Chemistry-driven Hit-to-lead Optimization Guided by Structure-based Approaches

Laurent Hoffer, Christophe Muller, Philippe Roche, Xavier Morelli

Molecular Informatics

Chemistry-Driven Hit-To-Lead Optimization Guided by Structure-Based Approaches

--Manuscript Draft--

Manuscript Number:	minf.201800059
Article Type:	Review
Corresponding Author:	Philippe Roche CRCM Marseille, FRANCE
Corresponding Author E-Mail:	philippe.roche@inserm.fr
Other Authors:	Laurent Hoffer, PhD
	Christophe Muller, PhD
	Xavier Morelli, PhD
Keywords:	hit-to-lead optimization; structure-based drug design; de novo design; library design; virtual screening
Manuscript Classifications:	Biological and Medicinal Chemistry; Drug design; Structure-activity relationships; Virtual screening
Suggested Reviewers:	Gisbert Schneider gisbert.schneider@pharma.ethz.ch
	Matthias Rarey rarey@zbh.uni-hamburg.de
Opposed Reviewers:	
Abstract:	For several decades, hit identification for drug discovery has been facilitated by developments in both fragment-based and high-throughput screening technologies. However, a major bottleneck in drug discovery projects continues to be the optimization of primary hits from screening campaigns to derive lead compounds. Computational chemistry or molecular modeling can play an important role during this hit-to-lead (H2L) stage by both suggesting putative optimizations and decreasing the number of compounds to be synthesized and evaluated. However, it is also crucial to consider the feasibility of organically synthesizing these virtually designed compounds. Furthermore, the generated molecules should have reasonable physicochemical properties and be medicinally relevant. This review focuses on chemistry-driven and structure-based computational methods that can be used to tackle the difficult problem of H2L optimization, with emphasis being placed on the strategy developed in our laboratory.
Author Comments:	Special Issue: 6th Strasbourg Summer School in Chemoinformatics Dear Editors, We are pleased to submit our contribution entitled "Chemistry-Driven Hit-To-Lead Optimization Guided by Structure-Based Approaches" for consideration for publication as a review article in Molecular Informatics as part of the special issue devoted to the 2018 Chemoinformatics Summer School in Strasbourg. The identification of hits for a given target has been greatly facilitated in the recent years, due to the progress in high-throughput screening mainly through automation and miniaturization. In addition, fragment-based approaches have proven efficient to rapidly identify small molecules that can bind a particular target. One of the main bottlenecks now resides in the efficient optimization of these numerous hits identified in early phase of drug discovery. This review focuses on in silico approaches dedicated to hit to lead optimization with a special emphasis on methods guided by the knowledge of the target 3D structure. After introducing the main challenges, we present in silico strategies and programs recently developed to tackle this critical issue in two chapters. Finally, the in silico steps of the

	DOTS strategy developed in our laboratory is described. We hope that you will find this review suitable for the special issue of Molecular Informatics. This review article has not been published and is not under consideration for publication elsewhere and authors have no conflicts of interest to disclose. Thank you for your consideration. We look forward to hearing from you at your earliest convenience. Sincerely yours, Drs Philippe Roche, Xavier Morelli Principal Investigators, CRCM
Section/Category:	Strasbourg2018
Additional Information:	
Question	**Response**
Dedication	
Submitted solely to this journal?	Yes
Has there been a previous version?	No

Review

Chemistry-Driven Hit-To-Lead Optimization Guided by Structure-Based Approaches

Laurent Hoffer[a], Christophe Muller[b], Philippe Roche[a,1] and Xavier Morelli[a,b,1]

[a] CNRS, INSERM, Institut Paoli-Calmettes, Aix-Marseille Univ, CRCM, Marseille, France

[b] Institut Paoli-Calmettes, IPC Drug Discovery, Marseille, France

[1]To whom correspondence should be addressed.

E-mail: philippe.roche@inserm.fr, xavier.morelli@inserm.fr

Abstract

For several decades, hit identification for drug discovery has been facilitated by developments in both fragment-based and high-throughput screening technologies. However, a major bottleneck in drug discovery projects continues to be the optimization of primary hits from screening campaigns to derive lead compounds. Computational chemistry or molecular modeling can play an important role during this hit-to-lead (H2L) stage by both suggesting putative optimizations and decreasing the number of compounds to be synthesized and evaluated. However, it is also crucial to consider the feasibility of organically synthesizing these virtually designed compounds. Furthermore, the generated molecules should have reasonable physicochemical properties and be medicinally relevant. This review focuses on chemistry-driven and structure-based computational methods that can be used to tackle the difficult problem of H2L optimization, with emphasis being placed on the strategy developed in our laboratory.

Keywords: hit-to-lead optimization; structure-based drug design; *de novo* design; library design; virtual screening

1. Introduction

The first step in a drug discovery project involves the identification of hit compounds that typically exhibit weak to moderate affinity for the biological target.[1] These primary hits are usually discovered either via the high-throughput screening (HTS) of large collections of diverse molecules that are of medium complexity or the use of a fragment-based technology in which small chemical libraries of low-molecular-weight fragments are considered.[2] There is a current trend toward fragment-based drug discovery (FBDD), since the optimization phase of hit compounds from HTS can be fastidious while also maintaining a reasonable molecular weight.[3] Moreover, FBDD approaches can cover a larger chemical space in the optimization phase, which leads to structural novelty and a higher probability of success. An increasing number of success stories have been reported in the past decade, and more than 30 compounds derived from fragments are currently in clinics or on the market.[4] Following orthogonal validation, the confirmed fragments/hits are improved in a subsequent stage called hit-to-lead (H2L).

