Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Structure-based virtual screening of vast chemical space as a starting point for drug discovery

Jens Carlsson[1] and Andreas Luttens[2,3]

Abstract

Structure-based virtual screening aims to find molecules forming favorable interactions with a biological macromolecule using computational models of complexes. The recent surge of commercially available chemical space provides the opportunity to search for ligands of therapeutic targets among billions of compounds. This review offers a compact overview of structure-based virtual screens of vast chemical spaces, highlighting successful applications in early drug discovery for therapeutically important targets such as G protein-coupled receptors and viral enzymes. Emphasis is placed on strategies to explore ultra-large chemical libraries and synergies with emerging machine learning techniques. The current opportunities and future challenges of virtual screening are discussed, indicating that this approach will play an important role in the next-generation drug discovery pipeline.

Addresses

[1] Science for Life Laboratory, Department of Cell and Molecular Biology, Uppsala University, BMC, Box 596, SE-751 24 Uppsala, Sweden

[2] Institute for Medical Engineering & Science and Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[3] Infectious Disease and Microbiome Program, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

Corresponding authors: Carlsson, Jens (jens.carlsson@icm.uu.se); Luttens, Andreas (aluttens@mit.edu)

(Carlsson J.), (Luttens A.)

Current Opinion in Structural Biology 2024, **87**:102829

This review comes from a themed issue on **New Concepts in Drug Discovery (2024)**

Edited by **Andrea Cavalli** and **Alessio Ciulli**

For a complete overview see the Issue and the Editorial

Available online xxx

https://doi.org/10.1016/j.sbi.2024.102829

Introduction

Structure-based drug discovery exploits the three-dimensional structural information of a biological target for ligand development. Computational modeling aiming to predict ligand interactions with a target binding site plays an increasingly important role in the drug discovery process [1]. The molecular docking method predicts the binding mode of a small molecule inside pockets of biological targets (*e.g.*, proteins) and can be used to screen large compound collections for ligands. While established experimental techniques such as high-throughput screening and DNA-encoded chemical libraries depend on the physical availability of compounds, virtual screening can extend beyond the constraints of tangible molecules [2]. Access to large libraries of building blocks and development of robust organic synthesis procedures enabled chemical vendors to offer massive and rapidly growing "make-on-demand" databases [3]. These combinatorial libraries led to a paradigm shift for virtual screening. Instead of evaluating the 10–20 million compounds that are available in stock, searches among billions of readily synthesizable compounds are now possible [4–7]. Notably, at the time of writing this review, the REAL (readily accessible) Space database offers more than 40 billion compounds, which the chemical supplier Enamine can synthesize and deliver for experimental testing within a few weeks. To utilize these libraries efficiently, ligand discovery campaigns need to transition from assaying diverse compound collections to pinpointing specific molecules within vast chemical spaces. However, searching through libraries of such magnitudes presents new challenges. Several strategies for navigation in vast chemical spaces have been developed and this review aims to cover the recent advances in structure-based virtual screening and the impact of machine learning on this field. We focus primarily on approaches that are accessible to the wider modelling community and have been applied successfully in drug discovery.

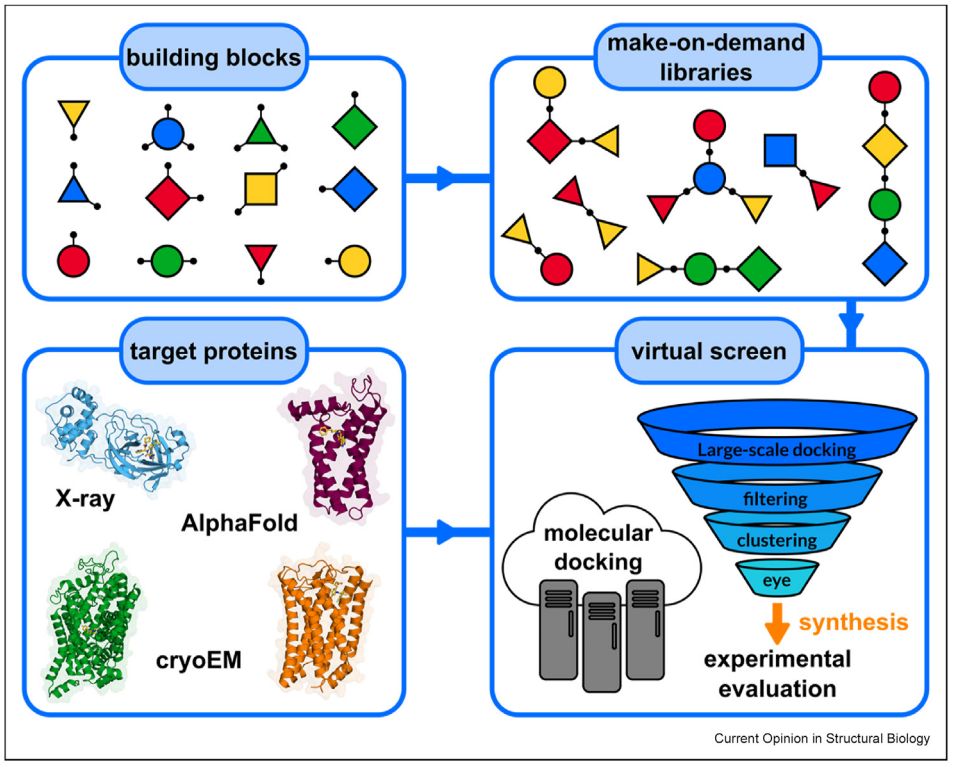
Ultra-large library docking

The most straightforward structure-based approach to evaluate ultra-large libraries is to simply dock each molecule to the target and select a subset of the top-scoring compounds for experimental evaluation. The most recent snapshot of ZINC22 (https://cartblanche.docking.org), an online database of purchasable compounds, contained nearly five billion molecules in ready-to-dock format, and the vast majority of these are available through make-on-demand synthesis [8]. Advancements in computing power and access to distributed cloud platforms enabled successful

