Nature Communications

https://doi.org/10.1038/s41467-026-68941-z

Article in Press

Developments and challenges in hit progression within fragment-based drug discovery

Received: 2 September 2024

Accepted: 7 January 2026

Cite this article as: Grosjean, H., Biggin, P.C. Developments and challenges in hit progression within fragment-based drug discovery. *Nat Commun* (2026). https://doi.org/10.1038/s41467-026-68941-z

Harold Grosjean & Philip C. Biggin

Developments and Challenges in Hit Progression within Fragment-Based Drug Discovery

Harold Grosjean[1,2,*] and Philip C Biggin[1*]

1. Structural Bioinformatics and Computational Biochemistry, Department of Biochemistry, University of Oxford, South Parks Road, OX1 3QU, Oxford, UK

2. Diamond Light Source Ltd, Harwell Science and Innovation Campus, OX11 0QX, Didcot, UK

* Corresponding authors: harold@foundersfunders.tech and philip.biggin@bioch.ox.ac.uk

Abstract

Fragment-based Drug Discovery (FBDD) is a proven methodology for the discovery of new therapeutics. After the identification of small molecular fragments, subsequent steps are guided by the "Design, Make, Test" (DMT) cycle. During the “Design” phase, chemical modifications are proposed that generate Structure-Activity Relationship information, improve interaction profiles and physico-chemical properties. In the “Make” phase, designs are synthesised into viable compounds, with an emphasis on feasibility, scalability and the incorporation of novel chemistries enabling broad chemical space sampling. Finally, the “Test” phase evaluates these compounds through a series of assays, identifying binders and enabling Structure-Activity Relationship models that guide subsequent designs. Within DMT cycles, fragment progression – the process of converting initial hits into more potent follow-up lead compounds - is an essential component, but has many challenges associated with it. Here, we review such challenges along with recent developments designed to mitigate them.

Introduction

Drug discovery is a complex process requiring substantial financial investment. Despite major technological advances, the pharmaceutical industry has experienced a consistent decline in R&D efficiency over the past six decades, with the number of new drugs approved per billion US dollars spent halving approximately every nine years, a trend referred to as *Eroom's Law ("Moore's" Law backwards)*[1]. This, in conjunction with an ever-expanding therapeutic landscape including evolving diseases, increasing pandemic threats and an aging population, underscores the imperative for improvements in the process of drug-discovery[1]. Fragment-Based Drug Discovery (FBDD) has emerged as a successful paradigm **(Fig. 1a)** in modern small molecule discovery[2,3] with 6 FDA approved compounds **(Fig. 2a)**[4], several more having entered clinical trials and even more used as chemical probes (see fragment to lead yearly summaries[5,6] and the Practical Fragments blog[7]). Fragments are not only pivotal in conventional small molecule drug discovery but have also become crucial in the development of novel drug modalities such as protein-protein interaction inhibitors[8], proteolysis-targeting chimeras (PROTACs)[9] and molecular glues[10]. Additionally, fragments can also be employed for more biology-focused tasks, including probing protein motions[11], binding site discovery[12] and druggability assessment[13], enabling a deeper understanding of protein targets. The small molecule drug discovery process begins with identifying a biological target and progresses through hit identification, hit-to-lead evolution, and lead optimisation before compounds enter preclinical *in vivo* and clinical testing. Among these stages, hit identification and hit-to-lead optimisation represent critical and resource-intensive bottlenecks, as they define both the chemical viability of candidate molecules and their potential for downstream development[14]. These phases are shaped by the paradigm of selecting an initial 'hit' compound from a molecular library, which is then iteratively elaborated into a lead series with improved potency, selectivity, and drug-like properties[15].

In contrast to high-throughput screening (HTS) where large libraries **(Fig. 1a)**, typically comprised of over a million molecules are screened, FBDD relies on a different, more focused approach. Rather than seeking to identify relatively large and potent hits as in HTS, FBDD relies on the sampling of diverse smaller molecular fragments **(Fig. 4)** that are ,somewhat arbitrarily, defined by the "Rule of 3" (molecular weight ≤300Da, logP ≤3,

hydrogen bond donors and acceptors ≤3)[3]. Hence, fragment library design or selection plays a foundational role in FBDD by ensuring that screened compounds are diverse, synthetically tractable, and optimised for downstream progression[16]. Fragments, with their small molecular size, achieve a more efficient chemical space exploration than larger and potentially more potent molecules used in HTS **(Fig. 5)**[17,18]. As molecular size increases, the quantity of potentially active molecules explodes, but the ability to explore all possibilities concurrently diminishes **(Fig. 1a)**[17]. The reduced molecular complexity of fragments also means they tend to have weaker potency and typically bind to targets with higher hit rates at millimolar to micromolar affinity. Although fragment hits have weak potencies, they are expected to show favourable efficiency metrics relative to their size. In practice, screening hits (fragment or HTS) are often evaluated using ligand efficiency (LE)[19], the binding energy per heavy atom (**Table 2**). Since a high LE (for example, ≥0.3 kcal·mol^{-1} per heavy atom) indicates that a fragment binds strongly for its size. Another useful metric is lipophilic ligand efficiency (LLE), which relates a compound's potency to its lipophilicity; a high LLE highlights fragment hits that achieve affinity without relying on excessive hydrophobicity. Together with the physicochemical constraints of the Rule of 3 that define fragment-like space, these metrics help prioritise high-quality fragment hits with efficient binding and room for transformation into lead compounds.

A fundamental limitation of FBDD lies in the intrinsically weak affinity of fragment hits, which, while enabling efficient chemical space exploration, introduces significant practical challenges for both computational and experimental fragment hit-to-lead strategies[20]. In particular, the weak binding affinities of fragments complicate the use of automation and modelling tools that are more readily applied to higher-affinity, drug-like molecules. For example, structure determination becomes more error-prone, structure-activity relationship (SAR) trends are often less distinct, and computational prioritisation methods, such as docking or free energy calculations, must contend with increased uncertainty due to flexible, unstable poses. Consequently, emerging techniques are increasingly focused on integrative strategies that combine orthogonal assays, structural insights, and advanced modelling strategies to better capture weak signals and guide early DMT cycles **(Fig. 3)**. These developments highlight that, rather than replacing traditional approaches,

automation and modelling in FBDD must be carefully adapted to fragment-specific limitations through innovation, iteration, and expert interpretation.

The inherently weak binding affinities of fragments, also implies that sensitive experimental identification methods using high target purity are typically required. In general, there are three main classes of approach that can be used to experimentally confirm hits against the target **(Fig. 1b)**[5]. (i) Biochemical techniques to evaluate a target's functional modulation (i.e. inhibition or activation) through enzymatic or binding assays[21,22]. (ii) Biophysical techniques for assessing binding strength with the target, typically derived from kinetic studies or thermodynamic properties[23,24]. (iii) Structural techniques to decipher binding modes in 3D at an atomic level[25,26]. These three assay classes tend to differ in throughput, sensitivity, requirements and the type of information they yield (**Table 1**)[27]. Biochemical assays, such as enzymatic or functional tests, are typically high-throughput and cost-effective, providing rapid activity-based readouts, sometimes from a single concentration experiment. Biophysical methods offer higher sensitivity and provide direct evidence of binding interactions, though often at lower throughput because they typically require a full titration. Structural techniques offer atomic-resolution insights into binding modes but require more substantial experimental investment to obtain suitable samples and are therefore of lower throughput[28]. In practice, these assays are often used in cascades or orthogonal combinations, where initial high-throughput methods serve as initial screens and more sensitive, lower-throughput assays validate or refine primary hits[29]. This integrative approach improves confidence in weak-binding fragment identification and supports hit selection by layering complementary readouts.

Pan-assay interference compounds (PAINS) are chemotypes that generate target-independent signals via colloidal aggregation, redox cycling, covalent reactivity, metal chelation or assay-reporter interference, leading to false-positive readouts in small molecule screens[30,31]. Typical PAINS-enriched motifs include catechols/quinones, rhodanines, Michael acceptors, azo dyes and related redox-active scaffolds[32]. Because fragments operate near detection limits, early PAINS filtration and orthogonal validation are essential to ensure that weak signals reflect genuine target engagement rather than assay artefacts[16] (**Table 2**). Throughout, we emphasise orthogonal validation of fragment

hits, preferably including a structural method when feasible, as crystallographic (or other structural) confirmation provides assay-independent evidence of a physically plausible pose and helps mitigate PAINS-type artefacts.

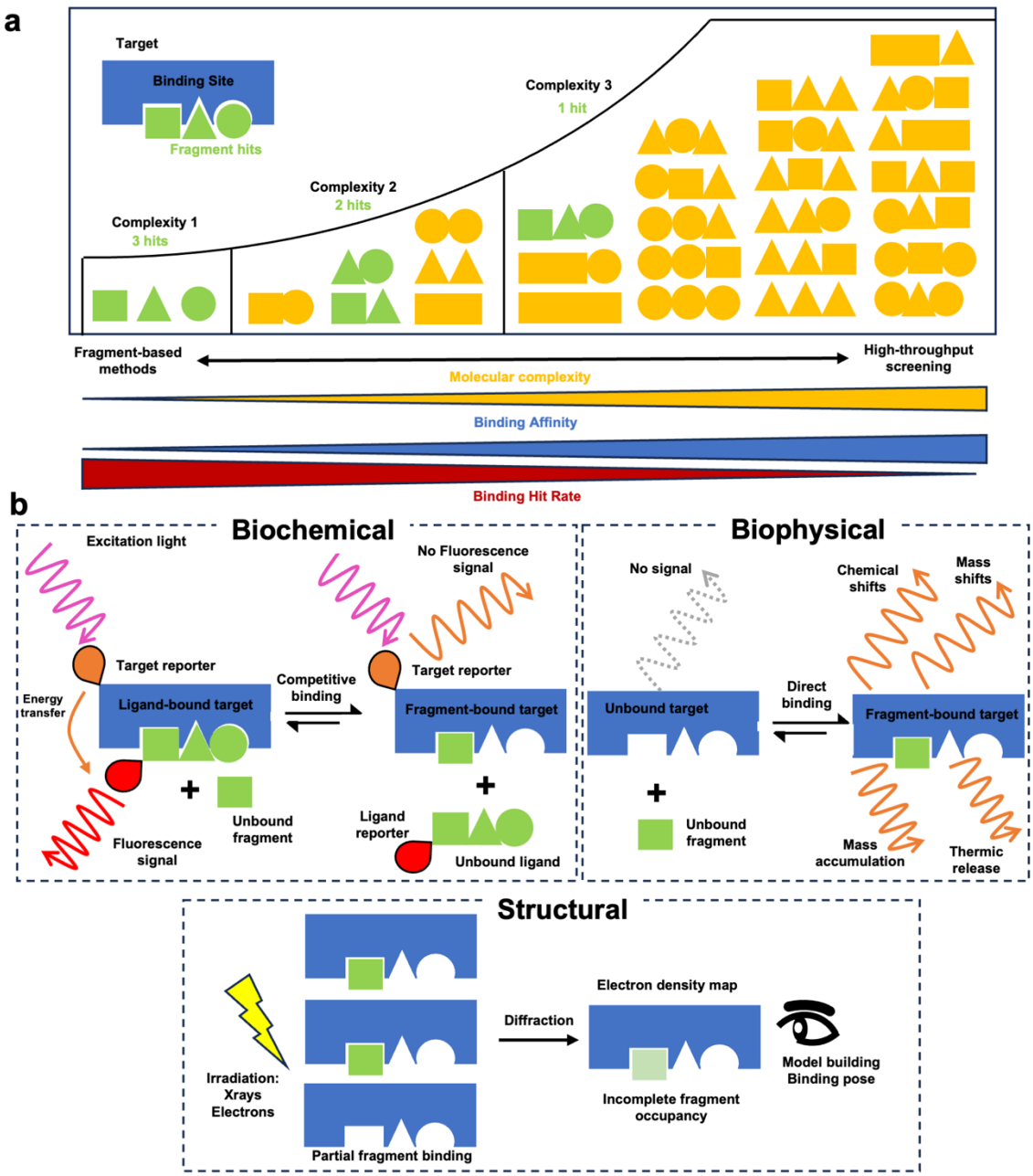

Figure 1: Foundational aspects of Fragment-Based Drug Discovery and fragment screening. a. illustrates how FBDD utilises simpler molecules than HTS, leading to higher hit rates. Screening "lower complexity" chemical spaces results in weaker binding affinities, due to a reduced coverage of the binding site, compared to the more complex molecules identified through HTS. b. Shows biochemical, biophysical, and structural methods can detect fragment-target interactions. Biochemical assays often rely on competitive displacement of a fluorescent ligand reporter, leading to changes in signal intensity due to energy transfer or quenching. Biophysical approaches detect binding through changes in intrinsic target properties such as mass (e.g., SPR, MS), heat (ITC), or local chemical environment (NMR), without the need for labelled ligands. Structural methods, such as X-ray crystallography or cryo-EM, infer protein-fragment conformations through diffraction data and electron density interpretation, allowing model building even with partial occupancy.

Design, Make, Test (DMT) cycles are iterative data-driven workflows that optimise initial hits into lead compounds **(Fig. 3)**. In this review we focus on DMT cycles within a typical FBDD workflow, starting from one or more initial fragment hits. The “Design” phase encompasses the step where fragment follow-up compounds are designed, often guided by SAR data and molecular models. Fragment hit follow-up design refers to the process of generating new analogues that aim to improve potency and map out the SAR landscape for lead development. Fragment modifications may also be prioritised accordingly to improve affinity, ADMET profiles,[33] synthetic feasibility and chemical stability[34]. The “Make” phase focuses on synthesising these designs in sufficient quantities and purities and also encompasses formatting compounds to ensure stability and compatibility with downstream experiments, adjusting parameters such as solvent and concentrations. In the “Test” phase, compounds are evaluated experimentally for properties such as biochemical activity, binding affinity, 3D conformation, selectivity, physicochemical parameters and key pharmacological properties. While traditional DMT steps were driven by medicinal chemistry insight and executed by specialists, they increasingly integrate computational modelling, not only to guide designs but also to help synthesis and interpret experimental data[35]. Each cycle iteratively refines compounds by reassessing objectives, strategically deploying appropriate tools (modelling, chemistry, assays) based on project stage and setting, and minimising bottlenecks to ensure successful hit-to-lead progression[36]. Industry laboratories often favour well-established, highly productive strategies regardless of monetary cost, while academic or early-stage ventures may instead adopt more experimental or cost-conscious approaches, leveraging emerging tools or theoretical frameworks to maximise impact within tighter constraints. Early DMT stages prioritise high-throughput, lower-cost approaches, while later stages leverage more targeted and data-rich tools to fine-tune the small molecule’s properties. This feedback loop enables continuous refinement with the aim of delivering potent lead compounds ready for preclinical evaluation[37,38].

Fragment hits, being chemically simpler than HTS hits, offer greater flexibility for hit optimisation and allow more creativity in improving affinity, selectivity, and physicochemical properties **(Fig. 5)**. However, their weak initial potency often necessitates additional rounds of DMT iterations **(Fig. 3)**. In contrast, HTS hits **(Fig. 1a)**,

while closer to lead-like compounds, may have more synthetic constraints due to their complex scaffolds, making modification more difficult. There are 3 types of fragment elaboration for follow-up design: growing (expanding one fragment see for example Murray et al.,[39]), linking (connecting two adjacent fragments – see Schuller et al.,[40] for example), or merging (combining overlapping fragments – see De Fusco et al.,[41] for example) **(Fig. 2)**[42]. Historically, one of the earliest and most influential fragment elaboration strategies was fragment linking, as exemplified by the "SAR by NMR" approach introduced by Shuker et al., in 1996[43]. In this *tour de force*, two fragments binding adjacent sites on the FK506-binding protein were identified using Nuclear Magnetic Resonance (NMR) spectroscopy. Each fragment was, first, individually optimised to establish basic SAR around it, then linked using spatial constraints, ultimately yielding a nanomolar binder. Although linking **(Fig. 2b)** can, in principle, result in substantial affinity gains, it remains challenging due to the strict geometric alignment required between fragments and the frequent loss of binding upon improper linkage[44,45]. Over time, fragment growing **(Fig. 2a)** emerged as a more practical and broadly adopted strategy. Growing is typically more tractable due to its incremental nature, the availability of SAR around a single chemotype, and greater compatibility with automation. In contrast, fragment merging **(Fig. 2c)** is technically more complex but has been proposed as a potentially faster route to potency if executed correctly [46].

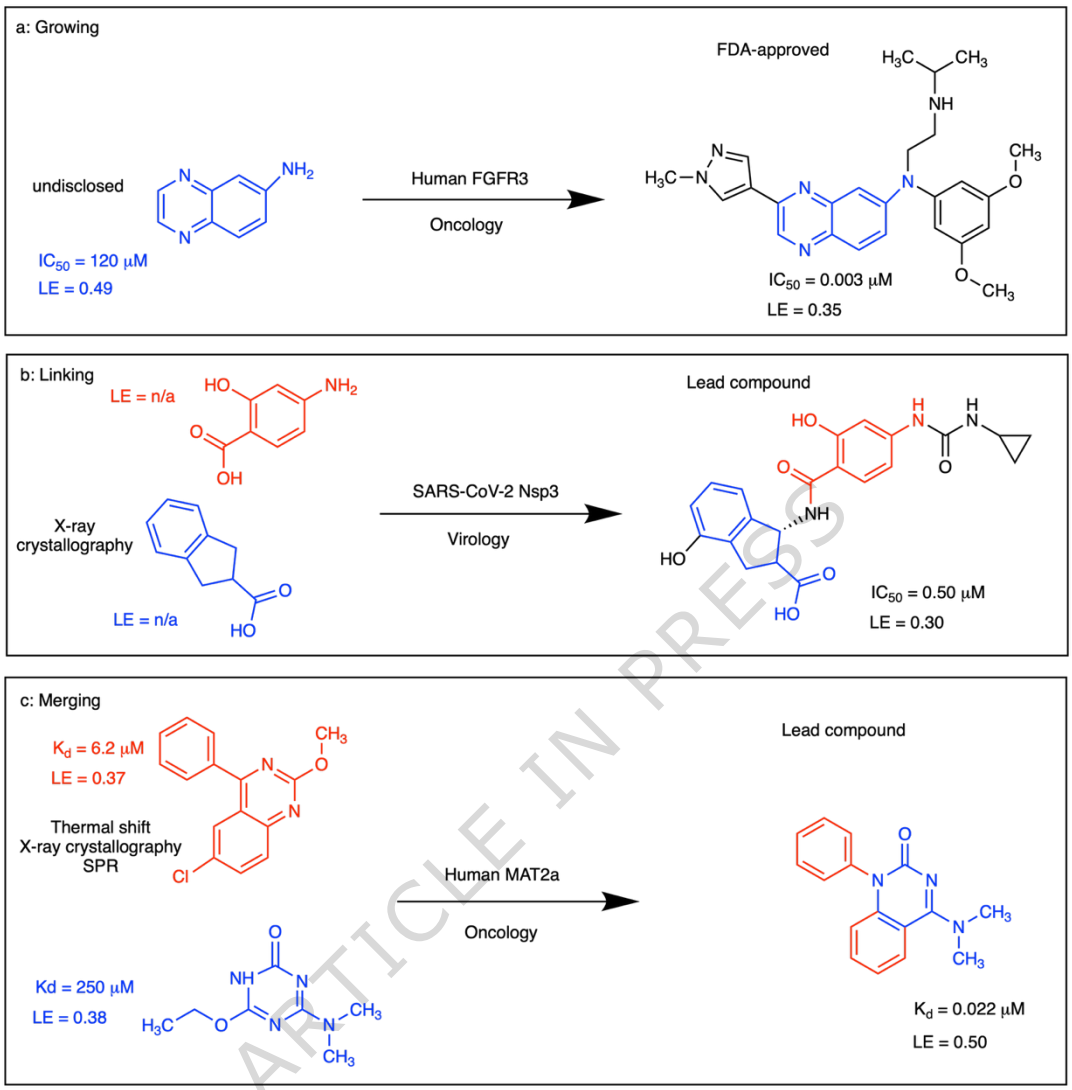

Figure 2: Fragment progression strategies of growing (a), linking (b), and merging (c). Examples indicate the initial screening strategy in black on the left, type of fragment progression, protein target, therapeutic area, affinities (IC_{50}, K_d), and resulting ligand efficiency (LE), defined as the ratio of binding affinity to the number of heavy atoms. Starting fragments are shown in red or blue, with additional decoration in black in the final compounds. LE=n/a for the second example because the fragments were obtained from crystallographic screening, which does not provide affinity information.

The DMT cycle is initiated after library design, screening, and hit validation which are stages that have been extensively reviewed elsewhere and are only briefly discussed here[27]. Here, the central challenge is to efficiently convert weak-affinity, low-complexity fragment hits into well-characterised, near-potent compounds suitable for further development. A key distinction we make between traditional medicinal chemistry-based DMT cycles and FBDMT lies in their starting point and objectives[47]. Classical SAR-based DMT optimisation generally involves smaller modifications to a relatively large, lead-like

scaffold to fine-tune potency, selectivity, or ADMET properties **(Fig. 5b)**. In contrast, FBDD DMT cycles must first chemically validate and establish a reliable SAR series around minimalist fragment hits that often lack clear binding modes, vectors for modification or sufficient binding affinity **(Fig. 5c)**. External data can also provide additional guidance, for example from other HTS or medicinal chemistry optimisation activities **(Fig. 5)**[48]. Similarly, novel fragment hits may provide a fresh outlook to compound series stuck at the optimisation stage[49,50].

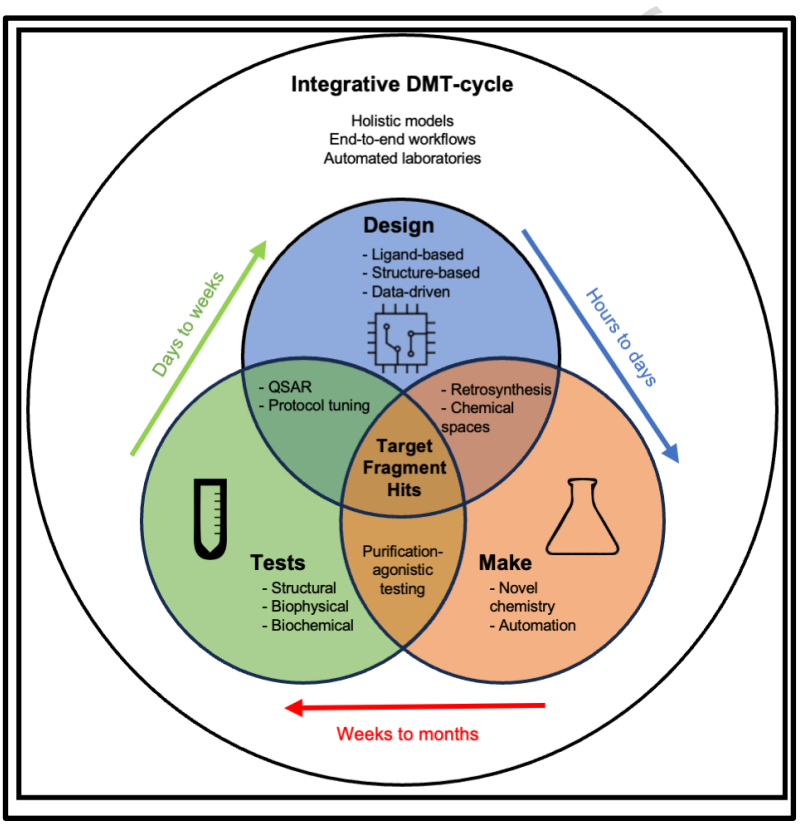

Figure 3: Efficient DMT cycles are required for optimal progression of fragment hits into lead compounds. DMT phases are interconnected, with overlapping regions spotlighting critical fields and elements conducive to methodological integration. The directional flow of the cycle is indicated by arrows, which also denote the typical duration required for each phase.

These DMT cycles are not merely sequential but continuous, with each stage shaping and informing the next through the methods and data it generates. The choice of design strategies, synthesis routes, and assay formats directly influences subsequent iterations,

making integration between stages critical for efficiency. Approaches may also span multiple stages, requiring fine coordination and integration across the cycle, as well as multidisciplinary expertise to navigate overlapping objectives and technical constraints. Moreover, the special demands of fragment progression have driven the development and repurposing of numerous enabling methods. Advances in structural biology methods now support structure-guided design even for challenging targets. AI/ML approaches also increasingly assist in proposing viable fragment hit elaboration strategies or prioritising designs[51] and (high-throughput) automated synthesis platforms shorten the Make phase, allowing more cycles within constrained timelines. Together, these innovations transform fragment-focused DMT into a sophisticated, technology-enabled workflow that effectively bridges the gap between initial fragment hits and the lead optimisation stage.

Outline of this review

Fragment-based drug discovery has been an active area of research for over two decades, and many reviews already cover the field's foundational concepts and advances[42,47]. These include discussions on screening methodologies[27], challenging protein targets[8], therapeutic area applications[52], synthesis[4], structural biology[53], and structure-agnostic follow-up strategies[54]. Simulation-based[55], AI-assisted [56] and other *in silico* approaches[57] to fragment identification and expansion have also been explored, as well as recent medicinal chemistry success stories. A foundational book by Erlanson and colleagues compiles historical and conceptual FBDD insights[58]. Rather than reiterating this substantial body of work, this review aims to contextualise FBDD within early fragment-based DMT (FBDMT) cycles, highlighting how emerging techniques are actively reshaping how low-affinity fragment hits are transformed into tractable chemical leads. In particular, we focus on how different aspects of FBDD integrate across DMT stages into dynamic and iterative processes. The aim is therefore not to examine any single area in exhaustive detail, but to provide a comprehensive and modern outlook on the field's current trajectory.

This review's perspective is shaped by the authors' core expertise in computer-assisted ligand design, structural biology (notably crystallography), biophysical assays, and purification-agnostic 'direct-to-biology' evaluation techniques as applied to FBDD.

Consequently, recent advances in these domains are explored in greater depth. Given the inherently holistic and interdisciplinary nature of drug discovery, we also refer to artificial intelligence, medicinal chemistry innovations, assay technologies, and synthetic methods, where relevant.

