For medicinal purposes

Gregory A. Petsko

Developments in techniques of pharmaceutical chemistry are poised to revolutionize the design of drugs beyond previous expectations. But what are our current expectations and what does the future hold?

Glendower: I can call spirits from the vasty deep.
Hotspur: Why, so can I, or so can any man; But will they come when you do call for them?
Henry IV, Part I

HOTSPUR, here in one of his typically sarcastic moods, had the irritating habit of asking just the right question. We must ask it too, for this is precisely the position in which the science of pharmaceutical chemistry finds itself as it ends its first century and faces the much-heralded millennium. Like Glendower, we claim that we can do many remarkable things. We can find drug targets more easily and in greater abundance than ever before. We can determine the three-dimensional structures of many of these targets, faster than was thought possible a decade ago. We can use that information to design completely new molecules. Sometimes we can even make them. And we can produce vast libraries of 'random' diverse compounds, and screen them, with a speed that chemists and biochemists of the previous generation could not have imagined. But does all this really change, in a profound way, our ability to make drugs?

For almost 100 years, a succession of new technologies have promised cheaper, faster drug development. Medicinal chemistry, protein crystallography, computer graphics and molecular biology have all "called spirits from the vasty deep". Yet the statistics — and economics — of drug discovery and development have remained daunting. For every approved drug (in the United States, anyway) an average of 6,200 chemical compounds will be synthesized, 21 of which will be tested for subacute toxicology. On average, 6.5 of these will be tested in humans and 2.5 will make it into phase III clinical trials (the last stage before approval). The entire process takes 12.8 years — a lifetime for children with diseases such as cystic fibrosis and ataxia telangiectasia — and costs about $350 million.

So why do several of the authors of the articles in this supplement talk as though a revolution had occurred during the past few years? I think it is because a confluence of new technologies has started to transform the way drugs are found and developed. In the accounts that follow, some of the world leaders in these new technologies describe these developments and current state of the art in natural products chemistry, high-throughput screening, combinatorial chemistry, antisense oligonucleotides and structure-guided drug design. What relationship does this have to the Shangri-La of 'rational' drug design? First, I don't like that term very much, because it suggests that all of the great drug discoveries of the past century were irrational. But if rational drug design means the ability to make drugs to order based on structural information about the target, then the new technologies take us very much closer to that goal, but not by the route we originally expected.

A new paradigm arises

Perhaps the most exciting time in science is when paradigms topple, and there is considerable evidence that a paradigm shift is occurring now in drug discovery. The historical approach of medicinal chemistry starting from natural product leads is being supplemented — note that I do not say supplanted — with screening of combinatorial libraries and structure-guided analogue development. Entirely new classes of compounds, such as antisense oligonucleotides, protein therapeutics and target genes themselves, are now open to development. And these diverse technologies are starting to mesh in a way that suggests an irreversible change in the culture.

Natural solutions

Nearly all 'wonder drugs' in use today are derived from natural products. While reminding us of this glorious past, Gregory Verdine, in his article in this supplement[1], points out that the pressures for faster and cheaper drug development are causing a change in the role that natural products will play in the future. Most natural products are extremely complex structurally, making their synthesis, and the synthesis of their analogues, a daunting task, even when cost and yield are not important. But for sheer chemical variety, biologically derived compounds are unrivalled. As Verdine indicates, drug discovery will always begin with attempts to find a molecule that causes a specific biological response: a 'hit'. But the trend is clearly towards rapid, high-throughput screening of large libraries of compounds. Collections of natural products, or of extracts containing them, must be on any such list. One fascinating possibility is that new libraries of natural products can be created by combinatorial shuffling of the genes coding for their biosynthetic enzymes, or by *in vitro* synthesis using mixtures of enzymes.

Apart from the proven track record of natural products as sources of drugs, such compounds are also invaluable springboards to new chemistries. It is doubtful that the β-lactam ring of the penicillins, for example, would have been discovered by synthetic chemists just making molecules at random. And, as Verdine illustrates, natural products are invaluable for defining new drug targets. In addition to taxol*, which showed that microtubule stabilization was a viable approach to cytotoxicity, cyclosporin's role in the identification of the calcineurin signal transduction pathway and the importance of various toxins in revealing the numerous classes of ion channels are just a few of the many recent examples that could be cited. Verdine reminds us that the sources of many of these extraordinary molecules are in danger of extinction, which makes the current efforts on the part of some politicians to gut environmental protection legislation incredibly short-sighted. (As just one example of what might be lost, consider the rain forest in the Hustein Mountains of Papua New Guinea. The Bahinemo and Bitara people who inhabit this region are under great pressure to allow clear-cutting of the kauri pines that cover the mountain slopes. But this forest is home to at least 1,237 types of flowering plants and 392 species of trees.) Fortunately, there remain some unexplored reservoirs of new molecules: the sea, for example, is just starting to be scrutinized.