prospective screens of hundreds of millions, and more recently even billions, of compounds against diverse targets [2,9]. Furthermore, developments in structural biology have led to an unprecedented wealth of high-resolution crystal and cryoEM structures of proteins that could serve as templates for structure-based modeling [10]. Large library docking has unveiled potent ligands targeting important therapeutic proteins such as receptors, transporters, and viral enzymes. Screens against membrane proteins have been remarkably successful [11–15], indicating that the binding pockets of important targets such as G protein-coupled receptors (GPCRs) are particularly suitable for structure-based methods (Figure 1 and Table 1). Recent discoveries include nanomolar modulators of the prostaglandin E2 receptor 4 (EP4R) and the α_{2A} adrenergic receptor (α_{2A}AR), which represent starting points for development of drugs to treat pain [14,15]. The onset of the COVID-19 pandemic prompted research groups to employ large library docking to identify antivirals, leading to the identification of potent inhibitors of several targets [16,17]. While the studies focused on GPCRs showcased high hit rates of up to 63%, the viral enzymes were more challenging targets due to their less well-defined binding sites. However, although hit rates and compound activities were typically 10-fold lower, novel inhibitors were discovered and experimentally determined structures often confirmed the predicted binding modes. The make-on-demand libraries also provide rapid access to analogs, which can be used to obtain structure–activity-relationships and accelerate hit-to-lead generation. However, despite the enormous number of compounds in the make-on-demand libraries, the availability of molecules bearing a specific chemotype can still be scarce. In such cases, focused databases based on commercially available building blocks can be constructed *in silico*, resulting in large libraries containing the desired scaffold [16,18].

As the computational cost of ultra-large docking screens is high, success is dependent on careful selection of target structure and parameters in the docking calculations. In order to screen billions of compounds, several

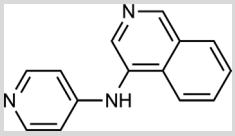
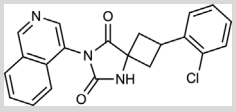
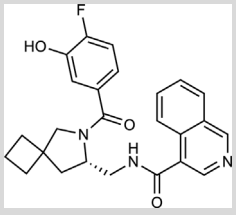
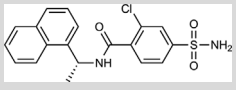
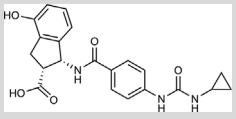
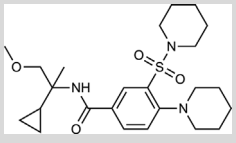
Figure 1



Ultra-large library docking scheme. Available building blocks are systematically coupled *in silico* to afford virtual make-on-demand libraries containing several billion compounds. Molecules in these databases can be obtained with available building blocks and chemical suppliers can rapidly synthesize selected compounds at a low cost. Interactions of molecules in a binding site of a biological target of interest are calculated using a scoring function (large-scale docking). Predictions are post-processed (filtering and clustering) and promising molecules are selected for experimental evaluation by visual inspection (eye). Target templates for large library docking screens can be experimental (X-ray, cryoEM) or AlphaFold2 structures. A more detailed description on this approach can be found in Ref. [19].

Table 1

Examples of successful virtual screening campaigns using different strategies to explore vast chemical space.

Compound	Structure	Activity	Target (hit rate[a])	template[b]	Ref
		Ultra-large library docking			
'9087		K_i: 1.7 nM	α_{2A}AR (63%)	cryoEM 6K41	[14]
19		K_D: 38 nM	Mpro (3%)	X-ray 6W63	[16]
		Machine-learning-accelerated virtual screening			
Z4927220858		IC_{50} : 11 μM	Mpro (10%)	X-ray 6W63	[45]
VPC-300195		IC_{50} : 15 μM	PLpro (10%)	X-ray 7LBR	[44]
		Fragment-based virtual screening			
Z8601		IC_{50} : 0.5 μM	Mac1 (33%)	X-ray 5RUE 5RSW	[48,49]
747		K_i: 0.9 nM	CB_2 (33%)	X-ray 5ZTY	[53]

[a] Percentage of active molecules among those experimentally tested.

[b] PDB accession codes of virtual screening templates.

approximations are made in docking algorithms. The protein structure is held rigid, and complexes are evaluated using rapid scoring functions that are unable to predict binding affinity accurately. Therefore, prior to engaging in large library docking, it is crucial to benchmark and optimize model performance using available experimental data [19]. Effective approaches include assessing if the selected binding site structure and docking parameters can prioritize known ligands over decoys (inactive compounds) and reproduce the key

interactions observed experimentally [20]. Encouragingly, the scores of top-ranked molecules in ultra-large virtual screens continuously improve as the size of the library increases, which suggest that access to larger fractions of chemical space will be beneficial [2,12,21]. This idea is further supported by the fact that hit rates also increased with improved docking scores and novel ligands with high potencies were identified [2,12]. However, the exploration of chemical libraries at this scale also poses new challenges. The many (largely unpublished) unsuccessful screens and the fact that most compounds predicted by virtual screening are inactive show that there is considerable room for improvement of scoring functions. Modeling of the expansion of virtual compound catalogs showed that the larger libraries could lead to accumulation of artifacts among the top-ranked molecules, compounds that exploit the flaws of the scoring function. As these molecules are rare events and often target-dependent, identification of such false positives may prove difficult [21]. Recent studies reported improvements of scoring functions for large library docking, such as interaction fingerprints and more accurate descriptions of ligand strain and receptor flexibility [22–24]. Another interesting approach that should be explored more thoroughly is to re-score a small set of top-ranked molecules from an ultra-large screen using more rigorous computational methods, *e.g.* by combining molecular docking and molecular dynamics free energy calculations [25,26].

For molecular docking to be successful, a high resolution target structure is required and therefore docking campaigns have traditionally relied on access to crystal structures. In recent years, advances in artificial intelligence (*e.g.* the AlphaFold2 and RoseTTAFold methods) have enabled protein structure predictions with near experimental accuracy, greatly expanding the number of structures available for structure-based ligand discovery [27–29]. While several retrospective studies have questioned the direct utility of these models for structure-based virtual screening [30–32], a recent study showed that prospective large library docking campaigns using AlphaFold2 models of the σ_2 and the 5-HT_{2A} receptors proved equally effective as those against experimental structures [33].

