Solution-phase synthesis in combinatorial chemistry: Applications in drug discovery

Richard Storer

The application of combinatorial chemistry principles to produce libraries of compounds, together with effective use of high-throughput biological screening, is set to revolutionize the way in which chemical leads for drug discovery programmes are generated and evaluated. In this overview, solution-phase synthesis is compared and contrasted with solid-phase synthesis, which remains the method of choice for most combinatorial library work. The author discusses a variety of specific approaches from different groups exemplifying the potential of solution-phase work.

The pharmaceutical industry is currently experiencing a period of intense self-examination to improve the efficiency of its drug discovery programmes. Pressure to be first-to-market with the best compounds requires close examination of all possible means to improve lead generation and to reduce time from lead discovery to identification of development candidates. The number of biological targets to be investigated will increase dramatically, and new approaches, both in chemistry and biology, are essential if such demands are to be met. The capacity to generate and evaluate much larger numbers of compounds of the widest possible diversity will be central to these new initiatives. High-throughput screening and combinatorial chemistry are currently the subject of exponential growth in the area of lead generation[1–5]. Lead evaluation and optimization are also beginning to benefit from the application of multiple-synthesis techniques and from the increasing use of automation in many aspects of organic synthesis.

Library approach and format

Some fundamental choices must be made in deciding which libraries of compounds to prepare and how they might best be synthesized. These decisions will be influenced by whether the library is to be used as a primary library for lead generation or as a focused library for lead optimization, although there is no defined boundary between the two. For lead generation allied to drug discovery, the most direct approach is to synthesize libraries that consist of structurally acceptable starting points for medicinal chemistry programmes. It is possible to control size, functionality and diversity within a library by appropriate selection of the sets of molecular building blocks from which the library is constructed. Decisions regarding library size and whether to construct it as mixtures (pools) or as single entities (discretes) will be influenced greatly by the purpose for which the library is designed. There are advantages and disadvantages with each approach, and specific requirements will determine which is most appropriate in any one case.

Pool generation using solid-phase chemistry uses the 'split and mix' procedure to generate one compound per bead. Generation of pools using solution-phase chemistry can employ mixtures of reactants, but adequate development work is required to ensure the reaction of all constituents under the conditions used. The products obtained are more difficult to analyse chemically for quality-control purposes.

Richard Storer, GlaxoWellcome Research and Development, Medicines Research Centre, Gunnels Wood Road, Stevenage, UK SG1 2NY. tel: +44 1438 763488, fax: +44 1438 763615, e-mail: RS0584@ggr.co.uk

 Copyright © Elsevier Science Ltd. All rights reserved. 1359-6446/96/$15.00. PII: S1359-6446(96)10023-4

The data obtained from screening pools will be less precise or reliable than those generated from testing of discrete compounds. Pools have an advantage when screen capacity is limited and when it would be impractical to generate equivalent numbers of discretes. Where the library is constructed and tested as pools, then a viable plan for 'decoding' to identify individual compounds of interest must be devised as part of the planning process.

The selection of specific target structures and the chemistry required to effect their preparation should ideally be considered together with the final major decision – whether to work in solid or solution phase. Solid-phase synthesis enables better control of purity by allowing extensive washing between steps. The use of excess reagent, when cost allows, can also be employed to drive reactions to completion. These advantages are of particular importance where multiple steps are involved in the synthesis of the target molecules. However, it can be difficult to control stoichiometry, and the use of solid phase imposes some chemical constraints and structural limitations. A suitable functional site is required for attachment to the solid support via an appropriate linker molecule. For most assays, cleavage from the support is required and, for many reactions that are currently employed, a residue resulting from the attachment to the linker is present in the final structure. Much effort is being devoted to expanding the range of developed chemistry for solid-phase work and in devising ways to circumvent the above limitations.

A wide range of reactions is potentially available for solution-phase library generation. In contrast to solid-phase work, no additional steps for attachment to or detachment from supports are required, and thus only the particular reaction of interest need be developed. There are no concerns about compatibility of the chemistry being conducted with either stability on the support or any tagging method employed. Furthermore, the reactions and techniques to be employed in solution-phase library generation are already familiar to all organic chemists, an especially important feature in helping to encourage rapid acceptance of the new approaches and technologies by nonspecialists. For short reaction sequences, in particular, this approach has much to recommend it, and it is easy to generate significant numbers and quantities of compounds in a format that is directly compatible with the requirements for most screens.

