



# HIGH THROUGHPUT SCREENING

**Mike Rippin**

**TESSELLA SUPPORT SERVICES PLC**

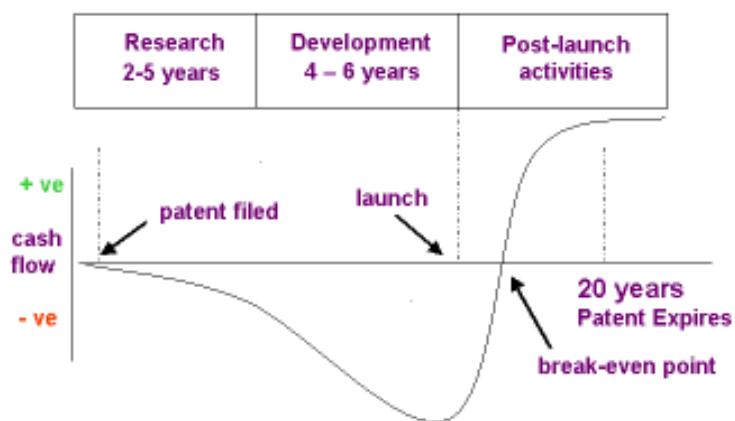
Issue V1.R1.M0

May 2003



## Background

Drug discovery and development is big money and big business. The benefit of being the first to market with a new drug can be as much as \$1bn revenue in the first year, with 75% market share for the lifetime of the drug. Being first to market is an essential objective for modern pharmaceutical companies to survive in this high risk business. Any process which may assist in speeding up time to market and concurrently optimising the revenue earning potential of the patent lifetime, is actively pursued and developed in the pharmaceutical industry.



The Drug Discovery process has evolved historically – originally relying a great deal on serendipity in the 50's and 60's, moving on to a more structured approach in the 80's which looked in detail at agonist-receptor interactions, gaining structure and activity information around compounds to feed forward in the Discovery process.

The modern approach to drug discovery, made possible in the last few years by great advances in technology, is very much “disease-oriented”. An attempt is made to understand more about a disease and the elements which cause it and to use this as a way of screening suitable compounds. This has been made possible due to advances in molecular and cellular biology, genetics and genomics.

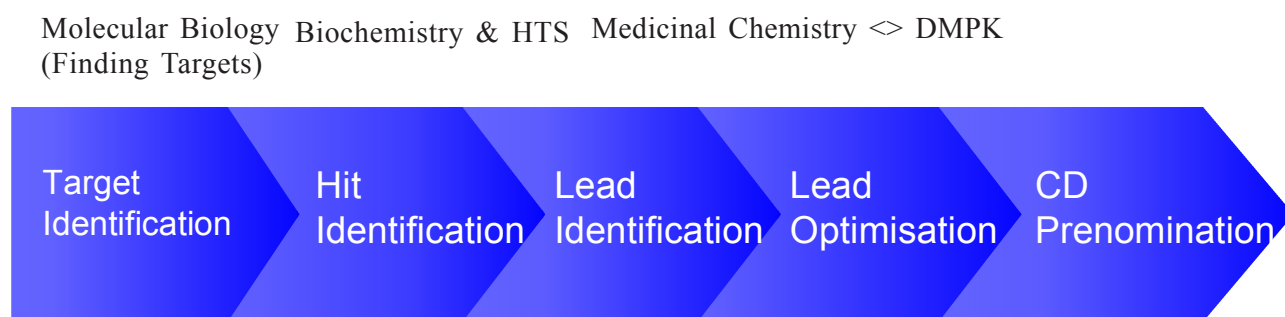
However, this approach still requires that the widest possible diversity of candidate samples be screened. One of the original drawbacks in the early days of drug discovery was lack of access to this diverse library. High Throughput Screening (HTS) refers to a modern process whereby *thousands* of samples can be screened *each day*, against a given (known or unknown) target (disease). This process has led to a significant improvement in “Lead Identification” – those samples which may eventually become candidate drugs – by firstly improving the spectrum of sample structures believed to have some possible effect and secondly by quickly eliminating

those samples which have no effect.

