Biotechnology and Drug Discovery: Two Years Into the Third Millennium

D. Bussiere* and K. Krause**

*D. Bussiere, Chiron Corporation, 4560 Horton Street, Emeryville, CA., USA, dirksen_bussiere@chiron.com
** K. Krause, Department of Biology and Biochemistry, University of Houston Houston, TX., USA, kkrause@uh.edu

ABSTRACT

The field of drug discovery is extremely complex and ever changing, with new methodologies and technologies being introduced constantly. The following paper provides an introduction to the field as well as the cutting-edge work being done in it, with special attention paid to structure-based drug design. It also serves to introduce the speakers who will be presenting at the computer-aided drug design session of ICCN 2002.

Keywords: Drug discovery, drug design, structure-based drug design, pharmaceutical discovery.

1 INTRODUCTION

The start of the 21st century finds the fields of biotechnology and drug discovery at the beginning of a true paradigm shift. In the last few years, the sets of carefully-planned, single experiments which built the foundation for most of the work done in these industries over the last twenty years are rapidly being replaced by high-throughput methodology in several scientific areas, including (but not limited to) DNA and genomic sequencing, protein analysis, high-throughput screening for inhibitor compounds, nuclear magnetic resonance (NMR), and X-ray crystallography [1]. The sheer amount of data generated by these new high-throughput methods will also require new computational methods of mining, processing, and interpreting the data, to allow the scientists to isolate the data that is relevant to the problem at hand. In addition, there is work underway to perform as much of the discovery and the development of new drug entities as possible using computational methodologies. The driving force behind this *in silico* approach is that such techniques are usually far more cost- and time-effective.

To effectively visualize these advancements, compare a traditional drug discovery cycle (Figure 1) used by many companies until the early-1980 s with the current biological target-based drug discovery cycle (Figure 2) used by most pharmaceutical and biotech companies today. In the traditional cycle, an initial lead compound is found by isolating a molecule (or other form of isolate, such as an organic extract from a plant or cell lysate) with a certain biological activity; often this was done by ethnobotany or by some lucky happenstance, as in the discovery of penicillin [2]. Modifications of this lead compound were planned using clues provided by a relatively crude structure-activity relationship (SAR) analysis (often just a visual alignment of available effective chemical structures) or by modification of the compounds using traditional medicinal chemistry techniques. The new, thus-modified compounds were then synthesized and retested. The cycle continued until the particular biological activity of the compound was maximized. This cycle, while often successful (many of our current drugs were discovered this way), was not rapid and did have major drawbacks. For example, what if the compound of interest could not be separated from the initial isolate? Or, as in the case of taxol, what if the isolated compound is so complex that it can not be modified chemically or even synthesized in large amounts for several years? Also, because this cycle exists independent of a known biochemical mode of action, the mechanism of drugs developed this way was often a mystery. A perfect example of this is aspirin. First commercially marketed in 1899, it was not until the early to mid-1990 s that scientists were able to determine that aspirin exerted its affect by inhibiting the COX-2 enzyme [3].

Contrast the traditional cycle to the current biological target-based cycle (Figure 2) [4]. Recognizing that many drugs are antagonists (inhibitors) of biological macromolecules-typically protein enzymes, the industry now chooses a representative biological target molecule before a drug discovery project is begun. The target molecule is a macromolecule that is crucial for the biological activity or process that is to be inhibited. In some cases, target selection is simply a matter of common sense and examination of the results from basic research. For example, human immunodeficency virus-1 (commonly known as HIV) is expressed as a single polypeptide within an infected host cell. This polypeptide is then processed by a virally-encoded protease; the processed proteins are then packaged and the virus lyses the infected cell. It was therefore easy to correctly predict that the HIV protease (the only one encoded by the virus) was critical in virus maturation and development. This has led to several highly effective therapeutics for HIV [5]. It is not always

Computational Nanoscience and Nanotechnology 2002, (www.cr.org), ISBN 0-9708275-6-3.

