Abstract

Abundance of data on molecular underpinnings of diseases will, with the highest possible resolution, define the properties of effective future therapeutics. Small molecules can play a big role in the future of disease-treatment options if chemistry that is used to synthesize them can be reinvented with this goal in mind. Our ability to identify novel chemical probes and therapeutics is directly linked to the quality of compound screening collections. Many modern screening libraries are fundamentally inefficient because they contain compounds that exhibit redundant biological activity, or none at all. An ideal library, therefore, comprises small molecules chosen for the uniqueness not of their chemical structures, but of their biological performances. Populating biological performance space more broadly will necessitate venturing into previously unexplored chemical space. To this end, one possible research avenue is to focus attention on previously under-explored photochemical reactions. Photochemical transformations are ideal entry points to complex and unique chemical scaffolds, many of which are highly underrepresented in modern screening libraries and are difficult to access by conventional means. In addition, generating collections of chemically tagged small molecules remains a significant challenge, notwithstanding the benefits of such collections in terms of ease of handling, performing binding studies (by DNA sequencing, e.g., in the case of DNA-encoded libraries), etc. Main challenges are centered around the development of suitable aqueous chemistry that would allow achieving sufficient diversity of structures present in those collections without interfering with the structure of the tag. Finally, one of the biggest challenges in studying compound’s interactions within complex biological systems is identification of its molecular binding partners. Advances in mass spectrometry have allowed for unprecedented scope and resolution in these studies.

Contents

1 Rapid biological evaluation of structurally complex photochemically generated molecules 2
1.1 Hypothesis 2
1.2 Compound synthesis 2
1.3 Rapid biological fingerprinting 3
1.3.1 Cell painting annotation 4
1.3.2 Transcriptomic annotation 4
1.4 Computational evaluation 4
2 Library of DNA-encoded macrocycles; screening for protein binders 5
2.1 Hypothesis 5
1 Rapid Biological Evaluation of Structurally Complex Photochemically Generated Molecules

1.1 Hypothesis

Many disease-targeting initiatives have reinforced the concept that “new chemistry yields new therapeutic opportunities.” Building upon this idea, we identified photochemistry as an underutilized strategy that facilitates the rapid generation of structurally complex chemical scaffolds from simple precursors. Using the pipeline described herein, we will be able to evaluate the biological performance of the members of our photochemistry-derived library and assess the extent to which covering new chemical space increases the chances of discovering new biological performance space.

Compounds generated through photochemistry typically display greater three-dimensional character, which might allow them to interact with the three-dimensional surfaces of biological macromolecules with more selectivity and affinity. This project aims to test the hypothesis that photochemically generated, topographically complex molecules exhibit distinct biological performances from compounds generated by conventional means, and that libraries comprising these compounds would cover a larger span of biological space. The overall structure of the project (Figure 1) is divided into three parts: 1) compound collection synthesis, 2) biological annotation, and 3) analysis of the obtained compound profiles and comparison to previously profiled compound collections.

Figure 1: Overall structure of the proposal.

1.2 Compound synthesis

This aim details the synthesis of compounds to be profiled in aim 2 of the project. It relies on the use of previously underutilized photochemical transformations to access complex chemical scaffolds that can be further elaborated to furnish diverse compound collections. The goal of this aim is to identify and execute reaction sequences that, overall, would yield 120 structurally distinct compounds. Strategy. The synthetic approach consists of three stages (Figure 2). In stage 1, simple precursors will be used to access substrates that are flat, absorb ultraviolet (UV) light, and contain strategically positioned chemical groups to facilitate further elaboration. Stage 2 is the key complexity-generating step in which the substrates from the stage 1 are exposed to UV light, yielding molecules with polycyclic, bridged, three-
dimensional chemical scaffolds. Finally, in stage 3, these scaffolds are further elaborated by using reactions that exploit additional functionalities within the molecule. Figure 2 shows two examples of this approach. Example 1 is anchored around a photochemical rearrangement of pyrroles to vinyl aziridines. This powerful transformation yields a tricyclic bridged ring system poised for many further functionalization reactions (some of which are depicted in the final panel). Example 2 transforms simple alkyl pyridinium salts into fused bicyclic allylic amines with three stereocenters and a cis ring fusion. This scaffold can be elaborated in one or two steps toward the molecules shown in stage 3, among others. In a similar manner, we will subject four more classes of readily accessible substrates to photochemical conditions and subsequently elaborate them to afford twenty molecules per photochemical pathway. The chosen synthetic pathways will conform to the following criteria: 1) Reaction sequences will be between 3 and 5 steps long; 2) starting materials will be readily available; 3) the operative photochemical reaction will result in a massive increase in complexity and produce an underrepresented chemical scaffold; 4) there should be ample possibility for post-photochemical elaboration of these scaffolds. Potential setbacks are to be anticipated due to the uncertainties of venturing into unexplored chemical space. Although there is strong literature precedent for many of our proposed photochemical transformations, the particular assembly of functional groups that we envisioned might fall outside of the scope of the described reactions. Some of these difficulties will likely be addressable by reaction optimization. Particularly stubborn reaction sequences that fail the optimization attempts would be replaced by other transformations that satisfy the above-outlined criteria.