Several deep-learning-based protein-ligand docking methods have recently emerged, holding great promise for enhanced speed and accuracy [34–36]. However, their performance is contingent on access to training data of sufficient size and quality, and concerns have been raised about the ability of the models to generalize to dissimilar targets and to generate physically valid ligand conformations [37]. While rapid advancements have been made, the application of deep-learning-based docking methods in prospective virtual screening requires further exploration to assess the value of these approaches in early drug discovery.

Machine-learning-accelerated virtual screening

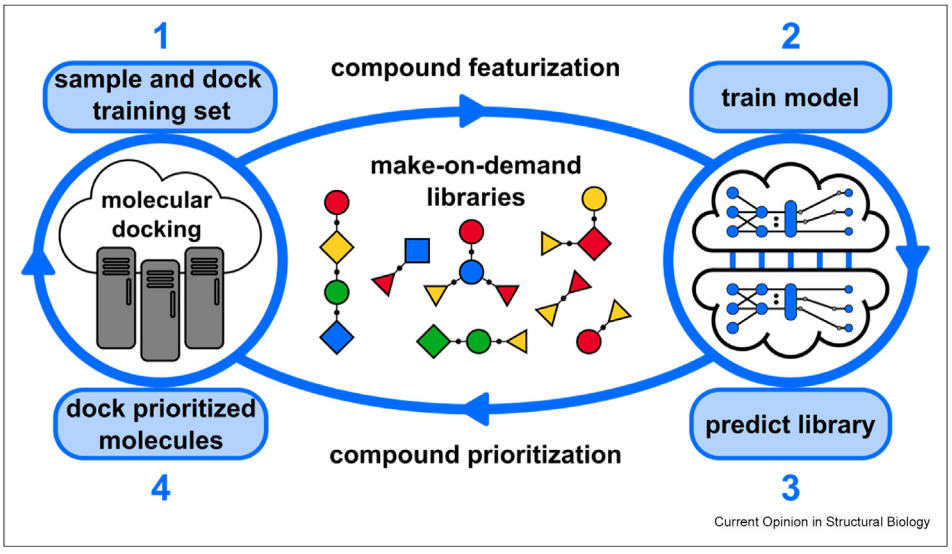
The make-on-demand chemical libraries continue to expand rapidly, and the computational resources associated with this growth are becoming prohibitive for large library docking. Recently developed strategies for traversing these enormous libraries rely on incorporation of machine-learning methods in the virtual screening pipeline [38–42], enabling efficient exploration of the largest available make-on-demand libraries. By training machine-learning models on docking results for a small subset of the library, regions in chemical space with enhanced likelihoods of having favorable scores for the target can be prioritized (Figure 2). Machine learning models trained on molecular docking data have been shown to reduce the number of compounds to dock in multi-billion-scale virtual libraries by up to 100-fold and simultaneously enrich top-scoring molecules [38]. Whereas large library docking has been employed to identify ligands of diverse targets, there are only a few successful applications of machine-learning-accelerated docking screens [25,43–45]. Recent studies using this approach discovered micromolar inhibitors of two SARS-CoV-2 proteases (Table 1) [44,45].

While machine learning can rapidly evaluate large chemical spaces, there are potential drawbacks of this strategy. More sophisticated algorithms, such as message-passing neural networks or large language models, also bear significant computational cost and their potential might not be fully realized due to the poor accuracy of docking scoring functions [41]. In addition, molecules in which machine learning models express higher confidence generally display significant structural overlap with compounds labeled as active in the training set, thereby limiting the capacity of these approaches to find novel molecules [41]. Furthermore, as many top-scoring molecules will be inactive in experiments, their presence in the training set may cause machine learning models to propagate, or even elevate, the high false positive rate of virtual screens. While machine-learning-accelerated docking screens increasingly find their way into early-phase drug discovery projects, their ability to identify ligands merit further investigation.

Fragment-based virtual screening

Notwithstanding the accelerating expansions of virtual compound catalogs, libraries of such magnitudes cover just a minuscule fraction of the drug-like chemical space. The size of the entire drug-like chemical space has been estimated to amount to 10^{60}, and the number of plausible molecules that populate this space drastically increases with the number of atoms composing them [46]. As a result, when new building blocks become available for combinatorial chemistry, the fractions of make-on-demand libraries that consist of larger molecules are most prone to increase. The number of

Figure 2



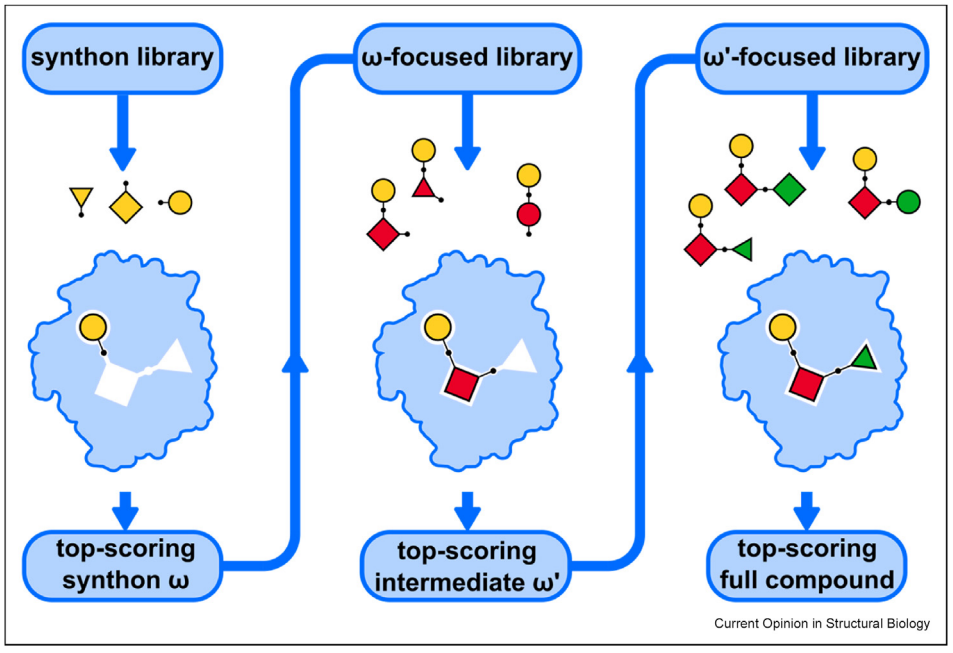
Machine-learning-accelerated virtual screening scheme. (1) A sample of the make-on-demand library is docked to a target of interest. Molecules are represented as numerical descriptors (features) and their corresponding docking scores are used to generate a training set. (2) A machine learning model is trained based on this data and (3) is then used to predict the scores of the remainder of the make-on-demand library. (4) Promising molecules are prioritized for explicit docking calculations, followed by either refinement of the machine learning models (active learning), or selection of candidates for experimental evaluation. A more detailed description of this approach can be found in Ref. [38].

