References and Notes

1. D. E. Cane, C. T. Walsh, C. Khosla, *Science* **282**, 63 (1998).
2. S. Omura *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **98**, 12215 (2001).
3. H. B. Bode, B. Bethe, R. Hofs, A. Zeeck, *ChemBioChem* **3**, 619 (2002).
4. Y. Tang, T. S. Lee, C. Khosla, *PLoS Biol.*, **2**, 227 (2004).
5. H. D. Mootz, D. Schwarzer, M. A. Marahiel, *ChemBioChem* **3**, 490 (2002).
6. B. Shen, *Curr. Opin. Chem. Biol.* **7**, 285 (2003).
7. C. T. Walsh *et al*., *Curr. Opin. Chem. Biol.* **5**, 525 (2001).
8. T. A. Keating *et al*., *ChemBioChem* **2**, 99 (2001).
9. R. M. Kohli, C. T. Walsh, *Chem. Commun.* **2003**, 293, (2003).
10. R. Pieper, G. Luo, D. E. Cane, C. Khosla, *Nature* **378**, 263 (1995).
11. H. M. Patel, C. T. Walsh, *Biochemistry* **40**, 9023 (2001).
12. D. A. Miller, L. Luo, N. Hillson, T. A. Keating, C. T. Walsh, *Chem. Biol.* **9**, 333 (2002).
13. B. A. Pfeifer, S. J. Admiraal, H. Gramajo, D. E. Cane, C. Khosla, *Science* **291**, 1790 (2001).
14. B. A. Pfieifer, C. C. Wang, C. T. Walsh, C. Khosla, *Appl. Environ. Microbiol.* **69**, 6698 (2003).
15. C. Khosla, P. B. Harbury, *Nature* **409**, 247 (2001).
16. N. Wu, S. Y. Tsuji, D. E. Cane, C. Khosla, *J. Am. Chem. Soc.* **123**, 6465 (2001).
17. S. E. O'Connor, C. T. Walsh, F. Liu, *Angew. Chem. Int. Ed. Engl.* **42**, 3917 (2003).
18. T. A. Keating, C. G. Marshall, C. T. Walsh, *Biochemistry* **39**, 15522 (2000).
19. T. L. Schneider, B. Shen, C. T. Walsh, *Biochemistry* **42**, 9722 (2003).
20. K. Shin-ya *et al*., *J. Am. Chem. Soc.* **123**, 1262 (2001).
21. R. Jansen, H. Irschik, H. Reichenbach, V. Wray, G. Höfle, *Liebigs Ann. Chem.* 759 (1994).
22. R. S. Roy, A. M. Gehring, J. C. Milne, P. J. Belshaw, C. T. Walsh, *Nat. Prod. Rep.* **16**, 249 (1999).
23. L. Du, C. Sanchez, M. Chen, D. J. Edwards, B. Shen, *Chem. Biol.* **7**, 623 (2000).
24. D. Konz, A. Klens, K. Schorgendorfer, M. A. Marahiel, *Chem. Biol.* **4**, 927 (1997).
25. N. J. Hillson, C. T. Walsh, *Biochemistry* **42**, 766 (2003).
26. S. A. Sieber, M. A. Marahiel, *J. Bacteriol.* **185**, 7036 (2003).
27. T. Brautaset *et al*., *Chem. Biol.* **7**, 395 (2000).
28. C. N. Boddy, T. L. Schneider, K. Hotta, C. T. Walsh, C. Khosla, *J. Am. Chem. Soc.* **125**, 3428 (2003).
29. J. W. Trauger, R. M. Kohli, H. D. Mootz, M. A. Marahiel, C. T. Walsh, *Nature* **407**, 215 (2000).
30. J. W. Trauger, R. M. Kohli, C. T. Walsh, *Biochemistry* **40**, 7092 (2001).
31. R. M. Kohli, J. W. Trauger, D. Scwarzer, M. A. Marahiel, C. T. Walsh, *Biochemistry* **40**, 7099 (2001).
32. R. M. Kohli, C. T. Walsh, M. D. Burkart, *Nature* **418**, 658 (2002).
33. R. M. Kohli, M. D. Burke, J. Tao, C. T. Walsh, *J. Am. Chem. Soc.* **125**, 7160 (2003).
34. S. A. Sieber, J. Tao, C. T. Walsh, M. A. Marahiel, *Angew. Chem. Int. Ed. Engl.* **43**, 493 (2004).
35. H. D. Mootz *et al*., *J. Am. Chem. Soc.* **124**, 10980 (2002).