We use a DMT-centric perspective highlight the unique challenges of fragment-based hit progression and how it can deliver lead-like series ready for classical medicinal chemistry optimisation. A recurring theme throughout this review is also how the inherently low affinities of fragment hits complicate both experimental validation and computational modelling, limiting the applicability of strategies developed for larger, drug-like molecules. We begin with a discussion on Fragment Library Design, emphasising key factors that define screening efficiency and the utility of fragments for subsequent elaboration. Next, we discuss how one can obtain and select fragment hits, focusing on the methodologies used to identify and validate them while addressing the inherent challenges posed by the weak binding affinities of fragments. We also include a subsection on covalent fragments in FBDD, reflecting growing interest in fragment libraries functionalised with reactive warheads. Following these pre-DMT stages, we then introduce DMT as an adaptive knowledge acquisition paradigm for compound optimisation, highlighting recent trends including automated labs and open-source collaborations. Following this, the Design Phase **(Fig. 3)** section discusses computational methods to prioritise follow-ups. We also explore recent developments in machine learning to dynamically transfer complex scoring schemes to ligand-based methods. The Make phase **(Fig. 3)** highlights advancements in synthetic and high-throughput approaches around fragments, which together facilitate faster and broader chemical exploration around hits. We also discuss bottlenecks related to simplicity of chemical reactions, solvent usage, scalability and purification. In the Test phase **(Fig. 3)**, we focus on recent developments in purification-agnostic approaches (sometimes referred as “direct-to-biology” methods), including both structural and kinetic methods, allowing for rapid validation of fragment follow-ups, although potentially at the expense of data quality. Although several of these assays may also be used for initial fragment screening, we discuss them in the “Test” section as they synergise especially well with follow-up evaluation workflows and feed results directly back into the Design phase. We also explore synergies between FBDD and other methodologies, such as

DNA-encoded libraries. While several recent computational and automation-driven techniques are highlighted across all DMT phases, we acknowledge that many remain at early stages of practical adoption. Our aim is to review these emerging approaches not as universally established standards, but as promising strategies whose broader impact may unfold over time. Finally, we reflect on the challenges and future directions in FBDD, as well as strategies for optimising the fragment hit-to-lead transition.

Foundations for FBDMT: fragment library design and selection

All FBDD campaigns are initiated by the screening of a fragment library to identify hits **(Fig. 1b)**. Fragment library design (or selection) involves composing a diverse collection of molecules that satisfy fragment-like criteria and is therefore critical for enabling compound progression via DMT cycles[16]. A well-designed fragment library maximises the chances of discovering initial hits by covering broad chemical diversity whilst ensuring fragments have suitable physicochemical characteristics and synthetic tractability.

Designing fragment libraries for efficient integration with DMT stages

The library should cover a diverse range of pharmacophoric combinations (i.e., different arrangements of key functional groups such as hydrogen bond donors/acceptors, aromatic rings and charged moieties) distributed across fragments to explore a broad spectrum of interaction types. This can include fragments bearing pharmacophores, enabling them to interact with diverse binding sites or pocket shapes. As such fragment screening can sample binding site interaction opportunities more exhaustively, revealing information that can then be employed in subsequent “Design” phases. This functional diversity is crucial because structurally different fragments can often exploit the same interaction patterns, leading to functional redundancy[59,60].

Employing fragments that permit easy follow-up synthesis at the “Make” phase, thus providing good access to the neighbouring chemical space, is crucial for efficiently SAR data and increasing the likelihood of finding potent follow-up compounds. For example, fragments from the Diamond–SGC–iNEXT (DSI)-poised library display chemical moieties with “poised” bonds, meaning these bonds are strategically selected to enable rapid and versatile diversification using robust, high-yielding and widely applicable reactions using commonly accessible building blocks. These poised bonds also facilitate fragment

deconstruction, allowing one to dissect, elaborate and analyse key substructures for deeper insights into SAR **(Fig. 4a)**[61]. The DSI-poised was employed in the development of fragment follow-up inhibitors against the second bromodomain of the pleckstrin homology domain interacting protein (PHIP2), a protein overexpressed in lethal cancers[62]. This approach allowed rapid synthetic expansion yielding the first compounds with measurable IC_{50} values paired with good LEs and structural data[61]. Fragments libraries can also be designed to facilitate experimental testing. For example, the FragLites library uses halogenated compounds that enhance diffraction signals and hence facilitate pose determination by X-ray crystallography as demonstrated in a study with cyclin-dependent kinase 2 **(Fig. 4b)**[63].

Fragment libraries can also be tailored to improve drug-like properties and provide better starting points for lead optimisation. This includes reducing planarity and increasing sp^3 carbon content, which is associated with improved solubility and pharmacokinetics[64] (**Fig. 4c**), though improvement of properties is context-dependent and also influenced by factors like crystal packing and solvation behavior[65]. Recent analyses has, however, challenged the predictive value of the sp^3 fraction metric in predicting whether a compound will succeed in development[66]. In addition, the relatively small size of fragment libraries, compared to HTS, enables for rigorous quality control, including purity checks, solubility assessments, and redox activity screening, to ensure that only high-quality fragments are retained for assays. This helps avoid issues such as compound degradation or non-specific binding[67].

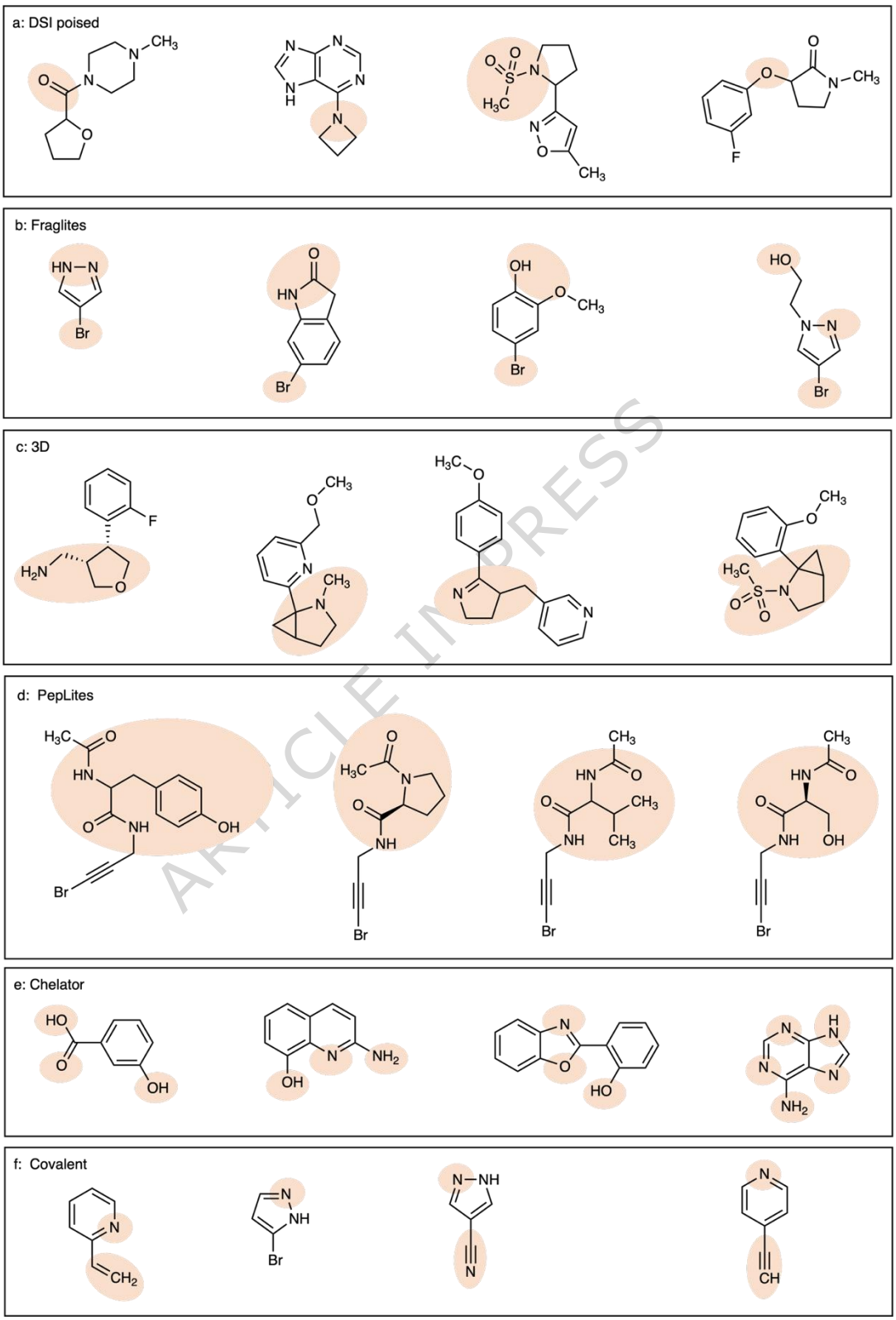

Figure 4: Illustrative library fragments for efficient DMT cycle integration and functional binding. Highlighted zones indicate key functionalities of illustrative library fragments. (a) Poised bonds enable rapid synthesis and decomposition. (b) Hydrogen bond donor-acceptor doublets and halogens enhance electron density fitting for X-ray crystallography. (c) Fragment library with increased three-dimensional (3D) complexity to solubility, pharmacokinetics, and specificity. (d) Peptide bonds and side chains mimic amino acid interactions.(**e**) Chelating atoms target metalloproteins through metal coordination. (f) Electrophilic warheads and electron withdrawing aromatic atoms enable covalent interactions for targeting reactive protein residues.

Functionally relevant fragment libraries for precision screening

The library can also be selected to ensure the fragment hits are pertinent to the biological context such as the PepLites library that mimics amino acid binding motifs **(Fig. 4d)**. The PepLites library was screened against bromodomain proteins BRD4 and ATAD2 enabling the efficient development of lead-like follow-up compounds that recapitulate interactions of natural polypeptide bromodomain binders[68]. This target-pharmacophore paradigm for fragment library design can be applied to any family of drug targets with conserved binding site features[69], such as kinases, which are a well-known class of drug targets with a conserved hydrogen bond donor and acceptor motif. Similarly, when dealing with metalloprotein it is also possible to screen fragments that have chelating capabilities hence promoting functionally relevant binding at the metal coordination centre **(Fig. 4e)** [70].

The covalent minifrags library, which selectively target cysteine residues often found in protease active sites, provides another example of a precision screening library **(Fig. 4f)**[71]. These fragments were recently used in a crystallographic and electrophilic fragment screening of the SARS-CoV-2 main protease, identifying merging opportunities of covalent and non-covalent fragment hits leading to a potent lead series[72]. Covalent fragments have also led to the identification of Sotorasib[73], an FDA-approved drug, against non-small cell lung cancer with the compound irreversibly attached to a cysteine at an allosteric binding site[74].

Foundations for FBDMT: hit finding and validation

An essential step in FBDD is the identification of hits through screening, typically from dedicated fragment libraries. While FBDD has the advantage of efficiently sampling chemical space, this comes at the cost of working with low-affinity binders. As such, one of the trickiest aspects of FBDD is to apply sensitive strategies that can reliably and unambiguously identify these weak binding events[29]. Several complementary methods exist to detect fragment binding, each offering distinct types of readouts, ranging from structural to kinetic or thermodynamic, shaping how subsequent follow-up design is conducted (**Fig. 1b**). Importantly, many of these same assays can be re-used later for follow-up validation in a context-, project-, and target-dependent manner - for example,

evaluating follow-ups directly as crude reaction mixtures by NMR, SPR, or crystallography but these applications are covered in later sections.

Obtaining fragment binding pose information with structural assays

Structural assays provide full three-dimensional information about both the target and ligand conformation. While they do not directly inform on binding affinity, these structural insights are invaluable for guiding fragment follow-up designs and downstream applications such as docking, rescoring, or molecular dynamics simulations[53].

X-ray crystallography provides atomic-scale insights into binding modes (**Table 1**)[75]. Usually, crystals are soaked with high-concentration ligand solvent stocks for high-throughput evaluation[76-78]. This process generates many diffraction datasets (**Fig. 1b**) that can be collectively processed to unmask low occupancy binding events, increasing hit rates[79], although obtaining suitable crystals for screening can still be prohibitively laborious[80]. The soaking approach is practical for fragment screening, as a single crystal form can be used across many compounds without the need to re-optimise crystallisation conditions for each[81]. However, soaking has limitations: the crystal packing must permit solvent channel access for the fragment to reach the relevant binding site or “hot spot”. Moreover, the crystal system itself should represent a solution-like low-energy conformation of the protein; otherwise, fragments may bind in non-physiological modes or fail to bind altogether [82]. A high-throughput X-ray crystallographic fragment screening, via soaking, approach was used to identify weakly binding hits against the SARS-CoV-2 main protease (Mpro)[72]. It revealed 71 fragment hits, including covalent binders, at the catalytic site and the dimer interface with hits cross-validated via mass spectrometry (MS). This study also showed that different crystal systems can yield different binders. It is important to note that fragments resolved crystallographically can yield no signal in solution assays[82-86]. This discrepancy likely arises because X-ray crystallography can detect weak binders at millimolar affinities due to high soaking concentrations, which exceed the solubility or sensitivity limits of most solution-phase assays[87]. Moreover, orthogonal assays operate under distinct physical principles and experimental conditions, such as buffer composition or protein immobilisation, which can affect the binding detection landscape and reduce hit overlap between techniques [84].

Cryo-electron microscopy (EM) can now also be applied for fragment studies[88] pas it is an attractive alternative to investigate physiologically relevant conformations[89] (**Table 1**). However, Cryo-EM is not yet scalable, but does have the potential to resolve important but difficult drug targets such as membrane proteins[25]. Cryo-EM is better suited to larger complexes and difficult targets such as membrane proteins, with routine applications for >150–200 kDa assemblies and, with careful specimen preparation, useful maps for ~80–100 kDa proteins; scaffolding strategies can extend below this range[90]. Currently, cryo-EM resolves more poorly than X-ray crystallography implying more ambiguous fragment poses possibly confounding designs[26]. However it should be noted that resolution has a different meaning between the two techniques resolution: in crystallography, it reflects diffraction limits from ordered lattice planes, whereas in cryo-EM, it refers to internal consistency within reconstructed particle images, so care should be given when comparing resolutions[91]. Pose visibility is highly resolution-dependent: unambiguous small-molecule placement typically requires local resolution $\lesssim$2.5–3.0 Å; side-chain modelling that supports ligand interpretation is more reliable at ≤3.2–3.5 Å, whereas 3–4 Å maps often leave fragment poses ambiguous[92,93]. In such cases, modeling can assist (density-restrained docking/real-space refinement, MD-based flexible fitting, and emerging ML tools), but these methods cannot substitute for insufficient map signal and require stringent validation (map–model metrics, alternative hypotheses) and orthogonal confirmation[94,95]. Progress is nonetheless being made in pushing the limits of cryo-EM to resolutions and throughput [96] useful for structure-based drug discovery and integration with AI/ ML systems hence, poising it to become an important FBDD tool in the future[97].

While full 3D structural models of protein–ligand complexes can also be obtained using Nuclear magnetic resonance (NMR) spectroscopy experiments, this is not typically performed due to practical limitations such as size constraints, isotope-labeling requirements, and complex data interpretation. Hence NMR is better suited to identifying interacting residue from chemical shifts indicating approximate binding location and nature of interactions. This implicit structural information can subsequently be mapped onto pre-existing structural models (**Table 1**). For example, in the development of Myeloid cell leukemia 1 (Mcl-1) inhibitors, the primary fragment screen was performed using NMR[98]. Resulting hits were further examined with advanced NMR techniques, enabling

the mapping of interactions between fragment hits and the binding site[99]. These interactions guided binding pose predictions and subsequent structure-based fragment follow-up designs. These models were crucial for obtaining SAR data, ultimately leading to the development of a potent and selective inhibitor, S64315, which progressed to clinical trials[100]. Moreover, recent developments in automated chemical shift assignment pipelines and chemoinformatic integration have made the analysis of protein-ligand interactions by NMR increasingly scalable and accessible to non-expert users[101].

Besides providing ligand poses, structural assays are also useful in inferring protonation states, informing water molecules conservation, or resolving novel conformational states[12,102-105], information crucial for accurate structure-based modelling. For example, a study compared room-temperature to cryogenic crystallographic fragment screening against PTP1b, which revealed altered binding modes, changes in solvation, and novel binding sites for room-temperature screens, thus providing a comprehensive panel of information for rational structure-based follow-up design (discussed later). Structural methods can resolve covalent bonds formed between electrophilic fragments and nucleophilic residues, providing unambiguous information on warhead positioning and residue engagement. Structural assays also delineate the landscape of ligandable pockets, whether these are functionally critical (e.g., catalytic or allosteric sites) or structurally favorable but non-functional[12]. Fragments often bind to biophysically favorable “hot spots” that may not directly modulate target activity[106]. To ensure meaningful follow-up design, it is crucial to validate structural fragment hits thoroughly, as pursuing binders at non-functional sites may lead to “silent binders” which are biophysically potent compounds with no biochemical activity. However, these “silent binders” may represent good candidates for PROTAC development[107]. In a “structure-first” paradigm, PAINS-type interference (**Table 2**) is intrinsically reduced because the evidence of binding derives from a physically plausible, well-supported pose (clear difference density, sensible geometry/interactions, appropriate B-factors/RSCC) rather than from assay signal generation. While structural methods still require careful model validation and functional follow-up, a robust pose provides strong reassurance that the interaction is specific rather than an assay artefact[108].

Although structural techniques offer useful 3D insight into ligand binding poses, they often require substantial experimental investment, such as high yield protein expression, optimisation of crystallisation or EM gridding, or isotopic labelling making them harder to implement than solution-phase assays, which can often be deployed with greater ease across a broader range of conditions and targets.

Solution assays for measuring and confirming weak fragment binding

Solution assays are experimental techniques performed in solution to measure either the biochemical activity or biophysical interaction of compounds with the drug target. These assays can provide information on activity, affinity, or binding kinetics in a native-like environment, without the constraints imposed by crystalline lattices or immobilisation on surfaces as in most structural methods. A variety of such assays are available to measure fragment binding, chosen based on throughput, sensitivity, and target characteristics and the type of data acquired, balancing practical accessibility in academia with instrumentation-rich environments in industry **(Table 1)**. In practice, the choice of assay is often guided by logistical and practical constraints rather than accuracy alone.

For instance, surface plasmon resonance (SPR) is an optical technique that monitors interactions between fragments and an immobilised protein on the sensor surface via refractive index changes. It is widely used in FBDD owing to its high sensitivity, moderate protein requirements and its ability to screen at high analyte concentrations to detect weak fragment interactions[109] (typically low–high micromolar) **(Table 1)**. However, SPR readouts can be perturbed by non-specific mass accumulation and surface immobilisation effects, which warrants careful controls and orthogonal validation[110]. In addition to extracting affinity and kinetic constants, qualitative features across a dilution series, including maximum response amplitude and sensogram curve shapes help distinguish true binders from promiscuous compounds[111].

Techniques such as NMR, SPR, and isothermal titration calorimetry (ITC) can directly provide binding affinity, whereas others like thermal shift assays have more implicit readouts (**Table 1**)[112-114]. Solution assays are also useful at cross-validating hit from structural readouts[85], to gain additional information about their binding affinities, and, where possible, cross-validate these positives structurally to guard against assay-specific

artefacts. Concordant readouts across independent solution and structural modalities provide orthogonal validation, increasing confidence in true target engagement, enabling early triage of artefactual or promiscuous binders, and helping to prioritise series for progression. Notably, solution-phase biochemical and some biophysical assays are also susceptible to PAINS mechanisms (e.g., fluorescence quenching/enhancement, enzyme-coupled steps, surface-induced aggregation) and therefore benefit from routine counter-screens (detergent, redox controls) and substructure filters prior to progression [115].

Overall, NMR is widely regarded as a gold standard for assessing fragment binding against small to medium size targets in solution and has played an historical role in FBDD since the introduction of the "SAR-by-NMR" approach[43]. NMR titrations can provide both affinity information and residue-level interaction maps through chemical shift perturbations, making it uniquely suited for hybrid quasi-structural and affinity mapping **(Table 1)**. NMR can also screen fragment cocktails, performing multiple experiments at once, increasing throughput. For example, in the combined workflow described by Johnson et al.,[116], fragment cocktails were screened using protein-observed and ligand-observed NMR techniques, allowing for time-efficient screening across multiple proteins. This multiplexing approach enabled efficient hits identification, followed by deconvolution using protein-resolved NMR to pinpoint specific fragment-protein interactions[116]. For fragment-hit validation purposes, both ligand- and protein-observed NMR are routinely deployed[117] as orthogonal confirmation to verify direct binding and to eliminate false positives originating from biochemical and biophysical readouts caused, for example, by fluorescence interference or immobilisation artefacts.

Mass spectrometry (MS) is particularly suited for studying covalent fragments **(Table 1)** due to the non-reversible nature of the interaction with the target, expressed as clear mass shift readouts[118]. This advantage was demonstrated in a study where a library of electrophilic fragments was screened against cysteine-containing proteins using intact protein LC-MS. For instance, covalent fragments targeting OTUB2 and NUDT7 formed covalent bonds with the target cysteine residues, enabling direct readouts of binding events and subsequent structure-based optimisation to potent inhibitors[119].

Table 1: Comparison of fragment screening methods commonly used in FBDD workflows. Methods are classified by type, throughput, sensitivity limit (as the weakest binding affinity typically detectable), sample requirement per assay unit, readout, confidence level, and recommended use within the drug discovery cycle. The table highlights suitability for primary screening, hit validation, binding mode elucidation, kinetic and thermodynamic profiling, and design support. All the numbers reported are indicative of order of magnitude and may be subject to variation due to target dependence.

	Type	Throughput	Sensitivity Limit	Amount of target sample	Readout	Confidence	Us
	Structural	~100–1000 per week	Mid mM	~1-10 µg per drop (single use or cocktail)	X-ray diffraction, Electron density	High (clear binding mode, false negatives if low occupancy)	Binding m Reso identifyin
	Structural	~10–50 per week	Low–Mid µM	~1–5 µg per grid (single use)	Electron diffraction, Electron density	Moderate (resolution limits, may miss weak binders)	Binding m Diffi
	Biophysical	~100–1000 per day	High mM	~0.1-0.6 mg per tube (single use or cocktail)	Ligand chemical shifts	High (low false positives, may miss very tight binders)	Primary v
	Biophysical/ Semi-structural	~10–50 per day	High mM	~0.5–1 mg per tube (per titration)	Protein chemical shifts	High (site mapping, low artifacts, Decreases with protein size)	Hit valida mapping
	Biophysical	~100–1000 per day	High mM	<1 µg per sample (single use)	Mass/charge shifts	High (careful control needed)	Prima Covale
	Biophysical	~10–50 per day	Mid µM	~0.3–1 mg per well (single use)	Heat release/ absorption	High (quantitative, false negatives for weak binders)	Thermod Affi
	Biophysical	~100–1000 per day	Mid µM	~1–10 µg per well (single use)	Melting temperature shifts	Moderate (artifacts, needs confirmation)	Prima Orthogo
n	Biophysical	~100–1000 per day	Mid µM	~25–50 µg per chip (reusable)	Mass accumulation, Real-time kinetics	High (immobilisation artifacts possible)	Kinetic p ranking, C
l	Biophysical	~500–1000 per day	Mid µM	~10–50 µg per immobilisation (reusable)	Mass accumulation, Real-time kinetics	High (resolves fast off-rates, immobilisation artifacts possible)	Kinetic p ranking, C
	Biophysical	500–1000 per day	Mid µM	~10–50 µg per immobilisation (reusable)	Optical interference Kinetics	Moderate (some sticking artifacts)	Hit vali
s	Biophysical	~100–500 per day	Mid mM	<0.01 µg per well (single use)	Thermophoresis	Moderate–High (fluorescence artifacts)	Hit vali
	Biochemical	~1000–10,000 per day	Low µM (competition)	~10–100 ng per well (single use)	Luminescence	Low–Moderate (signal interference)	Prima Functi
y	Biochemical	~1000–10,000 per day	Low µM (Functional IC_{50}s)	~0.1–1 µg per well (single use)	Activity readout	Moderate (PAINS risk)	Primary sc activity, Fu
	Biochemical	~1000–10,000 per day	Low µM	~10–100 ng per well (single use)	FRET signal	Moderate (labelling artifacts)	Primary sc activity, Fu
	Biochemical	~1000–10,000 per day	Low µM (competition)	<0.1 µg per well (single use)	Polarisation change	Moderate (fluorescence interference)	Primary sc activity, Fu

Challenges and opportunities of computational approaches for fragment prioritisation

It is in theory possible to use computational methods to prioritise fragments, but this still appears challenging to do quantitatively as submissions to the 7th Statistical Assessment of Proteins and Ligands (SAMPL) challenge revealed poor performance[102]. The authors argued that recovering crystallographic fragment hits using computational methods was hard likely due to, low to no activity, paired with small sizes resulting in flat energy landscapes making scoring functions unable to recover binders and poses correctly. Additionally, as also demonstrated elsewhere, selecting water molecules that enable accurate docking predictions appeared challenging[120].

There have, however, been examples of prospective computational fragment identification with methods including structure-based screening and molecular dynamic (MD) simulations. For example, Protein kinase CK1δ fragment hits were discovered via docking of a fragment database with MD simulations employed to refine binding poses. The selected compounds were then validated using an enzymatic assay resulting in the identification of novel scaffold with sub-micromolar activities[121]. Similarly, long-timescale MD simulations, likely inaccessible to most MD users due to high computing demands, prospectively discovered binding poses for fragments targeting tyrosine phosphatase 1b (PTP1b) novel allosteric sites with predicted binding poses confirmed with X-ray crystallography[122]. Such computationally demanding modelling studies are enabled by purpose-built platforms, such as the Anton supercomputer [123] or Folding@home[124], which are generally available to specialised academic laboratories or well-funded organisations. Supervised molecular dynamics (SuMD) accelerates MD-based binding pose predictions. Through iterative simulation and frame selection, SuMD guides fragments toward the binding site, enabling efficient exploration of fragment binding poses. Ferrari et al.,[125] used HT-SuMD to screen fragment hits for the Bcl-XL oncological target, dynamically sampling the binding site and prioritising binders, which were validated via NMR[125]. Furthermore, fragments screened via biophysical or biochemical methods lack pose information, hindering follow-up design. Computational approaches, like MD simulations coupled with Hidden Markov Models analysis, address this limitation more effectively than traditional docking, which struggles to identify correct low-energy poses[126]. Recent work

by Schmitz et al., (2024) showed that multiple short, unbiased MD replicas of fragment diffusion can recover binding poses and estimate absolute binding free energies. This was achieved by counting bound versus unbound configurations and applying the law of mass action to estimate equilibrium constants, enabling practical fragment prioritisation[127].