Typical H2L involves chemical modifications around the validated hit to optimize its affinity for the target to become a lead compound.[5] These optimization phases can be completed by exploiting a trial-and-error strategy, and several cycles are usually needed to reach a suitable affinity. Moreover, other features may be preferred, such as improving the physicochemical properties of the compounds and enabling or maintaining a degree of selectivity with respect to undesired related targets. In practice, successful H2L optimization can improve the binding constants by several orders of magnitude. However, because of the numerous parameters that must be considered, the H2L process can also be unsuccessful, time-consuming, and highly expensive.[6]

Various computer-based approaches have been developed to overcome bottlenecks during H2L.[7] For instance, ligand-based methods, such as the Quantitative Structure–Activity Relationship (QSAR), can be used to optimize a series of compounds by exploiting available experimental data and calculated descriptors.[8] However, this review will focus on computational structure-based approaches [9], such as molecular docking, *de novo* design (DnD) and pharmacophore screening. These methods rely on structural data to generate models, and success stories have recently been reported.[10] In practice, X-ray crystallography is the primary approach to determine the binding mode of a compound in its binding site at the atomic level.[11] Such structural data greatly facilitates the H2L optimization by clearly identifying nearby protein sub-pockets around the engaged fragment that could be used in the optimization phase.

The most intuitive approach for optimizing a fragment using an *in silico* structure-based approach is the transposition of the growing paradigm from FBDD to the molecular modeling context: The affinity of a compound for its target is increased by adding chemical moieties that are able to create new favorable contacts while maintaining its original binding mode (Figure 1A). Two additional H2L strategies used in FBDD, namely, linking and merging, can also be virtually mimicked (Figures 1B & 1C).[12] In contrast to the growing concept where a single fragment is required, merging and linking consist in the covalent assembly of two non-overlapping fragments, either directly or *via* a spacer of variable length.[13] Merging and linking strategies are less frequently used than growing and are also more challenging because they require two hits and the conservation of their original orientation after their fusion into a single compound. However, when these strict criteria are validated, merging and linking can lead to outstanding improvements in affinity.[14] In practice, these two approaches are primarily

applicable to fragments because merging or linking two compounds with moderate to high molecular weight would result in molecules that are too large to be used as probes or drugs.

Computational-based tools that are able to address growing, linking or merging strategies are of primary interest during the H2L stage.[12, 13, 15] For instance, a virtual-focused library that is around the validated fragment can be generated within the first step. Subsequently, this library can be virtually screened using a constrained docking strategy to mimic the growing paradigm and maintain the original interactions. The best putative optimizations are finally selected in the top hit list using a scoring function. Alternatively, classical “unconstrained docking” of the focused library can be performed. Pharmacophore filters[16] or interaction fingerprints[17] are subsequently used to extract putative optimizations that maintain the original binding mode. Finally, *de novo* design algorithms [18, 19], with the binding mode of the validated fragment serving as the starting seed, can also be used to sample the cavity and generate putative optimized compounds.

A critical point is determining how to handle the creation of new covalent bonds made during the virtual H2L optimization step. A first attempt to generate reasonable structures has been made by the RECAP[20] or BRICS[21] methods. For instance, molecules in RECAP are fragmented around specific bond types, and new terminal atoms are flagged to capture their previous chemical environment. Virtual fragments with complementary flags can subsequently be merged to design new virtual compounds. However, there is no assessment regarding their synthesis, since neither organic chemistry rules nor the availability of building blocks (BBs) were used during the process. This approach leads to either time-consuming follow-up efforts to devise synthesis routes or even the inability to produce the virtually generated compounds. Consequently, it is of critical importance to address the synthesis tractability of virtually

generated compounds in the context of prospective drug discovery projects.[18, 22] In this effort, generic reaction schemes “reactants → products” can be encoded to mimic common organic synthesis routes.[23, 24] These chemistry rules can subsequently be retained during the design of new molecules to facilitate the synthesis stage.

Several remarkable computational structure-based H2L methods were described in the past decade, but they lacked this fundamental requirement of ensuring the synthesizability of the generated compounds for use in drug discovery projects.[25, 26] *De novo* design or alternative structure-based H2L methods, which approached this critical issue, are presented in this review, with an emphasis on the *in silico* steps of the DOTS strategy developed in our laboratory.[27]

2. Virtual H2L methods relying on the “*de novo* design” concept

By definition, *de novo* design tools can automatically build compounds “from scratch” within a binding site of known 3D structure using predefined sets of substructures and rules governing their linkage.[18, 19] In theory, DnD methods enable the exploration of a considerably larger chemical space, which is in contrast to classic virtual screening (VS), where only commercially available compounds are tested. Although conceptually similar, DnD tools can be distinguished by their algorithms, ranking/scoring functions, convergence criteria, and branch-pruning strategies. This methodological approach, developed in the early 1990s, was initially appealing but is currently not widely practiced by chemoinformaticians.[28, 29] Indeed, DnD methods often exhibit several critical drawbacks that limit their application to prospective cases. The primary drawbacks include the following: 1) low reliability in predicting affinity, 2) potentially poor physicochemical properties of the designed compounds and, most importantly, 3) issues regarding the compounds’ synthetic tractability.[18]

As a response to the first criticism, *in silico* H2L methods that rely on DnD approaches, as introduced below, were upgraded to use a pre-positioned fragment in the binding site as the original seed. Starting from an experimentally validated fragment with a known binding mode is likely to increase the reliability of their predictions.

The third criticism of DnD approaches - the major drawback - concerns the difficulty in synthesizing certain of the suggested compounds because the rules governing bond creation during the process did not take into account any organic chemistry knowledge. This primary issue was approached by incorporating virtual reaction schemes in the workflow to ensure the synthetic accessibility with more confidence. Both ligand-based and structure-based DnD tools were developed to address this problem. **DOGS**[24] and **SYNOPSIS**[30] are popular examples of such chemistry-driven ligand-based DnD methods. However, other structure-based DnD approaches that rely on both substructure seed prepositioned in the binding site and encoded chemistry knowledge exist and are discussed below.

The **SPROUT** program was one of the first published DnD software in the 1990s, but several major upgrades have been recently added.[29] This program is now able to take a validated hit or fragment as seed and to incorporate chemistry-based rules during the virtual design process. SPROUT was successfully used to design inhibitors of dihydroorotate dehydrogenase enzyme[31], and more recently BACE-1 inhibitors, based on a synthetic Suzuki reaction scheme.[32]

LigBuilder[33] is another validated tool that was recently updated to tackle the usual drawbacks from DnD software. The current version (2.0), is able to perform both *ab initio* design and lead optimization of a compound of interest while considering both synthesis accessibility and filter-

based drug-likeness of the designed compounds.[34] LigBuilder was successfully applied to generate nanomolar inhibitors of Cyclophilin A.[35]

Similarly, the **AutoGrow** tool[36] was also updated to include organic chemistry rules within the operators of its evolutionary-based algorithm.[37] In this instance, the popular Autodock Vina docking engine is used for both conformational sampling and scoring stages.[38] Physicochemical filters were also added to reject non drug-like compounds.

