Published in final edited form as:

J Muscle Res Cell Motil. 2003 ; 24(4-6): 251–260.

Genomics, proteomics and bioinformatics of human heart failure

C.G. DOS REMEDIOS^{1,*}, C.C. LIEW², P.D. ALLEN³, R.L. WINSLOW⁴, J.E. VAN EYK⁵, and M.J. DUNN⁶

¹Muscle Research Unit, Institute for Biomedical Research, University of Sydney, Sydney 2006, Australia

²Cardiovascular Genome Unit, Brigham and Women's Hospital, Boston, MA 02115, USA

³Department of Anesthesia, Brigham and Women's Hospital, Boston, MA 02115, USA

⁴Center for Cardiovascular Bioinformatics and Modeling, The Johns Hopkins University School of Medicine and Whiting School of Engineering, Baltimore, MD 21218, USA

⁵Department of Physiology, Queen's University, Kingston, Ont., Canada, K7L 3N6

⁶Department of Neuroscience, Institute of Psychiatry, University of London King's College, De Crespigny Park, SE5 8AF London, United Kingdom

Abstract

Unraveling the molecular complexities of human heart failure, particularly end-stage failure, can be achieved by combining multiple investigative approaches. There are several parts to the problem. Each patient is the product of a complex set of genetic variations, different degrees of influence of diets and lifestyles, and usually heart transplantation patients are treated with multiple drugs. The genomic status of the myocardium of any one transplant patient can be analysed using gene arrays (cDNA- or oligonucleotide-based) each with its own strengths and weaknesses. The proteins expressed by these failing hearts (myocardial proteomics) were first investigated over a decade ago using two-dimensional polyacrylamide gel electrophoresis (2DGE) which promised to resolve several thousand proteins in a single sample of failing heart. However, while 2DGE is very successful for the abundant and moderately expressed proteins, it struggles to identify proteins expressed at low levels. Highly focused first dimension separations combined with recent advances in mass spectrometry now provide new hope for solving this difficulty. Protein arrays are a more recent form of proteomics that hold great promise but, like the above methods, they have their own drawbacks. Our approach to solving the problems inherent in the genomics and proteomics of heart failure is to provide experts in each analytical method with a sample from the same human failing heart. This requires a sufficiently large number of samples from a sufficiently large pool of heart transplant patients as well as a large pool of non-diseased, non-failing human hearts. We have collected more than 200 hearts from patients undergoing heart transplantations and a further 50 non-failing hearts. By combining our expertise we expect to reduce and possibly eliminate the inherent difficulties of each analytical approach. Finally, we recognise the need for bioinformatics to make sense of the large quantities of data that will flow from our laboratories. Thus, we plan to provide meaningful molecular descriptions of a number of different conditions that result in terminal heart failure.

Introduction

In this review we outline our plans for achieving a comprehensive analysis of the molecular defects present in end-stage failing human hearts. Our approach has been to select the best available techniques and practitioners in the fields of genomics, proteomics and bioinformatics

*To whom correspondence should be addressed: Tel.: +61-2-93513209; Fax: +61-2-93512813; E-mail: crisdos@anatomy.usyd.edu.au

and combine them into a coherent multifaceted investigation in which the defects (or potential defects) of any one technique are minimised by information from other, complementary techniques.

Information from the genomics of human heart will be derived from gene arrays based on either short oligonucleotide sequences or long cDNA nucleotides. Proteomics information has traditionally arisen from two-dimensional gel electrophoresis, a technique that has recently benefited from narrow range first dimension *pI* separations and mass spectrometry to identify the separated proteins. A more recent and in many ways more problematic proteomic approach is the separation, identification and rough quantification of proteins based on the protein equivalent to gene arrays. Commercial protein arrays have only now become available and none are yet specialised for examining myocardial proteins.

Finally, we recognize the need to integrate, codify and standardize the large amounts of information that are likely to come from this study. We describe our approaches in the expectation that International colleagues will be tempted to join with us and contribute to the project. Together we should make speedier progress towards understand the molecular nature of changes associated human heart failure.

Gene arrays

Gene array technology for analysis of changes in gene expression in failing human hearts is now well advanced. Initial sequencing of the human genome suggested it contains an estimated 32,000 genes (Lander *et al.*, 2001; Venter *et al.*, 2001) and more recently this figure has been revised downwards (Collins *et al.*, 2003).

