easy to do protein and nucleic-acid alignments, as well as to design primers, to search motifs and to perform multiple sequence analysis. One example is MacVector, for the Apple Macintosh and the Windows version DS Gene, produced by Accelrys in San Diego, California. MacVector has been available for many years and is continually being improved. The reasonably priced Jellyfish from LabVelocity in San Francisco, California, is also available for both Mac and PC. Like MacVector, Jellyfish will generate primers, oligonucleotides, cloning constructs and restriction maps, as well as perform BLAST searches and sequence alignments. When sequence data are downloaded, so are the annotations. Another option is the Visual Cloning 3 package from Redasoft in Bradford, Ontario, which the company claims simplifies sequence analysis, sequence editing and presentation, as well as offering access to online tools through its integrated web-browser interface and an array of plain-language 'wizards' to guide the user through the process.

Proteomics has benefited from software that not only allows rapid analysis of two-dimensional electrophoresis gels, but also draws together other functions, such as data mining, into one easy-to-use package. Progenesis from Nonlinear Dynamics in Newcastle upon Tyne, UK, comes in three modules. The first allows data annotation, the second does the image analysis and data logging, and the third is the data-mining component in which, for example, the data from different gels can be compared. Users can generate pick-lists of interesting spots


Making sense of 2D gels.

from the data-mining results and send them back to the image-analysis module to drive spot-picking from the original gel.

Paddy Lavery, Nonlinear's bioinformatics marketing manager, says that a future version of the software will allow users to import raw mass-spectrometry traces or peptide map lists into the program and to search against internal or external peptide sequence databases. The user will be able to store the search results and link them back to the gel and the sample ID of any spot in the pick list.

## Making predictions

As the number of sequences of known function increase, predictions of genes, protein structure and protein function from sequence information are getting more accurate. There is a trend towards collections of integrated, coordinated suites of gene-prediction programs, many of which can be tried out on the web. Softberry, for example, offers a number of gene-prediction programs that can be accessed over the web, including FGENESH,

which the manufacturers claim is fast, sensitive and accurate. The suite also includes FGENESH_C, which searches for similar cDNAs, and FGENESH+, which will find similar proteins. The fully automated FGENESH++C will automatically annotate any genome (other than human) to a standard claimed to be indistinguishable from manual annotation, using a battery of complementary techniques.

The accuracy of programs that predict protein structure from sequence has improved over the past few years and these are gradually becoming more widely used. A number of easy-to-use academic and commercial protein-structure prediction programs are available. PROSPECT Pro from Bioinformatics Solutions uses the 'threading' method, which threads the query sequence onto all known protein folds from the Protein Data Bank to find the one that gives the best-fitting structure. The program builds on this established strategy by allowing the user to feed in experimental data, such as constraints on the threading, and it checks its own results using a neural network.

Predicting the localization of a protein within the cell is also a help in identifying the function of a gene product. Softberry's ProtComp illustrates the trend in bioinformatics software to integrate diverse computational approaches. The program uses clues in a protein's sequence to guess at where it is localized within a cell, mostly by using neural networks to check sequence elements for tell-tale localization-specific motifs.

Another important issue is monitoring and controlling the flow of data as they are

# GENOMIC MERGERS

According to Celera Genomics in Rockville, Maryland, the next step after the sequencing of the human genome will be 'merging technologies'. Celera, recently forged agreements with Applied Biosystems in Foster City, California, under the umbrella of the Applera Corporation to work towards integrating all aspects of genomics. The companies are now developing an array of predesigned and prevalidated assays for genes identified on completion of the sequencing of the human genome. There are already assays for more than 18,500 human genes known to be expressed, as well as kits for over 125,000 single-nucleotide polymorphisms (SNPs) — their goal is to have 200,000 SNP assays by the end of the year.

"The most common complaint we hear from scientists is: 'but we're not bioinformaticians'," says Ramin Cyrus, a senior director for Celera. Anthony Kerlevage, a senior director at Celera, likens the situation to modern word processors. "They are packed with features, but most of us just use them to write letters," he smiles. The solution? This month, Celera and Applied Biosystems launched the 'myScience' portal. This website allows users to upload data from instruments, databases or laboratory information management systems (LIMS), and store them in personalized workspaces. Users can then analyse their data with an array of tools, guided by predesigned workflows. These workflows act like 'wizards' — guiding the


Which gene? Assays for human genetics are big business.

user through the options of which web-based tools are available. The site itself is free (Applied Biosystems hopes that it will tempt users to purchase the company's assays), and is designed to complement the subscription-based Celera Discovery System, which offers deeper analyses. **S.B.**