Strategies for Facilitated Forward Chemical **Genetics**

Sonya M. Khersonsky and Young-Tae Chang*[a]

Dedicated to Professor D. H. Kim on the occasion of his 70th birthday

Introduction

Completion of the Human Genome Project has provided the scientific community with a wealth of information to further our knowledge of the gene function. Currently, an enormous effort is being applied to studying the relationship between the mapped genes and human disease. In the past, this enormous endeavor has been mainly left to classical geneticists. However, since the gene products, proteins, are dy-

Figure 1. Forward chemical genetics approach a) Design and synthesis of small molecule library; b) screening for novel phenotypes; and c) target-protein isolation and identification.

namic and multifunctional, it has become clear that the study of genes alone may not be sufficient to see the whole picture. This has stimulated the field of chemical genetics at the interface between chemistry and biology. Chemical genetics, inspired by classical genetics, makes use of small molecules as mutation-inducing agents to study protein function. One approach of chemical genetics, forward chemical genetics, is gaining recognition as a powerful strategy.^[1-6] Forward chemical genetics is a three-step process (Figure 1). First, a collection of small molecules is designed and synthesized.^[7] The small molecules are then screened in a model organism for the ability to perturb/change a biological process. Once an interesting phenotype is found, the small molecule responsible for it is used to isolate the target protein(s). Finally, upon identification of the target protein, its identity can be linked with the previously observed phenotype, thus gaining better understanding of that particular biological process. In addition, biologically active small molecule can provide important structural information for further development of novel therapeutic agents.

Even though the concept of using small molecules to study cellular processes is not new, in practice, the forward chemical genetics approach is still not systematic. Consequently, systematic strategies at every step of the forward chemical genetic process would greatly accelerate the study of protein function and development of novel therapeutic agents. In this review, we will discuss the strategies and research tools that have been developed to accelerate and systematize the field of forward chemical genetics.

Library Design

The design of the library is the first and a very crucial step in the forward chemical genetics process; this step determines the success of the library. Libraries could be designed around a natural product scaffold or a known drug scaffold.^[8,9] A more traditional approach, used widely before the advent of combinatorial chemistry, is to use natural product libraries to study protein function.

Natural product library design

Natural products have undergone thousands to millions of years of evolution and natural selection. As a result, they should offer a rich source of biologically active compounds. However, in many cases, using natural products to study protein function can be very challenging. In order to construct a library, samples need to be iteratively extracted from a natural source and tested for biological activity until an active compound is found.^[5,10] Other limitations, such as low abundance and influence of other compounds in the crude mixture, may further compromise the activity. Cumbersome structure elucidation of large and complex natural products may further hamper the progress of the study. Even though these limitations do result in a labor-intensive chemical genetics process, this strategy should not be completely overlooked. According to the cover story of Chem. Eng. News' November 2003 issue, natural product drug discovery is on the verge of a comeback.[11] This is due to advances in separation technologies and more rapid and sensitive structure elucidation techniques. Natural products are irrefutably important for library design. Around 60% of the antitumor and anti-infective drugs either on the market or in the later stages of clinical trials are derived from natural products.^[12] One innovative method for accelerating natural product isolation and analysis was introduced by Eldridge et al.^[13] They developed an automated high-throughput method in which each plant sample was separated by parallel four-channel preparative HPLC and then analyzed by parallel eight-channel LC-ELSD-MS. As a demonstration, they used

[[]a] S. M. Khersonsky, Dr. Y.-T. Chang Department of Chemistry, New York University New York, NY 10003 (USA) Fax: (+1)212-260-7905 E-mail: yt.chang@nyu.edu

Taxus brevifolia extract to isolate 36 000 fractions containing detectable compounds that were then screened for biological activity.

Another approach undertaken by a collaboration of Analyti-Con Discovery and Aventis Pharma was to abandon the lengthy and laborious fractionation screens that require the use of crude plant extracts and instead to construct and screen a pure natural compound library.^[14] Using automated methods in combination with computer software to isolate and purify the compounds and elucidate their structures, within a period of 18 months, they amassed a collection of 4000 partially characterized, nonredundant natural compounds in quantities of \geq 5 mg per compound and \geq 80% purity.