Equipment and automation

The specific equipment to be used for library generation and the degree of automation to be employed are important considerations. The principle should be to automate where appropriate; many smaller libraries might most appropriately be generated manually using simple equipment. Equipment suitable for the generation of solid-phase libraries on various supports, such as resins and pins, is commercially available and will continue to develop rapidly. Some specialist companies, such as Affymax[1] (Palo Alto, CA, USA), have developed equipment that is compatible with their own niche technologies. For automated solution-phase library synthesis, a selection of liquid-handling robots that have been custom-modified for library generation are now available. Some commercially available equipment caters for a variety of simple synthetic operations, allowing the preparation of relatively modest numbers (tens) of compounds for focused libraries or building block generation. Specialist companies are actively working to produce purpose-built equipment designed for library generation, and many options will become available commercially in the next few years.

Quality control and decoding

Chemical analysis and quality control are of paramount importance for all libraries. Where solution-phase synthesis of pools is undertaken, the difficulty of adequate chemical analysis increases with the number of components per pool. Although NMR has some utility for very small pool sizes, which can be a useful option during development work, a combination of MS and HPLC has emerged as the most suitable combination of techniques in most cases.

Whichever method is used to construct a pooled library, appropriate decoding methods need to be included in the design to allow access to information about active components. Many elaborate methods have been devised for decoding solid-phase, bead-based libraries, including tagging procedures. Such techniques are facilitated by the fact that pooled libraries prepared by the 'split and mix' approach consist of one compound per bead. Pools made by solution chemistry must either be subjected to fractionation in order to identify the active components in an active pool or there must be some provision in the library design to facilitate decoding.

Solution-phase synthesis of pooled libraries

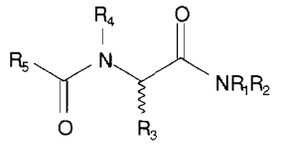
Solution-phase work so far described has included libraries both of pools and of discrete compounds. Where the libraries have been developed as pools, a range of strategies (described below) to provide convenient methods for identification of the active components have been attempted.

A design that would allow a chemical decoding procedure was employed in the 'diamide trimer' library (Figure 1) prepared by the GlaxoWellcome group[6,7] (Stevenage, UK). Amide bond formation was used to produce a trimer library by coupling sets of carboxylic acids and amines to the amine and acid functionality respectively of a central amino acid to give compounds of type (**1**). Using sets of 80 carboxylic acids, 50 amino acids and 40 amines produced a library of 160,000 'diamide trimers'. The library was generated as 4,000 pools of 40 compounds using the 40-amine mixture as the pool origin. The synthesis was carried out, in part, manually but also involved the use of custom robotic equipment for dispensing liquid and subsequent evaporation of solvents and volatile materials. The principle of the chemistry allowed a reversible element to provide for a simple chemical decoding of active pools.

The library was produced in an 80 well per plate format, which paralleled the microtitre format employed for screening. Thus each plate represented a single amino acid, each unique two-dimensional well address on each plate represented a single carboxylic acid, and each well had the same representation of the 40 amine mixture. A number of controls were implemented to ensure that screening data were as reliable as possible. Reactants and coupling agents that were likely to be present in the final products were prescreened to ensure that they did not interfere with the assays. The developed automated procedure was also used to prepare a compound of the library class that was known to have biological activity, and the activity of the product was confirmed. Such controls are a very important feature of the development programme.

The chemistry to decode active pools was undertaken as indicated above to produce the constituents of each pool of interest as discretes for rescreening. In some cases, a range of activities in the components provided some tentative SAR data, but more reliably indicated the most promising compounds to prepare in pure, fully characterized form. In other cases, no compound of significant activity was found, the activity of the pool having resulted presumably from additive weak activities. Good medicinal chemistry lead structures were identified in a number of areas and are exemplified by compounds shown in Figure 2 – an HIV protease inhibitor (**2**; IC_{50} = 10 μg/ml) and an NK_1 antagonist (**3**; pK_i = 7.3).