HTS is usually involved in the first phase after a target has been identified and is currently one of the few areas in the whole drug discovery process where significant time savings can be made in finding candidate drugs. For this reason HTS has seen widespread adoption amongst all of the major pharmaceutical companies.

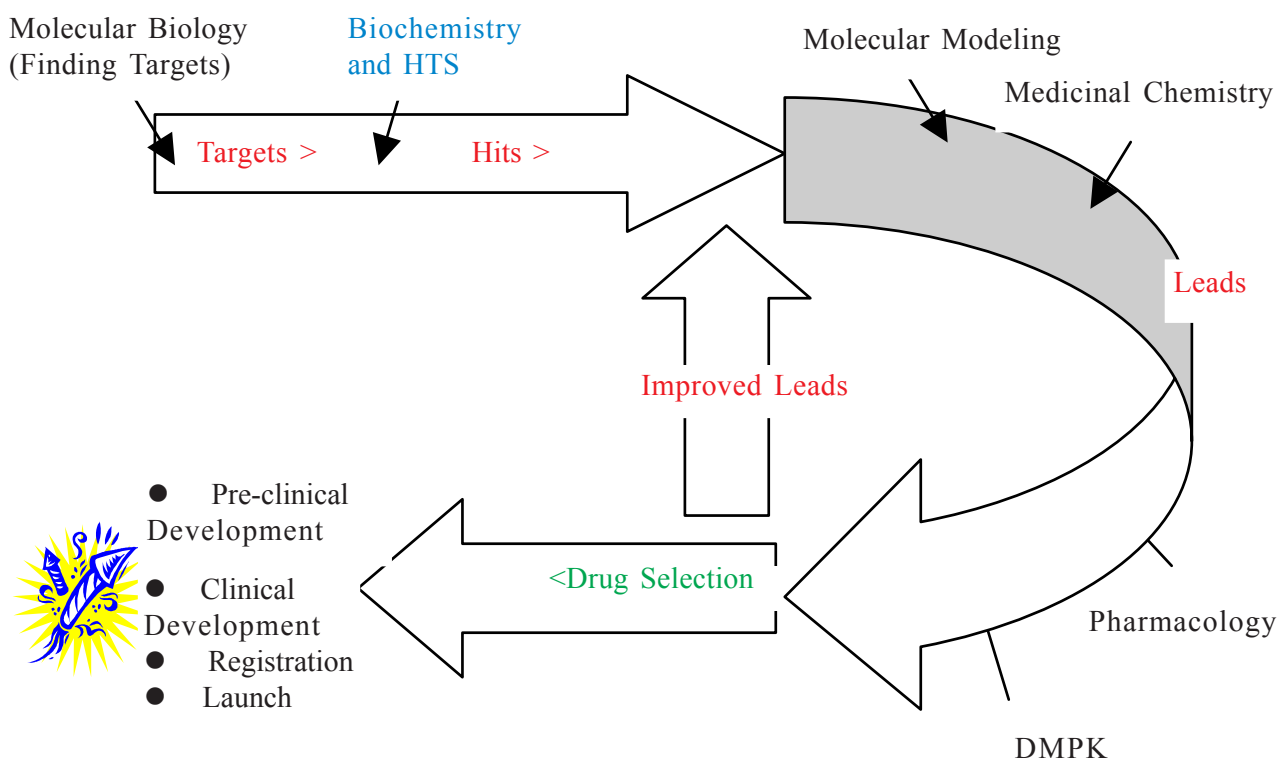
### Drug Discovery and Development Process

The modern Drug Discovery process has a number of stages, shown simplistically below:



- Identify potential “targets” – Molecular Biology
- Identify potential “hits” – Biochemistry and HTS
- Identify Leads – Structural Biology and Molecular Modelling
- Optimise Leads – Medicinal Chemistry and DMPK
- Drug Development:
  - Preclinical Development
  - Clinical Development
  - Registration
  - Launch

The Discovery process is ‘cyclical’. The optimisation stage leading to candidate drugs (CDs) is iterative, as the compound is successively modified in some way and re-analysed and tested.