Beccari *et al* developed a suite of programs called **LiGen** for DnD that both handles the chemical rules for designing accessible compounds and efficient docking with pharmacophore constraints.[39] Each module can be used either separately or combined in a complex workflow procedure. This toolbox is reported to tackle most issues in the DnD field while being efficient from a computational point of view. However, there is no case study to highlight its ability in either retrospective or prospective cases.

Cheron *et al* reported the development of **OpenGrowth** for the computer-based H2L optimization of compounds under binding site constraints.[40] In this study, the probability that a given fragment is connected to another one in a reference drug dataset is used to guide the design process. At the end of the process, designed molecules exhibit both reasonable synthetic accessibility and good physicochemical properties, although organic synthesis routes are not explicitly taken into account. A new version, including the implementation of an explicit synthetic accessibility score during the design process, is planned.

It should be noted that many DnD tools were only validated in retrospective case studies. However, all these programs exhibit promise and should be useful for prospective projects aiming to design active and accessible compounds.

3. Alternative approaches for *in silico* H2L optimization

Chevillard *et al* optimized several fragment-like compounds into low micromolar ligands for the β2-adrenergic receptor target.[41] The **PINGUI** method relies on both structural data and a series of 58 encoded organic chemistry rules published by Hartenfeller *et al* for designing putative accessible ligands by merging an original fragment with compatible BBs.[23] In practice, the SEED program is used to dock pre-processed BBs (called surrogates) where the reactive center is modified according to the considered synthesis reaction.[42] For instance, a methylamine group will replace the aldehyde function from the BB if the reductive amination synthesis scheme is selected. In a second step, putatively interesting surrogates, where the reactive center is sufficiently close to the one from the original hit without any overlap with other atoms, are identified. Next, final products corresponding to the coupling of the hit and selected surrogates are generated and docked using the DOCK software.[43] A last modeling step involving the refinement and rescoring of selected poses using the SZYBKI method is performed.[44]

The **LeadOp+R** method was developed to perform structure-based H2L optimization with synthetic accessibility.[45] This method relies on approximately 200 encoded chemical reactions while allowing multi-steps design. The pipeline starts with a query structure pre-positioned in the binding site and user-defined preferred ligand-receptor interactions. New 1D/2D compounds are created by combining the current structure with BBs by using encoded chemistry rules without knowledge of the binding site at this stage. 3D conformers are subsequently generated for each compound, superimposed to the shared reference substructure in the binding site, and evaluated for their ability to make additional favorable contacts without any clashes. The cycle stops when required interactions between both entities

are fulfilled. Several physicochemical filters are used to discard compounds with undesirable properties. They successfully applied their approach on two retrospective projects: several known potent inhibitors were designed for each target, while suggested synthetic routes shared steps with published ones.

An original strategy, relying on both pharmacophore and docking concepts, allowed Schulz *et al* to discover several covalent inhibitors of the enteroviral 3C protease target.[46] First, a new pharmacophore feature type in the **LigandScout** software was developed to look for predefined reactive functions in the vicinity of a cysteine residue.[47] Next, a 3D-pharmacophore was designed to catch fundamental interactions with the enteroviral 3C protease while looking for covalent binders using this new feature. Then, a library of fragment-like compounds was screened against the 3D-pharmacophore. Several fragment hits were identified and experimentally validated using mass spectrometry. The best one was further investigated but exhibited some instability. To bypass this issue, a scaffold hopping strategy, using simple SMARTS-based substructure search, was successfully employed to find alternative binding cores that still contain the required features. A protocol was later developed for the rational design of optimized analogs of the best new hit, while maintaining the fundamental interactions with the protein. Thus, a virtual library was generated by coupling the hit and commercially available BBs using encoded chemistry reaction rules. LigandScout and GOLD programs were used to identify putative optimizations that would occupy adjacent sub-pockets. Finally, the authors reported both reversible and irreversible inhibitors of the enteroviral 3C protease target.

Evers *et al* from the Sanofi-Aventis company reported the **CROSS** method for either rescaffolding or the optimization of compounds using explicit handling of organic reactions.[48]

The CROSS approach relies on the BROOD software to quickly identify pre-processed fragments that could replace an undesirable core by using “exit vectors” and 3D-shape analysis.[49] The main advantage is that these pre-processed fragments are directly connected to available BBs and specific chemical reactions for easier continued investigation. In addition, the use of chemical protection groups is also allowed to facilitate the results in the synthesis steps. Finally, generated compounds are post-processed using molecular docking and ADME-Tox predictions. While the main usage is rescaffolding, the CROSS method can also be employed in virtual H2L optimization using either a growing or linking strategy. Indeed, linking and rescaffolding are similar concepts, where a linker is selected to connect two moieties while maintaining their original orientation.

AutoCouple was recently reported as a useful tool to expand the chemical space in hit optimization.[50] In the first step, a diversity-oriented library is designed by virtually coupling one hit that includes a reactive function, with a list of commercially available BBs using encoded chemistry rules. In a second step, the rDock program is used to screen the library while adding constraints on the original moiety to maintain the reference binding mode during the conformational sampling stage.[51] The authors applied this strategy to the design of potent CBP bromodomain inhibitors. An acetyl benzene moiety served as the reference substructure able to mimic the acetylated lysine from histone tails, which are recognized by this epigenetic reader. Different organic reactions were considered to build the virtual library that was docked by rDock in the CBP binding site with constraints on the reference moiety. The best poses were minimized using the CHARMM molecular mechanic program[52] during a post-processing stage, before the final selection of target compounds to be synthesized was made. This study led to the discovery of several nanomolar inhibitors for the CBP target. The predicted binding

mode of several compounds was successfully confirmed by solving the structure of the complex using X-ray crystallography.