that simple, however, especially when the biological activity to be inhibited is not parasitic in nature, as it is in the case of a virus, or when the biochemical pathway to be manipulated has tens to thousands of inter-connected parts. An example of this is the complex metabolic signaling network within each of our cells. In these networks, kinases, which add phosphate groups to certain amino acids, and phosphatases, which remove phosphate groups from certain amino acids, effect their particular enzymatic function in response to various stimuli (such as the actions of other kinases and phosphatases). This phosphorylation and de-phosphorylation turns various enzymes off and on. The high degree of inter-connectivity and complexity in such networks makes target determination very difficult, but given the importance of these signaling networks in disease states, there is research being undertaken on such networks. One of the authors, Dirksen Bussiere, will be discussing target selection and drug discovery against such a kinase target for diabetes in his talk entitled, The Development of Selective GSK3 Inhibitors . Dr. Bussiere will also be speaking on the industrial approach to solving problems in structure based drug design. Certainly the only real successes to date in SBDD have come from pharmaceutical companies. This is due, in no small part, from the reported $250 million investment required per marketable drug. Only pharmaceutical companies have these resources to apply to SBDD problems. As Dr. Bussiere guides us through the current state-of-the-art methodologies used in industry, he will help us spot those industrial advances that might help aid academic programs devoted to drug design. As advances are communicated between academic and industrial settings, the cost per drug should decrease. Then over the next decade cost savings in the drug design cycle will allow both industry and academics to target projects that previously were felt to be too expensive.

The basis of target selection is quite complex. Often it relies upon the results of gene knockouts and other biological manipulations to examine the results of inhibiting a particular target. Increasingly, however, genomics, bioinformatics, and the vast database of genomic sequence data are used to select suitable targets and to predict the activities of the as-of-yet uncharacterized genes and proteins [6].

2 TARGET SELECTION AND HIGH-THROUGHPUT SCREENING

Assuming that a project team is able to select a suitable target, several technologies come into play. First the gene of the target must be cloned, expressed and purified. The initial lead compound(s) is then isolated by a variety of techniques grouped under the term of high-throughput screening, where hundreds to millions of compounds are rapidly examined for binding to and inhibiting the purified target. Until quite recently, this was done by a high-throughput modification of the typical enzyme assay, where the readout was colormetric in nature or by the detection of the transfer of radiation in each well of a 96 or 384 well plastic dish. Increasingly, complex biophysical techniques are being used for such screening. These techniques offer the added feature that they are able to detect compounds that bind efficiently to the molecule without necessarily filling the active site of the enzyme. Two such techniques utilize methodologies previously used to find the structures of macromolecules, these being X-ray crystallography and Nuclear Magnetic Resonance (NMR). Screening by crystallography [7] or SAR by NMR [8], both developed at Abbott Laboratories, allow semi-high-throughput and direct visualization of the binding of compounds in the environment of the macromolecule. The price of such detail is the added expense of the methods and the increased need for large amounts of purified macromolecule. A new biophysical method which appears promising is to be discussed by Dr. Holger Ottleben, an international expert on the use of surface plasmon resonance imaging of chemical microarrays used for drug discovery. Dr. Ottleben states, Graffinity Pharmaceuticals AG has pioneered the generation and application of chemical microarrays. Chemical microarrays are arrays of small organic compounds and represent a novel approach towards the analysis of chemical libraries. We are currently world wide the only company that is able to produce chemical microarrays and to apply this technology for drug discovery. Chemical microarrays can be used to analyse the interaction of proteins with organic compounds in a miniaturised and high-throughput fashion. Parallel SPR imaging allows the direct analysis of binding events without the need of reporter systems or tags and is therefore suited for function blind screening.

His presentation will discuss, the requirements for successfully screening ligand protein interactions on arrays of immobilised organic compounds and will reflect on the importance of surface chemistry for microarrays. In addition, a description of the chemical contents and diversity which can be presented on chemical microarrays will be given along with case studies on the results of SPR-imaging of chemical microarrays against different drug targets. Finally, the perspectives for microarray aided drug discovery will be outlined.

3 COMPUTATIONAL CHEMISTRY AND VIRTUAL SCREENING

Perhaps one of the most ambitious undertakings in the last 30 years is the ongoing attempt to model chemical systems *in silico*. These simulations began with relatively simple inorganic or organic molecules. The field began quite humbly in the late 1960 s with a simulation of a box of 216 waters on one of the fastest computers of the time[9]. Now, such simulations often involve proteins of

Computational Nanoscience and Nanotechnology 2002, (www.cr.org), ISBN 0-9708275-6-3.

10-30 kiloDaltons of molecular mass immersed in boxes or spheres of thousands of water molecules and are run on desktop workstations or personal computers. Similarly, the toolbox of the computational chemist now includes techniques for estimating the strength of binding of a small molecule to a biological macromolecule, the ability to do complex electrostatic calculations on macromolecules, and techniques which allow the modeling of novel macromolecules from other template structures of some homology. There are many other techniques, far too numerous to list here, as the field is very dynamic and approximates the advancement seen in personal computing in terms of evolution in available techniques.

