Locus Discovery: from structure to hit in weeks

Dawn Eringis and Bruce Goldman

Exploring the vast expanse of chemical space within organic compounds makes designing small molecules a difficult and time-consuming task, even when the 3D structure of a protein target is known. Locus Discovery's supercomputer cluster and proprietary algorithms are guided solely by a protein's structure and can generate many high-affinity compounds within weeks, ranked according to binding strength.

Dawn Eringis*
Locus Discovery
4 Valley Square
512 Township Line Road
Blue Bell, PA 19422, USA
tel: +1 215 358 2008
fax: +1 215 358 2010
*e-mail: deringis@locusdiscovery.com
Bruce Goldman
e-mail: brucegoldman@earthlink.net

▼ Locus Discovery uses proprietary computational drug-design methodology to discover drugs more quickly and efficiently, and with greater rates of success than previous drug-design approaches. Starting with only the 3D structural coordinates of a protein and proceeding from first principles, the technology can identify the biologically relevant active binding site and then design virtual libraries of potent, drug-like small-molecule compounds.

The impossibility of covering chemical space with preformed compounds is the Achilles heel of combinatorial chemistry. There are approximately 10^{80} designable organic compounds with molecular weights below 600 Da. Even if 99.999% could somehow be ruled out in advance, those remaining would outnumber the atoms in the Universe. Yet, in the entire history of drug research, perhaps only 100 million such compounds have ever been synthesized; figuratively speaking, much less than a drop of water in all the world's oceans.

It is generally agreed that knowledge of protein structures will streamline drug discovery. The National Institutes of Health (NIH) Protein Data Bank (PDB; http://www.rcsb.org/pdb) now holds 20,000 such structures, with a comparable number possibly residing in private databases [1]. However, designing a decent lead molecule can take years once a target protein's structure has been derived. Getting hits via computational approaches requires that a co-crystal complex of the protein and its *in vivo* ligand (or a good surrogate) be produced, and virtual libraries of chemicals screened. These hits must then be synthesized, their true affinities ranked experimentally, and medicinal chemistry efforts undertaken to optimize affinity and the absorption, distribution, metabolism, excretion and toxicity (ADMET) properties of the molecules to convert the hit into leads.

Financial and business strategy

This challenge has provided an opening for Locus Discovery (http://www.locusdiscovery.com), a privately held company. Locus's computer algorithms significantly shorten the pharmaceutical industry's multi-year lead-molecule searches into months. Locus's computational technology is exclusively licensed from the Sarnoff Corporation. Sarnoff spun-off Locus in September 1999, with Prism Venture Capital as its primary financial backer (Sarnoff retains 13% ownership of Locus). Locus's President and Chief Executive Officer, Nick Landekic, came onboard in February 2000. Other key Locus executives are William Moore, Chief Scientific Officer and Vice President of R&D; Nancy Barnabei, Vice President of Finance; Dawn Eringis, Director of Business Development; and Barbara Schilberg, Vice President and General Counsel. Duane Mason of Prism and Jeff Casdin of Cooper Hill, another major investor, sit on Locus's Board of Directors, as do Landekic, Carmen Catanese (Sarnoff), and Robert Gussin, former Chief Scientific Officer of Johnson & Johnson.

Locus has raised US$88 million in three rounds of venture financing since its spinoff from Sarnoff in late-1999, with Prism investing US$5 million in the seed round. In November 2000, Locus raised another US$42.5 million. All

1359-6446/02/$ – see front matter PII: S1359-6446(01)02165-1

institutional investors from those two previous financings participated in Locus's most recent round, led by Liberty Wanger Asset Management, which closed in November 2001 and brought in an additional US$40 million. Investors in the privately held company include Delphi Ventures, ING Furman Selz, INVESCO Global Health Sciences Fund, Amerindo Investment Advisors, Johnson & Johnson Development Corporation, Dresdner Kleinwort Capital, Tredegar Investments, Life Sciences Venture Fund, First Tier Biotechnology Partners and Origin Capital.