commercially available fragments (molecules with less than 17 heavy atoms) remains relatively small, but as there are orders of magnitude fewer molecules of this size, the chemical space coverage in fragment libraries is high. The intrinsically lower molecular complexity of fragments increases their likelihood to complement a binding site, giving rise to relatively high hit rates in experimental screens [47]. However, fragments typically lack the potency or selectivity displayed by larger molecules and extensive chemical elaborations are required to obtain a lead candidate. Although structure-based virtual screening of fragment libraries has rarely been used in the initial step of ligand discovery campaigns, several recent studies successfully used docking of large fragment libraries (Table 1) [48,49]. For example, a docking screen against Mac1, a non-structural protein involved in viral replication of SARS-CoV-2, resulted in an impressive number of fragment-bound crystal structures. Hits from crystallographic fragment screening have also been successfully used as seeds for design of larger and more complex leads by virtual screening of ultra-large chemical libraries [16,50,51].

Novel virtual screening methods based on the fragment-based approach have emerged and are able to explore the largest available make-on-demand libraries [52–54]. In the V-SYNTHES approach, a small library of fragments is first docked to the target binding site. By leveraging predicted binding modes of top-scoring fragments, or so-called synthons, the regions of commercial chemical space containing larger molecules that encompass the synthon can be prioritized for subsequent docking (Figure 3). This approach addresses a major limitation of large library docking, which demands computational resources that scale linearly with the size of the chemical library. By breaking down the challenge of modelling large compounds into a series of smaller steps, explorations of libraries that are several orders of magnitude larger than what can currently be achieved by explicit docking are enabled. Recent studies demonstrated how synthon-based virtual screening successfully identified low-nanomolar modulators of cannabinoid receptors and inhibitors of the ROCK1 enzyme by computationally evaluating only a small fraction (<0.1%) of the multi-billion chemical library [53,54].

A recently developed machine learning method (FRAME) is based on stepwise elaboration of fragments using a structure-guided approach. The study demonstrated how deep learning can be used to identify viable growing vectors based on an initial fragment-bound protein structure and assess diverse chemical elaborations through the utilization of SE(3)-equivariant neural networks [55]. The use of this approach and variants thereof [56] may lead to compounds with limited synthetic accessibility, which is a major disadvantage compared to the make-on-demand libraries. Encouragingly, recent years have seen substantial progress in

Figure 3



Fragment-based virtual screening scheme. A small library of fragments, or synthons, is docked to the target of interest. Top-scoring synthons (ω) are selected, and chemical libraries focused on this scaffold are prioritized for subsequent virtual screening. In multiple iterations (ω'), top-scoring molecules of greater size and complexity are identified. A more detailed description of this approach can be found in Refs. [53,54].

automated and accurate retro-synthetic analysis tools that help prioritize compounds generated by *de novo* algorithms [57,58].

Conclusion and outlook

Structure-based virtual screening has become an important tool to identify starting points for drug discovery and complements other technologies for hit identification. Make-on-demand chemical libraries will likely reach more than one trillion compounds in the next few years, providing access to unexplored regions of chemical space. Moreover, accurate protein structure prediction by deep learning will increasingly contribute to successful applications of virtual screens to novel biological targets. However, efficient searching through vast virtual libraries is growing more complex due to the required computational resources, creating a need for novel and efficient screening approaches. Although the use of machine learning in virtual screening pipelines is rapidly increasing, the widespread use of these methods in ligand discovery campaigns still has to gain momentum. Fragment-based traversal of chemical space is a computationally efficient technique to identify ligands and benefits greatly from complementary data obtained by biophysical screening methods and protein structure determination for complexes. Structure-based virtual screening offers numerous avenues to find bioactive compounds in large chemical libraries and its significance in early-phase drug discovery projects will likely grow in the foreseeable future.

Declaration of competing interest

J.C. is a founder of DareMe Drug Discovery Consulting.

Data availability

No data was used for the research described in the article.

Acknowledgments

J.C. received funding from the Swedish Research Council, the Swedish Cancer Society, the Olle Engkvist foundation, the Knut and Alice Wallenberg foundation, and the Swedish Strategic Research Program eSSENCE. A.L. was supported by a postdoctoral scholarship from the Knut and Alice Wallenberg foundation (KAW2022.0347).

References

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest
** of outstanding interest

1. Sadybekov AV, Katritch V: **Computational approaches streamlining drug discovery**. *Nature* 2023, **616**:673–685, https://doi.org/10.1038/s41586-023-05905-z.

2. Lyu J, Wang S, Balius TE, Singh I, Levit A, Moroz YS, O'Meara MJ, Che T, Algaa E, Tolmachova K, Tolmachev AA, Shoichet BK, Roth BL, Irwin JJ: **Ultra-large library docking for discovering new chemotypes**. *Nature* 2019, **566**:224–229, https://doi.org/10.1038/s41586-019-0917-9.

3. Grygorenko OO, Radchenko DS, Dziuba I, Chuprina A, Gubina KE, Moroz YS: **Generating multibillion chemical space of readily accessible screening compounds**. *iScience* 2020, **23**:11, https://doi.org/10.1016/j.isci.2020.101681.

4. https://marketing.emolecules.com/explore.

5. https://enamine.net/compound-collections/real-compounds/real-database.

6. https://wuxibiology.com/drug-discovery-services/hit-finding-and-screening-services/virtual-screening/.

7. https://www.otavachemicals.com/products/chemriya.

