# LigBuilder V3: A Multi-Target *de novo* Drug Design Approach

Yaxia Yuan[1], Jianfeng Pei[2]* and Luhua Lai[1,2,3]*

[1] Beijing National Laboratory for Molecular Sciences, State Key Laboratory for Structural Chemistry of Unstable and Stable Species, College of Chemistry and Molecular Engineering, Peking University, Beijing, China, [2] Center for Quantitative Biology, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China, [3] Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China

With the rapid development of systems-based pharmacology and poly-pharmacology, method development for rational design of multi-target drugs has becoming urgent. In this paper, we present the first *de novo* multi-target drug design program LigBuilder V3, which can be used to design ligands to target multiple receptors, multiple binding sites of one receptor, or various conformations of one receptor. LigBuilder V3 is generally applicable in *de novo* multi-target drug design and optimization, especially for the design of concise ligands for protein targets with large difference in binding sites. To demonstrate the utility of LigBuilder V3, we have used it to design dual-functional inhibitors targeting HIV protease and HIV reverse transcriptase with three different strategy, including multi-target *de novo* design, multi-target growing, and multi-target linking. The designed compounds were computational validated by MM/GBSA binding free energy estimation as highly potential multi-target inhibitors for both HIV protease and HIV reverse transcriptase. The LigBuilder V3 program can be downloaded at "http://www.pkumdl.cn/ligbuilder3/".

**Keywords:** *De novo* design, Multi-target drug design (MTDD), multi-target drug optimization, Dual-functional inhibitors, LigBuilder

## INTRODUCTION

For most of the twentieth century, drug discovery process was dominated by a reductionist "one disease, one target, one molecule" philosophy (Alcaro et al., 2019). Researchers and pharmaceutical industries around the world have been struggling to develop highly specific regulators against particular targets, which are generally expected to achieve higher potencies while reducing the risk of off-target related side effects (Eaton et al., 1995; Morphy and Rankovic, 2009; Hughes et al., 2011). Although successful drugs have been brought to market with this approach, new drug R&D aiming novel targets was noticeable slowdown and fewer drugs were approved over the last decades (Scannell et al., 2012; Ramsay et al., 2018), which implies the limitation and deficiency of previous single-target drug discovery strategy. Due to the complexity of biological network (Gerstein et al., 2012), disease usually involves multiple factors and biological pathways, so agents that directly interfere individual molecular targets often lack effectiveness at treating complex diseases (Brown and Superti-Furga, 2003; Kamb et al., 2007; Cavalli et al., 2008; He et al., 2016). Moreover, the upstream components of pathways have to be regulated if only one target is aimed at in a multiple pathology related disease, which is more likely to cause unexpected side effects. Consequently, researchers and pharmaceutical industries have been turning their attention to develop therapies that modulate multiple targets simultaneously (Reddy and Zhang, 2013; Zhang et al., 2017; Kumar and Sharma, 2018). Combination therapy and multi-target therapy were proposed to address this problem.

Combination drugs, which is defined as a concerted pharmacological intervention of multiple targets with several compounds, have been used increasingly to treat many types of diseases, such as viral and bacterial infection, cancer, hypertension, and atherosclerosis (Giles et al., 2014; Von Hoff et al., 2014; Blonde et al., 2015; Lu et al., 2018). Although the combination therapy is proposed to set up a new direction for drug discovery, it is not a new concept. In fact, using multi-component mixture extracted from natural products is a historical therapy in traditional medical treatments. Besides, the highly active antiretroviral therapy (HAART) (Lu et al., 2018), which is also known as the "AIDS cocktail," has been the first-line anti-AIDS treatment since the end of last century (Bhatti et al., 2016). Many combination drugs have been launched to market and proved to be effective therapies for complex diseases, however, poor patient compliance has been raised especially in treatment of asymptomatic diseases such as hypertension (Eisen et al., 1990). An alternative way to simplify drug dosing is to mix multiple drug components into single co-formulated tablet, but different PK/PD property of each component may complicate the formulation and raise the risk of drug-drug interaction, and increase the risk and cost of such fix dose combinations strategy (Morphy and Rankovic, 2009).

Multi-target drug, which is defined as single compound that interacts with multiple targets simultaneously, has been paid much attention recently. Multi-target therapy is expected to be new and more effective medications for a variety of complex diseases even with relatively weak activity (Korcsmaros et al., 2007; Zimmermann et al., 2007). The uniform chemical component of multi-target drug will introduce lower risk of drug-drug interaction comparing with multi-components strategy. Moreover, although the discovery process of multi-target drug will be more complicated in the design and optimization stage due to the increased constraints from multiple targets, the risk and costs for the most expensive clinic trial stage are in principle similar with traditional single-target drug development. Consequently, many methods for multi-target ligand discovery were developed (Morphy et al., 2004; Zhan and Liu, 2009; Abdolmaleki et al., 2017; Zhang et al., 2017), such as multi-target QSAR (González-Díaz et al., 2006), fragment linker strategy (Morphy and Rankovic, 2006), framework combination (Morphy and Rankovic, 2006; Chen et al., 2011), and common pharmacophore based virtual screening and cross screening (Wei et al., 2008). Among them, framework combination and cross screening are both widely used approaches for discovering of multi-target lead (Morphy and Rankovic, 2005, 2009, 2010; Wu et al., 2012; Lepailleur et al., 2014; Bottegoni et al., 2016). Framework combination approach is based on the integration of multiple compounds via the fusion of common or similar sub-structure. Although the combined molecule from this approach is usually much smaller than directly linking two distinct structures with flexible chain, the median ligand efficiency is typically lower than general preclinical compounds which may lead to possible poor oral pharmacokinetics (Morphy and Rankovic, 2007). An alternative way is to screen multiple targets with the same compound library and select the consensus hints, namely, cross screening (Geppert et al., 2010). Although reported compounds

derived by cross screening are better in ligand efficiency than that of framework combination approach, they are still statistically less efficient than general preclinical compounds. Considering the requirement of interacting with distinct binding sites, we are not surprising in the relative low ligand efficiency of multi-target compounds designed by the above methods (Morphy and Harris, 2012). Therefore, it is critical for multi-target compounds to be "highly integrated" that could make the most of each component group in multiple interactions. Moreover, the optimization of multi-target lead is far more complicated than that of single-target lead, because the "optimization landscape" of multi-target lead is no longer a simple stepwise "group-activity" profile in single-target lead optimization. The requirement of binding affinity balance for multiple binding will significantly reduce the available chemical space of the lead structure, as a result, stepwise optimization in multi-target optimization easily leads to "the blind alley," namely, local minima. The increased dimensions in "optimization landscape" of multi-target lead optimization make the stepwise strategy less efficient, and implies that a more global and extensive structure sampling is necessary in optimization, which may be difficult to be achieved by manual work. It also suggests that a "one-step" design rather than routine "optimizing-bioassay" cycle is more suitable for multi-target drug discovery process. Therefore, the efficient discovery strategy of "highly integrated" ligand for unrelated targets remains challenging and a general strategy of multi-target rational drug design for dissimilar targets needs to be developed.

We developed an innovative multi-target design method, called LigBuilder V3, which enables the *de novo* design and molecular optimization algorithm to handle multiple targets. The chemical space exploration algorithm inherited from LigBuilder V2 (Yuan et al., 2011) has been upgraded to explore more sophisticated structure space of multi-target ligands. As we design the multi-target ligands from scratch with the consideration of multiple interactions of each component group, high ligand efficiency is expected to be achieved with this *de novo* design approach, which is very important for multi-target drugs. Multi-target lead optimization is also implemented in LigBuilder V3, which can help researchers to find possible multi-target optimization solutions. Furthermore, we apply an "ensemble linking" strategy to promote the efficiency of "fragment linking" algorithm and make it available in linking fragments for multi-target design, which is helpful in highly efficient recombination of known ligands and framework combination.

