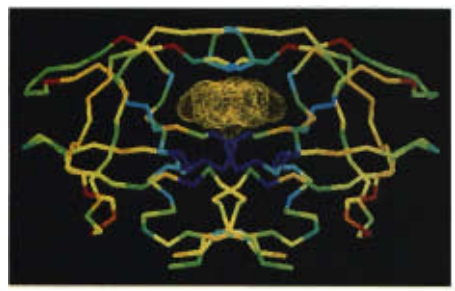
Bioinformatics

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

is that it capitalizes on the vast range of solution chemistry available in the chemical literature.

But solid-phase synthesis, too, has its strengths, in particular facile purification and easier automation. At Research Triangle Park, the strategy is thus to investigate how to construct novel pharmacophores and building blocks with which to make drug-like organic compounds, using solid-phase supports.

A major challenge is finding what these large libraries contain, and Glaxo Wellcome is making rapid advances in methodology for monitoring solid-phase chemistry. Affymax has developed a method for obtaining high-quality ^{1}H NMR spectra of organic molecules covalently attached to beads using magic angle spinning in a conventional spectrometer, which allows the non-destructive monitoring of solid-phase reactions. An alternative approach entails the use of ^{13}C-enriched building blocks coupled with ^{13}C NMR spectroscopy, which again permits very rapid analyses of the reaction during synthesis without destruction of precious samples.

Glaxo Wellcome is also looking at the opportunities created by the various uses of combinatorial chemistry. Large libraries, such as those based on beads, are appropriate for initial screening to identify lead compounds, while smaller libraries can be used for project-specific optimization of leads.

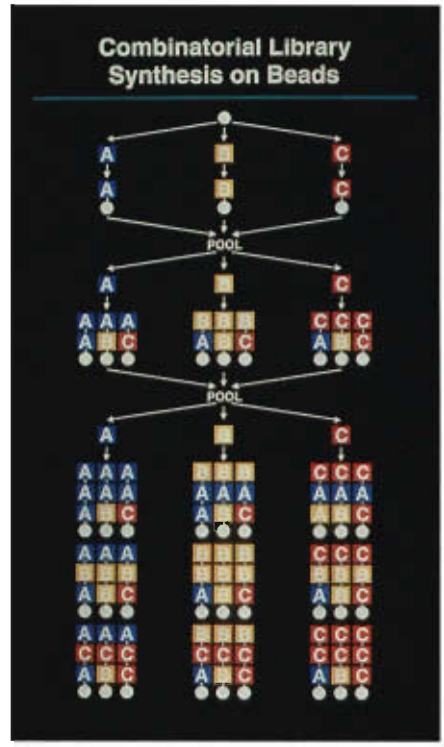
At Affymax, researchers are tailoring systems to particular needs. If, for example, there is no information on compounds interacting with a target of interest, such as the protein encoded by a gene sequence from the human genome programme, then the best approach would be random primary library screening (RPLS). For RPLS, high chemical diversity is required, so multiple libraries, each with 50,000 or more molecules, may be needed.

To produce such libraries, automated equipment controlled by custom-designed software is used to perform many reactions in parallel, maximizing the number of compounds produced by using a split-pool synthesis with bead supports. Once a particular chemistry has been adequately rehearsed, one chemist may be able to synthesize a library of 50,000 compounds in a day.

The task is then to isolate the few hits in such a library. If the target can bind to a library molecule while still attached to the bead, then extracting the desired bead and compound from the library should be relatively easy. Some companies are already taking this approach, using fluorescent targets, picking out the fluorescent bead–ligand combination, and analysing it by mass spectroscopy. This may require a fairly large bead, however, necessitating large quantities of reagents, and thus increased costs, as well as limiting the number of molecules that can be synthesized in a particular library.

Tagging beads

To permit smaller beads, Affymax has developed an approach which gives each



bead a unique code describing its history during the synthesis. Clearly the chemistry used to generate the code must be inert to the chemistries used to generate the combinatorial library. An early example used oligonucleotides, produced by a route that had no effect on the synthesis of the library and vice versa, each building block in the library being associated with a unique codon in the oligonucleotide attached to the bead. Following screening, the oligonucleotide attached to the active bead could then be amplified by PCR and sequenced, identifying the structure of the active compound.

