Pure Appl. Chem., Vol. 78, No. 1, pp. 1–14, 2006.
doi:10.1351/pac200678010001

Evolutionary approaches for the discovery of functional synthetic small molecules*

Zev J. Gartner

Department of Chemistry, University of California, Berkeley, CA 94702, USA

Abstract: Directed evolution is a powerful method for the laboratory discovery of nucleic acids and proteins with desired functional properties. A hallmark of this approach is the iterative translation, selection, amplification, and diversification of genetic information. The potential of evolutionary methods to impact the discovery of *synthetic small molecules* has recently been explored by a variety of laboratories. Four methods encompassing some or all of the hallmarks of evolution are discussed, including dynamic combinatorial chemistry (DCC), genetic algorithms (GAs), DNA display, and DNA-templated synthesis (DTS).

Keyword: DNA-templated synthesis; directed evolution; genetic algorithms; dynamic combinatorial chemistry; DNA display; nanotechnology; bottom-up synthesis.

The rational design of molecules to serve as catalysts, drugs, and materials, remains challenging despite our knowledge of many of the physical laws that govern molecular interactions [1,2]. As an alternative to purely design-based approaches, many chemists specifically interested in the discovery of functional molecules have come to rely on more empirical methods, using iterated cycles of design, synthesis, activity assays, and structure–activity relationship (SAR) formulation [3] (Fig. 1). However, this traditional chemical approach to discovery is both material-intensive and time-consuming. These two factors severely limit the number of possible molecules that can be designed, synthesized, and screened to optimize a desired chemical function.

The introduction of combinatorial chemistry along with developments in high-throughput screening have dramatically increased the number of molecules that can be designed, synthesized, and assayed for function in a limited amount of time. For example, the solid-phase split-pool synthesis of a library exceeding 10^6 molecules has been reported [4]. Additionally, chip-based fluorescent screens have been developed that in principle allow the analysis of millions of small molecule–protein interactions per day [5]. Although these technologies are promising, they do not differ fundamentally from the traditional chemical approach, and are thus bound by the same basic limitations.

In contrast to the chemist's approach to discovery described above, functional molecules arise in nature through the process of evolution: iterated cycles of translation, selection, amplification, and diversification of genetic material (Fig. 1). The basic principles of evolution can be harnessed in vitro, and experiments using these principles are typically referred to as directed evolution. Any system capable of evolving in vitro must

i. sort "fit" molecules from "unfit" molecules (a selection) and
ii. occasionally diversify the molecules surviving a selection.

Pure Appl. Chem.* **78, 1–64 (2006). A collection of invited, peer-reviewed articles by the winners of the 2005 IUPAC Prize for Young Chemists.

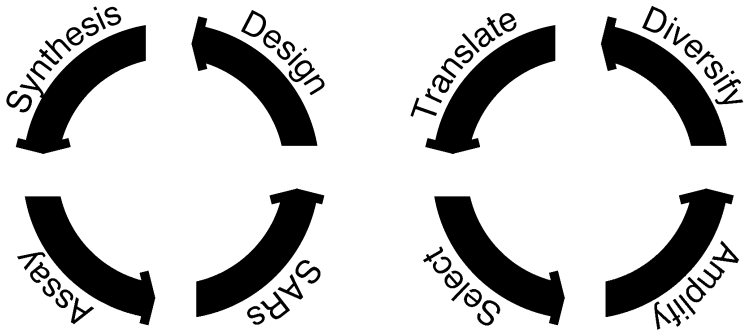

Fig. 1 A chemist's (left) vs. nature's (right) approach to discovering functional molecules.

In the case that the selected molecules may not be directly amplified and diversified, the evolving system must also

iii. translate an amplifiable information carrier into selectable molecules and
iv. link these molecules with the information that encodes them.

The power of directed evolution to generate proteins or nucleic acids with improved or novel function is well established [6–8]. A typical experiment begins with a diverse pool of an amplifiable information carrier (DNA or RNA), which is then translated into a corresponding collection of molecules (either other nucleic acids using polymerase enzymes, or proteins using the ribosome) in a manner that preserves the linkage between information carrier and translated structure. These molecules are then subjected en masse to a selection for functional properties, from which the most "fit" molecules emerge. The encoding nucleic acids are subsequently amplified by polymerases, and in the process, the genetic information is diversified by the rare introduction of errors, cross-overs, and reorganizations. This cycle is repeated, and the pool of nucleic acids is continually enriched for those that encode molecules of increasing fitness. The system is said to be evolving when its theoretical diversity (or "sequence space" in the case of a biological polymer) is larger than the number of unique molecules present at any single iteration of the cycle.

The steps of translation, selection, amplification, and diversification of evolution-based approaches are analogous to the steps of synthesis, screening, formulating SARs, and design used by chemists. However, the two methods differ fundamentally in the way they manipulate complex mixtures, evaluate their properties, and search for improved function (Fig. 2). In traditional approaches to discovery, chemists must design, synthesize, and assay molecules as discrete species or in small pools that must be chemically tagged and subsequently deconvoluted [9]. This contrasts with evolutionary approaches in which all manipulations are performed on diverse and heterogeneous pools of molecules (so-called "one pot" manipulations).

A second fundamental difference between chemists' and nature's approach to discovery lies in the way molecular fitness is evaluated. Not only must chemists use the inefficient manipulations described above, but many copies of each library member must be available in order to assay activity. The requirement for large numbers of molecules is a result of the sensitivity of the screens available to chemists, the best of which can typically detect about one picomole (10^{12} copies) of a single active structure [10]. By contrast, in vitro selections sort according to the properties of *single molecules*—that when linked to an information carrier such as DNA—can be amplified by the polymerase chain reaction (PCR). Since fewer than 10^{-20} moles (10^4 copies) of DNA are routinely amplified using PCR, selections are at least eight orders of magnitude more sensitive than typical screens.