Expanding the FBDD toolbox with covalent fragments

Covalent fragments bear electrophilic "warheads" capable of irreversibly reacting with nucleophilic residues in a protein. This covalent tethering enhances effective affinity of fragment by “stapling” fragments to a reactive residue, therefore addressing core challenges relating to fragment screening and FBDD: low affinity and unspecific binding[128].

Early campaigns primarily targeted cysteine due to its high nucleophilicity and low natural abundance, with acrylamides and chloroacetamides being popular warheads[129]. More recent developments have broadened the scope: sulfonyl fluorides can target tyrosine[130,131], while boronic acids and aldehydes can engage serine[132] and lysine[133], respectively. These novel warheads are “tuned” to be reactive enough for labelling specific residues without indiscriminately reacting with off-target nucleophiles, thereby improving selectivity and reducing toxicity risk[134]. Chemists must also ensure stability of covalent fragments: some warhead-bearing fragments are unstable in storage or physiological conditions. For example, certain sulfonyl fluorides can hydrolyse in DMSO or water, although they have the added advantage of being usable moieties for combinatorial chemistry [135].

Screening covalent fragments often requires specialised methods. Intact protein MS is commonly used to detect covalent adducts by measuring protein mass shifts, with tandem MS enabling the mapping of the labeled residue[136]. Traditional fragment screening assays (NMR, SPR, thermal shift, etc.) can also be adapted . For instance, differential ^{19}F-NMR or changes in protein melting temperature may indicate covalent adduct formation, although irreversible binding violates some kinetic assumptions of standard analyses[137]. Biochemical assays **(Table 1)** such as time-dependent enzyme inhibition can also flag covalent interactions[138]. Crystallographic soaking is also applicable, particularly for

confirming covalent bond formation and fragment orientation from electron densities[72]. Because covalent binding violates equilibrium assumptions, careful validation is needed. Furthermore, hit triage is less straightforward – a covalent "hit" doesn't always mean a meaningful interaction if it simply reacts with available amino acids non-specifically. Typical controls include mutation of the targeted residue or profiling against reactive thiols such as glutathione. Techniques like quantitative irreversible tethering (qIT) allow kinetic comparison of fragment reactivity against protein versus nonspecific thiols[139]. Another consideration is that covalent fragments add complexity to ADMET profiles. The irreversible binding can lead to long residence times that complicate dose optimisation and may trigger proteolysis or immune clearance of the adduct-laden protein.

A recent campaign targeting the peptidyl-prolyl isomerase Pin1 highlights this strategy's power. Dubiella et al., identified covalent hits using MS-based fragment screening, leading to Sulfopin, a selective inhibitor of Pin1 that covalently modifies Cys113 via a sulfolane, chloroacetamide warhead. Sulfopin, now a potential therapeutic for various cancers, demonstrated strong selectivity, low toxicity, and tumor regression in mouse models, despite Pin1's challenging binding site[140]. This example illustrates how weak-binding fragments, combined with warhead tuning and careful optimisation, can yield potent covalent inhibitors against elusive targets. Overall covalent fragments are an increasingly important class of tool compounds, including for traditionally "undruggable" targets with shallow or dynamic binding pockets[141].

Key factors in selecting fragment hits for follow-up

Identifying numerous fragment hits is common, with hit rates in FBDD being typically higher than for HTS and selecting which fragment(s) to follow-up on is a pivotal step that has a significant impact on the project's success. Historically, elaborations of selected fragments were performed by medicinal chemists that had an instinct for compound viability and over the years retrospective analysis has also provided insight into which factors define "good" fragments (**Table 2**). For example, favourable binding thermodynamics[142], which involves enthalpy-driven binding through directed polar interactions, leading to more specific and stable binding to the target, also sometimes via the rearrangement of water molecules[106,143] has been explored in detail. The desolvation

of apolar groups in fragments contributes favorably to binding entropy, enhancing binding affinity, while non-polar/hydrophobic interactions, driven by dispersion forces, may also contribute to binding affinity[143]. High LE is desirable, with a value >0.3 kcal·mol^{-1}·heavy atom^{-1} considered optimal, as it reflects a favorable free energy contribution to target binding per atom[144]. This highlights the importance of solution assays readouts to select fragments as they enable LE determination (**Table 2**). Ranking co-structures by computed binding free energies may also provide surrogate binding values to assay data.

Fragments should be easily modifiable to enable rapid and efficient chemistry during the hit-to-lead optimisation stages (**Table 2**) and must allow for elaborations at chemical vectors to increase affinity and modulate desired characteristics without causing clashes with the binding site. Another key consideration is the biological relevance of the pocket where the fragment binds. Fragments located in functional or allosteric pockets are often prioritised over those in solvent-exposed or peripheral sites with limited optimisation potential. Hotspots defined by pockets also influence selectivity—especially in families like kinases, where targeting allosteric over conserved ATP sites can improve specificity[145]. Early identification of such sites via structural assays, supported by orthogonal validation (e.g., site-directed mutagenesis), can guide hit progression toward chemically tractable and target-specific regions[146]. Hence, the 3D binding pose is therefore a helpful driver, highlighting the added value of structural information to inform chemical optimisation.

Despite these guidelines, there is no one-size-fits-all approach to fragment selection. The choice of fragments remains context-dependent, requiring careful consideration of the biological target , available experimental or computational capacity, and the project objectives with the available data, such binding pose, affinity and activity (or absence thereof), further limiting the fragment selection process[36]. Other variables may also be considered now including intellectual property or medicinal chemistry restraints. Several fragments could also be collectively selected to provide a data subset for virtual screening, for example via pharmacophore searches[147]. While it may seem attractive to systematically explore follow-ups for all fragment hits, such approach is not practical and likely undesirable considering not all hits may represent high-quality precursors[4].

Table 2 Overview of common guidelines for selecting fragment hits for progression into lead-like compounds. Each property is described with its significance, the stage(s) of the Design-Make-Test cycle where it is most relevant, and its importance classified as Essential, Important, or Nice to Have. Importance levels reflect how readily the property can be modulated later through synthetic elaboration or addition during fragment-to-lead optimisation.

Property	Description	Origin	DMT Stage	Importance
Favourable Thermodynamic Signature [148,149]	Enthalpy-driven binding via polar contacts and desolvation of apolar groups, with low entropy contributions from dispersion forces enhancing binding.	Derived from binding studies and retrospective hit analysis.	Design Test	Important
Ligand Efficiency (LE) [19]	High ligand efficiency (≥0.3) ensures optimised binding energy relative to molecular size, crucial for effective fragment progression.	Retrospective hit analysis and structural optimisation principles.	Design Test	Important
3D Complexity [150]	Fragments with sp^3-hybridised carbons and out-of-plane functional groups improve target specificity and enable diverse interactions.	Advances in library design and chemical diversity studies.	Design	Nice to have
High Solubility [151]	Fragments with solubility >10 mM in DMSO facilitate high-concentration screening, crucial for robust assay performance.	Screening requirements for diverse experimental conditions.	Test	Important
Synthetically Sociable [4]	Fragments with robust synthetic methods enabling growth vector modification and abundant analogues for rapid SAR exploration, streamlining hit-to-lead optimisation.	Concept of sociable fragments derived from synthetic and structural analyses.	Make	Essential
Hotspot Binding [145]	Fragments target protein hotspots with complementary electrostatics and interaction, though hotspot validation is often difficult early on.	Protein-fragment interaction and electrostatic modelling studies.	Design Test	Important
Lack of PAINS and Toxicophores [152]	Fragments excluding reactive groups, toxicophores, and PAINS ensure assay reliability and avoid false positives.	Computational filtering and experimental validation for fragment libraries.	Design Test	Nice to have
Diversity in Chemical Space [59,153]	Libraries should include diverse pharmacophoric and shape profiles to maximise exploration of binding potential.	Library design principles aimed at enhancing discovery efficiency.	Design	Important
Stability and Reactivity [154,155]	Chemically stable fragments minimise aggregation or degradation under screening conditions, ensuring reliable results and follow-up chemistry.	Empirical testing for robust performance across conditions.	Make Test	Essential
Hydrophobic and Hydrophilic Balance [156,157]	Fragments should exhibit polar features for enthalpy-driven binding, while allowing scope for hydrophobic additions to enhance entropy during optimisation.	Derived from successful fragment-to-lead progressions.	Design Test	Nice to have

Fragment-based DMT cycles

Once fragment hits are selected, they serve as starting points for FBDMT cycles **(Fig. 3)**. The DMT cycle is an iterative process that organises the progression of initial hits into leads with desired characteristics **(Fig. 5)**. The cycle involves three core steps: Design, where modifications to the chemical structure are proposed, typically based on data-driven insights; Make, where the designed molecules are synthesised; and Test **(Table 1)**, where the compounds are evaluated for activity, affinity, pose and/ or other relevant properties **(Fig. 3)**. Each iteration proposes optimised chemical structures and, as an intrinsic trial-and-error process, generates SAR data to guide subsequent designs. By continuously refining the chemical space explored, DMT cycles aim to rationally identify lead compounds that are both biologically active and synthetically accessible **(Fig. 5)**. While the DMT paradigm is often described as a set of discrete steps, in practice, it is dynamic because of significant interplays between stages. Therefore, methods and objectives must be continuously revisited and cross-integrated to optimally achieve molecular optimisation **(Fig. 3)**. Overall, FBDMT aims to tackles fragment limitations by combining early orthogonal validation, rational and data-driven Design, faster and broader Make, crude-tolerant or insight-rich Test readouts yielding more, faster, and more informative cycles.

Initial chemical space sampling around fragment hits for SAR model building

Robust orthogonal validation of fragment hits is recommended, ideally using complementary assays[27]. Because poor hit overlap between methods is common[84,85], concordant activity across assays provides stronger evidence of true, reproducible binding and supports progression of those fragments and underlying interactions. Then, in the early stages of the DMT cycle, a primary focus is generating broad SAR data around selected fragment hits (**Table 2**) which provides critical information about the chemical groups and vectors that alter binding interactions, guiding the subsequent design phases. Building SAR model from fragment screening data alone is typically not attempted due to the weak and sparse binding information typical of fragments paired with the lack of a central scaffold.

Once fragment binding is confirmed, analogues, often sourced from catalogues of readily available molecules, are screened to cheaply and rapidly explore nearby chemical space[54]. An approach also known as “SAR-by-catalogue”. Even though these compounds are commercially available, their sourcing is categorised under the Make phase, as ordering, packaging, and shipping can involve significant time delays and financial costs. These analogue series help identify tractable SAR, ideally supported by structurally conserved binding poses and affinity data in the low micromolar range. At this point, fragment-derived series can be prioritised for more intensive design and synthesis efforts. This “SAR-by-catalogue” approach has the added value of being relatively easily automatable to maximise efficiency of initial chemical space mapping around confirmed fragment hits, especially when supported by fragment-bound crystal structures. For instance, the Frag4Lead workflow uses docking-based virtual screening of catalogue compounds, from MolPort, selecting 28 follow-ups expanded from 5 individual fragment hits, from over 10,000 candidates against endothiapepsin. 10 were confirmed via crystallography and 5 with increased affinity as measured with ITC. This cost-effective and automated approach accelerates SAR generation and series selection, bypassing custom synthesis, and positions fragment hits as high-quality seeds for initial chemical exploration. Validation assays at this stage prioritise potency and scalability, with structural validation reserved for the most promising candidates to guide subsequent design and synthesis[158]. A complementary strategy involves sourcing fragment libraries directly from synthetic intermediates, enabling immediate access to follow-up compounds. Huschmann et al., (2024) applied this concept by screening a natural product-derived intermediate library via crystallography against endothiapepsin, allowing rapid validation and elaboration of hits using existing route-accessible analogues[159]. Structural analysis of these follow-ups also enabled early scaffold triage, highlighting the efficiency of this approach for both SAR expansion and rational deselection.

In contrast, SAR, or even quantitative SAR (QSAR) models [160], may be inferred directly from an initial HTS provided sufficient data quality and hit rate. In some cases, HTS hits are clustered to identify fragment-like scaffold around which SAR can be inferred. An example of SAR generation from HTS data can be seen in a study targeting Plasmodium falciparum PKG, where 1.7 million compounds were screened[161]. After applying

selectivity and cytotoxicity filters, clustering techniques grouped 6,086 hits into 79 chemical series. The authors identified a fragment-like thiazole scaffold as a promising antimalarial starting point—a reverse path to SAR where a core fragment is inferred from larger compounds with higher-affinity. Because of the scale of HTS campaigns, structural assays are not used as primary screening methods and come later in the process **(Fig. 5b)**. Consequently, when starting from fragment, SAR must be acquired through additional DMT rounds while SAR may be directly available from HTS screening data. While FBDD may start one step behind HTS in terms of direct SAR availability, FBDD offers the possibility of using lower throughput screening methods (**Table 1**), such as structural or biophysical assays, as primary screening methods, because of the significantly smaller size of fragment libraries compared to HTS libraries.

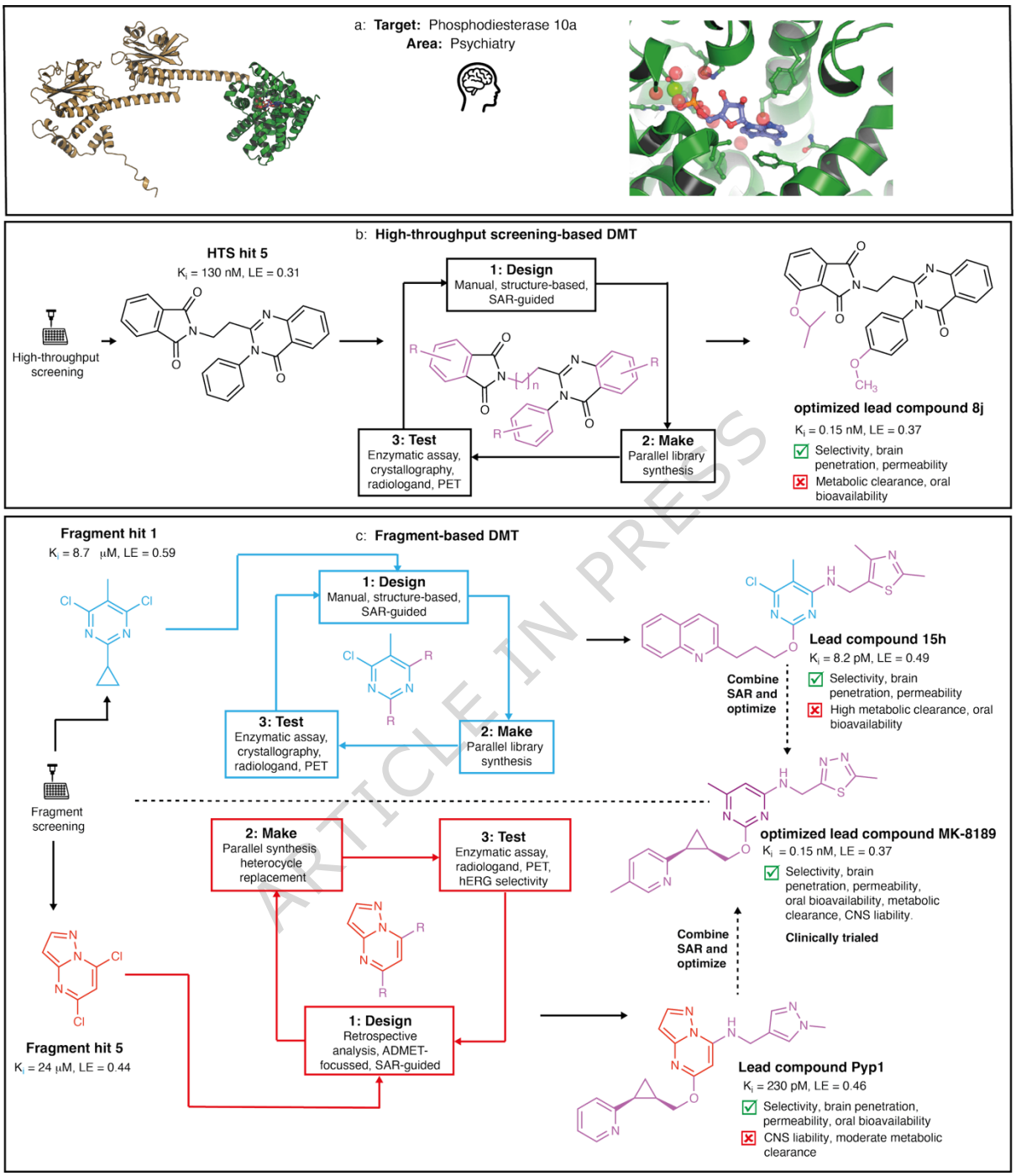

Figure 5: Comparative Design-Make-Test (DMT) strategies for phosphodiesterase 10A (PDE10A) inhibitor discovery. a Target and context: The phosphodiesterase, PDE10A, is a psychiatric target. The catalytic domain (green), substrate (lilac) and AlphaFold-predicted full-length model (beige) are shown as cartoon. b. High-throughput screening-based DMT: Screening of a large compound library identified a potent hit (HTS hit 5). SAR-driven design and parallel synthesis led to compound 8j, which showed some good properties, but was limited by high metabolic clearance and poor oral bioavailability. c. Fragment-based DMT: Screening of a fragment library yielded low-affinity but ligand-efficient fragments (fragment hit 1 and fragment hit 5). Fragment hit 1 (blue) was elaborated with iterative SAR cycles, leading to compound 15h, which displayed potent activity but suffered from clearance issues. This knowledge complemented a second FBDD cycle from fragment hit 5 (red) that led to Pyp-1. The final optimised compound (MK-8189) progressed to clinical trials.

Lead series identification and optimisation

As the DMT cycle progresses, the objective shifts from obtaining SAR information to improving the potency of compounds, often targeting low nanomolar affinities - a significant improvement from the weaker interactions typical of initial fragment hits **(Fig. 5)**[14]. Early lead optimisation also focusses on avoiding overly toxic or reactive groups to ensure that potential lead candidates maintain favourable drug-like properties. It is recommended to work on multiple series in parallel, keeping backup options in case the leading series encounters problems during ADMET evaluations[162]. As optimisation continues, the chemistry tends to become increasingly complex, requiring more precise modifications around the elaborated fragment scaffold. Once lead series have achieved the desired potency, additional physico-chemical properties and medicinal chemistry parameters become the focus of optimisation.

This transition marks the point where the exercise becomes increasingly multi-parametric. A cascade of assays, from simple *in vitro* tests to more complex *in vivo* evaluations **(Fig. 5)**, is employed, with only the most promising compounds progressing to later stages[163]. The DMT paradigm remains valuable here, but it now applies to a broader array of parameters, including selectivity and pharmacokinetics. Even when ligands are potent and selective for their therapeutic target, they may still bind to off-targets such as hERG **(Fig. 5c)**, CYP P450 enzymes, or transcription factors, which can lead to adverse effects such as cardiotoxicity, altered drug metabolism, or unintended gene expression. It is therefore crucial to ensure inactivity against these anti-targets, which can be considered as equally important as activity against the intended target[164]. At this advanced stage, while the core chemistry is often established, fine-tuning specific properties requires precise modifications that can be difficult to achieve synthetically[165]. A study on human adenosine A3 receptor (hA3R) antagonists illustrates the seamless integration of DMT cycle stages in lead optimisation. Advanced modelling techniques, including free energy perturbation (FEP) and MD simulations, were used to predict the binding affinities of modified analogues from initial fragment hits[166,167]. This prioritised 25 new analogues for custom synthesis to optimise parameters like selectivity and metabolic stability. These

compounds were evaluated through an assay cascade, yielding potent, selective A3R antagonists with nanomolar affinities and low toxicity ready for *in vivo* evaluation[168].

Another detailed case study of phosphodiesterase 10A (PDE10A) inhibitors **(Fig. 5a)** illustrates how multiple DMT cycles, whether initiated from HTS or fragment-based methods, can complement one another throughout optimisation[52]. In this example, HTS identified a potent scaffold requiring minimal modifications to address metabolic liabilities, but the resulting compound suffered from high clearance and low bioavailability **(Fig. 5b)**[169]. This lead compound was nonetheless used as a Positron Emission Tomography (PET) probe to assess target engagement and brain penetration *in vivo*, leveraging the high affinity of the scaffold despite its suboptimal pharmacokinetic properties[49]. Building on this, a FBDMT approach expanded small hits into potent leads, with one early FBDD-derived compound suffering from similar liabilities as the HTS series **(Fig. 5c)**[170]. Knowledge gained from these 2 campaigns (i.e HTS and FBDD) was used to steer a second fragment hit through an improved design-make-test sequence, ultimately yielding MK-8189, a compound with optimised Central Nervous System (CNS) exposure, bioavailability, and metabolic stability that progressed to clinical trials[48]. Notably, the pyrimidine core of the initial fragment hit was later replaced with a purine-derived scaffold from an alternative fragment, mitigating CNS liabilities and exemplifying how fragment motifs can be interchanged later in the campaign to address ADMET limitations **(Fig. 5c)**. This demonstrates how sequential or complementary DMT campaigns, even when initiated from distinct starting points, can complement each other to overcome bottlenecks, enrich SAR, and guide lead progression across multiparametric drug discovery landscapes.

Emerging strategies in integrative fragment-based DMT cycles

The integration of automation technologies with holistic models has the potential to significantly enhance DMT cycles[164] by optimising multiple parameters in real-time[171]. Currently, most DMT steps are executed by domain experts and orchestrated by medicinal chemists, requiring substantial human intervention. With the emergence of advanced computational methods and automation[172], the need for manual input is expected to decrease significantly, ultimately enabling scientists to focus on higher-value,

creative tasks. While autonomous and integrated closed-loop workflows are yet to be fully implemented, they have the potential streamline molecular design, chemical synthesis, and biological testing within compressed timelines[173,174]. Fragments, with their modularity, are well-suited starting points for these future technologies, which could significantly accelerate drug discovery and improve decision-making efficiency[175].

Deploying such technology, however, requires significant financial investment, often beyond the reach of academic labs or underfunded disease areas like antiviral and antibiotic research. Open-source initiatives and collaborative networks have emerged as possible alternatives, by "dividing to conquer" the DMT workflow for hit-to-lead progression, although hindered by organisational challenges[176]. A prime example is the COVID Moonshot, a fully open-science, crowd-sourced effort to target SARS-CoV-2's main protease[177]. This endeavour highlighted how fragment-based drug discovery can benefit from combining automation with public collaboration. The COVID Moonshot leveraged a large-scale crystallographic fragment screen to initiate iterative DMT cycles, with design, synthesis, and testing distributed across independent institutions[72]. Publicly open design submissions led to over 18,000 compound ideas, 2,400 synthesised molecules, and 490 ligand-bound X-ray structures leading to strong starting lead compounds for further antiviral development[178]. Indeed, the process culminated with nomination of compound ASAP-0017445 has as a pre-clinical candidate which was designed as a direct-to-generic and hopefully globally accessible treatment against coronavirus[179]. In this case, the synergy between automation, such as machine learning-guided synthesis, and community inputs enabled the DMT process[180]. In the following sections, we review developments within each phase of the DMT cycle.

Design

Now that initial fragment hits have been identified and the DMT framework established, we turn to the Design phase of the cycle: how fragment follow-up hypotheses are generated, prioritised, and selected for synthesis and testing. In the Design phase, the goal is to suggest and prioritise fragment follow-up compound hypotheses that can be synthesised, tested and optimised. This includes generating SAR data, incorporating structural assay insights, and applying computational methods to suggest modifications.

In the context of FBDD, the Design phase aims to sample the vast chemical space accessible through fragment elaboration while accounting for the inherently low affinity of initial hits. Nowadays, computational methods are playing a growing role by providing increased throughput and objectivity in terms of prioritising compounds based on quantitative criteria defined in scoring functions[35]. Ideally, a hierarchical approach is used whereby computationally inexpensive methods process the initial high volume of compounds followed by more demanding techniques for smaller subsets **(Fig. 6)**[35]. Simpler protocols, including catalogue searches or molecular docking, are increasingly employed in industry due to their ease of use and scalability, whereas more advanced workflows, such as enhanced MD-based approaches or bespoke generative AI tools, are often pioneered in academic settings before being commercialised through early-stage venture companies. In practice, the design phase often proceeds in parallel with analogue synthesis and SAR exploration, with X-ray crystallography pursued even before a potent lead series is obtained. Full structure-based design may hence begin only once higher-

affinity compounds emerge, reflecting the iterative and, sometimes empirical, nature of fragment campaigns and the progressive integration of structural insights.

Combining fragment-driven ligand approaches through cheminformatics and AI/ML for fast chemical space exploration

Establishing the initial follow-up chemical space from which to sample from is a pivotal step, which will usually start with libraries composed of commercially available compounds or those created from combinatorial reactions, potentially around fragments **(Fig. 7)**[17,181]. Follow-up libraries should be filtered to ensure suitable physicochemical properties, chemical accessibility and stability while also avoiding compounds that are reactive, unviable due to issues in ADMET, ensuring a high-quality chemical space[182]. The Enamine REAL Space library is frequently used to initiate such in-silico campaigns.