In principle, since the sequence of each gene is known, it can be precisely probed using gene arrays that contain either robotically-arrayed full length (or nearly full length) cDNA cloned probes, or using shorter (approximately 25 base pair) oligonucleotides that are synthesized on glass. The assumption that in practice each probe accurately quantifies the expression of the corresponding gene may be overly simplistic. For eukaryotes, the correlation between the quantity of mRNA and the quantity of the corresponding protein can be poor (Le Naour *et al.*, 2001). Despite these concerns, gene arrays remain the most extensive source of reliable information on the molecular status of an organ like the human heart.

Gene arrays start with intact RNA

In order to make proper use of any microarray technology, it is essential to have a sufficient quantity of intact RNA, isolated from human tissue samples representing a 'normal' or 'disease' state. Typically, the RNA is converted to cDNA and fluorescently labelled. This cDNA 'probe' is then applied to the microarray and allowed to hybridize overnight. At least 10 µg of RNA per array is required in order to generate a reliable signal on the CardioChip. 10–100 ng of RNA is sufficient for the AffyChip, if a second amplification step is used, for example with biopsy samples.

Oligonucleotide and cDNA arrays

Two DNA microarray technologies are currently in use for the study of human disease. The first is the Hu133 'AffyChip', an oligonucleotide array manufactured by Affymetrix that contains 22,000 transcripts representing the major transcripts of the human genome. The second, technology is the spotted cDNA arrays. These are both commercially available or can be manufactured in-house (Barrans *et al.*, 2002; Hwang *et al.*, 2002). One example of the latter is the 12,000-element 'CardioChip' (Barrans *et al.*, 2001). The CardioChip consists of unique cDNA clones of heart libraries that have been partially sequenced (i.e. ESTs) and annotated. This particular array is customized for the study of heart failure. It contains known

cardiovascular-related genes as well as a broad selection of novel genes and ESTs derived from our laboratory's heart cDNA library. The layout consists of 2761 known genes (25.5% of the array), 3406 matched ESTs (31.4%), 4489 novel ESTs (41.4%) and 192 bacterial controls (1.8%). A representative CardioChip (Figure 1) compares Cy3 (red) labelled non-diseased and Cy5 (green) labelled failing heart samples.

Gene clustering

The major advantage of DNA microarray technology lies in its ability to profile and compare thousands of genes simultaneously between several mRNA populations. However, the utility of the DNA microarray extends beyond the concept that it is a novel tool for large-scale transcript profiling and identifying differences in expression between single genes.

It is the wealth of data generated from a series of transcript profiling experiments that makes this technique so powerful. Groups of data can be analyzed to determine the level of 'relatedness' between genes or samples in multiple dimensions. Basically, using a set of expression fingerprints or profiles, similarities and differences in gene expression are capitalized on in order to group or cluster different mRNA populations or genes into discrete, related sets or bins. This is extremely powerful as the clusters of co-regulated genes often belong to the same biological pathway, or even to the same protein complex. Thus the clusters of mRNA populations defined by their 'expression fingerprint' provide a means of defining differences between samples that would not otherwise be possible.

Gene clustering between mRNA populations can be viewed from two perspectives. Clustering over a time-course can identify genes with a similar function because the expression of functionally related genes tends to be regulated, and thus expressed, in a similar manner (e.g. genes involved in G1 of the yeast mitotic cell-cycle). This is an important concept, as even genes with no known function can now be placed into a biological pathway.

Alternatively, clustering genes between several mRNA populations from a single time point can identify a vast number of associations between the expression of a cluster of genes and a biological phenotype such as marker or sentinel genes. Thus, through clustering the complete sets of gene expression data generated from samples, several gene clusters will be identified and examined to determine the specific relationship with each particular mRNA population. Furthermore, these results can be extended to define different etiologies of cardiomyopathies.

A focus on different causes of human heart failure

Over the past 4 years, we have examined a limited number of explanted heart samples representing hypertrophic and dilated cardiomyopathies (Barrans *et al.*, 2002; Hwang *et al.*, 2002) and suggested that 200–300 genes may be involved in the processes of developing heart failure. Through cluster analysis of gene profiles, donor (n = 6), hypertrophic (n = 3) and dilated (n = 6) samples can be clustered distinctively.

Over the past decade or so, we have accumulated a collection of over 250 explanted hearts from patients undergoing heart transplantations. Based on our available samples we have selected: (1) non-familial idiopathic dilated cardiomyopathy; (2) doxorubicin-poisoned myocardium; (3) peripartum cardiomyopathy; (4) hypertrophic cardiomyopathy; (5) failure due to known genetic causes (including Eisenmenger, Ebstein, and Holt-Oram syndromes, Beckers muscular dystrophy) and (6) ischemic heart disease particularly the diffuse type. Multiple samples have been harvested from left and right ventricles and atria as well as other tissues such as coronary arteries and aorta.