Natural-product-like and known-drug-like libraries

The first goal for both natural-product-like and known-druglike library design is to select a scaffold. Here, several aspects have to be considered: the scaffold should have a good number of diversity points for diversity-oriented synthesis (DOS), it should allow for fairly undemanding synthetic manipulations, and it should be rigid to minimize the entropic cost of binding to proteins.

An example of successful scaffold selection was demonstrated by Schultz and co-workers. The design of their scaffold, purine, was based on a natural inhibitor of several cyclindependent kinases (CDKs), olomoucine, in search of a more potent inhibitor (Scheme 1 a).^[15, 16] Since the purine ring is present in numerous biological molecules, it was expected that the diverse purine libraries would yield an abundance of bioactive compounds when screened in a variety of assays. And indeed, not only a group of more potent CDK inhibitors was found by screening the trisubstituted purine library, $[15, 16]$ but also other biologically active compounds such as the microtubule inhibitor myoseverin;^[17,18] estrogen sulfotransferase (EST) inhibitors;^[19] carbohydrate sulfotransferase inhibitors;^[20] diminutol, microtubule dynamics regulator;[21] inositol 1,4,5-triphosphate-3-kinase (IP3K) inhibitors;[22] purmorphamine, with osteogenesis-inducing activity; $[23]$ and reversine that converts differentiated cells into progenitor cells^[24] (Scheme 1 b-i). Purine, therefore, is capable of interacting on a variety of cellular pathways, underscoring its versatility as a scaffold.

Nicolaou et al. constructed a combinatorial library based on a template of 2,2-dimethylbenzopyran, which is found in a number of natural products with diverse biological activities.^[25, 27] They synthesized a 10000-membered library as well as several smaller libraries using the split-and-pool method and IRORI NanoKan optical encoding system (Scheme 2 a). They later used copies of the synthesized library and extended the diversity and library size by introducing additional diversity sites through solution-phase chemistry (Scheme 2b).^[27] A highthroughput antibacterial screen of one of these libraries yielded three benzopyran-derived cyanostilbenes 1-3 that were active against several Methicillin-resistant Staphylococcus aureus (MRSA) strains.[28] A more focused and thorough structure-activity relationship (SAR) study of compound 1 provided information on the structural features necessary for activity and uncovered several potent compounds.^[29] In another study, they screened the 10 000-membered library in search of farnesoid X receptor (FXR) activators (transcriptional sensor for bile acid, the primary product of cholesterol metabolism) in a cellbased reported assay.^[29] As a result of this screen, they found a number of active compounds that were further optimized to result in a set of potent FXR agonists.

Another strategy to design a natural-product-like library was described by Schreiber and colleagues.[3] Instead of using a natural product template per se, they used a natural product, (-)-shikimic acid, and modified it into a tetracyclic template (Scheme 3). This strategy renders the natural-product template more rigid, whilst retaining the structural features common to natural products. This template allowed for a large number of diversity modifications to build a small-molecule library from which activators of a TGF-ß-responsive reporter gene in mammalian cells were identified.^[3]

Using an already known drug scaffold is another approach to library design. Sulfonamides belong to a class of drugs that has a wide range of pharmacological effects such as antibiotic, hypoglycemic, diuretic, and antihypertensive.^[30] The versatility of sulfonamide-containing compounds makes them an interesting group for the study of protein function. For that reason, sulfonamide derivatives are one of the most widely studied types of drug analogues. Owa and colleagues designed a sulfonamide-focused library that was used in phenotypic screens to discover two novel anticancer drug candidates.^[31] Other sulfonamide-focused libraries have been designed and used in screens to discover novel blood coagulant thrombin inhibitors and antimalarial compounds (Scheme 4).^[32,33]

Phenotype Screening

The next step of the forward chemical genetics process is to screen the small-molecule library in a phenotypic assay. An efficient assay system entails a fully automated high-throughputscreening format of large libraries in the physiological context of a model organism. To date, a number of systems have been used, including Arabidopsis plant, Danio rerio (zebra fish), Drosophila (fruit fly), Caenorhabditis elegans, yeast, and mice.^[34,35] Some advancements toward automating and miniaturizing the screening step have been made.