36. Z. Hojati *et al*., *Chem. Biol.* **9**, 1175 (2002).
37. S. Weist *et al*., *Angew. Chem. Int. Ed. Engl.* **41**, 3383 (2002).
38. B. Wilkinson *et al*., *Chem. Biol.* **7**, 111 (2000).
39. Y. J. Yoon *et al*., *Chem. Biol.* **9**, 203 (2002).
40. Y.-Q. Cheng, G.-L. Tang, B. Shen, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 3149 (2003).
41. D. Panda, K. DeLuca, D. Williams, M. A. Jordan, L. Wilson, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 9313 (1998).
42. P. R. August *et al*., *Chem. Biol.* **5**, 69 (1998).
43. T.-W. Yu *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **99**, 7968 (2002).
44. A. Rascher *et al*., *FEMS Microbiol. Lett.* **218**, 223 (2003).
45. B. K. Hubbard, C. T. Walsh, *Angew. Chem. Int. Ed. Engl.* **42**, 730 (2003).
46. K. C. Nicolaou, C. N. Boddy, S. Brase, N. Winssinger, *Angew. Chem. Int. Ed. Engl.* **38**, 2096 (1999).
47. M. Oliynyk *et al*., *Mol. Microbiol.* **49**, 1179 (2003).
48. Y. Sun *et al*., *Chem. Biol.* **10**, 431 (2003).
49. H. J. Kwon *et al*., *Science* **297**, 1327 (2002).
50. A. M. van Wageningen *et al*., *Chem. Biol.* **5**, 155 (1998).
51. J. Recktenwald *et al*., *Microbiology* **148**, 1105 (2002).
52. J. Pootoolal *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **99**, 8962 (2002).
53. M. Sosio, S. Stinchi, F. Beltrametti, A. Lazzarini, S. Donadio, *Chem. Biol.* **10**, 541 (2003).
54. M. Sosio *et al*., *Microbiology* **150**, 95 (2004).
55. D. Bischoff *et al*., *Angew. Chem. Int. Ed. Engl.* **40**, 4688 (2001).
56. A. Holtzel *et al*., *J. Antibiot.* **55**, 571 (2002).
57. Z. A. Hughes-Thomas, C. B. W. Stark, I. U. Bohm, J. Staunton, P. F. Leadlay, *Angew. Chem. Int. Ed. Engl.* **42**, 4475 (2003).
58. W. Liu *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **100**, 11959 (2003).
59. J. L. Hansen *et al*., *Mol. Cell* **10**, 117 (2002).
60. D. Hoffmeister *et al*., *Chem. Biol.* **7**, 821 (2000).
61. C. Mendez, J. A. Salas, *Trends Biotechnol.* **19**, 449 (2001).
62. L. L. Remsing *et al*., *J. Am. Chem. Soc.* **124**, 1606 (2002).
63. D. Hoffmeister, G. Dräger, K. Ichinose, J. Rohr, A. Bechthold, *J. Am. Chem. Soc.* **125**, 4678 (2003).
64. H. C. Losey *et al*., *Biochemistry* **40**, 4745 (2001).
65. H. C. Losey *et al*., *Chem. Biol.* **9**, 1305 (2002).
66. C. Walsh, C. L. Freel Meyers, H. C. Losey, *J. Med. Chem.* **46**, 3425 (2003).
67. A. Li, J. Piel, *Chem. Biol.* **9**, 1017 (2002).
68. H. Chen *et al*., *Biochemistry* **40**, 11651 (2001).
69. F. Pojer, S.-M. Li, L. Heide, *Microbiology* **148**, 3901 (2002).
70. H. Chen, C. T. Walsh, *Chem. Biol.* **8**, 301 (2001).
71. M. G. Thomas, M. D. Burkart, C. T. Walsh, *Chem. Biol.* **9**, 171 (2002).
72. C. R. Hutchinson, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 3010 (2003).
73. Supported by NIH grants GM 20011, GM 49338, AI 42738. Special thanks to D. Vosburg for preparation of artwork.

