# Convergent automated parallel synthesis

David G. Powers and David L. Coffen

As combinatorial chemistry approaches its second decade of widespread use, the developmental pathways that could shape its future are becoming increasingly apparent. Progress down one of these pathways is driven by the desire to make and test more complex molecules containing intricate ring systems, stereochemical features and elaborate presentations of functional groups – molecules that resemble real drugs and natural products. The realization of this desire requires the recruitment of more of the methods and principles of complex molecule synthesis into combinatorial chemistry, thereby creating hybrid operating systems such as 'convergent automated parallel synthesis'.

Synthetic organic chemistry is a vast, complex and expanding web of interrelated reagents, reactions, procedures, protecting groups, catalysts, purification methods and characterization methods. While it is generally assumed that synthetic organic chemists can make any (sufficiently stable) molecule, complex molecules – particularly those in the related categories of drugs and natural products – tend to present difficult synthetic problems. This ever-present reality drives the constant search for new, better and more sophisticated reactions and methods to apply to such problems. Occasionally, chemists seek ways to achieve productivity gains by leveraging the existent body of knowledge, rather than adding to it incrementally. Such efforts have produced two distinct and powerful sets of levers, those that are conceptual and those that are mechanical. The conceptual levers are largely the result of the efforts of Corey, and are listed as follows[1–3]:

- Computer-assisted planning
- Retrosynthetic analysis
- Convergent pathways
- Electronic search tools for reactions, methods and analogies.

The mechanical levers can be ascribed to the emergence of combinatorial chemistry and may be listed as:

- Parallel processing
- Laboratory automation
- Solid-phase organic synthesis
- Split-and-combine synthesis protocols.

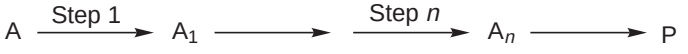
Medicinal and natural products chemists routinely employ the conceptual levers to expedite the synthesis of complicated single compounds, whereas combinatorial chemists typically rely on mechanical levers to synthesize multi-thousand, or even multi-million, compound libraries of simple or modular compounds. We believe that the simultaneous application of conceptual and mechanical leverage has significant potential, and will now review instances where convergence, parallel processing and automation have been applied in concert to support this belief.

## Conventional use of convergence

A linear synthesis starts with a readily available starting material (A), transforms it into intermediate ($A_1$) using the first synthesis step, transforms $A_1$ into $A_2$ using the second step, and proceeds through each of the succeeding steps until the product (P) is attained:

**David G. Powers***, ArQule, Inc., 200 Boston Ave., Medford, MA 02155, USA. *tel: +1 781 395 4100, fax: +1 781 395 1225, e-mail: dpowers@arqule.com. **David L. Coffen**, Discovery Partners International, 11149 North Torrey Pines Road, La Jolla, CA 92037, USA. tel: +1 619 546 1300, e-mail: dcoffen@irori.com

 1359-6446/99/$ – see front matter  PII: S1359-6446(99)01364-1 




There are thousands of examples of linear syntheses. The stereospecific preparation of cortisone from the precursor deoxycholic acid (Fig. 1) is a quintessential example of this type of strategy[4].

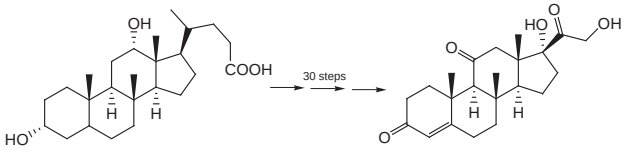



***Figure 1.*** *Illustration of the linear synthesis of cortisone.*

In contrast, convergent syntheses have two or more discrete starting points (materials): A, B, C, etc. Each starting material is carried through an independent series of transformations that lead to a series of intermediates: $A_n$, $B_n$, $C_n$, where $n$ is the number of steps in each series. At the appropriate stages in the synthesis, these intermediates are merged to produce late intermediates or a final product, both of which are usually far more complex than the individual intermediates (Fig. 2).