## METHOD AND ALGORITHM

### Data Structure and Definition

LigBuilder V3 implements the same genetic algorithm (GA) (Fraser, 1957; Bremermann, 1958; Holland, 1975; Whitley, 1994) used in LigBuilder V2. GA is an optimization algorithm inspired by the process of natural selection, and it mimics the evolution of a population under selection pressure. LigBuilder V3 uses the overlapping generation model of GA, that is, new generation of individuals are evolved from previous population and then replace their parents with GA iteration. For a typical overlapping generation model of GA, roulette wheel selection approach
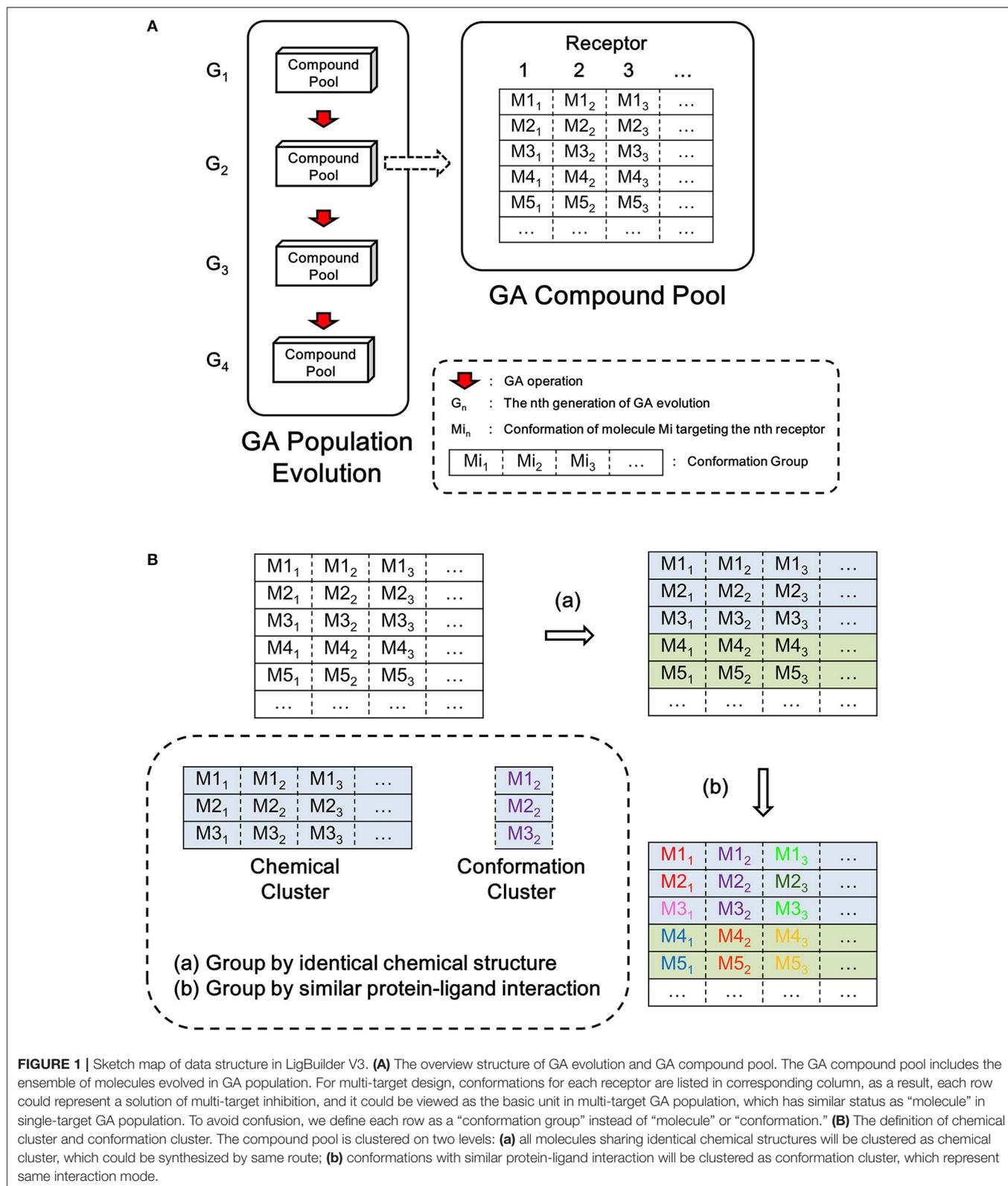
**FIGURE 1 |** Sketch map of data structure in LigBuilder V3. **(A)** The overview structure of GA evolution and GA compound pool. The GA compound pool includes the ensemble of molecules evolved in GA population. For multi-target design, conformations for each receptor are listed in corresponding column, as a result, each row could represent a solution of multi-target inhibition, and it could be viewed as the basic unit in multi-target GA population, which has similar status as "molecule" in single-target GA population. To avoid confusion, we define each row as a "conformation group" instead of "molecule" or "conformation." **(B)** The definition of chemical cluster and conformation cluster. The compound pool is clustered on two levels: **(a)** all molecules sharing identical chemical structures will be clustered as chemical cluster, which could be synthesized by same route; **(b)** conformations with similar protein-ligand interaction will be clustered as conformation cluster, which represent same interaction mode.

is used to select 10% members from current population as parent for evolving next generation, and all members in current population will be discarded. To balance quality and diversity of population, LigBuilder exempt the top 10% members in current population from elimination, that is, these top members will be directly transferred to next generation. So the quality of member

in offspring generation will be better, at least equal to parent generation. We define the GA compound pool as the ensemble of molecules in the newest generation of GA population evolution. The overview of the data structure used in GA evolution is described in **Figure 1A**.

As each molecule produced by multi-target design method involves multiple proteins, we should consider the multiple conformations of the molecule that bind to its corresponding targets. This is different to single target drug design method. To avoid confusion, we use "Conformation Group" instead of "Molecule" and "Conformation" to indicate the multiple binding conformations of multi-target molecule in this manuscript (**Figure 1A**).

We also define the "chemical cluster" and "conformation cluster" to depict the relationship in chemical structure and binding conformation among multi-target molecules (**Figure 1B**). From the perspective of synthesis, molecules with the same chemical structure could be considered as identical. But from the perspective of protein-ligand interaction, the conformations of ligand must be taken into account because the binding of ligand is based on spatial interaction between atoms from ligand and protein. Therefore, we cluster all conformations at two levels: (1) chemical clusters: each conformation of a chemical cluster shares the same two-dimensional (2D) structure, and they could be synthesized via the same reactions estimated by the synthesis-accessibility analysis module inherited from LigBuilder V2; (2) conformation clusters: all the conformations in a conformation cluster also share the same 2D structure, thus the conformation cluster is a subset of chemical cluster. All conformations in a conformation cluster are similar with each other, so they could be consider as sharing same interaction mode. Although all conformations of a conformation cluster are interchangeable from the perspective of interaction mode, we have to keep these "duplicates," because they may provide necessary local perturbation, for example, the members in a conformation cluster may have various orientations of hydrogen atoms. The orientations of hydrogen atoms usually have little effect on protein-ligand binding except being involved in hydrogen bond forming, but it is much sensitive in further evolution of molecules because the hydrogen atom is responsible for growing site for connecting newly added fragments.

## Multi-Target Seed Structure Mapping

Seed structure is the starting point structure for lead optimization. The preparation of seed structures for single target lead optimization is straightforward, however, additional steps are needed for preparation of seed structures for multi-target design. As each "multi-target seed structure" indicates a conformation group which is composed by the different binding conformation of the ligand to each target, therefore it is necessary to make one-to-one correspondence between atoms of each member in the conformation group. Because only hydrogen atoms are possible connection site in the whole design process, the seed structure mapping is based on the mapping of hydrogen atoms, namely, hydrogen mapping. Due to the symmetry of molecule, there may be more than one possible solution of hydrogen mapping between two structures.



**FIGURE 2 |** Symmetry in hydrogen mapping. **(A)** Example for the symmetry of structure. 2 $C_2$ symmetry axis of 1,4-dichlorobenzene conduce to 4 possible hydrogen mappings. **(B)** Example for the symmetry of group. A $C_3$ symmetry axis of methyl group of acetic acid conduces to three possible hydrogen mappings. The hydrogen atom colored in red and blue indicates the first hydrogen and last hydrogen in hydrogen mapping, respectively.

As depicted in **Figure 2**, two types of symmetry should be taken into account, i.e., the hydrogen symmetry of molecule and the hydrogen symmetry of group. The molecular hydrogen symmetry refers to the rotation symmetry of all hydrogen in the molecule, and the hydrogen symmetry of group refers to the rotation symmetry of multiple hydrogen atoms that connected to one heavy atom. **Figure 2A** shows two $C_2$ symmetry axises of 1,4-dichlorobenzene, which conduce to 4 possible hydrogen mappings. **Figure 2B** shows a $C_3$ symmetry axis of the methyl group of acetic acid, which conduces to 3 possible hydrogen mappings. We should note that although some molecules such as the acetic acid are not chiral, the potential chirality is taken into account for the hydrogen mapping in LigBuilder V3, because the further growing operation may bring in chirality to the carbon atom. In other words, both 2D topological and three dimensional (3D) structural information are considered in hydrogen mapping.

Although all the hydrogen atoms in ligand are possible fragment growing site, not every hydrogen atom could serve as growing site because of steric hindrance or user's preference. Thus, LigBuilder V3 only reserves the possible hydrogen mappings with maximal growing sites mapped, named growing site mapping, which increase the possibility of further growing operation to the greatest extent. If there is no possible

**FIGURE 3 |** Sketch map of the growing operation. **(A)** The seed structure in the binding site and chosen fragment (dashed box) in the building block library (solid box). All potential growing sites are colored in blue. **(B)** The randomly chosen growing site of seed structure and chosen fragment are colored in red. **(C)** The chosen fragment is moved and then attached to the seed structure. The favorable conformations are determined by uniformly 3 degree-step sampling of the torsion angle along the newly formed bond, which is colored in red.

hydrogen mapping or growing sites mapping, the corresponding conformation group will be ignored. However, if there are more than one solutions of rational hydrogen mappings, LigBuilder V3 will regard them as different seed structures and use them independently in subsequentially design process.

For the case that involves more than two targets, LigBuilder V3 makes hydrogen mapping between the conformation for the first target and each of the remaining targets one by one. As a result, all the rest conformations of the conformation group are mapped to the first conformation, so it is feasible to the find the common growing site mappings of the whole conformation group.

## Multi-Target Growing

Lead optimization is the fundamental function of LigBuilder series. Both LigBuilder V1 and V2 provide the "Growing" strategy, which generates derivatives based on the lead structure (i.e., "seed" structure) that has been pre-placed into the binding pocket. In the present study, we extend the "Growing" strategy to multi-target growing (multi-target lead optimization).

**Figure 3** is the sketch map of single-target growing operation, which is the basis of multi-target growing operation. The gray area on the left in **Figure 3A** represents the binding site of the target, and the benzene is a representative seed structure.