Seeing through a sieve

In the future, the starting point for the drug discovery process will usually be high-throughput screening of libraries of compounds to find hits. As James Broach and Jeremy Thorner explain in their article in this supplement[2], such screening can often be accomplished without precise identification of a macromolecular target. Increasingly, cell-based assays will be used

*The word 'Taxol' is a registered trademark of Bristol-Myers Squibb, which offers the term paclitaxel instead.

to find lead compounds that have a desired effect, giving immediate information about bioavailability and cytotoxicity. If positive selection can be incorporated into the screen, it is possible to examine literally hundreds-of-thousands of compounds in a few days. One key point emphasized by Broach and Thorner is the desirability of finding surrogate organisms to use in place of human cells to reduce costs and help in the development of selection schemes. They make a compelling case for yeast as the organism of choice. The recent completion of the sequence of the entire yeast genome, plus the ease of genetic and biochemical manipulation in this simple microorganism, reinforce their recommendation. And there are yeast functional homologues for many important human genes, and it is easy to identify them by simple genetic methods[3]. Both Verdine and Broach and Thorner call for the development of assays tailored to produce very specific, target-based readouts. Such an approach, particularly when combined with, say, repertoires of yeast strains systematically constructed to have markers for key pathways, may reduce the need for affinity-based column chromatography methods for target identification.

The right combination

What is molecular diversity? This is a question that the pharmaceutical industry is just beginning to try to answer. What has brought this question to the forefront is the advent of the third enabling technology, combinatorial chemistry. There are two reasons why this technology is changing the culture of drug discovery like no other. First, the rate-limiting step in the development of suitable lead compounds is often synthetic chemistry. Combinatorial methods provide a number of hits very rapidly with little synthetic effort. Second, as the capacity of high-throughput screening increases, there is a concomitant increase in the demand for new compounds to screen. Combinatorial chemistry can provide enormous libraries of such compounds faster and more cheaply than methods based on extracting natural substances. But, as Joseph Hogan reminds us in his article on combinatorial chemistry in this supplement[4], we still do not know how many compounds we need to make — and screen — to ensure our chances of finding one or more hits for any particular target. Hogan offers sound arguments that this question may be irrelevant as phrased. More important than how many compounds one can make, or how much 'diversity space' they cover (however that is estimated), may be what kinds of compounds they are.

The core of the problem?

Hogan has an interesting and, I think, very important angle on this issue. He suggests that the scaffold used in generating the combinatorial library may be as vital as the choice of substituents attached to it. Most combinatorial compound arrays are made by random attachment of various functional groups to a common core. This scaffold has often been chosen for ease in synthesis or low toxicity. But Hogan points out that the scaffold itself can confer biological activity. This is almost certainly true in the case of peptides, which have a range of biological activities that are almost unmatched in chemistry.

How might the scaffold play such an important role? In a recent and, I think, seminal article entitled 'What makes a binding site a binding site'[5], Dagmar Ringe offers an intriguing possible answer. She suggests that the ability of a potential ligand to displace the bound water from a site on the surface of its target molecule is as crucial as the interactions that it can make with that site. The peptide backbone is likely to be quite effective at displacing solvent, in part because it can make the same sort of interactions with proteins and nucleic acids that water does. Scaffolds with similar properties may allow a wide range of targets to be hit with libraries that are small enough for very rapid and inexpensive synthesis and screening.

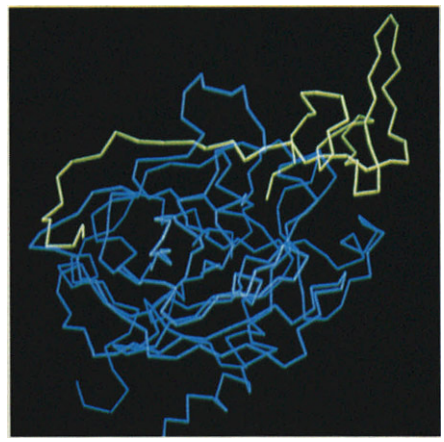

FIG. 1 The crystal structure of the complex of thrombin (blue) with the protease inhibitor hirudin (green). In addition to inserting a few amino-terminal residues into the active site of thrombin (right side), the leech protein hirudin extends its carboxy terminus around the target enzyme to interact with an exosite (left side) a considerable distance away. This double-headed interaction makes hirudin highly specific. See ref. 9 for details.