A variety of groups have worked to devise alternative ways in which to simplify the identification of the active and most-active compounds from pools prepared using solution chemistry. Smith and coworkers[8] investigated an orthogonal approach in a library consisting of amides and esters resulting from reaction of 40 acid chlorides with a set of 40 nucleophiles comprising alcohols and amines. The sets were selected in such a way that the hydrochloric acid liberated during the reaction was scavenged by a tertiary amine centre in the



1

(a)

$A_{mix} + B \longrightarrow A_{mix} - B \longrightarrow A_{mix} - B - C_1$

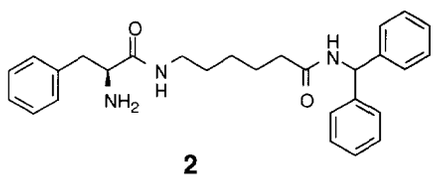
Manual Automated $A_{mix} - B - C_2$

(b)

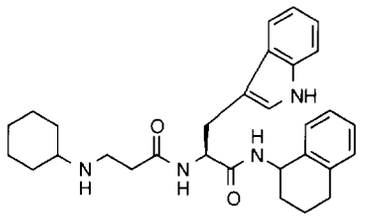
$C_x + B \longrightarrow C_x - B \longrightarrow C_x - B - A_1$

Manual Automated $C_x - B - A_2$

Figure 1. *Principle of reversible decoding. 'A' represents the amine set, 'B' represents the amino acid set and 'C' represents the carboxylic acid set. (a) Manual preparation of feedstock dimers, automated synthesis of final trimers. (b) Decoding reverse direction synthesis, maintain automated synthesis of final trimers.*



2



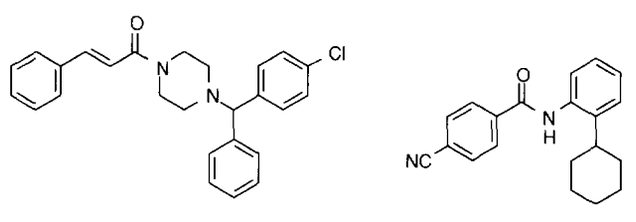
3

Figure 2. *Lead structures from the diamide trimer library: an HIV protease inhibitor (**2**) and an NK_1 antagonist (**3**).*

nucleophile molecule. Each individual nucleophile was reacted with an equimolar mixture of the acid chlorides, and each individual acid chloride was reacted with an equimolar mixture of the nucleophiles. The complete library was thus prepared as 80 sample mixtures, each containing a nominal 40 compounds. The design strategy determined that each possible product should appear in a unique pair of samples. Knowledge of positive samples from screening both halves of the library thus allows the structure of active compounds to be deduced immediately. The potential of this approach was demonstrated by the identification of compounds from the library which provided weak leads with properties of NK_3 antagonism (**4**; IC_{50} = 60 μM) and matrix metalloproteinase-1 inhibition (**5**; IC_{50} of 55 μM) when prepared and tested in pure form (Figure 3).

In a similar approach, Pirrung and Chen[9] have prepared an 'indexed combinatorial library' of 54 carbamates from a set of nine alcohols and a set of six isocyanates. The library was prepared as 15 sublibraries in which each of the alcohols and isocyanates was reacted with an equimolar mixture of the other set of reactants. The product mixtures were screened as inhibitors of acetylcholinesterase from the electric eel, and their activities used as 'indices' to the rows or columns of a two-dimensional matrix reflecting the activities of individual products. Selective synthesis of carbamates from the most active row and column demonstrated that the most active cell in the matrix could be identified using this sublibrary approach. The most potent compound prepared was the previously unknown *O*-succinimidyl *N*-methylcarbamate (**6**; Figure 4). Such a procedure would be generally applicable to any chemistry that can be carried out in a similar modular manner.

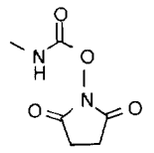
A variation on this theme, where all synthetic work is not carried out at the beginning of the exercise, was provided by



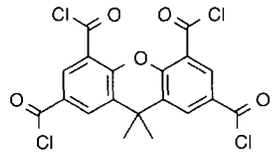
4 **5**

Figure 3. *Amide leads from the orthogonal library approach: an NK_3 antagonist (**4**) and a matrix metalloproteinase-1 inhibitor (**5**).*

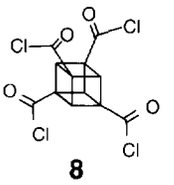
Figure 4. *Novel acetylcholinesterase inhibitor from the indexed library approach.*



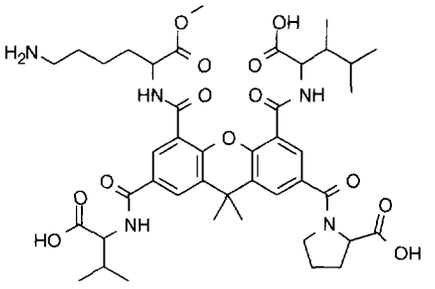
6



7



8



9

Figure 5. *Core templates (**7**, **8**) used by Rebek and coworkers[10,11] in library synthesis and the novel trypsin inhibitor subsequently identified (**9**).*