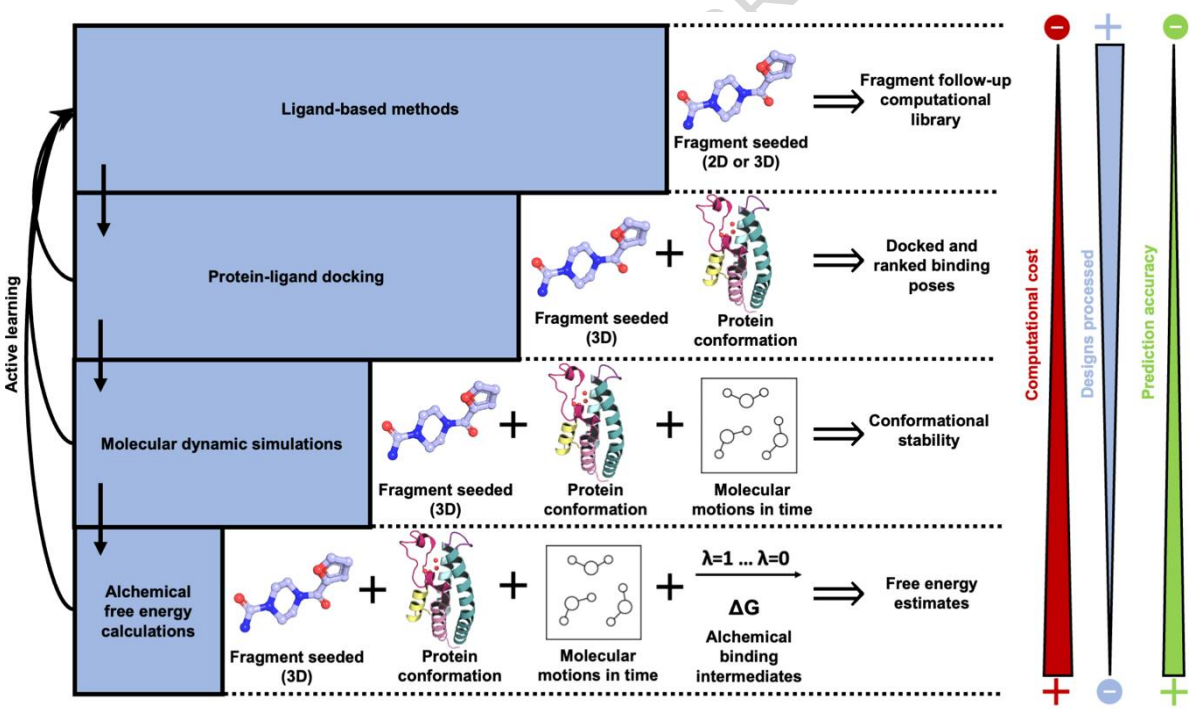

Figure 6: Hierarchical computational cascade for fragment-seeded follow-up design and prioritisation. Methods utilisation evolves from high-volume, cost-effective ligand-based screening to high-precision, structure-based approaches focused on predicting accurate ligand-binding properties. It integrates variables like hits, poses, protein structure and dynamics (represented by sandals, symbol time) and alchemical states (depicted as the ouroboros, symbol of alchemistry) to enhance the reliability of predictions, albeit at the cost of increased computational demand. Fragments are pivotal, providing initial hits and binding poses to inform the computational design process. Active learning is emphasised, illustrating how insights from more intensive computational methods can be used in the construction of virtual ligand-based QSAR models and facilitating the iterative exploration of chemical spaces.

It is a synth-on-demand virtual catalogue: compounds are enumerated in silico but are readily synthesisable via validated reaction rules and a large, curated set of building

blocks[183]. In its 2025 release, REAL Space comprises 78 billion accessible, drug-like products, being 4 to 5 orders of magnitude larger than a 1-million-member HTS deck and broadly comparable or slightly larger in scale to DNA-encoded libraries. Beyond sheer size, REAL Space spans broad scaffold and chemotype diversity arising from heterogeneous building blocks and reaction classes, exposing multiple elaboration vectors around fragment-like building block cores while remaining within drug-like property ranges[184]. Typical delivery time is about 3 to 4 weeks with an about 80% synthesis success, enabling rapid "SAR-by-catalogue" follow-ups once virtual triage is complete[183]. In the workflows discussed here, REAL-space triage is part of Design (virtual selection/prioritisation), whereas synthesis/procurement sits in Make. We note that REAL can also be used for virtual hit-finding prior to Make, but our emphasis in this section is on using these spaces to propose follow-ups seeded by fragment information.

Accounting for potential ADMET liabilities, early on, is particularly important, as such issues remain a leading cause of late-stage failure, contributing to about 40–45% of clinical attrition[185]. Additionally, follow-up compound cost is also a key prioritisation factor. Recent machine learning approaches, including retrosynthetic-pathway-based models and graph-based predictors, can help estimate compound prices or synthetic-route-derived cost metrics[186,187].

Ligand-based approaches, known for their cost-effectiveness (in computational terms) **(Fig. 6)**, link chemical features to activity outcomes but neglect target structure information and range from basic rule-based systems to advanced machine learning algorithms[35]. Incorporating fragment hits (**Table 2**) into these approaches allows for the extraction of follow-ups that mirror binding characteristics and explores the adjacent chemical space **(Fig. 6)**[188]. Conceptually, this is straightforward when growing ligands **(Fig. 2a)**, but can be extended to more complex graph-based methods for fragment merging **(Fig. 2c)**[189]. For example, one study demonstrated the use of a graph database, the Fragment Network, to explore chemical space around fragment hits. This method was applied to retrospectively identify commercially available fragment merges with micromolar IC_{50} values against Mpro and *Mycobacterium tuberculosis* transcriptional regulator EthR[189]. The Fragment Network provided a computationally efficient way to merge fragments while ensuring follow-ups are available in the catalogue and maintained

multiple crystallographic interactions, thus driving binding affinity. However, catalogue searches inherently cap the accessible design space to pre-enumerated products, and fragment merging, while powerful for recovering interactions, is often perceived as less synthetically tractable than fragment growing and linking **(Fig. 2)**. Synthesis-directed frameworks such as Syndirella[190] address both issues by planning multi-step, digitised routes from fragment-derived scaffolds and pursuing reactant-based procurement. This enables hundreds to thousands of congeneric designs per scaffold beyond catalogue space and integrates with plate-based, direct-to-biology workflows **(Fig. 8b)**, and delivers substantial per-compound cost reductions (reported ≈80% vs product purchasing).

As chemical spaces grow exponentially, challenges arise in tasks like virtual screening[191]. The need for navigating these “ultra-large” spaces, has driven the development of generative models and variational autoencoders (VAEs). These models avoid exhaustive enumeration by sampling from learned probabilistic distributions, efficiently generating novel structures not found in precompiled libraries. VAEs, create chemical latent spaces that encode molecular structures into low-dimensional representations, enabling the exploration of diverse molecular features like 3D complexity and chirality. VAEs can reconstruct valid structures while optimising for properties like bioactivity and synthetic feasibility[192]. An extension of VAEs is seen in Fragment-based Sequential Translation (FaST), which combines fragment-based methods with machine learning to optimise hits[193]. FaST employs a VAE to learn a distributional of molecular fragments, enabling molecular refinement via additions and deletions. By starting from known active molecules, FaST navigates chemical space to optimise multiple objectives, such as bioactivity, novelty, and synthetic feasibility. It efficiently generates diverse, high-quality follow-ups while maintaining desirable fragment-derived characteristics, showcasing the power of machine learning to balance exploration and optimisation within vast chemical spaces. Ligand-based generative models can also be seeded with fragment scaffolds to refine the search space and optimise multiple parameters simultaneously[194,195].

Leveraging fragment structural data in docking simulations for follow-up design

Structure-based methods use the 3-dimensional structure of the target to guide the design process **(Fig. 6)**. Structures therefore provide material for model preparation and docking

simulations **(Fig. 6)**[196,197]. A wide variety of scoring functions have been developed, ranging from physics-based to empirical and machine learning-driven approaches, but most conventional docking workflows do not natively account for experimentally derived fragment poses.[198]

Tailored protocols are often required to incorporate such structural inputs, and when properly implemented, these can significantly enhance design decisions by preserving known binding geometries or interactions, derived from fragment poses, thus ensuring compatibility with the target site. However, typically these methods make some key assumptions such as the receptor remains in the same conformation, which must be amenable to ligand binding and that key water molecules are optimally positioned within the binding site (if indeed they are included at all). These simplifications can obscure the dynamic nature of molecular interactions, potentially confounding the selection of follow-ups[199]. Selected follow-ups can then be processed to ensure the resulting poses recover as much of the available experimental information **(Fig. 4b)**. This might involve fitting to a key fragment, rescoring using shape and pharmacophoric overlay or employing electron densities directly[104,147,158,200,201]. For example, a study utilised poses from 4 fragments bound to protein kinase A (PKA) to guide template-based docking screen which identified 40 active compounds showed with the most active follow-up achieving a 13,500-fold gain in affinity compared to original fragments[147]. These considerations become even more relevant when designing follow-up from covalent fragments. Standard docking workflows are typically designed for non-covalent interactions and may not account for the irreversible bond formation or its structural constraints. Covalent bond formation restricts the conformational freedom of the ligand, simplifying the sampling landscape for follow-up design but requiring specialised tools, such as covalent docking protocols or restrained warhead positioning, to model the system accurately[202].

Recently, generative models, have been proposed to generate compounds in 3 dimensions directly using binding site structure. However, current *de novo* generative 3D approaches tend to generate chemically unviable compounds and unphysical binding poses[203,204]. Including experimentally resolved fragment conformations has been shown to improve generative models by enhancing 3D predictions. For example, AutoFragDiff**,** a fragment-based diffusion model, integrates chemically viable fragments to generate 3D

structures with more plausible binding poses[205]. Similarly, FragGen combines geometric constraints and machine learning to assemble synthesisable compounds, successfully delivering nanomolar kinase inhibitors with improved drug-likeness and synthetic accessibility[206]. Generative models, particularly deep learning-based approaches, are also becoming increasingly able to perform fragment linking **(Fig. 2b)** by connecting fragments in 3D space[119]. DeLinker, for instance, uses a graph-based model to optimise linkers while preserving the original fragments' binding poses and critical interactions within the binding site. This method significantly improves 3D similarity to known bioactive compounds, ensuring spatial orientation of fragment pairs is maintained[203]. These models efficiently link fragments, enabling the creation of lead-like compounds with greater accuracy and target relevance.

While these approaches hold great promise, it is important to note that their broad application in FBDD remains at an exploratory stage. The rapid expansion and diversity of machine learning methodologies, while a sign of innovation, also reflects the current uncertainty about which strategies provide consistently actionable insights across diverse target classes. Moreover, the sparsity of structural and affinity data, relative to the immense diversity of chemical space and targets, presents a key limitation that still hinders the generalisability of many AI-driven approaches. In cases where affinity data is available but structural information is lacking, it may be possible to generate synthetic structural data by leveraging fragment poses obtained, for example, via X-ray crystallography. Such synthetic data can then be used to retrain models or fine-tune them around previously unseen chemical series, potentially enhancing predictive accuracy. This strategy may be particularly well suited for emerging co-folding approaches, which currently struggle with generalisation[207-209]. Additionally, while generative approaches are

advancing rapidly, many remain difficult to integrate into routine pipelines due to challenges in synthesisability, interpretability, and validation.

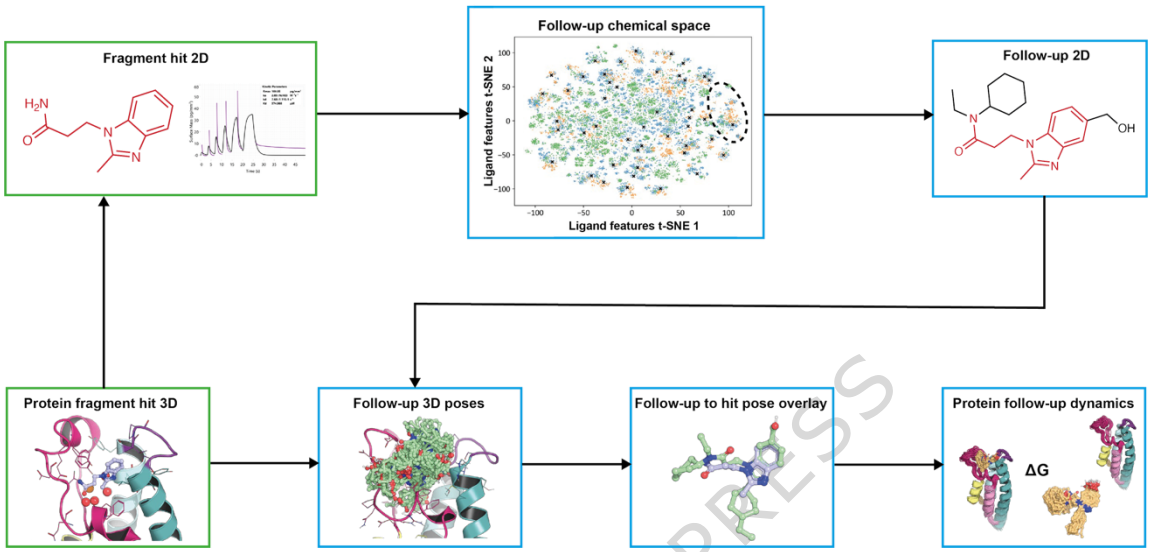

Figure 7: Exemplar workflow illustrating how fragment screening data fit within a typical structure-based design workflow for follow-up design. Fragments are screened with crystallography and validated with a solution assay (or vice versa). Selected fragment hits (in red) are used to seed follow-up searches, resulting in a mapped chemical space. Fragments are represented as black crosses, while different search methods (green, blue, orange) yield a variety of follow-up compounds. The follow-ups are then docked into the experimentally resolved bound receptor structure, using the native fragment pose to select the optimal docked pose. Molecular dynamics simulations are employed to estimate the dynamics and energetics of the receptor-follow-up complexes, providing a basis for ranking and selecting the most promising compounds for validation. Colour-coded frames match those in Fig. 3: design (blue), make (orange), test (green), and overlaps as combined colours.

Follow-up selection with molecular dynamic simulations

Docked poses of follow-up compounds or those obtained experimentally can be used to seed more intensive physics-based MD simulations, exploring conformational variation and estimate binding free energies **(Fig. 8a)**. Methods like Molecular Mechanics Generalised Born Surface Area (MMGB/SA) offer a computationally efficient approach for estimating the free energy of ligand binding, despite certain approximations[210]. If one has access to performant computational infrastructures, then free energy perturbations (FEP) allows follow-ups to be ranked according to predicted affinities[55]. Accurate MD-based predictions require a reliable structural model of the ligand-protein complex, as errors in pose selection can propagate through the calculations[211]. Including fragment pose information, such as from crystallographic data or validated docking results, can therefore enhance prediction reliability[210]. Structures of bound fragments serve as scaffolds for the

network of alchemical transformations required in relative binding free energy (RBFE) calculations given that the bound pose represents a global energy minimum and is not subject to changing upon elaboration[212,213].

For example, a recent study optimised fragments binding to a G protein-coupled receptor (GPCR) utilised MD simulations combined with RBFE calculations. The authors employed those computational tools to predict the pose of a micromolar fragment[167] initially identified via NMR screening against the A1 adenosine receptor (A1AR)[214]. They then employed RBFE to predict binding free energies for a series of elaborated compounds also providing an estimate of the SAR. The resulting compounds showed significant improvements in binding affinity, with one compound experimentally achieving a 41,000-fold increase in affinity and a 40-fold gain of selectivity[215]. Absolute binding free energy calculations may also provide additional confidence of bound and elaborate poses, but they are known to be challenging for fragments due to their weak affinity[210,216].

While MD simulations and free energy methods are powerful for non-covalent ligands, they are unable to account for covalent bond formation. To rigorously model covalent binding events, more advanced quantum mechanical (QM) or QM/MM methods are often required to calculate the energy landscape of bond formation[217]. Nevertheless, when running conventional MD simulations on covalent complexes derived from docking or X-ray structures, the covalent linkage inherently restricts ligand flexibility and reduces conformational sampling: an advantage for generating stable trajectories and refining elaborated poses[218].

Sampling with conventional MD for fragment binding kinetics remains challenging due to the small size and weak affinities of fragments, which result in high flexibility and unstable binding modes, complicating affinity estimates. The long-time scales required to capture binding and unbinding events, even for fragments with millimolar affinities, highlight this issue. For instance, in the FKBP study, microsecond-long simulations were necessary to observe multiple binding events and obtain statistically reliable estimates of binding affinities and kinetic parameters[219].

MD simulations and free energy calculations can also help dissect the thermodynamic contributions of fragment linking **(Fig. 2b)**[45]. As showed by Yu et al.,[45] these methods

allow a more detailed understanding of how each factor affects binding free energy changes. For example, fragment linking studies often show that expected gains from linking fragments can be offset by unfavorable interactions or binding mode changes[45]. These detailed insights may enable more rational molecular designs, as specific thermodynamic contributions can be targeted for optimisation rather than being folded into a single binding free energy term.

Despite many promising applications, the practical adoption of MD-based workflows in fragment optimisation remains limited. These methods require substantial computational infrastructures, careful system preparation, adapted analytics and expertise in setting up and analysing simulations. Small molecule parameterisation, particularly for novel scaffolds, can involve intricate quantum mechanical calculations, and errors at this stage can undermine predictive accuracy. Moreover, due to the intrinsically low affinities and high flexibility of fragments, long sampling times are often necessary to capture meaningful binding events, further increasing the barrier to widespread deployment in early-stage FBDD pipelines.

Combining active learning with structure-based predictions for dynamic virtual screening

Structure-based predictions, such as docking score or alchemical free energies, can be fed into an active learning loop, where these computationally expensive scores are used to train faster ligand-based models for rapid and incremental ligand-based scoring for further virtual chemical space exploration **(Fig. 6)**[220,221]. This enables to score large chemical space to optimise fragments at a fraction of the computational cost required by exhaustive structure-based screening approachces[222].

For example, Thompson sampling[223], an active learning method, was showed effective in guiding virtual screening of the Enamine REAL library, where reference shapes were extracted from pairs of crystallographically resolved fragment hits targeting the SARS-CoV-2 macrodomain. This approach enabled rapid screening of the 22-billion-molecule library with only a fraction of the computational effort, ultimately identifying a low micromolar fragment mergers with improved membrane permeability[224]. Another case is of dynamic chemical space sampling is Link-INVENT, a generative linker design tool that integrates reinforcement learning with structural data from docking simulations using fixed

fragment poses. By leveraging the learned docking scores, Link-INVENT optimises fragment linkers, by incorporating real-time on the structural fit of the generated linkers[225]. More computationally expansive free energy methods can also be used in conjunction with active learning. For example, a study used active learning with RBFE calculations to recover high-affinity phosphodiesterase 2 (PDE2) inhibitors from a large library. In each iteration, FEP calculations evaluated a small fraction of compounds, and the resulting affinities trained a ligand-based machine learning model[220]. This iterative process efficiently recovered high-affinity binders with only a small subset of the library evaluated with costly free energy calculations. Although fragments were not used, they could enhance the workflow by providing reference poses and affinity values as suggested elsewhere[226,227]. If only poses are available, absolute binding free energy calculations could be employed to estimate reference affinity values.

While computational design approaches show great promise, their routine adoption in fragment optimisation workflows remains limited, with many strategies still in proof-of-concept stages or used in parallel with expert-driven design. Expert review of selected follow-ups remains a critical step to manually discard unviable molecules (from a synthetic or compositional standpoint) (**Table 2**) that may have passed all computational filters and is still commonly practiced before proceeding into the “Make” part of the cycle[228].

Make

The “Make” phase encompasses all activities required to generate designed compounds in sufficient quantity and quality for testing. This includes route planning, synthesis strategies, and adaptations to scale-up or purify challenging intermediates. FBDD places particular emphasis on Make strategies that enable rapid, cost-effective, and high-throughput generation of diverse follow-up analogues.

The “Make” stage, in the DMT cycle **(Fig. 3),** is often rate-limiting due to unexpected reactions, low yields, and labour-intensive purification. In hit-to-lead campaigns, heavier synthetic medicinal chemistry efforts are typically deployed only after sufficient SAR and potent compounds have been established, once confidence in the series has grown and the risk of non-productive outcomes is reduced. Early-stage synthesis efforts often rely on brute, rapid and cost-efficient approaches for exploratory purposes. As confidence

increases, the chemistry progressively shifts from broad and exploratory to more targeted bespoke synthesis strategies enabling fine tuning compound properties. Overall, traditional medicinal chemistry workflows, though still common, tend to be slow, requiring extensive manual optimisation and quality control. Automated methods (**Fig. 8a**) streamline synthesis, purification, and reaction optimisation by reducing human intervention and improving scalability, making them ideal for modern drug discovery workflows. Although automated synthesis and robotics are gaining traction, these approaches are not yet widely implemented in fragment campaigns and often support, rather than replace, traditional medicinal chemistry efforts. In line with these automation-enabled workflows, the Make stage increasingly employs synthesis or assembly modes that are compatible with purification-agnostic assays, including combinatorial libraries or DELs. These strategies prioritise throughput and design-space coverage by minimising upfront purification and exploiting modular chemistries and robotics, thereby delivering sufficient material or identifiers for rapid evaluation. Their construction belongs to Make, while the associated assay readouts, deconvolution, and hit resolution are developed in the Test section, where results loop back into Design. Thus, the methods discussed in this section are applicable to single-compound elaborations and/ or the construction of such libraries.

Computationally guided synthesis around fragment scaffolds

Computational tools now aid synthesis planning by predicting reaction routes using available building blocks to generate optimisable schemes [229,230]. Predictive retrosynthesis, now often powered by deep learning models trained on reaction databases, deconstructs molecules into simpler components and maps optimal synthetic pathways. These algorithms can be constrained to start from fragments or close analogues, mapping possible reactions between a hit fragment and a designed molecule[231]. These tools provide chemists with plausible reaction templates to initiate synthetic planning, thereby reducing the preparatory workload although this may constrain routes to well documented reaction types.

Multiple commercial and non-commercial solutions are now available to support such tasks. Platforms like SPAYA™ (https://iktos.ai/solution/spaya) offer a user-friendly

interface and integrability with computational systems[232], allowing one to constrain intermediates or starting materials, such as fragment hits or analogues, when searching for optimal synthetic routes. Additionally, the integration of computational tools with synthetic workflows allows for real-time refinement of predictive models[233,234]. Recent developments have also demonstrated it is possible to combine synthesis with on-line analytics – in a so-called "chemputable" framework[235]. For instance, an integrative platform combining geometric deep learning and high-throughput reaction screening to predict reaction yields and regioselectivity with high accuracy, has been recently developed by Nippa et al.,[234] identifying opportunities for small molecule structural diversification. Coupled with fragment-based approaches, this can lead to more exhaustive chemical space sampling for the identification of potent follow-up compounds[234].

High-throughput robotic platforms for intensive fragment elaboration

Advances in high-throughput experimentation and automation have made generating experimental data for machine learning and enlarge chemical space sampling feasible **(Fig. 8a)**[236]. The introduction of novel robotics technologies has facilitated the synthesis of increasingly large arrays of compounds **(Fig. 8a)**[237,238]. For instance, the use of acoustic droplet ejection technology has enabled the automated synthesis of large and diverse compound libraries[239]. This approach was exemplified by the parallel synthesis around 16 different fragment-like scaffolds on several 384-well plates, resulting in a high-throughput generation of diverse small molecules, demonstrating the platform's capability to efficiently explore novel chemical space regions. Beyond single-transformation libraries, a complementary "cluster synthesis" strategy batches chemically diverse reactions by overlapping operational windows (time/temperature) on a single robot, enabling multi-reaction campaigns that boost structural diversity and platform utilisation (e.g., 27 name reactions and 135 molecules executed in three campaigns)[240]. While these arrays are primarily synthesised using well-established chemistry, which may restrict the diversity of reactions employed, they can also play a crucial role in reaction optimisation tasks[236].

Flow synthesis has also emerged as a modular approach to combinatorial compound synthesis. Pijper et al.[241], demonstrated that continuous flow platforms, integrating up to eight synthetic methodologies, can generate structurally diverse compounds in a single workflow. By incorporating various chemistries, including metallaphotoredox couplings, the platform produces chemically varied fragment-derived compounds at a throughput of up to four compounds per hour. Paired with a high-throughput purification system, this setup bypasses the typical bottleneck in purification, automating and streamlining the process, reducing human intervention, and improving overall efficiency.

Similarly, there has also been progress in affordable end-to-end synthesis platform, which includes compound purification, a known bottleneck at the "Make" stage because of human intervention, need for quality control and solvent and hardware usage[237]. For example, an automated synthesis platform paired with novel liquid-liquid extraction techniques and quality control has been recently developed, by Abdiaj et al.,[237], enabling the unattended synthesis and purification of drug-like molecules from fragment-like starting points, creating an end-to-end high-throughput production line. The platform utilises the uncommon Negishi coupling reaction, resulting in compounds with enriched sp^3-hybridised carbons, which are frequently present in small molecules reaching the clinic[237].

Expanding fragments with novel synthetic methodologies

Significant strides in organic methodologies and catalysis have broadened the toolkit of available reactions **(Fig. 8a)**. This is particularly relevant for fragment elaboration (**Table 2**) at previously unapproachable chemical vectors. For instance, C-H activation techniques have significantly advanced fragment elaboration strategies. It was showed that about 80% of fragment growth originates from carbon-centric elaboration vectors while the non-reacting moieties engage with the receptor through polar interactions[242]. Tolinapant, a clinical candidate against colorectal cancer that inhibits apoptosis proteins (IAPs)[243], was designed using C-H activation techniques. A fragment was identified via NMR screening with structure-based optimisation guiding the addition of groups at aliphatic carbons. One of these modifications included the addition of a simple methyl group, which resulted in a 60-fold increase in potency, highlighting how small and precise

modifications at carbon atoms can significantly progress fragments into potent compounds[244]. Recent advances in carbon–heteroatom bond-forming methodologies, such as photocatalyst-enabled late-stage functionalisation using sulfonamide or carbamate precursors to install nitrogen or oxygen substituents with high site selectivity, also enable analogous site-selective modifications to be performed at scale, expanding their practicality for medicinal chemistry campaigns[245]. It is critical that such insights of synthetic sociability (**Table 2**) are captured during the design phase to ensure that follow-ups can access these newly available regions of chemical space **(Fig. 7)**.

Recent advancements in synthetic methodologies have also made previously inaccessible chemical groups available for drug discovery. For example, novel 1,1'-bicyclopropenyl derivatives are now accessible via synergistic Au/Ag dual-catalyzed cyclopropenyl cross-coupling [246]. These highly strained bicyclic benzene isomers open unexplored regions of chemical space. Their modular synthesis under mild conditions, coupled with functional group tolerance, enables the incorporation of new groups like trifluoromethyl, potentially broadening access to follow-up designs implying a potentially greater diversity of protein-ligand interactions. Modular “click” chemistry offers a simple, plate-based way to grow fragment cores quickly[247]. In a two-step sequence, a common handle on a fragment (a primary amine) is first converted into an azide using a bench-stable reagent, and then “clicked” with an alkyne partner using a copper catalyst to form a triazole link. Because these relatively novel reactions are fast, high-yielding, and tolerant of many functional groups, large sets of analogues can be made in hours and may be clean enough to test directly in biology, cutting both cost and cycle time. This has been demonstrated in FBDD: a copper-catalyzed click campaign expanded a fragment against the viral protein LANA, delivering SAR information and multiple low-micromolar follow-ups that were confirmed with orthogonal assays[248].