Molecules in the solid box on the right are privileged fragments, which could serve as the building blocks for assembling new structure. Although all hydrogen atoms are feasible for attaching fragments, only a few of them are potential connection site without steric hindrance. Taking **Figure 3A** for example, hydrogen atoms of benzene face to the vacant region of binding site are colored in blue, which indicate the potential growing sites, and the others near to the receptor atoms will be ignored. Meanwhile, all the hydrogen atoms of building blocks will be considered as potential connection site by default. Users can also assign or block certain "growing sites" on seed structures and building blocks to customize the style of molecule. As the seed structure and building block library has been prepared, LigBuilder will randomly choose a fragment from the building block library (the dashed box in **Figure 3A**), and then randomly choose a potential growing site on the seed structure and the chosen building block, respectively (red hydrogen atoms in **Figure 3B**). The building block will be attached to the seed structure along the direction of selected hydrogen atoms (red hydrogen in **Figure 3B**). With uniformly 3 degree-step sampling of the torsion angle along the newly formed bond (red bond in **Figure 3C**), several favorable conformations with local minimal energies will be reserved as candidates in consideration of the flexibility of molecule. GA is applied

**FIGURE 4** | The sketch map of growing process. The multi-target growing operation could be considered as multiple synchronous single-target growing operation. The fragments grown in each step are colored in red (Only one conformation group of each generation are showed here as representatives of the compound pool).
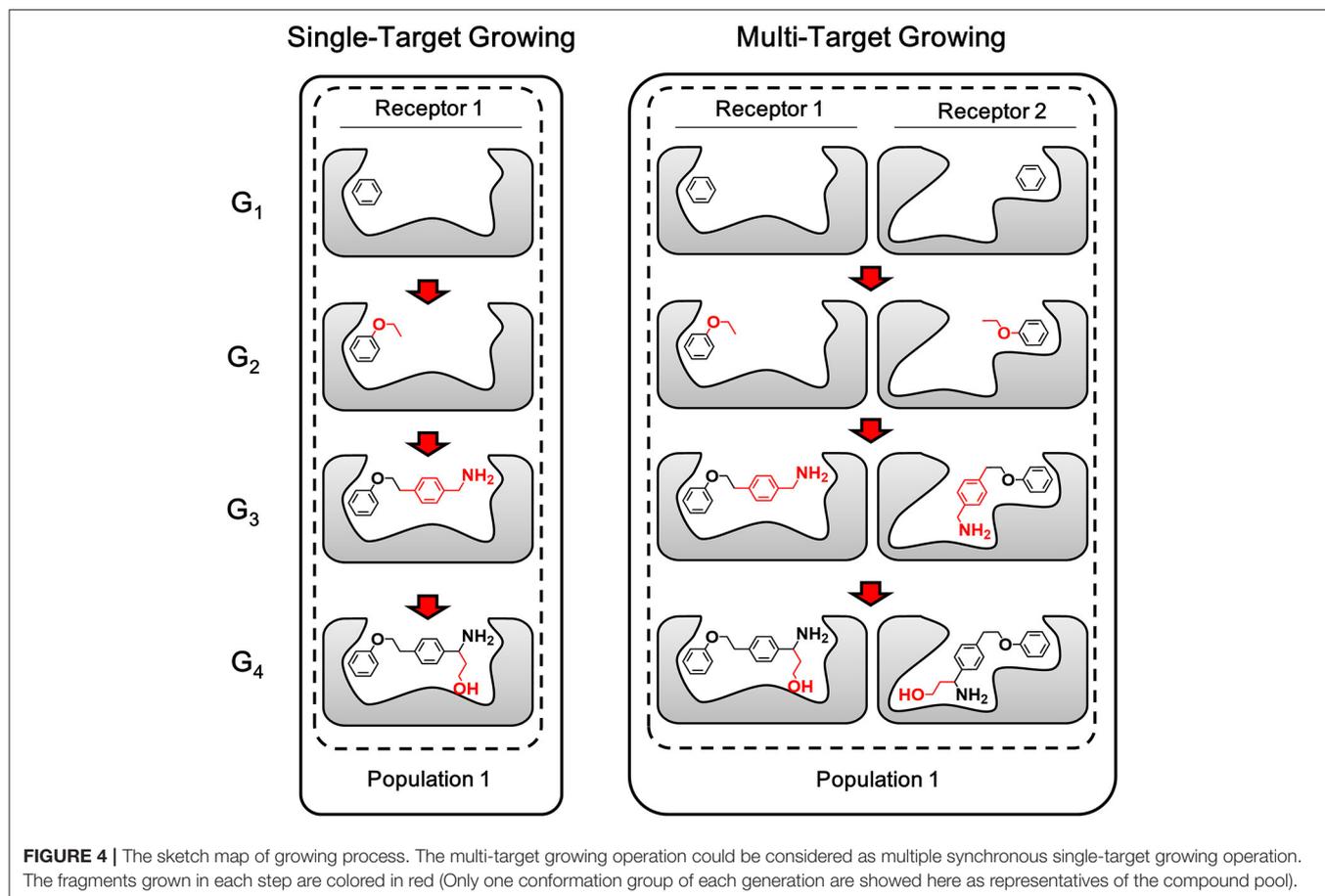
to select elites from these candidates, and these elites will serve as the seed structures for the next growing cycle. This repeated process for each ligand continues until: (1) the ligand is fully designed and there is no available space for adding any new chemical group; (2) the ligand reaches the limitation of molecular weight, which is 480 Da by default; (3) the GA generation number reaches a maximal number, which is 15 by default.

Different from single-target design, the lead structure for multi-target design should be prepared as "seed" conformation group, which is composed by the binding conformation of the lead structure to each target. With a simultaneous operation of growing chemically identical building block on the same site of each member in the conformation group, compounds generated by LigBuilder V3 are expected to be capable of binding to multiple targets. As depicted in **Figure 4**, multi-target growing could be considered as multiple synchronous single-target growing operation. The identical building block and the same growing site in the growing operation will maintain the consistency of 2D structures of the conformation group. Meanwhile, the 3D conformation of ligand is only restrained by its corresponding targets, that is, the conformation in each conformation group is optimized and evaluated independently. Therefore, this strategy could utilize the flexibility of ligand

to improve the capability of multi-target binding. Genetic Algorithm (GA) is also applied to manipulate the growing cycle in the same manner as single target growing.

## Ensemble Linking

Although assembling several bioactive fragments to generate potent ligand is very promising, the computational method of linking proximal fragments covalently is fraught with challenges. To avoid affecting respective bioactivity, the orientation and position of fragments should be changeless. Therefore, the feasibility of linking is severely limited by the rigid restriction of bond length and bond angle in molecule. Besides, the unfavorable energy of torsion may further reduce the feasibility. Although there may be some solutions existed in huge chemical space, the low efficiency in finding these solutions narrows the application of fragments linking. As a result, there are few successful cases of fragments linking, except using flexible chain as linker. However, although flexible chain could be used to relax the rigid restriction of linking, it may increase the amount of accessible conformation of the structure which brings in unfavorable entropy change during binding process, thus the linked fragments usually do not bind as the same degree as the sum of the individual fragments.

**FIGURE 5 |** The sketch map of linking and ensemble linking process. The red arrow indicates the growing operation while blue arrow indicates the linking operation. In accordance with the color of arrows, atoms, and bonds formed in growing and linking operation are colored in red and blue, respectively. For previous linking algorithm, the aggregation of certain fragments is regarded as a whole "seed," so the linking process will be constrained by every fragment in the aggregation. Instead, ensemble linking algorithm applies a more flexible linking strategy, which attempts to derive new structures from each of seed fragments independently and then find possible way of linking among these structures. The different populations indicated different independent GA threads.

Moreover, the excessive flexibility of structure may reduce the specificity of ligand.

The ensemble linking algorithm is developed in LigBuilder V3 to improve the efficiency of fragments linking process, which is expected to make this method more practical. To demonstrate the details of the new algorithm, the sketch map of the linking algorithm used in previous versions of LigBuilder series and ensemble linking algorithm used in LigBuilder V3 are compared in **Figure 5**. Previous linking algorithm applies a direct strategy of linking, which aims to linking certain fragments with many building blocks. Instead, ensemble linking algorithm applies a more flexible linking strategy, which attempts to derive new structures from each of seed fragments independently and then find possible way of linking among these structures. Although both algorithms are capable of generating the same final structure in **Figure 5**, ensemble linking strategy is expected to be more efficient. For the general linking algorithm, it is straightforwardly requiring that all the given fragments should be linked, which is usually hard especially for multi-target linking. To overcome this problem, ensemble linking algorithm is based on extra linking fragments, and automatically find the apportioned combination of fragments, which would improve the possibility of finding solution for linking fragments. To be specific, for general linking algorithm, the number of fragments used for linking

is limited, for example, user poses 3 fragments into the ligand binding site, and the linking algorithm attempts to find suitable linkers to connect all these 3 specific fragments. For ensemble linking, user could pose several thousands of fragments into the ligand binding site, and the ensemble linking algorithm attempts to find suitable linkers to connect any 3 fragments among all available fragments. Obviously, the ensemble linking algorithm will significantly increase the possibility of finding suitable solution for linking 3 fragments comparing with general linking algorithm. Besides, with dissociation of combined seed fragments in linking algorithm, the whole linking process would be more robust, it would not be dragged by improper derivation or conformation of individual fragment. Moreover, it raises the possibility of comprehensive utilization of more bioactive fragments without exhaustive combination. As a result, LigBuilder V3 could be applied to find possible solutions of linking among hundreds of fragments, which further improves the success rate of linking.