Some chemistries, however, are incompatible with oligonucleotide synthesis. To cope with these situations, Affymax has developed an alternative approach, called a hard tag protocol, using robust chemistries that are resistant to degradation. The process involves building a polymeric tag which forms secondary amines reactive towards dansyl groups when degraded by acid. These can then be identified by their retention times on HPLC. In this case, the building blocks used are encoded by the particular secondary amine used. Whatever the encryption scheme used, however, its presence means that huge libraries can be made and screened in small volumes, greatly reducing chemical consumption and providing a significant advantage over simpler combinatorial technologies.

Using this principle, Glaxo Wellcome can create encoded synthetic libraries (ESL) in one of two ways. With tethered ESL (tESL), compounds remain on the beads throughout, and the active compound is identified by separating and decoding its bead. Although this is a rapid screening method, it does require that the active compound can bind its target (which must be in solution) while attached to its bead.

To overcome these limitations, Affymax has developed soluble ESL (sESL), a further variation on the theme with much broader applications. In this case, the compound is cleaved from the bead to yield an aqueous solution of the molecule, making it accessible to any pharmaceutical assay. One version of this approach uses a photochemically cleavable linker between each bead and its compound, allowing precise cleavage to release only the desired fraction of the chosen compounds.

Using the split-pool approach with beads on an instrument with 36 reaction vessels, 46,656 compounds can be made in three chemical steps from three classes of building block. Such a library can then be screened by aliquoting pools of, say, 100 beads, each bearing a unique compound, cleaving 50% of the compound from each bead using UV light, and removing the resulting solution for screening. Beads from any pool that gives a positive result can then be redistributed individually, using special bead-dispensing equipment designed by Affymax, before cleaving the remaining compound from each bead and repeating the assay. Active beads can then be decoded, and the compound resynthesized to validate the result.

Directed primary libraries

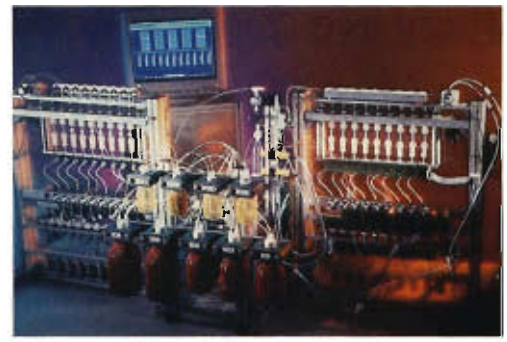
When chemists have some idea of what kind of molecule might interact with a target, such as X-ray data relating to the target, a different approach can be pursued. Then the likely choice will be a focused or directed primary library. Such libraries may be an order of magnitude smaller than in RPLS.

If the target is a metalloproteinase, for example, a zinc chelating group would be a good starting point. Prior knowledge is thus used to increase the chances of success in a focused library. Libraries for focused primary screening may contain only a few thousand members, often all containing a common pharmacophore that has historically yielded compounds active against related targets. Beyond the shared pharmacophore, maximum chemical diversity can be generated in a way that is believed to be compatible with retention of the pharmacophore's activity.

It is now over a year since Glaxo Well-

come acquired Affymax, and the benefits are already apparent. The parent company's chemists from Research Triangle Park, Stevenage and Verona are being trained in solid-phase chemistry and automation at the Affymax sites in Palo Alto and Santa Clara. Stevenage chemists have already built two Affymax ESL synthesizers, which are now producing libraries; a similar technology transfer is also under way at Research Triangle Park. In addition, many Glaxo Wellcome biological screens have been transferred to Affymax.

One of Affymax's major attractions to Glaxo Wellcome was that its approach goes beyond organic chemistry: it integrates combinatorial chemistry with engineering and instrumentation associated with novel assay formats. Affymax is now charged with inventing enabling technologies for efficient drug discovery and transferring them to the rest of Glaxo Wellcome for exploitation. The combinatorial revolution has shaken up traditional approaches to the final stages of lead optimization, and the Group's Combinatorial Lead Optimisation Programme is testing just how far down the drug discovery and development pathway the new technologies can be applied.



Lead optimization has traditionally relied on using single-compound synthesis, but this technique is rapidly being replaced by parallel synthesis of discrete compounds. One approach is array chemistry, a system involving dozens of parallel reactions for establishing structure/activity relationships. Glaxo Wellcome has developed proprietary parallel reactors, with efficient heating, cooling and stirring capability, and these are spreading rapidly throughout the labs. Automated, and therefore faster, purification equipment has also been developed. Automated HPLC means that large batches of compounds can be purified each day, a previously unthinkable feat.