Stockwell and colleagues developed an interesting method to facilitate the screening step and the study of the lead compound's mode of action.^[36] They hypothesized that screening compounds with known and dissected biological activities and pathways will allow for rapid elucidation of novel biological targets and mechanisms. Thus, the authors assembled a library of 2036 structurally diverse compounds with a wide range of known biological activities and compiled all published information on their activities. As a case study, the compounds in this annotated compound library (ACL) were screened for their effects on tumor-cell proliferation, and the results were compared to the commercial-source library. They found that 1% of the ACL compounds have fourfold selectivity for killing tumor cells as opposed to primary cells, compared with only 0.01% of the commercial-source compounds. Next, they set out to

MINIREVIEWS

Scheme 1. a) Design of the 2,6,9-trisubstituted purine library; b) CDK1 and CDK2 inhibitors (Purvalanol A, B and Compound 52, 52Me and other 2,6,9-trisubstituted purine CDK inhibitors acting at different cell-cycle phases; c) microtubule-binding compounds; d) estrogen sulfotransferase inhibitor; e) carbohydrate sulfotransferase inhibitor; f) diminutol, microtubule dynamics regulator; g) IP3K inhibitor; h) purmorphamine, osteogenesis-inducing activity; i) reversine, converts differentiated cells into progenitor cells.

Scheme 2. Library based on a 2,2-dimethylbenzopyran template. a) Template for the solid-phase method; b) Template for the extension of diversity sites by solution-phase chemistry; c) compounds displaying activity against MRSA strains.

VIBIO(

Scheme 3. Natural-product template used in the synthesis of a small-molecule library by Schreiber and colleagues.

 R^c = acid building blocks or skip codon

a) Potential building-block-coupling sites and b) synthesized library compounds.

multiwell-plate format.^[5, 6, 38, 39] Schreiber and colleagues used this method to identify several small molecules that specifically modulate various aspects of embryonic development in a zebrafish system, such as development of the central nervous system, the cardiovascular system, the neural crest, and the ear.^[40] Myoseverin, a microtubule-binding molecule, was found by Rosania et al. from

screening a 2,6,9-trisubstituted

Scheme 4. Novel sulfonamide a) anticancer compounds, b) blood coagulant thrombin inhibitors, and c) antimalarial derivative.

uncover mechanisms of action associated with tumor-cell proliferation. For that, they used their compiled information and developed automated algorithms to accelerate the analysis. Among the 85 active compounds, 28 mechanisms were statistically over-represented. In addition to known antitumor mechanisms, other mechanisms, unrelated to antitumor action were found; this underscores the effectiveness of this method.

Another contribution to the facilitation of the screening step was made by Kapoor et al., who designed a small-molecule probe for the dynamics of cell division.^[37] In their strategy, the compound of interest is "photocaged" with a photolabile protecting group, ortho-nitrobenzyl ether (Scheme 5). The compound remains inert while it is being equilibrated with the cells until its release through a photolysis reaction to perturb the function of the target.

Another novel approach involves the use of an automated microscope. Together with image-analysis software, this approach allows for high-throughput screening of cells in a purine library using phase-contrast microscopy.^[18] Shair and co-workers identified an active molecule that perturbs protein trafficking by screening a library of 2946 small molecules that were similar to natural compounds in a mammalian-cell phenotypic assay.^[41]

Target Isolation and Identification

Once an interesting phenotype is found, the next step is to link this phenotype to the protein function. Enormous research efforts have been focused on this step of the forward chemical genetics process. Here, the lead compound needs to either have a built-in functional group or has to be fitted with a handle in order to fish out the target protein(s). Most commonly used methods include tethering the active compound onto an agarose support or labeling it with a biotin tag. $[42]$ These methods then require protein separation by gel electrophoresis and detection by western blotting. Some innovative approaches to expedite the target-isolation step have been developed.

For example, one approach was designed to accelerate the attachment of the lead compound to an

Scheme 5. Photocaged small-molecule probe.

agarose support without having to perform extensive and laborious SAR studies to fit the lead compound with a handle, which can often lead to the loss of the compound's activity. Chang and co-workers designed a triazine library with a builtin linker containing an amino functionality, which, after the phenotype testing, can be directly used for attaching the compound to an agarose support (Scheme 6a).^[43] Using this method, Chang et al. found compound 1 and its more potent

S. M. Khersonsky and Y.-T. Chang

derivative, encephalazine, by a phenotypic screen of zebra-fish embryos; 1 was then directly tethered to an agarose support to fish out several ribosomal accessory proteins or their complexes as the target (Scheme 6b).