1.3 Rapid biological fingerprinting

To evaluate the biological performance of our library, we will create multidimensional profiles of our compounds as proxies for their biological activities in cells. To provide the resolution necessary to distinguish between all of the possible biological activities of our compounds, it is important that our profiles be of sufficient size and that constituent features are sufficiently orthogonal. With this requirement in mind, we will take advantage of two separate and orthogonal techniques: cell painting and L1000, which capture changes in cell morphology and gene expression, respectively. From
these signatures we will be able to determine if our compounds, first, have any measurable effect in cells, and second, have an activity that is unique.

1.3.1 Cell painting annotation

Cell painting is a single-cell imaging assay capable of generating profiles comprising hundreds of quantifiable cellular features. Compound-treated human cells (osteosarcoma U-2 OS cell line) are painted with a set of dyes that stain different subcellular compartments and structures (nuclei, mitochondria, cytoskeleton, etc.), imaged by a microscope, and analyzed via CellProfiler to quantify the morphological cellular features. These measured features will be compared to the features in vehicle-treated cells to determine the cellular response to compound treatment.

1.3.2 Transcriptomic annotation

L1000 is a transcriptomic experiment that measures the levels of 1,000 mRNA transcripts. These transcripts were specifically chosen such that they explain 80% of the variance seen in gene expression data, which raises throughput and lowers cost.

1.4 Computational evaluation

We will use the biological profiles generated through described annotations to evaluate the biological performance of the photochemically generated library by two metrics: diversity within the library (intra-library diversity) and distinctness compared to other libraries (inter-library diversity). Using the features measured in biological annotation experiment, we will calculate the correlation values (e.g. Pearson, Spearman) for each compound pair to quantify the similarities of their biological profiles. The matrix of compound-compound correlations represents the diversity of our collections biological performances. For example, a compound collection with high correlation values indicates low performance due to redundant biological activity. In a similar vein, a collection with low correlation values consists of compounds with distinct biological activities, and will have a high performance. We will first evaluate the uniqueness of biological activities within our photochemically generated library. This diversity metric can then be compared to the performance diversities within other compound collections that have previously been annotated in this manner; roughly 30,000 compounds have already been subjected to this type of annotation. Random samplings of equal number will be taken from each library to avoid any effects due to library size. This metric will give us a sense of the spread of the library within biological performance space. Our second set of analyses will evaluate whether our photochemically generated library covers a portion of biological performance space distinct from other libraries. Hypothetically, if two libraries cover distinct and finite areas of biological performance space, a collection comprising members from both libraries will have greater performance diversity than a collection comprising members from only one. Alternatively, if these two libraries are completely overlapping, there will be no gain in performance diversity when sampling from both. We will apply this model to our library to determine whether there is a resolvable separation in biological performance space between our library and others that have already been annotated. Again, we will take random samplings of equal size from each and both libraries being compared to account for library size effects.
Compounds in modern collections are arrayed as spatially separated instances of purified, neat substances or stock solutions at milligram scale. This arrangement requires tremendous investment to maintain, including dedicated personnel, automation, giant cold storage facilities, etc. Alternatively, the compound collections can be synthesized as a mixture in physical space, that can be deconvoluted by means of an appropriate chemical tag. This chemical tag can, within its structure, encode either only its identity, or the chemical steps that were performed to arrive at the molecule. Thanks to their ability to be amplified, double-stranded deoxynucleic acids are attractive as tags encoding diversity-oriented compound collections. Another possible tagging strategy is halocarbon tagging. Halocarbons can be used as more chemically inert tags of much smaller size by taking adventage of extremely high sensitivity of electron-capture detector for halogananated hydrocarbons, and relying on resolving powers of gas chromatograph. One major limitation of currently available “tagged” compound libraries is the chemistry used for their synthesis. One of the consequences of this scarcity of reactions that work in the presence of a tag is that these collections are often just “appendage-diverse,” i.e. the combinatorial power is attained only by introducing hundreds of different substituents at each of 4 or 5 different diversification sites, thus reaching the nominally high numbers of chemically distinct, albeit structurally highly similar, molecules. A major advance would be to expand the build-couple-pair paradigm to the tagged small-molecule mixtures, thus expanding the skeleton-diversity and three-dimensionality of molecules that populate these collections.

2.1 Hypothesis

Macrocycles have many attractive properties from the vantage point of binding to bio-macromolecules; conformational preorganization and flexibility being two important ones. Many natural, biologically active compounds contain a macrocycle within their structure (e.g. rapamycin, cyclosporine, bryostatin). The topographical properties of macrocycles allow them to bind large protein surfaces that lack “deep” binding pockets and in such a way interfere with important protein—protein interactions that dominate intracellular signaling. If we could synthesize a collection of DNA-tagged macrocycles using split-pool, and build-couple-pair strategies, we could probe the binding potential of these molecules on important protein targets that currently lack any small molecule perturbagens.

