Protein arrays for assessment of target selectivity 
transforming knowledge of the human genome into a lead optimisation tool

Genomics activity has been criticised for having elucidated a plethora of potentially interesting drug targets, each awaiting further validation, but having not yet revolutionised the pharmaceutical industry or replaced traditional drug discovery pipelines. Here, we detail applications for improving target selectivity of known therapeutic molecules to known targets by employing technologies not previously accessible to the pharmaceutical industry, ie appropriately-designed protein biochips. Protein micro-arrays can be used to screen lead molecules (antibodies, protein biomolecules and small molecule drugs), and iterations thereof, for the most specific target binders prior to toxicological, pre-clinical and clinical testing. Improvements in target selectivity linked to optimised recognition profiles following exposure to millions of potential binding sites derived from the human genome should translate into reduced adverse drug effects and shorter times to registration.

Proteomics is now generally hailed as the next phase of genomic discovery. Although there has been tremendous progress in the technologies employed for protein characterisation, protein and antibody micro-array technologies offer many advantages over traditional proteomics technologies. These include parallel analyte detection, miniaturisation, low cost, reproducibility, low level of operator expertise required for analysis, speed of fabrication, ease of distribution, reduction in analyte volume and sensitivity of detection – not to mention the ability to process large clinical cohorts measured in thousands of patients. The latter is critical to provide the statistical confidence necessary to afford accurate assessment of drug efficacy and/or detection of novel targets, particularly those associated with multigenic or low incidence disease entities. Protein and antibody micro-array technologies and their applications have been reviewed recently. However, like genomics, much of the current activity in proteomics has focused on the discovery of new targets or novel diagnostic markers for a particular disease entity.
Target selectivity screening and the ranking of lead molecules

If these same protein micro-array technologies can be applied to lead optimisation, then the impact on improved drug development can be brought substantially nearer term than that seen to date in the age of genomics. More important still is the ability to impact at the higher value end of the drug discovery chain, ie improved target selectivity during lead optimisation. Such deliverables become possible as a byproduct of a detailed knowledge of the human genome. Therapeutic molecules most often have their mode of action directed against the protein products of genes and not the nucleic acid code. Here, we will demonstrate the feasibility of transforming open reading frames detected within the human genome, either directly or through amplification of cloned complementary DNA, into recombinant proteins with a view to better detailing target recognition in the presence of an increasingly significant number of human recombinant proteins present on-array. Indeed, more than an
estimated 665 million different 5-mer epitopes or drug-binding sites could be contained on a single protein micro-array containing 5,000 different recombinant proteins or domains of 300 amino residues each (Gestel and Humphery-Smith, in preparation). A peptide array designed to display such diversity is not yet practicable with respect to the size of array required, time and cost. To afford a good representation of these binding motifs, a population of recombinant proteins is randomly and covalently immobilised in a three-dimensional hydrogel matrix atop of a glass substrate (Figure 1). For maximal utility to lead optimisation, the choice of proteins included on such arrays should include non-candidate proteins, known positive controls to allow ranking of results and an expanded Near Target Space (NTS), as shown in Figure 2. The objective must then be directed towards enhanced specificity of binders as part of lead optimisation (Figure 3).

The likelihood of unforeseen side-effects becoming apparent following clinical release of new therapeutic molecules should be reduced as a result of improved techniques for target selectivity optimisation. With the availability of such tools, chemical iterations of lead molecules and/or members of a particular family of molecules derived from screening chemical libraries should first be subjected to such screening. Examples relevant to the screening of therapeutic antibodies, protein biomolecules, and small molecule drugs are presented in Figure 4. Examples shown clearly demonstrate reliable mathematical ranking (ie on-array replicates) of an individual binder with respect to large numbers of potential targets. These reduction-to-practice experiments were conducted in parallel and combined with up to 12 on-array replicates to provide healthy levels of statistical confidence in the rankings obtained (Figure 4c).

