Target-based drug discovery: from protein structure to small-molecules by MCR chemistry
Wang, Yuanze

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2018

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Chapter 1

General Introduction and Scope of this Thesis
1. Phenotypic-based Drug Discovery

Drug discovery can generally be viewed as a challenging process that includes the identification of active substances with a desirable therapeutic effect on certain diseases and optimization of those substances to increase their efficacy, pharmacokinetics and safety. Historically, phenotypic-based screening strategies, also known as classical pharmacology or forward pharmacology, were the main drug discovery paradigms used in both academic research centers and the pharmaceutical industry (Figure 1.1). The phenotypic approach could be specifically defined as the screening of a large number of randomly selected molecules in a system-based and target-agnostic approach that monitors the desirable change in phenotype. Thus, the active ingredient was obtained without any knowledge of the biological target. More often, an effort was made to identify the target after the active substance was identified. 1-3

![Figure 1.1 Evolution of drug screening and lead discovery (adapted from reference 1).](image)

2. Target-based Drug Discovery

Since the late 1980s, technological advances in molecular biology and sequencing of the human genome, which allowed rapid production of large quantities of biological targets initiated a new era of molecular target-based screening approach for modern drug discovery.

Target-based or hypothesis-based drug discovery is an approach in which small molecules are synthesized to target a known pathological/physiological pathway, which is hypothesized to be involved in the treatment of a particular disease. This idea came from the conclusion that a drug has to interact with biological macromolecules of the human body (enzymes, receptors, channels, etc.). Thus, biological macromolecules are used by the scientist in the lead discovery
screening, rational design and optimization. After that, the selected lead compounds were tested in cell and animal experiments to prove their efficacy. It was believed that this approach would result in an increased productivity and improved efficiency toward development of novel treatments for a validated target as it avoids the mass screening of banks of stored compounds. Moreover, according to a recent analysis, 78 of 113 first-in-class drugs approved by FDA from 1999 to 2013 were discovered through the target-based approach.\(^4\)

2.1 High Throughput Screening

The drug-discovery process usually involves high throughput screening (HTS), wherein a large number of compound libraries are tested for identifying a few ‘active hits’ to modify the target. These ‘hits’ provide starting points for drug design and are then followed up to undergo several iterative screening runs to evaluate their properties, including off-target toxicity. The ‘validated hits’ are then used as ‘lead compounds’ for further studies. HTS has become one of the dominant paradigms for lead identification and has gained widespread popularity among both the academic researchers and industrial scientists over the last three decades.\(^5\) The main goal of the HTS technique is to accelerate the drug discovery process and reduce the costs of drug development. Meanwhile, it is of vital importance that HTS also takes advantage of the changes in chemical synthesis strategy. It is obvious that combinatorial chemistry and parallel synthesis enable the efficient generation of a vast number of novel compounds.

With the increasing popularity of HTS, the technology required for HTS has also evolved rapidly. Using robotics, liquid handling devices, data processing software and high content imaging, an Ultra High-Throughput Screening process allows the researcher to conduct 100,000 assays per day. Initially, the libraries were arrayed into 96 micro-well plates, but now these have been replaced by high-density microplates with up to 3456 wells per plate.

The quality of the compounds in the HTS library can be assessed by several parameters. One of the most important parameters is the Lipinski’s Rule of Five: the cLogP not greater than 5, molecular mass less than 500 Daltons, no more than 5 hydrogen bond donors and no more than 10 hydrogen bond acceptors. Some other parameters, such as lipophilicity and ligand efficiency are also used to assess the drug-likeness of the compounds.\(^6\)
2.2 Fragment based drug discovery

HTS has considerably expanded the number of high quality lead molecules that could be evaluated for many drug discovery programs. However, the drug development process still suffered from high attrition rates as optimization of the screening hits could be problematic due to large size and inappropriate physical properties. Additionally, hit rates were generally low when some intractable biological targets were screened. In the context of the great demand to find small molecule drugs more efficiently, Fragment based drug discovery (FBDD) has emerged as a complementary strategy for our drug discovery arsenal.\textsuperscript{7-11} As a new promising approach, FBDD has several advantages. In FBDD, each atom of the fragment is involved in the desired binding interactions with the target. In other words, although the binding affinity of the fragment is much weaker than the HTS hit, it actually forms interactions with higher quality. Thus, the optimization of the fragment into a higher affinity lead is much easier than optimizing the HTS hit which contains some mismatched groups. In addition, the size and other physical properties such as lipophilicity and polar surface area can be better controlled.

