Cutting-Ed
Crystall

The demand for protein crystallography in drug discovery is driven by the need to understand exactly how small-molecule drug candidates bind to their protein targets. "High-affinity ligands often induce conformational changes in protein structure," says Lance Stewart of Emerald BioStructures, Inc. (Bainbridge Island, WA). "Crystal structures allow scientists to observe such changes and exploit them in subsequent rounds of drug development. It is important to confirm predicted binding modes for ligands as they are improved through medicinal chemistry." The design of small-molecule drugs needs to be monitored with high-resolution pictures of the bound molecules. Whether the ligands bind as predicted or not, the pictures can be used to take the process forward. "Having information allows the scientist to convert a potential negative into a positive," adds Stewart.

"If you can understand the detailed atomic structure of a protein and how it interacts with other molecules, then this knowledge can aid the design of new molecules that bind better and more specifically," echoes Rod Hubbard of the drug discovery company RiboTargets (Cambridge, U.K.). Proteins perform most of the important chemistry of cells, and most therapeutic molecules work by blocking or modifying the way in which a protein works. Now, as we navigate the postgenomic world, there is a growing need to get an atom-by-atom snapshot of protein targets. This is the realm of structural genomics, and high-throughput X-ray crystallography (HTC) could be the key.

Putting HTC in its place

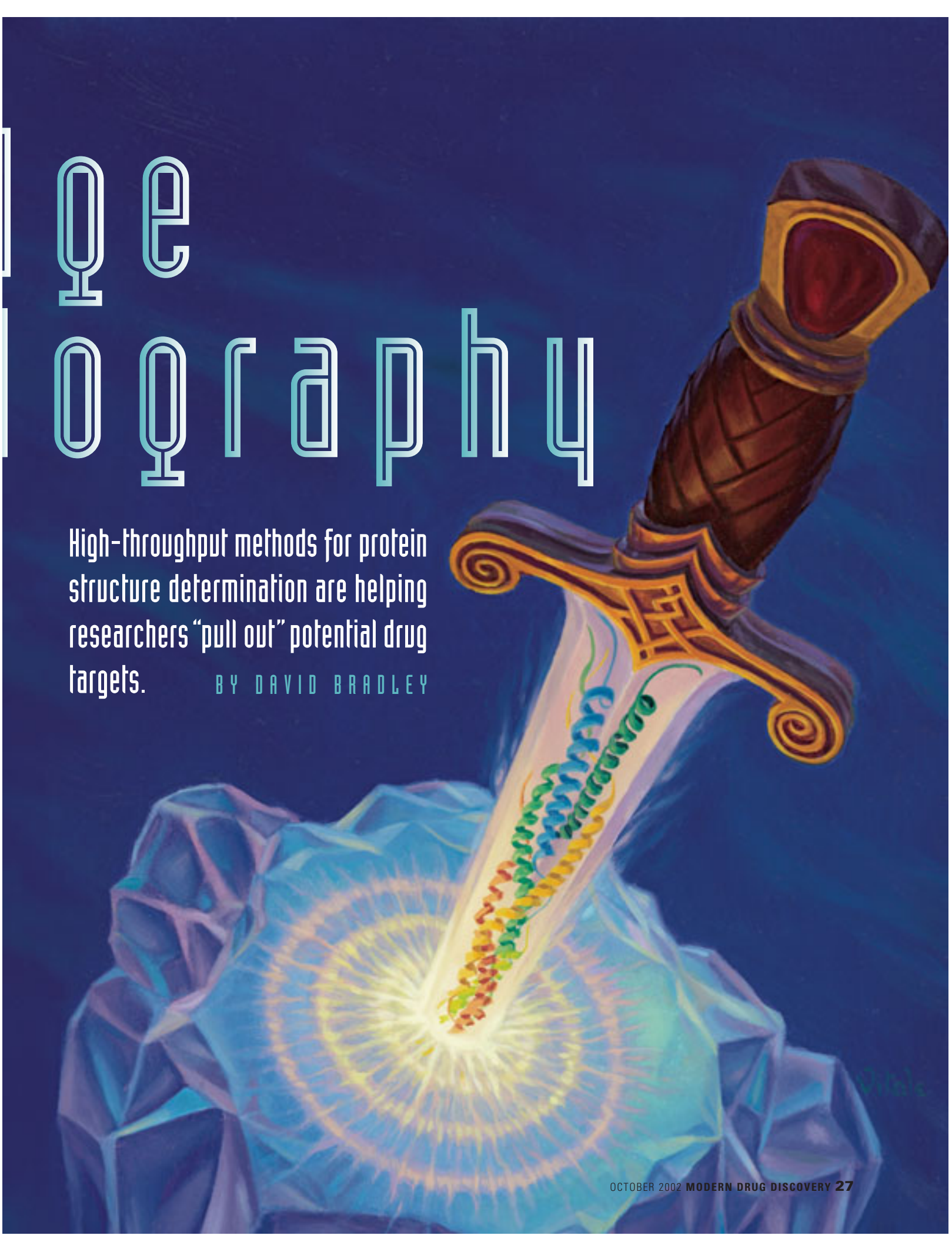
There was a time not long ago when elucidating a single protein structure would suffice to earn a student his or her doctorate. Those halcyon days are long gone, and although crystal structure determination is not yet a routine task, innovations in robotics, crystallization techniques, and processing the vast data streams from X-ray data-collection devices are beginning to make protein crystallography a semiautomated and high-throughput task. This once entirely specialist trade could one day become as immediately accessible and transparent to the drug designer as are spectroscopic and chromatographic methods today.