The elaboration of covalent fragment hits, also, demands specialised synthetic considerations. Reactive warheads, such as acrylamides, chloroacetamides, and sulfonyl fluorides, are typically introduced at later stages to minimise side reactions and preserve compound integrity during synthesis[249]. Hits originating from disulfide “tethering” screens must often be re-synthesised with more drug-like warheads (e.g., stable electrophiles like acrylamides) for further development[250]. Medicinal chemists also optimise warhead

positioning relative to the fragment core; short linkers are often preferred to achieve optimal geometric alignment for covalent bond formation at the nucleophilic site. These synthetic adaptations are critical to balancing reactivity, selectivity, and stability, and often require iterative fine-tuning of the warhead's functionality. Modern synthetic chemistry has thus become essential to successfully advancing covalent fragment leads[251].

Reaction scalability and solvent selection as important links between synthesis and validation

A critical challenge in fragment-based drug discovery is the scalability of synthetic reactions for producing large quantities of compounds for testing assays. Addressing this, a photoredox-mediated cross-dehydrogenative coupling approach, as demonstrated by Grainger et al., showcases a nanogram-to-gram workflow, which begins with nano-scale HTS of reaction conditions, optimisation and scales up using continuous flow chemistry[252]. This method allows for the rapid production of gram-scale quantities of compounds with sp^2–sp^3 privileged architectures, such as heterocyclic amines, which are particularly useful in medicinal chemistry. The combination of scalability and high functional group tolerance makes this approach invaluable for enabling large-scale synthesis, further enabling the drug discovery process by providing enough material for validation.

Furthermore, the choice of solvent can impact the outcome of assays, as solvents may have inherent affinity for the target, potentially outcompeting the ligand under study. Given that solvent molecules are almost infinitely more concentrated than the ligand, they can saturate the binding site, necessitating the use of appropriate controls in solution assays[253-255]. Such baseline controls may be feasible with solution-based methods, including SPR, which has been extensively used to study small molecule binding to bromodomains[254]. This issue is particularly relevant for fragments due to their inherently low potencies. For example, DMSO, a polar aprotic solvent commonly used in synthesis and storage, can interfere with assay readouts by binding to the target site, such as bromodomains. DMSO mimics the acetylated lysine binding motif, which can prohibit crystallographic ligand binding by saturating the pharmacologically relevant binding site[61]. However, baseline controls are not feasible with crystallography, thereby requiring the use

of alternative solvents, such as ethylene glycol highlighting the importance of target knowledge when preparing compound libraries (**Table 1**).

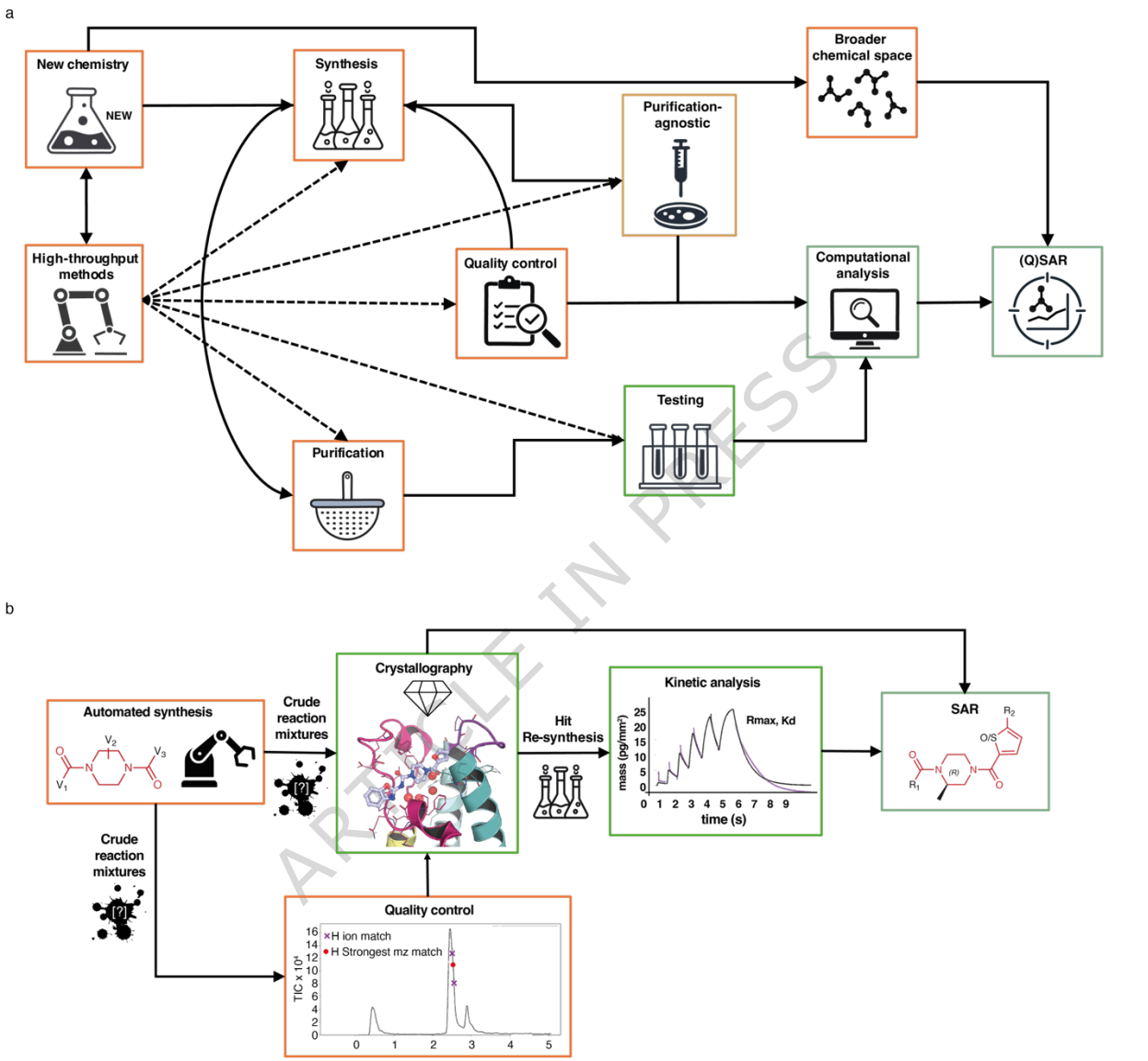

Figure 8: a. Integration of "Make" (orange) and "Test" (green) stages within DMT Cycles. Synthesis of novel chemical entities facilitates the exploration of an expanded chemical space. High-throughput methods impact synthetic and testing workflows, enhancing efficiency through increased data generation, crucial for robust QSAR modelling. So-called "purification-agnostic" approaches simplify the transition from synthesis to testing by minimizing the resources required for compound purification. **b**. Automated synthesis and quality control combine with purification-agnostic testing approaches for rapid fragment elaboration. Automated chemistry is performed around a fragment hit (in red) using crude reaction mixtures of unknown composition, which are submitted to crystallographic validation. Quality control results enable the tracking of samples expected to potentially bind. Hits from crude reaction mixtures are then independently repurified before confirmation with solution assays. The structural and affinity data can then be used to build robust SAR models around the fragment.

Test

The “Test” phase ideally evaluates synthesised fragment follow-up compounds for binding, activity, and pose at early DMT stage with other relevant ADMET properties tested later on. It includes biochemical, biophysical, and structural assays for affinity and pose determination, supported by appropriate analytics to ensure data quality, interpretability, and decision-making. The Test phase in FBDD demands assays that ideally combine high-throughput capacity to evaluate large follow-up arrays with sufficient sensitivity to reliably detect the low-affinity interactions typical of fragment-derived compounds.

To test compounds, validation assays typically require pure products, either self-manufactured or procured from vendors **(Fig. 1 & Table 1)**. Yet, as previously discussed, compound synthesis can be rate limiting within DMT cycles, resulting in challenges generating large numbers of molecules **(Fig. 3)**. An alternative to circumvent this bottleneck involves employing crude reaction mixtures directly[238,256] in purification-agnostic validation assays, often referred as “Direct-to-Biology” approaches **(Fig. 8b)**. In this context, fragment hits provide ideal scaffolds for high-throughput synthetic elaboration, thus allowing large-scale testing of the newly defined chemical follow-up space **(Fig. 8b)**[257]. In what follows we evaluate outputs of Make stage and feed results back into Design.

Scalable crystallographic evaluation of fragment elaborations in crude reaction arrays

While crystallography can be used at hit finding to obtain initial poses, here we consider its Test-stage role: a scalable, potentially orthogonal readout applied to post-hit design and synthesis. In recent years, crystallography has undergone transformative improvements enabling large-scale evaluation of compounds via the use of crude libraries. Because parallel/automated synthesis produces plate-based arrays and crystallography is tolerant to reaction by-products, the number of testable elaborations per Make cycle has increased substantially thus placing new throughput demands on the

structural pipeline. In parallel, advances in crystal systems, automated beamlines, sample handling, and data processing now permit HTS of hundreds of compounds, making crystallography a powerful tool for large-scale compound evaluation and structure-based drug discovery **(Table 1)**[77,258]. These gains in throughput, coupled with the inherently high sensitivity of X-ray crystallography, make it well-suited to purification-agnostic workflows where numerous follow-up compounds must be assessed, even when present at low concentrations due to unpurified reaction mixtures or suboptimal synthetic yields.

For example, Sutanto et al.,[259] showed that X-ray crystallography also pairs well with high-throughput synthesis by applying it to crystallographic fragment hits of the SARS-CoV-2 main protease. Crude reaction mixtures were soaked onto crystals, with crystallography serving as the primary assay to evaluate follow-up binding, followed by revalidation of repurified hits using solution assays. In addition, it was recently showed that, when resolved with high-throughput crystallography, crude reaction arrays of compounds can yield large-scale structural SAR insights into the elaboration space **(Fig. 8a)**[238]. X-ray crystallography proves advantageous in testing crude arrays, due to its ability to resolve reaction products within the binding site or protein-protein interaction interface[260], given sufficient resolution and quality control **(Fig. 8a)**. For example, hundreds of follow-up compounds were generated from a fragment identified crystallographically using an affordable liquid dispensing device (Opentrons®) targeting PHIP2[238]. These crude reaction mixtures were soaked directly onto protein crystals at high-throughput levels. Quality control of the mixtures, via tandem Liquid Chromatography-Mass Spectrometry, enabled tracking which samples had the desired final product, and this information was then used to extract SAR data from the resulting structural data or lack thereof **(Fig. 8b)**. Follow-up work demonstrated that even noisy data from crude assay readouts can generate actionable SAR models by identifying conserved chemical features linked to crystallographic binding events[261]. This approach enabled direct integration of SAR insights into subsequent computational design phases, facilitating the rapid identification of improved follow-up compounds. Importantly, the study illustrated how SAR extracted from in-house crystallographic evaluation of crude reaction mixtures can guide virtual screening of commercially available compounds, thus further accelerating DMT cycles.

However, crystallographic testing of crudes may occasionally be obscured by resolving lower-affinity species highlighting the need to account for potential false negatives with orthogonal assays as demonstrated in study on the heat shock protein 90 (HSP90) and pyruvate dehydrogenase kinase isoform 2 (PDHK2)[262]. Unlike solution experiments, such as NMR[263], where small molecules compete in equilibrium and the most stable binders dominate the readout, crystallographic soaking involves diffusion through the crystal lattice. Differently to cocrystalisation, multiple species present in a reaction mixture may enter with different kinetics, potentially leading to partial occupancies and/ or superimposed electron densities[262]. This may further complicate the interpretation of binding modes. Hence, high-throughput crystallographic testing of crude reaction mixtures should be coupled with robust quality control to ensure the observed binding species corresponds to the intended product. Indeed, crude mixtures may also contain reactive or unstable intermediates that can mislead structure-based interpretation or mask liabilities, as highlighted by Cramer et al., where degradation products interfered with assay interpretation and derailed follow-up efforts[264]. Accordingly, orthogonal assays, appropriate controls and structural triage steps remain essential to avoid erroneous conclusions in early-stage optimisation.

A prominent recent example highlighting both opportunities and challenges in fragment-based structural data comes from an industry-led crystallographic study on fatty acid-binding proteins[265]. Among 229 high-resolution structures, about 15% contained ligands chemically distinct from those initially soaked, due to reagent mislabeling or chemical transformations occurring in crystallography[266]. The dataset also demonstrated how subtle conformational changes—particularly in the ligand portal region—depend on ligand identity, impacting functional properties such as membrane association and nuclear import[267]. To mitigate risks of propagating incorrect ligand assignments in fragment-to-lead campaigns, the authors recommended stringent validation criteria, including chemical verification and occupancy filtering before structural data is used in computational design or machine learning[266]. However, the call for strict curation sparked debate. Jaskolski et al., argued for caution, proposing that large-scale crystallographic fragment screening datasets, many of which rely on multi-dataset approaches such as PanDDA, be clearly annotated or archived separately due to complexity and potential

misinterpretation by non-specialists[79,268]. Conversely, Weiss et al., defended these datasets, asserting their value in capturing compositional and conformational heterogeneity that can inform novel chemical strategies and improve computational methods[269]. Fraser further encouraged embracing "messy," ambiguous data to reveal hidden dynamic phenomena and enhance ligand discovery workflows[270]. Together, these perspectives underscore a central tension in fragment progression within DMT cycles: the need to balance conservative validation to avoid artefactual design decisions with openness to weak, ambiguous, or heterogeneous data that may seed valuable hypotheses or novel strategies[271]. Such complex and "messy" datasets also present opportunities to enhance AI-driven computational tools, training algorithms to handle real-world uncertainties inherent in fragment-bound structures[270]. In practice, orthogonal validation of fragment identity and binding markedly increases confidence that low-occupancy or ambiguous density reflects true target engagement. When fragments are fully validated, these borderline cases will be truly informative rather than artefactual and can be prioritised for follow-up. As crystallographic fragment-based screening campaigns scale in throughput and complexity, establishing shared best practices around annotation, triage, and interpretation of fragment- bound structures will be essential[272].

Non-structural purification-agnostic assays for fast SAR mapping around fragments and follow-up hit finding

Furthermore, the application of crude reaction products can be extended to off-rate evaluation via kinetic methods such as SPR or Grating-Coupled Interferometry (GCI) **(Table 1)**[110,256,273,274]. Off-rates are determined by the strength of target-ligand intermolecular interactions and are independent of concentration variances, thus minimising concentration-related uncertainties[256,262]. Equilibrium affinity (K_d) reflects the balance of both association (k_{on}) and dissociation (k_{off}). By contrast, the off-rate (k_{off}) isolates only the dissociation step and therefore serves as a practical proxy for binding stability, with residence time given by 1 over k_{off} with smaller k_{off} values indicate stronger binding. For example, a fragment hit for the bromodomain-3 extra terminal (BRD3-ET) domain was identified using NMR[275]. Fragment follow-ups were then purchased and tested as pure compounds with NMR to identify tolerated elaboration vectors hence providing early but important SAR information. Parallel chemistry was applied to these

vectors, and the resulting reaction mixtures were tested using off-rate screening. This structure-agnostic approach identified higher-affinity follow-ups, by about 30-fold, from the crude mixtures. These hits were resynthesised and tested in their pure form, revealing similar off-rates to those initially observed from crudes showing that the crude evaluation was not generating random signal. However, large-scaled interpretation of purification-agnostic assays from crudes is likely challenging as multiple species may contribute to binding creating readout uncertainties. Thus, when a follow-up hit emerges from crude testing, it generally necessitates resynthesis to validate the initial findings, thereby feeding back into the "Make" phase **(Fig. 8a)**[238,276]. Although retrospective, SAR analysis has been explored, collective data from crude evaluation of fragment follow-ups is typically not directly utilised in subsequent design phases, a potential avenue to optimise hit-to-lead transitions, by avoiding independent compound resynthesis.

One possible avenue for deconvoluting these noisy readouts is through affinity chromatography–mass spectrometry, which enables precise quantification of binding interactions within complex mixtures under equilibrium conditions[277]. Additionally, in a study on a crude library of histone deacetylase inhibitors, MS was successfully used with affinity selection for microsomal metabolic stability assays [278]. The crude readouts showed excellent correlation with those from pure compounds, demonstrating that crude mixtures can effectively serve as a proxy in this context. This highlights the potential for evaluating crude reaction mixtures in a lead optimisation context.

NMR spectroscopy can also be highly effective for testing crude or complex mixtures, potentially informing about structural interactions and offering orthogonal and sensitive readouts of binding without requiring compound purification. For example, Wu et al., (2013) illustrated the broader utility of NMR in high-throughput mixture screening by applying ligand-observed NMR techniques, to over 10,000 compounds screened in mixture formats. This enabled the reliable detection of weak binders, which were later confirmed individually in pure form[279]. Later, Larda et al., (2023) introduced an integrative workflow for hit-to-lead development against undruggable targets such as HRas. Their "NMR for SAR" approach combined fluorine NMR (to detect changes in signal shape) and relaxation-based assays (to assess binding), along with both protein- and ligand-detected NMR, to monitor solubility, aggregation, target stability, and interactions across multiple

SAR cycles[280]. Alboreggia et al., (2023) also demonstrated how protein-observed NMR, specifically monitoring the aliphatic region of the target protein spectrum, can be used for mixture-based screening of focused combinatorial libraries (~125,000 compounds) against hMcl-1. This enabled identification and optimisation of active motifs directly from complex mixtures, relying on cumulative chemical shift perturbations caused by structurally related analogues sharing a common scaffold[281]. These examples underscore that NMR-based assays offer a robust and versatile platform for decoding SAR from impure or complex matrices, complementing kinetic methods, and pave the way for broader adoption of purification-agnostic hit finding strategies.

In the broader context of follow-up identification, protein-directed dynamic combinatorial chemistry (DCC) offers an efficient strategy for discovering high-affinity ligands. In DCC, reactive fragment building blocks reversibly assemble under thermodynamic control, allowing the protein target to shift the equilibrium toward the most favourable binders. This obviates the need for purified intermediates or affinity quantification of weak binders[282]. For example, a study used fragment-derived acylhydrazones dynamically assembled and screened in situ against α-glucosidase. Enriched products were identified via LC-MS by comparing peak intensities between protein-templated and blank libraries and validated through enzymatic inhibition assays. Iterative cycles of fragment growth and library evolution yielded nanomolar inhibitors without resynthesising intermediates[283]. Additionally, MS-based assays have proven effective for evaluating covalent fragment follow-ups directly from crude reaction mixtures. For instance, Wilders et al., developed a parallel synthesis platform combined with intact-protein LC-MS to accelerate SAR cycles on reactive fragments targeting SARS-CoV-2 MPro, enabling the identification of both covalent and non-covalent leads without prior purification[284].

Protein-based affinity or enzyme inhibition alone does not guarantee cellular activity with discrepancies commonly arising from poor intracellular exposure (limited permeability, active efflux), suboptimal ionization/solubility (pKa/logD), nonspecific sequestration, or rapid clearance that prevents sufficient target engagement in cells. A clear FBDD example is the DNA gyrase GyrB program: fragments were screened by ligand-/protein-observed NMR against the ATP site, yielding a millimolar pyrrole hit that seeded a focused library. Iterative X-ray structures then guided optimisation to micromolar ATPase inhibitors, yet

the early leads showed little or no antibacterial (cellular) activity, exposing an exposure gap between enzyme and cells[285]. Importantly, this case also illustrates a limitation of structural readouts: a well-defined crystallographic pose can confirm pocket binding and rationalise SAR, but it remains agnostic to permeability and efflux, so structurally validated potency may still fail to translate in cells. Progression addressed these liabilities by tuning physicochemistry and exposure: varying the 3-piperidine substituent and installing 4-thiazole substituents that form a seven-membered intramolecular H-bond to modulate acidity and permeability, enhancing cellular antibacterial potency, greater activity in efflux-deficient E. coli highlighted the role of efflux, and improved systemic exposure through reduced Mrp2 recognition lowered clearance, leading to the clinical candidate AZD5099[286].

Potential synergies between fragment-based and as DNA-encoded libraries readouts

Other purification-agnostic approaches, such as DEL-based or fragment photo-active screening have also been employed, demonstrating versatility in testing methodologies[276]. While DEL construction can be categorised as a Make-stage activity[287], we discuss DELs here as we on hit resolution and readout rather than the underlying on-DNA chemistry.

SAR models have, previously been extracted from DEL results, with starting building blocks initially identified through crystallographic fragment screenings. For example, X-ray crystallography identified BRD4 fragment hits that mimic the interaction features of acetyl-lysine, the native substrate. These hits were expanded and validated with SPR and further confirmed via initial SAR[288]. Then, one fragment was expanded via on-DNA Suzuki–Miyaura cross-coupling, linking the isoxazole warhead to a 42-member “poised” DNA-encoded library (NUDEL). The focused DEL appended amino-acid-like side chains and heteroaryl caps (e.g., alanine plus a methyl-amidopyrazole), enabling additional contacts in the binding site while systematically sampling complementary building-block combinations[289]. This approach rapidly optimised a weakly binding fragment into a structurally resolved 51 nanomolar inhibitor with good ADMET characteristics also yielding key binding site insights used for retrospective SAR. [289] This example illustrates the synergies of fragment-based drug discovery with other methods to enhance the

identification of follow-ups lead compounds. This study further illustrates how carefully selecting functionally relevant fragments and validating them through SAR and orthogonal assays can lead to a successful conversion of a weak binders into lead-like compounds.

DEL readouts are inherently noisy, requiring advanced analytics to extract useful and actionable insights, with noise arising from factors such as incomplete or variable synthesis yields, tag mis-encoding or truncations, differential PCR amplification efficiencies, sequencing errors, and target-independent binding. Uncertainty-aware machine learning models, as shown by Lim et al.,[290] can denoise DEL experiments, providing more reliable binding affinity predictions. Inference models can also pinpoint productive building blocks linked to significant binding outcomes, as demonstrated by Zhang et al.,[291] Additionally, combining DEL-binding data with docking poses using Graph Neural Networks with self-attention mechanisms, as highlighted by Shmilovich et al.,[292], further refines docking predictions, aligning them more closely with experimental affinity measurements. These recent DEL advancements may be enhanced by integrating fragment data. Indeed, co-crystal structures of fragments can refine structure-based and QSAR models derived from DEL data, while fragment affinities may provide baseline scoring metrics for evaluating the binding propensity of DEL building blocks.

Overall, data from testing phases evaluate fragment follow-ups potency and structure, facilitating SAR model development, which are crucial for subsequent designs **(Fig. 3)**. Hits identified from DEL and other purification-agnostic approaches often undergo re-purification, effectively looping back into the Make stage before the next round of design. This workflow adjustment can thus appear as a Design–Make–Test–Make iteration, illustrating the occasionally modular nature of modern FBDMT workflows. As the DMT cycles progress, the quest for compounds with suitable parameters for in vivo evaluation gradually shifts away from initial fragment space, refining compound selection and diminishing the relevance of initial screening hits **(Fig. 2)**. The introduction of novel considerations, particularly ADMET characteristics, underscore the project's advancement from hit identification to lead optimisation, a phase that extends beyond the scope of this review[42,293].

Summary

Fragments enable efficient chemical space exploration, facilitating the development of potent follow-up compounds with lead-like characteristics. FBDD however has limitations. The initial selection of fragments plays a pivotal role in shaping the hit-to-lead strategy, yet remains a case-dependent, multi-objective decision that balances factors including physico-chemical properties, binding affinity, biological activity, selectivity, chemical tractability, chemotype novelty, ADMET properties and other features such as intellectual property. Fragment libraries are also often tailored to specific goals, such as synthetic accessibility, or focused on particular target classes, further influencing selection strategies. While many projects apply rigorous and objective selection criteria internally, there is currently no universally accepted paradigm guiding library design, and fragment selection across the field. This signals an area for further development, where holistic, end-to-end models could help formalise and streamline early-stage decision-making, through *in silico* follow-up modelling, and by accounting for multiple objectives simultaneously. Crucially, establishing a fragment hit and building a meaningful SAR series relies on orthogonal validation approaches, an effort facilitated by advances in high-throughput chemistry, computational modelling, and experimental testing.

Defining and refining objectives across different DMT iterations is similarly a case-specific and often subjective process. In practice, an immediate, pragmatic goal is to establish a lead series through follow-up of validated fragment hits, with tractable SAR to guide subsequent optimisation. End-to-end workflows and automated laboratories powered by artificial intelligence for decision-making[294,295] have the potential overcome this bottleneck. However, much progress remains before such proof-of-concept work can be translated into practical, streamlined, and broadly deployable approaches. Fragments, with their modularity, serve as ideal scaffolds for multi-task optimisation across these workflows, yet their intrinsically weak initial affinities remain an important bottleneck in early iterations.

In the design phase, fragment data can empower computational methods, particularly those aimed at integrating activity (affinity and ADMET) predictions with chemical accessibility[15]. Evolving fields like machine learning and free energy calculations, where

fragment structural and affinity data can be used to train models and benchmark simulations, are also showing promise to better prioritise follow-up compound *in silico*[194,220]. Generative AI appears particularly powerful for sampling ultra-large chemical spaces and addressing multi-parameter optimisation challenges. However, there are still significant pitfalls, including chemical stability considerations or unphysical binding poses with fragment data possibly constraining search spaces, thus hindering model development[296]. Newly established active learning protocols for transforming costly structure-based scoring into more efficient ligand-based approaches are showing promise, with fragment data enhancing predictions by providing high-quality reference poses and scaffolds.

The “Make” phase in FBDD is also evolving. Novel synthetic methodologies, in particular C-H activation, provide access to previously unexplored chemical spaces[242]. Advances in high-throughput instrumentation and automation, facilitate the synthesis of larger follow-up arrays, thereby allowing for more extensive explorations[236,237,239]. Fragments and analogues serve as high-quality starting points for large scaled synthetic elaboration. However, the large-scale production of compound arrays can lead to scalability bottlenecks, particularly during purification[4,237,238]. Retrosynthetic algorithms are increasingly valuable for compound scoring for accessibility, offering pathways to identify compounds derived from fragments or analogues, while constraining routes to achieve more feasible and efficient synthesis plans. In parallel, increasing efforts in developing drug candidates from covalent fragments, especially through modern and site-selective synthetic strategies, are enabling the targeting of difficult proteins with improved precision and reactivity control.