This large collection will help us to define the biological pathways and clusters of genes related to the etiology of heart failure. For example, we can now select multiple examples of idiopathic

dilated cardiomyopathy that are essentially free of coronary vascular disease. Importantly, we now have nearly 60 examples of non-diseased, donor hearts that can act as ‘controls’ for the failing hearts.

AffyChip or CardioChip: which is better?

We are aware that a major question currently concerns which of the two microarray technologies is superior? Actually, we feel that *neither* is superior and that both can have ideal or equivalent situations for use. Thus one or the other should be selected based on an appropriate context. Of course, there are inherent advantages and disadvantages to both technologies. These are summarized as follows:

We feel that *both* technologies should be used concertedly. They can complement each other and allow us to take the best approach to identify the genes/pathways involved in investigations into the molecular basis of heart failure (Table 1).

Two-dimensional gel electrophoresis

Proteomics is the study of the protein complement present in a cell at any given moment in time. In part they reflect the genome, but they also mirror changes due to further cellular processes resulting in co- and post-translational modification (PTM) events, including for example acetylation, phosphorylation and oxidation. Figure 2 is an example of a silver-stained 2-D PAGE gel where the proteins from a non-failing, ‘normal’ samples of human ventricle are separated in the first dimension on a 3–10 pH gradient and in the second dimension on a 10% acrylamide gel.

The excitement surrounding proteomics focuses on the potential to elucidate the molecular mechanisms that control normal cell and organ functions as well as, the potential to identify proteins responsible for initiation and progression of disease. In this respect, the proteins that change as well as those that do not are key to our understanding of the subtle and complex manner in which the cells adapt to pathophysiological stimuli resulting in disease.

This is especially true with heart failure, a complex multi-factorial disease/disorder ultimately diagnosed on the basis of the inability of the heart to maintain sufficient cardiac output to meet the metabolic needs of the body.

Potentially, there are many sub-proteomes within cardiomyocytes capable of causing reduced cardiac output. Proteomic analysis can certainly help in sorting out which families of proteins are affected. The outcome of a proteomic experiment is a long list of proteins can change (or not change) due to the heart failure process. The ultimate goal is to differentiate the protein changes that are causative versus those that occur as a result/effect of the disease. To discern ‘cause vs. effect’ requires foresight in experimental design and quantitative analyses of the proteomic data.

Proteomic experimental design

There are several experimental strategies that can assist in addressing this issue. The first is based on the assumption that protein(s) altered early after the initiation or trigger of heart failure may subsequently cause downstream protein changes. This connectivity is key to understanding the far-reaching cellular dynamics and interconnections.

To obtain this information it is prudent to undertake temporal protein profiling. In other words, to carry out in-depth proteomic analysis on a series of tissue samples, preferably from the same patient or animal model, obtained over the time course of the disease progression.

It would be even more powerful to match clinical and physiological measurements (such as cardiac output or ejection fraction), and pharmacological records obtained over the same time course. For example, using the reversible, fast, pacing-induced canine model of heart failure, proteomic analysis was carried out on dogs in end-stage heart failure. Cardiac output was tracked in these animals by echocardiography giving rise to physiologically relevant time points for the collection of proteomic data. In this study, we (Heinke *et al.*, 1998, 1999) showed that a relatively small number of proteins from a 2-D gel database of around 2000 protein spots changed with end-stage heart failure.

Of the 70 proteins that showed quantitative changes in the failing heart, 42 were present at reduced abundance, while 28 were increased. After identification, these proteins were assigned to three broad functional classes: (1) proteins associated with mitochondria and energy metabolism; (2) cytoskeletal and myofibrillar proteins; and (3) proteins associated with cellular stress responses (Corbett *et al.*, 1998).

An interesting aspect of this model is that after pacing was stopped, the hearts reverted to 'normal' function. An investigation into whether there is concomitant 'normalization' of the alterations to the cardiac proteome will of great interest.

Temporal profiling

Temporal profiling of the human myocardium proteome during the development of heart failure is not possible for many reasons. Thus, an alternative strategy is required that will provide equal insight into 'cause vs. effect' of various changes within the proteome of the cardiac myocyte.