The way in which evolution "performs" a search for fitness also contrasts fundamentally with the approach taken by chemists. In traditional approaches to discovery, a researcher must generate SARs in

	Chemist's Approach	Nature's Approach
Identifying Hits	individually in spatially separated *assays*	one pot *selections*
Material Quantity	> ~10^{12} copies of each molecule	<~1,000 copies of each molecule
Sample Size	≤10^{6} members per screen (CAS < 10^{8})	≤10^{15} members per selection
Generality	whatever can be synthesized	nucleic acids & proteins

Fig. 2 Comparing chemists' and nature's approaches to discovering functional molecules.

order to design new and potentially improved pools of molecules to be synthesized and assayed. This is typically done using empirical observations, but in certain cases, can be performed quantitatively using predefined scoring functions (QSARs). However, our ability to formulate correct SARs decreases dramatically as the complexity of the system increases (e.g., as with macromolecules such as proteins) [11]. Furthermore, bias may be introduced either by the researcher or by imperfect scoring algorithms, ultimately skewing the redesigned pool of molecules in a suboptimal manner. Evolution, on the other hand, randomly mutates and recombines the properties of the most-fit molecules resulting from a selection, precluding the necessity to generate SARs or fitness functions. In addition, these manipulations are performed on pools, making this process highly efficient. The random and efficient introduction of diversity into the evolving pool of molecules allows nonintuitive solutions to emerge that may have been passed over by lower throughput, more biased approaches.

The fundamental differences in the way chemists and nature manipulate complex mixtures, evaluate their properties, and search for improved functions, have several practical implications. First, evolutionary based approaches can search through vastly larger numbers of molecules, or "complexities," than traditional chemical approaches. This practical advantage of evolutionary approaches stems from the fact that all manipulations can be performed in one-pot, very few molecules of each species are needed, and an efficient search algorithm exists to navigate, rather than randomly sample, a complex and multidimensional "sequence space". For example, a researcher directing the in vitro evolution of nucleic acids or proteins can routinely diversify, translate, select, and amplify greater than 10^{15} different molecules in a single iteration of the evolutionary cycle. This number represents more than seven orders of magnitude greater complexity than is present in the entire CAS (Chemical Abstract Service) small-molecule database. More importantly, mutation of the most-fit molecules in subsequent iterations of the evolutionary cycle gradually narrows the search to the most active regions of sequence space identified in the initial iterations.

A second difference between the chemist's and evolutionary approaches to discovery, and the most important for the present discussion, is the generality of systems to which they can be applied. Traditional approaches are inherently more general in the types of structures that can be accessed because they are limited only in as much as chemists are limited by the tools of synthetic chemistry. Evolutionary approaches, on the other hand, can only be applied to molecules that can be directly amplified and diversified or alternatively, can be translated from and linked to an amplifiable information carrier. Until recently, the only methods for directly translating DNA (the prototypical amplifiable in-

formation carrier) into other chemical structures involved the use of polymerases or the ribosome, limiting the types of structures that could be evolved to nucleic acids and proteins, respectively.

The power of directed evolution to discover functional biological polymers is well established, but at the outset, it is not entirely obvious that analogous efforts to discover functional synthetic small molecules will benefit in the same way. Do the magnitudes of the two endeavors even compare? The enormous size of "small-molecule space" has been estimated at between 10^{60} and 10^{100} molecules [1], an equivalent level of complexity to that represented by all 45–75mer proteins or 100–165mer oligonucleotides. The challenge of exploring such diversity is well within the grasp of evolutionary methods, as functional proteins and nucleic acids have been discovered by evolving de novo libraries of similar length sequences [12,13]. However, applying the power of evolution to the exploration of small-molecule diversity has until recently been hampered by the lack of a general method to translate an amplifiable information carrier such as DNA into arbitrary chemical structures.

Four approaches are discussed below that bring part or all of the aspects of evolutionary methods to bear on the challenge of chemical discovery. Two of these, target-templated chemical libraries (such as those based on dynamic combinatorial chemistry, DCC) and genetic algorithms (GAs), sidestep the problem of translating genetic information into chemical structures so as to benefit where possible from other steps of the evolutionary cycle. DNA display and DNA-templated synthesis (DTS) are two other recently developed methods that allow the translation of DNA into arbitrary chemical structures. As a result, they have enabled the full evolutionary cycle to be applied to the discovery of synthetic small molecules of defined function.

TARGET-TEMPLATED CHEMICAL SYNTHESIS

Systems based upon target-templated chemical synthesis (TTCS) were pioneered by Lehn and coworkers in the 1990s and can be broadly divided into two categories: those libraries assembled under equilibrating conditions (commonly referred to as DCC) and libraries assembled irreversibly (referred to here as target-accelerated chemistry, TAC). Both methods use the addition of a target molecule (template) to skew product distributions of a library of reactants toward those products that interact most favorably with the target, effectively enriching the population for functional molecules. Whether the reaction being used is under thermodynamic or kinetic control determines the mechanism of this effect. Under thermodynamic control, the species with the most favorable binding energetics among an equilibrating pool are stabilized via noncovalent interactions with a template and are thereby removed from solution (Fig. 3). Restoration of the equilibrium within the pool of unbound molecules biases the population toward those that interact most strongly with the target, thereby increasing the representation of the best binders [14]. Under kinetic control, stabilization of a particular subset of transition states by the template accelerates reactions between the best binding starting materials, leading to the accumulation of products containing those components [15,16] (Fig. 4). In either case, when the reaction is stopped, the distribution of products is compared to control experiments where the template is omitted. Reaction products that appear in greater quantity in the presence of the template than without are likely to be the best binders. Libraries assembled by TTCS allow the synthesis and assaying of an array of compounds to be performed with a single manipulation, in a manner analogous to evolution. However, no strategies have been proposed to introduce new diversity into the pool of reactants, to relate the assembled structures to easily read amplifiable information carriers, or to iterate their synthesis and analysis.