Additionally, major advancements have been made in the computational modeling of absorption, distribution, metabolism, and excretion (ADME) predictions for small molecule compounds [10]. While it is currently possible to discover and further improve potential inhibitors to be better binders to a macromolecule, it is still quite difficult to predict which compounds will be able to survive the effects of an organism s metabolism without being rapidly being cleared. The development of algorithms to predict molecules that have suitable ADME parameters is of critical importance to the pharmaceutical industry.

One of the major advancements in computational chemistry came from the work of Brian Shoichet and Irwin Kuntz and their coworkers at University of California-San Francisco. Underpinning the development of structure based drug design is an idea fostered by the German organic chemist Emil Fischer that enzymes and substrates (or inhibitors) often fit together like a lock and key. Translating this simple two-dimensional notion into the complex three-dimensional world of proteins and organic molecules has resulted in the development of software that docks inhibitor candidates into a target molecule s active site, optimizing the fit. In this way thousands of candidates can be screened *in silico* before any bench synthesis is begun. Shoichet and Kuntz were able to develop a program that can dock a small-molecule compound into the active site of a target macromolecule using a scoring function that incorporated goodness of fit of the compound based on physical forces [11]. In recent years, this technique has advanced significantly and is now used in almost every drug discovery effort. Our speaker in this area, Dr. John Irwin comes from a laboratory specializing in the development of this type of software. He is delivering a paper entitled. Molecular Docking and Drug Design that also appears in this volume.

4 STRUCTURAL BIOLOGY AND STRUCTURE-BASED DESIGN

In the last decade, structural biology has transformed the face of biomedical research, including that of drug discovery. Due to the improvement of equipment and experimental methodology, as well as the advent of cloning and protein expression, and the availability of rapid, inexpensive computers, the determination the spatial coordinates of all (or most) of the atoms in a biological macromolecule has become commonplace. The aforementioned techniques of X-ray crystallography and NMR are the most widely used techniques, but others, such as neutron diffraction, do exist. These techniques allow us to view the molecules that make up a living organism at an extraordinary resolution. To solve a macromolecular structure (be it protein, DNA, or RNA) by NMR involves specific isotopic labeling of an overproduced form of the molecule, collection of data by various pulse sequences, and assignment of resonances within that data [12]. Likewise, to solve a novel macromolecular structure by X-ray crystallography involves overproduction of the native molecule, as well as specific selenomethionine labeling of the molecule in a manner similar to that used in NMR. This is followed by collection of the appropriate data sets and generation of initial electron density maps [13]. With the initial electron density maps in hand, either a scientist, or if the data is of high-quality, a computer, determines and refines the structure. Solving a structure by either of the major techniques used to be very time consuming, but recent developments in the field of structural genomics promise to shorten this time to days or weeks for some proteins. Structural genomics seeks to automate most of the aforementioned steps, essentially, industrializing the two techniques. These new automatic pipelines will be used to generate many such novel protein structures in a high-throughput manner. It is very likely, however, that several proteins and protein complexes will require very complex biochemistry to generate suitable samples for NMR or crystallography and that the structure of these proteins will be solved in the traditional manner.

The typical structure-based drug design cycle incorporates all of the aforementioned technique to allow a project team to rapidly develop potent inhibitors of a target molecule. The three-dimensional structure of the macromolecule is solved by X-ray crystallography or NMR (crystallography is the more common) either in the presence or absence of initial lead compounds, which are discovered by high-throughput screening or by basic research. Once the structure of the target has been obtained, it is also possible to perform virtual screening of the target to obtain information on which compounds would favor the binding sites contained within the target; this allows one to construct biased libraries for high-throughput screening. Regardless of how initial lead compounds are obtained, one must realize that the design cycle is iterative (Figure 3). The modeling methodologies are not yet accurate enough to allow one to model a compound into its binding site in a target and accurately predict its interactions with the target and the local solvent. Thus, one must have a structure of the current best lead compund bound to the target as a starting point to each round of the cycle. Modeling can then be used to propose

Computational Nanoscience and Nanotechnology 2002, (www.cr.org), ISBN 0-9708275-6-3.

potential compounds with improved properties that are then synthesized by chemists on the project team. These compounds are then assayed for activity and compounds that show the desired characteristics can then be brought through the next round of the cycle. The cycle is not flawless: it is not uncommon, in one of the author s experience, for 30-70% of the designed compounds (depending on the project) to not show an improvement in binding, which is why several compounds are designed for each iteration of the cycle. However, the ability to see the binding of the compound via a structural technique does speed the drug discovery process significantly.