Testing methodologies, such as crystallography, are adapting, especially with the introduction of testing crude reaction mixtures from fragment elaboration and other purification-agnostic methods, like DELs[238,275,289]. It is our view that these advances bring about the challenge of deconvoluting noisy experimental data to build reliable and actionable SAR models, highlighting the need for sensitive testing methods, usage of complementary modelling tools, scalable validation assays, and interpretable data analysis techniques. In FBDD, testing must offer both high sensitivity to detect weak

binding events typical of early fragment follow-ups and sufficient throughput to handle increasingly large follow-up libraries.

Outlook

Beyond FBDD, the integration of these advancements into end-to-end, automated drug discovery pipelines represents an aspirational yet increasingly plausible goal. Modular workflows combining machine learning, robotics, and real-time decision-making are actively being developed to address inefficiencies and resource constraints linked to various DMT stages, fostering a paradigm shift toward closed-loop systems. Fragments, having already demonstrated their success in delivering FDA-approved drugs and clinical candidates, continue to serve a productive tool in modern discovery efforts. Their simplicity, modularity, and ability to seed chemical space exploration make them particularly well-suited for incorporation into these automated and computationally driven approaches. Covalent FBDD are also showing promise thanks to novel reactive moieties, however, computational tools for modelling these remain underdeveloped and often inaccessible to non-experts. Even when robustly implemented, these integrated and modular DMT advances would primarily strengthen medicinal chemistry by accelerating the conversion of validated fragment hits into coherent lead series with tractable SAR. That said, many of these workflows remain at an early stage of adoption, with practical implementation still limited to select case studies or pilot systems. Equitable access to these technologies remains a challenge, particularly for underfunded therapeutic areas such as antivirals or neglected diseases. In parallel, openly sharing high-quality fragment data, reaction protocols, and design strategies[213] across all stages of the DMT cycle, including structure-based designs, synthetic reactions[297], and assay outcomes[272], will be essential to benchmark emerging methodologies and train the next generation of AI-driven drug discovery[298].

Looking ahead, the continued progression of FBDD with emerging technologies will be pivotal in producing treatments for currently unmet medical needs. By leveraging fragments as modular scaffolds, the drug discovery community can amplify their impact through holistic, increasingly technology-enabled DMT workflows. These approaches may not only increase efficiency but also ensure that the process remains adaptive and

accessible to a broad range of research contexts. Addressing significant bottlenecks will require a multidisciplinary approach, combining computational chemistry, synthetic methodologies, and innovative testing strategies, ensuring that FBDD continues to deliver on its promise as a cornerstone of modern drug discovery.

Author Contribution

HG and PCB conceived and wrote the paper.

Acknowledgements

We thank the Wellcome Trust and the Diamond Light Source (grant ID: COL0108 H.G.) for support. We thank Tobias Krojer, Kate Fieseler and Oleg Federov for critical reading of the manuscript. We also thank Frank von Delft and XChem group members, including Lizbé Koekemoer, Anthony Aimon, Warren Thomson, Ruben Sanchez Garcia, William Bradshaw, Conor Wild, Mihaela Smilova, Charlie Tomlinson, Romain Talon and the I04-1 beamline scientists for their guidance. We also acknowledge Irfan Alibay, Zhiyi Wu, Rocco Meli, Charlie Cook, and Cameron Anderson, John Spencer, David Mobley, Mehtap Işık, Thomas Grimes, Jack Scantlebury, John Chodera, Charlotte Deane, Paul Brennan, Brian Marsden, Storm Hassell-Hart and Edward Fitzgerald. H.G. also thank Anaïs de Bergeyck.

Competing Interests

The authors declare no competing interests

References

1 Scannell, J. W., Blanckley, A., Boldon, H. & Warrington, B. Diagnosing the decline in pharmaceutical R&D efficiency. *Nat. Rev. Drug. Discov.* **11**, 191-200 (2012). https://doi.org/10.1038/nrd3681

2 Lamoree, B. & Hubbard, R. Current perspectives in fragment-based lead discovery (FBLD). *Essays Biochem.* **61**, 453-464 (2017). https://doi.org/10.1042/EBC20170028

3 Rees, D. C., Congreve, M., Murray, C. W. & Carr, R. Fragment-based lead discovery. *Nature Rev. Drug Discov.* **3**, 660-672 (2004). https://doi.org/10.1038/nrd1467

4 St. Denis, J. D., Hall, R. J., Murray, C. W., Heightman, T. D. & Rees, D. C. Fragment-based drug discovery: opportunities for organic synthesis. *RSC Med. Chem.* **12**, 321-329 (2021). https://doi.org/10.1039/D0MD00375A

5 Osborne, J., Panova, S., Rapti, M., Urushima, T. & Jhoti, H. Fragments: where are we now? *Biochem. Soc. Trans.* **48**, 271-280 (2020). https://doi.org/10.1042/BST20190694

6 Walsh, L. *et al.* Fragment-to-lead medicinal chemistry publications in 2021. *J. Med. Chem.* **66**, 1137-1156 (2023). https://doi.org/10.1021/acs.jmedchem.2c01827

7 Erlanson, D. in *Practical Fragments* (2022).

8 Wang, Z.-Z., Shi, X.-X., Huang, G.-Y., Hao, G.-F. & Yang, G.-F. Fragment-based drug discovery supports drugging 'undruggable' protein–protein interactions. *Trends Biochem. Sci.* **48**, 539-552 (2023). https://doi.org/https://doi.org/10.1016/j.tibs.2023.01.008

9 Michaelides, I. N. & Collie, G. W. E3 Ligases Meet Their Match: Fragment-Based Approaches to Discover New E3 Ligands and to Unravel E3 Biology. *J. Med. Chem.* **66**, 3173-3194 (2023). https://doi.org/10.1021/acs.jmedchem.2c01882

10 Brosey, C. A. *et al.* Chemical screening by time-resolved X-ray scattering to discover allosteric probes. *Nature Chem. Biol.* **20**, 1199-1209 (2024). https://doi.org/10.1038/s41589-024-01609-1

11 Saeed, A. A., Klureza, M. A. & Hekstra, D. R. Mapping Protein Conformational Landscapes from Crystallographic Drug Fragment Screens. *J. Chem. Inf. Model.* **64**, 8937-8951 (2024). https://doi.org/10.1021/acs.jcim.4c01380

12 Krojer, T., Fraser, J. S. & Von Delft, F. Discovery of allosteric binding sites by crystallographic fragment screening. *Curr. Opin. Struc. Biol.* **65**, 209-216 (2020). https://doi.org/10.1016/j.sbi.2020.08.004

13 Aretz, J. *et al.* Chemical fragment arrays for rapid druggability assessment. *Chem. Comms.* **52**, 9067-9070 (2016). https://doi.org/10.1039/C5CC10457B

14 Southey, M. W. Y. & Brunavs, M. Introduction to small molecule drug discovery and preclinical development. *Front. Drug Discov.* **3** (2023). https://doi.org/10.3389/fddsv.2023.1314077

15 Rácz, A. *et al.* The changing landscape of medicinal chemistry optimization. *Nature Reviews Drug Discovery* **24**, 870-887 (2025). https://doi.org/10.1038/s41573-025-01225-1

16 Bon, M., Bilsland, A., Bower, J. & McAulay, K. Fragment‐based drug discovery–the importance of high‐quality molecule libraries. *Mol. Oncol.* **16**, 3761-3777 (2022). https://doi.org/10.1002/1878-0261.13277

17 Hall, R. J., Mortenson, P. N. & Murray, C. W. Efficient exploration of chemical space by fragment-based screening. *Prog. Biophys. Mol. Biol.* **116**, 82-91 (2014). https://doi.org/10.1016/j.pbiomolbio.2014.09.007

18 Leach, A. R. & Hann, M. M. Molecular complexity and fragment-based drug discovery: ten years on. *Curr. Opin. Chem. Biol.* **15**, 489-496 (2011). https://doi.org/https://doi.org/10.1016/j.cbpa.2011.05.008

19 Schultes, S. *et al.* Ligand efficiency as a guide in fragment hit selection and optimization. *Drug Discov. Today: Tech.* **7**, e157-e162 (2010). https://doi.org/https://doi.org/10.1016/j.ddtec.2010.11.003

20 Murray, C. W. & Rees, D. C. The rise of fragment-based drug discovery. *Nat. Chem.* **1**, 187-192 (2009). https://doi.org/10.1038/nchem.217

21 Barker, J., Courtney, S., Hesterkamp, T., Ullmann, D. & Whittaker, M. Fragment screening by biochemical assay. *Exp. Opin. Drug Disc.* **1**, 225-236 (2006). https://doi.org/10.1517/17460441.1.3.225

22 Kutchukian, P. S. *et al.* Large scale meta-analysis of fragment-based screening campaigns: Privileged fragments and complementary technologies. *SLAS Disc.* **20**, 588-596 (2015). https://doi.org/10.1177/1087057114565080

23 Ciulli, A. in *Protein-Ligand Interactions* Vol. 1008 (eds Mark A. Williams & Tina Daviter) 357-388 (Humana Press, 2013).

24 Robson-Tull, J. Biophysical screening in fragment-based drug design: a brief overview. *Biosci. Horiz. Int. J. Stud. Res.* **11** (2018). https://doi.org/10.1093/biohorizons/hzy015

25 Maveyraud, L. & Mourey, L. Protein X-ray Crystallography and Drug Discovery. *Molecules* **25**, 1030 (2020). https://doi.org/10.3390/molecules25051030

26 Zhu, K.-F. *et al.* Applications and prospects of cryo-EM in drug discovery. *Mil. Med. Res.* **10**, 10 (2023). https://doi.org/10.1186/s40779-023-00446-y

27 Kirkman, T., dos Santos Silva, C., Tosin, M. & Bertacine Dias, M. V. How to Find a Fragment: Methods for Screening and Validation in Fragment-Based Drug Discovery. *ChemMedChem* **19**, e202400342 (2024). https://doi.org/https://doi.org/10.1002/cmdc.202400342

28 Coyle, J. & Walser, R. Applied Biophysical Methods in Fragment-Based Drug Discovery. *SLAS Discov.* **25**, 471-490 (2020). https://doi.org/https://doi.org/10.1177/2472555220916168

29 Mashalidis, E. H., Śledź, P., Lang, S. & Abell, C. A three-stage biophysical screening cascade for fragment-based drug discovery. *Nat. Prot.* **8**, 2309-2324 (2013). https://doi.org/10.1038/nprot.2013.130

30 Dahlin, J. L. *et al.* PAINS in the Assay: Chemical Mechanisms of Assay Interference and Promiscuous Enzymatic Inhibition Observed during a Sulfhydryl-Scavenging HTS. *Journal of Medicinal Chemistry* **58**, 2091-2113 (2015). https://doi.org/10.1021/jm5019093

31 Aldrich, C. *et al.* The Ecstasy and Agony of Assay Interference Compounds. *ACS Central Science* **3**, 143-147 (2017). https://doi.org/10.1021/acscentsci.7b00069

32 Capuzzi, S. J., Muratov, E. N. & Tropsha, A. Phantom PAINS: Problems with the Utility of Alerts for Pan-Assay INterference CompoundS. *Journal of Chemical Information and Modeling* **57**, 417-427 (2017). https://doi.org/10.1021/acs.jcim.6b00465

33 Lee, K. L. *et al.* Discovery of Clinical Candidate 1-{[(2S,3S,4S)-3-Ethyl-4-fluoro-5-oxopyrrolidin-2-yl]methoxy}-7-methoxyisoquinoline-6-carboxamide (PF-06650833), a Potent, Selective Inhibitor of Interleukin-1 Receptor Associated Kinase 4 (IRAK4), by Fragment-Based Drug Design. *J. Med. Chem.* **60**, 5521-5542 (2017). https://doi.org/10.1021/acs.jmedchem.7b00231

34 Maurer, T. S., Edwards, M., Hepworth, D., Verhoest, P. & Allerton, C. M. N. Designing small molecules for therapeutic success: A contemporary perspective. *Drug Discov. Today* **27**, 538-546 (2022). https://doi.org/10.1016/j.drudis.2021.09.017

35 Sadybekov, A. V. & Katritch, V. Computational approaches streamlining drug discovery. *Nature* **616**, 673-685 (2023). https://doi.org/10.1038/s41586-023-05905-z

36 McCardle, K. Drug discovery with limited resources. *Nat. Comp. Sci.* **3**, 815-815 (2023). https://doi.org/10.1038/s43588-023-00546-8

37 Bergström, F. & Lindmark, B. Accelerated drug discovery by rapid candidate drug identification. *Drug Discov. Today* **24**, 1237-1241 (2019). https://doi.org/10.1016/j.drudis.2019.03.026

38 Wesolowski, S. S. & Brown, D. G. in *Meth. Princ. Med. Chem.* (ed Jörg Holenz) 487-512 (Wiley, 2016).

39 Murray, C. W., Newell, D. R. & Angibaud, P. A successful collaboration between academia, biotech and pharma led to discovery of erdafitinib, a selective FGFR inhibitor recently approved by the FDA. *MedChemComm* **10**, 1509-1511 (2019). https://doi.org/10.1039/C9MD90044F

40 Schuller, M. *et al.* Fragment binding to the Nsp3 macrodomain of SARS-CoV-2 identified through crystallographic screening and computational docking. *Sci. Adv.* **7**, eabf8711 https://doi.org/10.1126/sciadv.abf8711

41 De Fusco, C. *et al.* Fragment-Based Design of a Potent MAT2a Inhibitor and in Vivo Evaluation in an MTAP Null Xenograft Model. *Journal of Medicinal Chemistry* **64**, 6814-6826 (2021). https://doi.org/10.1021/acs.jmedchem.1c00067

42 Kirsch, P., Hartman, A. M., Hirsch, A. K. H. & Empting, M. Concepts and core principles of fragment-based drug design. *Molecules* **24**, 4309 (2019). https://doi.org/10.3390/molecules24234309

43 Shuker, S. B., Hajduk, P. J., Meadows, R. P. & Fesik, S. W. Discovering High-Affinity Ligands for Proteins: SAR by NMR. *Science* **274**, 1531-1534 (1996). https://doi.org/10.1126/science.274.5292.1531

44 Bedwell, E. V., McCarthy, W. J., Coyne, A. G. & Abell, C. Development of potent inhibitors by fragment-linking strategies. *Chem. Biol. Drug Des.* **100**, 469-486 (2022). https://doi.org/https://doi.org/10.1111/cbdd.14120

45 Yu, H. S. *et al.* General theory of fragment linking in molecular design: Why fragment linking rarely succeeds and how to improve outcomes. *J. Chem. Theory. Comput.* **17**, 450-462 (2021). https://doi.org/10.1021/acs.jctc.0c01004

46 Fearon, D. *et al.* Accelerating Drug Discovery With High-Throughput Crystallographic Fragment Screening and Structural Enablement. *Appl. Res.* **4**, e202400192 (2025). https://doi.org/https://doi.org/10.1002/appl.202400192

47 Xu, W. & Kang, C. Fragment-Based Drug Design: From Then until Now, and Toward the Future. *J. Med. Chem.* **68**, 5000-5004 (2025). https://doi.org/10.1021/acs.jmedchem.5c00424

48 Layton, M. E. *et al.* Discovery of MK-8189, a Highly Potent and Selective PDE10A Inhibitor for the Treatment of Schizophrenia. *J. Med. Chem.* **66**, 1157-1171 (2023). https://doi.org/10.1021/acs.jmedchem.2c01521

49 Cox, C. D. *et al.* Discovery of [11C]MK-8193 as a PET tracer to measure target engagement of phosphodiesterase 10A (PDE10A) inhibitors. *Bioorg. Med. Chem. Letts.* **25**, 4893-4898 (2015). https://doi.org/https://doi.org/10.1016/j.bmcl.2015.05.080

50 Woolford, A. J. A. *et al.* Fragment-Based Approach to the Development of an Orally Bioavailable Lactam Inhibitor of Lipoprotein-Associated Phospholipase A2 (Lp-PLA2). *J. Med. Chem.* **59**, 10738-10749 (2016). https://doi.org/10.1021/acs.jmedchem.6b01427

51 Powers, A. S. *et al.* Geometric Deep Learning for Structure-Based Ligand Design. *ACS Cent. Sci.* **9**, 2257-2267 (2023). https://doi.org/10.1021/acscentsci.3c00572

52 Chan, B. W. G. L. *et al.* Fragment-based drug discovery for disorders of the central nervous system: designing better drugs piece by piece. *Front. Chem.* **12** (2024).

53 Murray, C. W. & Blundell, T. L. Structural biology in fragment-based drug design. *Curr. Opin. Struct. Biol.* **20**, 497-507 (2010). https://doi.org/https://doi.org/10.1016/j.sbi.2010.04.003

54 Erlanson, D. A., Davis, B. J. & Jahnke, W. Fragment-Based Drug Discovery: Advancing Fragments in the Absence of Crystal Structures. *Cell Chem. Biol.* **26**, 9-15 (2019). https://doi.org/https://doi.org/10.1016/j.chembiol.2018.10.001

55 Bissaro, M., Sturlese, M. & Moro, S. The rise of molecular simulations in fragment-based drug design (FBDD): an overview. *Drug Discov. Today* **25**, 1693-1701 (2020). https://doi.org/10.1016/j.drudis.2020.06.023

56 Yoo, J., Jang, W. & Shin, W.-H. From part to whole: AI-driven progress in fragment-based drug discovery. *Curr. Opin. Struct. Biol.* **91**, 102995 (2025). https://doi.org/https://doi.org/10.1016/j.sbi.2025.102995

57 de Souza Neto, L. R. *et al.* In silico Strategies to Support Fragment-to-Lead Optimization in Drug Discovery. *Front. Chem.* **Volume 8 - 2020** (2020).

58 Erlanson, D. A. & Jahnke, W. *Fragment‐based Drug Discovery Lessons and Outlook*. (Wiley‐VCH Verlag GmbH & Co. KGaA, 2016).

59 Carbery, A., Skyner, R., von Delft, F. & Deane, C. M. Fragment libraries designed to be functionally diverse recover protein binding information more efficiently than standard structurally diverse libraries. *J. Med. Chem.* **65**, 11404-11413 (2022). https://doi.org/10.1021/acs.jmedchem.2c01004

60 Wollenhaupt, J. *et al.* F2X-Universal and F2X-Entry: Structurally Diverse Compound Libraries for Crystallographic Fragment Screening. *Structure* **28**, 694-706.e695 (2020). https://doi.org/10.1016/j.str.2020.04.019

61 Cox, O. B. *et al.* A poised fragment library enables rapid synthetic expansion yielding the first reported inhibitors of PHIP(2), an atypical bromodomain. *Chem. Sci.* **7**, 2322-2330 (2016). https://doi.org/10.1039/C5SC03115J

62 de Semir, D. *et al.* PHIP as a therapeutic target for driver-negative subtypes of melanoma, breast, and lung cancer. *Proc. Natl. Acad. Sci.* **115**, E5766-E5775 (2018). https://doi.org/10.1073/pnas.1804779115

63 Wood, D. J. *et al.* FragLites—minimal, halogenated fragments displaying pharmacophore doublets. An efficient approach to druggability assessment and hit generation. *J. Med. Chem.* **62**, 3741-3752 (2019). https://doi.org/10.1021/acs.jmedchem.9b00304

64 Klein, H. F., Hamilton, D. J., de Esch, I. J. P., Wijtmans, M. & O'Brien, P. Escape from planarity in fragment-based drug discovery: A synthetic strategy analysis of synthetic 3D fragment libraries. *Drug Discovery Today* **27**, 2484-2496 (2022). https://doi.org/https://doi.org/10.1016/j.drudis.2022.05.021

65 Docherty, R., Pencheva, K. & Abramov, Y. A. Low solubility in drug development: de-convoluting the relative importance of solvation and crystal packing. *J. Pharm. Pharmacol.* **67**, 847-856 (2015). https://doi.org/https://doi.org/10.1111/jphp.12393

66 Churcher, I., Newbold, S. & Murray, C. W. Return to Flatland. *Nature Rev. Chem.* **9**, 140-141 (2025). https://doi.org/10.1038/s41570-025-00688-5

67 Lucas, S. C. C. *et al.* Fragment screening at AstraZeneca: developing the next generation biophysics fragment set. *RSC Medicinal Chemistry* **13**, 1052-1057 (2022). https://doi.org/10.1039/D2MD00154C

68 Davison, G. *et al.* Mapping ligand interactions of bromodomains BRD4 and ATAD2 with FragLites and peplites—halogenated probes of druglike and peptide-like molecular interactions. *J. Med. Chem.* **65**, 15416-15432 (2022). https://doi.org/10.1021/acs.jmedchem.2c01357

69 Sydow, D., Schmiel, P., Mortier, J. & Volkamer, A. KinFragLib: Exploring the Kinase Inhibitor Space Using Subpocket-Focused Fragmentation and Recombination. *J. Chem. Inf. Model.* **60**, 6081-6094 (2020). https://doi.org/10.1021/acs.jcim.0c00839

70 Agrawal, A. *et al.* Chelator Fragment Libraries for Targeting Metalloproteinases. *ChemMedChem* **5**, 195-199 (2010). https://doi.org/https://doi.org/10.1002/cmdc.200900516

71 Keeley, A., Ábrányi-Balogh, P. & Keserű, G. M. Design and characterization of a heterocyclic electrophilic fragment library for the discovery of cysteine-targeted covalent inhibitors. *MedChemComm* **10**, 263-267 (2019). https://doi.org/10.1039/C8MD00327K

72 Douangamath, A. *et al.* Crystallographic and electrophilic fragment screening of the SARS-CoV-2 main protease. *Nat. Comms.* **11**, 5047 (2020). https://doi.org/10.1038/s41467-020-18709-w

73 Ostrem, J. M., Peters, U., Sos, M. L., Wells, J. A. & Shokat, K. M. K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions. *Nature* **503**, 548-551 (2013). https://doi.org/10.1038/nature12796

74 de Langen, A. J. *et al.* Sotorasib versus docetaxel for previously treated non-small-cell lung cancer with KRASG12C mutation: a randomised, open-label, phase 3 trial. *The Lancet* **401**, 733-746 (2023). https://doi.org/https://doi.org/10.1016/S0140-6736(23)00221-0

75 Patel, D., Bauman, J. D. & Arnold, E. Advantages of crystallographic fragment screening: Functional and mechanistic insights from a powerful platform for efficient drug discovery. *Prog. Biophys. Mol. Biol.* **116**, 92-100 (2014). https://doi.org/10.1016/j.pbiomolbio.2014.08.004

76 Cornaciu, I. *et al.* The automated crystallography pipelines at the EMBL HTX facility in grenoble. *J. Vis. Exp.*, 62491 (2021). https://doi.org/10.3791/62491

77 Douangamath, A. *et al.* Achieving efficient fragment screening at XChem facility at diamond light source. *J. Vis. Exp.*, 62414 (2021). https://doi.org/10.3791/62414

78 Lima, G. M. A. *et al.* FragMAX: the fragment-screening platform at the MAX IV laboratory. *Acta Cryst. Sec. D Struc. Biol.* **76**, 771-777 (2020). https://doi.org/10.1107/S205979832000889X

79 Pearce, N. M. *et al.* A multi-crystal method for extracting obscured crystallographic states from conventionally uninterpretable electron density. *Nat. Comms.* **8**, 15123 (2017). https://doi.org/10.1038/ncomms15123

80 Ng, J. T., Dekker, C., Reardon, P. & von Delft, F. Lessons from ten years of crystallization experiments at the SGC. *Acta Cryst. Sec. D, Struc. Biol.* **72**, 224-235 (2016). https://doi.org/10.1107/S2059798315024687

81 Kaščáková, B., Koutská, A., Burdová, M., Havlíčková, P. & Kutá Smatanová, I. Revealing protein structures: crystallization of protein-ligand complexes – co-crystallization and crystal soaking. *FEBS Open Bio* **15**, 542-550 (2025). https://doi.org/https://doi.org/10.1002/2211-5463.13913

82 Wienen‐Schmidt, B., Oebbeke, M., Ngo, K., Heine, A. & Klebe, G. Two methods, one goal: Structural differences between cocrystallization and crystal soaking to discover ligand binding poses. *ChemMedChem* **16**, 292-300 (2021). https://doi.org/10.1002/cmdc.202000565

83 Fischer, M., Shoichet, B. K. & Fraser, J. S. One crystal, two temperatures: Cryocooling penalties alter ligand binding to transient protein sites. *ChemBioChem* **16**, 1560-1564 (2015). https://doi.org/10.1002/cbic.201500196

84 Schiebel, J. *et al.* One question, multiple answers: Biochemical and biophysical screening methods retrieve deviating fragment hit lists. *ChemMedChem* **10**, 1511-1521 (2015). https://doi.org/10.1002/cmdc.201500267

85 Schiebel, J. *et al.* Six Biophysical Screening Methods Miss a Large Proportion of Crystallographically Discovered Fragment Hits: A Case Study. *ACS Chem. Biol.* **11**, 1693-1701 (2016). https://doi.org/10.1021/acschembio.5b01034

86 Ehrmann, F. R. *et al.* Soaking suggests "alternative facts": Only co-crystallization discloses major ligand-induced interface rearrangements of a homodimeric tRNA-binding protein indicating a novel mode-of-inhibition. *PLOS ONE* **12**, e0175723 (2017). https://doi.org/10.1371/journal.pone.0175723

87 Vantieghem, T., Osipov, E. M., Beelen, S. & Strelkov, S. V. Crystal engineering of the hepatoma-derived growth factor-related protein 2 PWWP domain towards crystallographic fragment screening. *Acta Crystallographica Section F* **81**, 358-364 (2025). https://doi.org/doi:10.1107/S2053230X25006302

88 Saur, M. *et al.* Fragment-based drug discovery using cryo-EM. *Drug Discovery Today* **25**, 485-490 (2020). https://doi.org/https://doi.org/10.1016/j.drudis.2019.12.006

89 Wang, H.-W. & Wang, J.-W. How cryo-electron microscopy and X-ray crystallography complement each other. *Protein Science* **26**, 32-39 (2017). https://doi.org/https://doi.org/10.1002/pro.3022

90 Wu, M. & Lander, G. C. How low can we go? Structure determination of small biological complexes using single-particle cryo-EM. *Current Opinion in Structural Biology* **64**, 9-16 (2020). https://doi.org/https://doi.org/10.1016/j.sbi.2020.05.007

91 Dubach, V. R. A. & Guskov, A. The Resolution in X-ray Crystallography and Single-Particle Cryogenic Electron Microscopy. *Crystals* **10** (2020).