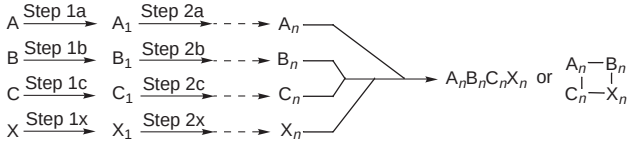



***Figure 2.*** *Flow chart depicting the preparation and confluence of advanced synthetic intermediates (A, B, C…up to X intermediates) with n number of steps in a convergent synthesis.*

The late stages of a synthesis might entail multiple synthetic steps, and convergence of the various series of intermediates might happen at one, two, or several points along the pathway. A useful analogy can be drawn to a system of tributaries feeding a larger and more powerful river. An example of simple convergence using two starting points is found in the case of perindopril, a representative member of the angiotensin-converting enzyme (ACE) inhibitor, antihypertensive class of drugs (Fig. 3)[5]. A further example is the synthesis of vitamin $B_{12}$, one of the most complex molecules ever prepared synthetically,

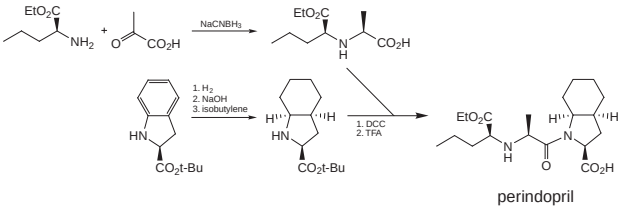



***Figure 3.*** *Convergent synthesis of perindopril.*

which entailed multiple starting points and multiple points of convergence (Fig. 4). The four intermediates – A, B, C and D – were each prepared from commercial chemicals via discrete multi-step sequences and combined in a powerful convergent strategy to obtain cobyric acid, a precursor of vitamin $B_{12}$ (Ref. 6).

The gains in function accessed by convergence include higher overall yields, significantly reduced quantities of material required in the early steps of the synthesis, explicit control of relative stereochemistry in postconvergence intermediates and rapid generation of complexity.

## Parallel processing applied to organic synthesis

In one respect, parallel processing has been applied in organic chemistry laboratories for decades: every chemist who has ever set up two or more reactions to run overnight in a fume hood implicitly understands the basic concept. Contemporary use of the term, however, relates to the simultaneous conduct of hundreds or, more typically, thousands of individual chemical reactions within the context of combinatorial chemistry and carried out under a given set of circumstances (described in the paragraphs that follow). Parallel processing requires that chemists no longer think in terms of individual reagents to be combined in a synthesis step, but, instead, plan synthetic chemistry in terms of reagent sets ('reagent sets' is used interchangeably with 'building block sets'). A reagent set is defined in such a way that each of its members contains a specific functional group pertaining to a specific synthetic reaction and the reactivity profile of any member of this group is comparable to that of all other members of the set. For example, all of the 1,3-dienes capable of participating predictably in a Diels–Alder reaction would constitute a reagent set. For any given reagent set, one or more complementary sets must be identified as well. Thus, all dienophiles capable of participating predictably in a Diels–Alder reaction constitute the reagent set complementary to the diene set. A set of cyclic anhydrides would be complementary to a set of primary and secondary