Locus's business model aims to provide three sources of revenue in the future: (1) computational drug design collaborations, in which Locus, using its computational capacity and power, designs libraries of compounds for protein targets (whether proprietary or publicly available) of interest to partners; (2) out-licensing of lead candidates from Locus's internal drug discovery programs at the point of *in vivo* proof-of-principle; and (3) retention of certain products for further development and commercialization. The company's first corporate partnership – to design lead compounds against three still undisclosed proteins for Aventis (Bridgewater, NJ, USA) – was put in place between Sarnoff (before it spun-off Locus) and Hoechst Marion Roussel (before it merged with Rhône-Poulenc to become Aventis). That deal carries a maximum potential cash value of US$79 million plus royalties on sales. Having assembled one of the world's most powerful parallel-processing supercomputer clusters in July 2001, and with its staff having increased to approximately 40 employees as of year-end 2001, the company is now gaining the capacity to handle additional partnerships and plans to hold discussions with several other companies in spring 2002, with animal data in hand.

Though it began life only in late 1999, Locus has already initiated a dozen internal drug-discovery programs. With a burn rate of approximately US$1 million per month (projected to roughly double over the course of 2002), Locus's cash reserves, now around US$70 million, should last a long time. Its methodology requires only a high-resolution protein structure [crystallographic or nuclear magnetic resonance (NMR)] and no additional experimental input.

Scientific background

Locus can test virtual, low molecular weight fragments at any protein target for which the structure is known, simultaneously identifying the biologically relevant binding site of the protein, virtually assembling high-affinity molecules from the fragments, and predicting the binding affinities of the molecules to the target.

Employing quantum-mechanical principles and other mathematical constructs, the Locus software calculates binding affinities of small virtual molecular fragments – separately or joined together – for a protein. Locus routinely uses approximately 150 organic fragments on its first pass, with an additional 5000 or so on the 'shelf' as needed. The approach samples tens of millions of unique molecular structures within each protein binding site, and finally builds the molecules and ranks them by affinity.

This pairing of molecular structures with binding-affinity data provides good quantitative structure–activity relationship (QSAR) data, enhancing lead optimization and increasing the likelihood of obtaining a selective therapeutic candidate. Without having synthesized a single molecule, one can predict how changes in small-molecule structures will affect binding strengths, and which portions of those molecules can be modified to improve pharmacokinetics or ADME characteristics. Locus can also bias the molecules synthesized towards those that conform to numerous metrics that predict bioavailability [2].

Producing the extensive combinations and permutations arising from 150 organic fragments (expected to rise to 500 or 600 by year-end with additional fragment optimization) necessitates immense computing power. Locus has assembled a parallel-processor supercomputer cluster comprising 2048 1GHz parallel processors and possessing a peak performance of 2.05 trillion calculations per second, one of the largest and fastest non-military Linux-based supercomputer clusters in the world (this claim is supported by comparison against the rankings of the top 500 supercomputer clusters listed on the Top Five Hundred Computer Cluster's website, http://www.top500.org/lists/2001/11).

Technology validation

Although Locus's algorithms and compounds remain proprietary, the company has generated increasing evidence that they perform as predicted. Locus validated its computational approach against four enzymes whose binding sites had been previously identified by conventional means. Locus's software independently located the same binding sites and, in several cases, reproduced chemical structures of recognized drugs or drug candidates. In addition, the software designed new, biologically active small molecules for each enzyme. As an example, human neutrophil elastase (HNE) is implicated in emphysema, asthma and adult respiratory distress syndrome (ARDS). Interestingly, Ono Pharmaceutical's (Osaka, Japan) HNE inhibitor ONO5046, now undergoing Phase III clinical trials in Japan, and Merck's (Whitehouse Station, NJ, USA) commercial acquired immune deficiency syndrome (AIDS) compound Crixivan, directed at human immunodeficiency virus (HIV)-1 protease (another Locus validation target), both ranked below several corresponding Locus candidates with respect to binding affinity.

