Once upon a time, not long ago, HIV/AIDS was a scourge, killing anyone who contracted the deadly virus. Now, many people are living with the disease, which they control with drugs initially developed in the 1980s and early 1990s using an approach called computer-aided drug design—the use of computer models to find, build, or optimize drug leads.

Armed with information about the 3-D structure of HIV protease, an enzyme essential to the HIV reproductive cycle, computational researchers designed molecules in silico to precisely fit the shape of the enzyme’s active site—as though fitting a key to a lock. The resulting drugs, potent inhibitors of HIV protease and the HIV life cycle, were brought to market in record time and revolutionized the treatment of HIV/AIDS.

Around the same time, another anti-viral—Relenza, which treats influenza and was a forerunner to Tamiflu—was also designed using these methods. These HIV and flu drugs are among the best known success stories of computer-aided drug design (see page 23 for both stories).

Since those early successes, computer modeling has become an integral part of drug discovery. “Almost everything that has recently moved forward from big pharmaceutical companies to market has involved some sort of collaboration with computational chemistry. It’s like asking, were there chemists involved? Of course there were. It is part of the process,” says Tara Mirzadegan, PhD, head of the computer-aided drug design group at Johnson & Johnson.
“Almost everything that has recently moved forward from big pharmaceutical companies to market has involved some sort of collaboration with computational chemistry. It’s like asking, were there chemists involved? Of course there were. It is part of the process,” says Tara Mirzadegan.

Quite often, computers play a role without making the big splash they did with Relenza and the protease inhibitors. That’s probably because no drug is created solely in silico; the computer is just one of many tools in this process. But as algorithms evolve, computing power explodes, and scientists solve a greater number of 3-D protein structures, computer-aided design has the potential to dramatically cut the cost and time of drug discovery. How? By narrowing down the field of compounds that might help treat a particular disease; by assembling novel drug molecules to disrupt specific disease pathways; and by providing new attack routes against traditionally difficult drug targets. Computers are also increasingly playing a role in optimizing drug leads for bioavailability and safety.

Despite the over-hype of computers as the saviors of drug development companies, many still expect this process to bear important fruit. Computer-aided drug design played a critical role in the design of several drugs that are now in late preclinical or early clinical development. Only time will tell which of these, if any, will emerge as drug success stories.

VIRTUAL SCREENING

How it works: In the ideal situation, the 3-D structure of the target molecule (usually an enzyme or receptor) is known, allowing scientists to directly visualize drug-target interactions in silico. Structure-based methods have evolved in two directions since Relenza and the HIV proteases—virtual screening and fragment-based design.

In virtual screening, the 3-D structure of a target is screened against libraries of potentially active small molecules. The computer “docks” each compound, or ligand, into the target’s active site and scores its geometric and electrostatic fit.

Considerable progress has been made in docking programs in the last two decades, but scientists agree that the problem is complex and that they have yet to find a perfect solution. To start with, the ligand and protein target are often pictured as a rigid lock and key—but in fact they are dynamic, moving objects that continually change shape and adjust their shapes in response to each other.

“Imagine taking a fluffy ball and trying to mold it to optimally fit some kind of a binding site. There are just way too many configurations,” says Dimitris K. Agrafiotis, PhD, vice president of...
EARLY EXAMPLES: ANTI-VIRAL DRUGS

Relenza and the HIV protease inhibitors stand out as the two classic examples of computer-aided drug design. Relenza was developed through a collaboration of Australian scientists, including Jose N. Varghese, PhD, head of structural biology at CSIRO Molecular and Health Technologies. In 1983, Varghese and his colleagues used X-ray crystallography to solve the 3-D structure of the enzyme neuraminidase, one of two potential protein targets on the surface of flu. Neuraminidase plays a critical role in the flu life cycle: after the virus replicates within a host cell, neuraminidase releases the newly formed viral progeny by cleaving a bond between the viral surface protein hemagglutinin and a sugar on the host cell surface, sialic acid.

A series of structural experiments revealed important insights. The active site of the enzyme was highly conserved in all strains of flu—both human and animal; the virus routinely escaped antibody recognition by mutating around the periphery of the active site but never changing the active site itself.

“Because it was so highly conserved, it seemed clear to us that it must have a very important function,” Varghese says. “So, clearly if one made a molecule that went in there and blocked that site, it would be pretty effective.”

