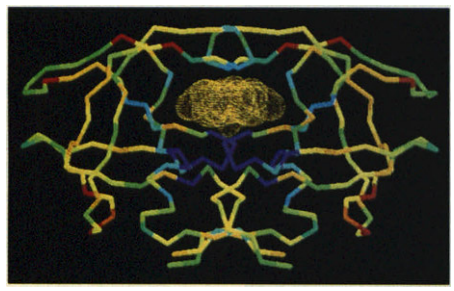
Bioinformatics

On 10th June, Silicon Graphics announced its largest single sale of workstations in the UK: 100 workstations and two servers bought by Glaxo Wellcome Research & Development to expand access to bioinformatics data. There can be little doubt that the computer has become the bench scientist's most important tool for drug discovery.

Although molecular biology has played a role in drug discovery for many years, the wide availability of databases for gene and protein sequences, such as Genbank and the Brookhaven Database, has raised it from an exotic speciality to a necessary tool. With this change has come an increasing requirement for bioinformatics if researchers are to manipulate and compare sequence data in the search for similarities, targets and models. This validation of gene targets demands sophisticated computing if sequence and structure data are to be integrated, molecules are to be modelled, and patient data are to be made accessible.



Glaxo Wellcome is exploring several approaches to bioinformatics, ranging from specialist applications requiring intensive support to robust and user-friendly programs designed for general biological applications. All scientists who need to analyse molecular biology and protein structure data use bioinformatics. But whereas genetics staff may use the system several times a day, molecular biologists may only need it once or twice a week. Even this will change, however, as bioinformatics, like e-mail, becomes part of daily research life.

User-friendly programs have been developed so that scientists from many disciplines can access appropriate information with a minimum of training. The latest innovation is the creation, with Oxford Molecular, of an on-line bioinformatics library that also provides access to public software, data bases and on-line resources for the storage, retrieval and analysis of genetic information.

In the early stages of a drug discovery project, bioinformatics is used to collect data about a particular disease or biochemical process involved in pathology. During this phase, scientists must use many different data collections, with information on disease phenotypes, genetic linkage, genomic sequences, protein sequences and structural data. Many of the data collections are run by universities, hospitals and government laboratories, although valuable information may also be held in the company's own data banks.

One such project is the Alzheimer's research programme at Glaxo Wellcome, which culminated in the recent discovery of a gene responsible for early-onset Alzheimer's disease. Not only did bioinformatics investigations suggest that the gene encodes an integral transmembrane protein, they also drew attention to a related protein in the nema-tode *C. elegans* involved in intracellular signalling, suggesting that the encoded protein might be involved in the protein trafficking pathways responsible for processing the amyloid precursor protein in Alzheimer's.

Glaxo Wellcome scientists also use bioinformatics to model disease processes, often based on incomplete information. Sequence data from several microbial genome projects, for example, can be analysed automatically and systematically using standard bioinformatics tools. Not only can the user specify a target organism and determine whether a metabolic pathway of interest is present, but the program can fill gaps in the available genomic information using sequences from the nearest available evolutionary neighbour. Although further experimentation is required to validate the results, these techniques can be used to evaluate the utility of different experimental approaches or to eliminate experiments that are unlikely to provide useful information.

But bioinformatics is not only about analysing gene and protein sequences. To exploit its power in managing data generated using more traditional drug discovery approaches, another project underway at Glaxo Wellcome aims to use bioinformatics to link receptor sequence data and small-molecule activity. When a related protein is found, libraries of small molecules can then be trawled for likely activity, directing searches to those molecules active against the original receptor.

The fundamental scientific information available to the pharmaceutical industry is undergoing an explosive increase. Bioinformatics will be a major part of the information technology systems that make this information available for drug discovery. □

Combinatorial chemistry

Combinatorial chemistry is based on the simple premise that the greater the diversity of compounds tested, the better the chance of finding one that can be developed into a drug. Last year, Glaxo Wellcome demonstrated its belief in the technology by acquiring the pioneering company Affymax. In addition to the purchase of Affymax, Glaxo Wellcome has assembled strong combinatorial chemistry teams at its major research sites in Europe and the USA.

Combinatorial chemistry can improve drug discovery by

- increasing the efficiency with which novel leads are generated
- assisting the optimization of previously identified leads
- generating molecules for target validation, independent of their value as potential drug candidates.

An early example of combinatorial chemistry involving peptide synthesis produced a library of more than 25 billion different compounds; it was built in the pursuit of synthetic vaccines. Although they are relatively simple to make, peptides are not very useful to the drug industry.

In the past few years, combinatorial chemistry has moved ahead rapidly, and a combination of improved chemistry techniques and automated instrumentation has now made it possible to generate many different classes of drug molecules. Affymax scientists have devised combinatorial syntheses of highly functionalized pyrrolidines, 4-thiazolidinones and β-lactams, for example, all in formats amenable to creating and screening libraries of tens of thousands of compounds. To apply combinatorial chemistry efficiently, the technology must be integrated with engineering and instrumentation to facilitate the synthesis and screening of the library.

Glaxo Wellcome is encouraging its scientists to develop a broad range of combinatorial technologies, with different approaches being developed at Affymax in Palo Alto, California, at Research Triangle Park in North Carolina, and at Stevenage in the UK. From Glaxo Wellcome's point of view, this is not overkill, but reflects the view that any one technique is unlikely to be ideal in all situations. The technology is developing at breakneck speed, and no one can now say which of the techniques being developed will eventually prove best.

Solution- and solid-phase libraries

At Stevenage, Glaxo Wellcome's chemists lead the world in solution-phase libraries. The attraction of solution-phase synthesis