The second strategy is to undertake proteomic analysis of the same disease (or a particular phenotype) produced by different causative factors. For example, heart failure can be caused by ischemia, viral infection, genetic disorders and structural problems with the heart (valve insufficiency), each ultimately producing the same disease phenotype.

The same approach can be used where multiple animal models of the same disease are analysed. Regardless of whether experiments are conducted using human or animals, this second approach is based on the assumption that a disease caused by different triggers will have overlapping (common) protein changes. These represent either convergent points for the various mechanism or end-effectors of the disease phenotype.

It is these commonalities in mechanisms that may reveal key pathways or groups of proteins/targets for therapeutic intervention. However, one must keep in mind that in some cases the same end phenotype or function may arise by different groups of proteins. In the heart, contractile dysfunction can arise either through a change of Ca-handling proteins and/or the contractile/myofilament proteins.

An example is the comparison of two different animal models of myocardial stunning. Myocardial stunning is defined as reversible cardiac dysfunction due to brief ischemic episode. In the rat isolated heart model, the myofilament protein, troponin I, is selectively proteolysed while in the *in vivo* swine model, the Ca-handling protein, phospholamban, undergoes target phosphorylation (for a review see Murphy and Van Eyk, 2001). Both defects are able to produce the same overt contractile dysfunction, the hallmark of stunning. Therefore, although the phenotype is the same, a different sub-proteome is being altered. The lesson to be learned from this example is that unless one understands the physiological consequences of the protein changes, the link between disease pathways may not become clear.

Proteomic analysis of human end-stage heart failure has been carried out (Corbett *et al.*, 1998), in which samples of failing hearts resulting from both dilated cardiomyopathy and ischemic heart disease were compared with a small number of samples from 'normal' unused transplant donor hearts.

In spite of the inherent heterogeneity of the 2-D gel protein profiles obtained from this human population, we were able to show significant changes in some 93 proteins (80 decreased, 13 increased in abundance) from a total of a 2-D gel database of around 1600 protein spots. Interestingly, identification of these altered proteins showed them to be in the same three functional classes as described in our proteomic studies of heart failure in the paced dog model (Heinke *et al.*, 1998, 1999).

A major limitation to these studies is the lack of large numbers of clinically well-defined failing hearts. The greatly enlarged collection of human heart samples now available to us provides the power to weed out the background noise from the common disease-induced changes. What would be truly insightful is to have several hundreds heart samples to analyse using proteomic techniques. We are nearly there.

Proteomic technologies – 2DGE

In order to understand the underlying molecular mechanisms of heart disease, it is critical that the most extensive observation of the proteome is achieved and that proteins are identified and PTMs characterized. Broad-based proteomics involves the systematic characterization of myocardial proteins most often by two-dimensional gel electrophoresis (2DGE) followed by mass spectrometry, although alternative and complementary technologies exist. We and others have found that this type of broad-based screening represents a major experimental advance and is ideally suited to documenting the subtle, complex and dynamic changes to the proteome.

The cornerstone of proteomic analysis is 2DGE. This technique separates intact proteins in the first dimension (isoelectric focusing, IEF) based on their intrinsic *pI*s, and in the second dimension (SDS-PAGE), according to their molecular masses. With 2DGE, we are limited by the solubility of the proteome and therefore, we need to optimise the homogenization and IEF steps using various combinations of detergents (i.e. Stanley *et al.*, in press) to maximize the quantity and types of proteins loaded onto the gel.

In some cases, optimization is required for a single key protein of interest (Labugger *et al.*, 2002). Furthermore, protein spot resolution and increased coverage of the proteome can be achieved using a variety of overlapping narrow pH gradient IEF gels for the first dimension (i.e. Westbrook *et al.*, 2001) as well as by varying the extent of crosslinking used for the SDS PAGE dimension. Even so, with myocardial tissue we still have the problem of the highly abundant myofilament proteins dominating the myocyte proteome. Therefore, it can be helpful to use of modified extraction methods to remove the myofilament sub-proteome and enrich for the cytoplasmic and membrane proteins (Arrell *et al.*, 2001).

Alternative approaches involve isolation of a particular organelle (i.e. mitochondria – Taylor *et al.*, 2003) or of a binding partner complex i.e. protein kinase C (Edmonson *et al.*, 2002). In any case, complex 2DGE gel image analysis is carried out to compare multiple gels and indicate protein changes (reviewed in Dowsey *et al.*, in press). Proteins are identified by in-gel digestion and mass spectrometry analysis of the resulting peptide fragments.