A synthetic analog of sialic acid was known to inhibit neuraminidase, but without sufficient potency. Using the crystal structure of neuraminidase bound with this analog, the researchers set out to design a better inhibitor in silico. Computer predictions revealed that a particular guanidinium-for-oxygen substitution would give tight binding. Synthesis of this compound—Relenza—turned out to be tricky, but eventually succeeded.

“It bound in nanomolar binding, so it was very tight, and it certainly blocked the virus replication right down to its tracks,” Varghese says.

Relenza was licensed to GlaxoSmithKline Inc. in 1990 and approved by the FDA in 1999. Following their lead—and capitalizing on a patent oversight, according to Varghese—Gilead Sciences developed the better-known neuraminidase inhibitor, Tamiflu (marketed by Roche). Both drugs may be important in the fight against bird flu, Varghese says.

Development of the HIV protease inhibitors lagged behind that of the neuraminidase inhibitors by several years, but the former won FDA approval sooner (in the mid-1990s) because of the pressing medical need.

Dale Kempf, PhD, who is now a distinguished research fellow in Global Pharmaceutical Research and Development at Abbott, was involved in Abbott’s development of ritonavir (brand name Norvir), which started in late 1987.

“It’s one of the first examples of the application of genomics for drug design,” he says. When the HIV genome was sequenced and published in the mid-1980s, several groups recognized characteristic sequences suggestive of a protease enzyme.

Interestingly, the gene encoded only half a protein, which led Kempf and others to realize that the protease must be composed of a dimer—two identical halves that come together to form one active site. This provided a key structural insight even before X-ray crystal structures of the protease were available: the active site had to have a particular type of symmetry, known as C2 or two-fold symmetry (rotation 180 degrees around a central axis yields the identical structure).

Kempf’s group used that insight to create a computer model of the protease active site and to design possible inhibitors in silico by starting with a known substrate, chopping off half of the substrate, and rotating the remaining half by 180 degrees.

“And when we went into the lab and made those compounds, they turned out to be very potent inhibitors,” Kempf says.

Using a combination of the X-ray crystal structures of HIV protease (which had since become available) and computer graphics, they modified these compounds in silico to visualize how certain substitutions would improve characteristics like bioavailability. The first compound with sufficient oral bioavailability, ritonavir, was synthesized in 1991.

In 1996, the FDA approved ritonavir in record time (72 days). The total development time—about eight years—was roughly half that of a typical drug, due both to the structure-based approach and to the FDA’s accelerated review. Several other HIV proteases emerged around the same time, including saquinavir (Roche) and nelfinavir (developed by Agouron, now a subsidiary of Pfizer). These drugs helped to revolutionize the treatment of HIV.
informatics at Johnson & Johnson Pharmaceutical Research & Development. “Small molecules—unless they’re very small—tend to be very flexible. They flop around a lot. They can assume a multitude of conformations in 3-D.” If a molecule has five rotatable bonds, then each bond can rotate at many different angles, creating a lot of freedom to take on unique conformations.

Most docking programs now account for the flexibility of the ligand by sampling its many conformations and docking each one, but adequately accounting for the flexibility of the target protein is a much more challenging problem. Adding protein flexibility exponentially increases computing demands.

“The state of the art today is coming up with sensible simplifications that make the problem computationally tractable but still meaningful,” Agrafiotis says.

Besides the flexibility of the protein, many docking programs do not adequately account for the influence of water—which surrounds all molecules in living systems. “The mathematical models for defining water and how it shapes itself around the receptor and the drug molecule are still pretty unclear,” says Kent Stewart, PhD, a research fellow in structural biology at Abbott.

In addition, the algorithms estimate binding energies using classical Newtonian physics, rather than quantum physics—which also reduces accuracy. “You can calculate the binding energies from some sort of Newtonian point of view, treating atoms as sort of balls attached to springs. Or you can treat it from a quantum mechanical point of view. Now the quantum mechanical calculations, as you can imagine, are horrendous,” says Jose N. Varghese, PhD, head of structural biology at CSIRO Molecular and Health Technologies. “At this stage, it is a computational challenge.”