VIEWPOINT

Organic Chemistry in Drug Discovery

Malcolm MacCoss[1]* and Thomas A. Baillie[2]

The role played by organic chemistry in the pharmaceutical industry continues to be one of the main drivers in the drug discovery process. However, the precise nature of that role is undergoing a visible change, not only because of the new synthetic methods and technologies now available to the synthetic and medicinal chemist, but also in several key areas, particularly in drug metabolism and chemical toxicology, as chemists deal with the ever more rapid turnaround of testing data that influences their day-to-day decisions.

Numerous changes are now occurring in the pharmaceutical industry, not just in the way that the industry is perceived, but also in the rapid expansion of biomedical and scientific knowledge, which affects the way science is practiced in the industry. The recent changes in the way that synthetic chemistry is practiced in this environment center around new scientific advances in synthetic techniques and new technologies for rational drug design, combinatorial chemistry, automated synthesis, and compound purification and identification. In addition, with the advent of high-throughput screening (HTS), we are now faced with many targets being screened and many hits being evaluated. However, success in this arena still requires skilled medicinal chemists making the correct choices, often with insight gleaned from interactions with computational chemists and structural biologists, about which "hits" (*1*) are likely to play out as true "lead" (*1*) structures that will meet the plethora of hurdles that any drug candidate must surmount.

In the recent past, the usual flow of information that was generated regarding any new compound prepared in the laboratory of a drug discovery company followed a paradigm similar to that shown in Fig. 1. This scheme was driven by the need to get the initial information on a compound first, before deciding whether its properties met appropriate criteria before moving onto the next evaluation step. Such a linear sequence of events, although sparing of the number of compounds taken down the pathway, often meant that a considerable amount of time passed (several weeks) before it was known whether a particular change in a molecule was in fact a useful transformation, or whether it was a potency-enhancing change in the primary in vitro assay but was perhaps a liability in a downstream evaluation. Thus, the delay in getting appropriate feedback to the synthetic chemist meant that decisions about which molecules to prepare in the next round of synthesis were not guided by input from downstream data. With the advent of faster synthetic technologies, including advances in nuclear magnetic resonance (NMR) methods,

[1]Department of Basic Chemistry, Merck Research Laboratories, 126 East Lincoln Avenue, Rahway, NJ 07065, USA. [2]Department of Drug Metabolism, Merck Research Laboratories, Sumneytown Pike, West Point, PA 19486, USA.

*To whom correspondence should be addressed. E-mail: malcolm_maccoss@merck.com

rapid separations, and automated syntheses, the cycle time for synthetic manipulation of analogs has decreased dramatically. In addition, in the same time frame, advances have been made in the ability to assay compounds, both in vitro and in vivo, at a much greater speed than was previously possible, and so the current paradigm has shifted toward that shown in Fig. 2, where it is now feasible to generate a tremendous amount of relevant data on a newly synthesized compound within 1 week of its initial preparation. This process allows for a much better-informed set of decisions, as one considers the next round of molecules that need to be prepared.