For the remainder of the session we have gathered a group of researchers with complementary but quite distinct approaches to research in structure based drug design. As mentioned previously, in any SBDD project the single most important decision made is the proper choice of target molecule. Simple accounting makes this plain. If the average bacterial genome contains 1000 — 2000 genes and if only 10% of those genes are potentially drug design targets, then a drug design project target all of these genes would cost more than 25 billion dollars. Culling the herd of targets is, therefore, needed. Speaking on the subject of target selection, Dr. James Musser will describe an approach to this problem that involves filtering the targets through successively more stringent criteria. First, total bacterial genes are mapped, next secreted proteins are identified and finally those proteins that evoke an immune response are elucidated. As expected the complete bacterial genomes are studied using modern computational techniques. Then, complementary nucleotides for all of the pathogen s genes can be embedded onto DNA chips. These chips can be used to pan the cytosolic pool of RNA to identify those genes that are *transcribed* during conditions that mimic infection. This experiment results, in effect, in the production of a genome wide Northern blot. The proteome of a given pathogen can be further narrowed by screening the antibodies produced during infection to identify those proteins that are targeted by the host for antibody production. Dr. Musser argues that antibodies are more likely to target those proteins that are important in the pathogenic process, and thus strong drug design targets.

Dr. James Sacchettini heads a large academic group working on developing new therapies for important infectious diseases including tuberculosis. He notes that, *Mycobacterium tuberculosis* infections are responsible for one in four avoidable adult deaths in developing countries. While there are a number of effective strategies available for treating tuberculosis, current therapies are greatly complicated by the long duration of drug therapy necessary to kill persistent bacteria. For example, the long multi-drug regimen commonly used to treat tuberculosis results in greater than 50% noncompliance, in developing countries. Dr. Sacchettini s studies focus on the discovery of new anti-tuberculosis agents that target persistent mycobacteria. These agents have the potential to reduce the duration of drug treatment. In this work we combine genetics, structural biology and cell biology to provide the identity of enzymes essential for maintenance of a chronic mycobacterial infection. Several proteins have now been identified and characterized. The three dimensional structures of these drug targets have provided a template for the design of new inhibitors, which now are being tested for *in vitro* and *in vivo* activity against persistence models.

Dr. Timothy Palzkill designs inhibitors to β-lactamase using the powerful combination of phage display coupled with *in situ* peptide synthesis. He states, Protein-protein interactions are involved in most biological processes and are important targets for drug design. Over the past decade, there has been an increased interest in the design of small molecules that mimic functional epitopes of protein inhibitors. The production of TEM-1 β-lactamase is one of the most common mechanisms of bacterial resistance to β-lactam antibiotics. The β-lactamase inhibitory protein (BLIP) is a 165 amino acid protein that is a potent inhibitor of TEM-1 β-lactamase (K_i = 0.1 nM). To aid in the development of new inhibitors of β-lactamase, the gene encoding BLIP was randomly fragmented and DNA segments encoding peptides that retain the ability to bind TEM-1 β-lactamase were isolated using phage display. In addition, a phage display library containing random sequence peptides was also screened for sequences that bind to β-lactamase. A number of candidate peptides were isolated from the phage display libraries and found to weakly inhibit β-lactamase (K_i ~ 500 μM). The binding affinity of several peptides was optimized using SPOT synthesis to screen >1,000 peptide variants for increased binding affinity. The combination of phage display and SPOT synthesis has facilitated the design of peptides that mimic the BLIP-β-lactamase protein-protein interaction and thereby inhibit β-lactamase activity.

Dr. Kurt Krause, also an author on this paper, is the principal investigator on an academic drug design project focussed on a single target molecule, alanine racemase. Bacteria use alanine racemase to catalyze the synthesis of D-alanine, an essential component of the cell wall in Gram-positive bacteria, Gram-negative bacteria, and mycobacteria. Alanine racemase is a proven drug target for successful antibiotic design, because it has been repeatedly shown that inhibition of this enzyme will result in the death of the target bacteria. In fact, one antibiotic that targets alanine racemase, cycloserine, is available clinically. This antibiotic was developed prior to any structural information about ALR and its clinical use is limited due to frequent and severe side effects. Using structure based drug design, we hope to design more specific and potentially less toxic inhibitors. Because of increasing resistance to traditional antibiotics, it is important to explore new options for treating bacterial infections. This project involves collaboration between groups in microbial genetics, crystallography,

Computational Nanoscience and Nanotechnology 2002, (www.cr.org), ISBN 0-9708275-6-3.

computational biochemistry/drug design and organic synthesis. Progress, to date, in each of these areas will be reviewed is his presentation.