Scheme 6. a) Tagged-triazine-library scaffold; b) active compounds found in the forward chemical genetics screen with zebra-fish embryos and used to isolate the target protein.

In some cases the affinity and specificity of small molecules to their target proteins are rather low; this can lead to nonspecific interactions rendering the identification of the primary binding partners difficult.^[44] In this case, the use of affinity matrices to isolate the target protein is not sufficient. To address this problem, Oda et al. developed a comprehensive method that involves quantitative proteome analysis.^[44] The authors first identified a total of 285 proteins from affinity-matrix work using two-dimensional HPLC-MS/MS analysis. They went on to perform two-dimensional differential in-gel electrophoresis (2D-DIGE) and the isotope-coded affinity tag (ICAT) method to identify a number of common proteins. They further performed array-based transcription profiling, which, in addition to their proteome analyses, implicated metabolic enzyme proteins as the possible target. Finally, they carried out surface plasmon resonance (SPR) analysis using the active small molecule and several metabolic enzymes discovered through the proteome analysis to find that cytosolic MDH was the target protein.

Yet another innovative approach that omits the agarose step altogether involves a covalent modification of a target protein with a library compound. Here, the library compounds are designed to contain an electrophilic group such as an epoxide, fluorophosphonate, sulfonate ester, or vinyl sulfone as well as a tag. $[42, 45-50]$ The use of fluorescence or radioactively tagged compounds allows for direct visualization of target proteins by scanning the gel, thus eliminating the western-blot step and further accelerating target-isolation step.^[45-47] By using the reactive compound, the target protein is covalently modified and can then be easily isolated from the mixture by using the built-in tag. This approach utilizes activitybased protein profiling, which is complementary to bindingbased affinity-matrix methods. For example, Cravatt and coworkers designed a library of biotinylated and rhodamine-

tagged sulfonate esters and used them to profile the reactivity of a proteome based on properties other than protein abundance (Scheme 7).^[48]

When designing such a scaffold, the choice of electrophilic group is key. More specifically, its reactivity should be carefully considered, since the selectivity of the target protein toward the compound will be compromised if the electrophilic group is too reactive.

Conclusion

Forward chemical genetics is proving to be a powerful approach to studying protein function. However, since it has been estimated that there are over 30 000 protein-encoding genes in the human genome, innovative strategies will be needed to accelerate forward chemical genetics work. Significant progress has already been made in all three steps of the forward chemical genetics process; a) small-molecule library synthesis, b) screening for interesting phenotypes, and c) isolation and identification of the target protein. With the advent of combinatorial chemistry, generation of small-molecule libra-

Scheme 7. Examples of activity-based probes.

ries has become rapid, once the scaffold is designed. Phenotype screening is probably still the bottleneck step of the process; thus new robust high-throughput screening methods are required. A number of useful research tools for the final targetprotein-isolation step have been introduced, but more generalized strategies are still needed in order to target diverse groups of proteins.

Strong collaborative efforts between chemists and biologists to make this field more systematic will greatly accelerate forward chemical genetics. The post-genome-project era is an exciting time for the scientific community; the forward chemical genetics approach holds great promise to be an integral player in the study of protein function and ultimately human disease.

Keywords: bioorganic chemistry \cdot forward chemical genetics \cdot phenotype screening \cdot proteins \cdot small-molecule libraries