Rebek and coworkers[10,11]. Their library approach used rigid multifunctional cores and monomer sets with comparable chemical reactivity towards the core molecules (Figure 5). Validation of the chemistry was by preparation of representative small sublibraries. Synthesis of (hopefully) all constituents as complex mixtures is followed by several rounds of iterative sublibrary generation and screening. They employed as cores

the 9,9-dimethylxanthene and cubane tetracid chlorides (**7**) and (**8**) and an initial set of 21 amines. Results from this exercise then determined the constitution of the next sublibrary to be generated. Thus, increasingly potent compounds in progressively smaller pools can be identified. The potential of this approach was demonstrated by the identification from this work of a micromolar inhibitor (**9**) of trypsin.

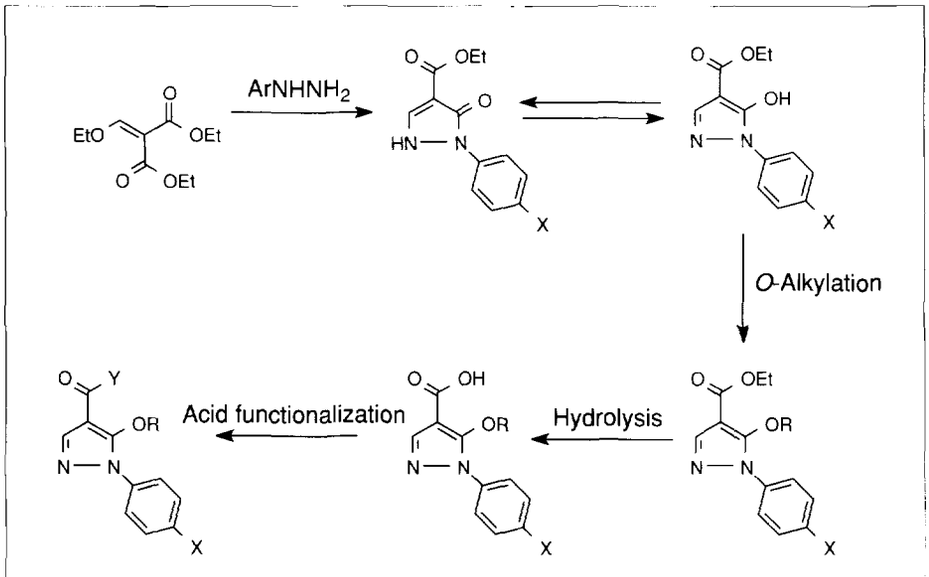
Pooled pyrazole libraries of up to 1,000 members have been described[12] by a group from SmithKline Beecham (Harlow, UK) and were prepared using a variety of reaction types, including amidation, esterification, saponification, Mitsunobu, alkylation, decarboxylation and Claisen rearrangements (Scheme 1). Their policy has been to try to ensure that components are present in equal quantities whilst using multistep sequences and to envisage an iterative procedure for the identification of the structures of lead compounds from the active pools.

Solution-phase synthesis of libraries of single compounds

Where the available capacities allow, generation and screening of libraries of individual compounds rather than pools affords, in addition to the benefits discussed earlier, direct knowledge of the structures of active compounds. Where it is feasible to screen at several concentrations, it is also possible to obtain a more precise indication of which compounds are more active than others. Furthermore, for most SAR work, single-compound synthesis is essential. Where it is desirable to test pools, then robotic mixing of discretes of different structural classes has advantages. This procedure allows for rapid rescreening of the components of an active pool and should reduce the risk of activity in a pool resulting from additive weak activities of the constituents.

The GlaxoWellcome group has devoted considerable effort to extending the range of solution-phase chemistry employed in library generation beyond amide bond formation. The range currently includes 1,4-additions, alkylations, carbonyl/active methylene condensations, halide displacements and carbonyl derivative formation. The objective is to establish a toolkit of useful reactions developed in a suitable form for use in automated synthesis, and these reactions have been used to prepare libraries of a variety of different compound classes, largely as single compounds. The synthesis of a variety of heterocyclic systems has been a high priority[13], and the libraries prepared to date are exemplified by aminothiazoles (**10**) and oxadiazoles (**11**) which include the known anti-inflammatory compound (**12**) and the histamine H_3 antagonist (**13**), respectively (Figure 6). This exemplifies the important principle of including in the library, wherever possible, one or more compounds of the class that are known to have specific biological actions and which thus provide valuable standards when the library is screened with the appropriate assays.