ILLUSTRATION: RAOUL VITAL

qe
oqraphy

High-throughput methods for protein structure determination are helping researchers "pull out" potential drug targets.

BY DAVID BRADLEY



Advertisers

Combinatorial chemistry and high-throughput screening can take the research process only so far in generating lead candidates. There comes a point at which a more detailed appraisal of the proteins becomes essential. In short, understanding how a small molecule might interact with a protein or how to design such a molecule from first principles can happen much faster if a precise high-resolution atomic structure of the protein is available.

Stewart points out that the narrowest bottleneck along the road from target gene to marketable drug is determining the three-dimensional protein structure. NMR and other techniques provide important insights along the way, but accelerating along the route to a complete picture will only be possible with automated protein crystallography. This technique not only will provide new markers but also will offer insight into the function of the myriad members of the proteome and so generate new landmarks en route.

Five steps to crystal nirvana

The solving of a protein crystal structure is not nearly as simple as the solution of a small-molecule NMR spectrum, as one might imagine. It involves a five-point plan (Figure 1). The first step is the expression of enough protein for analysis. In some cases, this step requires the use of in vitro transcription/translation kits or the subcloning of genes into various bacteria, yeast, or cultured insect cells to express the proteins at levels reaching 50% of total protein without killing the host cells. The second challenge is purifying the protein using a potentially large number of extraction and isolation steps to yield protein that must be both stable (under various conditions) and high-concentration (in the mg/mL range). The third step is generally the toughest—producing a sufficiently high-quality crystal by whatever means possible. Data collection with an X-ray diffractometer, the fourth step, is usually a skilled but straightforward task, while the fifth and final step to crystal heaven is conversion of the mass of X-ray data into a refined model of the protein using dedicated, proprietary, and sometimes home-brewed computer software.

The third step often proves to be the most problematic, but many years of experience have led to a vast knowledge base and literature on finding the right conditions to turn a mass of protein chains into crystal form. Indeed, the Biological Macromolecular Crystallization Database (BMCD) is the first port of call—after the local technicians—for many researchers hoping to find the means to crystallize a stubborn solubilized protein. The BMCD includes information on how 2526 biological macromolecules were crystallized. As Stewart points out, however, the BMCD entries only list crystallization experiments that yielded crystals from which a structure could be determined; an immeasurable number of crystallizations taking place daily are never published. The first tentative steps in organizing this information have taken shape in the form of the Parallel Experiment Planning System at the University of Pittsburgh, Abbott Laboratories' CrystaLEAD, and Emerald BioStructures' Crystal Monitor.