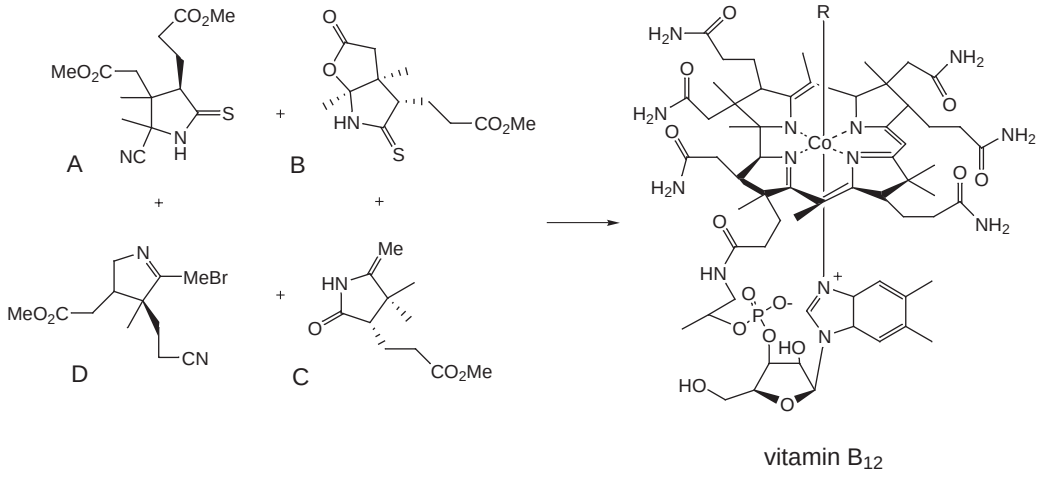



***Figure 4***. *Convergent strategy used in the synthesis of vitamin $B_{12}$.*

amines for the preparation of an amido acid library, a set of acetophenones would be complementary to a set of benzaldehydes for the preparation of a chalcone library (see below) and so on. The structures in Fig. 5 depict a particular Diels–Alder reaction, but a set of ten dienes and a set of ten dienophiles, when crossed in the matrix pattern shown, would result in 100 Diels–Alder adducts, all produced simultaneously by parallel processing.

An obvious but important point is that, while the number of members in a reagent set increases by addition, the number of products increases by multiplication. Thus, while the use of ten dienes and ten dienophiles requires the preparation and manipulation of only 20 stock solutions, the number of products is $10 \times 10 = 100$. When the array chemistry is such that three reagent sets are employed, the number of combinations – i.e. the number of products made – now bears a cubic relationship to the number of members in the sets. Thus, for a reaction of the type A + B + C → P, ten members in each of the building block sets A, B and C (a total of just 30 reagents) afford 1000 compounds in the resultant product array P, e.g. the array derived via the boronic acid Mannich reaction [Fig. 6; Stabile-Harris, M. *et al.* (1998) Proceedings of the *216th National ACS Meeting*, Boston, MA, #229].

It should be noted, however, that it is not necessary for the confluence of the synthetic scheme to occur in a

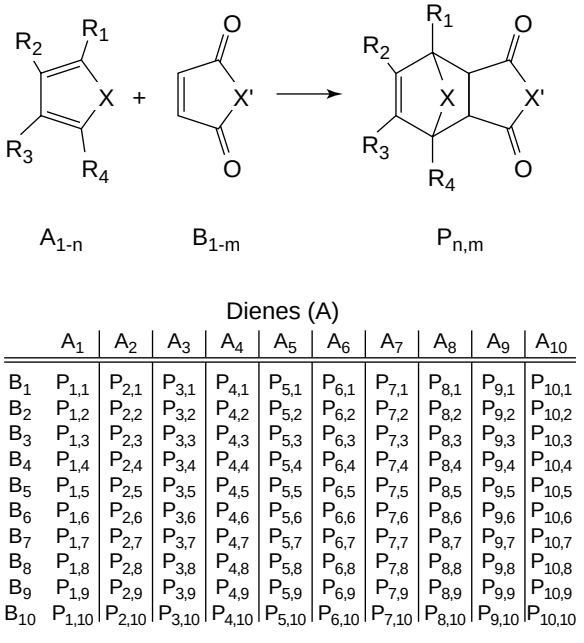



***Figure 5.*** *Illustration of the number of Diels–Alder adducts (P) that would be derived from the reaction (or cross) of all dienes (A) with all of the functionally complementary dieneophiles (B). 10A × 10B = 100 products from 20 reagents.*