Glaxo Wellcome is equipping itself to remain the leader in combinatorial chemistry, and its latest research projects are heavily influenced by the new technology. But the traditional methods that served it so well in the past are not being abandoned. Instead, the company is testing the new technologies, and integrating those that work into its activities throughout its research laboratories. □

Systematization of research

Systematization of research follows naturally from the acceptance that knowledge gained from one drug target can be transferred to related targets. While hardly a startling concept, previous drug discovery projects tended to focus single-mindedly on one target. This concentrated minds on the matter at hand and led to many notable pharmacological successes, but it is an inadequate strategy for responding to today's biomedical opportunities. Systematization draws on knowledge of the human genome by combining advances in gene expression, automation, combinatorial chemistry and bioinformatics.

Over the past ten years, numerous families of genes have emerged. The majority of drug targets, however, come from only four of these:

- G-protein-coupled or seven-trans-membrane-domain (7 TM) receptors (estimated total: 5,000)
- Nuclear (hormone) receptors (estimated total: >150)
- Ion channels (estimated total: 1,000s)
- Enzymes (estimated total uncertain because of low homology between members).

Of the top 100 pharmaceutical drugs, as defined by the International Marketing Survey audit sheets in 1995, 18 bind to 7 TM receptors, 10 to nuclear receptors and 16 to ion channels. The remainder generally inhibit enzymes.

The system-based approach attempts to transfer the knowledge gained from working on one drug target to other related targets. Much of the molecular technology required to work with one 7 TM receptor, for example, is useful for other 7 TM receptors, including cloning and expression systems, together with information about their structures and ligands.

Systematization requires a significant commitment of time and resources. It allows efficiencies to be gained through economies of scale, but only if the target families are of significant size, richness and diversity of therapeutic value. For this reason, not all receptor and enzyme classes are candidates for this approach. In addition, a "learning opportunity" is created whereby past successes allow rapid attack on new targets.

At Glaxo Wellcome, the system-based approach is represented by two target classes in particular: 7 TM receptors and nuclear receptors. Both classes are appropriate, as they have already proven to be a rich source of valid drug targets with small-molecule ligands that are bioavailable, non-toxic and efficacious. In addition, generic technologies for the efficient discovery of new agonists and antagonists for these receptors are well advanced, and many validated targets within these receptor families remain for which no drugs currently exist.

Glaxo Wellcome already has leading drugs aimed at 7 TM receptors, including salmeterol (a β_2 adrenergic receptor agonist used to treat asthma), ranitidine (an H_2 receptor antagonist that blocks acid secretion), and sumatriptan (a $5HT_{1D}$ agonist for the treatment of migraine). The company thus has a database of molecules already directed at these proteins, making the chemistry necessary to reach untapped receptors particularly suitable for the new approach.

Unlike membrane-bound receptors, the nuclear receptors are intracellular and control the activity of target genes directly. One of the first results of applying the systems-based approach to this class of target at Glaxo Wellcome has been the demonstration that thiazolidinediones are potent and selective activators of peroxisome proliferator-activated receptor γ (PPARγ), a nuclear receptor recently shown to function in adipogenesis.

Thiazolidinediones are known to have antidiabetic properties, increasing the insulin sensitivity of target tissues in animal models of non-insulin dependent diabetes mellitus. *In vitro*, they promote differentiation of preadipocyte and mesenchymal stem cell lines into adipocytes, but the molecular basis for this effect was unknown. The finding that they activate PPARγ not only provides the first known high-affinity ligand for the receptor, but also strongly suggests that PPARγ is a molecular target for the adipogenic effects of thiazolidinediones, as well as raising the possibility that PPARγ is a target for their therapeutic actions.

By using the systems-based approach, Glaxo Wellcome intends to be in a strong position to generate useful drugs against novel targets as they are discovered. It will also have access to a rich source of quality ligands for validating orphan receptor function. □

Glaxo Wellcome would like to thank the following for helping with the supplement: Jürgen Lehmann, Alan Baxter, David Brown, Philip Connolly, Mario Geysin, Michael Hayes, Russell Howard, Jonathan Knowles, Melanie Lee, Andrew Lyall, Nuala Moran, James Niedel, Elizabeth Rees, David Saussy, Joel Shaffer, Robert Short, Mike Tarbit, Michael Ward, Emma Weitkamp, Russell Williamson.