5 CONCLUSION

Obviously, it is impossible to present all of the current work being done in the field of modern-drug discovery, either in this review or in the one-day session. However, this session should allow the participant to get exposure to, and a sense of, current topics of interest to the research community in each area. It is the hope of the organizers that this session will allow the participants to gain a perspective on the opportunities and challenges that each area holds.

REFERENCES

[1] T. Peat, J. Newman and D. Bussiere, Curr. Opin. Drug Disc. & Devel., 2000. 3, 399-407, 2000.

[2] P. Case and M. Balick, Scientific American, 270, 82-87, 1994.

[3] G. O'Neill et al., Agents Actions Suppl, 46, 159-168, 1995.

[4] J. Greer et al., J. Med. Chem., 35, 1035-1054, 1994.

[5] A. Wlodawer and J. Vondrasek, Ann. Rev. Biophys. Biomol. Struct., 27, 249-284, 1998.

[6] E. Maggio and K. Ramnarayan, Trends Biotechnol., 19(7), 266-272, 2001.

[7] V. Nienaber et al., Nat. Biotechnol., 18(10), 1105-1108, 2000.

[8] S.B. Shuker et al., Science, 274(5292), 1531-1534, 1996.

[9] M. Gerstein and M. Levitt, Scientific American, 279, 100-105, 1998

[10] M. Wessell and S. Mente, Ann. Reports Med. Chem., 36, 257-266, 2001.

[11] B. Shoichet et al., Science, 259(5100), 1445-15550, 1993.

[12] K. Wuthrich and G. Wider, Curr. Opin. Struct. Biol., 9(5), 594-601, 1999.

[13] B. Shoichet and D. Bussiere, Curr. Opin. Drug Disc. & Devel, 3, 408-422, 2000.

FIGURES

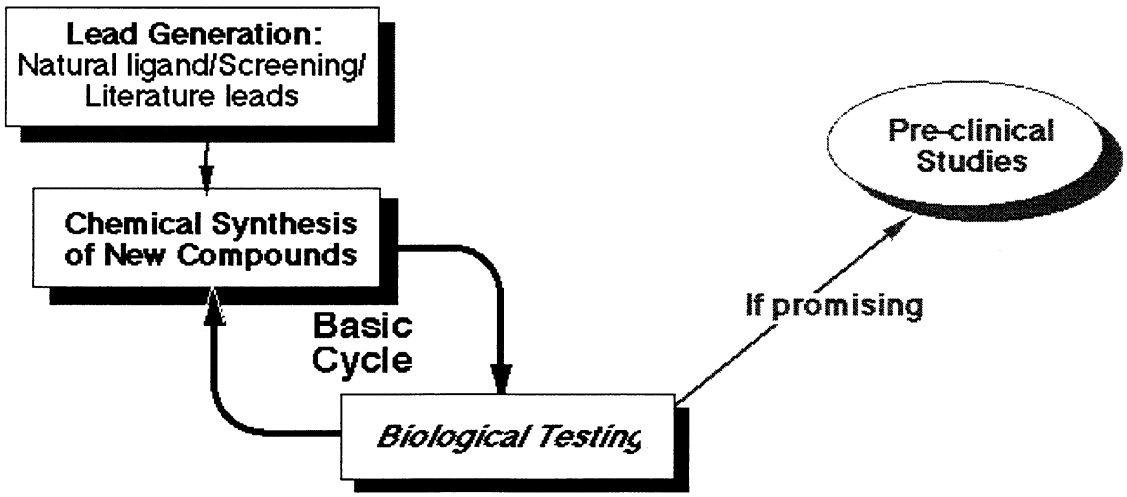

Figure 1. An example of the traditional drug discovery cycle. The figure was provided to the author several years ago by Dr. Jonathan Greer and Dr. Charles Hutchins (Abbott Laboratories).

