DNA-Encoded Chemical Libraries

Isolation of Specific Binding Molecules

The isolation of specific binding molecules is a central problem in the drug discovery process. Libraries of organic molecules, covalently coupled to DNA tags that serve as amplifiable identification bar codes, represent a new tool for the efficient identification of ligands to target proteins of choice. The implementation of next generation high throughput sequencing now allows the identification of binding molecules from DNA-encoded libraries of unprecedented size.

Most drug development programs start with the isolation of a binding molecule to a pharmacologically relevant target protein. Technologies for the isolation of such binding molecules are therefore of central importance for drug discovery. In high throughput screening campaigns, large collections of chemical compounds are individually screened in miniaturized in vitro assays in parallel for binding to the target protein of interest, typically requiring the availability of an enzymatic assay or a ligand for a displacement assay. These discovery campaigns are sometimes unsuccessful, thus preventing drug development even though suitable target protein for pharmacological intervention may be available.

DNA-encoded chemical libraries represent a new avenue for the synthesis and screening of collections of chemical compounds of unprecedented size and quality [1-3]. In DNA-encoded chemical libraries, each chemical compound is covalently linked to a DNA fragment containing a specific sequence serving as identifier bar code (Fig. 1a). This linkage of each library member to a corresponding DNA tag allows the facile identification of binding molecules after affinity capture procedures on an immobilized target protein of choice (fig. 1b).

DNA-encoded chemical libraries bear certain similarities to phage display libraries of polypeptides, such as antibodies. In antibody phage display, recombinant antibody fragments are physically linked to phage particles that bear the gene coding for the attached antibody, which is equivalent to a physical linkage of a "phenotype" (the protein) and a "genotype" (the gene encoding for the protein).
Affinity capture procedures allow the enrichment of phage particles displaying antibodies with binding affinities to the target protein of interest, which can be identified by DNA sequencing [4].

The binding affinities of the antibodies that can be isolated to a given antigen are largely dependent on the number of diverse antibody mutants that are present in the phage library [5]. The larger the antibody library, the higher the chances to isolate potent binders.

In analogy, the challenge for the successful use of DNA-encoded chemical libraries in drug discovery programs lies in the judicious synthesis of good quality libraries of large size. We have applied different strategies for the construction of large DNA-encoded chemical libraries. One approach, based on the combinatorial self-assembly of smaller sub-libraries with distinct chemical groups at the 5' and 3' extremity of partially complementary strands, yields large "dual-pharmacophore" libraries. Alternatively, split-and-pool methods characterized by the introduction of a DNA code after each synthetic step allow the construction of large "single-pharmacophore" libraries (fig. 2).

Dual pharmacophore libraries of large size can be formed through the self-assembly of smaller sub-libraries, which consist of organic molecules conjugated to DNA oligonucleotides that contain both an identifier sequence and a constant hybridization domain [6]. If different sub-libraries have complementary hybridization domain sequences, individual members of these sub-libraries can form heterodimers. For example, the self-assembly (heterodimerization) of two sub-libraries containing 1,000 members would yield 1,000,000 different combinations, i.e. 1,000,000 different chemical entities. Obviously, similar procedures can be used for the affinity maturation of known lead binding compounds, which can be paired with libraries of binding molecules for the identification of synergistic
We applied this approach in order to improve the potency of the trypsin inhibitor benzamidine, which was conjugated to an oligonucleotide, and paired to a sub-library containing 700 members. The best inhibitors isolated showed IC50 value in the low nanomolar range, improving the potency by 2-3 orders of magnitude compared to the starting benzamidine inhibitor. The best inhibitor displayed selectivity among closely related serine proteases, exhibiting a 40- and 6,500-fold lower potency towards thrombin and factor Xa, respectively, compared with the inhibition of trypsin [7].

The screening of binding molecules from DNA-encoded chemical libraries does not require the availability of a biochemical assay. We isolated a class of portable binders to human albumin, the most abundant protein in human serum, with dissociation constants in the nanomolar range [8]. While many organic molecules exhibit some binding affinity to albumin, the quest for a portable albumin binder, which can be covalently conjugated to another molecule while still retaining its binding affinity to albumin, had so far only been successful in the case of albumin-binding peptides [9, 10]. These technologies have been devised to expose the body to adequate concentrations of the therapeutic agent for a sufficiently long period of time, thus improving efficacy and reducing the number of injections. The isolated small organic molecule albumin binders may replace albumin-binding peptides in many applications, and show superior clinical applicability because of their smaller size.

We conjugated our most potent albumin binder to fluorescein (product name Albufluor) with the aim to develop an innovative reagent for angiographic procedures. Fluorescein is widely used in ophthalmology as a contrast agent in fluorescein angiography of the eye fundus in patients with retinal disorders [8]. However, the short circulatory half-life of fluorescein requires injection of high doses (200-500 mg per patient) with non-negligible side effects [8] and impedes the comparative study of both eyes. In animal studies, Albufluor had a >100-fold longer circulatory half-life compared to Fluorescein, and allowed the visualization of smaller blood vessels which could not be detected when using Fluorescein as an imaging agent. To our knowledge, Albufluor is the first compound derived from a molecule isolated from DNA-encoded chemical libraries that entered clinical development.

While a suitable linker has to be found for the conjugation of synergistically acting compounds isolated from dual pharmacophore libraries, the use of identified pharmacophores.
molecules from single pharmacophore libraries is more straight-forward (fig. 2). The synthesis of single pharmacophore libraries requires the use of split-and-pool synthesis and alternating cycles of chemistry on the displayed organic molecule, and addition of blocks of identifier DNA sequences on the nascent oligonucleotide. With the advent of high throughput next generation sequencing [11], it is now possible to decode libraries of hundreds of thousands up to millions of compounds in parallel a single sequencing run. We have recently reported for the first time the implementation of next-generation high-throughput sequencing for the identification of compounds in DNA-encoded chemical libraries [12] (fig. 3).

The methodologies described in the PNAS article are compatible with the synthesis of very large (>106 compounds) DNA-encoded chemical libraries based on three building blocks. The implementation of next-generation sequencing for DNA-encoded chemical libraries allows the handling of chemical libraries of this size, whose construction is currently underway in our laboratories.

In contrast to conventional screening procedures such as high-throughput screening, biochemical assays are not required for binder identification, allowing the isolation of binders to a wide range of proteins that were historically difficult to tackle with conventional screening technologies. The availability of binders to such pharmacologically important, but so-far "undruggable" target proteins promises to facilitate the development of new generations of drugs for diseases that could not be treated thus far.

References
Authors

Prof. Dr. Dario Neri, ETH Zurich, Dr. Samu Melkko, Philochem, Dr. Luca Mannocci, Philochem

Contact

ETH Zürich
Wolfgang-Pauli-Str. 10 -16
8093 Zürich
Switzerland
Phone: +41 44 632 1111
Telefax: +41 44 632 1010