Other groups have developed methods for the synthesis of libraries of single compounds using solution-phase chemistry. The group at Arqule[14] (Medford, CT, USA) has developed libraries based on azlactone chemistry and employing multistep procedures (Scheme 2) to give compounds of type (**14**).

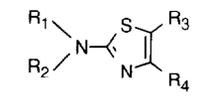


Scheme 1. *Synthesis of pyrazole libraries.*

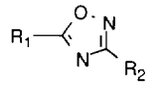
Aminimides have also received attention from the Arqule group. These are zwitterionic substances having no net charge and which are freely soluble in both water and organic solvents. Compounds are prepared in systematic arrays where each member bears a formal structural relationship to its neighbours to facilitate identification of trends in activity. Specific examples of interesting compounds that have been successfully identified from using the approaches of this group include an HIV-1 proteinase inhibitor (**15**; K_i = 100 nM; Figure 7). The reactions employed are largely

three-step sequences, but the group believes that, with suitable chemistry, up to five steps may be possible while retaining adequate control of purity. As further clean, robust reactions are identified, it should be increasingly possible to assemble multistep sequences that can produce solution-phase libraries meeting required specifications.

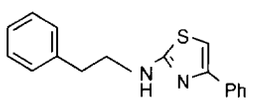
Panlabs[15] (Bothell, WA, USA) are producing libraries commercially using solution-based chemistry. Their approach is to concentrate on the synthesis of libraries of small organic molecules as individual compounds for lead generation or follow-up. They too have developed a toolkit of useful reactions, including the formation of esters, alcohols, amides, ureas, carbonates and carbamates. Reaction validation is by single-compound synthesis and appropriate quality control of a representative sample to a predetermined level, with HPLC and MS being the preferred combination of techniques. In a collaboration with Tripos Inc. (St Louis, MO, USA), they now offer a service for library design using Tripos software and subsequent synthesis using Panlabs technology.



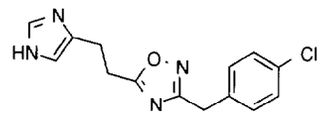
10



11



12



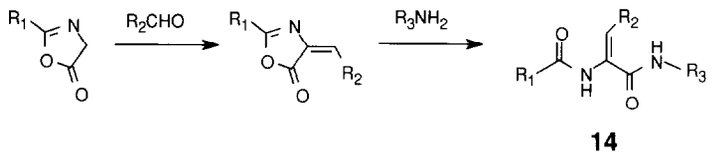
13

Figure 6. *Examples of heterocyclic compounds from array synthesis. Such libraries are exemplified by aminothiazoles (**10**) and oxadiazoles (**11**), which include the known anti-inflammatory compound (**12**) and the histamine H_3 antagonist (**13**), respectively.*

Libraries of single compounds in lead optimization

Most organic synthesis carried out in medicinal chemistry groups in pursuit of lead optimization is currently performed in solution phase. In future, as more leads are generated using solid-phase chemistry, lead optimization using solid phase will become more important. The use of solution-phase array synthesis of individual molecules is rapidly becoming an accepted part of lead optimization programmes. Such focused libraries may be of an appropriate size, from tens to thousands, for a particular requirement. The scale on which the components are prepared will also depend on the status of the project; larger quantities of smaller numbers will be required as the project progresses towards a development candidate. Although it is always desirable to prepare libraries in as pure a form as possible, the essential required levels of purity also depend on the status of the project. In the early stages of lead generation or optimization, lower standards may be sufficient to allow identification of those compounds most worth pursuing in a fully purified and characterized form.

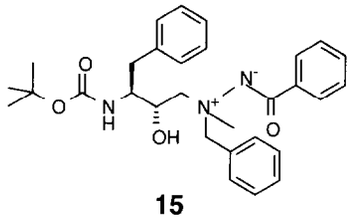
For later-stage optimization, libraries of pure compounds are required for the generation of reliable quantitative data. There are a variety of ways in which this objective can be realized. The identification of robust, clean, high-yielding reactions is clearly of great importance. A great many reactions should prove eminently suitable for automated solution-phase synthesis. Automated purification, for example by preparative HPLC, is also possible, and it is easy to envisage the generation of libraries of compounds to whatever level of purity is appropriate. Other possibilities for improving the quality of solution-phase libraries will continue to be explored. The use of supported reagents (e.g. triphenylphosphine[16]) is clearly a most attractive option, allowing the possibility of removing byproducts and excess reagent easily. In this way it should be possible to enjoy some of the benefits of solid-phase approaches whilst retaining the versatility of solution-phase work.