Combinatorial approaches have undergone an important transformation in recent years. Earlier techniques based on solid-state synthesis that led to mixtures of compounds are being challenged by a new generation of methods that produce large libraries of individual, pure compounds of known chemical structure. The advantages of such defined libraries are obvious. Yet, as Hogan points out, any combinatorial approach has a tremendous additional benefit: screening a diverse array of compounds against a single target gives a huge amount of structure/activity data. And even knowledge of what doesn't work can be of great use in later medicinal chemistry efforts. Doubtless, combinatorial approaches will also be useful in the later stages of drug development, when hits are being transformed into leads and leads are being optimized for bioavailability, low toxicity and so forth. In such cases, the libraries will be directed ones rather than vast diversity arrays.

Making sense of antisense

Of course, nature has been performing combinatorial chemistry for aeons. All biopolymers are formed in this fashion. Pharmaceutical chemists are starting to imitate this strategy to make derivatives and mimetics of these polymers as grist for the screening mill, as well as for specific applications. In their article in this supplement, Matteucci and Wagner[6] examine one example in detail: the antisense oligonucleotides. When these molecules were first introduced in the 1980s, with much fanfare (and just as much dispute over patents), they were trumpeted as the ideal way to inactivate disease-causing genes, particularly in viral genomes. But if the history of science teaches us anything, it should be that no one thing is ever the answer to everything. Antisense oligonucleotides have run into many problems, ranging from surprisingly poor specificity to negligible bioavailability. Matteucci and Wagner do not make light of these problems, but show that the decade of effort into antisense oligonucleotides is beginning to pay off in promising approaches to solutions. That is exciting, because oligonucleotides and their analogues lend themselves naturally to the creation of large random libraries, and it is easy to devise screening methods for selecting among them.

Structure versus function

It seems likely that, as combinatorial chemistry is applied more widely in drug discovery, we will begin to learn the rules for molecular recognition. Nothing would make more of an impact on the final enabling technology discussed in this supplement: structure-based drug design. It is fashionable to criticize the use of structural information to design drugs as slow and costly, yet the bottleneck in the pathway from purified target to a drug is seldom at the step of structure determination, or designing something to fit that structure. Crystallographic and nuclear magnetic resonance methods of structure determination, as Tom Blundell points out in his article[7], have become rapid and inexpensive compared with 10 years ago, and there are excellent computational tools available and under development to aid the design of molecules to fill any specified binding pocket. Calling spirits

from the vasty deep is no problem. Getting them to come — actually making what we design and it being useful — is the real trick. Designs are not leads, and leads are a long way from drugs.

Yet Blundell reminds us that, despite these difficulties, there have been a number of dramatic success stories. One of them, whose impact may reverberate for decades, is the design of effective antiviral agents for HIV based on the crystal structure of its protease. Structural information was invaluable in taking hits and converting them to leads and, eventually, approved drugs. Structural approaches have also been very valuable in those increasing number of cases where the proposed drug is a macromolecule. For example, the successful 'humanizing' of monoclonal antibodies is based on knowledge of their structures. And structural information is vital for another reason: it has become increasingly apparent that analogous compounds do not always bind analogously to the same target. Attempts to deduce structure–function relationships in the absence of structural data are apt to give confusing or misleading information in such cases. But when structures are available, such alternative binding modes can help to map the binding surface.

Room for improvement

There are, however, major areas that remain in need of similar breakthroughs. To my knowledge, there is no small-molecule drug that has yet been designed to disrupt a protein–protein interaction. Such targets will be of increasing importance as our understanding of signal transduction and transcriptional regulation deepens. And we have only recently begun to exploit the possibilities for combination drug therapy — which may be the only good strategy for antiviral agents — thanks to the pressure from the AIDS crisis. It is high time that regulatory agencies realized the need for a more open-minded view of this approach. Yet another opportunity for new developments is in computational tools for identifying binding sites and mapping their characteristics. Nature has already learned to exploit much more of the surface of a target macromolecule than just the active site, as the crystal structure of the complex of thrombin with the anticoagulant hirudin so dramatically illustrates (Fig. 1). Computational tools such as those described by Blundell can locate the sites that hirudin uses, but the correct sites will be merely two among hundreds of potential sites that are identified by scanning the thrombin structure. Recently, Ringe and co-workers have described an experimental approach to mapping the actual binding surface of any crystalline macromolecule[8] (Fig. 2). At the very least, this new method should provide the data needed to guide the development of improved computational techniques for doing the same thing.