The importance of optimised protein recovery and quality assurance of recombinant proteins employed in cross-reactivity screening

There is a well-known adage in analytical chemistry that states ‘garbage in equals garbage out’. Nowhere is this likely to be more true than in efforts to clarify target specificity in the absence of cross-reactivity. Cross-reactivity can be due to conservation of a particular binding site within the human proteome found on protein isoforms derived from the same ORF or as a result of sequence and/or structural similarity. Based on earlier work dealing with unique ‘signature peptides’9, predictions have shown numerous linear epitopes to be present on hundreds and thousands of occasions (data not shown) within the human proteome, not to mention those containing highly conserved post-translational modifications such as phosphorylation, glucosylation, myristylation, palmitoylation, etc. However, most critical is the ability to produce high purity, quality assured recombinant human proteins. This is a non-trivial exercise. Current practice involves a long list of quality control steps on our recombinant proteins to assure purity and fidelity of product. Noteworthy is that at every step there is an attrition rate. An example of end-product purity is shown in Figure 5. Laboratory-based in vitro molecular biology is far more error prone than a
Figure 4
Examples of fluorescent detection of A a highly specific recognition pattern for a high affinity antibody for x6 replicates of its target immobilised on-array in the absence of cross-reactivity for another 130 on-array targets; B the same result presented in A but plotted as a graph showing the lowest to the highest (left-to-right) affinity binders and their respective signal intensities and Standard Errors (dark bars); C a highly cross-reactivity protein/protein interaction based on up to 12 on-array replicates of a single protein (Bovine Serum Albumin) interacting with the same 130 different recombinant protein targets as in A; and D a small molecule of 400 Daltons along with a 390 Dalton linker molecule attached on the non-active side and subsequently linked to a fluorescent label following exposure to x6 replicates of 361 different recombinant proteins on a single array. The red circle indicates those binders showing above background target interactions, but less than the biotin-streptavidin positive control. It is noteworthy that small molecule/protein interactions will possess far lower affinities than the latter control positive

Average of ≥ 12 replicates plus standard error

Drug Discovery World Winter 2002/3
Figure 5
Optimised recovery of recombinant proteins visualised on silver-stained poly-acrylamide gels following dual affinity enrichment. Similar samples produce clean spectrographs by both MALDI-TOF MS and ESI-MS/MS.

References

Continued on page 24

similar process occurring in living cells. Molecular biologists know that results must be confirmed on agarose gels at every step of a cloning procedure, yet still errors persist and these must be discarded by methodical screening. Without this attention to protein purity, the results obtained for target selectivity are rendered immediately uninterpretable. Although induction can significantly upregulate the abundance of recombinant protein expression, other cellular constituents significantly contaminate signals obtained during cross-reactivity assessment, ie binding to impurities in the protein sample placed on array. Herein lies the need for routine dual affinity enrichment of recombinant proteins. The following steps are involved in quality assurance of recombinant proteins placed on arrays, namely verification of:

- PCR product – on gel.
- Entry clone – on gel.
- Expression clone – on gel.
- Vector design to ensure only recovery of proteins in the correct Reading Frame & absence of any read-through phenomenon.
- Dual affinity enrichment for enhanced protein purity.
- DNA sequencing of cloned insert (even if starting from fully-sequencecloned).
- Absence of 5' & 3' UTRs (untranslated regions).
- Protein purity and Mr – on gel.
- Concentration – level of expression & standardisation thereof across array.
- MALDI-TOF MS peptide mass fingerprinting.
- MALDI-TOF MS Total mass.

- ESI-MS-MS sequencing tagging (HTS implementation is currently problematic, but this may change in not too distant future).