Finally, the Diversity-Oriented Target-focused Synthesis (**DOTS**) is an integrated strategy developed in our laboratory for generic H2L optimization relying on the growing paradigm.[27] This strategy involves molecular modeling (chemical library design and structure-based VS) and robotic-based experimental stages (diversity-oriented *de novo* synthesis and *in vitro* evaluation and validation). The virtual steps are an upgrade from a former computational optimization method[26] relying on the RECAP algorithm[20], which did not consider the synthetic accessibility of the designed compounds. The general DOTS workflow can be summarized as follows (Figure 2): 1) hit identification and characterization of its binding mode using structural biophysics method, such as X-ray crystallography; 2) design of a virtual focused library around the hit using a database of commercially available BBs, encoded organic chemical rules, and post-processing of the library to extract a diverse subset of representative compounds that also possess reasonable physicochemical properties without any undesirable functions to medicinal chemists; 3) constrained VS of the library with the S4MPLE tool[26, 53] to identify the best putative optimizations that create additional favorable contacts while maintaining the original binding mode; 4) parallel synthesis of the compounds using a chemistry robot; and 5) *in vitro* evaluation with a robotic screening workstation. The *in silico* part of DOTS relies on real chemical knowledge and allows for the production of theoretically accessible and diverse compounds, while exploring the chemical space around the hit, and matching regular physicochemical and medicinal chemistry-like features. The DOTS strategy was successfully applied to the optimization of a previously reported xanthine core that binds the first bromodomain (BD1) of the BRD4 protein.[54] Several sub-micromolar inhibitors were developed and validated in one cycle of optimization, with the best one displaying a K_d value of 190 nM.

X-ray crystallography was used to solve the structure of the best inhibitor in complex with the BRD4(BD1) protein, and confirmed the predicted binding mode. Several features, including linking optimization and design of covalent inhibitors, are currently under development in order to address all available H2L strategies.

4. Summary and Outlook section

This review contributes to an overview of structure-based computational H2L methods that tackle the synthetic accessibility of virtually generated compounds. The different programs discussed in this review, and their main characteristics, are summarized in Table 1. Most of them currently encode real chemistry knowledge to design molecules that can be synthesized with high probability in one or two steps. Such approaches help scientists design accessible and optimized compounds that perfectly fit the binding site of the target.

Despite the shared ability to handle virtual chemistry, these *in silico* structure-based H2L methods rely on various strategies. *De novo* design algorithms start from a given compound as the original seed to perform the optimization process, while alternative approaches use docking or pharmacophore concepts to identify promising compounds from a virtual focused library that was designed around the hit to optimize.

In certain cases, a single reaction scheme is considered, but it is still possible to explore a large chemical space around the starting hit due to the high number of commercially available BBs. Moreover, target compounds selected using these approaches can also be produced in parallel using robotic platforms when a single reaction is considered, as exemplified in the DOTS approach.

Computational H2L can be useful to optimize hits in a time-efficient and cost-effective manner, as highlighted by successful cases described in this review. These *in silico* methods clearly should play a larger role in drug discovery in both academic and pharmaceutical environments.

5. Acknowledgments

We would like to thank Prof. A. Varnek and Dr. D. Rognan for the organization of the 2018 International Summer School in Chemoinformatics.

Figure legends

Figure 1. Selected examples of fragment growing, linking and merging strategies. 2D structure, 3D ligand-protein complex (with PDB code) and binding constants (IC_{50} or AC_{50}) are provided for each fragment and optimized compound. Small organic compounds are displayed in ball-and-stick representation. Fragments are displayed in pink and optimized compounds are shown in cyan. Direct hydrogen bonds between optimized ligands and protein are represented by green dashed lines. **A.** Growing example for the development of Phosphodiesterase (PDE) inhibitors.[55] **B.** Linking example leading to the discovery of a Pyruvate kinase (PKM2) activator.[56] **C.** Merging example towards the development of small molecule inhibitors of Mycobacterium tuberculosis transcriptional repressor protein (EthR).[57]

Figure 2. Schematic workflow of the DOTS strategy. 1) Following hit identification, the binding mode is characterized using structural biophysics methods such as X-ray crystallography. **2)** A virtual focused library is conceived by combining a database of functionalized BBs with an activated form of the original hit using SMARTS-encoded medicinal chemistry-relevant organic synthesis rules. The raw library is then filtered to extract a diverse set of compounds with reasonable physicochemical properties. **3)** The focused library is virtually screened under constraints with S4MPLE to identify compounds that create additional favorable contacts while maintaining the original binding mode. **4)** The selected compounds are synthesized using an automated parallel chemistry robot. **5)** Compounds are finally evaluated *in vitro* with a robotic screening workstation and the best molecules are further validated using orthogonal methods.

6. References

[1] V. Erakovic Haber, R. Spaventi, *Prog Mol Subcell Biol* **2017**, *55*, 91-104; J. P. Hughes, S. Rees, S. B. Kalindjian, K. L. Philpott, *Br J Pharmacol* **2011**, *162*, 1239-1249; B. Munos, *Nature Reviews Drug Discovery* **2009**, 959-968.

[2] J. P. Renaud, M. A. Delsuc, *Current Opinion in Pharmacology* **2009**, *9*, 622-628; R. Macarron, M. N. Banks, D. Bojanic, D. J. Burns, D. A. Cirovic, T. Garyantes, D. V. Green, R. P. Hertzberg, W. P. Janzen, J. W. Paslay, U. Schopfer, G. S. Sittampalam, *Nat Rev Drug Discov* **2011**, *10*, 188-195.

[3] R. Lahana, *Drug Discovery Today* **1999**, 447-448.