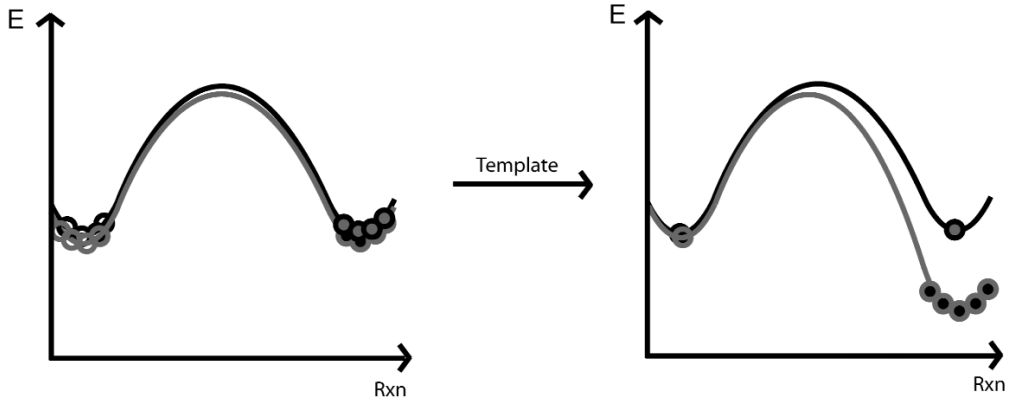

Fig. 3 Dynamic combinatorial chemistry. An equilibrating landscape of reacting molecules is biased toward those that interact most strongly with a template.

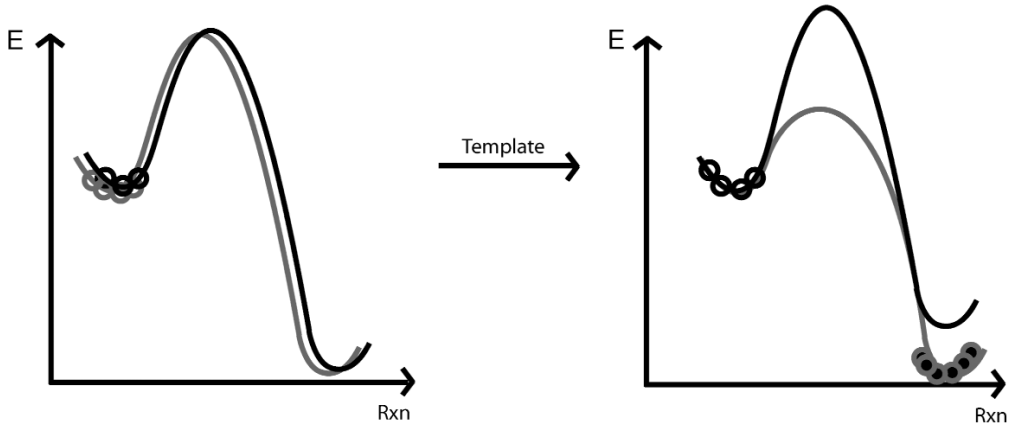

Fig. 4 Target-accelerated chemistry. Reactions between building blocks that interact most strongly with a target are accelerated, leading to the accumulation of their products.

The use of DCC is growing and has been reviewed extensively [17] (Fig. 5). Thus far, applications have been limited by the small number of useful chemical reactions that are reversible under mild conditions (around 10 have been suggested), and by the difficulty of deconvoluting the complex mixtures that result from these experiments [18]. Nonetheless, various studies have validated the ability of dynamic combinatorial libraries to identify functional ligands for proteins and nucleic acids, as well as receptors for organic molecules and metal ions [17]. For example, in an early application of DCC, a mixture of 22 starting materials and products equilibrating via imine exchange was selectively enriched for one molecule in the presence of the target enzyme carbonic anhydrase [19]. Encouragingly, the enriched compound closely resembled a known inhibitor of the enzyme. In another example of dynamic combinatorial libraries employing hydrazone exchange, the authors enriched for a molecule with an IC_{50} of 2.8 nM for acetylcholinesterase from a virtual library of 105 products and starting materials [20]. Bis-cationic inhibitors of *Bacillus subtilis* HPr kinase/phosphatase were identified from similar libraries [21]. DCC was also used to identify optimal linkers for dimers of vancomycin with therapeutic potential in vancomycin-resistant *enterococci* infections [22]. The investigators found several compounds with antibacterial activities improved by several orders of magnitude over monomeric vancomycin.

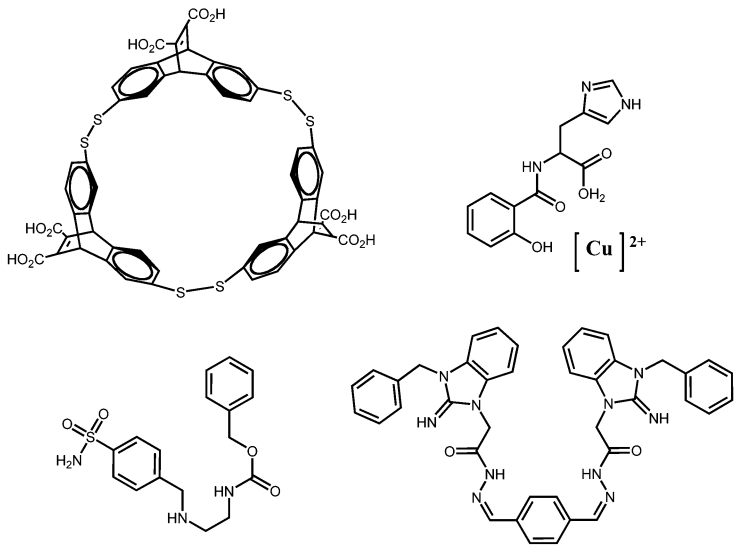

Fig. 5 Four molecules discovered using TTCS. Clockwise from upper left: a receptor for *N*-methyl morphine, an RNA binder, an inhibitor of the *B. subtilis* HPr kinase/phosphatase, and a carbonic anhydrase binder.

DCC has also been used to find synthetic receptors for small organic molecules. In one example, synthetic small-molecule receptors with moderate affinity ($K_d > 10^{-3}$) for acetylcholine and other biologically important quaternary ammonium salts were templated from pools of linear and cyclic pseudopeptides [23]. Building on these results, receptors based on cyclic aromatic disulfides were isolated with higher affinity ($K_d > 10^{-5}$) and modest selectivity for 2-methylisoquinolinium or *N*-methyl morphine [24].