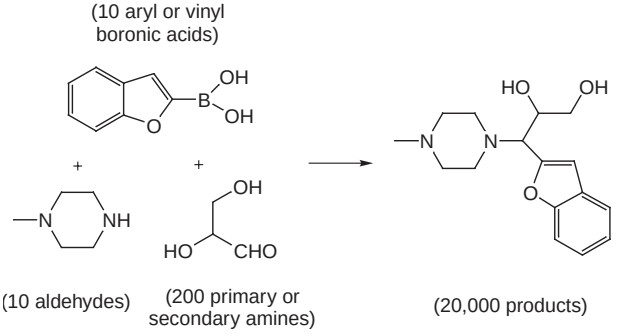



***Figure 6.*** *Boronic acid Mannich condensation. An example of the exponential relationship between the number of starting materials used for a library and the number of resultant products.*

single step. Should the chemistry involved require it, a three-component array will arrive at the same total constituency if two of the reagent sets are first combined to form an array of advanced intermediates, which can then be treated iteratively, in array format, with a functionally appropriate third reagent set. The successful implementation of parallel processing in organic synthesis also requires that reaction conditions that are compatible with all complementary combinations of individual reagents in the sets are defined prior to library production. This requirement places special emphasis on the role of process development. Adaptation of conditions and methods that performed well for a single reaction in the laboratory hood to a method that performs well against a broadly diverse set of reagents on an automated platform is imperative for the successful implementation of this strategy. The development of suitable methods is frequently the most time-consuming part of the overall process. Traditional process development considerations – including solvent or solvent combinations, reaction temperatures, reaction times, agitation rates, the effects of air and moisture, tolerance for non-participating functional groups, chemo-, regio- and stereo-selectivities, the reactivity range of set members, product isolation, purification and appropriate analytical methods – must be rigorously investigated beforehand.

As a final point, the implementation of parallel processing must recognize that one chemist, or even a small group of chemists, cannot physically set up and run thousands of reactions and isolate, purify and analyze thousands of products simultaneously. Traditional laboratory notebooks, sample vialing methods and labeling methods cannot cope with the task of recording thousands of synthetic and analytical results, or track materials. All of this requires extensive deployment of laboratory automation: robotics, programmable fluid-handling devices, automated analytical systems and, most importantly, high-capacity electronic data management systems.

## Automation in the synthetic laboratory

The development of robotic workstations that perform single-unit operations is one strategy that has enabled the implementation of parallel processing in the organic chemistry laboratory[7]. These individual workstations, when taken in aggregate, can be used to perform, in parallel, all of the typical manipulations required in organic synthesis. The first physical step in the preparation of an array is the preparation of equimolar solutions of the reagent sets to be used. The identities of the desired reagents are read into a weighing and dissolution system using a combination of barcode and spreadsheet technologies. The robotics platform first tares a series of vials to be used and stores that information. The production chemist is prompted by the system software with a calculated amount of each material necessary for the array. It is not necessary for the operator to be exact in the measurements as the system reweighs all of the vials, and the software component of the system calculates and drives the delivery of the appropriate amount of solvent necessary to bring each reagent up to a standard concentration. Reagent racks created by this method are then transferred in sequence, according to a synthetic protocol, to a fluid-handling robot whose function is to dispense the reagents in a matrix according to a predetermined method. The method used on the fluid-handling platform defines the reagent combinations found within each reaction vessel in the array. Both the creation of stock solutions at the weighing and dissolution station and the dispensation of the reagents into the individual reaction vessels can be performed under an inert atmosphere. A schematic of the deck of a chemical synthesis workstation is shown in Fig. 7.

The reaction blocks, containing sets of reactions, are then cycled through an appropriate temperature–time profile, as specified by the process development work, using a thermally controlled agitation module. Manufacturing control systems have been implemented to ensure the reproducibility of a method should an array need to be re-prepared. Upon completion, reactions can be concentrated using centrifugal evaporators or put through a work-up protocol as warranted. Postsynthetic work-up procedures