The processes of CD pre-nomination and the final release of a drug, are lengthy indeed and governed by very strict rules and regulations to ensure drug safety and effectiveness (see earlier diagram). There is less scope for reducing the time to market in this stage of the Drug Development process. However, in Drug Discovery, new technologies *are* providing a means for candidate drugs to be discovered and developed more quickly. In the following sections we discuss what is involved in HTS, what the benefits really are and importantly where the discovery process needs to go next in order to fully maximise the potential of this technology.

### HTS in the Drug Discovery Process

HTS is usually the first stage after Target Identification. Some target has been identified which either causes, or is a component in, a disease. The aim is to find a drug which negates the effect causing the disease. As was historically the case, this is still very much an approach whereby a number of compounds are “screened” against the target and the ones showing the biggest positive effect are taken on for more detailed analysis. However, modern HTS differs in many respects:

- Principally the volume of compounds screened is enormous, thousands or tens of thousands a day.

- ❑ The initial selection of compounds to screen is not random or based on some tentative understanding of an existing drug's previous efficacy for some other disease. Modern techniques study the target as well as systematically characterising all compounds screened against all targets and this information is used to identify the most suitable sub-set.
- ❑ The process is highly automated using sophisticated laboratory equipment, robots and chemical analysis techniques.

### Screening

Screening is the term given to the process of analysing the “effect” of a series of compounds on a given target. The “effect” can vary depending on the stage of screening, or the nature of the target. The term screening is independent of volume, single compounds can be screened in isolation.

Screens will be run by loading up multi-well plates with as many samples as required, one in each well. Plates come in a number of sizes, the most common being 96 and 384 wells. However, modern assays are so sophisticated that plates with as many as 3456 are used. In addition, several plates may be created for a single screen, sometimes *hundreds* for a large HTS screen. The distribution of samples within and across plates can also vary.



Given a set of plates, automated screening systems have been developed to run the experimental measurements. These systems employ robots for sampling wells on plates in the correct order and running the appropriate physical analyses. The nature of the physical measurements varies as well, though fluorescence, radioactivity and LC/MS (Liquid Chromatography / Mass Spectrometry) measurements are common.



Low Throughput Screening (LTS) describes the screening of between 1-100 compounds in one pass and is typical of the screens run during Lead Optimisation. In LTS, potentially quite detailed experiments will be run to characterise accurately the effect of a compound on a target.

### **The Primary Screen**

HTS in the first instance generally attempts to screen thousands of compounds at a time with a view to obtaining one simple data value for each sample at a single fixed known dose – whether it is “active” and has an effect on the target, or it is “inactive”. It is this simple “pass-fail” test that makes the process of screening 1000’s of compounds at a time possible. The subsequent process of sample elimination can be automated, given rules about what constitutes the threshold between active and inactive. These kinds of screens are often referred to as primary screens, they will likely be the first screens run against the target of interest. This “pass-fail” approach raises questions:

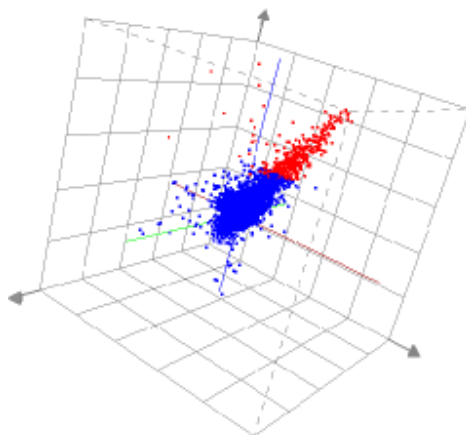
- What should the threshold be set to?
- What about experimental error? Some compounds may be classed as inactive erroneously.

The latter problem can occur for a number of reasons:

- Sample degradation before the screen is run
- Errors in sample quantities
- Machine error in measuring the appropriate physical properties
- Edge effects or other drift effects due to the position of a well on a plate.