Unlike DCC, libraries based on TAC are not limited to the use of reversible chemical reactions. These libraries take advantage of cases in which the transition state of a chemical reaction closely resembles the enzyme-bound product. For example, when the reactive groups between two species represent a small fraction of the total surface interacting with a target, the target stabilizes the transition states that lead to the synthesis of the best binders. Applying this strategy, sulfonamide inhibitors of carbonic anhydrase, bearing substantial resemblance to those described by Huc and coworkers, were isolated from a pool of reacting thiols and benzylic chlorides [15]. In another report, binders of acetylcholinesterase with pM affinities were isolated from a pool of azides and alkynes capable of reacting to form libraries of triazoles [16].

While the basic principles of DCC and TAC have been validated, these methods suffer from several intrinsic difficulties that must be overcome to enable their more general application. These difficulties include the limited diversity of available building blocks that have appropriate chemical handles for target-templated synthesis, as well as the difficulty of deconvoluting the complex mixtures that result from TTCS-based libraries. Combining thermodynamically controlled reactions with kinetically controlled reactions could address the issue of building block diversity, and enable the construction of larger TTCS-based libraries. Using a variation of this concept, Lehn and coworkers built "multi-tiered" libraries via two orthogonal and reversible reactions: metal coordination and imine exchange [20,25]. In addition, an iterative process that continually substitutes unfit building blocks, guided by the results of previous iterations, may increase library diversity and reduce the difficulty in deconvoluting the selected pool of molecules. While such a strategy may allow TTCS-based libraries to be implemented in an iterated manner analogous to evolution, it would still require extensive manipulation of the evolving system, as well as the selection of new building blocks after each iteration of the process (Fig. 6).

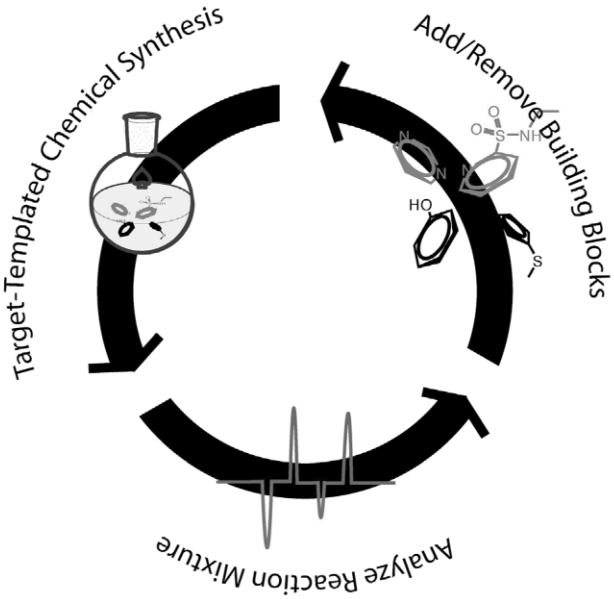

Fig. 6 Iterated TTCS libraries. Building blocks that are consumed and appear in templated products (peaks above baseline in chromatography trace) remain in the evolving pool, while those not being used (peaks below baseline in chromatography trace) are removed and replaced with new, potentially improved building blocks.

GENETIC ALGORITHMS

When the chemical structures under study can not be directly amplified and diversified as in TTCS libraries, the information inherent in the structures must be represented and manipulated in another form. This may be done in silico using GAs. Genetic algorithms are computational methods first developed by Holland and coworkers in the 1970s as evolution-inspired optimization techniques [26]. GAs have been shown to efficiently find solutions to problems characterized by ill-defined and complex functions of large numbers of variables [27]. When applied to chemical systems, GAs assume that a molecule's fitness can be expressed as a function of the fitness of its components (building blocks). A major advantage of GAs is that a priori knowledge of this fitness function is not needed in order for the algorithms to be applied successfully [27]. To be implemented, however, GAs require a method of encoding and diversifying chemical information that's reflective of the manner in which the evolving structures will be assembled. As such, molecules are typically encoded based on the connectivity of their building blocks. This approach allows diversification by mutation and recombination to be applied to the encoded molecules in silico in a manner analogous to that applied to DNA during evolution.

GAs operate on virtual libraries composed of all possible combinations of the available building blocks chosen by the researcher. These "virtual libraries" are directly analogous to the "sequence space" of a biological polymer. In either case, only a small fraction of the virtual library is ever created in the laboratory. Rather than exhaustively synthesizing and evaluating each potential molecule, GAs navigate through this virtual space of compounds by suggesting the synthesis of a tiny but diverse fraction of the library. These molecules are then synthesized and assayed for activity. The information encoding the best molecules is then recombined and mutated by the GA, providing a new set of compounds to be synthesized and assayed. Amplification of the information encoding the most-fit individuals, a hallmark of evolving systems, does not occur explicitly, but is not needed because the relative fitnesses of the encoded molecules are retained in silico. Therefore, GAs substitute the steps of SAR formulation and design typically used by chemists, for the evolutionary step of diversification, while retaining the need to

synthesize and assay individual molecules using traditional methods. The net result is the ability to navigate through large virtual libraries in a more efficient, unbiased manner.

In an early application of GAs toward the discovery of functional molecules, Singh and coworkers reported the identification of substrates for the protease stromolysin from a virtual library of 64 000 000 hexapeptides [28]. They reported a modest 3-fold increase in susceptibility of the best peptides to cleavage by stromolysin after five generations (defined as one round of synthesis and assays, followed by application of the GA). In all, a total of 60 peptides were synthesized and assayed per generation, 300 in total. However, the authors acknowledge that the number of generations was probably insufficient to allow the population to converge on an optimal solution. Using a similar approach, but a reduced theoretical library size, Yokobayashi and coworkers evolved inhibitors of trypsin from hexapeptides of Phe, Lys, Ile, and Thr (16 384 possible peptides) [29]. Average inhibitory activity per generation was increased from 16 to 50 % after six successive applications of their GA.

Weber and coworkers undertook a more ambitious application of GAs, searching through a virtual library of Ugi four-component coupling products for inhibitors of the protease thrombin [30] (Fig. 7). From a total of 160 000 possible molecules, a sub-μM binder was identified after 18 generations. Fewer than 400 compounds (<0.25 % of the virtual library) were synthesized to achieve this result. The authors attribute their success to the larger number of generations used relative to Singh and Yokobayashi.