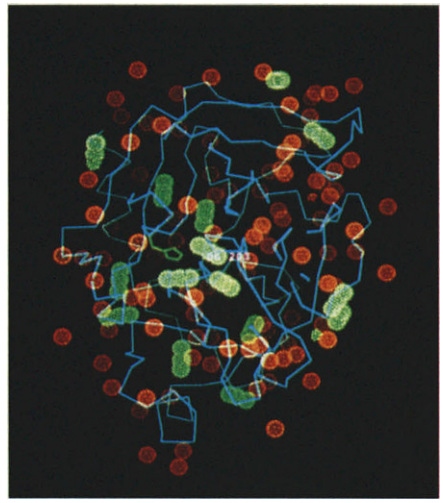

FIG. 2 Crystal structure of the potential drug target elastase (blue) in the water-free organic solvent acetonitrile. The orange spheres are water molecules that remain bound to the protein, even in 100 per cent water-miscible non-aqueous solvent. The green blobs are the positions of acetonitrile molecules that bind to the protein in a small number of specific sites. Similar structures in a variety of organic solvents allow the binding surface of elastase to be mapped out. See ref. 8 for details.

What about the gene as the drug? Although it is true that one could imagine some sort of gene therapy for almost any disease, such technology is clearly in its infancy and faces huge technical and ethical barriers to its widespread application. Undoubtedly it can and will be made to work, especially for some of the more common genetic diseases, but its real impact on the pharmaceutical industry is for the future.

The sceptical chemist

It would be folly to assume that these new technologies will put medicinal chemists out of business, or even dilute their importance. The primary goal of most of the methods I have been discussing is to find lead compounds. There is a world of difference between a lead and a drug. Optimization for low toxicity, good pharmacokinetics, oral bioavailability and other such properties will still require the efforts of skilled synthetic organic chemists. Far from taking away their weapons, these new methods give them ammunition, in the form of more lead compounds and new approaches and guides to altering them.

Quo vadis?

I liken the problem of designing a drug to that of finding a street address in Tokyo. As any visitor to Japan may know, many areas of that old city have no street signs and no house numbers. But if one knew the right neighbourhood, and could knock on every door in that neighbourhood, the problem would be solved. Structural techniques and high-throughput screening offer the hope that we can start our search for new drugs in the right neighbourhood. Libraries of natural products, synthetic oligonucleotides and combinatorial small organic compounds give us the capability to knock on literally thousands of doors. Our designs no longer have to be perfect, or even nearly perfect, the first time. This is the recipe for a revolution.

So I believe we really do stand on the threshold of a new era in pharmaceutical chemistry. To summarize, here is what I think the pattern will be over the next 20 years or so. Target identification will come primarily from genomics and basic cell biology research, aided by natural products that define new pathways and molecules to be inhibited or activated. If there is any structural information about the desired target, or information about some molecules — usually natural products — that are already known to bind to that target, we will make small directed libraries of compounds that will contain one or more high nanomolar leads. If all we have is an assay, we will be able to convert it to a high-throughput screen and fish out several hits from large diversity libraries and pools of natural products. Optimization of these hits into leads will proceed by medicinal chemistry, abetted where necessary by directed combinatorial methods of making analogues and high-resolution structures of lead compound/target complexes.

Development from leads to drugs will follow much the same strategy. Clinical trials will be conducted with several compounds in parallel, with the poorer performers dropping out until the best drug emerges at the end. Even though we still might not quite be able to make 'designer' drugs, I believe that these new tools for drug discovery — all of which derive from basic research, by the way — that are outlined in the articles that follow will enable us to find leads and develop them into drugs two to three times faster than has been possible.

We always could call spirits from the vasty deep. This time, they just might come. □

Gregory A. Petsko is the Director of the Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, Massachusetts 02254-9110, USA.

1. Verdine, G. L. *Nature* **384** Suppl., 11–13 (1996).
2. Broach, J. R. & Thorner, J. *Nature* **384** Suppl., 14–16 (1996).
3. Kranz, J. E. & Holm, C. *Proc. natn. Acad. Sci. U.S.A.* **87**, 6629–6633 (1990).
4. Hogan, J. C. Jr *Nature* **384** Suppl., 17–19 (1996).
5. Ringe, D. *Curr. Opin. struct. Biol.* **5**, 825–829 (1995).
6. Matteucci, M. D. & Wagner, R. W. *Nature* **382** Suppl., 20–22 (1996).
7. Blundell, T. L. *Nature* **384** Suppl., 23–26 (1996).
8. Allen, K. N. *et al. J. phys. Chem.* **100**, 2605–2611 (1996).
9. Vitali, J. *et al. J. biol. Chem.* **267**, 17670–17678 (1992).