Methods of scoring how well a small molecule fits a protein’s active site also must trade off between speed and accuracy. “The scoring function that we use has many shortcuts and approximations,” says Mirzadegan. Her group will virtually dock the company’s one million proprietary compounds (which it has purchased or developed over the years) against a given target, and pick the highest ranked 10,000 for biological testing. “We cannot afford docking one compound per day. That would be one

Cancer Interrupted. This three-dimensional computer graphic shows a drug candidate (MET tyrosine kinase inhibitor) bound to its target protein. MET receptor tyrosine kinase controls cell growth, division, and motility and is implicated in a range of cancers, including renal cell carcinoma, gastric cancer, lung cancer, glioblastoma and multiple myeloma. Courtesy of SGX Pharmaceuticals, Inc.
“The state of the art today is coming up with sensible simplifications that make the problem computationally tractable but still meaningful,” says Dimitris K. Agrafiotis.
DOCK THIS: *In Silico* Drug Design Feeds Drug Development

**Fragment-based design**. Drug companies, such as SGX Pharmaceuticals, screen hundreds of fragments in their fragment libraries and identify hits that serve as the building blocks for novel drug candidates. Knowledge of the binding mode of each fragment to its target is combined with advanced computational tools to produce “engineered” drug leads. For example, in this series, a hit is first identified through crystallographic screening (yellow); then chemical groups (red and pink) are added to the bound fragment to increase its binding affinity. Courtesy of SGX Pharmaceuticals, Inc.

Distributed computing is key to developing better, more accurate algorithms for computer-aided drug design, says Vijay Pande. “It allows us to do calculations otherwise impossible,” Pande says.

**Fragment-Based Design**

Fragment-based methods take a “Lego” approach to drug design. In a lab, scientists create chemical libraries of small compounds, or fragments—perhaps one-third the size of a typical drug—that are easily linked together. They then screen the libraries for binding activity experimentally, using high-throughput X-ray crystallography (or NMR or mass spectrometry); when a fragment binds to the target, the crystallography provides an exact 3-D picture of the bound fragment in the active site. Next, with the help of computer modeling, fragments are turned into potent drug leads by adding new chemical groups to the initial core fragment or by stitching together several fragments that bind to different points in the active site.

“I think this approach is showing quite good promise,” Varghese says. “In fact, with the advent of these modern synchrotrons, scientists can do this fairly quickly—and a lot of pharmaceutical companies are moving in this direction.”

The approach offers a combinatorial advantage: “Instead of having a database of say four million compounds that a really large company would have, you take compounds that are say one-third of the size, and explore them combinatorically. If you explored ten fragments in three different positions, you’d actually explore 1000 combinations. So with a database of something like 400 compounds, you can explore a chemical space that is in the several millions,” says Sir Tom Blundell, FRS, FMedSci, professor and chair of biochemistry at the University of Cambridge. In 1999, Blundell co-founded Astex Therapeutics to do fragment-based methods; the company is now testing a kinase inhibitor—a type of cancer drug—in clinical trials.

“The experiment is really one of using crystallography to do your screening. So you’ve pushed the crystallography technology to the point where you can do it so rapidly that it becomes effective to use as a screening tool,” says Siegfried Reich, PhD, vice president of drug discovery at SGX Pharmaceuticals, another company that uses fragment-based methods. (Reich previously helped develop the HIV protease inhibitor nelfinavir at Agouron.) When it was founded in 1999, SGX was named Structural Genomix and its aim was to use high throughput X-ray crystallography to solve a record number of protein structures. But this was not sustainable as a business model. So, in 2000, the com-
pany changed its name to SGX Pharmaceuticals and put its crystallography power to use in drug discovery.

One of their lead candidates is a new inhibitor of BCR-ABL, a perpetually active kinase enzyme involved in chronic myelogenous leukemia, or CML. The BCR-ABL inhibitor Gleevec has had enormous success in treating CML patients, but 20 percent are resistant to Gleevec. So scientists at SGX cloned, expressed, purified, and crystallized the Gleevec-resistant protein. Then they screened their fragment library against the wild type and mutant versions of BCR-ABL to find compounds active against both. The fragment hit that eventually led to their lead candidate started with a low binding affinity of just 10 micromolars (i.e., a fairly high concentration of compound was required to bind at least half the protein).