One approach to compensate for this is to run all plates in triplicate, such that every compound is measured three times. If a given sample has any real effect, it should

measure approximately the same result on all three plates. This provides some guard against bad plates, provides a simple mechanism for eliminating random results whereby every measurement for a sample is different (within some defined rules) and provides a simple average and accuracy estimate to work with when determining which samples to filter. A visual aid for viewing this data is to plot a 3D graph as shown below:



Blue represents the range selected and red the range deselected. Outliers are identified by off-diagonal axis data points. Such tools can be very useful in identifying good samples.

The output samples from the primary screen are called “hits” and these move into the next stage of processing.

### **The Secondary Screen**

Given a significantly filtered set of compounds from the primary screen (the hits), a secondary screen will be run which aims to investigate the effect of the compounds on the target in more detail. The physical nature of these screens will be different and more varied, and generally each compound will have a number of separate measurements made on it, for example, to build up a dose-response curve or a time-series measurement.

The processing of these secondary screens is still highly automated and aims to reduce the number of interesting samples down to about 5% of the original number. This process aims at turning the set of hits into a smaller set of “leads” – compounds which have potential for further development into drugs. This “hit to lead” (HTL) stage of the Discovery Process is a varied, yet still a relatively high throughput stage.

These leads then feed into the Lead Optimisation phase, during which the same lead compounds are modified and re-synthesised to produce similar variants which are themselves run through the HTS/HTL screens. It is during this stage that analysis and testing of the smaller sample sets is carried out, for example, DMPK (drug metabolism and pharmacokinetics) determines as much about the compounds as possible as early as possible. This information, even if negative in the sense that it eliminates a particular compound, can be used in further cycles of the Lead Optimisation process by providing information about what characteristics that compound had which made it unsuitable, therefore affecting subsequent synthesis of refined leads.

### Design of Screening Experiments

As already mentioned, there are a number of experiments performed on plates of samples. However, there are a number of common features to all plate based experiments which aim to ensure accuracy and quality control:

- When looking to measure a physical activity, a plate will be designed so that some of the wells will have “known” responses. These will act as quality controls to ensure that either there were no errors in dosing or measurement, or to normalise out the results across the plate.
- There will be wells containing no samples as a way of measuring “background” activity which is also used to re-normalise the data.
- There will be an arrangement of wells containing samples under test.
- This arrangement depends entirely on the type of assay and may be any one of the following:
  - One sample and one measurement per well – this will be the case in the primary HTS screen.
  - The same sample in multiple wells with different concentrations – this allows the potency of the compound to be determined and is investigated in secondary screens or HTL experiments.
  - One sample per well, but multiple measurements made on the same sample over time.

This is only a very basic set of plate-based assay characteristics, the full range of experiments being very large and expanding as new techniques are devised.

The kinds of experiments run in plate-based assays are simple, but often ingenious in their simplicity. Examples include:

- Inhibition through competitive binding to known receptor types, where the relative level of binding of a compound is measured against the level of binding of a known binding material.

- Inhibition of protein production – compounds are added to wells containing known live cells and the concentration of protein is measured after a given period of time. If a compound manages to inhibit protein production, the protein concentration will be lower with active samples.
- Time series measurements to investigate ion channel activity.

### **Benefits of HTS**

What then does high throughput screening offer?:

- The ability to screen hundreds of thousands of compounds in controlled and repeatable experiments.
- The potential ability to optimise the compound lead selection by eliminating compounds with no measurable activity very early on.
- The consequence of this is that more effective drugs can be developed more quickly.
- Higher levels of confidence in the selected CDs.
- The time to market can be reduced minimising costs and maximising patent lifetime.

### **What's the Catch? – Sample Selection**

Although HTS offers many things, there is one fundamental problem. Millions of compounds can be, and have been synthesised and are available for new screens.

When a new target is being investigated it is physically impossible to screen all compounds, even for relatively tight subsets, as there are just too many. Indeed, whereas HTS may be efficient overall, it is expensive, and unnecessary costs need to be kept down. This is accentuated with the inclusion of Combinatorial Chemistry, whereby tens of thousands of compounds may be synthesised based on a base chemical structure and a defined set of reactions with base reactants.