8. Tingle BI, Tang KG, Castanon M, Gutierrez JJ, Khurelbaatar M, Dandarchuluun C, Moroz YS, Irwin JJ: **ZINC-22 – a free multi-billion-scale database of tangible compounds for ligand discovery**. *J Chem Inf Model* 2023, **63**:1166–1176, https://doi.org/10.1021/acs.jcim.2c01253.

9. Gorgulla C, Boeszoermenyi A, Wang ZF, Fischer PD, Coote PW, Padmanabha Das KM, Malets YS, Radchenko DS, Moroz YS, Scott DA, Fackeldey K, Hoffmann M, Iavniuk I, Wagner G, Arthanari H: **An open-source drug discovery platform enables ultra-large virtual screens**. *Nature* 2020, **580**:663–668, https://doi.org/10.1038/s41586-020-2117-z.

10. Nakane T, Kotecha A, Sente A, McMullan G, Masiulis S, Brown PMGE, Grigoras IT, Malinauskaite L, Malinauskas T, Miehling J, Uchański T, Yu L, Karia D, Pechnikova EV, de Jong E, Keizer J, Bischoff M, McCormack J, Tiemeijer P, Hardwick SW, Chirgadze DY, Murshudov G, Aricescu AR, Scheres SHW: **Single-particle cryo-EM at atomic resolution**. *Nature* 2020, **587**:152–156, https://doi.org/10.1038/s41586-020-2829-0.

11. Ballante F, Kooistra AJ, Kampen S, de Graaf C, Carlsson J: **Structure-based virtual screening for ligands of G protein-coupled receptors: what can molecular docking do for you?** *Pharmacol Rev* 2021, **73**:527–565, https://doi.org/10.1124/pharmrev.120.000246.

12. Alon A, Lyu J, Braz JM, Tummino TA, Craik V, O'Meara MJ, Webb CM, Radchenko DS, Moroz YS, Huang XP, Liu Y, Roth BL, Irwin JJ, Basbaum AI, Shoichet BK, Kruse AC: **Structures of the σ_2 receptor enable docking for bioactive ligand discovery**. *Nature* 2021, **600**:759–764, https://doi.org/10.1038/s41586-021-04175-x.

13. Singh I, Seth A, Billesbølle CB, Braz J, Rodriguiz RM, Roy K, Bekele B, Craik V, Huang XP, Boytsov D, Pogorelov VM, Lak P, O'Donnell H, Sandtner W, Irwin JJ, Roth BL, Basbaum AI, Wetsel WC, Manglik A, Shoichet BK, Rudnick G: **Structure-based discovery of conformationally selective inhibitors of the serotonin transporter**. *Cell* 2023, **186**:2160–2175.e17, https://doi.org/10.1016/j.cell.2023.04.010.

14. Fink EA, Xu J, Hübner H, Braz JM, Seemann P, Avet C, Craik V, Weikert D, Schmidt MF, Webb CM, Tolmachova NA, Moroz YS, Huang XP, Kalyanaraman C, Gahbauer S, Chen G, Liu Z, Jacobson MP, Irwin JJ, Bouvier M, Du Y, Shoichet BK, Basbaum AI, Gmeiner P: **Structure-based discovery of nonopioid analgesics acting through the α_{2A}-adrenergic receptor**. *Science* 2022, **30**:6614, https://doi.org/10.1126/science.abn7065.

15. * Gahbauer S, DeLeon C, Braz JM, Craik V, Kang HJ, Wan X, Huang XP, Billesbølle CB, Liu Y, Che T, Deshpande I, Jewell M, Fink EA, Kondratov IS, Moroz YS, Irwin JJ, Basbaum AI, Roth BL, Shoichet BK: **Docking for EP4R antagonists active against inflammatory pain**. *Nat Commun* 2023, **14**:8067, https://doi.org/10.1038/s41467-023-43506-6.
This paper reports a large-library docking screen leading to the discovery of nanomolar antagonists of the prostaglandin E2 receptor 4 with anti-inflammatory activity.

16. Luttens A, Gullberg H, Abdurakhmanov A, Vo DD, Akaberi D, Talibov VO, Nekhotiaeva N, Vangeel L, De Jonghe S, Jochmans D, Krambrich J, Tas A, Lundgren B, Gravenfors Y, Craig AJ, Atilaw Y, Sandström A, Moodie LWK, Lundkvist Å, van Hemert MJ, Neyts J, Lennerstrand J, Kihlberg J, Sandberg K, Danielson UH, Carlsson J: **Ultralarge virtual screening identifies SARS-CoV-2 main protease inhibitors with broad-spectrum activity against coronaviruses**. *J Am Chem Soc* 2022, **144**:2905–2920, https://doi.org/10.1021/jacs.1c08402.

17. Singh I, Li F, Fink EA, Chau I, Li A, Rodriguez-Hernández A, Glenn I, Zapatero-Belinchón FJ, Rodriguez ML, Devkota K, Deng Z, White K, Wan X, Tolmachova NA, Moroz YS, Kaniskan HÜ, Ott M, García-Sastre A, Jin J, Fujimori DG, Irwin JJ, Vedadi M, Shoichet BK: **Structure-based discovery of inhibitors of the SARS-CoV-2 Nsp14 N7-Methyltransferase**. *J Med Chem* 2023, **66**:7785–7803, https://doi.org/10.1021/acs.jmedchem.2c02120.

18. Kaplan AL, Confair DN, Kim K, Barros-Álvarez X, Rodriguiz RM, Yang Y, Kweon OS, Che T, McCorvy JD, Kamber DN, Phelan JP, Martins LC, Pogorelov VM, DiBerto JF, Slocum ST, Huang XP, Kumar JM, Robertson MJ, Panova O, Seven AB, Wetsel AQ, Wetsel WC, Irwin JJ, Skiniotis G, Shoichet BK, Roth BL, Ellman JA: **Bespoke library docking for 5-HT2A receptor agonists with antidepressant activity**. *Nature* 2022, **610**:582–591, https://doi.org/10.1038/s41586-022-05258-z.