The work of Weber and coworkers is of particular interest because the molecules evolved by their GAs were nonpeptidic. Therefore, it was not immediately obvious whether these compounds would benefit from evolution in a manner analogous to proteins and nucleic acids—the substrates for evolution in natural systems. Recognizing this, they undertook the exhaustive synthesis of 15 360 molecules composed of three of the four dimensions of diversity from their previous library [31]. Each of these molecules was assayed for its ability to inhibit thrombin in order to better understand the topology of the functional landscape being searched by the GA. Weber and coworkers then performed simulated evolution experiments to optimize the various parameters of the GA. They found that of the 15 360 molecules in their virtual library, only 9 compounds (0.059 %) had IC_{50} values below 1 μM, 54 (0.352 %) had IC_{50} between 1 and 10 μM, and 675 (4.395 %) had IC_{50} values between 10 and 100 μM. For this small library, the authors determined that their GA found one of the 9 sub-μM thrombin inhibitors 284 % more efficiently when compared to a high-throughput random screening method. However, they also concluded that the improvement in efficiency would rise dramatically as the size and complexity of the virtual library increased.

Work in the area of GAs has two important implications for the prospects of applying evolutionary principles to chemical discovery. First, it has established that the benefits of evolutionary searches apply equally well to synthetic molecules as to nucleic acids and proteins. Second, it suggests that the larger numbers of molecules accessible using evolutionary methods will accentuate the advantages of evolution-based genetic search algorithms.

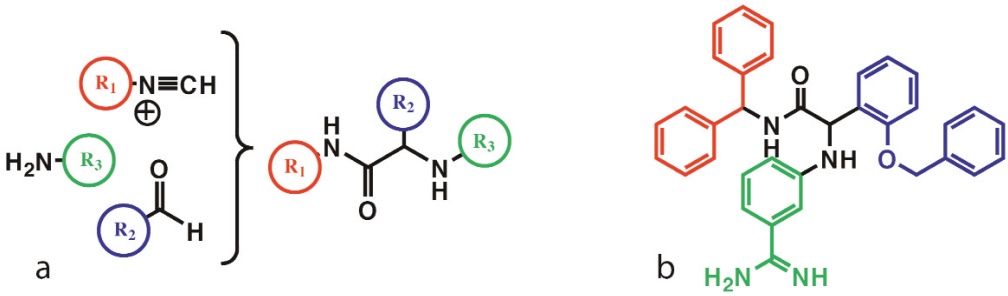

Fig. 7 Ugi-type three component coupling reactions. (a) The Ugi-type three component coupling reaction used by Illgen et al. [31]. (b) Sub-μM thrombin binder discovered by Weber et al. using a GA [30].

DNA DISPLAY

An ideal application of evolutionary methods would allow for the seamless, one-pot, and iterative application of the evolutionary operations of selection, amplification, diversification, and translation to libraries of arbitrary chemical structure. TTCS libraries fall short of this ideal, as useful chemistries are limited, iteration has yet to be demonstrated, and deconvolution of a complex mixture must follow every experiment. Similarly, computation-based evolutionary methods necessitate the laborious synthesis and screening of individual compounds after every application of the GA.

Bringing the possibility of performing directed evolution on arbitrary chemical structures closer to a reality, Halpin and Harbury have presented "DNA display" as a method of translating and establishing a physical link between an amplifiable information carrier and libraries of small molecules [32]. To do so, they take advantage of single-stranded DNA's chemical stability, ease of amplification, and ability to hybridize to complementary sequences. Like other combinatorial technologies, a tag (in this case, DNA) is used to decode the synthetic history of an individual member of the library [4,33]. However, unlike most other combinatorial approaches, the DNA tag is also a determinant of which sequence of reactions are used to synthesize a given molecule. It is installed *before* synthesis where it *di-*

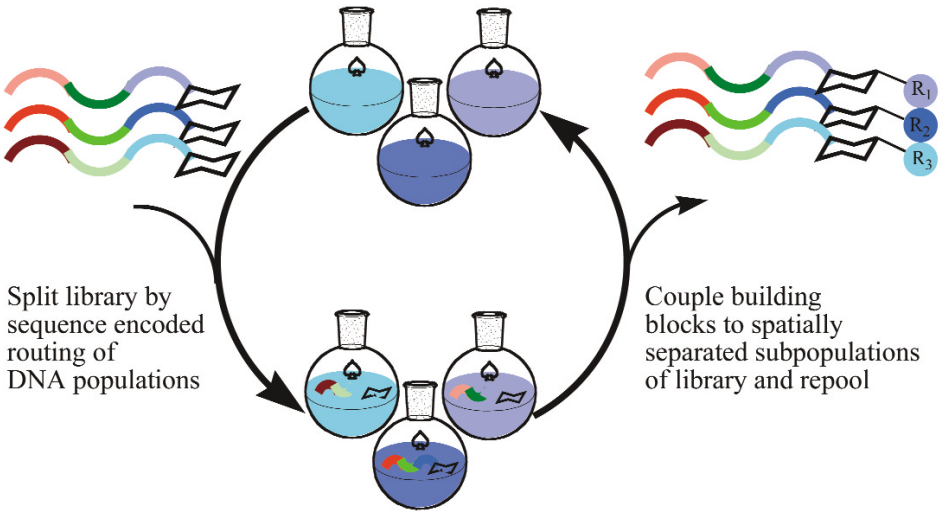

Fig. 8 DNA display. DNA libraries are split into subpopulations according to their sequences. Each subpopulation, having in common a particular DNA codon, is subjected to a particular set of reaction conditions. The reacted subpopulations are re-pooled and subjected to further manipulations.

rects the chemical fate of each growing molecule in the library as it moves through a split-pool synthetic scheme.