The ensemble linking algorithm will generate many derivative candidates based on the linking fragments to enhance the possibility for finding solution, obviously, although the possibility of linking increases with the number of candidates, the computation cost also will increase by the same rate. So it is important to generate candidates more effective rather than

**FIGURE 6 |** The flow chart of linking process in LigBuilder V3. The structures of initial compound pool are randomly selected from the seed pool in the initialization stage. Then LigBuilder V3 perform the growing and linking operation on each compound pool for producing new generation of compound pools. The growing operation brings in derivation on molecules in previous compound pool and result in a "Growing pool," which is colored in red. Then the linking operation will be performed for finding possible way of linking between these newly formed structures and previous existed structures. The "previous existed structures" in each generation is denoted with the dashed box. As similar with growing operation, the linking results will be collected into a "Linking pool" which is colored in blue. At last, the "Growing pool" and "Linking pool" will be merged into a new generation of compound pool.

increase the number of candidates to improve the efficiency. As all candidates in the GA population have similar molecular weight because they are generated by GA evolution with same number of generations, the linking possibility will reach the peak when candidates in the population occupy about half of the binding site. But it rapidly falls when candidates in the population are too large to be integrated in limited space of binding site. In addition, the linking possibility is also low when the candidates in the population are too small which may make them far away from each other for linking. So we applied a stagger strategy that operating several independent GA threads simultaneously, meanwhile, the starting of each GA threads are staggered so as to make them be in various generation of GA process. That is, ensemble linking algorithm will not only perform "intra-linking" among candidates in a GA process, but also perform "inter-linking" among candidates in different GA threads. With this strategy, the high diversity of molecular weight distribution among all candidates could bring in higher linking possibility and efficiency.

As depicted in **Figure 6**, each generation of ensemble linking can be decomposed into two steps, i.e., the growing step and the linking step. LigBuilder V3 performs the growing operation on

all compounds from each compound pool in the growing step, and then finds possible way of linking between the newly formed compounds and all previous existed compounds including seed pool (dashed box in **Figure 6**) in the linking step. Although compounds generated in both steps will be collected together into new generation of compound pools, the compounds generated in the linking step (linking pools in **Figure 6**) will have a certain level of priority in GA process, which make the ensemble linking algorithm trends to link fragments rather than grow for derivation. To be specific, the compounds generated in the linking step indicates a "linking" operation is occurred, on the contrary, compounds generated in the growing step do not link with other fragments in this step. So LigBuilder will elevate the fitness score of compounds from linking step, which encourage the linking behavior. The structures of initial compound pool are randomly selected from the seed pool. After the initialization, LigBuilder V3 will repeat the ensemble linking process for each ligand until: 1) the ligand is fully designed and there is no available space for any new chemical group; 2) the ligand reaches the limitation of molecular weight, which is 480 Da by default; 3) the GA generation number reaches a maximal number, which is 15 by default.

**FIGURE 7 |** The multi-target ensemble linking operation could be considered as multiple synchronous single-target ensemble linking operation. The red arrow indicates the growing operation while blue arrow indicates the linking operation. In accordance with the color of arrows, atoms, and bonds formed in growing and linking operation are colored in red and blue, respectively.
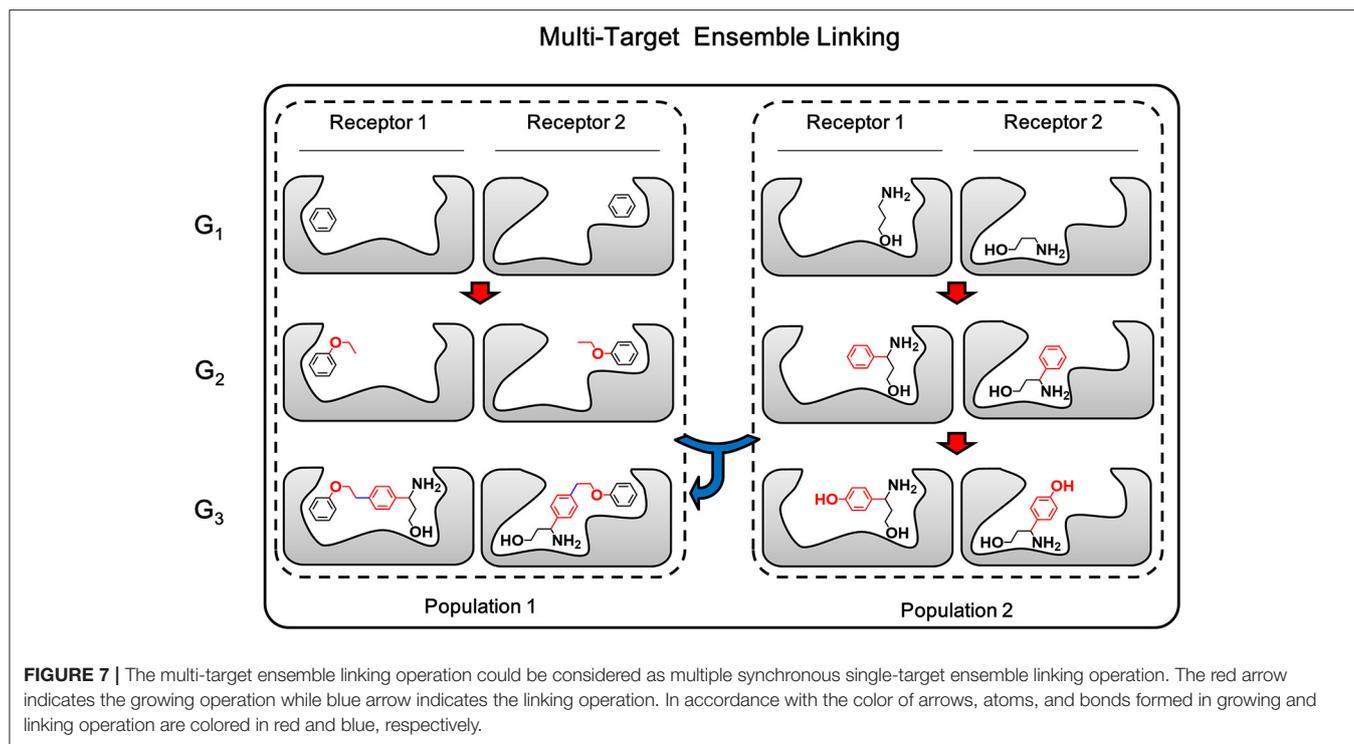
## Multi-Target Linking

A further challenge lies in designing multi-target ligand is linking fragments that interacting with multiple targets. Although many successes occurred in designing single target ligand by fragments linking strategy, few research focus on multi-target linking method. Comparing with lead compound, potential active fragments are much easier to pick by fragment-based approach, such as NMR, DSF, X-ray crystallography, surface plasmon resonance and mass spectrometry (Mashalidis et al., 2013). In addition, computational methods such as fragment docking (Wang et al., 2015) or CrystalDock (Durrant et al., 2011) are also effective ways to identify lead-fragments. Moreover, small fragments are much more likely to interact with multiple targets due to its lower specificity. Therefore, it is feasible and promising to design multi-target ligand by integrating several fragments. So we try to improve our ensemble linking algorithm to handle multi-target fragments linking in LigBuilder V3.

As with the multi-target growing algorithm, the fragments for multi-target linking should also be prepared as "seed" conformation group. Because the ensemble linking algorithm in LigBuilder V3 handle the fragments independently, users no longer have to predetermine which fragments will be linked together in the stage of seed preparation. That is, the procedure of conformation group preparation for linking is same as that for growing. The only difference is that at least 2 conformation groups should be prepared for linking and at least 1 for growing.

Multi-target linking algorithm is based-on the ensemble linking algorithm described in above section, therefore both the growing step and linking step which make up the ensemble linking algorithm will be extended to multiple targets. The

growing step of multi-target ensemble linking is exactly the same with multi-target growing, while the linking step applies the same strategy of "multi-target operation" used in multi-target growing. As described above, multi-target growing operation is a simultaneously operation of growing chemical identical building block on the same site of each member in the conformation group. In a similar way, multi-target linking operation in linking step is a simultaneously pairwise operation of linking corresponding members from two conformation groups on the same linking sites (**Figure 7**). That is, the first member of conformation group A will be linked with the first member of conformation group B. Then the second member of both conformation group will be linked together on the same linking sites of the first member. This pairwise process will be repeated until all members have been linked or any failure occurs due to steric hindrance or molecular tension. As a result, these linked structures are expected to be capable of binding to multiple receptor, while they share identical chemical structures.

## Multi-Target *de novo* Design

LigBuilder V3 inherits the "Chemical Space Exploring Algorithm" (CSEA) from LigBuilder V2 to create novel scaffolds and structures. In LigBuilder V2, CSEA works in the following way: (1) an $sp^3$ carbon with 4 hydrogen atoms will be randomly posed in the binding site and serve as the starting point of constructing new molecules with growing operation; (2) newly designed molecules will be split into fragments; (3) fragments with high predicted binding affinity, that is, high contribution fragments will be selected for updating the "seed structure pool," which is used to supply seeds for subsequent design cycles;

**FIGURE 8** | Schema of Chemical Space Exploring Algorithm (CSEA). A randomly positioned sp$^3$ carbon will be taken as starting point of growing operation for generating the initial compound pool. Then fragments extracted from compound pool will be used to construct a seed structure pool, which contains seed structures that more favorable than the initial sp$^3$ carbon. As new compound pools could be generated based-on structures in seed structure pool, while fragments extracted from these newly formed compounds will be used to update the seed structure pool, LigBuilder V3 avoids the dependency of initial structure and be promising in exploring larger chemical space.