Obviously, the successful design of a fragment library is the basis of the fragment-based approach.\textsuperscript{12} The library should be designed on the basis of several considerations such as sample relevant chemical space, appropriate complexity, synthetic tractability, fragment hit rate and knowledge of binding target. More importantly, fragments in the library typically obey a rule of three (RO3): molecular mass \(< 300\text{ Da}\), the number of hydrogen bond donors \(< 3\), the number of hydrogen bond acceptors \(< 3\) and cLogP \(\leq 3\). Another consideration is the number of the fragments included in the library, normally on the order of \(10^3\), because these fragments cover better chemical space than the compounds included in the HTS libraries.

As the fragment normally forms relatively weak interaction with the target macromolecules (100 \(\mu\text{M}\) to 10 \(\text{mM}\) range), sensitive biophysical techniques need to be used. According to a report by S. D. Pickett and coworkers, surface plasmon resonance (SPR), NMR spectroscopy and fluorescence-based thermal shift assay are the most popular techniques used for fragment-based drug discovery, followed by X-ray crystallography, mass spectrometry (MS) and isothermal titration calorimetry (ITC). Other useful techniques include capillary electrophoresis (CE), biolayer interferometry (BLI) and microscale thermophoresis (MST). In our lab,
the fragment screening strategy involves a three-stage cascade of biophysical screens: 1) preliminary screening using medium throughput technique: differential scanning fluorimetry (DSF), 2) validation of the binding of fragment hits with target protein by ITC or MST, 3) characterization of the fragment-target interaction by X-ray crystallography (Figure 1.2).

Figure 1.2 Three-stage cascade of biophysical screens for fragment based drug discovery.

To improve the potency of the validated fragment hits, three main approaches have been successfully established with the knowledge of the fragment binding mode: fragment linking, fragment merging and fragment growing. Fragment linking requires covalently joining two or more fragments that are known to bind in different but close pockets to form a novel-scaffold ligand on the same target. Fragment merging involves the incorporation of the structural features of other overlapping ligand or substrates into the fragment. Fragment growing refers to the modification of the fragment by chemical synthesis to incorporate more interactions. More frequently, fragment growing was used as the preferred fragment-optimization strategy, as fragment linking and merging are more challenging.
3. Proteins as drug target

Target-based drug discovery normally starts with the identification and validation of a drug target which can be used for lead screening. It is very important to have enough evidence to support the drug target selected before it is used for the drug discovery. A good drug target includes three core factors: 1) it participates in a crucial biological process linked to diseases; 2) it contains binding sites which can interact with drug-like molecules; 3) its structure and function has been clearly characterized and is different with known targets. The vast majority of the approved drugs use proteins as their targets. In a recent survey, the number of protein targets for all classes of available marketed drugs was found to be 324, including both human proteins and proteins from pathogenic organisms. In terms of potential drug targets, it was estimated that there are 2000-3000 druggable proteins in humans.

3.1 Protein expression and purification

The structural biology studies, such as the ligand-protein binding assay and the X-ray crystallization experiment, need large amount of proteins. In general, for the production of recombinant protein, *Escherichia coli* (*E. coli*) is used as the preferred host because it has several advantages: 1) simple and convenient; 2) rapid and cheap; 3) labeling methods for NMR are well established; 4) a wide variety of commercial products are available. The protein expression starts with introducing the plasmid that encodes the protein of interest into the *E. coli* cell, then the expression is induced and the presence of the protein is finally checked by SDS-PAGE analysis. It is important to establish the optimal growth and expression conditions with small-scale cultures before the protein is produced in large-scale. Most recombinant proteins can be expressed at high levels in *E. coli* when the host strain and vectors are carefully chosen, and the culture conditions are properly controlled.

If the produced protein is stable and soluble, the purification can be easily achieved by the affinity tag column, ion-exchange chromatography and size exclusion chromatography. However, many polypeptide gene products are expressed in an insoluble form that lack functional activity. This happens especially when the protein is expressed at high levels, leading to aggregation and formation of inclusion bodies. The formation of inclusion bodies can also be influenced by the vector and host cell selected, the nature of the protein, as well as the growth and induction conditions. The inclusion bodies need to be purified
under denaturing conditions instead of standard purification methods, which are based on the protein’s native solubility. It is usually necessary to renature the inclusion bodies to reconstruct its three dimensional structure before it is used for structural studies. In this case, the refolding experiments need to be designed empirically for each individual protein.15-18

3.2 Protein crystallization

The understanding of the molecular mechanism of the protein function by its three-dimensional structure is a very important tool for the modern biotechnological research.19 Protein crystallization is one of the most powerful ways for structure determination, which have significant impact on the rational drug design. Therefore, the field of protein crystallography has undergone an enormous expansion in recent decades, which was indicated by the rapid growth of protein structures deposited in the Protein Data Bank (PDB).