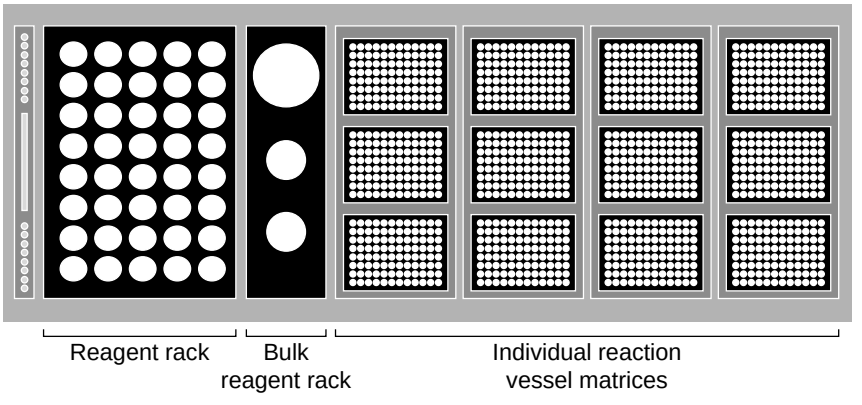



***Figure 7.*** *Schematic of the deck of an automated synthesis workstation.*

that have been automated include filtration of solid by-products or scavenger resins, liquid–liquid extractions and high-throughput preparative chromatography using intelligent fraction collection (mass spectroscopy used as a detection method in combination with a laboratory information management system containing starting material information). After the library has been physically created, high-throughput mass spectral and high-performance liquid chromatography (HPLC) analyses are performed to assess whether purity criteria have been met[8]. Depending on the library size, between 25% and 100% of the library compounds are routinely sampled for analysis. If appropriate, culling of failures and reformatting of the library can be performed.

The simultaneous production of thousands of compounds – hundreds of thousands annually – creates an enormous flow of structural and analytical data that must be correlated with the identity and location of the individual samples. One data-management system that has proven satisfactory is predicated on formatting materials in patterns that mimic a microtiter plate format. The materials are then maintained in this format at every stage of production, inventory storage and in high-throughput screening. Using 96-well plates, each well in a numbered plate is further identified by its row–column intercept. The reagent combinations for that particular address are defined by both the fluid-handling method applied and the particular racks of reagents used. The reagent combinations are captured online and transferred to customized software for generation of a complete compound structure file. This is accomplished using the structures of the reagents together with a suitable algorithm for the chemistry used to combine them. Thus, a structural database for each array is created in which individual structures displayed in electronic form are linked to individual samples through a plate–row–column descriptor, which also defines the precise location of that sample. The database is searchable by this plate–row–column descriptor, thereby providing instant links between screening results, chemical structures, methods of synthesis and analytical data. Systems of this nature can produce in the order of 500,000 spatially addressed, single organic compounds annually.

## Convergent automated parallel synthesis

Having outlined the basic precepts of convergence, parallel processing and laboratory automation, we will now attempt to show, primarily with examples, the powerful synergy to be realized when all three techniques are used in concert. (A powerful example of convergent synthesis performed in parallel, in solution phase, by the method of mixtures, without the aid of laboratory automation can be found in Ref. 9.) Far more complex, polyfunctional, drug-like molecules become readily accessible, with the combinatorial libraries thereby derived frequently featuring carbon–carbon bond-forming processes, stereochemical features and, in some instances, what can be called a natural products 'patina'.

### *Preparation of serine protease inhibitors*

Serine proteases offer attractive molecular targets for the treatment of both autogenous and infectious disease conditions. Two of the several compound classes recognized as general serine protease inhibitors are α-ketoamides and aminomethyleneoxazolones, and an array of compounds incorporating both of these features was desired, e.g. the final product in Fig. 8 (Refs 10–12).