[4] C. W. Murray, D. C. Rees, *Nature Chemistry* **2009**, *1*, 187-192; P. J. Hajduk, J. Greer, *Nature Reviews Drug Discovery* **2007**, *6*, 211-219; D. A. Erlanson, S. W. Fesik, R. E. Hubbard, W. Jahnke, H. Jhoti, *Nat Rev Drug Discov* **2016**, *15*, 605-619; M. J. Harner, A. O. Frank, S. W. Fesik, *J Biomol NMR* **2013**, *56*, 65-75; A. J. Souers, J. D. Leverson, E. R. Boghaert, S. L. Ackler, N. D. Catron, J. Chen, B. D. Dayton, H. Ding, S. H. Enschede, W. J. Fairbrother, D. C. Huang, S. G. Hymowitz, S. Jin, S. L. Khaw, P. J. Kovar, L. T. Lam, J. Lee, H. L. Maecker, K. C. Marsh, K. D. Mason, M. J. Mitten, P. M. Nimmer, A. Oleksijew, C. H. Park, C. M. Park, D. C. Phillips, A. W. Roberts, D. Sampath, J. F. Seymour, M. L. Smith, G. M. Sullivan, S. K. Tahir, C. Tse, M. D. Wendt, Y. Xiao, J. C. Xue, H. Zhang, R. A. Humerickhouse, S. H. Rosenberg, S. W. Elmore, *Nat Med.* **2013**, *19*, 202-208. doi: 210.1038/nm.3048. Epub 2013 Jan 1036.

[5] R. Deprez-Poulain, B. Deprez, *Curr Top Med Chem* **2004**, *4*, 569-580; R. Poulain, D. Horvath, B. Bonnet, C. Eckoff, B. Chapelain, M. C. Bodinier, B. Deprez, *J. Med. Chem.* **2001**, *44*, 3391-3401.

[6] A. Alanine, M. Nettekoven, E. Roberts, A. W. Thomas, *Comb Chem High Throughput Screen* **2003**, *6*, 51-66; K. H. Bleicher, H. J. Böhm, K. Müller, A. I. Alanine, *Nat Rev Drug Discov* **2003**, *2*, 369-378.

[7] G. Sliwoski, S. Kothiwale, J. Meiler, E. W. Lowe, *Pharmacol Rev* **2014**, *66*, 334-395; W. L. Jorgensen, *Acc Chem Res* **2009**, *42*, 724-733; W. Yu, A. D. MacKerell, *Methods Mol Biol* **2017**, *1520*, 85-106.

[8] T. Wang, M. B. Wu, J. P. Lin, L. R. Yang, *Expert Opin Drug Discov* **2015**, *10*, 1283-1300.

[9] T. Wang, M. B. Wu, R. H. Zhang, Z. J. Chen, C. Hua, J. P. Lin, L. R. Yang, *Curr Top Med Chem* **2016**, *16*, 901-916; P. Śledź, A. Caflisch, *Curr Opin Struct Biol* **2018**, *48*, 93-102.

[10] B. O. Villoutreix, R. Eudes, M. A. Miteva, *Comb Chem High Throughput Screen* **2009**, *12*, 1000-1016; S. Kalyaanamoorthy, Y. P. Chen, *Drug Discov Today* **2011**, *16*, 831-839.

[11] R. L. M. van Montfort, P. Workman, *Essays Biochem* **2017**, *61*, 431-437.

[12] R. J. Bienstock, *Methods Mol Biol* **2015**, *1289*, 119-135.

[13] L. Hoffer, J. P. Renaud, D. Horvath, *Comb Chem High Throughput Screen* **2011**, *14*, 500-520.

[14] V. Borsi, V. Calderone, M. Fragai, C. Luchinat, N. Sarti, *Journal of Medicinal Chemistry* **2010**, *53*, 4285-4289; J. J. Barker, O. Barker, S. M. Courtney, M. Gardiner, T. Hesterkamp, O. Ichihara, O. Mather, C. A. Montalbetti, A. Müller, M. Varasi, M. Whittaker, C. J. Yarnold, *ChemMedChem* **2010**, *5*, 1697-1700.

[15] R. Law, O. Barker, J. J. Barker, T. Hesterkamp, R. Godemann, O. Andersen, T. Fryatt, S. Courtney, D. Hallett, M. Whittaker, *Journal of Computer-Aided Molecular Design* **2009**, *23*, 459-473; K. Loving, I. Alberts, W. Sherman, *Current Topics in Medicinal Chemistry* **2010**, *10*, 14-32.

[16] T. Langer, R. D. Hoffmann, *Expert Opin Drug Discov* **2006**, *1*, 261-267.

[17] G. Marcou, D. Rognan, *J Chem Inf Model* **2007**, *47*, 195-207.

[18] G. Schneider, U. Fechner, *Nature Reviews Drug Discovery* **2005**, *4*, 649-663.

[19] M. Hartenfeller, G. Schneider, *Methods Mol Biol* **2011**, *672*, 299-323.

[20] X. Q. Lewell, D. B. Judd, S. P. Watson, M. M. Hann, *Journal of Chemical Information and Computer Sciences* **1998**, *38*, 511-522.

[21] J. Degen, C. Wegscheid-Gerlach, A. Zaliani, M. Rarey, *Chemmedchem* **2008**, *3*, 1503-1507.

[22] P. Schneider, G. Schneider, *J Med Chem* **2016**, *59*, 4077-4086; B. B. Masek, D. S. Baker, R. J. Dorfman, K. DuBrucq, V. C. Francis, S. Nagy, B. L. Richey, F. Soltanshahi, *J Chem Inf Model* **2016**, *56*, 605-620; S. Takeda, H. Kaneko, K. Funatsu, *J Chem Inf Model* **2016**, *56*, 1885-1893.

[23] M. Hartenfeller, M. Eberle, P. Meier, C. Nieto-Oberhuber, K. Altmann, G. Schneider, E. Jacoby, S. Renner, *Journal of Chemical Information and Modeling* **2011**, *51*, 3093-3098.

[24] M. Hartenfeller, H. Zettl, M. Walter, M. Rupp, F. Reisen, E. Proschak, S. Weggen, H. Stark, G. Schneider, *PLoS Comput Biol* **2012**, *8*, e1002380.

[25] W. J. Allen, B. C. Fochtman, T. E. Balius, R. C. Rizzo, *J Comput Chem* **2017**, *38*, 2641-2663; G. F. Hao, W. Jiang, Y. N. Ye, F. X. Wu, X. L. Zhu, F. B. Guo, G. F. Yang, *Nucleic Acids Res* **2016**, *44*, W550-556; P. Maass, T. Schulz-Gasch, M. Stahl, M. Rarey, *Journal of Chemical Information and Modeling* **2007**, *47*, 390-399; J. Degen, M. Rarey, *ChemMedChem* **2006**, *1*, 854-868; Y. Li, Z. Zhao, Z. Liu, M. Su, R. Wang, *J Chem Inf Model* **2016**, *56*, 435-453.