To crystallize a protein, we first need to purify large quantities of proteins (5-50 mg) with high purity and homogeneity. There are several methods to check the quality of the protein. The CD spectroscopy can help us to detect if the protein is correctly folded. The presence of aggregates could be determined by techniques such as dynamic light scattering (DLS). These techniques are crucial because sometimes the unfolded protein might also exist in a soluble form. In addition, additives like salts, metals and co-factors are sometimes necessary for the protein stabilization, which can be identified by the established DSF buffer screening methods. It is vital to keep in mind that protein quality is one of the key factors for the success of the crystallization.

Even when soluble protein with good quality is available, finding crystallization conditions for a new protein is still sometimes like searching for a needle in a haystack, because it is a complex and multi-parametric process. The most basic factors for crystallization include protein concentration, the presence of salt, precipitate and the pH of the buffer. Among these factors, the role of the precipitate is to bind with water molecules in order to bring the solution into supersaturation. Ammonium sulfate and polyethylene glycol are two commonly used precipitates.

From the crystallization phase diagram, we can see that there are four areas distinguished (Figure 1.3). The precipitation zone, where the protein is super saturated and precipitates; the nucleation zone is less supersaturated, where
nucleation will not take place but it is good for crystals to grow; the metastable zone is the best area, where large and well-ordered crystals might form and the under-saturated zone is an area, where the protein will never crystallize, because the concentration is too low. The process of the protein crystallization proceeds in two steps: the formation of nuclei and crystal growth. In an ideal experiment, once a critical nucleus has formed, the system might go into the metastable zone naturally as the protein concentration drops. If this is not the case, there are several methods like (i) microbatch, (ii) vapor diffusion, (iii) dialysis and (iv) free interface diffusion which can help to reach the nucleation zone and the metastable zone.

4. Small molecules synthesized by MCR chemistry

4.1 Ugi reaction

Multicomponent reactions (MCRs) are generally defined as a one-pot process during which three or more starting materials react in a single chemical step to form a product that incorporates all or most of the atoms of the reactants. Through the years, MCRs have attracted considerable interest in organic chemistry due to their convergence, atom economy, efficiency, shortened reaction time and
simplicity. Meanwhile, MCRs have also met renewed interest for medicinal chemists, because they are also powerful synthetic strategies for the construction of biologically interesting scaffolds with huge chemical diversity and molecular complexity. Among MCRs, the Ugi four-component reaction (Ugi-4CR) is one of the most widely used reactions which based on the peculiar reactivity of isocyanides. The Ugi-4CR involves a condensation of a carbonyl component (aldehyde or ketone), an amine, a carboxylic acid and an isonitrile to afford the peptide-like α-acylaminoamides with a newly created stereogenic center.

4.2 Mechanism of Ugi reaction

After the pioneer work of Ivar Ugi on the discovery of the Ugi reaction in 1959, the first mechanistic proposal was postulated by Ugi himself. Almost all the synthetic developments around this reaction rely on this mechanistic assumption as depicted in Scheme 1.1, path A. The initial step of the reaction is the formation of an imine 3 by condensation of the amine 1 with the oxo-component (aldehydes/ketones) 2 with loss of one equivalent of water. The imine 3 is then activated by proton exchanging with carboxylic acid 4. Subsequent nucleophilic addition of the isocyanide 6 with its terminal carbon to the iminium ion 5 forms intermediate nitrilium ion 7. This intermediate then undergoes a second nucleophilic addition by the carboxylate anion gives intermediate imidate 9. The last step is an irreversible Mumm rearrangement with transfer of the acyl imidate to generate the final product 10. In this mechanism, all the reaction sequence are in equilibrium except for the last step, which is considered as the driving force for the total reaction sequence.

Scheme 1.1 Two possible mechanisms for the Ugi reaction
Recently, an alternative mechanism raised the debate by proposing the formation of hemiaminal as a key intermediate (See Scheme 1.1, path B). The first step of this pathway is still the formation of imine which is activated by acid component. However, instead of isocyanide addition, the carboxylate anion first attacks the iminium ion to form intermediate hemiaminal 8, which is followed by isocyanide insertion to furnish the same imidate intermediate 9.