It should be stressed that an awareness of the potential downstream obstacles to successful drug development is an important consideration in the chemist's decision-making process. Based on a rationalization of experimental and computational approaches, Lipinski *et al.* presented the "rule of five" in the mid 1990s, which is an excellent working hypothesis for predicting good druglike properties in new compounds (*2*, *3*). Thus, close attention needs to be paid to molecular weights, as well as to the physicochemical properties of lead molecules, such as lipophilicity (logP) and aqueous solubility (which will affect oral bioavailability and the feasibility of generating a parenteral formulation), together with animal pharmacokinetics, which can be extrapolated with caution to predict corresponding behavior in humans. The latter is particularly important in providing some assurance that the candidate drug molecule will exhibit linear pharmacokinetics in humans, with appropriate dose size and elimination characteristics for the intended route and frequency of drug administration.

Preliminary absorption, distribution, metabolism, and excretion (ADME) studies of lead compounds in animal species also provide information on routes of clearance (such as renal, biliary, or metabolic), which is helpful in guiding the selection of compounds that exhibit a balance between elimination pathways and thus would not be unduly dependent on a single organ for excretion. At the same time, in vitro data are provided on the interaction of drug candidates with human cytochrome P-450 (CYP) enzymes, so that CYP inhibitors and inducers are identified at an early stage, and due consideration is given to the attendant risk that such candidates may cause drug-drug interactions in the clinic. In cases where oxidative metabolism by CYP enzymes is likely to be an important mechanism of drug clearance in humans, it is preferable to have contributions from multiple isoforms, as opposed to a single CYP (again, to minimize the potential for drug-drug interactions), whereas it is particularly undesirable for metabolism to be catalyzed solely by an enzyme, such as CYP2D6 or CYP2C19, that exhibits genetic polymorphism (potentially leading to large individual variability in drug pharmacokinetics and clinical response where metabolism is the major route of clearance). Moreover, if the therapeutic target resides within the central nervous system (CNS), it becomes important to determine whether the structural series of interest serve as substrates for the efflux transporter P-glycoprotein and thereby are denied access to brain tissue in vivo. By obtaining such information in the discovery phase, potentially serious liabilities in a given structural series become evident at the outset, and informed decisions can be made accordingly to redirect chemistry efforts.

The chemist also needs to be conversant with issues of toxicology, given that the primary cause of failure of drug candidates in early development continues to be preclinical toxicity. Although the potential for genotoxicity can be assessed directly through a number of in vitro assays, the same does not hold true for end-organ toxicities (such as drug-induced liver damage) or immune-mediated toxicities (idiosyncratic reactions) (*4*). However, based on the premise that some (but certainly not all) drug-related adverse events appear to be mediated by a chemically reactive, electrophilic metabolite or metabolites, as opposed to the parent drug itself, it may be argued that the generation of such electrophiles is an undesirable feature of any drug candidate. By means of appropriate in vitro "trapping" experiments and assessments of covalent binding of lead drug candidates to protein, both in vitro and in vivo, it is usually possible for the medicinal chemist, working closely with colleagues in drug metabolism, to identify routes of metabolic activation and, through appropriate structural modification, to minimize this potential liability (*5*). Moreover, before selection of a lead compound for development, information also will be available from in vivo studies in animals aimed at assessing selected off-target pharmacological activities of the compound of interest, including effects on the CNS and cardiovascular systems. It is true that different pharmaceutical companies generate and weigh the above types of information to different extents in selecting lead candidates for progression into development. At Merck, all of the above characteristics are taken into account in arriving at this key decision, which requires considerable experience and sound judgment on the part of the group of senior scientists collectively charged with this responsibility.

This new paradigm has led to a different type of decision-making by chemistry group leaders. As noted above, the results from a preliminary evaluation of the pharmacological, pharmacokinetic, metabolic, and toxicological profile of a series of molecules usually will expose any serious deficits that would hinder or even preclude successful development of a drug candidate. As a result, such "flawed" compounds, or sometimes entire structural series, are dropped from further consideration, and development resources are conserved as a result.