[26] L. Hoffer, J. P. Renaud, D. Horvath, *J Chem Inf Model* **2013**, *53*, 836-851.

[27] L. Hoffer, Y. V. Voitovich, B. Raux, K. Carrasco, C. Muller, A. Y. Fedorov, C. Derviaux, A. Amouric, S. Betzi, D. Horvath, A. Varnek, Y. Collette, S. Combes, P. Roche, X. Morelli, *submitted for publication* **2018**.

[28] H. Böhm, *J Comput Aided Mol Des* **1992**, *6*, 61-78; A. Caflisch, A. Miranker, M. Karplus, *J Med Chem* **1993**, *36*, 2142-2167; S. Rotstein, M. Murcko, *Journal of Medicinal Chemistry* **1993**, *36*, 1700-1710; D. Westhead, D. Clark, D. Frenkel, J. Li, C. Murray, B. Robson, B. Waszkowycz, *J Comput Aided Mol Des* **1995**, *9*, 139-148; R. DeWitte, E. Shakhnovich, *Journal of the American Chemical Society* **1996**, 11733-11744; G. Lauri, P. A. Bartlett, *J Comput Aided Mol Des* **1994**, *8*, 51-66.

[29] V. Gillet, A. P. Johnson, P. Mata, S. Sike, P. Williams, *J. Comput. Aided Mol. Des.* **1993**, *7*, 127-153.

[30] H. Vinkers, M. de Jonge, F. Daeyaert, J. Heeres, L. Koymans, J. van Lenthe, P. Lewi, H. Timmerman, K. Van Aken, P. Janssen, *J Med Chem* **2003**, *46*, 2765-2773.

[31] M. Davies, T. Heikkilä, G. A. McConkey, C. W. Fishwick, M. R. Parsons, A. P. Johnson, *J Med Chem* **2009**, *52*, 2683-2693.

[32] N. Y. Mok, J. Chadwick, K. A. Kellett, E. Casas-Arce, N. M. Hooper, A. P. Johnson, C. W. Fishwick, *J Med Chem* **2013**, *56*, 1843-1852.

[33] R. Wang, Y. Gao, L. Lai, *Molecular modeling annual* **2000**, *6*, 498-516.

[34] Y. Yuan, J. Pei, L. Lai, *J Chem Inf Model* **2011**, *51*, 1083-1091.

[35] S. Ni, Y. Yuan, J. Huang, X. Mao, M. Lv, J. Zhu, X. Shen, J. Pei, L. Lai, H. Jiang, J. Li, *J Med Chem* **2009**, *52*, 5295-5298.

[36] J. D. Durrant, R. E. Amaro, J. A. McCammon, *Chemical Biology & Drug Design* **2009**, *73*, 168-178.

[37] J. D. Durrant, S. Lindert, J. A. McCammon, *J Mol Graph Model* **2013**, *44*, 104-112.

[38] O. Trott, A. J. Olson, *J Comput Chem* **2010**, *31*, 455-461.

[39] A. R. Beccari, C. Cavazzoni, C. Beato, G. Costantino, *J Chem Inf Model* **2013**, *53*, 1518-1527.

[40] N. Chéron, N. Jasty, E. I. Shakhnovich, *J Med Chem* **2016**, *59*, 4171-4188.

[41] F. Chevillard, H. Rimmer, C. Betti, E. Pardon, S. Ballet, N. van Hilten, J. Steyaert, W. E. Diederich, P. Kolb, *J Med Chem* **2018**, *61*, 1118-1129.

[42] N. Majeux, M. Scarsi, J. Apostolakis, C. Ehrhardt, A. Caflisch, *Proteins* **1999**, *37*, 88-105.

[43] T. J. Ewing, S. Makino, G. A. Skillman, I. D. Kuntz, *Journal of Computer-Aided Molecular Design* **2001**, *15*, 411-428.

[44] B. A. Ellingson, M. T. Geballe, S. Wlodek, C. I. Bayly, A. G. Skillman, A. Nicholls, *J Comput Aided Mol Des* **2014**, *28*, 289-298.

[45] F. Y. Lin, E. X. Esposito, Y. J. Tseng, *Front Pharmacol* **2018**, *9*, 96.

[46] R. Schulz, A. Atef, D. Becker, F. Gottschalk, C. Tauber, S. Wagner, C. Arkona, A. A. Abdel-Hafez, H. H. Farag, J. Rademann, G. Wolber, *J Med Chem* **2018**, *61*, 1218-1230.

[47] G. Wolber, T. Langer, *J. Chem. Inf. Mod.* **2005**, *45*, 160-169.

[48] A. Evers, G. Hessler, L. H. Wang, S. Werrel, P. Monecke, H. Matter, *J Med Chem* **2013**, *56*, 4656-4670.

[49] L.-h. Wang, A. Evers, P. Monecke, T. Naumann, *Journal of Cheminformatics* **2012**, *4(Suppl 1)*, O20.

[50] L. Batiste, A. Unzue, A. Dolbois, F. Hassler, X. Wang, N. Deerain, J. Zhu, D. Spiliotopoulos, C. Nevado, A. Caflisch, *ACS Cent Sci* **2018**, *4*, 180-188.

[51] S. Ruiz-Carmona, D. Alvarez-Garcia, N. Foloppe, A. B. Garmendia-Doval, S. Juhos, P. Schmidtke, X. Barril, R. E. Hubbard, S. D. Morley, *PLoS Comput Biol* **2014**, *10*, e1003571.

[52] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, M. Karplus, *Journal of Computational Chemistry* **1983**, *4*, 187-217.

[53] L. Hoffer, D. Horvath, *J Chem Inf Model* **2013**, *53*, 88-102; L. Hoffer, C. Chira, G. Marcou, A. Varnek, D. Horvath, *Molecules* **2015**, *20*, 8997-9028.

[54] B. Raux, Y. Voitovich, C. Derviaux, A. Lugari, E. Rebuffet, S. Milhas, S. Priet, T. Roux, E. Trinquet, J. C. Guillemot, S. Knapp, J. M. Brunel, A. Y. Fedorov, Y. Collette, P. Roche, S. Betzi, S. Combes, X. Morelli, *J Med Chem* **2016**, *59*, 1634-1641.