Halpin and Harbury use DNA tags to sequence specifically fractionate a pool of DNA-linked small molecules (Fig. 8) [34]. The tags are composed of PCR primer binding sites flanking one or more DNA "codons," each of which is associated with a particular chemical step and building block in a library synthesis. For each codon, a corresponding reaction flask is prepared, to which is immobilized a single complementary oligonucleotide sequence. When the library of DNA tags is allowed to equilibrate between these flasks, they hybridize to their complimentary sequences and the library is fractionated. This is analogous to the "splitting" of beads into smaller pools as performed before each round of a traditional split-pool synthesis, except that the splitting is directed by DNA hybridization. Once the library has been fractionated, oligonucleotides from each subpopulation are eluted onto a cationic support, subjected to solid-phase synthesis, and re-pooled prior to the next chemical step [35]. Library synthesis is complete after sufficient rounds of DNA-based splitting, synthesis, and pooling have allowed the translation of all codons on the DNA tags into the chemical structures synthesized on their ends. The resulting DNA-linked small molecule libraries are amenable to in vitro selection and may be amplified using PCR.

To demonstrate their technology, Halpin and Harbury synthesized a library of 1 million unnatural acylated pentapeptides, and subjected them to a selection for binding to monoclonal antibody 3-E7, raised again [Leu]enkephalin. After only two rounds of translation, selection and amplification, their initially diverse library had converged on a group of peptides which closely matched the [Leu]enkephalin consensus sequence. While the selected structures for the most part consisted of the naturally abundant amino acids, the library contained several amino acids with non-natural side chains and D stereochemistries, in addition to a variety of acyl capping groups that would have been difficult or impossible to incorporate by in vitro ribosomal translation systems [36].

DNA display has the potential to be a tremendously powerful method, yet it remains to be seen whether libraries based upon nonpeptidic scaffolds prove as robust as the peptide-based schemes described in their proof-of-principle work. However, should chemistry not prove limiting, libraries of unprecedented size and diversity should be accessible using the same commonly available chemical building blocks used in traditional combinatorial synthesis. Such libraries may be used to discover new catalysts, chemical switches, sensors, and materials [32].

DNA-TEMPLATED SYNTHESIS

Gartner, Liu, and coworkers have described DNA-templated organic synthesis as another method for the seamless, one-pot, and iterative application of evolutionary methods toward synthetic-organic molecules [37]. Whereas in DNA display, DNA tags are routed by their sequence to spatially separated reaction flasks for modification, DNA templates in DTS recruit complementary DNA-linked reagents to their proximity via Watson–Crick base-pairing (Fig. 9). All templates and reagents coexist in a single dilute solution. Only upon sequence-specific annealing is the effective molarity of complementary templates and reagents increased ($>10^9$ fold in some cases) to a level allowing chemical reactions to proceed [37] (Fig. 10).

The generality of DTS has been well established [38,39] (and references therein). It is compatible with amide bond formation in addition to many other useful chemical transformations [40] and does not strictly require functional group adjacency in the context of annealed templates and reagents [37,41]. DTS can also be performed in multistep or library formats [42–46]. When used in combination with flexible linkers, it has been shown that the nucleic acid component of templates and reagents do not effect the outcome of a chemical reaction apart from their ability to modulate effective molarities [47]. Methods have been developed to efficiently conduct DTS in organic solvents [48] and to perform DNA-templated functional group transformations, allowing non-DNA-linked reagents to be included in DTS synthetic schemes in a manner analogous to DNA display [49].

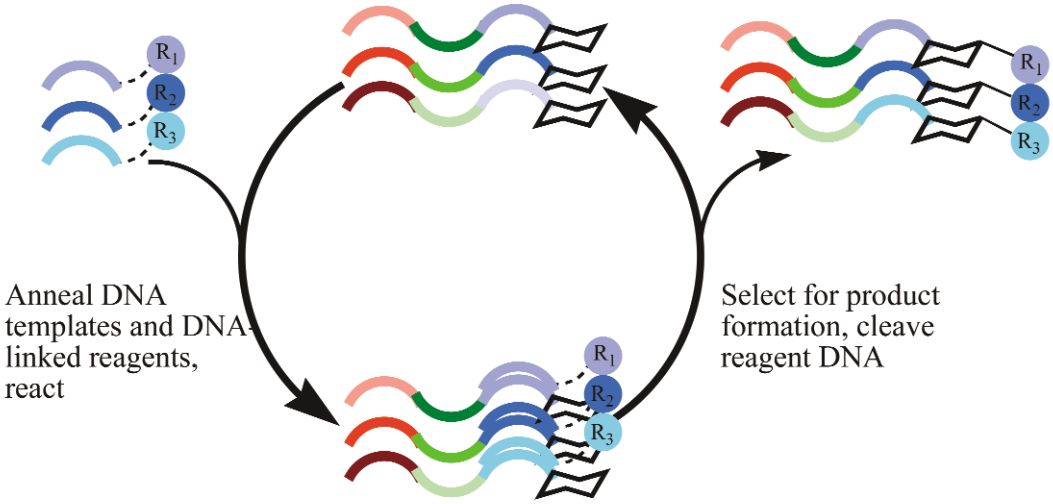

Fig. 9 DNA-templated syntheses. Pools of template libraries are combined with pools of DNA-linked reagents. Complementary sequences form duplexes and their functionalized ends react. Products are purified away from starting material by selection for bond formation and the decoding reagent DNA is chemically cleaved to liberate the functionalized templates for subsequent manipulations.

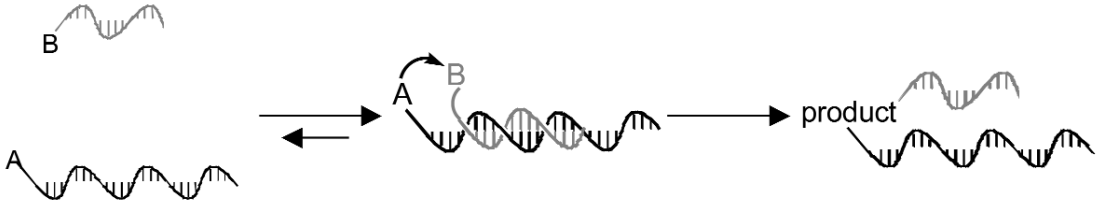

Fig. 10 A model for DTS. Reagent A-linked DNA hybridizes to complementary reagent B-linked DNA increasing the effective molarity of A relative to B, and allowing them to react.