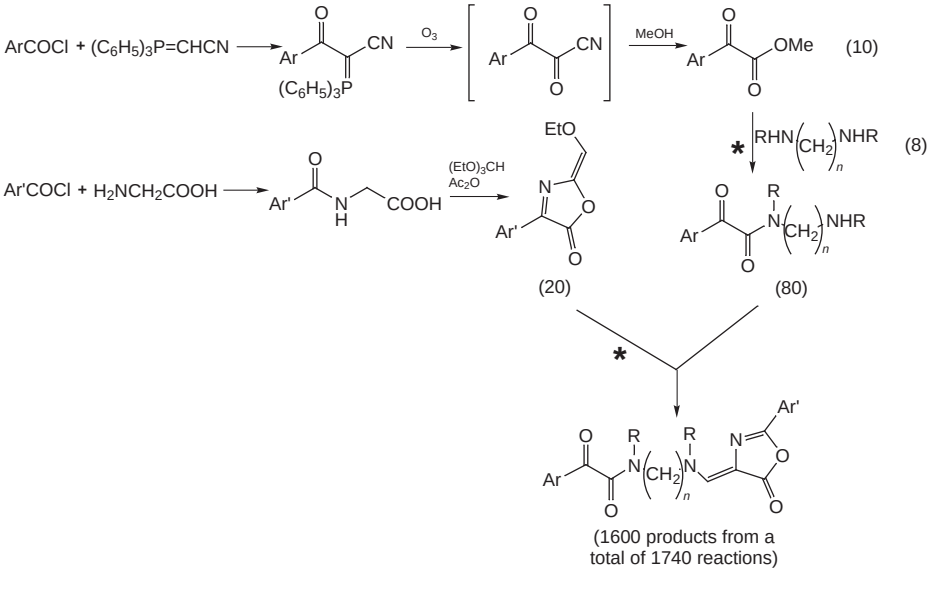



***Figure 8.*** *Convergent strategy applied in the automated preparation of an array of anticipated protease inhibitors. (The figures in parentheses denote the number of possible compounds at each stage of the reaction.) *Denotes automated synthesis step.*

The structural elements Ar, Ar′, R and the linker $(CH_2)_n$ were designed to be independently variable. Execution of this synthesis plan entailed linking the effector molecules with symmetrical diamines; this was made possible by the serendipitous discovery that methyl α-ketoesters provide a method for clean, high-yield monofunctionalization of symmetrical diamines[13]. Preparation of the α-ketoester reagent set was carried out in one branch of this convergent scheme using Wasserman's method[14]. Preparation of the oxazolone reagent set in the second branch was effected using Stammer's procedure[15]. The overall reaction scheme for the synthesis of this array is depicted in Fig. 8.

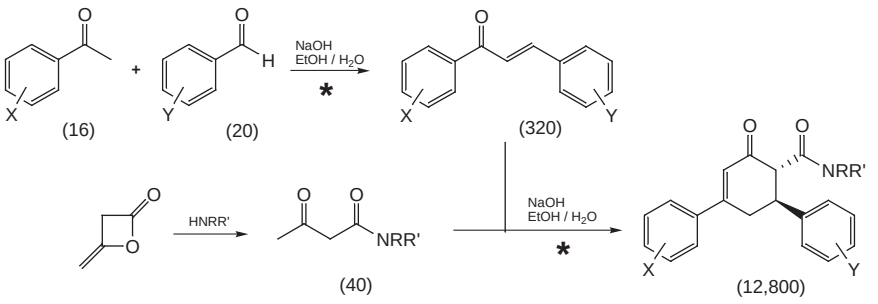



***Figure 9.*** *Convergent strategy applied in the automated preparation of an array of trans-trisubstituted cyclohexenones. (The figures in parentheses denote the number of possible compounds at each stage of the reaction.) *Denotes automated synthesis step.*

The convergent nature of this seven-step scheme allowed for its rapid execution (e.g. six full-time-employee weeks). A linear synthesis of the 1600 compounds would have required far greater effort (7 × 1600 = 11,200 reactions). Once again, it is worth noting the benefit of parallel processing: 1600 compounds were prepared by the manipulation of only 38 reagents (ten α-ketoesters, eight diamines and 20 ethoxymethyleneoxazolones).