This preoccupation with quality control must also be linked to a significant throughput of production, as potentially every user will possess different requirements with respect to the protein content associated with NTS. The protein inventory associated with non-candidate space can be increased through time, but one cannot afford to wait many years for the synthesis of a protein repertoire required for a specific application in lead optimisation and thus rapid, high quality synthesis of numerous proteins is an obligatory pre-requisite for the implementation of such technologies to lead optimisation. The latter involves expression of sequence homologues, tertiary structural homologues and tertiary homologues detected by threading algorithms. Some 200-300 such proteins are likely for any target molecule, particularly when the NTS is expanded by splice variants and the numerous potential post-translation modifications afforded by expression in multiple expression vector hosts, such as bacterial, yeast, insect and mammalian systems. In mid-2002, our production capacity in Escherichia coli was upwards of 1,000 successful recombinant proteins from any 1,500 randomly-chosen Human ORFs within 6-8 weeks following primer design and synthesis, whether the starting material was genomic sequence alone (ie in silico detected ORFs) or cDNA clones. Both have been reduced to practice, but the latter is associated with less attrition, particularly as a result of less undesirable PCR products. In all the quality control measures listed above one must expect to encounter attrition due to errors or low efficiencies obtained during amplification, cloning, transcription, translation and affinity enrichment. Successful production of an intended recombinant protein for chip-based applications is currently assessed as recovery of at least 100-200mg of protein. For other applications, lots of 10mg can be produced for applications such as immunogen production, immuno-arrays or structural studies.

Recombinant proteins are then immobilised on to a standard microscope format. Contract printing procedures allow for up to 5,000 to 6,000 different elements on chip, be they different proteins or more replicates of less proteins. The latter is most desirable if one is intending to reliably rank experimental outcomes. Nanotechnologies and non-contact printing methodologies can further increase the number of elements included on a
Proteomics

Continued from page 22 single protein biochip. The virtues of array-based assays with respect to many competing technologies include:

- Protein purity ensured (critical to data interpretation).
- Standardised protein abundance & accessibility.
- On-array replicates of target (reproducibility of assay).
- Inter-array reproducibility.
- Biomolecular interactions mathematically ranked.
- Inclusion of known target as positive control.
- Inclusion of target homologues to assess target selectivity.

In addition, an important advantage is the ability to titrate the concentration and time of potential ligands across the array, as opposed to the more simplistic ‘Yes/No’ responses obtained from techniques such as affinity capture or the yeast two hybrid approach. Protein-protein interactions are dependent upon time; concentration of both target and ligand; binding affinities, both on- and off-; the specificity or lack thereof for the association being studied; the physiological context, eg cleavage or activation of precursor proteins; and the influence of intra-cellular location. Furthermore, protein arrays have some conspicuous advantages over cell-based bioassays. These include the consistency and reproducibility of assay with respect to:

- Temporal expression, ie variation in heterologous DNA sequence means that maximal expression of recombinant proteins is rarely synchronous.
- Location of protein gene product.
- Target accessibility.
- Multiple batches of arrays constructed from the same high-abundance proteins being encountered in all cells and tissues (most evident on images of 2D electrophoresis gels). Variability in abundance and accessibility; cell and tissue heterogeneity; and can thus give rise to false positives on arrays. Can be contrasted with an undetectable signal that should have resulted from an interaction between a critically-important, low abundance, house-keeping gene interacting with its high-affinity binder. Using currently available technologies, the latter could go undetected during lead optimisation studies, and possibly even following toxicological testing and clinical trials with the resultant and obvious serious ramifications to patients and the pharmaceutical group involved. Until accessibility, homogeneity, purity and concentration of analytes are standardised, interpretation is difficult to interpret.