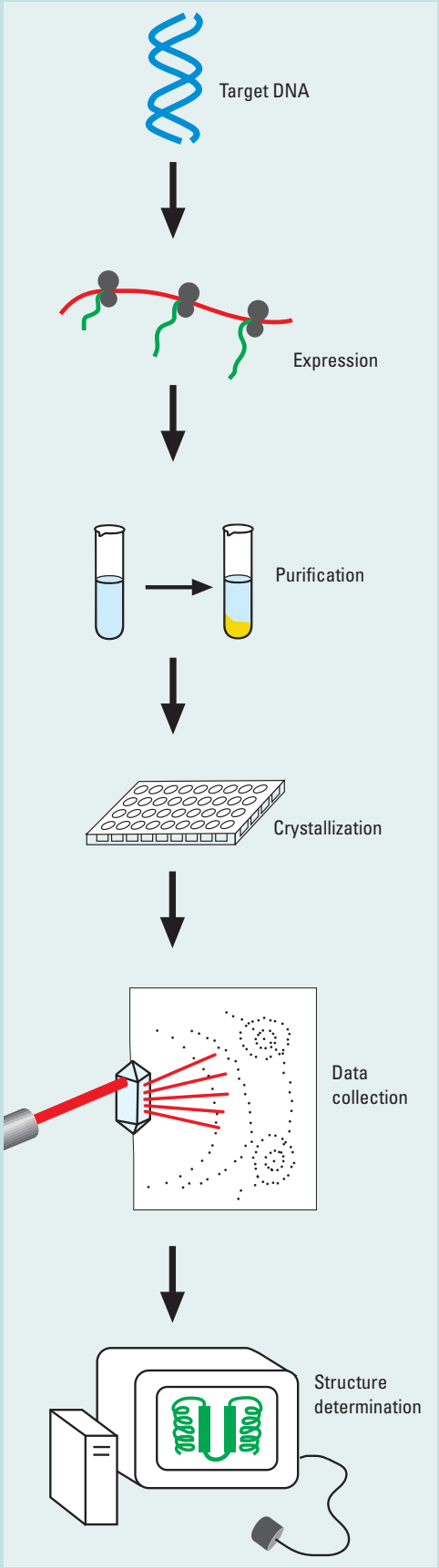


Figure 1. A five-point plan. Getting from gene to protein structure involves five basic steps, each with its own unique challenges: protein expression, protein purification, protein crystallization, data collection, and structure determination.

Commercializing crystallography

Raymond Stevens, a professor of molecular biology at the Scripps Research Institute, founded the biotechnology company Syrrx, Inc., with a view to commercializing the HTC developments from his laboratory. The company's approach is to focus on weaving the various steps in parallel and automating the whole crystallization process. According to Syrrx researcher Chris Behnke, each step will ultimately be streamlined, but he points out that not every step will be improved by automation, because some steps are performed just as efficiently manually as automatically. Last May, pharmaceutical giant Roche licensed Syrrx technology for HTC of undisclosed protein targets covering several therapeutic areas.

Exploiting this crystallization folk knowledge is key to automating the crystallography process, although Hubbard adds that producing sufficient quantities of pure, homogeneous protein is probably even more important. The robotic system being developed by the Hauptman-Woodward Medical Research Institute (Buffalo, NY) is used to test 40,000–60,000 crystallization experiments a day. Similarly, Emerald BioStructures' automated system can cocrystallize and image 10,000 dif-

Crystal-clear information

Corporate HTC

Abbott Laboratories	www.abbott.com
Accelrys Inc.	www.accelrys.com
Affinium Pharmaceuticals (formerly Integrative Proteomics)	www.integrativeproteomics.com/html/home/home.shtml
Astex Technology	www.astex-technology.com
Bristol-Myers Squibb Co.	www.bms.com/landing/data/index.html
Corvas International, Inc.	www.corvas.com
3-Dimensional Pharmaceuticals, Inc.	www.3dp.com
Emerald BioStructures, Inc.	www.emeraldbiostructures.com
Exegenics, Inc.	www.exegenicsinc.com
Exelixis, Inc.	www.exelixis.com
Genencor International, Inc.	www.genencor.com
Genetics Institute	www.genetics.com/genetics/genetics/index.htm
Millennium Pharmaceuticals	www.mlnm.com
Pfizer, Inc.	www.pfizer.com/main.html
Procter & Gamble	www.pg.com
RiboTargets	www.ribotargets.com
Schering AG	www.schering.de/eng/index.html
Structural GenomiX	www.stromix.com
Syrrx, Inc.	www.syrrx.com
Tularik, Inc.	www.tularik.com/about/about.php
Vertex Pharmaceuticals, Inc.	www.vpharm.com/frame09.html