Glycoprotein 41

Locus's fifth validation target was glycoprotein gp41, a structural coat protein found in HIV and numerous other viruses. Dimerization of gp41 is essential to viral fusion with cell membranes and, consequently, to infectivity. No commercial drug is yet available that targets gp41, which lacks a traditional binding site. For the past decade, Massachusetts Institute of Technology (Cambridge, MA, USA) scientist Peter Kim (now Head of Research at Merck) has used several experimental approaches to identify a 'critical assembly region' on the protein that is essential for gp41 dimerization.

In April 2001, Locus initiated a full-bore gp41 drug discovery program. In under four weeks, Locus software 'found' the same gp41 region deduced by Peter Kim's group and predicted several novel small molecules that prevent gp41 dimerization. Less than six months later, Locus chemists had synthesized 12 ligands with molecular weights of between 300 and 500 Da, several of which have proved very potent, over a six-day period, in blocking HIV fusion with HeLa cells exposed to the compounds. A standard assay employing antibodies to two specific HIV coat proteins (whose locations inside or outside of the cell can be visualized) showed that, in concentrations below 100 μm, the Locus compounds conveyed dose-dependent protection against fusion of the HeLa cells. Because HIV fusion with susceptible cells is essential for infection, it is no surprise that several of these same compounds also protected HeLa cells in culture from being killed by HIV. Based on these results, Locus is proceeding to produce a second generation of compounds for development.

Erythropoietin mimetics

A second discovery program is focussed on small-molecule erythropoietin (EPO) mimetics. EPO is a bulky peptide that triggers dimerization of two separate protein constituents of the EPO receptor on the cell surface. The receptor's full activation requires exact alignment of these two receptor proteins. Several structural characterizations of the human EPO receptor, based on complexes of the receptor with molecules that triggered only partial activation, have been made public in the past five years. Locus designed several small molecules that triggered proliferation in three different human stem-cell lines: UT-7 (commercially available from many vendors) and two types of pre-erythroid cells, CD34+ and CD36+, which are isolated from fresh cord blood. Proliferation was observed in three ways, over a 48-h period for each cell line: first, via monitoring uptake of a fluorescent dye (which varies with the number of cells); second, by plating the cells out and counting them; and third, by measuring ATPase activity (which tracks with cell count). Results using all three measures were similar: the triggering of proliferation by Locus-designed compound concentrations at the picomolar range, comparable to natural EPO itself. Assays also show that these compounds trigger differentiation of pre-erythroid cells into mature erythrocytes.

Proceeding from a more recently characterized structure of the EPO receptor dimer complexed with human EPO – a fully activated and therefore more appropriate conformation for EPO-mimetic drug discovery – Locus has initiated a second EPO program and designed small molecules whose chemical structures differ considerably from those of the first group. Locus has therefore shifted its resources into the second program.

Other programs

Along with gp41 and EPO, Locus has initiated several programs, including arthritis, inflammation, cancer and antibacterial programs. The company has already identified binding sites for several of these targets, and plans, by year-end 2002, to have obtained animal data for at least ten of its internal discovery programs and to have completed lead optimization in several of them. An additional 15 programs have been prioritized that can be ramped up quickly if any of the first-tier programs fail or if more resources become available.

Concluding remarks

Locus's criteria for a target's selection are that its structure be available and biologically validated and that it presents significant commercial opportunities. The exponentially growing number of protein structures with well-validated connections to disease processes is already sufficient to keep Locus researchers busy for several decades.

References

1 Berman, H.M. *et al.* (2000) The protein data bank. *Nucleic Acids Res.* 28, 235–242

2 Lipinski, C. (1999) Drug structures and properties, past and present: can we design drugs with beautiful properties? http://www.iainm.demon.co.uk/spring99/lipins.pdf