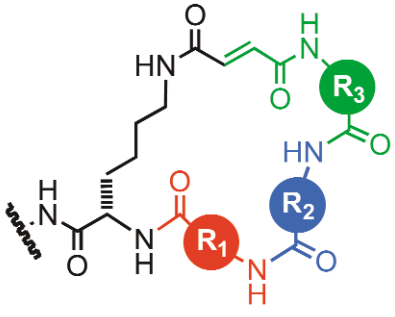

Fig. 11 Macrocyclic peptide fumaramide library. These structures were targeted because of the availability of numerous unprotected amino acid building blocks, their partially constrained nature, and their ability to capture proximal nucleophiles (found in many enzyme active sites) [43] as a consequence of their fumaramide functionality.

While not strictly necessary [49], the use of DNA-linked reagents has several practical advantages. First, purification strategies based upon selection for bond formation or cleavage allow products to be separated from starting materials independent of their structure or chemical properties [42]. Second, hydrophobic reagents can be made soluble in water as a consequence of their coupling to a

charged polymer such as DNA and the low concentrations necessary for performing DTS. Water is known to accelerate a variety of chemical reactions passing through transition states having a negative $\Delta V^{\dagger}$ (a property of most coupling reactions) or polar character relative to the ground state as a consequence of its high internal pressure and dielectric [50].

The ability to modulate effective molarities between multiple DNA-linked reagents in a single flask also enables new modes of reactivity not available to traditional chemical approaches. As a consequence of the low reagent concentrations used for DTS, multiple incompatible reactions can be directed in a single reaction vessel [44,51], and one-pot multistep DTSs can be performed using temperature gradients as the only external control [46]. This aspect of DTS also enables the use of chemical reagents that might have a tendency to homo-couple or polymerize at higher concentrations, having practical implications in the context of macrocyclization reactions, bifunctional reagents, and transition-metal-catalyzed cross-coupling reactions [52]. Indeed, the synthetic utility of several new transition-metal-catalyzed reactions is only revealed at the low concentrations used for DTS [52].

To highlight the power of DTS, Gartner and coworkers undertook the synthesis of a structurally diverse library of 65 macrocyclic peptide fumaramides, a novel class of molecules that are not generally accessible using known biological catalysts (Fig. 11). Despite unique challenges associated with the synthesis of macrocycles [53], the key cyclization step in their synthetic scheme was found to be efficient and general in the DNA-templated format. After three rounds of DTS and two rounds of selection for carbonic anhydrase affinity, the amplified DNA templates were enriched almost entirely for sequences coding for an aryl sulfonamide amino acid known to bind to carbonic anhydrase.

Like the method of Halpin and Harbury, the chemisty used for the synthesis of DTS libraries must be compatible with DNA. However, the subset of reactions falling into this category is surprisingly large and growing to include transformations not accessible to chemists running reactions at higher concentrations. This should allow DTS libraries to search through areas of "small-molecule space" that may be under explored at the present. In addition, DTS may form the foundation for a new "bottom-up" approach for the synthesis of nano- and meso-scale structures.

CONCLUSION

The physical linkage of a small molecule to a DNA template that both encodes and directs its synthesis lies at the heart of both DNA display and DTS. It is the unifying element enabling the seamless translation, selection, amplification, and diversification of genetic information. Indeed, DNA display and DTS utilize template architectures that are nearly identical. Such commonalities may allow the two techniques to be combined, so as to take advantage of their most powerful elements in a single reaction scheme. Both technologies are still in their infancy, and it will be exciting to follow their future development and application. Such applications will likely include the discovery of new catalysts [32,52], materials [54], and chemical transformations [52].

ACKNOWLEDGMENTS

Thanks to IUPAC for their recognition of this work, as well as to Bristol-Myers Squibb, Harvard University, and the National Science Foundation for their generous funding. I would also like to thank Prof. David Liu for helpful comments in the preparation of this manuscript.

REFERENCES

1. G. Schneider and U. Fechner. *Nat. Rev. Drug Discov.* **4**, 649 (2005).
2. T. Lanio, A. Jeltsch, A. Pingoud. *Protein Eng.* **13**, 275 (2000).
3. P. W. Erhardt. *Pure Appl. Chem.* **74**, 703 (2002).
4. D. S. Tan, M. A. Foley, M. D. Shair, S. L. Schreiber. *J. Am. Chem. Soc.* **120**, 8565 (1998).