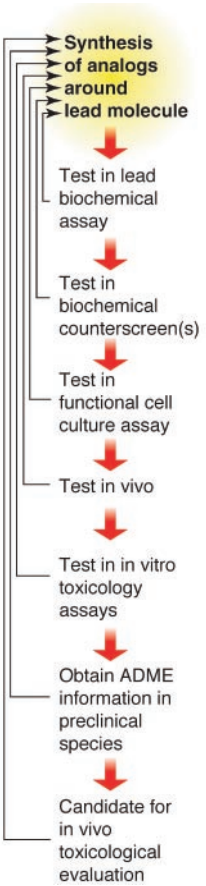

Fig. 1. Linear path to drug candidate, with numerous feedback loops designed to provide information for target selection in the next round of synthesis.

In the majority of cases, however, there is no single factor that would lead to the exclusion of a molecule from further consideration, and the decision to advance a given compound needs to be based on a critical assessment of the relative attributes and potential liabilities of that molecule. Admittedly, the availability of more, rather than less, information on each drug candidate can introduce an element of ambiguity into the chemist's decision-making process. For instance, if a structural change leads to increased potency in the lead biochemical assay, but the compound is less orally bioavailable in a rodent, has more activity on a biochemical counterscreen, and is less potent in a toxicity assay, then the decision to continue exploring that avenue is less clear.

Of course, all knowledge is useful and so the ongoing detailed compilation of structure-activity relationships (SARs) across many assays is already helping our understanding of what types of functionality are responsible for binding to various CYPs, cardiac ion channels, transporters, nuclear receptors responsible for CYP induction, etc. In fact, the medicinal chemist has always had to make judgments regarding such data, but in the current environment the task is to make such decisions rapidly and to know how to weigh the data as they come in from different sources. These decisions can, of course, also be influenced by the nature of the target itself, because a tolerance for a particular toxicity might well be different for diseases with such different profiles as obesity versus a particular cancer.

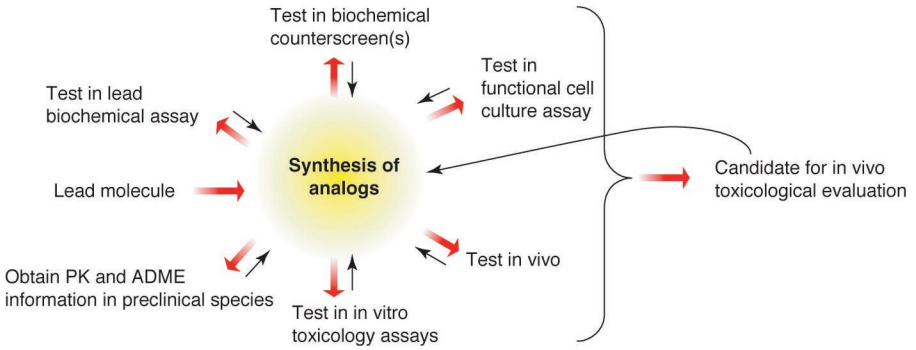

Fig. 2. Nonlinear time-optimized path to drug candidate, with numerous feedback loops designed to provide optimal information on the next round of synthesis.

At Merck, as at several other pharmaceutical companies, we have found that the most fruitful approach to the selection of new drug candidates is to identify the key issues of a lead compound, based on early screening data, and then to focus on minimizing these deficiencies by informed chemical intervention, bearing in mind SAR data for the pharmacological target. For example, potent CYP inhibition in a lead compound may be localized to a single functional group, which may then be replaced by a noninhibitory substituent. Likewise, CYP induction (for example, through activation of the nuclear transcription factor PXR), metabolism to a reactive electrophile, or unwanted cardiovascular activities (for example, ion channel activity that may lead to adverse cardiac effects in vivo, such as that reflected by prolongation of the QT interval on an electrocardiogram) (*6*) may be traced to specific structural motifs that can be successfully engineered out of the lead structure. This multidisciplinary approach to drug discovery, with organic chemistry serving as the cornerstone of the process, is far removed from the linear paradigm of former years (Fig. 1).