This is where the medicinal chemists and structural biologists sit down with the computational chemists, Reich says. Computational chemists virtually build new compounds by adding chemical groups to the starting fragment. For example, they might try linking all the different simple alkyl amines to one of the fragment’s “chemical handles” (sites on the fragment that easily bind to other chemical groups), Reich explains. The computer calculates the binding affinity for each iteration, until it finds one with tight binding. Specialized versions of docking programs are used to calculate the binding affinities. But because you already know exactly how the fragment binds, you start with more information than in virtual screening.

By elaborating their initial lead in this way, SGX got their first hit down to nanomolar potency—i.e. very little of the compound was required in order to bind the protein—in about three months. “That gives you a flavor for how fast this can go,” Reich says.

**TRICKY TARGETS**

Docking algorithms and fragment-based methods work well on soluble enzymes that are easily crystallized and contain well-defined pockets where ligands can bind—but many diseases instead involve membrane-bound receptors or protein-protein interactions.

Membrane-bound receptors transmit signals from outside to inside the cell. Because the proteins are embedded in the membrane, they cannot easily be crystallized and it is difficult to solve their structures. For example, 25 percent of the top 100 drugs on the market today target G-protein coupled receptors—including the dopamine and serotonin receptors in the brain—but the structure of only one mammalian G-protein coupled receptor is known.

When structural information is unavailable, computational chemists use ligand-based methods to hunt for new drug leads. They superimpose a set of ligands with known activity against the target and compare their structural and chemical features. A common pattern, called a pharmacophore, emerges—key functional groups (such as hydrogen bond donors, electrostatic charges, and hydrophobic patches) must be in certain positions. This fingerprint is then used to virtually screen libraries for novel compounds with similar patterns. Ligand-based methods pre-date the structure-based methods and have helped develop many drugs, including drugs to treat high blood pressure, pain, and depression.

Protein-protein interactions occur via surfaces that are often featureless and shallow, and binding affinities can be quite large—so it’s hard for small molecules to disrupt these interactions, says Arthur Olson of Scripps Research Institute. You have to find or design drugs that can bind to multiple footholds, or hot spots, on the protein surface, which is challenging, he says. “I think that this is an area that is really still in its infancy.”

But some progress is being made. Kent Stewart of Abbott Labs hopes to control BCL-2, a protein that is overexpressed in certain cancers. It blocks apoptosis (programmed cell death) and thus keeps cancer cells alive. Compared to HIV, Stewart says, which has an actual cave you can dock a molecule into, on
BCL-2, “there’s no such thing as a cave; it’s a very flat and open surface, so it’s hard to get molecules that actually stick,” So, using a fragment-based approach, scientists at Abbott linked together two fragments that bind to the BCL-2 protein surface, resulting in a potent compound that can disrupt the protein-protein interaction. The compound is now in late preclinical development. The compound is in late preclinical development. Courtesy of Abbott.

Some companies have made these difficult targets their niche area. For example, Polymedix’s mission is to develop drugs against membrane-bound targets, protein-protein interactions, and membrane-protein interactions, using a suite of computational tools specifically developed for these aims (by professors William DeGrado, PhD, and Michael Klein, PhD of the University of Pennsylvania).

Polymedix is working on a new line of antibiotics that mimic the action of defensins—natural proteins found in the body that kill bacteria.

“They work similarly to a needle or a corkscrew going into a balloon. They directly attack and perforate the bacterial cell membrane,” says Nicholas Landekic, MBA, President, CEO, and co-founder of Polymedix. Because they do not target bacterial proteins—which can easily evolve to escape drug pressures—defensin-like drugs should not engender bacterial resistance, he says.

Scientists at Polymedix built a computational model of a defensin protein inserted into a bacterial cell membrane (a peptide-membrane interaction). Then they virtually transformed the defensin protein into a drugsized compound. By swapping amino acid groups for chemically analogous small molecule groups, they shrunk the protein while preserving its chemical interactions (electrostatics, lipophilicity, etc.) within the membrane.

The result: drug leads one-tenth the size of the defensins, but about 100-fold more potent and 1000-fold more selective. “So we’ve been able to improve on nature,” Landekic says. The compounds are now being tested in animal studies.

“We’ve spent less than 14 million dollars to date since starting Polymedix, so in terms of an efficiency and efficacy rate, I think that’s pretty good,” he adds.