19. Bender BJ, Gahbauer S, Luttens A, Lyu J, Webb CM, Stein RM, Fink EA, Balius TE, Carlsson J, Irwin JJ, Shoichet BK: **A practical guide to large-scale docking**. *Nat Protoc* 2021, **16**:4799–4832, https://doi.org/10.1038/s41596-021-00597-z.

20. Stein RM, Yang Y, Balius TE, O'Meara MJ, Lyu J, Young J, Tang K, Shoichet BK, Irwin JJ: **Property-unmatched decoys in docking benchmarks**. *J Chem Inf Model* 2021, **61**:699–714, https://doi.org/10.1021/acs.jcim.0c00598.

21. Lyu J, Irwin JJ, Shoichet BK: **Modeling the expansion of virtual screening libraries**. *Nat Chem Biol* 2023, **19**:712–718, https://doi.org/10.1038/s41589-022-01234-w.

22. Gu S, Smith MS, Yang Y, Irwin JJ, Shoichet BK: **Ligand strain energy in large library docking**. *J Chem Inf Model* 2021, **61**:4331–4341, https://doi.org/10.1021/acs.jcim.1c00368.

23. Fischer M, Coleman RG, Fraser JS, Shoichet BK: **Incorporation of protein flexibility and conformational energy penalties in docking screens to improve ligand discovery**. *Nat Chem* 2014, **6**:575–583, https://doi.org/10.1038/nchem.1954.

24. Fassio AV, Shub L, Ponzoni L, McKinley J, O'Meara MJ, Ferreira RS, Keiser MJ, de Melo Minardi RC: **Prioritizing virtual screening with interpretable interaction fingerprints**. *J Chem Inf Model* 2022, **62**:4300–4318, https://doi.org/10.1021/acs.jcim.2c00695.

25. * Gutkin E, Gusev F, Gentile F, Ban F, Koby SB, Narangoda C, Isayev O, Cherkasov A, Kurnikova MG: **In silico screening of LRRK2 WDR domain inhibitors using deep docking and free energy calculations**. *ChemRxiv preprint ChemRxiv* 2024, https://doi.org/10.26434/chemrxiv-2023-lnzvr-v2.
This paper reports a workflow that combines machine learning and physics-based computational methods to identify ligands of the WDR domain of LRRK2, a target for which no inhibitors are known.

26. Feng M, Heinzelmann G, Gilson MK: **Absolute binding free energy calculations improve enrichment of actives in virtual compound screening**. *Sci Rep* 2022, **12**, 13640, https://doi.org/10.1038/s41598-022-17480-w.

27. Google DeepMind AlphaFold Team: ***Isomorphic labs team: performance and structural coverage of the latest, in-development AlphaFold model***. https://storage.googleapis.com/deepmind-media/DeepMind.com/Blog/a-glimpse-of-the-next-generation-of-alphafold/alphafold_latest_oct2023.pdf.

28. ** Krishna R, Wang J, Ahern W, Sturmfels P, Venkatesh P, Kalvet I, Lee GR, Morey-Burrows FS, Anishchenko I, Humphreys IR, McHugh R, Vafeados D, Li X, Sutherland GA, Hitchcock A, Hunter CN, Baek M, DiMaio F, Baker D: **Generalized biomolecular modeling and design with RoseTTAFold all-atom**. *bioRxiv preprint bioRxiv* 2023, https://doi.org/10.1101/2023.10.09.561603.
This paper presents RoseTTAFold All-Atom, which enables modelling of full biological assemblies that contain several distinct molecule types, such as proteins, DNA, RNA, small molecules, and metals.

29. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D: **Highly**

accurate protein structure prediction with AlphaFold. *Nature* 2021, **596**:583–589, https://doi.org/10.1038/s41586-021-03819-2.

30. Karelina M, Noh JJ, Dror RO: **How accurately can one predict drug binding modes using AlphaFold models?** *Elife* 2023, **12**, RP89386, https://doi.org/10.7554/eLife.89386.

31. Wong F, Krishnan A, Zheng EJ, Stärk H, Manson AL, Earl AM, Jaakkola T, Collins JJ: **Benchmarking AlphaFold-enabled molecular docking predictions for antibiotic discovery**. *Mol Syst Biol* 2022, **18**, e11081, https://doi.org/10.15252/msb.202211081.

32. Zhang Y, Vass M, Shi D, Abualrous E, Chambers JM, Chopra N, Higgs C, Kasavajhala K, Li H, Nandekar P, Sato H, Miller EB, Repasky MP, Jerome SV: **Benchmarking refined and unrefined AlphaFold2 structures for hit discovery**. *J Chem Inf Model* 2023, **63**:1656–1667, https://doi.org/10.1021/acs.jcim.2c01219.

33. ** Lyu J, Kapolka N, Gumpper R, Alon A, Wang L, Jain MK, Barros-Álvarez X, Sakamoto K, Kim Y, DiBerto J, Kim K, Tummino TA, Huang S, Irwin JJ, Tarkhanova OO, Moroz Y, Skiniotis G, Kruse AC, Shoichet BK, Roth BL: **AlphaFold2 structures template ligand discovery**. *bioRxiv preprint bioRxiv* 2023, https://doi.org/10.1101/2023.12.20.572662.
This paper challenges the conclusions of retrospective analyses of molecular docking screens against AlphaFold2 models. Prospective docking screens against AlphaFold2 models and experimental structures resulted in comparable hit rates.

34. Corso G, Stärk H, Jing B, Barzilay R, Jaakkola T: **Diffdock: diffusion steps, twists, and turns for molecular docking**. *arXiv preprint arXiv* 2022, https://doi.org/10.48550/arXiv.2210.01776.

35. Lu W, Wu Q, Zhang J, Rao J, Li C, Zheng S: **TANKBind: trigonometry-aware neural NetworKs for drug-protein binding structure prediction**. *arXiv preprint arXiv* 2022, https://doi.org/10.1101/2022.06.06.495043.

36. Bryant P, Kelkar A, Guljas A, Clementi C, Noé F: **Structure prediction of protein-ligand complexes from sequence information with Umol**. *bioRxiv preprint bioRxiv* 2023, https://doi.org/10.1101/2023.11.03.565471.

