Today* **28**, 103557 (2023).

38. M. L. Verdonk, V. Berdini, M. J. Hartshorn, W. T. M. Mooij, C. W. Murray, R. D. Taylor, P. Watson, Virtual screening using protein-ligand docking: avoiding artificial enrichment. *J. Chem. Inf. Comput. Sci.* **44**, 793–806 (2004).

39. J. J. Irwin, K. G. Tang, J. Young, C. Dandarchuluun, B. R. Wong, M. Khurelbaatar, Y. S. Moroz, J. Mayfield, R. A. Sayle, ZINC20-A free ultralarge-scale chemical database for

ligand discovery. *J. Chem. Inf. Model.* **60**, 6065–6073 (2020).

40. B. J. Bender, S. Gahbauer, A. Luttens, J. Lyu, C. M. Webb, R. M. Stein, E. A. Fink, T. E. Balius, J. Carlsson, J. J. Irwin, B. K. Shoichet, A practical guide to large-scale docking. *Nat. Protoc.* **16**, 4799–4832 (2021).

41. I. S. Knight, S. Naprienko, J. J. Irwin, Enrichment Score: a better quantitative metric for evaluating the enrichment capacity of molecular docking models, *arXiv [q-bio.QM]* (2022). http://arxiv.org/abs/2210.10905.

42. M. M. Mysinger, B. K. Shoichet, Rapid context-dependent ligand desolvation in molecular docking. *J. Chem. Inf. Model.* **50**, 1561–1573 (2010).

43. H. Park, P. Bradley, P. Greisen Jr, Y. Liu, V. K. Mulligan, D. E. Kim, D. Baker, F. DiMaio, Simultaneous optimization of biomolecular energy functions on features from small molecules and macromolecules. *J. Chem. Theory Comput.* **12**, 6201–6212 (2016).

44. A. V. Sadybekov, V. Katritch, Computational approaches streamlining drug discovery. *Nature* **616**, 673–685 (2023).

45. P. Škrinjar, J. Eberhardt, J. Durairaj, T. Schwede, Have protein-ligand co-folding methods moved beyond memorisation?, *Bioinformatics* (2025). https://www.biorxiv.org/content/10.1101/2025.02.03.636309v2.

46. J. Lyu, N. Kapolka, R. Gumpper, A. Alon, L. Wang, M. K. Jain, X. Barros-Álvarez, K. Sakamoto, Y. Kim, J. DiBerto, K. Kim, I. S. Glenn, T. A. Tummino, S. Huang, J. J. Irwin, O. O. Tarkhanova, Y. Moroz, G. Skiniotis, A. C. Kruse, B. K. Shoichet, B. L. Roth, AlphaFold2 structures guide prospective ligand discovery. *Science* **384**, eadn6354 (2024).

47. M. Holcomb, Y.-T. Chang, D. S. Goodsell, S. Forli, Evaluation of AlphaFold2 structures as docking targets. *Protein Sci.* **32**, e4530 (2023).

48. M. Karelina, J. J. Noh, R. O. Dror, How accurately can one predict drug binding modes using AlphaFold models? *Elife* **12** (2023).

49. M. M. Attwood, D. Fabbro, A. V. Sokolov, S. Knapp, H. B. Schiöth, Trends in kinase drug discovery: targets, indications and inhibitor design. *Nat. Rev. Drug Discov.* **20**, 839–861 (2021).

50. P. Cohen, D. Cross, P. A. Jänne, Kinase drug discovery 20 years after imatinib: progress and future directions. *Nat. Rev. Drug Discov.* **20**, 551–569 (2021).

51. J. Lyu, S. Wang, T. E. Balius, I. Singh, A. Levit, Y. S. Moroz, M. J. O'Meara, T. Che, E. Algaa, K. Tolmachova, A. A. Tolmachev, B. K. Shoichet, B. L. Roth, J. J. Irwin, Ultra-large library docking for discovering new chemotypes. *Nature* **566**, 224–229 (2019).

52. F. Gentile, J. C. Yaacoub, J. Gleave, M. Fernandez, A.-T. Ton, F. Ban, A. Stern, A. Cherkasov, Artificial intelligence-enabled virtual screening of ultra-large chemical libraries with deep docking. *Nat. Protoc.* **17**, 672–697 (2022).

53. A. A. Sadybekov, A. V. Sadybekov, Y. Liu, C. Iliopoulos-Tsoutsouvas, X.-P. Huang, J. Pickett, B. Houser, N. Patel, N. K. Tran, F. Tong, N. Zvonok, M. K. Jain, O. Savych, D. S. Radchenko, S. P. Nikas, N. A. Petasis, Y. S. Moroz, B. L. Roth, A. Makriyannis, V. Katritch,

Synthon-based ligand discovery in virtual libraries of over 11 billion compounds. *Nature* **601**, 452–459 (2022).