Thus, while many new technologies such as combinatorial chemistry, rapid analog synthesis, automated synthesis, open access liquid chromatography mass spectrometry, and high-speed automated high-performance liquid chromatography (to name but a few) are now affecting medicinal chemistry, their main effect has been to shorten the cycle time of synthetic operations. This, in turn, has led to a profound difference in the way in which a medicinal chemistry project progresses through the system. Different companies have embraced these new technologies in different ways (*7*–*10*). For instance, some invested heavily in the mid-1990s in combinatorial chemistry and made this technology a key driver of their efforts to discover new leads and to expand their existing sample collections, particularly when traditional sources of compounds failed to deliver new leads. Others have used these technologies in appropriate projects and have forged alliances with smaller companies that specialize in such efforts, thus freeing up their internal operations to use their historical institutional knowledge of medicinal chemistry, but now guided by more information, as depicted in Fig. 2. This approach has led to more outsourcing of research medicinal chemistry than was common practice a few years ago (*11*–*13*).

It should be noted that pharmaceutical companies have sample collections filled with molecules that were prepared many years ago for old discovery programs. Even if those molecules did not advance the program for which they were initially made, they were designed at the time by medicinal chemists in the hope of interacting with some type of proteinaceous domain (such as an enzyme, heterotrimeric G protein–coupled receptor, ion channel, etc). It is not unusual, therefore, for these molecules to be the starting point of new medicinal chemistry programs when they show up as hits in a new HTS screen. Thus, because of the rapid synthetic cycle times, a medium-sized group of medicinal chemists can now advance several different lead classes at the same time and thus potentially shorten the timelines for developing a hit or lead into a true drug candidate. Usually, it is not clear at the start of a project what the downstream toxicological, metabolic, or off-target liabilities of a particular lead class are likely to be, and so different structural classes can now be investigated simultaneously to allow for data-driven decisions.

When experienced medicinal chemists are asked to reflect on why various programs were advanced more quickly than others, they will invariably agree that it was because of the nature and quality of the starting hit or lead. One of the most difficult properties to build into a newly discovered lead molecule is the desired pharmacokinetic (PK) profile, particularly in the case of orally dosed compounds. In recent years, the resources available for early PK evaluations in rodents have been increased, both for single compounds and, where appropriate, with the use of cassette dosing methods (*14*). Such rapidly obtained information on newly synthesized compounds is one of the most important factors in the quest to shorten the times from lead molecules to drug candidates.

One must constantly be aware that the rapid synthesis of large numbers of molecules that are laden with ADME, physical property, or toxicological shortcomings may provide intriguing hits or leads, but they may not shorten the time to the elaboration of such a hit into a drug candidate. In fact, as noted above, most experienced medicinal chemists would prefer to start in a structural series that has inherently good ADME properties, albeit with poor potency on the target receptor, and then set about improving the potency on the target, rather than working in the other direction (starting with a potent molecule that requires modification to optimize ADME and toxicological properties, which requires optimization of several, often opposing, structural parameters within the predefined tight structure-activity boundaries required for potency), although the history of drug discovery is replete with examples of both. A good recent example of this situation from our laboratories has been the development of the orally active substance P antagonist EMEND (aprepitant) (*15*) (Fig. 3).