- [1] S. L. Schreiber, Chem. Eng. News 2003, 81, 51-61.
- [2] S. L. Schreiber, Bioorg. Med. Chem. 1998, 6, 1127 1152.
- [3] D. S. Tan, M. A. Foley, B. R. Stockwell, M. D. Shair, S. L. Schreiber, J. Am. Chem. Soc. 1999, 121, 9073-9087.
- [4] D. S. Tan, Nat. Biotechnol. 2002 , 20 , $561 563$.
- [5] R. S. Lokey, Curr. Opin. Chem. Biol. 2003, 7, 91-96.
- [6] G. E. Ward, K. L. Carey, N. J. Westwood, Cell. Microbiol. 2002, 4, 471-482. [7] N. S. Gray, Curr. Opin. Neurobiol. 2001, 11, 608-614.
- [8] R. Breinbauer, I. R. Vetter, H. Waldmann, Angew. Chem. 2002, 114, 3002 -3015; Angew. Chem. Int. Ed. 2002, 41, 2878-2890.
- [9] P. M. Abreu, P. S. Branco, J. Braz. Chem. Soc. 2003, 14, 675-712.
- [10] U. Abel, C. Koch, M. Speitling, F. G. Hansske, Curr. Opin. Chem. Biol. 2002, $6, 453 - 458$.
- [11] A. M. Rouhi, Chem. Eng. News 2003, 81, 77-107.
- $[12]$ Y. Z. Shu, J. Nat. Prod. 1998, 61, 1053 1071.
- [13] G. R. Eldridge, H. C. Vervoort, C. M. Lee, P. A. Cremin, C. T. Williams, S. M. Hart, M. G. Goering, M. O'Neil-Johnson, L. Zeng, Anal. Chem. 2002, 74, $3963 - 3971.$
- [14] K. U. Bindseil, J. Jakupovic, D. Wolf, J. Lavayre, J. Leboul, D. van der Pyl, Drug Discovery Today 2001, 6, 840-847.
- [15] N. S. Gray, L. Wodicka, A. M. Thunnissen, T. C. Norman, S. Kwon, F. H. Espinoza, D. O. Morgan, G. Barnes, S. LeClerc, L. Meijer, S.-H. Kim, P. G. Schultz, Science 1998, 281, 533-538.
- [16] Y. T. Chang, N. S. Gray, G. R. Rosania, D. P. Sutherlin, S. Kwon, T. C. Norman, R. Sarohia, M. Leost, L. Meijer, P. G. Schultz, Chem. Biol. 1999, 6, $361 - 375.$
- [17] Y. T. Chang, S. M. Wignall, G. R. Rosania, N. S. Gray, S. R. Hanson, A. I. Su, J. Merlie, Jr., H. S. Moon, S. B. Sangankar, O. Perez, R. Heald, P. G. Schultz, J. Med. Chem. 2001, 44, 4497-4500.
- [18] G. R. Rosania, Y. T. Chang, O. Perez, D. Sutherlin, H. Dong, D. J. Lockhart, P. G. Schultz, Nat. Biotechnol. 2000, 18, 304-308.
- [19] D. E. Verdugo, M. T. Cancilla, X. Ge, N. S. Gray, Y. T. Chang, P. G. Schultz, M. Negishi, J. A. Leary, C. R. Bertozzi, J. Med. Chem. 2001, 44, 2683-2686.
- [20] J. I. Armstrong, A. R. Portley, Y. T. Chang, D. M. Nierengarten, B. N. Cook, K. G. Bowman, A. Bishop, N. S. Gray, K. M. Shokat, P. G. Schultz, C. R. Bertozzi, Angew. Chem. 2000, 112, 1359-1362; Angew. Chem. Int. Ed. 2000, 39, 1303 - 1306.
- [21] S. M. Wignall, N. S. Gray, Y. T. Chang, L. Juarez, R. Jacob, A. Burlingame, P. G. Schultz, R. Heald, Chem. Biol. 2004, 11, 135-146.
- [22] Y. T. Chang, G. Choi, Y. S. Bae, M. Burdett, H. S. Moon, J. W. Lee, N. S. Gray, P. G. Schultz, L. Meijer, S. K. Chung, K. Y. Choi, P. G. Suh, S. H. Ryu, ChemBioChem 2002, 3, 897-901.
- [23] X. Wu, S. Ding, Q. Ding, N. S. Gray, P. G. Schultz, J. Am. Chem. Soc. 2002, 124, 14520 - 14521.
- [24] S. B. Chen, Q. S. Zhang, X. Wu, P. G. Schultz, S. Ding, J. Am. Chem. Soc. 2004, 126, 410 - 411.