Other HTC

Advanced Photon Source	www.aps.anl.gov
Biological Macromolecular Crystallization Database	http://wwwbmcd.nist.gov:8080/bmcd/bmcd.html
Genomics Institute of the Novartis Research Foundation	www.gnf.org
New York Structural Genomics Research Consortium	www.nysgrc.org
Stanford's Blu-Ice	http://smb.slac.stanford.edu/datacollect/BluIce
Structural Genomics of Pathogenic Protozoa	www.sgpp.org
Structural Genomics	www.rcsb.org/pdb/strucgen.html
Tuberculosis HTC Consortium	www.doe-mbi.ucla.edu/TB

ferent proteins with their small-molecule ligands in a mere eight hours. The systems take a snapshot of the multiwell plates, and those that produce crystals rather than syrupy goo can then be taken on to the next stage. And the Agincourt robot developed by Syrrx and the Genomics Institute of the Novartis Research Foundation (San Diego) is set up for more than 2 million crystallization drops.

The time for trial and error

Chemist Marek Brzozowski of York University (U.K.) recognizes that protein crystallization is an integral part of crystal X-ray analysis. Although such robotic systems can handle tens of thousands of crystallizations, the process still involves a great deal of trial and error. His research focuses on the design and practical testing of crystallization procedures for non-water-soluble proteins and peptides in conditions that mimic the natural protein environment: membranes, micelles, and cytoplasmic vesicles.

Other researchers agree that HTC will open the floodgates for drug discovery. "Protein crystallographers are faced with a tremendous task of solving hundreds if not thousands of structures per year," explains team leader Steven Almo of the Albert Einstein College of Medicine (New York). "This flood of structural information may have a considerable impact on our ability to understand many fundamental unsolved scientific problems (e.g., protein folding) and, from a practical point of view, will give a unique opportunity for better understanding of various normal and pathogenic biological processes."

The ongoing structural genomics pilot project at the New York Structural Genomics Research Consortium is also helping to make the high-throughput crystallographic determination of three-dimensional protein structures a reality and prevent researchers from drowning in this flood. "The upshot," says Almo, is that "we can do structural genomics on the same scale that people do genome sequencing."

The crystallographic platform

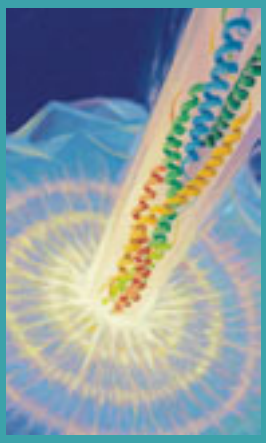
Synchrotron X-ray sources could provide the time reduction and resolution enhancement necessary for accelerating proteomics down the crystallography road. Although the likes of Abbott Laboratories' Automated Crystal Transport, Orientation, and Retrieval (ACTOR) system provides the software needed to make the most of the X-ray sources in digging out the atomic coordinates from a protein crystal, the BLU-ICE X-ray data collection software developed by Stanford Synchrotron Research Laboratory scientists in Palo Alto (CA) offers a model system for taking crystallography to the high-throughput level, according to Stewart and colleagues.

The use of bigger and better X-ray diffractometers, including those with synchrotron sources, then means that the same or better data are available from small crystals as were previously possible from a much larger chunk of crystallized protein only a couple of years ago. Indeed, conventional single-crystal X-ray crystallography requires a crystal between 100 and 300 μm long, but bright, focused

X-rays such as those sourced by Lawrence Livermore National Laboratory's Advanced Light Source can garner structural information from crystals around the 50-µm scale.

The AutoSolve software from Astex Technology (Cambridge, U.K.) automatically analyzes and interprets the electron density data that emerge from a diffractometer, and company founder and CSO Harren Jhoti describes Astex's HTC-based methodology (called HTX) as a screening tool to discover novel lead compounds. This brings X-ray crystallography into the realm of lead discovery and away from its historical lead optimization role.

Indeed, Jhoti, writing with Astex co-founders Tom Blundell and Chris Abell (both at Cambridge University, U.K.), recently described the endeavor in a paper published in *Nature Reviews Drug Discovery*. "Once the structure of the target has been solved, virtual screening, coupled with high-throughput X-ray crystallography, can be used to identify one or more weakly binding small-molecule fragments from compound libraries that consist of hundreds of small-molecule fragments," they wrote. "High-resolution definition of these binding interactions provides an information-rich starting point for medicinal chemistry. X-ray crystallography can then be used to rapidly guide the elaboration of the fragments into larger-molecular-weight lead compounds."