5. K. S. Lam and M. Renil. *Curr. Opin. Chem. Biol.* **6**, 353 (2002).
6. J. A. Bittker, K. J. Phillips, D. R. Liu. *Curr. Opin. Chem. Biol.* **6**, 367 (2002).
7. H. Lin and V. W. Cornish. *Angew. Chem., Int. Ed.* **41**, 4402 (2002).
8. S. V. Taylor, P. Kast, D. Hilvert. *Angew. Chem., Int. Ed.* **40**, 3310 (2001).
9. M. H. Ohlmeyer, R. N. Swanson, L. W. Dillard, J. C. Reader, G. Asouline, R. Kobayashi, M. Wigler, W. C. Still. *Proc. Natl. Acad. Sci. USA* **90**, 10922 (1993).
10. J. B. Doyon, T. M. Snyder, D. R. Liu. *J. Am. Chem. Soc.* **125**, 12372 (2003).
11. D. T. Jones. *Pharmacogenomics J.* **1**, 126 (2001).
12. E. H. Ekland, J. W. Szostak, D. P. Bartel. *Science* **269**, 364 (1995).
13. A. D. Keefe and J. W. Szostak. *Nature* **410**, 715 (2001).
14. S. Otto, R. L. Furlan, J. K. Sanders. *Drug Discov. Today* **7**, 117 (2002).
15. R. Nguyen and I. Huc. *Angew. Chem., Int. Ed.* **40**, 1774 (2001).
16. W. G. Lewis, L. G. Green, F. Grynszpan, Z. Radic, P. R. Carlier, P. Taylor, M. G. Finn, K. B. Sharpless. *Angew. Chem., Int. Ed.* **41**, 1053 (2002).
17. S. J. Rowan, S. J. Cantrill, G. R. Cousins, J. K. Sanders, J. F. Stoddart. *Angew. Chem., Int. Ed.* **41**, 898 (2002).
18. J. M. Lehn. *Chem. Eur. J.* **5**, 2455 (1999).
19. I. Huc and J. M. Lehn. *Proc. Natl. Acad. Sci. USA* **94**, 2106 (1997).
20. T. Bunyapaiboonsri, O. Ramstrom, S. Lohmann, J. M. Lehn, L. Peng, M. Goeldner. *Chembiochem* **2**, 438 (2001).
21. T. Bunyapaiboonsri, H. Ramstrom, O. Ramstrom, J. Haiech, J. M. Lehn. *J. Med. Chem.* **46**, 5803 (2003).
22. K. C. Nicolaou, R. Hughes, S. Y. Cho, N. Winssinger, H. Labischinski, R. Endermann. *Chemistry* **7**, 3824 (2001).
23. R. L. E. Furlan, Y.-F. Ng, G. R. L. Cousins, J. E. Redman, J. K. M. Sanders. *Tetrahedron* **58**, 771 (2002).
24. S. Otto, R. L. Furlan, J. K. Sanders. *Science* **297**, 590 (2002).
25. V. Goral, M. I. Nelen, A. V. Eliseev, J. M. Lehn. *Proc. Natl. Acad. Sci. USA* **98**, 1347 (2001).
26. J. H. Holland. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*, University of Michigan Press, Ann Arbor, MI (1975).
27. L. Weber. *Curr. Opin. Chem. Biol.* **2**, 381 (1998).
28. J. Singh, M. A. Ator, E. P. Jaeger, M. P. Allen, D. A. Whipple, J. E. Soloweij, S. Chowdhary, A. M. Treasurywala. *J. Am. Chem. Soc.* **118**, 1669 (1996).
29. Y. Yokobayashi, K. Ikebukuro, S. McNiven, I. Karube. *J. Chem. Soc., Perkin Trans. 1* 2435 (1996).
30. L. Weber, S. Wallbaum, C. Broger, K. Gubernator. *Angew. Chem., Int. Ed. Engl.* **34**, 2280 (1995).
31. K. Illgen, T. Enderle, C. Broger, L. Weber. *Chem. Biol.* **7**, 433 (2000).
32. D. R. Halpin and P. B. Harbury. *PLoS Biol.* **2**, E174 (2004).
33. S. Brenner and R. A. Lerner. *Proc. Natl. Acad. Sci. USA* **89**, 5381 (1992).
34. D. R. Halpin and P. B. Harbury. *PLoS Biol.* **2**, E173 (2004).
35. D. R. Halpin, J. A. Lee, S. J. Wrenn, P. B. Harbury. *PLoS Biol.* **2**, E175 (2004).
36. K. Josephson, M. C. Hartman, J. W. Szostak. *J. Am. Chem. Soc.* **127**, 11727 (2005).
37. Z. J. Gartner and D. R. Liu. *J. Am. Chem. Soc.* **123**, 6961 (2001).
38. X. Li and D. R. Liu. *Angew. Chem., Int. Ed.* **43**, 4848 (2004).
39. J. A. Walder, R. Y. Walder, M. J. Heller, S. M. Freier, R. L. Letsinger, I. M. Klotz. *Proc. Natl. Acad. Sci. USA* **76**, 51 (1979).
40. Z. J. Gartner, M. W. Kanan, D. R. Liu. *Angew. Chem., Int. Ed.* **41**, 1796 (2002).
41. Z. J. Gartner, R. Grubina, C. T. Calderone, D. R. Liu. *Angew. Chem., Int. Ed.* **42**, 1370 (2003).
42. Z. J. Gartner, M. W. Kanan, D. R. Liu. *J. Am. Chem. Soc.* **124**, 10304 (2002).

43. Z. J. Gartner, B. N. Tse, R. Grubina, J. B. Doyon, T. M. Snyder, D. R. Liu. *Science* **305**, 1601 (2004).
44. C. T. Calderone and D. R. Liu. *Angew. Chem., Int. Ed.* (2005). In press.
45. X. Li, Z. J. Gartner, B. N. Tse, D. R. Liu. *J Am Chem Soc* **126**, 5090 (2004).
46. T. M. Snyder and D. R. Liu. *Angew. Chem., Int. Ed.* (2005). In press.
47. X. Li and D. R. Liu. *J. Am. Chem. Soc.* **125**, 10188 (2003).
48. M. Rozenman and D. R. Liu. (2005). Submitted for publication.
49. K. Sakurai, T. M. Snyder, D. R. Liu. *J. Am. Chem. Soc.* **127**, 1660 (2005).
50. C.-J. Li and T.-H. Chan. *Organic Reactions in Aqueous Media*, John Wiley, New York (1997).
51. C. T. Calderone, J. W. Puckett, Z. J. Gartner, D. R. Liu. *Angew. Chem., Int. Ed.* **41**, 4104 (2002).
52. M. W. Kanan, M. M. Rozenman, K. Sakurai, T. M. Snyder, D. R. Liu. *Nature* **431**, 545 (2004).
53. R. B. E. L. Woodward, K. P. Nambiar, K. Sakans, B.-W. A.-Y. D. E. Ward, P. Balaram, L. J. Browne, C. H. C. P. J. Card, R. B. ChBnevert, A. Fliri, K. Frobel, D. G. G. H.-J. Gais, K. Hayakawa, W. Heggie, D. H. D. P. Hesson, I. Hoppe, J. A. Hyatt, D. Ikeda, K. S. K. P. A. Jacobi, Y. Kobuke, K. Kojima, V. J. L. K. Krowicki, T. Leutert, S. Malchenko, R. S. M. J. Martens, B. S. Ong, J. B. Press, G. R. T. V. Rajan Babu, H. M. Sauter, M. Suzuki, L. M. T. K. Tatsuta, E. A. Truesdale, I. Uchida, T. U. Y. Ueda, A. T. Vasella, W. C. Vladuchick, R. M. W. P. A. Wade, H. N.-C. Wong. *J. Am. Chem. Soc.* **103**, 3281 (1981).
54. D. Rosenbaum, Y. Brudno, D. R. Liu. (2005). Submitted for publication.