Merck and many other companies have worked in this area for many years. The field was stimulated in 1991 by the discovery of CP-96,345 by Pfizer scientists, which showed that a potent subnanomolar small molecule could selectively antagonize substance P at the NK-1 receptor (*16*). However, because of the difficulty in advancing structurally related molecules through the drug development process, presumably due largely to off-target activities, metabolism issues, and the need to penetrate the CNS, it took more than a decade before a small molecule was identified that had the appropriate properties to be a drug, and EMEND was launched by Merck in 2003 for the treatment of both acute and delayed-phase chemotherapy-induced nausea and vomiting. Based on our experiences and knowing the large number of other companies working in this area, it is very likely that tens of thousands of molecules have been prepared in

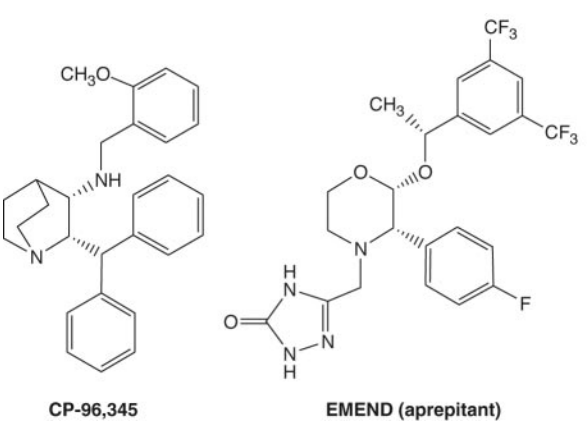

Fig. 3. Structures of CP-96,345 and EMEND (aprepitant).

the past decade as substance P antagonists, a large percentage of which likely exhibited subnanomolar potency at the NK-1 receptor, but only one has made it to market. This example highlights the difficulty of and resources needed to optimize the ancillary properties of potent inhibitors/antagonists so that they can become safe, viable medicines.

In the drug discovery process, we must also be cognizant of the interrelatedness of academic, government, and industrial research in the development of new drug entities. Despite large research budgets, the biomedical research carried out by pharmaceutical companies still represents only a small percentage of the overall worldwide research effort on diseases and approaches to their treatment. Academic and government laboratories, funded with public monies, often provide much basic research and fundamental insight into diseases that can direct researchers toward novel ways of attacking diseases. However, they are rarely organized (nor is it their mission) to embrace the drug discovery process in the multidisciplinary fashion outlined above that is the modern paradigm by which new hits or leads are first identified and then get transformed into new viable medicines. All of the above discussion speaks to one of the most important issues facing discovery medicinal chemistry today: the continuing need for excellent synthetic chemists. In large pharmaceutical companies, the drug discovery process is driven by multidisciplinary teams made up of the very best experts in each discipline, and chemistry is one key element in this. These teams have ready access to experts in other areas of biomedical science, and although chemists often end up as group leaders of discovery efforts, that usually occurs after much experience has been gained in the drug discovery process. The recent advances discussed above have put more tools in the chemist's toolkit, but in order to use these tools effectively, it invariably comes down to the ability to make the absolutely "correct" molecule in a timely and cost-effective manner. This process requires the very best organic chemistry skills, and we must continue to provide funding in the university system for training in these core skill sets to chemists in their graduate and postdoctoral studies if we are to continue to provide the very best in medicines for what is becoming an aging population.