37. * Buttenschoen M, Morris GM, Deane CM: **PoseBusters: AI-based docking methods fail to generate physically valid poses or generalise to novel sequences**. *Chem Sci* 2024, https://doi.org/10.1039/D3SC04185A.
This paper demonstrates that current deep learning-based docking methods often generate physically invalid molecular structures and do not outperform classical docking algorithms.

38. Gentile F, Agrawal V, Hsing M, Ton AT, Ban F, Norinder U, Gleave ME, Cherkasov A: **Deep docking: a deep learning platform for augmentation of structure based drug discovery**. *ACS Cent Sci* 2020, **6**:939–949, https://doi.org/10.1021/acscentsci.0c00229.

39. Gentile F, Yaacoub JC, Gleave J, Fernandez M, Ton AT, Ban F, Stern A, Cherkasov A: **Artificial intelligence-enabled virtual screening of ultra-large chemical libraries with deep docking**. *Nat Protoc* 2022, **17**:672–697, https://doi.org/10.1038/s41596-021-00659-2.

40. Yang Y, Yao K, Repasky MP, Leswing K, Abel R, Shoichet BK, Jerome SV: **Efficient exploration of chemical space with docking and deep learning**. *J Chem Theor Comput* 2021, **17**:7106–7119, https://doi.org/10.1021/acs.jctc.1c00810.

41. * Luttens A, Cabeza de Vaca I, Sparring L, Norinder U, Carlsson J: **Rapid traversal of ultralarge chemical space using machine learning guided docking screens**. *ChemRxiv preprint ChemRxiv* 2023, https://doi.org/10.26434/chemrxiv-2023-w3x36.
This paper presents an analysis of the effectiveness of combining machine learning models and molecular docking in virtual screening.

42. Graff DE, Shakhnovich EI, Coley CW: **Accelerating high-throughput virtual screening through molecular pool-based active learning**. *Chem Sci* 2021, **12**:7866–7881, https://doi.org/10.1039/d0sc06805e.

43. Tang M, Wen C, Lin J, Chen H, Ran T: **Discovery of novel A2AR antagonists through deep learning-based virtual screening**. *Artif Intell Life Sci* 2023, **3**, 100058, https://doi.org/10.1016/j.ailsci.2023.100058.

44. Garland O, Ton AT, Moradi S, Smith JR, Kovacic S, Ng K, Pandey M, Ban F, Lee J, Vuckovic M, Worrall LJ, Young RN, Pantophlet R, Strynadka NCJ, Cherkasov A: **Large-scale virtual screening for the discovery of SARS-CoV-2 Papain-like protease (PLpro) non-covalent Inhibitors**. *J Chem Inf Model* 2023, **63**:2158–2169, https://doi.org/10.1021/acs.jcim.2c01641.

45. Gentile F, Fernandez M, Ban F, Ton AT, Mslati H, Perez CF, Leblanc E, Yaacoub JC, Gleave J, Stern A, Wong B, Jean F, Strynadka N, Cherkasov A: **Automated discovery of noncovalent inhibitors of SARS-CoV-2 main protease by consensus deep docking of 40 billion small molecules**. *Chem Sci* 2021, **12**:15960–15974, https://doi.org/10.1039/D1SC05579H.

46. Bohacek RS, McMartin C, Guida WC: **The art and practice of structure-based drug design: a molecular modeling perspective**. *Med Res Rev* 1996, **16**:3–50, https://doi.org/10.1002/(SICI)1098-1128(199601)16:1%3c3::AID-MED1%3e3.0.CO;2-6.

47. Erlanson DA, Fesik SW, Hubbard RE, Jahnke W, Jhoti H: **Twenty years on: the impact of fragments on drug discovery**. *Nat Rev Drug Discov* 2016, **15**:605–619, https://doi.org/10.1038/nrd.2016.109.

48. * Schuller M, Correy GJ, Gahbauer S, Fearon D, Wu T, Díaz RE, Young ID, Carvalho Martins L, Smith DH, Schulze-Gahmen U, Owens TW, Deshpande I, Merz GE, Thwin AC, Biel JT, Peters JK, Moritz M, Herrera N, Kratochvil HT, , QCRG Structural Biology Consortium, Aimon A, Bennett JM, Brandao Neto J, Cohen AE, Dias A, Douangamath A, Dunnett L, Fedorov O, Ferla MP, Fuchs MR, Gorrie-Stone TJ, Holton JM, Johnson MG, Krojer T, Meigs G, Powell AJ, Rack JGM, Rangel VL, Russi S, Skyner RE, Smith CA, Soares AS, Wierman JL, Zhu K, O'Brien P, Jura N, Ashworth A, Irwin JJ, Thompson MC, Gestwicki JE, von Delft F, Shoichet BK, Fraser JS, Ahel I: **Fragment binding to the Nsp3 macrodomain of SARS-CoV-2 identified through crystallographic screening and computational docking**. *Sci Adv* 2021, **7**, eabf8711, https://doi.org/10.1126/sciadv.abf8711.
This paper reports how large-scale fragment docking against Mac1 was able to accurately predict fragment binding modes.

49. Gahbauer S, Correy GJ, Schuller M, Ferla MP, Doruk YU, Rachman M, Wu T, Diolaiti M, Wang S, Neitz RJ, Fearon D, Radchenko D, Moroz Y, Irwin JJ, Renslo AR, Taylor JC, Gestwicki JE, von Delft F, Ashworth A, Ahel I, Shoichet BK, Fraser JS: **Iterative computational design and crystallographic screening identifies potent inhibitors targeting the Nsp3 macrodomain of SARS-CoV-2**. *Proc Natl Acad Sci U S A* 2023, **120**, e2212931120, https://doi.org/10.1073/pnas.2212931120.

50. Müller J, Klein R, Tarkhanova O, Gryniukova A, Borysko P, Merkl S, Ruf M, Neumann A, Gastreich M, Moroz YS, Klebe G, Glinca S: **Magnet for the needle in haystack: "Crystal Structure First" fragment hits unlock active chemical matter using targeted exploration of vast chemical spaces**. *J Med Chem* 2022, **65**:15663–15678, https://doi.org/10.1021/acs.jmedchem.2c00813.

