# Dynamic combinatorial and protein-templated click chemistry in medicinal chemistry

Mondal, Milon

# Conclusions and Perspectives

*This chapter summarizes the key findings of the research presented in this thesis, followed by future perspectives for the broader field of medicinal chemistry and, in particular, the early stages of the drug-discovery process.*

## 7.1 Introduction

Aspartic proteases are a class of enzymes widely found in fungi, plants, vertebrates as well as HIV retro-viruses. These enzymes play a causative role in several diseases such as hypertension, amyloid disease, malaria, fungal infections and AIDS. In HIV, the aspartic protease has an essential role in maturation of the HIV-virus, making it a validated target for the treatment of AIDS. In eukaryotes, the aspartic protease renin has a role in hypertensive action, cathepsin D in tumorigenesis, and pepsin in the hydrolysis of acid-denatured proteins. Therefore, the enzymes of this class of aspartic proteases are considered as a rich source of therapeutic targets.

We have highlighted throughout this thesis that fragment-based drug design (FBDD) and structure-based drug design (SBDD) still constitute a number of challenges such as the risk associated with *de novo* SBDD and are time-consuming as they involve synthesis and validation of the binding mode of each derivative in the fragment/hit-optimization cycle. To overcome these hurdles, we combined FBDD or *de novo* SBDD projects with dynamic combinatorial chemistry (DCC) or protein-templated click chemistry (PTCC) to render the identification/optimization of hits/leads more efficient, using the aspartic protease endothiapepsin as a model system.

The main achievements described in this thesis are: 1) the development of a powerful technique that combines *de novo* SBDD and DCC for the rapid identification of novel hits, 2) the development of an efficient approach that combines fragment linking and DCC to accelerate hit-to-lead optimization, 3) optimization of an initial hit, 4) development of a technique that combines fragment growing and DCC for the rapid optimization of a fragment, and 5) the development of a method that combines fragment linking/optimization and PTCC to accelerate the hit-identification process using the aspartic protease endothiapepsin as a model enzyme.

Based on these accomplishments, it can be concluded that several combinations of computational and analytical techniques are synergistic and facilitate the identification/optimization of hits/leads for the aspartic protease endothiapepsin.

In this chapter, we present a brief overview of the key findings reported in this thesis as well as future prospects.

## 7.2 Research overview

In Chapter 1 of this thesis, we discussed the application of DCC in the discovery of binders of a range of protein targets. Since the first report of DCC applied to the discovery of binders for a protein, this elegant tool has been employed on a range of protein targets at various stages of medicinal-chemistry projects. A series of suitable, reversible reactions that are biocompatible have been established and the portfolio of analytical techniques is growing. Despite progress, in most cases, the libraries employed remain of moderate size. In Chapter 1, we discussed the most recent advances in the field of DCC applied to protein targets, paying particular attention to the experimental conditions and analytical methods chosen. We gave an overview of SBDD and FBDD and also discussed their limitations. We gave a brief introduction of a class of aspartic proteases and why it is considered as a rich source of drug targets that are notoriously difficult. Finally, we discussed the strategic combinations of computational studies and analytical techniques that accelerate the identification/optimization of hits/leads for the aspartic protease endothiapepsin.

In Chapter 2, we demonstrated for the first time that the combination of *de novo* SBDD and DCC is a powerful technique for the rapid identification of novel hits that inhibit the aspartic protease endothiapepsin. We exploited $^{1}$H-STD-NMR spectroscopy to identify the binders directly from the dynamic combinatorial library (DCL). Among the hits identified, the best ones exhibit $IC_{50}$ values in the low micromolar range. Subsequent cocrystal-structure determination confirmed our *in silico* prediction that either direct or water-mediated interactions with the catalytic dyad can be achieved. We reported the first example of acylhydrazone-based inhibitors of endothiapepsin and aspartic proteases in general.

In Chapter 3, we described that the synergistic combination of fragment linking and DCC is a powerful and efficient strategy to accelerate hit-to-lead optimization of the aspartic protease endothiapepsin. We chose two co-crystal structures of endothiapepsin with acylhydrazone-based hits as the starting point of fragment linking. We had previously identified these hits from an acylhydrazone-based DCL using the synergistic combination of *de novo* SBDD and DCC, reported in Chapter 2. We exploited LC-MS analysis to identify the best binders directly from the DCLs. The best binder exhibits an $IC_{50}$ value of 54 nM, representing a 240-fold improvement in potency compared to its parent hits. Subsequent co-crystallization studies validated our *in silico* prediction.

In Chapter 4, we designed a library of eight acylhydrazone-based inhibitors starting from a hit, which was reported in Chapter 2, by using SBDD, in particular, focusing on optimizing an amide–π interaction. These compounds inhibit the aspartic protease endothiapepsin with $IC_{50}$ values in the low micromolar region. The best compound displays an $IC_{50}$ value of 7.0 μM, which is two-fold more potent than the original hit. The increase in potency could be due to the strengthened amide–π interaction compared to the original hit, owing to the more strongly electron-withdrawing nature of the trifluoromethyl group as well as better lipophilic interactions. Furthermore, it has the potential to have increased metabolic stability compared to the original hit because of the presence of a trifluoromethyl instead of three methyl groups.

In Chapter 5, we demonstrated for the first time that the combination of fragment growing and DCC is a powerful technique for the rapid optimization of initial fragments for the aspartic protease endothiapepsin. Moreover, by using a fluorescence-based assay, we could directly screen the DCLs for active inhibitors. The advantages of this approach are that only very small amounts of protein are required compared with established analytical methods and that the protein only needs to be in the assay mixture for a short period of time, making this protocol ideal for precious and unstable proteins. Among the acylhydrazones identified, the most potent inhibitor displays an $IC_{50}$ value of 85 μM.