(4) a structure from "seed structure pool" is randomly select as the starting point of constructing new molecules with growing operation, then the 2–4 steps will be repeated. For LigBuilder V3, CSEA is extended to multi-target design purpose by applying multi-target growing operation instead of single-target growing operation. Meanwhile, the fragment extraction process of single conformation will also be replaced by fragment extraction of "conformation group" (**Figure 8**). With the self-circulation seed generating feature, CSEA can help to avoid the limitations associated with pre-assigned seed structures and explore a broader chemical space, thus greatly improving the novelty and efficiency of design.

As the seed structure pool is used to collect and provide initial fragments of design process, the quality of seed structure pool may significantly affect the design results. Since CSEA could provide a mass of potential seed fragments during the design process, the seed structure pool will be updated to achieve higher binding affinity while maintain diversity of seeds. Then the CSEA will have a higher starting point for generating potent structures, which in turn produce better seed fragments for updating the seed structure pool. Therefore, the seed structure pool will keep evolving during the whole design process, which iteratively optimize the performance of design.

## Seed Structure Extraction

The most direct way of extracting component fragments from chemical structure is splitting the molecule by iterating over all single bonds. However, the traversal extraction method would take a lot of computing time when handle millions of compounds,

which is a common order of magnitude in CESA process of LigBuilder V3. Therefore, we develop a simplified extraction algorithm for acceleration. To balance the representativeness of fragments and extraction speed, we only focus on the molecular scaffold and key interaction group, which are major determinants of molecular conformation and protein-ligand interaction. As a result, only five categories of fragments are considered in CESA: (1) Single atoms; (2) rigid scaffold (rigid chain and rigid ring system); (3) flexible scaffold (flexible ring system); (4) interaction group (e.g., carboxyl group); (5) scaffold with connected interaction group. If a fragment could be classified into more than one category, it will be put into the category with smallest category number. With this algorithm, LigBuilder V3 is also capable of extracting fragments from known ligands as seed structures, which is convenient for fragments linking or lead optimization.

For multi-target design, as the seed structure is composed of multiple conformations (conformation group) instead of single conformation, seed structures extracted from known ligands of each targets should be paired to construct conformation groups first. LigBuilder V3 extracts fragments from all known ligands in the same way of extracting fragments in single-target design, and then hydrogen mapping algorithm will be implemented among these fragments to find all possible combinations. That is, if a fragment is present in ligand of every targets, it is a common fragment and LigBuilder will mapping this fragment to construct multi-target fragments group as seed structures. However, if this fragment is absent in ligands for any one of the targets, it is not common fragment and the fragment will be discarded.

## Multi-Target Ligand Efficiency

Ligand efficiency (LE) is frequently used to prioritize hits from HTS, and it can be regarded as a guide for selecting efficient fragments for further optimization. It is much more important for multi-target design because the ligand efficiency is also the index of "integration degree," which is a more critical index for multi-target drugs. There are various definitions of LE but most widely used approximation is described as the average free energy of binding per heavy atom or average $pIC_{50}$ per heavy atom, which are demonstrated as follows (Hopkins et al., 2004):

$$LE = \frac{-\Delta G}{HAC} \quad or$$
$$LE = \frac{-\log(IC_{50})}{HAC} \quad (HAC \text{ is heavy atom count})$$

For multi-target ligand, multi-target ligand efficiency (MLE) could be derived with a similar form of LE. As the multi-target ligand causes multiple binding free energy, MLE could be described as the summation of average free energy of binding per heavy atom or summation of average $pIC_{50}$ per heavy atom, which are demonstrated as follows:

$$MLE_N = \frac{\sum -\Delta G_n}{HAC} \quad or$$
$$MLE_N = \frac{\sum -\log\left(IC_{50_n}\right)}{HAC} \quad (N \text{ is the number of target})$$

The ligand efficiency is not comparable if the target number is different, so we use the subscript for MLE to indicate the condition of rational comparison and make it distinct from LE of single-target ligands. We should note that the MLE is insufficient for the performance evaluation of multi-target ligand, because the uneven activity of targeting individual binding site may be obscured by summation. However, it is much complicated to evaluate the efficiency of multi-target ligand, because it depends on the specific biological network that it is involved in. So the MLE would be only considered as an index of average efficiency for selecting potential multi-target lead structures.

## Other Functional Modules

The other functional modules implemented in LigBuilder V3 are directly inherited from LigBuilder V2, including: (1) drug-like and privileged building blocks; (2) Toxic fragments; (3) Drug-like rules (for example, Lipinski rule); (4) Ligand-binding site detection module; (5) Synthesis analysis modules; (6) Scoring function; (7) LogP module; (8) GA fitness function (composed of scoring function, MLE, toxic fragments filter, and drug-like rules). It should be noted that the binding affinity predicted by scoring function for multi-target ligand is calculated by its average of binding affinity predicted for each target.

## Design HIV-PR/HIV-RT Dual-Functional Inhibitor

### Structural Preparation

The crystal structures of PR and RT used in this study are downloaded from the RCSB Protein Data Bank (Berman et al., 2003) (PDB code: 3A2O, Hidaka et al., 2009, and 4G1Q, Kuroda

et al., 2013, respectively), and both structures are complexes with potent inhibitors solved at high resolution (0.88 and 1.51 Å, respectively). The inhibitor binding sites of PR and RT were defined by binding site detection program Cavity (Yuan et al., 2013; Zhang et al., 2015; Xu et al., 2018), which provides the detailed definition for boundary of "design space." The drug-like and privileged building blocks used in this study were inherited from LigBuilder V2. Then three different design strategies were used to design dual-function inhibitors for PR and RT.

### De novo Design Approach

LigBuilder V3 inherited the seed generation and optimization algorithm from LigBuilder V2, that is, LigBuilder V3 could iteratively extract seed structures from designed compounds and use the extract seed structures for design new compounds. The GA parameters were set as follows: GA population size of 1,000, GA parent ratio of 10%, GA generation number of 12. Total 1 million candidate dual-functional compounds were generated by LigBuilder V3 with de novo design mode.

### Growing Approach

Growing approach is for optimization of prepared seed structures. We collected all protein-ligand binding complex of PR or RT from the RCSB Protein Data Bank (Berman et al., 2003), including 323 PR-ligand complexes and 141 RT-ligand complexes (Listed in **Table S1**). All the PR-ligand complexes were aligned to the PR structure complexed with KNI-1689 (PDB code: 3A2O, Hidaka et al., 2009), and all the RT-ligand complexes were aligned to the RT structure complexed with Rilpivirine (PDB code: 4G1Q, Kuroda et al., 2013) using Pymol (Schrodinger, 2010), which could ensure that all ligands in the complexes are also aligned according to the receptor alignment. With the fragment-extraction function of LigBuilder V3, fragments with no more than 20 heavy atoms were extract from these known PR or RT ligands. As the ligand efficiency is important for the seed structure, fragments with SLE index <0.1 were removed, and a total of 2,386 fragments for PR and 1,442 fragments for RT were obtained at this stage. Then fragments for PR and RT with the same 2D structure were paired with "hydrogen mapping" algorithm mentioned above, and a total of 3,506 paired fragments were prepared. The GA parameters were set as follows: GA population size of 1,000, GA parent ratio of 10%, GA generation number of 12. Total 100 K candidate dual-functional compounds were generated by LigBuilder V3 with growing design mode based on the prepared fragments. As the binding affinity is usually related to the size of molecule, large seed fragments are more competitive than small fragments especially for the genetic algorithm used in LigBuilder. So, each fragment was independently used as the seed structure with multiple runs of LigBuilder to avoid bias to large seed fragments.

### Linking Approach

Linking approach is for integrating key fragments into new compounds. The paired fragments used in this approach were prepared in the same way of growing approach. However, all paired fragments were used together in linking approach to maximize the possibility of finding ways for linking fragments.

It would be intuitive that the more fragments provided, the better performance would be expected. The GA parameters were set as follows: GA population size of 10,000, GA parent ratio of 10%, ensemble population number of 10, GA generation number of 12. Total 10 K candidate dual-functional compounds were generated by LigBuilder V3 with linking design mode based on the 3,506 prepared fragments.

## Post-processing

As multi-target ligand should bind to different proteins with the same chemical structure, ideally, each moiety of the ligand could contribute to its binding to all targets, so ligand efficiency would be important in evaluating a multi-target ligand. In this study, predicted pKd of all the output compounds are larger than 5.0, so only MSLE index were used to rank and select top 1,000 results with best ligand efficiency from multi-target design procedure for the three approaches. Because LigBuilder V3 only uses a fast empirical scoring function for estimating protein ligand binding affinity, in order to improve the accuracy of calculation, the total 3,000 selected compounds were further subjected to energy minimization and 100 ps short time molecular dynamic (MD) simulation by using the Amber package (Case et al., 2012) for

**TABLE 1 |** Binding free energy predicted by MM/GBSA method.

| | De novo design | | Growing approach | | Linking approach | | HIV-PR potent inhibitor[b] | HIV-PR weak inhibitor[c] | HIV-RT potent inhibitor[d] | HIV-RT weak inhibitor[e] |
|---|---|---|---|---|---|---|---|---|---|---|
| | Top 10[a] | Top 1 | Top 10 | Top 1 | Top 10 | Top 1 | | | | |
| PR | −28.2 | −35 | −28.4 | −32 | −34.9 | −35.0 | −68.2 | −21.9 | – | – |
| RT | −33.6 | −34.9 | −34.5 | −40.3 | −38.0 | −38.8 | – | – | −41.5 | −22.6 |
| Ave. | −30.9 | −35.0 | −31.5 | −36.2 | **−36.5** | **−36.9** | – | – | – | – |

Energy unit is kcal/mol.