2.2 Synthesis

To make the DNA-encoded macrolactam library, we would first establish the method of diversity-oriented synthesis of macrolactams. Our synthetic strategy is to make a macrolactam from four building blocks through four-step ligation involving native chemical ligation, KAHA ligation, SuFEx reaction, and thiol-ene reaction (Figure 3). These reactions fulfill many criteria that are needed for this sequence to be successfully performed. They have all been shown to be compatible with aqueous conditions, they do not generate difficult-to-remove byproducts, and they are reliable and versatile in their scope.

2.2.1 Fragment synthesis and their structural diversity

Bifunctional diversity unit DU1 can be synthesized from cycloalkanones through the nitrosobenzene-mediated oxidative C-C bond cleavage of the β-keto ester in its enolate form in a one-pot procedure (Figure 4). Variety of avail-
able cycloalkanones allows introduction of stereocenters along the perimeter of target macrolactams. DU2 can be synthesized from the multi-substituted olefins through the Fukuyama hydroxylation of amine following Sharpless asymmetric aminohydroxylation of olefins (if the pool of commercial vicinal amino-alcohols needs to be expanded).

Figure 3: Overview of the diversity oriented synthesis of the DNA-encoded macrolactam library.

Figure 4: Synthesis of diversity units.

2.2.2 Ligation reactions

Upon synthesis of building blocks, and attaching the head piece to DNA, water- and DNA-compatible reactions that allow high yield and little purification are required. 1) Native chemical ligation is the most powerful ligation reaction method to connect two synthesized peptides/proteins without the use of any protecting groups. This ligation works through the reaction of the C-terminal thioester on one peptide with an N-terminal cysteine of another peptide in water. 2) Thiol-ene reaction is a “click”-like reaction between a thiol and an olefin to form an alkylsulfide. 3) The SuFEx reaction was identi-
fied as the next generation click chemistry reaction. In line with traditional click chemistry reactions, SuFEx is a simple, water- and oxygen-friendly reaction that allows obtaining the desired product in high yield and with little purification.

4) The α-ketoacid-hydroxylamine (KAHA) ligation is an amide bond-forming reaction. This reaction proceeds in aqueous solution without any coupling reagents or catalyst, with water and carbon dioxide as the only byproducts.

2.3 Binding studies

Having synthesized the collection of DNA-tagged macrolactams, the ability of its members to bind to biological targets of interests would be assessed. I am particularly interested in measuring binding of these molecules to important neurological disease-related protein targets such as Alzheimer’s apolipoprotein ApoE$_4$ or prion protein PrP. Ability to degrade these proteins if a protein binder is found (through one of degradation methodologies available today, i.e. degronomides) presents a unique therapeutic opportunity in these important areas. Specifically, biological targets would be incubated with the library of DNA-tagged molecules. Upon equilibration of bound ligands, the protein target with ligands bound would be “extracted” out of the mixture with a solid-supported antibody. DNA-tags from bound ligands would then be amplified (same sequence for PCR amplification present on all DNA tags), and the identity of the bound ligands would be revealed through sequencing of the amplified tags.

3 Cellular, Label-Free Identification of Small Molecule Protein Targets

Full potential of phenotypic screening can only be realized through the identification of the targets of active compounds (direct binders and functionally-affected proteins). This remains an unsolved analytical problem of significant impact. Many current proposed solutions rely on chemical modifications of “bait” molecules with bio-orthogonal handles. This approach, however, requires substantial investment in understanding structure—activity relationships, and in many cases does not guarantee successful outcome. One worthwhile approach is to investigate changes that a label-free compound induces in the relevant biological context. To achieve this goal, I am interested in studying and developing methods that reveal physical associations of the bioactive compounds with its molecular targets in cells or in vivo. A recently developed proteomics-based, label-free method for target investigation (MS-CETSA) is an attractive method developed in this regard (Figure 5).[10]

As a test-case for the use of MS-CETSA in target identification of a small molecule, I started investigating a benzoyl sulfonamide that is toxic to stroma-dependent multiple myeloma cells. This molecule has proven difficult to study via conventional chemical tagging approaches. Instead, cells treated with the chemically unmodified compound were exposed to gradually elevating temperatures, and the thermal stability of proteomically-measurable proteome was assessed. Proteins whose stability, and therefore amount changes (because thermally unstable proteins precipitate out of solution as aggregates) in compound-treated versus DMSO-treated samples are potential target proteins, or proteins downstream from the molecular target of the compound. If this strategy is successful, it could present a much needed departure from ad hoc target ID efforts towards a more systematic and unbiased method to solve one of the central problems in discovery of biologically-active small molecules.
REFERENCES