### **Sample Selection**

The question then becomes, how do we optimise the process of sample selection for the initial screens of new targets? If it is not possible to screen 10 million samples, the “most promising” 100,000 need to be identified and screened. This decision operates on two levels:

- Choosing which set of samples from existing known libraries of compounds.
- Choosing which of the CombiChem samples to synthesise in the first place, before screening.

In short – screen smarter, not faster. This subject is currently a hot topic in drug discovery, and a successful solution to the problem will lead not only to a reduction in the number of compounds to screen, but also an increase in the quality of those samples, further reducing the Lead Optimisation timeframe. Recursive partitioning using Genetic Algorithms and QSAR (Quantitative Structure Activity Relationship) analysis are two techniques for identifying “best” screening samples. Both of these methods build on the structure activity (SAR) data from previous screens, identifying sub-sets of compounds with the most appropriate chemical composition to work with the appropriate target, the target having been characterised in the Molecular Biology phase prior to screening.

In particular, recursive partitioning is a statistical technique used to build SAR models from HTS data sets and associated chemical descriptors. Virtual Screening (sometimes referred to as “*In Silico*” screening) is then used to find additional active compounds, the activity being used as a fitness function for a genetic algorithm (GA). This GA then attempts to find subsets with a higher probability of being active. Principally, this is done on a computer, no compounds need to be synthesised and no HTS screens need to be run. It acts as a kind of virtual pre-screen before the business of physical screening needs to begin.

### **Role of IT in HTS**

As a highly automated process, the computing requirements within the HTS discipline are all pervasive and include at least the following systems:

- Workflow management software for experiment definition and execution.
- Database systems for compound libraries and results storage.
- Data query systems on the chemical and results databases for compound analysis, comparison and selection.
- Virtual (*in silico*) screening
- Sample-selection system.
- Post-processing software to gather, analyse and aggregate plate data.
- Data visualization and reporting systems.
- Software to facilitate integration across assay disciplines, enhancing the decision making processes in lead optimisation.
- Integration software across business processes, for example, for sample ordering and plate dispensing.
- Software to control and interface to the automated screening systems.

The above list is just a small sample from all of the detailed systems required, but shows how fundamentally dependent the HTS process is on the underlying computer infrastructure – it would not be possible to process the quantities of data otherwise.

## Business Process Change

What are the consequences of introducing HTS? Every pharmaceutical company has a way of working which will have evolved over the years to match the needs of that company and optimise work efficiency. Introduction of a highly automated system into a largely manual environment has a number of business and IT challenges:

- ❑ The working practice on the ground may need to change to accommodate the new system i.e. new protocols, standards, training.
- ❑ The existing research environment itself may require a change to integrate the new HTS function – the way in which samples are ordered, the supply chain for dispatching samples or the “feedback” loop within Lead Optimisation.
- ❑ A re-evaluation of the chemical storage system may be required in anticipation of several million new samples swamping the database.
- ❑ Seamless integration of HTS results data with data from HTL and DMPK and its subsequent query and retrieval, may well require a redesign of the data model and results database.
- ❑ The addition of several new HTS experiments with their intrinsic automated nature, may require a re-evaluation of the business research model, the notion and definition of experiments and tests, the formatting and reporting of data or the incorporation of GxP.

Business Process changes such as these represent challenges somewhat beyond the IT infrastructure required to implement them, they may very well require cultural changes within the company. However, they are likely to be necessary.

## Database and Query systems

A well designed and optimally tuned database system is essential for the implementation of a scalable HTS function but has its demands, Huge data volumes, in multiple formats, with single and multi-valued results, need to be reviewed, summarised and analysed alongside results cross-discipline e.g. with HTL and DMPK data. The design of such a system needs to match the overall business model of the company, providing the required flexibility without constraining other areas of research. This is no small challenge but is key to a successful HTS venture.