[a] Top 10 indicates the average of best 10 compounds; [b] FDA approved drug Darunavir with IC$_{50}$ of 0.15 nM (Shen et al., 2013). Complex structure was from PDB code 4LL3 (Kožíšek et al., 2014). [c] PR weak inhibitor with IC$_{50}$ of 2.3 μM (Jhoti et al., 1994). Complex structure was from PDB code 1HTE (Jhoti et al., 1994). [d] FDA approved drug Efavirenz with IC$_{50}$ of 41 nM (King et al., 2002). Complex structure was from PDB code 1FK9 (Ren et al., 2000). [e] RT weak inhibitor with IC$_{50}$ of 1.2 μM (Chan et al., 2017). Complex structure was from PDB code 5VQS (Chan et al., 2017). Bold values indicate the best average energy among results from three approaches.



**FIGURE 9 | (A)** Binding mode of FDA approved PR inhibitor Darunavir. The dimerized PR is showed in green and cyan cartoon style, and the Darunavir is showed in golden stick style (figure generated by Pymol, Schrodinger, 2010). **(B)** The 2D interaction figure for Darunavir binding with PR (figure generated by PoseView, Stierand and Rarey, 2010). **(C)** Binding mode of FDA approved RT inhibitor Efavirenz. **(D)** The 2D interaction figure for Efavirenz binding with RT.

estimating the binding affinity with MM/GBSA method (Rastelli et al., 2010).

## Application of LigBuiler V3 in Multi-Target Ligand Design

The concept prototype for growing mode algorithm of LigBuilder V3 has been experimental validated by designing COX2/LTA$_4$H dual-functional inhibitor, which resulted in a single ligand that binding to COX2 and LTA$_4$H with IC$_{50}$ of 7.1 and 7.0 μM, respectively (Shang et al., 2014). Although this work is based on a developing version of LigBuilder V3, and many manual interventions were involved due to immature of the algorithm, the success of this case suggests the feasibility of using LigBuilder V3 to design multi-target ligand. Moreover, LigBuilder V3 were further developed based on the knowledge learned from this case. Besides the improvement of multi-target growing algorithm, both multi-target *de novo* design approach and multi-target linking approach are realized in this version of LigBuilder V3. In this study, we have tested the LigBuilder V3 by designing dual-functional inhibitor targeting two well-characterized virus enzymes, HIV protease (PR) and HIV reverse transcriptase (RT) with all three design modes. As both PR and RT are important drug targets of clinical antiretroviral therapy, the multi-target strategy such as combination of nucleoside reverse transcriptase inhibitors (NRTI) and protease inhibitor (PI) shows significant advantage over each single component and has been broadly used for HIV treatment (Lu et al., 2018).

Consequently, researchers have been interesting in developing cocktail drug combinations, and pursue multi-target anti-HIV inhibitors for improving patient compliance. Matsumoto *et al.* have reported the strategy of linking PR and RT inhibitor by spontaneously cleavable linker (Matsumoto et al., 2000). Furthermore, scaffold merging strategy is successfully applied in designing multi-target anti-HIV inhibitors in recent years (Song et al., 2015; Sun et al., 2016). However, both the dependency of known inhibitors and specific requirement of molecular structure limit the practical applications of structure merging strategy. Therefore, we present a more universal solution of multi-target design with the example of designing dual-functional inhibitors for PR and RT by LigBuilder V3. The detailed methods and parameters are described in the Method and Algorithm section.

The top 1,000 compounds from each design modes were selected and subjected to 100 ps short time molecular dynamic simulation, then the binding affinity of each compounds were estimated by MM/GBSA method. The average binding affinity of the top 10 compounds and top 1 compound for each design modes were collected in **Table 1**. Although the designed multi-target compounds could not compare with the super potent PR and RT inhibitors with sub-nanomolar level activity, the designed compounds are predicted to be more potent than micromolar level inhibitor of both PR and RT, that is, these compounds are expected to be dual-functional inhibitor for PR and RT at sub-micromolar level activity for both targets.



**FIGURE 10** | Binding mode of the best compound from *de novo* design approach. **(A)** Binding mode with PR. **(B)** 2D interaction with PR. **(C)** Binding mode with RT. **(D)** 2D interaction with RT.
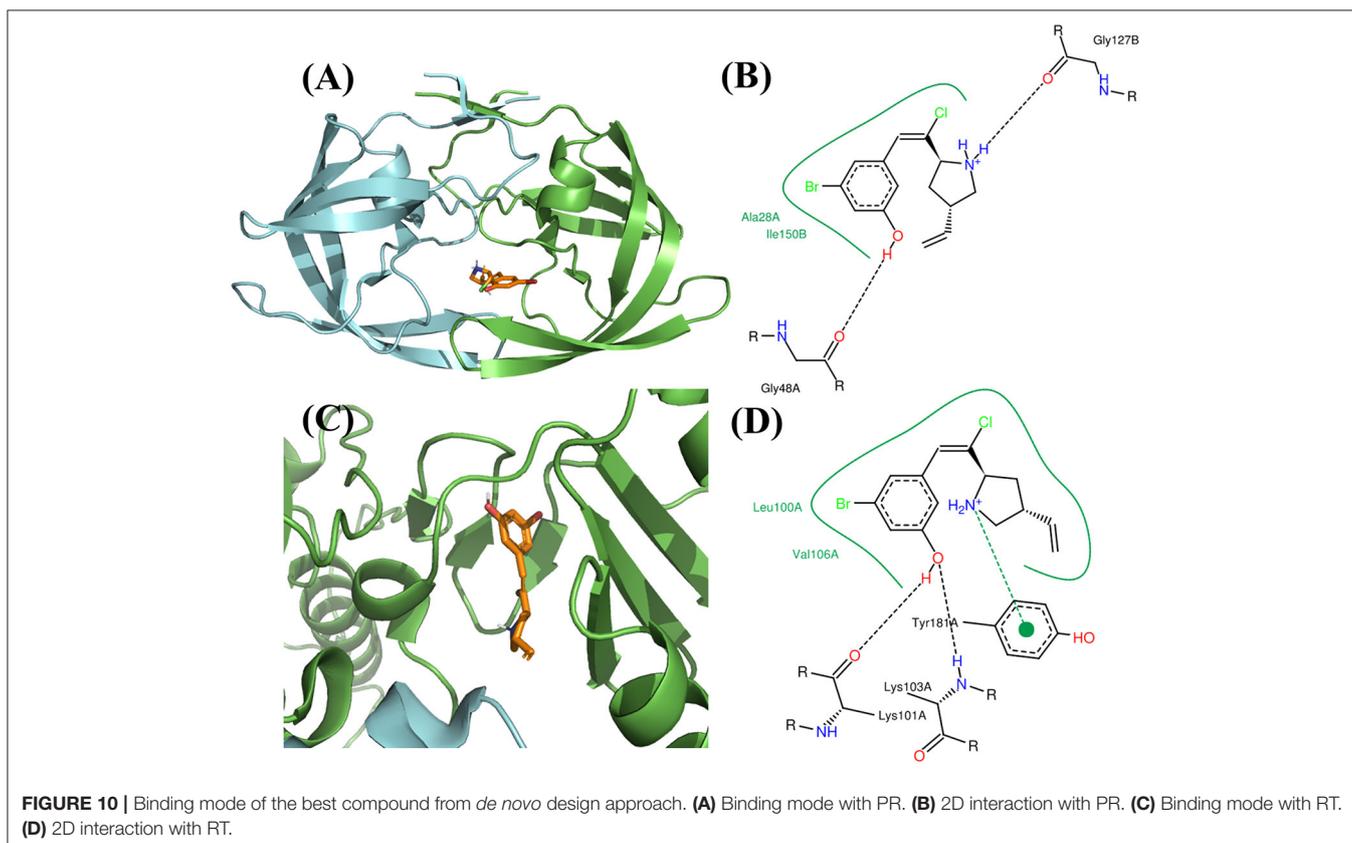
The binding mode of FDA approved PR inhibitor and RT inhibitor are depicted in **Figure 9**. Obviously, they adopt very distinct protein-ligand interaction modes. The best compound from *de novo* design approach is depicted in **Figure 10**. This compound forms distinct interaction comparing with known PR or RT inhibitors. It is highly compact and fully utilized its polar groups and hydrophobic groups to form interaction with PR and RT in different manner. The best compound from fragments growing approach is depicted in **Figure 11**. This compound is growing from a benzene ring which is one of the most common fragments in PR and RT inhibitors. The best compound from fragments linking approach is depicted in **Figure 12**. This compound is much bigger than compounds from *de novo* approach and growing approach, which indicates its relatively lower ligand efficiency. As linking approach is intensively pursing possible ways for linking provided fragments, the success of linking is more important than ligand efficiency,

so the algorithm is preferable to allow generating derivates with much lower ligand efficiency which may enhance the possibility of linking. Overall, all of these compounds are relatively small, and groups in these compounds usually contribute to the binding with different protein in different manner, which is the most need feature for designing highly compact multi-target ligand.

Essentially, the design process of LigBuilder is a kind of "random evolution" process implemented by genetic algorithm. So the quality of design result is expected to be improved along with the total computational time. As the output result could be unlimited, it is not realistic to achieve the maximal quality. For a fair comparison among the results from three strategies, we used different output number to ensure they consumed roughly similar computational time. So we designed 1 million, 100 K, and 10 K compounds for *de novo* approach, growing approach, and linking approach, respectively, which is roughly corresponding to the compound generating efficiency of three approaches in



**FIGURE 11 |** Binding mode of the best compound from growing approach. **(A)** Binding mode with PR. **(B)** 2D interaction with PR. **(C)** Binding mode with RT. **(D)** 2D interaction with RT. **(E)** Source of fragments in PR inhibitor and RT inhibitor for growing are colored in red.