14

Scheme 2. *Use of azlactones for library synthesis.*

Figure 7. *HIV proteinase inhibitor from an aminimide library.*



15

Conclusions

Solution-phase automated synthesis has a very important role to play in the future of drug discovery both for lead generation and for lead optimization. For the very large pooled libraries, solid-phase chemistry may become the preferred method of generation. Innovative chemical and design ideas may help in decoding pooled solution libraries, but having pools of very similar compounds will always present a problem, and synthesis of discretes and pooling of unlike compounds may be a preferred option. Solution-phase array synthesis, particularly directed towards obtaining pure compounds, will become a very important feature of synthetic chemistry in many areas of endeavour.

Automated solution-phase synthesis will prove to be of great benefit in the area of process research. Virtually all larger-scale work is carried out using solution-phase techniques, and the ability to conduct parallel array syntheses in pursuit of process optimization is potentially very attractive. The successful process research project will define robust, high-yielding reactions that would then become attractive options on which to base future libraries. As lead generation and optimization become more efficient, chemistry scale-up will be one of the areas to be subjected to increased pressure. Optimal use of such in-house expertise will be one way in which such pressure may be minimized.

REFERENCES

1 Gallop, M.A. *et al.* (1994) *J. Med. Chem.* 37, 1233–1251
2 Gordon, E.M. *et al.* (1994) *J. Med. Chem.* 37, 1385–1401
3 Terrett, N.K. (1994) *Tetrahedron* 51, 8135–8173
4 Thompson, L.A. and Ellman, J.A. (1996) *Chem. Rev.* 96, 555–600
5 Patel, D.V. and Gordon, E.M. (1996) *Drug Discovery Today* 1, 134–144
6 Merritt, A.T. (1995) Combinatorial Synthesis Symposium, July 20–21, University of Exeter, UK
7 Storer, R. (1995) IBC Conference on Synthetic Chemical Libraries in Drug Discovery, October 30–31, London, UK
8 Smith, P.W. *et al.* (1994) *Biomed. Chem. Lett.* 4, 2821–2824
9 Pirrung, M.C. and Chen, J. (1995) *J. Am. Chem. Soc.* 117, 1240–1245
10 Carell, T. *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33, 2059–2061
11 Carell, T., Wintner, E.A. and Rebek, J. (1994) *Angew. Chem. Int. Ed. Engl.* 33, 2061–2064
12 MacLachlan, W. (1995) IBC Conference on Synthetic Chemical Libraries in Drug Discovery, October 30–31, London, UK
13 Watson, S.P. *et al. Bioorg. Med. Chem. Lett.* (in press)
14 Hogan, J.C. (1995) IBC Conference on Synthetic Chemical Libraries in Drug Discovery, October 30–31, London, UK
15 Peterson, J.R. (1995) Exploiting Molecular Diversity: Small Molecule Libraries for Drug Discovery, January 23–25, La Jolla, California
16 Bernard, M. and Ford, W.T. (1983) *J. Org. Chem.* 48, 326–332

Contributions to Drug Discovery Today

Drug Discovery Today publishes topical information on all aspects of drug discovery – molecular targets, lead identification, lead optimization and the associated technologies – together with overviews of the current status of compound classes, approaches in specific therapeutic areas or disease states and novel strategies, such as gene therapy. Areas of pharmaceutical development that relate to the potential and viability of new drug candidates are also included, as are those relating to the strategic, organizational and logistic issues underlying pharmaceutical R&D.

Authors should aim for topicality rather than comprehensive coverage. Ultimately, articles should improve the reader's understanding of the field addressed and should therefore assist in the increasingly important decision-making processes for which discovery scientists are responsible.

Most articles appearing in *Drug Discovery Today* are commissioned. However, suggestions and proposals for full reviews or shorter items for the *Editorial*, *Monitor* or *Update* sections are welcomed; in the first instance, a tentative title and brief outline of the proposed article should be supplied. Typically, full reviews will extend to 4,000 words with up to 60 references. *Update* and *Monitor* items (news and views, reports of new technological advances, conferences, experimental methods, and critical assessment of important new literature and other media) do not usually exceed 1,000 words, and one or two figures plus up to ten references may be included. The *Editorial* represents a personal perspective on contemporary issues and controversies affecting R&D and the pharmaceutical industry.