Computer-aided methods can identify drug leads with potent activity against a target, but these compounds are far from being drugs. Drugs must also be bioavailable and safe. Safety problems derail many drugs late in development, so identifying potential safety snags early on could save considerable time and money.
“How well can we evaluate bioavailability and toxicity in silico? It’s pretty blunt and not a very popular answer: we don’t do very well,” Stewart says. “The biological mechanisms underlying bioavailability and toxicity are complex. So the mathematical models in those areas are still in their infancy,”

Olson agrees: We are a long way from being able to simulate a drug’s effect on the entire human body. “When you’re talking about toxicity, it’s much easier to give a compound to a rat than it is to dock against all possible proteins that are in the rat, even today,” he says. “But someday, you might be able to do that. We’re certainly creeping up on that.”

Computers do play a role today, however. Drugs must meet properties that fall under the ADME acronym: be Absorbed by the body, Distributed to the target tissues, and not Metabolized or Excreted too quickly. Software programs check molecules for key features (known as “Lipinski’s Rule of Five”) that are associated with favorable ADME profiles, such as having five or fewer hydrogen bond donors and a molecular weight below 500.

With enough computing power, scientists can also virtually screen a candidate compound against a large panel of proteins from the body, to make sure the compound will not cross react with other enzymes or receptors to cause side effects.

To ensure that molecules identified in the computer will have real-world

For the field to progress, says Anthony NcHollis, the current software needs to be more closely scrutinized—using prospective studies that directly compare the impact of computer-aided methods with more traditional drug design approaches.
value, computational scientists benefit from working closely with medicinal chemists during lead identification and optimization.

“Medicinal chemists would tell you that there’s lots of intuition involved, so it’s not all computational,” says Hans Wolters, PhD, associate director of informatics at XDx, Inc. For example, he says that as computer scientists became more involved in making drugs, the molecular weight of candidate compounds began to creep up precipitously—to sizes that would not be easily absorbed by the human body. Medicinal chemists help recognize this type of problem early in the process.

DEBATING THE IMPACT

In the past two decades, although computer-aided drug design has become an integral part of drug discovery, some remain skeptical as to whether these methods are delivering on their promise. The productivity of the pharmaceutical industry has actually declined in the past decade (The FDA approved 58 drugs from 2002 to 2004 compared with 110 from 1994 to 1996, according to the Tufts Center for Drug Development.) Though this is likely due to many factors—in particular, tightening safety standards and the enormous cost and time of clinical trials—the trend has left some wondering whether large investments in technology, including computer-aided drug design, are paying significant dividends.

Many modeling programs are unreliable, and they are not making a big difference in the real world, cautions Anthony Nicholls, President and CEO of OpenEye Scientific Software, which develops software for computer-aided drug design. “It’s all done on faith. It’s all done on the idea that ‘oh, we’re using computers, so it must be better,’” he says. “I think a lot of people are fooling themselves.” He believes that, for the field to progress, the current software needs to be more closely scrutinized—using prospective studies that directly compare the impact of computer-aided methods with more traditional drug design approaches.

Other scientists agree that the algorithms are still being refined, but have a more optimistic outlook. They say that progress is steady and that computer-aided design is already having an impact. Klaus Klumpp, PhD, an associate director at Roche (who was involved in the development of the HIV protease inhibitor saquinavir), points to a suite of emerging drugs for hepatitis C virus (HCV) as a case in point.

HCV was discovered in 1989 and the virus was difficult to grow, so structural information for HCV polymerase and HCV protease became available relatively late—in the mid-to-late 1990s. By this time, computer-aided drug design was well integrated into big pharmaceutical companies. Several companies quickly identified binding sites and designed inhibitors, many of which are now in early clinical trials. “It is expected to completely change the treatment paradigm for HCV infected patients,” Klumpp says.

Richard Casey, PhD, founder and chief scientific officer of RMC Biosciences, Inc., has also witnessed the dramatic effect that computers can have on drug design. His company provides computer-aided drug design services for small and mid-size pharmaceutical companies, which often lack in-house teams.

Recently, he made 3-D models and performed in silico docking studies for a mid-size pharmaceutical company that had identified active lead compounds but had no understanding of how they were binding the target, an RNA synthetase.

“When they saw this for the first time, it was the ‘aha’ effect: So that’s why this compound has high activity and this compound does not. It was a real eye-opener for them,” Casey says. “I think in the next seven to ten years, with the computational power that’s coming on line here pretty soon and the steady development in algorithms, computer-aided design is going to make a huge difference.” □