**FIGURE 12 |** Binding mode of the best compound from linking approach. **(A)** Binding mode with PR. **(B)** 2D interaction with PR. **(C)** Binding mode with RT. **(D)** 2D interaction with RT. **(E)** Source of fragments in PR inhibitor and RT inhibitor for linking are colored in red.

this project. Base on the data in **Table 1**, linking approach is most effective way of design high affinity ligands, and growing approach is also more effective than *de novo* approach. This is not surprising because known fragments would provide good starting points for derivation and significantly reduce searching space. The linking approach uses more known fragments which further improve its efficiency comparing with growing approach using only 1 fragment. However, the results from *de novo* approach demonstrated that this approach could achieve similar design performance to growing approach or linking approach if more computational resource is provided. As the *de novo* approach does not reply on known fragments, it would be very useful for design ligands for new targets or discover novel ligands for known targets. On the other hand, growing approach and linking approach also have their unique advantages comparing with *de novo* approach. Since compounds designed by growing or linking approach contain "validated active fragments," it would reduce the risk of "false positive," which is very common in computer-aided drug discovery. So the three strategies could be complementary in practical drug discovery projects.

## CONCLUSION

In this paper, we present the first *de novo* multi-target drug design program LigBuilder V3. In addition, building ligands from scratch, LigBuilder V3 also provides the feasibility of multi-target lead optimization and multi-target fragments linking. This

program is generally applicable in rational and elegant multi-target drug design and optimization, especially for the design of concise ligands for proteins targets with large difference in binding sites. The developing version of LigBuilder V3 was successfully applied in designing COX2/LTA$_4$H dual-functional inhibitors with micromolar level activity. In this study, we further demonstrated the three design strategies of LigBuilder V3 with computational evaluation of designing HIV-PR and HIV-RT dual functional inhibitors. We hope the concept and LigBuilder V3 can be validated by applications from the users in the future.

## DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

## AUTHOR CONTRIBUTIONS

JP and LL conceived the project. YY, JP, and LL designed the experiments, analyzed the results, and wrote the manuscript. YY performed the experiments.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2020.00142/full#supplementary-material

## REFERENCES

Abdolmaleki, A., Ghasemi, J. B., and Ghasemi, F. (2017). Computer aided drug design for multi-target drug design: SAR/QSAR, molecular docking and pharmacophore methods. *Curr. Drug Targets* 18, 556–575. doi: 10.2174/1389450117666160101120822

Alcaro, S., Bolognesi, M. L., García-Sosa, A. T., and Rapposelli, S. (2019). Multi-target-directed ligands (MTDL) as challenging research tools in drug discovery: from design to pharmacological evaluation. *Front. Chem.* 7:71. doi: 10.3389/fchem.2019.00071

Berman, H. M., Bourne, P. E., Westbrook, J., and Zardecki, C. (2003). *The Protein Data Bank in Protein Structure*. Boca Raton, FL: CRC Press. 394–410.

Bhatti, A. B., Usman, M., and Kandi, V. (2016). Current scenario of HIV/AIDS, treatment options, and major challenges with compliance to antiretroviral therapy. *Cureus* 8:e515. doi: 10.7759/cureus.515

Blonde, L., Jendle, J., Gross, J., Woo, V., Jiang, H., Fahrbach, J. L., et al. (2015). Once-weekly dulaglutide versus bedtime insulin glargine, both in combination with prandial insulin lispro, in patients with type 2 diabetes (AWARD-4): a randomised, open-label, phase 3, non-inferiority study. *Lancet* 385, 2057–2066. doi: 10.1016/S0140-6736(15)60936-9

Bottegoni, G., Veronesi, M., Bisignano, P., Kacker, P., Favia, A. D., and Cavalli, A. (2016). Development and application of a virtual screening protocol for the identification of multitarget fragments. *ChemMedChem* 11, 1259–1263. doi: 10.1002/cmdc.201500521

Bremermann, H. J. (1958). *The Evolution of Intelligence: The Nervous System as a Model of its Environment*. Seattle, WA: University of Washington, Department of Mathematics.

Brown, D., and Superti-Furga, G. (2003). Rediscovering the sweet spot in drug discovery. *Drug Discov. Today* 8, 1067–1077. doi: 10.1016/S1359-6446(03)02902-7

Case, D. A., Darden, T. A., Cheatham T. E. III, Simmerling, C. L., Wang, J., Duke, R. E., et al. (2012). *AMBER 12*. San Francisco, CA: University of California.

Cavalli, A., Bolognesi, M. L., Minarini, A., Rosini, M., Tumiatti, V., Recanatini, M., et al. (2008). Multi-target-directed ligands to combat neurodegenerative diseases. *J. Med. Chem.* 51, 347–372. doi: 10.1021/jm7009364

Chan, A. H., Lee, W.-G., Spasov, K. A., Cisneros, J. A., Kudalkar, S. N., Petrova, Z. O., et al. (2017). Covalent inhibitors for eradication of drug-resistant HIV-1 reverse transcriptase: from design to protein crystallography. *Proc. Natl. Acad. Sci. U.S.A.* 114, 9725–9730. doi: 10.1073/pnas.1711463114

Chen, Z., Wu, Y., Liu, Y., Yang, S., Chen, Y., and Lai, L. (2011). Discovery of dual target inhibitors against cyclooxygenases and leukotriene A4 hydrolase. *J. Med. Chem.* 54, 3650–3660. doi: 10.1021/jm200063s

Durrant, J. D., Friedman, A. J., and Mccammon, J. A. (2011). CrystalDock: a novel approach to fragment-based drug design. *J. Chem. Inf. Model.* 51, 2573–2580. doi: 10.1021/ci200357y

Eaton, B. E., Gold, L., and Zichi, D. A. (1995). Let's get specific: the relationship between specificity and affinity. *Chem. Biol.* 2, 633–638. doi: 10.1016/1074-5521(95)90023-3

Eisen, S. A., Miller, D. K., Woodward, R. S., Spitznagel, E., and Przybeck, T. R. (1990). The effect of prescribed daily dose frequency on patient medication compliance. *Arch. Intern. Med.* 150, 1881–1884. doi: 10.1001/archinte.1990.00390200073014

Fraser, A. S. (1957). Simulation of genetic systems by automatic digital computers ii. Effects of linkage on rates of advance under selection. *Aust. J. Biol. Sci.* 10, 492–500. doi: 10.1071/BI9570492

Geppert, H., Vogt, M., and Bajorath, J. (2010). Current trends in ligand-based virtual screening: molecular representations, data mining methods, new application areas, and performance evaluation. *J. Chem. Inf. Model.* 50, 205–216. doi: 10.1021/ci900419k

Gerstein, M. B., Kundaje, A., Hariharan, M., Landt, S. G., Yan, K.-K., Cheng, C., et al. (2012). Architecture of the human regulatory network derived from ENCODE data. *Nature* 489, 91–100. doi: 10.1038/nature11245

Giles, T. D., Weber, M. A., Basile, J., Gradman, A. H., Bharucha, D. B., Chen, W., et al. (2014). Efficacy and safety of nebivolol and valsartan as fixed-dose combination in hypertension: a randomised, multicentre study. *Lancet* 383, 1889–1898. doi: 10.1016/S0140-6736(14)60614-0

González-Díaz, H., Prado-Prado, F. J., Santana, L., and Uriarte, E. (2006). Unify QSAR approach to antimicrobials. part 1: predicting antifungal activity against different species. *Bioorg Med Chem.* 14, 5973–5980. doi: 10.1016/j.bmc.2006.05.018

He, B., Lu, C., Zheng, G., He, X., Wang, M., Chen, G., et al. (2016). Combination therapeutics in complex diseases. *J. Cell. Mol. Med.* 20, 2231–2240. doi: 10.1111/jcmm.12930

Hidaka, K., Kimura, T., Abdel-Rahman, H. M., Nguyen, J.-T., Mcdaniel, K. F., Kohlbrenner, W. E., et al. (2009). Small-sized human immunodeficiency virus type-1 protease inhibitors containing allophenylnorstatine to explore the S2′ pocket. *J. Med. Chem.* 52, 7604–7617. doi: 10.1021/jm9005115

Holland, J. (1975). *Adaptation in Natural and Artificial Systems*. Ann Arbor, MI: University of Michigan Press.

Hopkins, A. L., Groom, C. R., and Alex, A. (2004). Ligand efficiency: a useful metric for lead selection. *Drug Discov. Today* 9, 430–431. doi: 10.1016/S1359-6446(04)03069-7

Hughes, J. P., Rees, S., Kalindjian, S. B., and Philpott, K. L. (2011). Principles of early drug discovery. *Br. J. Pharmacol.* 162, 1239–1249. doi: 10.1111/j.1476-5381.2010.01127.x

Jhoti, H., Singh, O. M., Weir, M. P., Cooke, R., Murray-Rust, P., and Wonacott, A. (1994). X-ray crystallographic studies of a series of penicillin-derived asymmetric inhibitors of HIV-1 protease. *Biochemistry* 33, 8417–8427. doi: 10.1021/bi00194a005

Kamb, A., Wee, S., and Lengauer, C. (2007). Opinion - why is cancer drug discovery so difficult? *Nat. Rev. Drug Discov.* 6, 115–120. doi: 10.1038/nrd2155

King, R. W., Klabe, R. M., Reid, C. D., and Erickson-Viitanen, S. K. (2002). Potency of nonnucleoside reverse transcriptase inhibitors (NNRTIs) used in combination with other human immunodeficiency virus NNRTIs, NRTIs, or protease inhibitors. *Antimicrob. Agents Chemother.* 46, 1640–1646. doi: 10.1128/AAC.46.6.1640-164 6.2002

Korcsmaros, T., Szalay, M. S., Csaba, B., Kovacs, I. A., and Csermely, P. (2007). How to design multi-target drugs: target search options in cellular networks. *Expert Opin. Drug Discov.* 2, 799–808. doi: 10.1517/17460441. 2.6.799

Kožíšek, M., Lepšík, M., Grantz Šašková, K., Brynda, J., Konvalinka, J., and Rezáčová, P. (2014). Thermodynamic and structural analysis of HIV protease resistance to darunavir–analysis of heavily mutated patient-derived HIV-1 proteases. *FEBS J.* 281, 1834–1847. doi: 10.1111/febs. 12743

Kumar, A., and Sharma, A. (2018). "Computational modeling of multi-target-directed inhibitors against Alzheimer's disease," in *Computational Modelling Drugs Against Alzheimer's Disease*, ed K. Roy (New York, NY: Springer), 533–571.