51. Boby ML, Fearon D, Ferla M, Filep M, Koekemoer L, Robinson MC, , COVID Moonshot Consortium, Chodera JD, Lee AA, London N, von Delft A, von Delft F, Achdout H, Aimon A, Alonzi DS, Arbon R, Aschenbrenner JC, Balcomb BH, Bar-David E, Barr H, Ben-Shmuel A, Bennett J, Bilenko VA, Borden B, Boulet P, Bowman GR, Brewitz L, Brun J, Bvnbs S, Calmiano M, Carbery A, Carney DW, Cattermole E, Chang E, Chernyshenko E, Clyde A, Coffland JE, Cohen G, Cole JC, Contini A, Cox L, Croll TI, Cvitkovic M, De Jonghe S, Dias A, Donckers K, Dotson DL, Douangamath A, Duberstein S, Dudgeon T, Dunnett LE, Eastman P, Erez N, Eyermann CJ, Fairhead M, Fate G, Fedorov O, Fernandes RS, Ferrins L, Foster R, Foster H, Fraisse L, Gabizon R, García-Sastre A, Gawriljuk VO, Gehrtz P, Gileadi C, Giroud C, Glass WG, Glen RC, Glinert I, Godoy AS, Gorichko M, Gorrie-Stone T, Griffen EJ, Haneef A, Hassell Hart S, Heer J, Henry M, Hill M, Horrell S, Huang QYJ, Huliak VD, Hurley MFD, Israely T, Jajack A, Jansen J, Jnoff E, Jochmans D, John T, Kaminow B, Kang L, Kantsadi AL, Kenny PW, Kiappes JL, Kinakh SO, Kovar B, Krojer T, La VNT, Laghnimi-Hahn S, Lefker BA, Levy H, Lithgo RM, Logvinenko IG, Lukacik P, Macdonald HB, MacLean EM, Makower LL, Malla TR, Marples PG, Matviiuk T, McCorkindale W, McGovern BL, Melamed S, Melnykov KP, Michurin O, Miesen P, Mikolajek H, Milne BF, Minh D, Morris A, Morris GM, Morwitzer MJ, Moustakas D, Mowbray CE,

Nakamura AM, Neto JB, Neyts J, Nguyen L, Noske GD, Oleinikovas V, Oliva G, Overheul GJ, Owen CD, Pai R, Pan J, Paran N, Payne AM, Perry B, Pingle M, Pinjari J, Politi B, Powell A, Pšenák V, Pulido I, Puni R, Rangel VL, Reddi RN, Rees P, Reid SP, Reid L, Resnick E, Ripka EG, Robinson RP, Rodriguez-Guerra J, Rosales R, Rufa DA, Saar K, Saikatendu KS, Salah E, Schaller D, Scheen J, Schiffer CA, Schofield CJ, Shafeev M, Shaikh A, Shaqra AM, Shi J, Shurrush K, Singh S, Sittner A, Sjö P, Skyner R, Smalley A, Smeets B, Smilova MD, Solmesky LJ, Spencer J, Strain-Damerell C, Swamy V, Tamir H, Taylor JC, Tennant RE, Thompson W, Thompson A, Tomásio S, Tomlinson CWE, Tsurupa IS, Tumber A, Vakonakis I, van Rij RP, Vangeel L, Varghese FS, Vaschetto M, Vitner EB, Voelz V, Volkamer A, Walsh MA, Ward W, Weatherall C, Weiss S, White KM, Wild CF, Witt KD, Wittmann M, Wright N, Yahalom-Ronen Y, Yilmaz NK, Zaidmann D, Zhang I, Zidane H, Zitzmann N, Zvornicanin SN: **Open science discovery of potent noncovalent SARS-CoV-2 main protease inhibitors**. *Science* 2023, **382**, eabo7201, https://doi.org/10.1126/science.abo7201.

52. Piticchio SG, Martínez-Cartró M, Scaffidi S, Rachman M, Rodriguez-Arevalo S, Sanchez-Arfelis A, Escolano C, Picaud S, Krojer T, Filippakopoulos P, von Delft F, Galdeano C, Barril X: **Discovery of novel BRD4 ligand scaffolds by automated navigation of the fragment chemical space**. *J Med Chem* 2021, **64**:17887–17900, https://doi.org/10.1021/acs.jmedchem.1c01108.

53. * Sadybekov AA, Sadybekov AV, Liu Y, Iliopoulos-Tsoutsouvas C, Huang XP, Pickett J, Houser B, Patel N, Tran NK, Tong F, Zvonok N, Jain MK, Savych O, Radchenko DS, Nikas SP, Petasis NA, Moroz YS, Roth BL, Makriyannis A, Katritch V: **Synthon-based ligand discovery in virtual libraries of over 11 billion compounds**. *Nature* 2022, **601**:452–459, https://doi.org/10.1038/s41586-021-04220-9.
This paper reports the discovery of submicromolar ligands of two different drug targets using fragment-based navigation of commercial chemical space.

54. Beroza P, Crawford JJ, Ganichkin O, Gendelev L, Harris SF, Klein R, Miu A, Steinbacher S, Klingler FM, Lemmen C: **Chemical space docking enables large-scale structure-based virtual screening to discover ROCK1 kinase inhibitors**. *Nat Commun* 2022, **13**:6447, https://doi.org/10.1038/s41467-022-33981-8.

55. Powers AS, Yu HH, Suriana P, Koodli RV, Lu T, Paggi JM, Dror RO: **Geometric deep learning for structure-based ligand design**. *ACS Cent Sci* 2023, **9**:2257–2267, https://doi.org/10.1021/acscentsci.3c00572.

56. Ragoza M, Masuda T, Koes DR: **Generating 3D molecules conditional on receptor binding sites with deep generative models**. *Chem Sci* 2022, **13**:2701–2713, https://doi.org/10.1039/d1sc05976a.

57. Zhong W, Yang Z, Chen CY: **Retrosynthesis prediction using an end-to-end graph generative architecture for molecular graph editing**. *Nat Commun* 2023, **14**:3009, https://doi.org/10.1038/s41467-023-38851-5.

58. Gao W, Coley CW: **The synthesizability of molecules proposed by generative models**. *J Chem Inf Model* 2020, **60**:5714–5723, https://doi.org/10.1021/acs.jcim.0c00174.