With peptide mass fingerprinting, the observed mass spectrum of the peptides are compared in silico to the known protein databases. In the case where the species or isoform of the protein is not present, identification can only be made on the basis of amino acid sequence homology. Amino acid sequencing maybe carried out with MS/MS to validate protein identification.

One important advantage to 2DGE analysis is the ability to detect PTMs that result in a change in pI or M_r . However, it is still extremely challenging to characterize PTMs. Examples are phosphorylation of the elongation factor-2 (Jager *et al.*, 2002) or myosin light chain 1. The latter is a protein previously thought to be unphosphorylatable (Arrell *et al.*, 2001). Another example is the apparently aberrant ubiquitination of several cardiac proteins (Weekes *et al.*, 2003).

A high level of skill and expertise is required to generate identical or nearly identical 2DGE gels. The technique is expensive, time consuming and limited to soluble proteins. These inherent limitations have driven the development of alternative or complementary proteomic methods. Examples of such techniques are liquid chromatography (one-dimensional or two-dimensional (2D/LC)) of proteins (Neverova and Van Eyk, 2002) or peptides obtained from the global enzymatic digestion of the sample of interest (Shotgun and MudPit (review: McDonald and Yates JR 3rd, 2002)) or based on quantitative MS using methods such as ICAT (isotope coded affinity tag reagents) (see review: Turecek, 2002). Another possibility is protein arrays (described below).

As with 2DGE, all of these methods have technical limitations, but if used collectively they can expand the observable proteome. Thus, in the future it will become essential to use more than one method to analyse a particular proteome. Then limitations will be on sample size, technical expertise and funding.

Protein arrays

The conversion of gene array technology into protein arrays that can identify and quantify proteins is fraught with problems. There are distinct advantages for those that succeed but a series of technical challenges will have to be overcome before 'off-the-shelf' protein arrays become a reality. Protein arrays, particularly antibody arrays, can overcome the sensitivity problems of 2DGE proteomics and simultaneously offer the prospect of massive parallel processing of large numbers of proteins (Lal *et al.*, 2002). However, unlike the PCR reaction for nucleic acids, proteins cannot be amplified.

Advantages of protein arrays

As a strategy, it is better to analyse expressed proteins rather than the mRNAs that encode them. The advantages include the ability to detect PTMs and the certainty that a protein has been expressed in response to an altered gene expression. Several years ago we reported (Bennetts *et al.*, 1986) there was little or no correspondence between the amounts of mRNAs (determined from Northern blots) of α -cardiac and α -skeletal muscle actin isoforms and the total quantity of translated actin determined from one-dimensional PAGE gels. The same is true for cardiac myosin heavy chain expression, namely changes observed in myosin heavy chain mRNA did not correspond to changes at the protein level (Coumans *et al.*, 1997). Although this is clearly not true for all myocardial proteins, the correspondence of changes in gene arrays and proteomics in the analyses of eukaryotic cells can be quite poor (La Naour *et al.*, 2001).

The main challenge for protein array technology is the identification and quantification of proteins selectively retained on the arrays. This largely depends on ligand affinity.

Binding affinity and specificity

At present, specificity is mainly achieved using monoclonal antibodies. These, rather than polyclonals, are used because they can be reliably produced and have definable (although usually not precisely defined) epitopes. Their binding affinities need to be high, preferably in the picomolar range, but most monoclonal antibodies have lower affinities. The exceptional

affinity of biotin for streptavidin has made these binding partners a popular choice for downstream detection.

Binding specificity also must be considered. Monoclonal antibodies are not necessarily the obvious choice for a capture device. It is true that monoclonal antibodies have highly specific epitopes but these epitopes are not necessarily unique. In fact, cross reactions with unexpected proteins can and do occur. Some years ago, we defined the epitope of a monoclonal antibody to myosin light chains. It turned out to be an unusual PTM, trimethyl alanine (Boey et al., 1992). No other contractile protein has such a modification and we believed it to be unique. It was not. Another group also had reported an antibody directed against a nuclear protein which was almost certainly had the same epitope. Others have gone so far as to say that *most* monoclonal antibodies have some, limited degree of cross-reactivity.

An important issue is the preservation of activity of the capturing antibody. The chemical nature of the solid support is critical for the preservation of protein structure (Mitchell, 2002). More than 50% of antibodies are thought to denature on solid planar surfaces such as glass, and considerable effort has been spent on ameliorating the forces that distort protein structure.