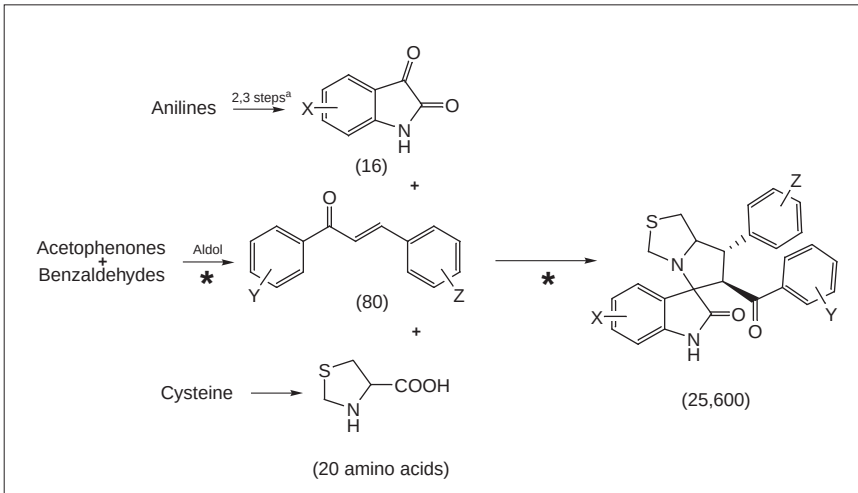



***Figure 10.*** *Convergent strategy applied in the automated preparation of an array of spirocyclic oxindole derivatives. (The figures in parentheses denote the number of possible compounds at each stage of the reaction.) *Denotes automated synthesis step. [a]For further detail, please see Ref. 17.*

### *Preparation of trans-trisubstituted cyclohexenones*

The condensation of primary and secondary amines with diketene readily affords a diverse set of acetoacetamides. This reagent set (40) converges with a complementary reagent set of chalcones (320) prepared by the aldol condensation of acetophenones (16) and aryl aldehydes (20). In the automated convergent step, a total of 12,800 tandem Michael-addition, Robinson-ring annelation reactions were carried out in parallel, giving the *trans*-trisubstituted cyclohexenone library depicted in Fig. 9 (Ref. 16).

### *Preparation of spiro-oxindoles*

The richness of chalcone chemistry in a convergent setting is further illustrated by the preparation of the spiro-oxindole library shown in Fig. 10.

Isatins are cyclic dicarbonyl amides whose chemistry revolves around the reactive nature of the α-carbonyl. The combination of this aldehyde equivalent with an amino acid, such as thiaproline (shown), followed by concomitant decarboxylation, results in the formation of a 1,3-dipole of sufficient reactivity that addition to the recalcitrant dipolarophile chalcone occurs without the aid of Lewis acids[17,18]. Although isatins are available from a variety of synthetic routes, most isatin syntheses start with aniline derivatives (for references and the most recent addition to the repertoire, see Ref. 19). Using a set of 20 isatin analogs, 20 amino acids and approximately 1000 chalcones could result in a library of these analogs that would number approximately 400,000. An array of 25,600 analogs was prepared from 80 chalcones, 20 amino acids and

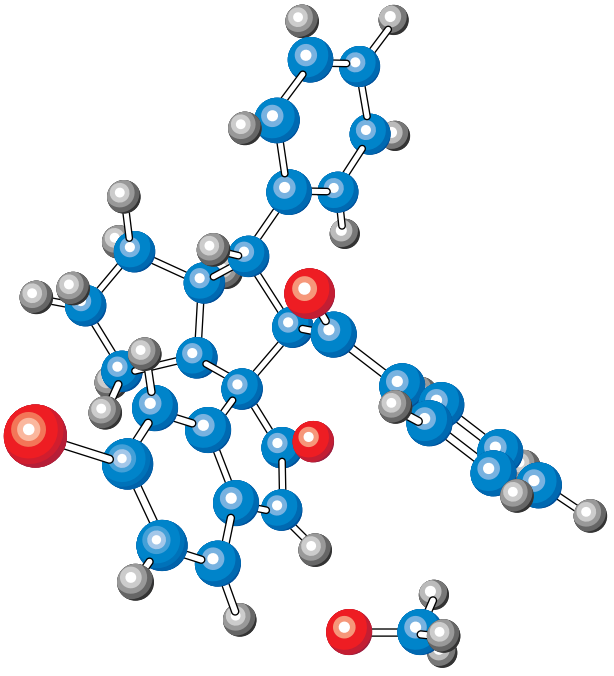


***Figure 11.*** *Single crystal X-ray structure of a spiro[pyrrolidine-2,3′-oxindole]. Each unit cell contains one methanol of crystallization.*

16 isatins[17]. The products were isolated as single racemates where four new stereo-centers had been generated in one transformation. Single-crystal X-ray analysis of the adduct from 5-bromoisatin, proline and chalcone (X = Y = H) proved that both the stereochemical features and the regiochemistry of addition are as shown in Fig. 11.