92 Zhou, W. *et al.* Cryo-EM structure-based selection of computed ligand poses enables design of MTA-synergic PRMT5 inhibitors of better potency. *Communications Biology* **5**, 1054 (2022). https://doi.org/10.1038/s42003-022-03991-9

93 Scapin, G., Potter, C. S. & Carragher, B. Cryo-EM for Small Molecules Discovery, Design, Understanding, and Application. *Cell Chemical Biology* **25**, 1318-1325 (2018). https://doi.org/https://doi.org/10.1016/j.chembiol.2018.07.006

94 Sweeney, A., Mulvaney, T., Maiorca, M. & Topf, M. ChemEM: Flexible Docking of Small Molecules in Cryo-EM Structures. *Journal of Medicinal Chemistry* **67**, 199-212 (2024). https://doi.org/10.1021/acs.jmedchem.3c01134

95 Lawson, C. L. *et al.* Outcomes of the EMDataResource cryo-EM Ligand Modeling Challenge. *Nature Methods* **21**, 1340-1348 (2024). https://doi.org/10.1038/s41592-024-02321-7

96 Cushing, V. I. *et al.* High-resolution cryo-EM of the human CDK-activating kinase for structure-based drug design. *Nat. Comms.* **15**, 2265 (2024). https://doi.org/10.1038/s41467-024-46375-9

97 Biggin, P. C. & Hubbard, R. E. Editorial overview - New Concepts in Drug Discovery (2025). *Curr. Opin. Struct. Biol.* **93**, 103069 (2025). https://doi.org/https://doi.org/10.1016/j.sbi.2025.103069

98 Szlávik, Z. *et al.* Structure-Guided Discovery of a Selective Mcl-1 Inhibitor with Cellular Activity. *J. Med. Chem.* **62**, 6913-6924 (2019). https://doi.org/10.1021/acs.jmedchem.9b00134

99 Szlavik, Z. *et al.* Discovery of S64315, a Potent and Selective Mcl-1 Inhibitor. *J. Med. Chem.* **63**, 13762-13795 (2020). https://doi.org/10.1021/acs.jmedchem.0c01234

100 Mukherjee, N. *et al.* MCL1 inhibition targets Myeloid Derived Suppressors Cells, promotes antitumor immunity and enhances the efficacy of immune checkpoint blockade. *Cell Death & Disease* **15**, 198 (2024). https://doi.org/10.1038/s41419-024-06524-w

101 Torres, F. *et al.* Protein–fragment complex structures derived by NMR molecular replacement. *RSC Med. Chem.* **11**, 591-596 (2020). https://doi.org/10.1039/D0MD00068J

102 Grosjean, H. *et al.* SAMPL7 protein-ligand challenge: A community-wide evaluation of computational methods against fragment screening and pose-prediction. *J. Comput-Aid Mol. Des.* **36**, 291-311 (2022). https://doi.org/10.1007/s10822-022-00452-7

103 Kneller, D. W. *et al.* Direct observation of protonation state modulation in SARS-CoV-2 main protease upon inhibitor binding with neutron crystallography. *J. Med. Chem.* **64**, 4991-5000 (2021). https://doi.org/10.1021/acs.jmedchem.1c00058

104 Ma, W. *et al.* Using macromolecular electron densities to improve the enrichment of active compounds in virtual screening. *Commun. Chem.* **6**, 1-12 (2023). https://doi.org/10.1038/s42004-023-00984-5

105 Skaist Mehlman, T. *et al.* Room-temperature crystallography reveals altered binding of small-molecule fragments to PTP1B. *eLife* **12**, e84632 (2023). https://doi.org/10.7554/eLife.84632

106 Bajusz, D. *et al.* Exploring protein hotspots by optimized fragment pharmacophores. *Nat. Comms.* **12**, 3201 (2021). https://doi.org/10.1038/s41467-021-23443-y

107 Gechijian, L. N. *et al.* Functional TRIM24 degrader via conjugation of ineffectual bromodomain and VHL ligands. *Nature Chemical Biology* **14**, 405-412 (2018). https://doi.org/10.1038/s41589-018-0010-y

108 Bolz, S. N., Adasme, M. F. & Schroeder, M. Toward an Understanding of Pan-Assay Interference Compounds and Promiscuity: A Structural Perspective on Binding Modes. *Journal of Chemical Information and Modeling* **61**, 2248-2262 (2021). https://doi.org/10.1021/acs.jcim.0c01227

109 Papalia, G. A. *et al.* Comparative analysis of 10 small molecules binding to carbonic anhydrase II by different investigators using Biacore technology. *Analytical Biochemistry* **359**, 94-105 (2006). https://doi.org/https://doi.org/10.1016/j.ab.2006.08.021

110 FitzGerald, E. A. *et al.* Identification of fragments targeting SMYD3 using highly sensitive kinetic and multiplexed biosensor-based screening. *RSC Med. Chem.* **15**, 1982-1990 (2024). https://doi.org/10.1039/D4MD00093E

111 Giannetti, A. M., Koch, B. D. & Browner, M. F. Surface Plasmon Resonance Based Assay for the Detection and Characterization of Promiscuous Inhibitors. *Journal of Medicinal Chemistry* **51**, 574-580 (2008). https://doi.org/10.1021/jm700952v

112 Chavanieu, A. & Pugnière, M. Developments in SPR fragment screening. *Exp. Opin. Drug Discov.* **11**, 489-499 (2016). https://doi.org/10.1517/17460441.2016.1160888

113 Dart, M. L. *et al.* Homogeneous Assay for Target Engagement Utilizing Bioluminescent Thermal Shift. *ACS Medicinal Chemistry Letters* **9**, 546-551 (2018). https://doi.org/10.1021/acsmedchemlett.8b00081

114 Navratilova, I. & Hopkins, A. L. Emerging role of surface plasmon resonance in fragment-based drug discovery. *Future Med. Chem.* **3**, 1809-1820 (2011). https://doi.org/10.4155/fmc.11.128

115 Baell, J. & Walters, M. A. Chemistry: Chemical con artists foil drug discovery. *Nature* **513**, 481-483 (2014). https://doi.org/10.1038/513481a

116 Johnson, J. A., Olson, N. M., Tooker, M. J., Bur, S. K. & Pomerantz, W. C. K. Combined Protein- and Ligand-Observed NMR Workflow to Screen Fragment Cocktails against Multiple Proteins: A Case Study Using Bromodomains. *Molecules* **25** (2020).

117 Ramon, C.-O. NMR Screening and Hit Validation in Fragment Based Drug Discovery. *Current Topics in Medicinal Chemistry* **11**, 43-67 (2011). https://doi.org/http://dx.doi.org/10.2174/156802611793611887

118 Sternicki, L. M. & Poulsen, S.-A. Fragment-based drug discovery campaigns guided by native mass spectrometry. *RSC Med. Chem.* **15**, 2270-2285 (2024). https://doi.org/10.1039/D4MD00273C

119 Resnick, E. *et al.* Rapid covalent-probe discovery by electrophile-fragment screening. *J. Am. Chem. Soc.* **141**, 8951-8968 (2019). https://doi.org/10.1021/jacs.9b02822

120 Bolcato, G., Bissaro, M., Sturlese, M. & Moro, S. Comparing Fragment Binding Poses Prediction Using HSP90 as a Key Study: When Bound Water Makes the Difference. *Molecules* **25** (2020).

121 Bolcato, G. *et al.* A computational workflow for the identification of novel fragments acting as inhibitors of the activity of protein kinase CK1δ. *Int. J. Mol. Sci.* **22**, 9741 (2021). https://doi.org/10.3390/ijms22189741

122 Greisman, J. B. *et al.* Discovery and validation of the binding poses of allosteric fragment hits to protein tyrosine phosphatase 1b: From molecular dynamics simulations to X-ray crystallography. *J. Chem. Inf. Model.* **63**, 2644-2650 (2023). https://doi.org/10.1021/acs.jcim.3c00236

123 Shaw, D. E. *et al.* in *Proc. Int. Conf. High Perf. Comp. Networking, Storage and Analysis* Article 1 (Association for Computing Machinery, St. Louis, Missouri, 2021).

124 Larson, S. M., Snow, C. D., Shirts, M. & Pande, V. S. Folding@Home and Genome@Home: Using distributed computing to tackle previously intractable problems in computational biology. *arXiv* **0901.0866** (2009).

125 Ferrari, F. *et al.* HT-SuMD: making molecular dynamics simulations suitable for fragment-based screening. A comparative study with NMR. *J. Enz. Inhib. Med. Chem.* **36**, 1-14 (2021). https://doi.org/10.1080/14756366.2020.1838499

126 Linker, S. M., Magarkar, A., Köfinger, J., Hummer, G. & Seeliger, D. Fragment Binding Pose Predictions Using Unbiased Simulations and Markov-State Models. *J. Chem. Theory. Comput.* **15**, 4974-4981 (2019). https://doi.org/10.1021/acs.jctc.9b00069

127 Schmitz, B., Frieg, B., Homeyer, N., Jessen, G. & Gohlke, H. Extracting binding energies and binding modes from biomolecular simulations of fragment binding to endothiapepsin. *Archiv der Pharmazie* **357**, 2300612 (2024). https://doi.org/https://doi.org/10.1002/ardp.202300612

128 Lu, W. *et al.* Fragment-based covalent ligand discovery. *RSC Chem. Biol.* **2**, 354-367 (2021). https://doi.org/10.1039/D0CB00222D

129 McAulay, K., Bilsland, A. & Bon, M. Reactivity of Covalent Fragments and Their Role in Fragment Based Drug Discovery. *Pharmaceuticals* **15** (2022).

130 Hillebrand, L., Liang, X. J., Serafim, R. A. M. & Gehringer, M. Emerging and Re-emerging Warheads for Targeted Covalent Inhibitors: An Update. *J. Med. Chem.* **67**, 7668-7758 (2024). https://doi.org/10.1021/acs.jmedchem.3c01825

131 Mukherjee, H. *et al.* Discovery and optimization of covalent Bcl-xL antagonists. *Bioorg. Med. Chem. Letts* **29**, 126682 (2019). https://doi.org/https://doi.org/10.1016/j.bmcl.2019.126682

132 Csorba, N., Ábrányi-Balogh, P. & Keserű, G. M. Covalent fragment approaches targeting non-cysteine residues. *Trends Pharmacol. Sci* **44**, 802-816 (2023). https://doi.org/https://doi.org/10.1016/j.tips.2023.08.014

133 Pettinger, J., Jones, K. & Cheeseman, M. D. Lysine-Targeting Covalent Inhibitors. *Ang. Chem. Int. Ed.* **56**, 15200-15209 (2017). https://doi.org/https://doi.org/10.1002/anie.201707630

134 Ishikita, H. Origin of the pK a shift of the catalytic lysine in acetoacetate decarboxylase. *FEBS Letts* **584**, 3464-3468 (2010). https://doi.org/https://doi.org/10.1016/j.febslet.2010.07.003

135 Tolmachova, K. A. *et al.* (Chlorosulfonyl)benzenesulfonyl Fluorides—Versatile Building Blocks for Combinatorial Chemistry: Design, Synthesis and Evaluation of a Covalent Inhibitor Library. *ACS Comb. Sci.* **20**, 672-680 (2018). https://doi.org/10.1021/acscombsci.8b00130

136 Filippakopoulos, P. *et al.* Histone Recognition and Large-Scale Structural Analysis of the Human Bromodomain Family. *Cell* **149**, 214-231 (2012). https://doi.org/https://doi.org/10.1016/j.cell.2012.02.013

137 Ghaby, K. & Roux, B. Kinetic Modeling of Covalent Inhibition: Effects of Rapidly Fluctuating Intermediate States. *bioRxiv*, 2025.2005.2028.656658 (2025). https://doi.org/10.1101/2025.05.28.656658

138 Boettcher, A. *et al.* Fragment-Based Screening by Biochemical Assays: Systematic Feasibility Studies with Trypsin and MMP12. *SLAS Discov.* **15**, 1029-1041 (2010). https://doi.org/https://doi.org/10.1177/1087057110380455

139 Craven, G. B. *et al.* High-Throughput Kinetic Analysis for Target-Directed Covalent Ligand Discovery. *Ang. Chem. Int. Ed.* **57**, 5257-5261 (2018). https://doi.org/https://doi.org/10.1002/anie.201711825

140 Dubiella, C. *et al.* Sulfopin is a covalent inhibitor of Pin1 that blocks Myc-driven tumors in vivo. *Nat. Chem. Biol.* **17**, 954-963 (2021). https://doi.org/10.1038/s41589-021-00786-7

141 Valenti, D., Hristeva, S., Tzalis, D. & Ottmann, C. Clinical candidates modulating protein-protein interactions: The fragment-based experience. *Eur. J. Med. Chem.* **167**, 76-95 (2019). https://doi.org/https://doi.org/10.1016/j.ejmech.2019.01.084

142 Williams, G., Ferenczy, G. G., Ulander, J. & Keserű, G. M. Binding thermodynamics discriminates fragments from druglike compounds: a thermodynamic description of fragment-based drug discovery. *Drug Discovery Today* **22**, 681-689 (2017). https://doi.org/https://doi.org/10.1016/j.drudis.2016.11.019

143 Rühmann, E., Betz, M., Heine, A. & Klebe, G. Fragment Binding Can Be Either More Enthalpy-Driven or Entropy-Driven: Crystal Structures and Residual Hydration Patterns Suggest Why. *J. Med. Chem.* **58**, 6960-6971 (2015). https://doi.org/10.1021/acs.jmedchem.5b00812

144 Schultes, S. *et al.* Ligand efficiency as a guide in fragment hit selection and optimization. *Drug Discovery Today: Technologies* **7**, e157-e162 (2010). https://doi.org/https://doi.org/10.1016/j.ddtec.2010.11.003

145 Smilova, M. D. *et al.* Fragment Hotspot Mapping to Identify Selectivity-Determining Regions between Related Proteins. *J. Chem. Inf. Model.* **62**, 284-294 (2022). https://doi.org/10.1021/acs.jcim.1c00823

146 Thomas, A. M. *et al.* Mutate and Conjugate: A Method to Enable Rapid In-Cell Target Validation. *ACS Chem. Biol.* **18**, 2405-2417 (2023). https://doi.org/10.1021/acschembio.3c00437

147 Müller, J. *et al.* 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.* **65**, 15663-15678 (2022). https://doi.org/10.1021/acs.jmedchem.2c00813

148 Ferenczy, G. G. & Keserű, G. M. Thermodynamics of fragment binding. *J. Chem. Inf. Model.* **52**, 1039-1045 (2012). https://doi.org/10.1021/ci200608b

149 Ferenczy, G. G. & Keserű, G. M. Thermodynamic profiling for fragment-based lead discovery and optimization. *Exp. Opin. Drug Discov.* **15**, 117-129 (2020). https://doi.org/10.1080/17460441.2020.1691166

150 Johnson, J. A. *et al.* Evaluating the Advantages of Using 3D-Enriched Fragments for Targeting BET Bromodomains. *ACS Med. Chem. Letts.* **10**, 1648-1654 (2019). https://doi.org/10.1021/acsmedchemlett.9b00414

151 Baybekov, S. *et al.* DMSO Solubility Assessment for Fragment-Based Screening. *Molecules* **26** (2021).

152 Baell, J. B., Ferrins, L., Falk, H. & Nikolakopoulos, G. PAINS: Relevance to Tool Compound Discovery and Fragment-Based Screening. *Aus. J. Chem.* **66**, 1483-1494 (2013).

153 Shi, Y. & von Itzstein, M. How Size Matters: Diversity for Fragment Library Design. *Molecules* **24** (2019).

154 Lucas, S. C. C. *et al.* Fragment screening at AstraZeneca: developing the next generation biophysics fragment set. *RSC Med. Chem.* **13**, 1052-1057 (2022). https://doi.org/10.1039/D2MD00154C

155 McGovern, S. L., Caselli, E., Grigorieff, N. & Shoichet, B. K. A Common Mechanism Underlying Promiscuous Inhibitors from Virtual and High-Throughput Screening. *J. Med. Chem.* **45**, 1712-1722 (2002). https://doi.org/10.1021/jm010533y

156 Edink, E., Jansen, C., Leurs, R. & de Esch, I. J. P. The heat is on: thermodynamic analysis in fragment-based drug discovery. *Drug Discov. Today: Tech.* **7**, e189-e201 (2010). https://doi.org/https://doi.org/10.1016/j.ddtec.2010.12.001

157 Ferenczy, G. G. & Keserű, G. M. On the enthalpic preference of fragment binding. *MedChemComm* **7**, 332-337 (2016). https://doi.org/10.1039/C5MD00542F

158 Metz, A. *et al.* Frag4Lead: growing crystallographic fragment hits by catalog using fragment-guided template docking. *Acta Cryst. Sec. D Struc. Biol.* **77**, 1168-1182 (2021). https://doi.org/10.1107/S2059798321008196

159 Huschmann, F. U. *et al.* Screening, Growing, and Validation by Catalog: Using Synthetic Intermediates from Natural Product Libraries to Discover Fragments for an Aspartic Protease Through Crystallography. *Crystals* **14** (2024).

160 Muratov, E. N. *et al.* QSAR without borders. *Chem. Soc. Rev.* **49**, 3525-3564 (2020). https://doi.org/10.1039/D0CS00098A

161 Penzo, M. *et al.* High-throughput screening of the Plasmodium falciparum cGMP-dependent protein kinase identified a thiazole scaffold which kills erythrocytic and sexual stage parasites. *Scientific Reports* **9**, 7005 (2019). https://doi.org/10.1038/s41598-019-42801-x

162 Provins, L., Jnoff, E. & Genicot, C. Back-up strategies in drug discovery: what, how and when? *Drug Discovery Today* **19**, 1808-1811 (2014). https://doi.org/https://doi.org/10.1016/j.drudis.2014.07.001

163 Born, J. R. *et al.* The Impact of Assay Design on Medicinal Chemistry: Case Studies. *SLAS Discovery* **26**, 1243-1255 (2021). https://doi.org/https://doi.org/10.1177/24725552211026238

164 Jenkinson, S., Schmidt, F., Rosenbrier Ribeiro, L., Delaunois, A. & Valentin, J.-P. A practical guide to secondary pharmacology in drug discovery. *J. Pharmacol. Toxicol. Meth.* **105**, 106869 (2020). https://doi.org/https://doi.org/10.1016/j.vascn.2020.106869

165 Mukaidaisi, M., Vu, A., Grantham, K., Tchagang, A. & Li, Y. Multi-Objective Drug Design Based on Graph-Fragment Molecular Representation and Deep Evolutionary Learning. *Front. Pharm.* **13** (2022). https://doi.org/10.3389/fphar.2022.920747

166 Chen, D., Ranganathan, A., Ijzerman, A. P., Siegal, G. & Carlsson, J. Complementarity between in Silico and Biophysical Screening Approaches in Fragment-Based Lead Discovery against the A2A Adenosine Receptor. *J. Chem. Inf. Model.* **53**, 2701-2714 (2013). https://doi.org/10.1021/ci4003156

167 Matricon, P. *et al.* Fragment optimization for GPCRs by molecular dynamics free energy calculations: Probing druggable subpockets of the A2Aadenosine receptor binding site. *Sci. Reps.* **7**, 6398 (2017). https://doi.org/10.1038/s41598-017-04905-0

168 Huang, X. *et al.* Hit-to-Lead Optimization of Heterocyclic Carbonyloxycarboximidamides as Selective Antagonists at Human Adenosine A3 Receptor. *J. Med. Chem.* **67**, 13117-13146 (2024). https://doi.org/10.1021/acs.jmedchem.4c01092

169 Hostetler, E. D. *et al.* Preclinical Characterization of the Phosphodiesterase 10A PET Tracer [11C]MK-8193. *Mol. Imag. Biol.* **18**, 579-587 (2016). https://doi.org/10.1007/s11307-015-0910-0

170 Shipe, W. D. *et al.* Discovery and Optimization of a Series of Pyrimidine-Based Phosphodiesterase 10A (PDE10A) Inhibitors through Fragment Screening, Structure-Based Design, and Parallel Synthesis. *J. Med. Chem.* **58**, 7888-7894 (2015). https://doi.org/10.1021/acs.jmedchem.5b00983

171 Fromer, J. C. & Coley, C. W. Computer-aided multi-objective optimization in small molecule discovery. *Patterns* **4**, 100678 (2023). https://doi.org/https://doi.org/10.1016/j.patter.2023.100678

172 Fehlis, Y. *et al.* Accelerating Drug Discovery Through Agentic AI: A Multi-Agent Approach to Laboratory Automation in the DMTA Cycle. *arXiv* **2507.09023** (2025). https://doi.org/https://doi.org/10.48550/arXiv.2507.09023

173 Gioiello, A., Piccinno, A., Lozza, A. M. & Cerra, B. The Medicinal Chemistry in the Era of Machines and Automation: Recent Advances in Continuous Flow Technology. *J. Med. Chem.* **63**, 6624-6647 (2020). https://doi.org/10.1021/acs.jmedchem.9b01956

174 Schneider, G. Automating drug discovery. *Nat. Rev. Drug Discov.* **17**, 97-113 (2018). https://doi.org/10.1038/nrd.2017.232

175 Descamps, C. *et al.* Growing and linking optimizers: synthesis-driven molecule design. *Briefings in Bioinformatics* **26**, bbaf482 (2025). https://doi.org/10.1093/bib/bbaf482

176 Müller, S. *et al.* Target 2035 – update on the quest for a probe for every protein. *RSC Medicinal Chemistry* **13**, 13-21 (2022). https://doi.org/10.1039/D1MD00228G

177 Chodera, J., Lee, A. A., London, N. & von Delft, F. Crowdsourcing drug discovery for pandemics. *Nature Chemistry* **12**, 581-581 (2020). https://doi.org/10.1038/s41557-020-0496-2

178 Griffen, E. J. *et al.* Open-science discovery of DNDI-6510, a compound that addresses genotoxic and metabolic liabilities of the COVID Moonshot SARS-CoV-2 Mpro lead inhibitor. *bioRxiv*, 2025.2006.2016.660018 (2025). https://doi.org/10.1101/2025.06.16.660018

179 Howes, L. Pan-coronavirus antiviral unveiled. *C&EN Global Enterprise* **103**, 6-6 (2025). https://doi.org/10.1021/cen-10309-scicon4

180 Boby, M. L. *et al.* Open science discovery of potent noncovalent SARS-CoV-2 main protease inhibitors. *Science* **382**, eabo7201 https://doi.org/10.1126/science.abo7201

181 Warr, W. A., Nicklaus, M. C., Nicolaou, C. A. & Rarey, M. Exploration of ultralarge compound collections for drug discovery. *J. Chem. Inf. Model.* **62**, 2021-2034 (2022). https://doi.org/10.1021/acs.jcim.2c00224

182 Benet, L. Z., Hosey, C. M., Ursu, O. & Oprea, T. I. BDDCS, the Rule of 5 and drugability. *Adv. Drug Del. Rev.* **101**, 89-98 (2016). https://doi.org/10.1016/j.addr.2016.05.007

183 Grygorenko, O. O. *et al.* Generating Multibillion Chemical Space of Readily Accessible Screening Compounds. *iScience* **23**, 101681 (2020). https://doi.org/https://doi.org/10.1016/j.isci.2020.101681

184 Bellmann, L., Penner, P., Gastreich, M. & Rarey, M. Comparison of Combinatorial Fragment Spaces and Its Application to Ultralarge Make-on-Demand Compound Catalogs. *Journal of Chemical Information and Modeling* **62**, 553-566 (2022). https://doi.org/10.1021/acs.jcim.1c01378

185 Sun, D., Gao, W., Hu, H. & Zhou, S. Why 90% of clinical drug development fails and how to improve it? *Acta Pharmaceutica Sinica B* **12**, 3049-3062 (2022). https://doi.org/https://doi.org/10.1016/j.apsb.2022.02.002

186 Sanchez-Garcia, R. *et al.* CoPriNet: graph neural networks provide accurate and rapid compound price prediction for molecule prioritisation. *Digital Discov.* **2**, 103-111 (2023). https://doi.org/10.1039/D2DD00071G

187 Abderrahmane, M., Tajmouati, H., Barros Ribeiro da Silva, V. & Perron, Q. Predicting the price of molecules using their predicted synthetic pathways. *ChemRxiv* **doi:10.26434/chemrxiv-2024-8wcfp** (2024). https://doi.org/doi:10.26434/chemrxiv-2024-8wcfp

188 Hall, R. J., Murray, C. W. & Verdonk, M. L. The fragment network: A chemistry recommendation engine built using a graph database. *J. Med. Chem.* **60**, 6440-6450 (2017). https://doi.org/10.1021/acs.jmedchem.7b00809

189 Wills, S. *et al.* Fragment merging using a graph database samples different catalogue space than similarity search. *J. Chem. Inf. Model.* **63**, 3423-3437 (2023). https://doi.org/10.1021/acs.jcim.3c00276

190 Fieseler, K. K. *et al.* Syndirella: Synthesis-directed fragment elaboration enables extensive binding site exploration beyond catalog compounds. . 2025; doi:10.26434/chemrxiv-2025-0jd39-v2 This content is a preprint and has not been peer-reviewed. *ChemRxiv* (2025). https://doi.org/0.26434/chemrxiv-2025-0jd39-v2

191 Vogt, M. Exploring chemical space — Generative models and their evaluation. *Art. Intel. Life Sci.* **3**, 100064 (2023). https://doi.org/https://doi.org/10.1016/j.ailsci.2023.100064

192 Ochiai, T. *et al.* Variational autoencoder-based chemical latent space for large molecular structures with 3D complexity. *Comm. Chem.* **6**, 249 (2023). https://doi.org/10.1038/s42004-023-01054-6

193 Chen, B., Fu, X., Barzilay, R. & Jaakkola, T. Fragment-based Sequential Translation for Molecular Optimization. *arXiv* **2111**, 01009 (2021).

194 Perron, Q. *et al.* Deep generative models for ligand‐based de novo design applied to multi‐parametric optimization. *J. Comp. Chem.* **43**, 692-703 (2022). https://doi.org/10.1002/jcc.26826

195 Parrot, M. *et al.* Integrating synthetic accessibility with AI-based generative drug design. *J. Cheminf.* **15**, 83 (2023). https://doi.org/10.1186/s13321-023-00742-8

196 Saikia, S. & Bordoloi, M. Molecular docking: Challenges, advances and its use in drug discovery perspective. *Current Drug Targets* **20**, 501-521 (2019). https://doi.org/10.2174/1389450119666181022153016

197 Scantlebury, J. *et al.* A small step toward generalizability: Training a machine learning scoring function for structure-based virtual screening. *J. Chem. Inf. Model.* **63**, 2960-2974 (2023). https://doi.org/10.1021/acs.jcim.3c00322

198 Meli, R., Morris, G. M. & Biggin, P. C. Scoring Functions for Protein-Ligand Binding Affinity Prediction Using Structure-based Deep Learning: A Review. *Front. Bioinf.* **Volume 2 - 2022** (2022).