## Virtual Screening

Virtual screening requires, as a key prerequisite, knowledge about binding criteria, e.g., knowledge about the binding site geometry, ligand binding modes, molecular similarity or bioisosteric functional group replacements. This information can then be used to predict the activity of the substrate. Virtual screening enables rapid computational screening of commercial databases, such as the Available Chemicals Directory database, in addition to corporate databases. When it is carried out for a corporate database in parallel with experimental HTS, virtual screening can prioritize re-testing of

single-point data hits and identify false-negatives for testing.

### **Data Acquisition**

Data may need to be imported from files generated by many different instruments. Data within the files may be organised in tables or plate-like arrays. Data for several plates may be stored in separate files, or in a single file for each experiment. The files will have different sized headers and footers, and the data may be located in different columns, with different separators. Data may also need to be imported directly from SQL databases.

Also, different types of plate may be used (i.e. different dimensions and numbers of wells per plate). Any satisfactory software tool will need to address these issues.

### **Data Processing**

This takes many forms, including:

- analysing and filtering the raw HTS inhibition data,
- graphing and processing HTL time series data,
- providing quality control systems which aid the diagnosis of potential sources of errors (due to edge effects, or reagent degradation, for example)
- normalisation of data
- provision of user-interactive systems allowing detailed modelling and aggregation of data at all levels of compound, plate and target
- providing assistance in identification of active samples.

A whole plethora of support tools may need to be developed to support these requirements.

### **Integration Infrastructure**

As touched on already, an HTS function is a fundamental part of the discovery process and as such needs to be integrated with it. This requires systems which interface both between research functions themselves and support functions such as compound management and ordering.

### **Data Sharing and Multi-Site Integration**

A primary objective of HTS is high throughput. This will generate a large amount of data for present and future consumption. However, this data will likely only be “local” to a given pharmaceutical site – stored on a local database, accessible on a local intranet. Data transfer and sharing *between* sites often occurs through email with Excel spreadsheet attachments or shared network drives. However, this rapidly becomes unfeasible with large volumes of data. It is also unfeasible to analyse the data against anything locally.

This is another challenge area for HTS – minimising the effort of duplication across sites, while maximising access to data in a seamless manner. Scientists on one site should not need to know that the data they are accessing derives from a site on the other side of the world and everyone can take advantage of the full library of compounds available in the company. Being the largest source of raw data, effective data sharing of HTS results is essential. The IT requirements for the development of a cross-site enterprise system are vast and complex involving novel technologies, unique design challenges and sophisticated distributed applications.

### **Surely Someone has done this by Now?**

There are available Commercial Off the Shelf (COTS) systems which go some way to providing “frameworks” for HTS systems. These systems provide:

- Workflow management tools for assay definition and tracking.
- Tools for raw data loading.
- Tools for data processing and final database uploading.
- Tools for results and chemical querying.

Successful implementation of such systems depends entirely on the precise requirements and existing business environment. Guidelines are difficult to present and several large pharmaceutical companies have *spent millions just evaluating* large commercial systems. As a general rule, however, “one size does not fit all” and a decision has to be made whether implementing a COTS system will complement or enhance existing working practices, or if it will force an unnecessary and oblique change so local business processes need to change to fit the system. The latter case would need to be considered very carefully. The total cost of ownership in a system like this could well exceed that of developing a new system from scratch, if large amounts of conversion and integration software needs to be written to implement it. COTS solutions can be extremely effective in areas where there is no defined process already in place, but if you already have existing systems, bespoke software developments to support and enhance HTS can be cost effective.

### **COTS vs Bespoke Solutions**

There are a number of arguments in favour of the bespoke approach to software:

- A bespoke solution will provide all the functionality required and will be tailored specifically towards the company’s preferences and working practices. Even a well-fitting COTS solution may only satisfy 80-90% of your requirements and might also require changes to your working practices. Competitive advantage can often be gained by using tools precisely optimised to your needs.