In conclusion, this last example serves to illustrate how the prudent use of conventional preparative chemistry can provide access to more sophisticated reagent sets, which then converge in multiple combinations to create large libraries of highly complex products. The 25,000 oxindole derivatives produced in this fashion share significant structural homology with the oxindole subset of indole alkaloids, vastly exceeding in number the meagre supply of naturally occurring compounds (<100) and potentially offering pharmacological richness such as that of the oxindole, gelsemine.

## REFERENCES


1 Corey, E.J. (1967) *Pure Appl. Chem.* 14, 1937–1945
2 Corey, E.J. (1971) *Annu. Rev. Chem. Soc.* 455–482
3 Corey, E.J. and Cheng, X.M. (1989) *The Logic of Chemical Synthesis*, Wiley
4 Sarrett, L.H. *et al.* (1946) *J. Biol. Chem.* 162, 601–607
5 Vincent, M. *et al.* (1982) *Tetrahedron Lett.* 23, 1670–1677
6 Eschenmoser, A. and Wintner, C.E. (1977) *Science* 196, 1410–1420
7 Rivero, R.A., Greco, M.N. and Maryonoff, B.E. (1997) in *A Practical Guide to Combinatorial Chemistry* (Czarnik, A.W. and Hobbs DeWitt, S., eds), American Chemical Society
8 Kyranos, J.N. and Hogan, J.C., Jr (1998) *Anal. Chem. News Features* 70, 389–395
9 Boger, D.L., Chai, W. and Jin, Q. (1998) *J. Am. Chem. Soc.* 120, 7220–7225
10 Munoz, B., Giam, C-Z. and Wong, C-H. (1994) *Bioorg. Med. Chem.* 2, 1085–1092
11 Benoiton, N.L., Hudesz, F. and Chen, F.M.F. (1995) *Int. J. Pept. Protein Res.* 45, 266–271
12 Filler, R. and Rao, Y.S. (1997) *Adv. Heterocyclic Chem.* 21, 175–182
13 Baldino, C.M. *et al.* (1997) *Synlett* 488–490
14 Wasserman, H.H. and Ho, W-B. (1994) *J. Org. Chem.* 59, 4364–4369
15 Bland, J.M., Stammer, C.H. and Varughese, K.I. (1984) *J. Org. Chem.* 49, 1634–1636
16 Powers, D.G. *et al.* (1998) *Tetrahedron* 54, 4085–4096
17 Fokas, D. *et al.* (1998) *Tetrahedron Lett.* 39, 2235–2238
18 Aly, M.F., Younes, M.I. and Metwally, S.A.M. (1994) *Tetrahedron* 50, 3159–3167
19 Kraynack, E.A., Dolgardy, J.E. and Gaeta, F.C.A. (1998) *Tetrahedron Lett.* 39, 7679–7682


## In short…

Diversa Corporation (San Diego, CA, USA) and Invitrogen Corporation (Carlsbad, CA, USA) have signed an agreement to exchange technologies and products in specified fields of use. Diversa is to receive an exclusive licence to use Invitrogen's TOPO® technology to clone nucleic acids from mixed populations and uncultured organisms. This technology will enable the efficient cloning of products, bypassing a number of intermediate steps required with current cloning methods. In return, Invitrogen will receive exclusive access to selected proprietary DNA-modifying organisms for use in researching new reagents. Lyle Turner, Invitrogen's Chairman and CEO, said 'Access to Diversa's novel enzymes may lead to the identification and development of reagents and products which may have improved performance characteristics resulting in the next generation of tools for molecular biology and genomics research.'