Kuroda, D. G., Bauman, J. D., Challa, J. R., Patel, D., Troxler, T., Das, K., et al. (2013). Snapshot of the equilibrium dynamics of a drug bound to HIV-1 reverse transcriptase. *Nat. Chem.* 5:174. doi: 10.1038/nchem.1559

Lepailleur, A., Freret, T., Lemaitre, S., Boulouard, M., Dauphin, F. O., Hinschberger, A., et al. (2014). Dual histamine H3R/serotonin 5-HT4R ligands with antiamnesic properties: pharmacophore-based virtual screening and polypharmacology. *J. Chem. Inf. Model.* 54, 1773–1784. doi: 10.1021/ci500157n

Lu, D.-Y., Wu, H.-Y., Yarla, N. S., Xu, B., Ding, J., and Lu, T.-R. (2018). HAART in HIV/AIDS treatments: future trends. *Infect. Disord. Drug Targets* 18, 15–22. doi: 10.2174/18715265176661705051 22800

Mashalidis, E. H., Sled,z, P., Lang, S., and Abell, C. (2013). A three-stage biophysical screening cascade for fragment-based drug discovery. *Nat. Protoc.* 8, 2309–2324. doi: 10.1038/nprot.2013.130

Matsumoto, H., Hamawaki, T., Ota, H., Kimura, T., Goto, T., Sano, K., et al. (2000). 'Double-drugs'—A new class of prodrug form of an HIV protease inhibitor conjugated with a reverse transcriptase inhibitor by a spontaneously cleavable linker. *Bioorg. Med. Chem. Lett.* 10, 1227–1231. doi: 10.1016/S0960-894X(00)00202-X

Morphy, J. R., and Harris, C. J. (2012). Designing multi-target drugs. *R. Soc. Chem.* doi: 10.1039/9781849734912

Morphy, R., Kay, C., and Rankovic, Z. (2004). From magic bullets to designed multiple ligands. *Drug Discov. Today* 9, 641–651. doi: 10.1016/S1359-6446(04)03163-0

Morphy, R., and Rankovic, Z. (2005). Designed multiple ligands. An emerging drug discovery paradigm. *J. Med. Chem.* 48, 6523–6543. doi: 10.1021/jm058225d

Morphy, R., and Rankovic, Z. (2006). The physicochemical challenges of designing multiple ligands. *J. Med. Chem.* 49, 4961–4970. doi: 10.1021/jm06 03015

Morphy, R., and Rankovic, Z. (2007). Fragments, network biology and designing multiple ligands. *Drug Discov. Today* 12, 156–160. doi: 10.1016/j.drudis.2006.12.006

Morphy, R., and Rankovic, Z. (2009). Designing multiple ligands-medicinal chemistry strategies and challenges. *Curr. Pharm. Des.* 15, 587–600. doi: 10.2174/138161209787315594

Morphy, R., and Rankovic, Z. (2010). "Design of multitarget ligands," in *Lead Generation Approaches in Drug Discovery*, eds Z. Rankovic and R. Morphy (Hoboken, NJ: John Wiley & Sons), 141–164.

Ramsay, R. R., Popovic-Nikolic, M. R., Nikolic, K., Uliassi, E., and Bolognesi, M. L. (2018). A perspective on multi-target drug discovery and design for complex diseases. *Clin. Transl. Med.* 7:3. doi: 10.1186/s40169-017-0181-2

Rastelli, G., Del Rio, A., Degliesposti, G., and Sgobba, M. (2010). Fast and accurate predictions of binding free energies using MM-PBSA and MM-GBSA. *J. Comput. Chem.* 31, 797–810. doi: 10.1002/jcc. 21372

Reddy, A. S., and Zhang, S. (2013). Polypharmacology: drug discovery for the future. *Expert Rev. Clin. Pharmacol.* 6, 41–47. doi: 10.1586/ecp.12.74

Ren, J., Milton, J., Weaver, K. L., Short, S. A., Stuart, D. I., and Stammers, D. K. (2000). Structural basis for the resilience of efavirenz (DMP-266) to drug resistance mutations in HIV-1 reverse transcriptase. *Structure* 8, 1089–1094. doi: 10.1016/S0969-2126(00)0 0513-X

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

Schrodinger, Llc, (2010). *The PyMOL Molecular Graphics System*, Version 1.3r1.

Shang, E., Yuan, Y., Chen, X., Liu, Y., Pei, J., and Lai, L. (2014). De novo design of multitarget ligands with an iterative fragment-growing strategy. *J. Chem. Inf. Model.* 54, 1235–1241. doi: 10.1021/ci500021v

Shen, Y., Altman, M. D., Ali, A., Nalam, M. N., Cao, H., Rana, T. M., et al. (2013). Testing the substrate-envelope hypothesis with designed pairs of compounds. *ACS Chem. Biol.* 8, 2433–2441. doi: 10.1021/cb400468c

Song, A., Yu, H., Wang, C., Zhu, X., Liu, K., and Ma, X. (2015). Novel dual small-molecule HIV inhibitors: scaffolds and discovery strategies. *Curr. Pharm. Des.* 21, 950–962. doi: 10.2174/13816128206661409290 95102

Stierand, K., and Rarey, M. (2010). PoseView–molecular interaction patterns at a glance. *J. Cheminform.* 2:P50. doi: 10.1186/1758-2946-2-S1-P50

Sun, X., Fan, N., Xu, W., Sun, Y., Xie, X., Guo, Y., et al. (2016). Design, synthesis and biological evaluation of caffeoyl benzanilides as dual inhibitors of HIV integrase and CCR5. *MedChemComm* 7, 2028–2032. doi: 10.1039/C6MD00311G

Von Hoff, D. D., Goldstein, D., and Renschler, M. F. (2014). Albumin-bound paclitaxel plus gemcitabine in pancreatic cancer. *N. Engl. J. Med.* 370, 479–480. doi: 10.1056/NEJMc13 14761

Wang, T., Wu, M.-B., Chen, Z.-J., Chen, H., Lin, J.-P., and Yang, L.-R. (2015). Fragment-based drug discovery and molecular docking in drug design. *Curr. Pharm. Biotechnol.* 16, 11–25. doi: 10.2174/13892010156661411222 04532

Wei, D., Jiang, X., Zhou, L., Chen, J., Chen, Z., He, C., et al. (2008). Discovery of multitarget inhibitors by combining molecular docking with common pharmacophore matching. *J. Med. Chem.* 51, 7882–7888. doi: 10.1021/jm80 10096

Whitley, D. (1994). A genetic algorithm tutorial. *Stat. Comput.* 4, 65–85. doi: 10.1007/BF00175354

Wu, Y., He, C., Gao, Y., He, S., Liu, Y., and Lai, L. (2012). Dynamic modeling of human 5-lipoxygenase–inhibitor interactions helps to discover novel inhibitors. *J. Med. Chem.* 55, 2597–2605. doi: 10.1021/jm20 1497k

Xu, Y., Wang, S., Hu, Q., Gao, S., Ma, X., Zhang, W., et al. (2018). CavityPlus: a web server for protein cavity detection with pharmacophore modelling, allosteric site identification and covalent ligand binding ability prediction. *Nucleic Acids Res.* 46, W374–W379. doi: 10.1093/nar/gky380

Yuan, Y., Pei, J., and Lai, L. (2011). LigBuilder 2: a practical de novo drug design approach. *J. Chem. Inf. Model.* 51, 1083–1091. doi: 10.1021/ci1 00350u

Yuan, Y., Pei, J., and Lai, L. (2013). Binding site detection and druggability prediction of protein targets for structure-based drug design. *Curr. Pharm. Des.* 19, 2326–2333. doi: 10.2174/13816128113191 20019

Zhan, P., and Liu, X. Y. (2009). Designed multiple ligands: an emerging Anti-HIV drug discovery paradigm. *Curr. Pharm. Des.* 15, 1893–1917. doi: 10.2174/138161209788453266

Zhang, W., Pei, J., and Lai, L. (2017). Computational multitarget drug design. *J. Chem. Inf. Model.* 57, 403–412. doi: 10.1021/acs.jcim.6b00491

Zhang, W., Yuan, Y., Pei, J., and Lai, L. (2015). "CAVITY: mapping the druggable binding site," in *Part of the Series Methods in Pharmacology and Toxicology*, ed W. Zhang (New York, NY: Humana Press), 111–132.

Zimmermann, G. R., Lehar, J., and Keith, C. T. (2007). Multi-target therapeutics: when the whole is greater than the sum of the parts. *Drug Discov. Today* 12, 34–42. doi: 10.1016/j.drudis.2006.11.008