References and Notes

1. In this discussion, a "hit" is defined as a nonoptimized structure obtained from some screening process on a target protein. It is often a very weak binder and is likely to have a nonoptimized pharmacokinetic profile. A "lead" is defined as a structure that has been derived from an early "hit" and, although still not fully optimized, has been shown to have some appropriate characteristics to be a precursor of a drug entity. Often a good lead will have shown some proof-of-concept activity in an in vivo pharmacological model, but will likely not have been fully optimized for pharmacokinetic properties or undesirable off-target activities.
2. C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney, *Adv. Drug Delivery Rev.* **23**, 3 (1997).
3. In the discovery setting, the rule of five (2) predicts that poor absorption or permeation of drugs is more likely when a drug molecule possesses either (i) more than 5 hydrogen bond donors, (ii) 10 hydrogen bond acceptors, (iii) a molecular weight greater than 500, or (iv) a calculated logP greater than 5.
4. J. Uetrecht, *Drug Discov. Today* **8**, 832 (2003).
5. D. C. Evans, A. P. Watt, D. A. Nicoll-Griffith, T. A. Baillie, *Chem. Res. Toxicol.* **17**, 3 (2004).
6. See (*17*) for an excellent review of the cardiovascular effects manifested by QT interval prolongation and the evaluation of drug candidates for this parameter.
7. See the cover story in *Drug Discov. Dev.* **6**, 30 (2003).
8. T. Koppal, *Drug Discov. Dev.* **6**, 59 (2003).
9. A. DePalma, *Drug Discov. Dev.* **5**, 50 (2002).
10. A. DePalma, *Drug Discov. Dev.* **6**, 51 (2003).
11. M. McCoy, J.-F. Tremblay, *Chem. Eng. News* **81**, 15 (2003).
12. A.M. Rouhi, *Chem. Eng. News* **81**, 75 (2003).
13. T. Koppal, *Drug Discov. Dev.* **6**, 22 (2003).
14. W. A. Korfmacher *et al.*, *Rapid Commun. Mass Spectrom.* **15**, 335 (2001).
15. J. J. Hale *et al.*, *J. Med. Chem.* **41**, 4607 (1998).
16. R. M. Snider *et al.*, *Science* **251**, 435 (1991).
17. R. Netzer, A. Ebneth, U. Bischoff, O. Pongs, *Drug Discov. Today* **6**, 78 (2001).

REVIEW

The Many Roles of Computation in Drug Discovery

William L. Jorgensen

An overview is given on the diverse uses of computational chemistry in drug discovery. Particular emphasis is placed on virtual screening, de novo design, evaluation of drug-likeness, and advanced methods for determining protein-ligand binding.

"Is there really a case where a drug that's on the market was designed by a computer?" When asked this, I invoke the professorial mantra ("All questions are good questions."), while sensing that the desired answer is "no". Then, the inquisitor could go back to the lab with the reassurance that his or her choice to avoid learning about computational chemistry remains wise. The reality is that the use of computers and computational methods permeates all aspects of drug discovery today. Those who are most proficient with the computational tools have the advantage for delivering new drug candidates more quickly and at lower cost than their competitors.

However, the phrasing of the question suggests misunderstanding and oversimplification of the drug discovery process. First, it is the rare case today when an unmodified natural product like taxol becomes a drug. It is also inconceivable that a human with or without computational tools could propose a single chemical structure that ends up as a drug; there are far too many hurdles and subtleties along the way. Most drugs now arise through discovery programs that begin with identification of a biomolecular target of potential therapeutic value through biological studies including, for example, analysis of mice with gene knockouts. A multidisciplinary project team is then assembled with the goal of finding clinical candidates, i.e., druglike compounds that are ready for human clinical trials, which typically selectively bind to the molecular target and interfere either with its activity as a receptor or enzyme. Molecular libraries are screened, and the resulting leads are optimized in a cycle that features design, synthesis and assaying of numerous analogs, and animal studies. Crystal structure determination for complexes of some analogs with the biomolecular target is often possible, which enables "structure-based drug design" (SBDD) and the efficient optimization of leads. The success of SBDD is well documented (*1*, *2*); it has contributed to the introduction of ~50 compounds into clinical trials and to numerous drug approvals. Minimally, the role of computation here is in the structure refinement using simulated annealing (*3*), development of the underlying molecular mechanics (MM) force fields, structure display, and building and MM evaluation of analogs. All top pharmaceutical companies have substantial structural biology and computational chemistry groups that are intertwined and participate on the project teams.

There is usually much "tweaking" toward the end of the preclinical period of drug discovery when a series of compounds

Department of Chemistry, Yale University, New Haven, CT 06520-8107, USA. E-mail: william.jorgensen@yale.edu