[55] G. L. Card, L. Blasdel, B. P. England, C. Zhang, Y. Suzuki, S. Gillette, D. Fong, P. N. Ibrahim, D. R. Artis, G. Bollag, M. V. Milburn, S. H. Kim, J. Schlessinger, K. Y. Zhang, *Nat Biotechnol* **2005**, *23*, 201-207.

[56] Y. Matsui, I. Yasumatsu, T. Asahi, T. Kitamura, K. Kanai, O. Ubukata, H. Hayasaka, S. Takaishi, H. Hanzawa, S. Katakura, *Bioorg Med Chem* **2017**, *25*, 3540-3546.

[57] P. O. Nikiforov, S. Surade, M. Blaszczyk, V. Delorme, P. Brodin, A. R. Baulard, T. L. Blundell, C. Abell, *Org Biomol Chem* **2016**, *14*, 2318-2326.

Table 1

Program	Search Method[a]	Availability[b]	Main Features[c]	Validation[d]	URL	Ref
SPROUT	DnD	Commercial	*ab initio* DnD + H2L (Growing) One of the first published DnD software	Prospective (BACE-1, Dihydrorotate dehydrogenase)	http://www.keymodule.co.uk	29
LigBuilder	DnD	Academics	*ab initio* DnD + H2L (Growing, Linking)	Prospective (Cyclophilin A)	http://mdl.ipc.pku.edu.cn/drug_design/work/ligbuilder.html	33,34
AutoGrow	DnD	Academics	*ab initio* DnD + H2L (Growing) External sampling engine: AutoDock Vina	Retrospective[e] (RNA Editing Lgase-1, PPARγ, Dihydrofolate reductase)	http://autogrow.ucsd.edu	36,37
LiGen™	DnD	In house	*ab initio* DnD + H2L (Growing, Linking) Pharmacophore-based Docking Computational speed (GPU)	N/A	http://www.dompe.com/Drug-Discovery-en	39
OpenGrowth	DnD	Academics	*ab initio* DnD + H2L (Growing) Graphical User Interface	Retrospective[e] (HIV-1 protease)	http://opengrowth.sourceforge.net	40
PINGUI	VS	Academics	H2L (Growing) External sampling engine: SEED	Prospective (β2 Adrenergic Receptor)	http://www.kolblab.org/scubidoo/pingui	41
LeadOp+R	VS	In house	H2L (Growing) Multi-step synthesis design	Retrospective (Tie-2 Kinase, 5-Lipoxygenase)	http://www.cmdm.tw/index.html	45

LigandScout	VS	Commercial	H2L (Growing) Design of Covalent Inhibitors Pharmacophore concept	Prospective (Enteroviral 3C Proteases)	http://www.inteligand.com/ligandscout	46,47
CROSS	VS	In house	H2L (Growing, Linking) Rescaffolding External sampling engine: BROOD	Retrospective[e] (Thrombin)	N/A	48
AutoCouple	VS	Academics	H2L (Growing) External sampling engine: rDock Computational speed	Prospective (Bromodomain)	http://github.com/Caflisch-Group/AutoCouple_Python-based	50
DOTS	VS	In house	H2L (Growing, Linking) Integrated strategy Design of Covalent Inhibitors External sampling engine: S4MPLE	Prospective (Growing with Bromodomain) Retrospective (Linking with FXa)	http://2p2idb.cnrs-mrs.fr/dots.html	27

[a] Main method to search optimized compounds. DnD: De novo Design; VS: Virtual Screening.

[b] Availability of the programs. Commercial: requires a Commercial license; Academics: program accessible via a server or scripts can be downloaded; In house: programs and scripts are not distributed.

[c] Main characteristic and features of the programs.

[d] Type of validation and biological targets in parenthesis.

[e] Analogs of active reference compounds have been generated but they were not experimentally validated.

Table 1. Hit to lead algorithms discussed in this review. The different approaches can be divided into two major categories according to the method used to optimize the original hit, *de novo* design (chapter 2) and virtual screening (chapter 3). All these methods rely on explicit synthesis rules to design new compounds except for OpenGrowth and they all use medicinal chemistry-like filters to discard molecules with undesirable properties.