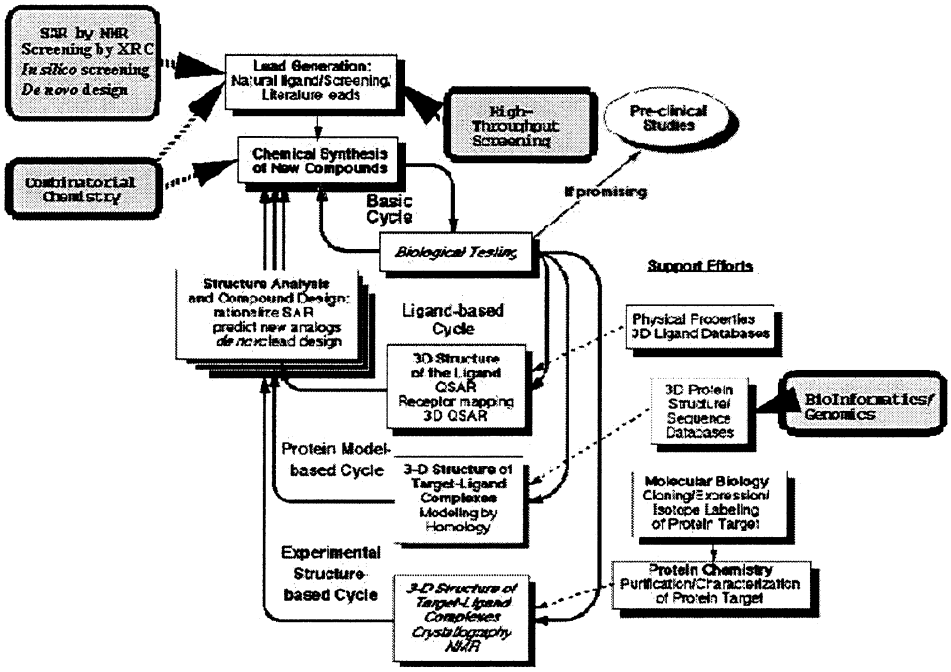

Figure 2. A modern drug discovery cycle. The figure is modified from a figure provided to the author several years ago by Dr. Jonathan Greer and Dr. Charles Hutchins (Abbott Laboratories).

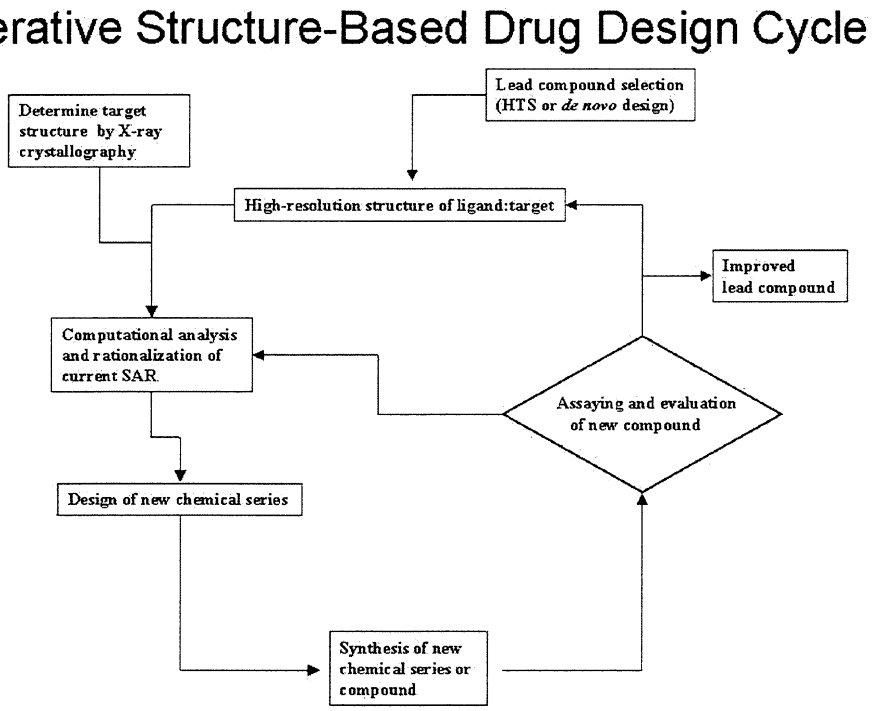

Figure 3. A simplified structure-based design cycle

Computational Nanoscience and Nanotechnology 2002, (www.cr.org), ISBN 0-9708275-6-3.