Scheme 1.2 The Ugi reaction started form electrophilic iminium species
(Structure adapted form reference 21-24)

The formation of the imine as the first step was generally admitted in both mechanisms. This was also demonstrated by the direct use of electrophilic iminium species in the Ugi reaction (Scheme 1.2). In 1982, Nutt and co-workers first reported the use of substituted 1-pyrrolines instead of the amine and aldehyde components to produce substituted prolyl peptides by an Ugi-type three-component reaction (U-3CR). In 2004, N-acylazinium salt formation in situ was reported by Lavilla and co-workers as a new source of iminium ion equivalents. In 2007, the use of the 3,4-dihydroisoquinoline dehydrogenated by in situ oxidation of the corresponding secondary amine tetrahydroisoquinoline in U-3CR was explored by Zhu and co-workers. Recently, Orru’s group reported stereoselective synthesis of highly functionalized, optically pure 3,4-substituted prolyl peptides by Ugi reaction starting from optically active 1-pyrrolines.
Scheme 1.3 Imidate intermediate isolated from Ugi reaction
(Structure adapted form reference 25-27)

In fact, the existence of imidate intermediate was also proved by several studies (Scheme 1.3). The first successful isolation of the imidate 11 and 12 has been reported by Ugi in 1971. In a research to explore thiols in the Ugi–Smiles reaction, Barthelon and co-workers surprisingly obtained thioimidate 13 in good yields when using methyl mercaptosalicylate as acid component. In 2009, Faggi and co-workers isolated a stable imidate 14 in the tautomeric enediamine form, which was further demonstrated by X-ray crystallography.

Figure 1.4 Energy profile of intermediate in Ugi reaction
(Structure adapted form reference 28)

In 2012, Chéron and co-workers reported the first theoretical calculation of Ugi reaction, which was performed at the M06-2X/6-31+G (d, p) level of theory
including ZPE corrections, based on Ugi’s proposal presented in Scheme 1.1, path A. Both Ugi-Mumm and Ugi-smiles reaction were studied with a realistic model and the calculation was computed separately in two solvents: methanol and toluene. According to the energy profile, optimization of the transition state (TS) for the insertion of the isocyanide in the hemiaminal leads to the TS of subsequent isocyanide addition to the iminium. Thus, path B proceeds through hemiaminal first fragmentation into the iminium and carboxylate anion and then the isocyanide addition. This was also confirmed by the intrinsic reaction coordinate (IRC) approach, which is a valuable tool to check if the given transition state is the expected transition state for the reaction of interest. Therefore, path A in Scheme 1.1 was considered as the privileged mechanistic pathway for the Ugi reaction. More importantly, two commonly accepted features in the Ugi mechanistic pathway were challenged by this theoretical research. Firstly, although the formation of imine as the first step is obviously confirmed, the involvement of iminium during isocyanide insertion was questioned by their calculation. Instead of activation of imine by a proton transfer, their computed results suggested that the formation of a hydrogen bonded complex between imine and acid substrate is more favorable. This conclusion was also in accordance with a recent NMR experimental study in imine activation reported by Fleischmann and co-workers. Secondly, according to the energy profile, the imidate solvated by a methanol dimer 9a lies at -33.6 kcal mol\(^{-1}\), whereas the highest TS 15 for Mumm rearrangement lies at -32.5 kcal mol\(^{-1}\), which means this final rearrangement can evolve easily with a barrier of only 0.9 kcal mol\(^{-1}\) (Figure 1.4). On the contrary, in the isocyanide addition step, it requires 19.8 kcal mol\(^{-1}\) of activation energy to reach a stable nitrilium-acetate ion pair. Therefore, they proposed that Ugi reaction should not be considered as an equilibratory reaction sequence driven by a final irreversible step as previously stated. Instead, the isocyanide addition step where a new stereogenic center was formed was the only rate-determining step.