In Chapter 6, we demonstrated that the strategic combination of fragment linking/optimization and PTCC is an efficient and powerful method that accelerates the hit-identification process for the aspartic protease endothiapepsin. We exploited the highly sensitive UPLC-TOF-SIM method to identify the triazole binders templated by the protein. The best binder inhibits endothiapepsin with an $IC_{50}$ value of 43 μM. Due to the limited solubility of the triazoles identified, we were not able to obtain any crystals of any triazole in complex with endothiapepsin. We reported the first example of triazole-based inhibitors for endothiapepsin. The great advantage of this approach is that, a catalytic amount of protein is sufficient to initiate and accelerate triazole formation from a sufficiently large library. Whereas this is an important consideration for precious proteins, a disadvantage is that the protein needs to be stable at room temperature for prolonged periods of time, greatly limiting the number of compatible targets.

## 7.3 Perspectives

The concept of developing several strategic combinations of computational and analytical techniques for the identification/optimization of hits/leads using the aspartic protease endothiapepsin as a model enzyme holds the potential to greatly influence the initial stages of the drug-discovery process. In this thesis, we were able to show that the synergistic combinations work well for endothiapepsin. Now all our strategies need to be applied to real drug targets of this class of enzymes such as HIV-protease, plasmepsin etc. as well as unrelated target classes.

In Chapter 2, we developed an efficient strategy for the rapid identification of novel hits for endothiapepsin that makes *de novo* SBDD less risky and therefore more widely applicable. To analyze the DCLs, we used $^{1}$H-STD-NMR spectroscopy, which has the advantage that it can directly detect the binders in solution. The application of this analytical method, however, is limited to weak to strong binders ($K_d$ > 10 nM). Therefore, $^{1}$H STD NMR spectroscopy is ideally suited for hit identification. Hit optimization tshould rely on a different analytical technique such as LC-MS or a fluorescence-based assay, which we have employed in Chapters 3 and 5, respectively. $^{1}$H-STD-NMR spectroscopy is also limited to small libraris. In case of larger libraries, it would be difficult to analyze the individual components from the library due to signal overlap in the STD-NMR spectra. $^{19}$F-STD-NMR spectroscopy should solve this problem. Although acylhydrazones are attractive for protein-templated DCC and our best hits exhibit $IC_{50}$ values in the low micromolar range, synthesis of bioisosteres might be required to develop a lead- or drug-like molecule. Acylhydrazones and the hydrazide building blocks represent potential toxicity issues.

In Chapter 3, we demonstrated a powerful approach for the rapid optimization of the initial hits to a lead using the combination of fragment linking and DCC. We also validated the binding mode of the lead by X-ray crystallography. Although our best lead exhibits an $IC_{50}$ value of 54 nM, with a 240-fold improvement in potency compared to its parent hits, we do not know the exact nature and bound conformation of the lead compound as the X-ray co-crystallization studies were done using a mixture of three diastereoisomers and the X-ray crystal structure was not fully resolved. Synthesis of all three diastereoisomers for biochemical evaluation and co-crystallization experiments of the best inhibitor with endothiapepsin are now required. Bioisosteres of the best lead should also be synthesized and tested against endothiapepsin.

In Chapter 4, we designed eight acylhydrazone-based inhibitors using SBDD starting from an initial hit reported in Chapter 2. In the optimization process, we modified only the aldehydic part of the acylhydrazone, which provided only two-fold increase in potency. The other moiety (hydrazide part) of the acylhydrazone should also be optimized using SBDD to get a lead-like compound, which might also require the growing of the compound. Subsequently, the acylhydrazone linker should be replaced with a suitable bioisostere to afford a drug-like compound.

In Chapter 5, we developed a strategic approach for the rapid optimization of initial fragment to hit for endothiapepsin. Although this technique accelerates the optimization of an initial fragment to a hit, the best compound shows an $IC_{50}$ value the upper double-digit micromolar range. Therefore, the best approach to optimize an initial fragment into a hit would be to grow the initial fragment and apply DCC to identify a hit with an $IC_{50}$ value in the upper micromolar range, followed by SBDD and DCC to get an improved hit with an $IC_{50}$ value in the lower micromolar range. Moreover, the binding mode of the hit identified should also be validated by X-ray crystallography.

In Chapter 6, we developed a method that combines fragment linking/optimization and PTCC to identify a triazole-based hit of endothiapepsin. Although our approach accelerates the hit-identification process, the best hit only displays an $IC_{50}$ value of 43 μM. In a second stage, a new library should be designed based on the best triazole hit identified, exploiting a combination of SBDD and PTCC to identify an optimized hit in lower micromolar to sub-nanomolar range. Due to the limited solubility of the triazoles identified, we could not obtain any crystals of any triazole in complex with endothiapepsin. So a second generation of triazoles should be designed containing more hydrophilic groups to increase their solubility in order to be able to validate the binding mode using protein X-ray crystallography. This approach would also represent a novel application of DCC, namely for the optimization of physicochemical properties rather than potency.

All hits/leads identified in this thesis should be tested against the real drug targets of the class of aspartic proteases such as HIV protease, plasmepsin, etc. Based on the literature, the structures of active compounds should be translated in a straightforward manner into inhibitors of the drug targets. Bioisosteres of the best hits/leads featuring the reversible acylhydrazone linker should also be synthesized and tested against these enzymes, affording more drug-like

compounds for further development.