Caveats

No one technology is likely to supply the pharmaceutical industry with the knowledge required to confirm target selectivity with respect to all possible potential targets presented within the human proteome. Thus, one must insist that at all times results obtained on array are confirmed by orthogonal approaches both in vitro and in vivo. In any case, this need to confirm experimental findings is likely to represent the status quo within the pharmaceutical industry. If one is employing recombinant proteins alone or in parallel there will be a number of caveats needing to be considered, be they employed on array or in solution. Recombinant proteins studied structurally one at a time by NMR or X-ray crystallography each suffer similar caveats, ie these problems are not unique to array-based proteomics. Highly insoluble and/or membrane-associated proteins remain a major challenge at every turn within the protein sciences. However, Fang, Frutos and Lahiri have suggested a path forward through the use of lipid arrays. Cellular compartmentalisation can mean interactions due to improved accessibility of targets are never encountered within living cells and can thus give rise to false positives on arrays. Protein complexes are thought to be important in driving much of biology, yet these complexes cannot be easily synthesised and/or immobilised. A saving grace with respect to the latter is that one can...
expect differential assays dependent upon interaction partners (total or partial) to produce a higher signal during differential screening than molecules not involved in interactions, ie between molecules associated as a complex or between motifs found on individual members of a protein complex and on-array targets. Whenever interaction partners are immobilised there exist caveats with respect to in solution assay. These can, however, be minimised through the use of random immobilisation (as opposed to strategies which present only one side of a molecule for interaction assay) and the immobilisation of targets in a three-dimensional, highly hydrophobic hydrogel environment. These hydrogel substrates are thought to best emulate solution-like properties. Co- and post-translational modifications of proteins need to be addressed during synthesis of recombinant proteins. This is best achieved through the use of different expression vector hosts such as bacterial, yeast, insect and mammalian cells for each Open Reading Frame. Thereafter, the challenge for all recombinant techniques is to synthesise appropriately-folded and conformationally-correct recombinant proteins, ie to emulate the structural/binding integrity of the native protein. NB: Emulation of, for example, enzymatic functional integrity may not be so easily emulated for numerous on-array analytes, whereby each has specific physiological requirements with respect to optimal pH, substrate, cleavage and activation of precursors). Production procedures for recombinant proteins should be designed to minimise each of the above-mentioned caveats. In so doing, a powerful new parallel technology can be applied to lead optimisation. Previously, such a tool was simply not available to the pharmaceutical industry and thus information-gathering on a similar scale would have been painstakingly slow.

**Detection strategies**

Numerous detection strategies have evolved over the years to detect and amplify signals associated with the analysis of intermolecular binding events between macromolecules, small molecules and between these two molecular classes. These will not be reviewed here. Because of the almost ubiquitous nature of fluorescent detection technologies now seen as existing equipment infrastructure in well-equipped molecular laboratories, we have chosen to concentrate on labelling antibody and protein macromolecules with the same or similar dyes to those employed for differential analysis on cDNA microarrays, namely Cy3/Cy5 or Alexa 488/546. Labelled detection of small molecules is not practicable due to steric hindrance linked to moieties often as large or bigger than the drug being analysed. Nonetheless, such small molecule interactions become accessible through radio-labelling, which is currently a routine practice during lead validation and lead optimisation in the pharmaceutical industry. However, for the latter to become feasible for large-scale screening of small molecules, these approaches must first be linked to non-labelled parallel screening technologies. Table 1 provides a brief overview of non-labelled approaches for the detection of small molecule binding events. Some of these methodologies should be able to be modified for parallel detection when interfaced with chip-based readers. An added dilemma for small molecule detection is the need

<table>
<thead>
<tr>
<th>Table 1: Non-labelled detection strategies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surface plasmon resonance12,13</td>
</tr>
<tr>
<td>Grating coupled surface plasmon resonance14,16</td>
</tr>
<tr>
<td>Colorimetric resonant reflection17,18</td>
</tr>
<tr>
<td>Colorimetric gold nanoparticle sensors19</td>
</tr>
<tr>
<td>Reflectometric interference spectroscopy20</td>
</tr>
<tr>
<td>Quartz crystal microbalance21</td>
</tr>
<tr>
<td>Magnetic tweezers22</td>
</tr>
<tr>
<td>Optical tweezers23</td>
</tr>
<tr>
<td>Atomic force microscopy24-27</td>
</tr>
<tr>
<td>Nanocantilevers28-30</td>
</tr>
<tr>
<td>Mach-Zehnder interferometry21</td>
</tr>
<tr>
<td>Ellipsometry32</td>
</tr>
<tr>
<td>Resonant mirrors33</td>
</tr>
<tr>
<td>Fibre optic34</td>
</tr>
<tr>
<td>Surface acoustic waves35</td>
</tr>
<tr>
<td>Microcalorimetry36</td>
</tr>
<tr>
<td>Electrochemical detection37</td>
</tr>
<tr>
<td>Thermal lens microscopy38,39</td>
</tr>
<tr>
<td>Hartman interferometry40,41</td>
</tr>
<tr>
<td>Mass spectrometry42,43</td>
</tr>
</tbody>
</table>

Continued from page 24


Continued on page 27
for high surface occupancy of target combined with good signal-to-noise ratio so as to detect the very small D mass associated with the binding of a small molecule to a significantly larger biomolecule. Here, a substrate employing a three-dimensional matrix has advantages over mono-layer immobilisation strategies.