Figure 1

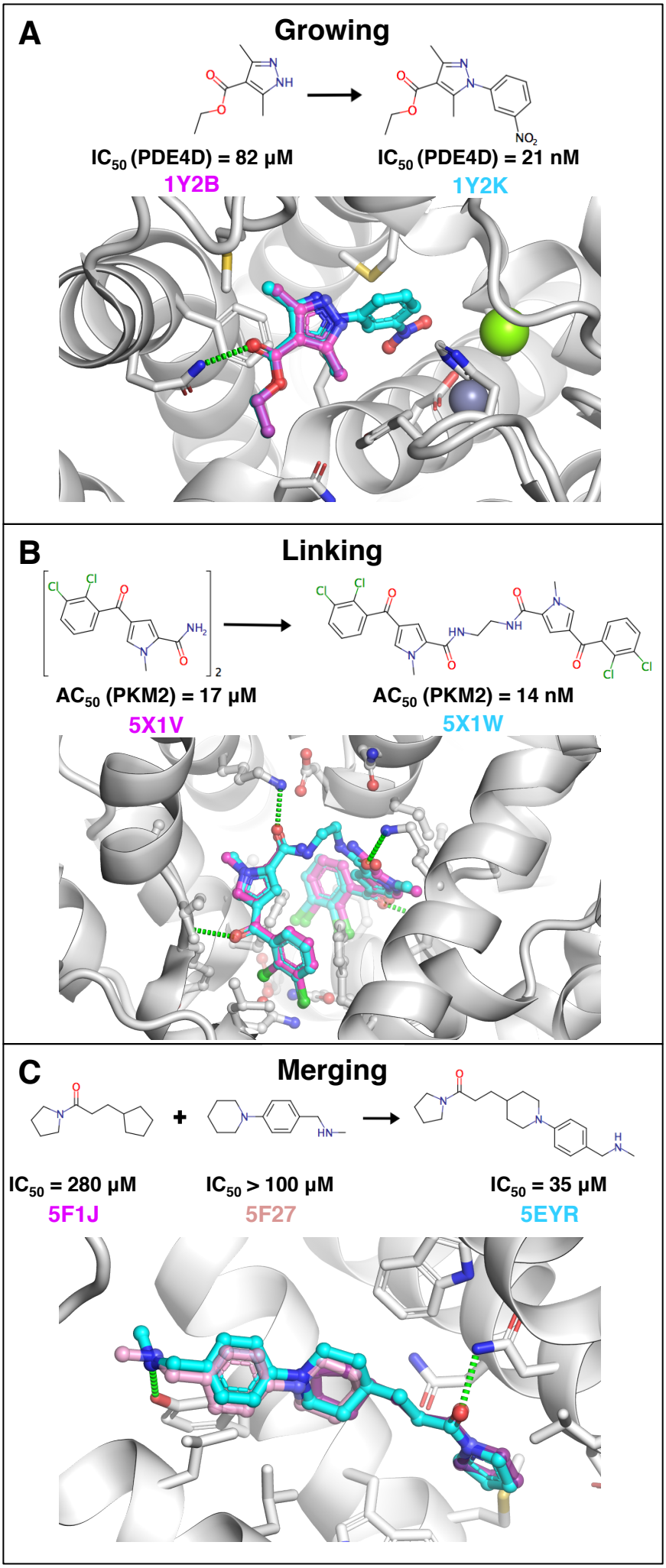
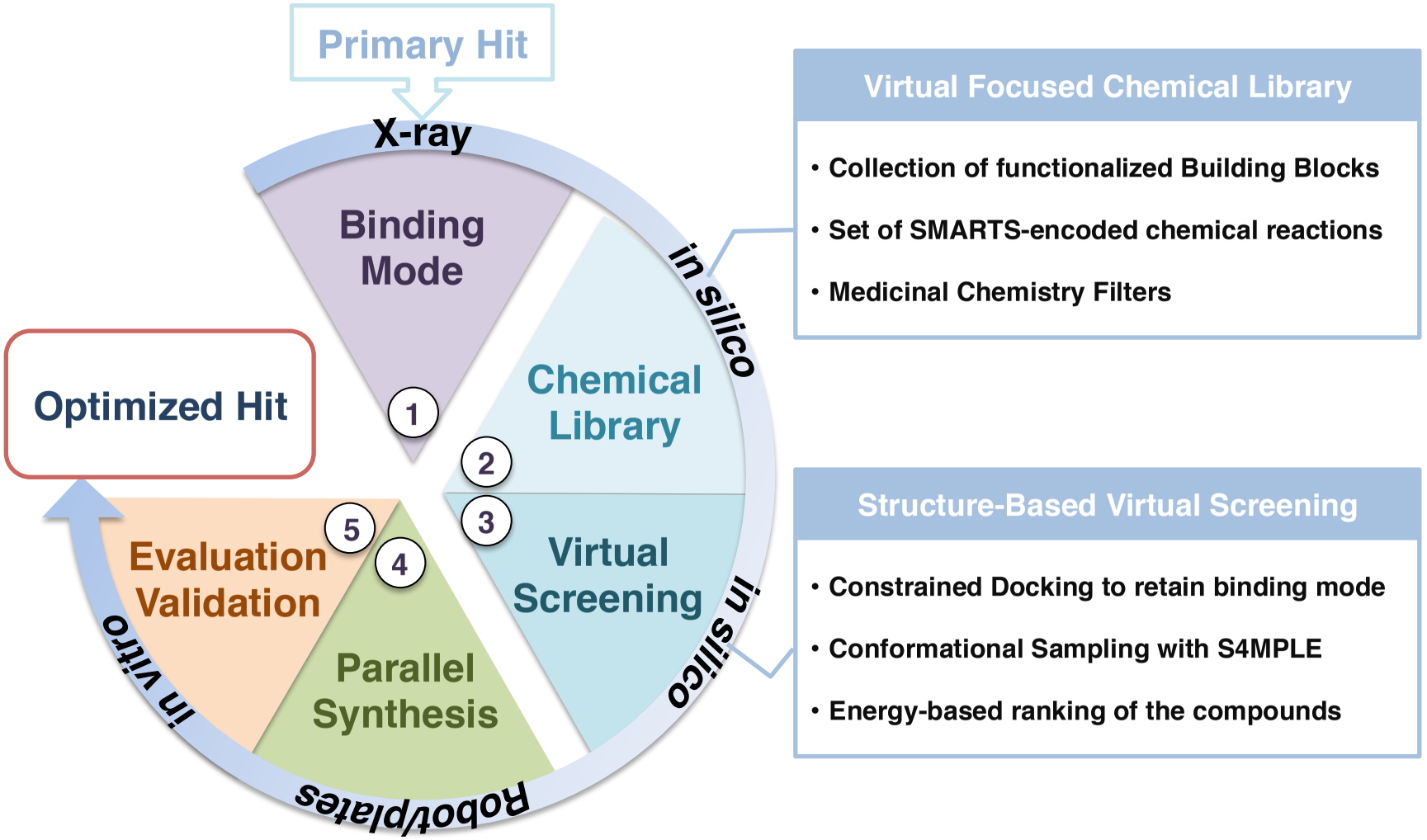

A
Growing
IC50 (PDE4D) = 82 μM
1Y2B
IC50 (PDE4D) = 21 nM
1Y2K
B
Linking
AC50 (PKM2) = 17 μM
5X1V
AC50 (PKM2) = 14 nM
5X1W
C
Merging
IC50 = 280 μM
5F1J
IC50 > 100 μM
5F27
IC50 = 35 μM
5EYR

Figure 2


Primary Hit
X-ray
Binding Mode
1
Chemical Library
2
3
Virtual Screening
4
Parallel Synthesis
5
Evaluation Validation
Optimized Hit
in silico
in silico
Robotplates
in vitro
Virtual Focused Chemical Library
• Collection of functionalized Building Blocks
• Set of SMARTS-encoded chemical reactions
• Medicinal Chemistry Filters
Structure-Based Virtual Screening
• Constrained Docking to retain binding mode
• Conformational Sampling with S4MPLE
• Energy-based ranking of the compounds