As in other areas of drug discovery, the processes of synthesis, analysis, and screening of potential drug leads and their targets are being accelerated through high-throughput technologies.

Tom Peat of Structural GenomiX (San Diego) certainly believes that genome-scale protein structure information will "advance the productivity of drug discovery by improving target selection and validation, and providing templates for drug design." The bioinformatics team at Structural GenomiX is working on selecting targets and generating DNA primers while automated PCR and crystallization products are fed to their proprietary synchrotron beam line at the Advanced Photon Source of Argonne National Laboratory (Chicago) for X-ray data collection. "We are testing a number of different data collection strategies, most of which center on collecting [multiple- or single-wavelength anomalous diffraction] data on selenomethionine-substituted proteins," explained Peat at the 4th International Conference on Molecular Structural Biology in Vienna in September 2001. The presence of the selenomethionine group allows for easier structure solution by providing a known reference point.

HTC get together

A group of major pharmaceutical players hopes to rapidly drive forward the development of software that will minimize human intervention in HTC. Sponsored by software provider Accelrys (San Diego), the HTC Consortium was formed in June 2000 with a three-year first-phase plan. Current membership includes Abbott Laboratories, Bristol-Myers Squibb Co., Corvas International, Inc., Emerald BioStructures, Inc., eXegenics, Exelixis, Inc., Genencor International, Inc., Genetics Institute, Millennium Pharmaceuticals, Inc., Pfizer, Inc., Procter & Gamble, Schering AG, Tularik, Inc., and Vertex Pharmaceuticals, Inc., which are focusing on several key technology development areas.

The first goal is to produce the methods and protocols for automating as much of the structure determination and refinement process as possible. Phasing (including molecular replacement) can now exploit the computational power available in most laboratories to perform many thousands of calculations to search for an answer. This can be combined with methods developed by Accelrys, Inc., over the past eight years that have significantly advanced the automation of electron-density fitting and refinement. "The key," explains RiboTarget's Hubbard, who is also HTC Consortium director, "is bringing these different software components together in a streamlined system, with validation methods for checking how well the automation has worked, and appropriate heuristics for making decisions about the next calculation to try."

"There are many computational crystallography challenges in achieving true high-throughput structure determination, not least [of which is] developing metrics that can discriminate between when the automated calculations are giving the right or wrong answers," he adds. "But perhaps the biggest impact is going to be in providing these methods within an informatics environment that can track and manage all the data that is being generated and provide the direct link into medicinal chemistry and the rest of the drug discovery process."

As in other areas of drug discovery, the processes of synthesis, analysis, and screening of potential drug leads and their targets are being accelerated through high-throughput technologies. "Structural genomics will parallel the developments in genomics," says Behnke. "At first, simply sequencing the genes was difficult, and now the big effort is combining the sequencing data and meta-analysis." The same is true for HTC, in which "the analysis of the structures of many [protein] family members in combination with a number of specific and nonspecific ligands will teach us an immense amount about how drugs interact with their protein targets, and how these protein families work," he continues. "This will lead to improvement in drug design and to technologies for predicting ligand–protein interactions, and that feeds into the virtual ligand screening process."

Further reading

Abbott, A. Proteomics: The society of proteins. *Nature* **2002,** *417,* 894–896.

Blundell, T. L.; Jhoti, H.; Abell, C. High-throughput crystallography for lead discovery in drug design. *Nature Rev. Drug Disc.* **2002,** *1,* 45–54.

Henry, C. M. Structure-based drug design. *Chem. Eng. News* **2001,** *79* (23), 69–74.

Lamzin, V. S.; Perrakis, P. Current state of automated crystallographic data analysis. *Nature Struct. Biol.* **2000,** *7,* 978–981.

Stewart, L.; Clark, R.; Behnke, C. High-throughput crystallization and structure determination in drug discovery. *Drug Disc. Today* **2002,** *7,* 187–196.

David Bradley is a freelance science writer at sciencebase.com in Cambridge, U.K. Send your comments or questions about this article to mdd@acs.org or to the Editorial Office address on page 3. ■