- [25] K. C. Nicolaou, J. A. Pfefferkorn, A. J. Roecker, G. Q. Cao, S. Barluenga, H. J. Mitchell, J. Am. Chem. Soc. 2000, 122, 9939-9953.
- [26] K. C. Nicolaou, J. A. Pfefferkorn, H. J. Mitchell, A. J. Roecker, S. Barluenga, G. Q. Cao, R. L. Affleck, J. E. Lillig, J. Am. Chem. Soc. 2000, 122, 9954-9967.
- [27] K. C. Nicolaou, J. A. Pfefferkorn, S. Barluenga, H. J. Mitchell, A. J. Roecker, G. Q. Cao, J. Am. Chem. Soc. 2000, 122, 9968 - 9976.
- [28] K. C. Nicolaou, A. J. Roecker, S. Barluenga, J. A. Pfefferkorn, G. Q. Cao, ChemBioChem 2001, 2, 460-465.
- [29] K. C. Nicolaou, R. M. Evans, A. J. Roecker, R. Hughes, M. Downes, J. A. Pfefferkorn, Org. Biomol. Chem. 2003, 1, 908-920.
- [30] J. Drews, Science 2000, 287, 1960 1964.
- [31] A. Yokoi, J. Kuromitsu, T. Kawai, T. Nagasu, N. H. Sugi, K. Yoshimatsu, H. Yoshino, T. Owa, Mol. Cancer Ther. 2002, 1, 275-286.
- [32] S. W. Kim, C. Y. Hong, K. Lee, E. J. Lee, J. S. Koh, Bioorg. Med. Chem. Lett. 1998, 8, 735 - 738.
- [33] A. Ryckebusch, R. Deprez-Poulain, M. A. Debreu-Fontaine, R. Vandaele, E. Mouray, P. Grellier, C. Sergheraert, Bioorg. Med. Chem. Lett. 2002, 12, $2595 - 2598.$
- [34] B. R. Stockwell, Nat. Rev. Genet. 2000, 1, 116-125.
- [35] J. H. Nadeau, P. J. Dunn, Curr. Opin. Genet. Dev. 1998, 8, 311-315.
- [36] D. E. Root, S. P. Flaherty, B. P. Kelley, B. R. Stockwell, Chem. Biol. 2003, 10, $881 - 892.$
- [37] S. Hotha, J. C. Yarrow, J. G. Yang, S. Garrett, K. V. Renduchintala, T. U. Mayer, T. M. Kapoor, Angew. Chem. 2003, 115, 2481-2484; Angew. Chem. Int. Ed. 2003, 42, 2379-2382.
- [38] T. U. Mayer, Trends Cell Biol. 2003, 13, 270-277.
- [39] R. A. Blake, Curr. Opin. Pharmacol. 2001, 1, 533-539.
- [40] R. T. Peterson, B. A. Link, J. E. Dowling, S. L. Schreiber, Proc. Natl. Acad. Sci. USA 2000, 97, 12 965 - 12 969.
- [41] H. E. Pelish, N. J. Westwood, Y. Feng, T. Kirchhausen, M. D. Shair, J. Am. Chem. Soc. 2001, 123, 6740-6741.
- [42] D. A. Jeffery, M. Bogyo, Curr. Opin. Biotechnol. 2003, 14, 87-95.
- [43] S. M. Khersonsky, D. W. Jung, T. W. Kang, D. P. Walsh, H. S. Moon, H. Jo, E. M. Jacobson, V. Shetty, T. A. Neubert, Y. T. Chang, J. Am. Chem. Soc. 2003, 125, 11 804 - 11 805.
- [44] Y. Oda, T. Owa, T. Sato, B. Boucher, S. Daniels, H. Yamanaka, Y. Shinohara, A, Yokoi, J. Kuromitsu, T, Nagasu, Anal. Chem. 2003, 75, 2159-2165.
- [45] D. Greenbaum, K. F. Medzihradszky, A. Burlingame, M. Bogyo, Chem. $Biol. 2000, 7, 569 - 581.$
- [46] M. Bogyo, S. Verhelst, V. Bellingard-Dubouchaud, S. Toba, D. Greenbaum, Chem. Biol. 2000, 7, 27-38.
- [47] M. P. Patricelli, D. K. Giang, L. M. Stamp, J. J. Burbaum, Proteomics 2001, $1, 1067 - 1071.$
- [48] G. C. Adam, B. F. Cravatt, E. J. Sorensen, Chem. Biol. 2001, 8, 81-95.
- [49] G. C. Adam, E. J. Sorensen, B. F. Cravatt, Nat. Biotechnol. 2002, 20, 805-
- 809. [50] G. Y. J. Chen, M. Uttamchandani, Q. Zhu, G. Wang, S. Q. Yao, ChemBio-Chem 2003, 4, 336-339.

Received: February 17, 2004