In 2014, the mechanism of the Ugi reaction was investigated by Eberlin and co-workers using electrospray ionization mass spectrometry (ESI-MS) with alternatively two imidazolium charge-tagged reagents (a carboxylic acid or an amine). The Ugi’s original mechanistic proposal (Scheme1.1, path A) was consolidated as the key intermediate 7 has been isolated and characterized. In addition, the energetics of the final Mumm rearrangement were calculated by Density Functional Theory (DFT) studies. It predicted a very low energy barrier from transient imidate 9 to the final product 10, which is consistent with both ESI-
MS/MS and TWIM-MS data. In the same year, Angelis and co-workers developed finely selected reaction conditions for ESI-MS characterization of Ugi reaction mechanism which can avoid the impact of ion tagged reactants on the reaction pathway. Remarkably, their data demonstrated that the formation of nitrilium ion 7 is kinetically favored and its formation is the rate-determining step. This conclusion not only strongly supported the original hypothesis of Ugi, but also agreed with the theoretical findings predicted by Fleurat-Lessard and co-workers.\(^{30}\)

### 4.3 Ugi reaction and its post-cyclizations

Although Ugi reaction stands out as a powerful method for the construction of compounds with great diversity, the backbone of these compounds are normally linear which thus lack the conformational constriction. In this context, the combination of Ugi reaction with a subsequent secondary transformation, typically a ring-forming process, has been proven to be an extremely powerful strategy for the synthesis of structurally diverse complex molecules, especially highly functionalized heterocyclic compounds. Inevitably, a large variety of reactions has been introduced for the post-Ugi transformations strategy such as acid/base-catalyzed cyclizations, cycloadditions, condensations, Ugi-deprotection-cyclization (UDC), S\(_{\text{N}}\)Ar reactions, macrolactonizations, S\(_{\text{N}}\)2 reactions, aryl couplings, ring closing metathesis, radical cyclization \emph{etc.} Therefore, all kinds of functionalized skeletons have been successfully constructed by this two-step procedure, as concluded in Scheme 1.4.
Scheme 1.4 Functionalized skeletons obtained by Ugi reaction and its post-cyclization. (Structure adapted form reference 31-81)
5. Aim and scope of this thesis

Refolding of proteins derived from inclusion bodies is very promising as it can provide a reliable source of target proteins of high purity. However, inclusion body-based protein production is often limited by the lack of techniques for the detection of correctly refolded protein. Thus, the selection of the refolding conditions is mostly achieved using trial and error approaches and is thus a time-consuming process. In chapter 2, we use the latest developments in the differential scanning fluorimetry guided refolding approach as an analytical method to detect correctly refolded protein. We describe a systematic buffer screen that contains a 96-well primary pH-refolding screen in conjunction with a secondary additive screen. Our research demonstrates that this approach could be applied for determining refolding conditions for several proteins. In addition, it revealed which “helper” molecules, such as arginine and additives are essential.

In chapter 3, we describe the fast and efficient synthesis of libraries of positional isomeric 1H-tetrazoles and 5H-tetrazoles for the purpose of testing binding hypothesis of isomeric tetrazoles in fragment-based drug discovery.

Isocyanide-based multicomponent reactions (IMCR) are by far the most versatile reactions that can construct relatively complex molecules by one-pot synthesis. More importantly, the development of post IMCR modifications significantly improves the scaffold’s diversity. In chapter 4, we describe the use of N-Boc protected hydrazine together with α-amino acid derived isocyanides in the Ugi-tetrazole reaction and its post-cyclization under both acidic and basic conditions. The cyclization in acidic conditions was conducted in a one pot fashion, which give 7-aminotetrazoloypyrazinone and tetrazolotriazepinone cyclic products. The post cyclization under basic conditions could selectively afford Boc-protected 7-aminotetrazoloypyrazinone products in yield from 38 to 87%.

In chapter 5, the Pomeranz-Fritsch reaction was for the first time successfully applied in the Ugi post-cyclization strategy by using aminoacetaldehyde diethyl acetal and electron rich aldehydes as starting materials. Isoquinoline derivatives and benzo[d]azepinone scaffolds with great diversity were constructed in moderate to good yield. In addition, the isoquinoline-tetrazoles and an alkaloid-like tetrazole-fused tetracyclic compound were synthesized by this method in a very efficient manner.

Encouraged by the results from chapter 5, the Schlittler-Müller modification was successfully used in the Ugi post-cyclization strategy in chapter 6. The acoustic
dispensing-enabled scouting enabled a pipeline of fast and efficient nL scale scouting to mg to gram scale synthesis. Isoquinoline derivatives and heterocyclic compounds were constructed by this method in moderate to excellent yield. These synthetic approaches were unprecedented, simple and efficient. Meanwhile, the substrate scope and functional group tolerance were proved to be exceptional.
References