**Conclusion**

Protein and antibody arrays are likely to find immediate application in areas such as target discovery, validation of target discovered by the genomic sciences, precocious diagnosis of disease, patient cohorting with respect to disease and treatment outcomes and replacement of diagnostic assays not currently conducted in a parallel fashion, eg ELISAs in a clinical and research setting. More importantly, however, we believe that the greatest immediate advantage to the development of novel therapeutic agents likely to be derived from an increased knowledge of the human genome will be through the use of protein chips emulating increasingly large portions of the human proteome for applications directed towards improved target selectivity during lead optimisation. Drug registration authorities globally remain on the look-out for such improvements in target selectivity testing procedures, ie so as to help reduce the likelihood of adverse drug effects associated with novel therapeutic agents. Indeed, the use of protein arrays during lead optimisation has the potential of offering-up a reliable ‘early cull’ technology, more reliable than their cDNA counterparts, and most importantly, help insure against potentially deleterious interactions going undetected prior to clinical testing and market release.

**Erik Wischerhoff** was until recently the head of surface chemistry of Glaucoma Proteomics BV, Odijk, The Netherlands, a company involved in second generation proteomics based on protein and antibody microarrays. Before joining Glaucoma, he was responsible for the development of biosensor surfaces at BioTil AG in Munich, Germany (August 1997-May 2000). From September 1994 to May 1997 he worked as a postdoc at the Université Catholique de Louvain in Louvain-la-Neuve (Belgium), synthesising special property polymers for surface modification and using them to build up multilayer films by electrostatic adsorption. He received his PhD in 1994 at the University of Mainz (Germany) for his studies on structure-property relationships of functionalised polymeric liquid crystals. His main scientific interests are the synthesis of specialty polymers, the tailored modification of surfaces with polymeric compounds and the preparation of bioactive surfaces.

**Ryuji Hashimoto** is a research scientist of Daiichi Pharmaceutical Co Ltd, Tokyo, Japan. He received his PhD in 1998 at Kyushu University (Fukuoka, Japan) for his studies on structural and functional analyses of insulin-like growth factors and their binding proteins. He worked at the University of Utrecht as a research scientist from June 2001 to June 2002. During this period, he also contributed to high-throughput recombinant protein production for the protein micro-array of Glaucoma Proteomics BV and drug target validation in a micro-array format. His main task in Daiichi is high-throughput drug screening based on protein chemistry and he is interested in drug development, especially drug target validation, using proteomics technologies.

**Until recently, Ian Humphery-Smith was Managing Director and Chief Scientific Officer of Glaucoma Proteomics BV, a company aspiring to produce protein chips emulating increasingly large portions of the human proteome for applications directed towards improved target selectivity during lead optimisation. Drug registration authorities globally remain on the look-out for such improvements in target selectivity testing procedures, ie so as to help reduce the likelihood of adverse drug effects associated with novel therapeutic agents. Indeed, the use of protein arrays during lead optimisation has the potential of offering-up a reliable ‘early cull’ technology, more reliable than their cDNA counterparts, and most importantly, help insure against potentially deleterious interactions going undetected prior to clinical testing and market release.**

**DDW**

**Proteomics**

**Erik Wischerhoff**

**Ryuji Hashimoto**

**Continued from page 25**


34 Bender, WJH, Desoy, RE, Miller, ML, Chapman, MS, Dillow, SA. Recognition of a chemical microsensor based on surface plasmon resonance on fiber optics modified by multilayer vapor deposition. Anal. Chem. 66: 963-70.


