Progress and issues for computationally guided lead discovery and optimization

William L. Jorgensen

INTRODUCTION

Since the late 1980s there have been striking advances, fueled by large increases in both industrial and NIH-funded academic research, that have revolutionized drug discovery. This period has seen the introduction of high-throughput screening (HTS), combinatorial chemistry, PC farms, Linux, SciFinder, structure-based design, virtual screening by docking, free-energy methods, absorption/distribution/metabolism/excretion (ADME) software, bioinformatics, routine biomolecular structure determination, structures for ion channels, G-protein-coupled receptors (GPCRs) and ribosomes, structure/activity relationships (SAR) obtained from nuclear magnetic resonance (SAR by NMR), fragment-based design, gene knockouts, proteomics, small interfering RNA (siRNA), and human genome sequences. The result is a much-accelerated progression from identification of biomolecular target to lead compound to clinical candidate. However, a serious concern is that the dramatic increase in drug discovery abilities and expenditures has not been paralleled by an increase in FDA approvals of new molecular entities.[1] High demands for drug safety, broader and longer clinical trials, too much HTS, too little natural products research, and effective generic drugs for many once-pressing afflictions have all been suggested as contributors.[2–4] Numerous corporate mergers and acquisitions may have also had adverse effects on productivity through distractions of reorganization and integration. Nevertheless, one should consider what the success would have been in the absence of the striking technical advances. Certainly, progress with some critical and challenging target classes such as kinases would have been greatly diminished, and the adverse impact on many cancer patients would have been profound. Indeed, further gains in the treatment and prevention of human diseases must require even more emphasis and commitment to fundamental research. As in other discovery enterprises, the answer is to drill deeper.

The topic of this volume focuses on one of the areas in drug discovery that has seen major transformation and progress: structure- and ligand-based design. The design typically features small molecules that bind to a biomolecular target and inhibit its function. The distinction stems from whether a three-dimensional structure of the target is available and used in the design process. Structure-based design can be carried out with nothing more than the target structure and graphics tools for building ligands in the proposed binding site. However, additional insights provided by evaluation of the molecular energetics for the binding process are central to most current structure-based design activities. Ligand-based design does not require a target structure but rather stems from analysis of structure/activity data for compounds that have been tested in an assay for the biological function of the target. One seeks patterns in the assay results to suggest potential modifications of the compounds to yield enhanced activity. The upside is that a target structure is not required; the downside is that substantial activity data are needed. My research group has focused on the development and application of improved computational methodology for structure-based design. Some of the experiences and issues that have been addressed are summarized in the following.

LEAD GENERATION

Both lead generation and lead optimization may be pursued through joint computational and experimental studies. As summarized in Figure 1.1, our approach has evolved to feature two pathways for lead generation, de novo design with the ligand-growing program BOMB (Biochemical and Organic Model Builder)[5] and virtual screening using the docking program GLIDE.[6] Fragment-based design, which involves the docking and linking together of multiple small molecules in a binding site, is another popular alternative.[7,8] Desirable compounds resulting from de novo design normally have to be synthesized, whereas compounds from virtual screening of commercial catalogs are typically purchased. In both cases, it is preferable to begin with a high-resolution crystal structure for a complex of the target protein with a ligand; though the ligand is removed, it is advisable to start from a complex rather than an apo structure, which may have side chains repositioned to fill partially the binding site. An extreme example occurs with HIV-1 reverse transcriptase (HIV-RT) for which the allosteric binding site for nonnucleoside inhibitors (NNRTIs) is fully collapsed in apo structures.[9]