- ❑ The flip side of this coin is that a bespoke solution will **not** provide superfluous functionality. Users will not be confused by a choice of unnecessary options (typically, 70% of functionality in COTS products is not used by users). i.e. you are not paying for functionality you don't require and which might distract your users.
- ❑ New functionality can be added quickly in the future to meet changing requirements. It can take a very long time to get vendors to add functionality to COTS products (or even to fix bugs in them).
- ❑ Owning the source code and IPR to a bespoke application means that there are no ongoing licence fees to consider (which can be very expensive for some COTS). A system can be rolled out to extra users for nominal cost.
- ❑ Bespoke applications permit the use of novel analysis techniques, which may give competitive advantage. With COTS systems you are using the same algorithms as your competitors.

### **Tessella's Experience**

Tessella has worked for most of the major pharmaceutical companies and has been heavily involved in developing IT systems for HTS, HTL and DMPK departments. We are always interested in hearing about new situations where advances in software can bring real business benefits to the drug discovery process.

**Tessella Support Services plc**  
**Creating Software for Science and Engineering**

Tessella's services range from feasibility studies, through system design, development, implementation and ongoing support. Our expertise includes:

Data Analysis Software  
Data Capture  
Simulation Software  
Advanced Graphics  
Systems Support  
Database Applications

**Other Technical Supplements available include:**

- |   |  |
|---|--|
| <input type="checkbox"/> Archiving of Electronic Info   | <input type="checkbox"/> Object Oriented Programming   |
| <input type="checkbox"/> Active Server Pages            | <input type="checkbox"/> Pocket PC                     |
| <input type="checkbox"/> Automated GUI Testing          | <input type="checkbox"/> Portable GUI Development      |
| <input type="checkbox"/> Bayesian Statistics            | <input type="checkbox"/> Printer Technology Guide      |
| <input type="checkbox"/> Beowulf Clusters               | <input type="checkbox"/> Real Time Systems             |
| <input type="checkbox"/> C++                            | <input type="checkbox"/> Regression Testing            |
| <input type="checkbox"/> Client-Server Technology       | <input type="checkbox"/> Security and the Internet     |
| <input type="checkbox"/> COM                            | <input type="checkbox"/> Simulation                    |
| <input type="checkbox"/> Computational Fluid Dynamics   | <input type="checkbox"/> Soft Computing                |
| <input type="checkbox"/> Computer Image Processing      | <input type="checkbox"/> Software Design Methodologies |
| <input type="checkbox"/> Decision Support Systems       | <input type="checkbox"/> Software Development Cycle    |
| <input type="checkbox"/> Electronic Data Capture        | <input type="checkbox"/> Software Documentation        |
| <input type="checkbox"/> Electronic Lab Notebooks       | <input type="checkbox"/> Software Portability          |
| <input type="checkbox"/> Excel                          | <input type="checkbox"/> Software Re-engineering       |
| <input type="checkbox"/> Extending the Life of Software | <input type="checkbox"/> Software Specification        |
| <input type="checkbox"/> Federal Drug Administration    | <input type="checkbox"/> SQL                           |
| <input type="checkbox"/> FORTRAN 90                     | <input type="checkbox"/> UNIX Inter-Process Comms      |
| <input type="checkbox"/> Grid Computing                 | <input type="checkbox"/> UNIX Systems Performance      |
| <input type="checkbox"/> High Throughput Screening      | <input type="checkbox"/> UNIX Workstations             |
| <input type="checkbox"/> Instrumentation                | <input type="checkbox"/> Visual Basic 6                |
| <input type="checkbox"/> Integrated Lab Systems         | <input type="checkbox"/> WAP                           |
| <input type="checkbox"/> J2EE                           | <input type="checkbox"/> Web Services                  |
| <input type="checkbox"/> Java                           | <input type="checkbox"/> Windows 2000 Services         |
| <input type="checkbox"/> Lims                           | <input type="checkbox"/> XML                           |
| <input type="checkbox"/> Linux                          | <input type="checkbox"/> X Windows                     |
| <input type="checkbox"/> Microsoft Net                  |  |



INVESTOR IN PEOPLE

**Tessella Support Services plc**

3 Vineyard Chambers, Abingdon, Oxon, OX14 3PX, England

Tel: (+44) (0) 1235 555511 Fax: (+44) (0) 1235 553301

E-mail: [info@tessella.com](mailto:info@tessella.com) Web Address: <http://www.tessella.com>