199 Fischer, M., Coleman, R. G., Fraser, J. S. & Shoichet, B. K. Incorporation of protein flexibility and conformational energy penalties in docking screens to improve ligand discovery. *Nat. Chem.* **6**, 575-583 (2014). https://doi.org/10.1038/nchem.1954

200 Leung, S., Bodkin, M., Von Delft, F., Brennan, P. & Morris, G. SuCOS is Better than RMSD for Evaluating Fragment Elaboration and Docking Poses. (Chemistry, 2019).

201 Bergues, N. *et al.* Template-Guided 3D Molecular Pose Generation via Flow Matching and Differentiable Optimization. *arXiv* **arXiv:2506.06305v1** (2025).

202 Silva, G. M. d. *et al.* Covalent docking and molecular dynamics simulations reveal the specificity-shifting mutations Ala237Arg and Ala237Lys in TEM beta-lactamase. *PLOS Comp. Biol.* **18**, e1009944 (2022). https://doi.org/10.1371/journal.pcbi.1009944

203 Imrie, F., Bradley, A. R., Van Der Schaar, M. & Deane, C. M. Deep generative models for 3D linker design. *J. Chem. Inf. Model.* **60**, 1983-1995 (2020). https://doi.org/10.1021/acs.jcim.9b01120

204 Thomas, M., Bender, A. & De Graaf, C. Integrating structure-based approaches in generative molecular design. *Curr. Opin. Struc. Biol.* **79**, 102559 (2023). https://doi.org/10.1016/j.sbi.2023.102559

205 Ghorbani, M., Gendelev, L., Beroza, P. & Keiser, M. J. Autoregressive fragment-based diffusion for pocket-aware ligand design. *arXiv* **2401**, 05370 (2023). https://doi.org/https://arxiv.org/abs/2401.05370

206 Zhang, O. *et al.* FragGen: towards 3D geometry reliable fragment-based molecular generation. *Chemical Science* **15**, 19452-19465 (2024). https://doi.org/10.1039/D4SC04620J

207 Valsson, Í. *et al.* Narrowing the gap between machine learning scoring functions and free energy perturbation using augmented data. *Comm. Chem.* **8**, 41 (2025). https://doi.org/10.1038/s42004-025-01428-y

208 Hsu, W.-T., Grevtsev, S., Douglas, T., Magarkar, A. & Biggin, P. C. Can AI-predicted complexes teach machine learning to compute drug binding affinity? *arXiv*, 2507.07882 (2025). https://doi.org/10.48550/arXiv.2507.07882

209 Škrinjar, P., Eberhardt, J., Durairaj, J. & Schwede, T. Have protein-ligand co-folding methods moved beyond memorisation? *bioRxiv*, 2025.2002.2003.636309 (2025). https://doi.org/10.1101/2025.02.03.636309

210 Alibay, I., Magarkar, A., Seeliger, D. & Biggin, P. C. Evaluating the use of absolute binding free energy in the fragment optimisation process. *Commun. Chem.* **5**, 105 (2022). https://doi.org/10.1038/s42004-022-00721-4

211 Aldeghi, M., Heifetz, A., Bodkin, M. J., Knapp, S. & Biggin, P. C. Accurate calculation of the absolute free energy of binding for drug molecules. *Chemical Science* **7**, 207-218 (2016). https://doi.org/10.1039/C5SC02678D

212 Cappel, D., Jerome, S., Hessler, G. & Matter, H. Impact of Different Automated Binding Pose Generation Approaches on Relative Binding Free Energy Simulations. *J. Chem. Inf. Model.* **60**, 1432-1444 (2020). https://doi.org/10.1021/acs.jcim.9b01118

213 Boby, M. L. *et al.* Open science discovery of potent noncovalent SARS-CoV-2 main protease inhibitors. *Science* **382**, eabo7201 (2023). https://doi.org/10.1126/science.abo7201

214 Chen, D. *et al.* Fragment Screening of GPCRs Using Biophysical Methods: Identification of Ligands of the Adenosine A2A Receptor with Novel Biological Activity. *ACS Chemical Biology* **7**, 2064-2073 (2012). https://doi.org/10.1021/cb300436c

215 Matricon, P. *et al.* Fragment-based design of selective GPCR ligands guided by free energy simulations. *Chem. Comms.* **57**, 12305-12308 (2021). https://doi.org/10.1039/D1CC03202J

216 Steinbrecher, T. B. *et al.* Accurate Binding Free Energy Predictions in Fragment Optimization. *J. Chem. Inf. Model.* **55**, 2411-2420 (2015). https://doi.org/10.1021/acs.jcim.5b00538

217 Mihalovits, L. M., Ferenczy, G. G. & Keserű, G. M. The role of quantum chemistry in covalent inhibitor design. *Int. J. Quant. Chem.* **122**, e26768 (2022). https://doi.org/https://doi.org/10.1002/qua.26768

218 Hayek-Orduz, Y. *et al.* Novel covalent and non-covalent complex-based pharmacophore models of SARS-CoV-2 main protease (Mpro) elucidated by microsecond MD simulations. *Sci. Reps.* **12**, 14030 (2022). https://doi.org/10.1038/s41598-022-17204-0

219 Pan, A. C., Xu, H., Palpant, T. & Shaw, D. E. Quantitative Characterization of the Binding and Unbinding of Millimolar Drug Fragments with Molecular Dynamics Simulations. *J. Chem. Theory. Comput.* **13**, 3372-3377 (2017). https://doi.org/10.1021/acs.jctc.7b00172

220 Khalak, Y., Tresadern, G., Hahn, D. F., de Groot, B. L. & Gapsys, V. Chemical space exploration with active learning and alchemical free energies. *J. Chem. Theory. Comput.* **18**, 6259-6270 (2022). https://doi.org/10.1021/acs.jctc.2c00752

221 Thompson, J. *et al.* Optimizing active learning for free energy calculations. *Art. Intellig. Life Sci.* **2**, 100050 (2022). https://doi.org/10.1016/j.ailsci.2022.100050

222 Sadybekov, A. A. *et al.* Synthon-based ligand discovery in virtual libraries of over 11 billion compounds. *Nature* **601**, 452-459 (2022). https://doi.org/10.1038/s41586-021-04220-9

223 Gentile, F. *et al.* Artificial intelligence–enabled virtual screening of ultra-large chemical libraries with deep docking. *Nature Protocols* **17**, 672-697 (2022). https://doi.org/10.1038/s41596-021-00659-2

224 Correy, G. J. *et al.* Exploration of structure-activity relationships for the SARS-CoV-2 macrodomain from shape-based fragment linking and active learning. *Sci. Adv.* **11**, eads7187 (2025). https://doi.org/10.1126/sciadv.ads7187

225 Guo, J. *et al.* Link-INVENT: generative linker design with reinforcement learning. *Dig. Discov.* **2**, 392-408 (2023). https://doi.org/10.1039/D2DD00115B

226 Bieniek, M. K. *et al.* An open-source molecular builder and free energy preparation workflow. *Comms. Chem.* **5**, 136 (2022). https://doi.org/10.1038/s42004-022-00754-9

227 Cree, B., Bieniek, M., Amin, S., Kawamura, A. & Cole, D. Active learning driven prioritisation of compounds from on-demand libraries targeting the SARS-CoV-2 main protease. *ChemRxiv* (2024). https://doi.org/ doi:10.26434/chemrxiv-2024-xczfb

228 Llompart, P. *et al.* Harnessing Medicinal Chemical Intuition from Collective Intelligence. *J. Med. Chem.* **68**, 10860-10876 (2025). https://doi.org/10.1021/acs.jmedchem.4c03066

229 Chen, Z., Ayinde, O. R., Fuchs, J. R., Sun, H. & Ning, X. G2Retro as a two-step graph generative models for retrosynthesis prediction. *Commun. Chem.* **6**, 102 (2023). https://doi.org/10.1038/s42004-023-00897-3

230 Struble, T. J. *et al.* Current and future roles of artificial intelligence in medicinal chemistry synthesis. *J. Med. Chem.* **63**, 8667-8682 (2020). https://doi.org/10.1021/acs.jmedchem.9b02120

231 Zhong, Z. *et al.* Recent advances in deep learning for retrosynthesis. *WIREs Comp. Mol. Sci.* **14**, e1694 (2024). https://doi.org/https://doi.org/10.1002/wcms.1694

232 Perron, Q. *et al.* Deep generative models for ligand-based de novo design applied to multi-parametric optimization. *Journal of Computational Chemistry* **43**, 692-703 (2022). https://doi.org/https://doi.org/10.1002/jcc.26826

233 Liu, J. & Hein, J. E. Automation, analytics and artificial intelligence for chemical synthesis. *Nat. Synth.* **2**, 464-466 (2023). https://doi.org/10.1038/s44160-023-00335-1

234 Nippa, D. F. *et al.* Enabling late-stage drug diversification by high-throughput experimentation with geometric deep learning. *Nat. Chem.* (2023). https://doi.org/10.1038/s41557-023-01360-5

235 Matysiak, B. M., Thomas, D. & Cronin, L. Reaction kinetics using a chemputable framework for data collection and analysis. *Angew. Chem. Int. Ed.* **n/a**, e202315207 (2023). https://doi.org/https://doi.org/10.1002/anie.202315207

236 Townley, C. *et al.* Enabling synthesis in fragment-based drug discovery (FBDD): microscale high-throughput optimisation of the medicinal chemist's toolbox reactions. *RSC Med. Chem.* **14**, 2699-2713 (2023). https://doi.org/10.1039/D3MD00495C

237 Abdiaj, I. *et al.* End-to-end automated synthesis of C(sp^3)-enriched drug-like molecules *via* negishi coupling and novel, automated liquid–liquid extraction. *J. Med. Chem.* **66**, 716-732 (2023). https://doi.org/10.1021/acs.jmedchem.2c01646

238 Grosjean, H. *et al.* Binding-Site Purification of Actives (B-SPA) Enables Efficient Large-Scale Progression of Fragment Hits by Combining Multi-Step Array Synthesis With HT Crystallography. *Ang. Chem. Int. Ed.* **n/a**, e202424373 (2025). https://doi.org/https://doi.org/10.1002/anie.202424373

239 Gao, L. *et al.* 'Chemistry at the speed of sound': automated 1536-well nanoscale synthesis of 16 scaffolds in parallel. *Green Chem.* **25**, 1380-1394 (2023). https://doi.org/10.1039/D2GC04312B

240 Le Vaillant, F. *et al.* Thinking outside the library: cluster synthesis of diverse molecules on a single robotic platform. *Chem. Sci.*, Advance article (2026). https://doi.org/10.1039/D5SC07668D

241 Pijper, B. *et al.* Multistep and multivectorial assembly line library synthesis in flow. *Chem Catalysis* **4** (2024). https://doi.org/10.1016/j.checat.2024.101118

242 Chessari, G. *et al.* C–H functionalisation tolerant to polar groups could transform fragment-based drug discovery (FBDD). *Chem. Sci.* **12**, 11976-11985 (2021). https://doi.org/10.1039/D1SC03563K

243 Chessari, G. *et al.* Fragment-Based Drug Discovery Targeting Inhibitor of Apoptosis Proteins: Discovery of a Non-Alanine Lead Series with Dual Activity Against cIAP1 and XIAP. *J. Med. Chem.* **58**, 6574-6588 (2015). https://doi.org/10.1021/acs.jmedchem.5b00706

244 Crawford, N. *et al.* Clinical Positioning of the IAP Antagonist Tolinapant (ASTX660) in Colorectal Cancer. *Molecular Cancer Therapeutics* **20**, 1627-1639 (2021). https://doi.org/10.1158/1535-7163.MCT-20-1050

245 Song, G. *et al.* General method for carbon–heteroatom cross-coupling reactions via semiheterogeneous red-light metallaphotocatalysis. *Nat. Comms.* **16**, 7045 (2025). https://doi.org/10.1038/s41467-025-61812-z

246 Li, X. & Waser, J. Forging 1,1′ -Bicyclopropenyls by Synergistic Au/Ag Dual-Catalyzed Cyclopropenyl Cross-Coupling. *J. Am. Chem. Soc.* **146**, 29712-29719 (2024). https://doi.org/10.1021/jacs.4c10996

247 Zhang, J. & Dong, J. Modular Click Chemistry Library: Searching for Better Functions. *Chin. J. Chem* **39**, 1025—1027 (2021).

248 Kirsch, P. *et al.* Fragment-Based Discovery of a Qualified Hit Targeting the Latency-Associated Nuclear Antigen of the Oncogenic Kaposi's Sarcoma-Associated Herpesvirus/Human Herpesvirus 8. *Journal of Medicinal Chemistry* **62**, 3924-3939 (2019). https://doi.org/https://doi.org/10.1021/acs.jmedchem.8b01827

249 Reddi, R. N. *et al.* Sulfamate Acetamides as Self-Immolative Electrophiles for Covalent Ligand-Directed Release Chemistry. *J. Am. Chem. Soc.* **145**, 3346-3360 (2023). https://doi.org/10.1021/jacs.2c08853

250 Kathman, S. G. & Statsyuk, A. V. Covalent tethering of fragments for covalent probe discovery. *MedChemComm* **7**, 576-585 (2016). https://doi.org/10.1039/C5MD00518C

251 Abdeldayem, A., Raouf, Y. S., Constantinescu, S. N., Moriggl, R. & Gunning, P. T. Advances in covalent kinase inhibitors. *Chem Soc. Rev.* **49**, 2617-2687 (2020). https://doi.org/10.1039/C9CS00720B

252 Grainger, R., Heightman, T. D., Ley, Steven V., Lima, F. & Johnson, C. N. Enabling synthesis in fragment-based drug discovery by reactivity mapping: photoredox-mediated cross-dehydrogenative heteroarylation of cyclic amines. *Chem. Sci.* **10**, 2264-2271 (2019). https://doi.org/10.1039/C8SC04789H

253 Hirano, A. *et al.* Technical capabilities and limitations of optical spectroscopy and calorimetry using water-miscible solvents: The case of dimethyl sulfoxide, acetonitrile, and 1,4-dioxane. *J. Pharm. Sci.* **109**, 524-531 (2020). https://doi.org/10.1016/j.xphs.2019.10.056

254 Navratilova, I. *et al.* Discovery of new bromodomain scaffolds by biosensor fragment screening. *ACS Med. Chem. Letts.* **7**, 1213-1218 (2016). https://doi.org/10.1021/acsmedchemlett.6b00154

255 Wernersson, S., Birgersson, S. & Akke, M. Cosolvent dimethyl sulfoxide influences protein–ligand binding kinetics via solvent viscosity effects: Revealing the success rate of complex formation following diffusive protein–ligand encounter. *Biochemistry* **62**, 44-52 (2023). https://doi.org/10.1021/acs.biochem.2c00507

256 Murray, J. B., Roughley, S. D., Matassova, N. & Brough, P. A. Off-rate screening (ORS) by surface plasmon resonance. An efficient method to kinetically sample hit to lead chemical space from unpurified reaction products. *J. Med. Chem.* **57**, 2845-2850 (2014). https://doi.org/10.1021/jm401848a

257 Bentley, M. R. *et al.* Rapid elaboration of fragments into leads by X-ray crystallographic screening of parallel chemical libraries (REFiL $_X$). *J. Med. Chem.* **63**, 6863-6875 (2020). https://doi.org/10.1021/acs.jmedchem.0c00111

258 Fearon, D. *et al.* Accelerating Drug Discovery With High-Throughput Crystallographic Fragment Screening and Structural Enablement. *Applied Research* **4**, e202400192 (2025). https://doi.org/https://doi.org/10.1002/appl.202400192

259 Sutanto, F. *et al.* Combining High-Throughput Synthesis and High-Throughput Protein Crystallography for Accelerated Hit Identification. *Angewandte Chemie International Edition* **60**, 18231-18239 (2021). https://doi.org/https://doi.org/10.1002/anie.202105584

260 Muñoz-Reyes, D. *et al.* Leveraging crystallographic fragment screening, algorithmic merging and digital chemistry to target NCS-1 protein-protein interactions for treatment of neurological disorders. *bioRxiv* **https://doi.org/10.1101/2025.06.18.660085** (2025). https://doi.org/https://doi.org/10.1101/2025.06.18.660085

261 Grosjean, H. *et al.* Structure-Activity-Relationships can be directly extracted from high-throughput crystallographic evaluation of fragment elaborations in crude reaction mixtures. *ChemRxiv* **10.26434/chemrxiv-2025-bg2ll** (2025). https://doi.org/10.26434/chemrxiv-2025-bg2ll

262 Baker, L. M. *et al.* Rapid optimisation of fragments and hits to lead compounds from screening of crude reaction mixtures. *Commun. Chem.* **3**, 122 (2020). https://doi.org/10.1038/s42004-020-00367-0

263 Alboreggia, G., Udompholkul, P., Baggio, C. & Pellecchia, M. Mixture-Based Screening of Focused Combinatorial Libraries by NMR: Application to the Antiapoptotic Protein hMcl-1. *J. Med. Chem.* **66**, 10108-10118 (2023). https://doi.org/https://doi.org/10.1021/acs.jmedchem.3c01073

264 Cramer, J. *et al.* A False-Positive Screening Hit in Fragment-Based Lead Discovery: Watch out for the Red Herring. *Ang. Chem. Int. Ed.* **56**, 1908-1913 (2017). https://doi.org/https://doi.org/10.1002/anie.201609824

265 Garman, E. Perspective on a large-scale ligand structure characterization. *Acta Cryst. Sec. D* **81**, 394-395 (2025).

266 Ehler, A., Bartelmus, C., Benz, J., Plitzko, I. & Rudolph, M. G. A high-resolution data set of fatty acid-binding protein structures. III. Unexpectedly high occurrence of wrong ligands. *Acta Cryst. Sect. D* **81**, 451-464 (2025).

267 Casagrande, F. *et al.* A high-resolution data set of fatty acid-binding protein structures. I. Dynamics of FABP4 and ligand binding. *Acta Cryst. Sect. D* **81**, 423-435 (2025).

268 Jaskolski, M., Wlodawer, A., Dauter, Z., Minor, W. & Rupp, B. Group depositions to the Protein Data Bank need adequate presentation and different archiving protocol. *Prot. Sci.* **31**, 784-786 (2022). https://doi.org/https://doi.org/10.1002/pro.4271

269 Weiss, M. S. *et al.* Of problems and opportunities—How to treat and how to not treat crystallographic fragment screening data. *Prot. Sci.* **31**, e4391 (2022). https://doi.org/https://doi.org/10.1002/pro.4391

270 Fraser, J. (2025).

271 Fraser, J. S. & Murcko, M. A. Structure is beauty, but not always truth. *Cell* **187**, 517-520 (2024). https://doi.org/https://doi.org/10.1016/j.cell.2024.01.003

272 Erlanson, D. A. *et al.* Where and how to house big data on small fragments. *Nat. Comms.* **16**, 4179 (2025). https://doi.org/10.1038/s41467-025-59233-z

273 Kartal, Ö., Andres, F., Lai, M. P., Nehme, R. & Cottier, K. waveRAPID—a robust assay for high-throughput kinetic screens with the creoptix WAVEsystem. *SLAS Disc.* **26**, 995-1003 (2021). https://doi.org/10.1177/24725552211013827

274 Liu, L. Efficient hit and lead compound evaluation strategy based on off-rate screening by surface plasmon resonance. *J. Med. Chem.* **57**, 2843-2844 (2014). https://doi.org/10.1021/jm5003815

275 Adams, L. A. *et al.* Rapid elaboration of fragments into leads applied to bromodomain-3 extra-terminal domain. *J. Med. Chem.* **66**, 5859-5872 (2023). https://doi.org/10.1021/acs.jmedchem.3c00137

276 Ma, H. *et al.* PAC-FragmentDEL – photoactivated covalent capture of DNA-encoded fragments for hit discovery. *RSC Med. Chem.* **13**, 1341-1349 (2022). https://doi.org/10.1039/D2MD00197G

277 Ng, E. S. M., Chan, N. W. C., Lewis, D. F., Hindsgaul, O. & Schriemer, D. C. Frontal affinity chromatography—mass spectrometry. *Nature Protocols* **2**, 1907-1917 (2007). https://doi.org/10.1038/nprot.2007.262

278 Tyagarajan, S. *et al.* Rapid Affinity and Microsomal Stability Ranking of Crude Mixture Libraries of Histone Deacetylase Inhibitors. *ACS Med. Chem. Letts* **15**, 1787-1794 (2024). https://doi.org/10.1021/acsmedchemlett.4c00345

279 Wu, B. *et al.* HTS by NMR of Combinatorial Libraries: A Fragment-Based Approach to Ligand Discovery. *Chem. Biol.* **20**, 19-33 (2013). https://doi.org/https://doi.org/10.1016/j.chembiol.2012.10.015

280 Larda, S. T. *et al.* Robust Strategy for Hit-to-Lead Discovery: NMR for SAR. *J. Med. Chem.* **66**, 13416-13427 (2023). https://doi.org/10.1021/acs.jmedchem.3c00656

281 Alboreggia, G., Udompholkul, P., Baggio, C. & Pellecchia, M. Mixture-Based Screening of Focused Combinatorial Libraries by NMR: Application to the Antiapoptotic Protein hMcl-1. *J. Med. Chem.* **66**, 10108-10118 (2023). https://doi.org/10.1021/acs.jmedchem.3c01073

282 Canal-Martín, A. & Pérez-Fernández, R. Protein-Directed Dynamic Combinatorial Chemistry: An Efficient Strategy in Drug Design. *ACS Omega* **5**, 26307-26315 (2020). https://doi.org/10.1021/acsomega.0c03800

283 Wu, Y., Liu, C. & Hu, L. Fragment-Based Dynamic Combinatorial Chemistry for Identification of Selective α-Glucosidase Inhibitors. *ACS Med. Chem. Letts.* **13**, 1791-1796 (2022). https://doi.org/10.1021/acsmedchemlett.2c00405

284 Wilders, H. *et al.* Expedited SARS-CoV-2 Main Protease Inhibitor Discovery through Modular 'Direct-to-Biology' Screening. *Ang. Chem. Int. Ed.* **64**, e202418314 (2025). https://doi.org/https://doi.org/10.1002/anie.202418314

285 Eakin Ann, E. *et al.* Pyrrolamide DNA Gyrase Inhibitors: Fragment-Based Nuclear Magnetic Resonance Screening To Identify Antibacterial Agents. *Antimicrobial Agents and Chemotherapy* **56**, 1240-1246 (2012). https://doi.org/10.1128/aac.05485-11

286 Basarab, G. S. *et al.* Optimization of Pyrrolamide Topoisomerase II Inhibitors Toward Identification of an Antibacterial Clinical Candidate (AZD5099). *Journal of Medicinal Chemistry* **57**, 6060-6082 (2014). https://doi.org/10.1021/jm500462x

287 Wang, X., Li, L., Shen, X. & Lu, X. Rational Design Strategies in DNA-Encoded Libraries for Drug Discovery. *Angewandte Chemie International Edition* **64**, e202511839 (2025). https://doi.org/https://doi.org/10.1002/anie.202511839

288 Hewings, D. S. *et al.* 3,5-Dimethylisoxazoles Act As Acetyl-lysine-mimetic Bromodomain Ligands. *J. Med. Chem.* **54**, 6761-6770 (2011). https://doi.org/10.1021/jm200640v

289 Salvini, C. L. A. *et al.* Fragment expansion with NUDELs – poised DNA-encoded libraries. *Chem. Sci.* **14**, 8288-8294 (2023). https://doi.org/10.1039/D3SC01171B

290 Lim, K. S. *et al.* Machine Learning on DNA-Encoded Library Count Data Using an Uncertainty-Aware Probabilistic Loss Function. *J. Chem. Inf. Model.* **62**, 2316-2331 (2022). https://doi.org/10.1021/acs.jcim.2c00041

291 Zhang, C. *et al.* Building block-based binding predictions for DNA-encoded libraries. *J. Chem. Inf. Model.* **63**, 5120-5132 (2023). https://doi.org/10.1021/acs.jcim.3c00588

292 Shmilovich, K., Chen, B., Karaletsos, T. & Sultan, M. M. DEL-Dock: Molecular Docking-Enabled Modeling of DNA-Encoded Libraries. *J. Chem. Inf. Model.* **63**, 2719-2727 (2023). https://doi.org/10.1021/acs.jcim.2c01608

293 Hughes, J., Rees, S., Kalindjian, S. & Philpott, K. Principles of early drug discovery. *Br. J. Pharm.* **162**, 1239-1249 (2011). https://doi.org/10.1111/j.1476-5381.2010.01127.x

294 Choung, O.-H., Vianello, R., Segler, M., Stiefl, N. & Jiménez-Luna, J. Extracting medicinal chemistry intuition via preference machine learning. *Nat. Comms.* **14**, 6651 (2023). https://doi.org/10.1038/s41467-023-42242-1

295 Vidler, L. R. & Baumgartner, M. P. Creating a Virtual Assistant for Medicinal Chemistry. *ACS Med. Chem. Letts* **10**, 1051-1055 (2019). https://doi.org/10.1021/acsmedchemlett.9b00151

296 Buttenschoen, M., Morris, G. M. & Deane, C. M. PoseBusters: AI-based docking methods fail to generate physically valid poses or generalise to novel sequences. *Chem. Sci.*, 10.1039.D1033SC04185A (2024). https://doi.org/10.1039/D3SC04185A

297 Kearnes, S. M. *et al.* The Open Reaction Database. *J. Am. Chem. Soc.* **143**, 18820-18826 (2021). https://doi.org/10.1021/jacs.1c09820

298 Kramer, C. *et al.* The Need for Continuing Blinded Pose- and Activity Prediction Benchmarks. *J. Chem. Inf. Model.* **65**, 2180-2190 (2025). https://doi.org/10.1021/acs.jcim.4c02296

Editorial Summary:
This review maps key bottlenecks in turning fragment hits into lead compounds and highlights emerging solutions: integrated Design-Make-Test cycles, high-throughput assays, automation and modelling to accelerate reliable hit-to-lead progression.

Peer Review Information: *Nature Communications* thanks Charles Wartchow, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.