Finding and arraying the right antibody

Thousands of monoclonal antibodies are available for use in protein arrays and the number increases daily. Search engines are quite good at finding commercial sources (e.g. www.abcam.com) but research laboratories generate many more antibodies. Even when an antibody is directed against a protein, more than one clone may be available, each with a different affinity and epitope.

Once the right antibody has been identified, its use in an array is normally very economical. 50–500 μ l can be delivered to an array surface, so it follows that the purchase of a single 0.5 ml sample of a monoclonal antibody will normally be sufficient to dispense onto very large numbers of arrays.

A reasonable estimate of the number of proteins expressed in a cardiomyocytes is about 10^4 , and gene arrays tell us that about 10% of these differ in non-diseased and heart failure (Barrans *et al.*, 2002). It is wise to employ more than one clone for those proteins whose expression is critically elevated (or depressed) and of course isotype and negative controls should be present in the array. At an average purchase price of about \$200 per antibody, clearly the development of a comprehensive protein array for myocardial proteins associated with human heart failure can be costly. The cost of each antibody in an array will be small (<1 cent per antibody) so the cost per array will be dominated more by the cost of the solid supports (e.g. nitrocellulose-coated glass FAST slides from Schleicher & Schuell Biosciences), which currently are expensive. Special, modified glass slides (e.g. from OptArray) are about twice as expensive.

Surface denaturation of antibodies can be overcome by either capturing them by their Fc portions, or by immobilizing Fab fragments by their free cysteinyls that face away from the antigen-combining surface. Accelr8 Technology (www.accelr8.com) and Prolinx (www.prolinx.com) have developed novel three-dimensional coatings for solid supports that may reduce denaturation while increasing ligand densities. They also increase the cost.

Commercially available antibody arrays (BD Biosciences, Amersham) are arrayed on modified glass and at present contain only limited numbers (about 250–500) of antibodies. Like the early gene arrays, they will rapidly progress into more extensive devices while maintaining a unit cost of a few hundred dollars. Several companies now offer custom antibody arraying services but the range of antibodies from these companies is currently limited and tends to be specialised (e.g. for chemokines and cytokine receptors).

Initially, our choice of antibodies for a human cardiac protein array will be based on those suggested by the gene arrays (200–300 genes) and those identified by 2DGE (about 70 proteins). This number will be manageable in the start-up phase and can be expanded later.

Blocking indiscriminate binding

When a solid support is used to immobilize an antibody, the unoccupied surfaces generally must be blocked to prevent attachment of non-specific proteins. Usually a universal protein such as powdered skim milk can be used as effectively as more costly proteins like bovine serum albumin. No blocking protein will be ideal but skim milk is particularly problematic where biotin–streptavidin complexes are used in downstream detection because of the presence of biotin.

A better solution is to avoid the blocking process all together. One approach is to control the hydrophobicity–hydrophilicity of the capturing surface. Hydrophobic surfaces bind most proteins better than hydrophilic surfaces but separation on this basis alone is unlikely to avoid the need to block. Two companies mentioned above have developed surfaces that do not seem to require the blocking step. Accelr8 Technology uses a cross-linked matrix combined with the selective capture of ligands via streptavidin-linked surfaces (to capture of biotinylated proteins), or via thiol- or amine-reactive surfaces to covalently attach proteins. ProLinx Inc. uses polymer ‘brushes’ containing phenyl(di)boronic acid that cross-reacts with proteins pre-modified with salicylhydroxamic acid.

Detection of bound proteins

Antibodies can be used to detect ligands captured by primary antibodies in what amounts to a solid state ELISA assay. There are at least two problems with this approach. Firstly, the detecting antibody must not have the same epitope as the capture antibody. Also, the antibody must have a sufficiently high affinity to detect low abundance components in the array. This can be achieved by labelling the second antibody with a fluorescent probe (these are often commercially available) but this approach can run into sensitivity problems. Clearly the assay is limited by the weakest binding affinity.

Another approach is to biotinylate the second antibody and then detect the biotin moiety with a streptavidin conjugated to a visible probe like HRP or a fluorescent label. Streptavidin is now available linked to magnetic microbeads (www.miltenyibiotec.com) or to streptavidin-conjugated Q-Dots™ (www.qdots.com). With these, levels of detection can be substantially lowered since single Qdots and microbeads can be unambiguously identified and quantified. None of these solutions compare with detection by mass spectrometry (see below).

Advantages of protein arrays over 2DGE gels

A major driver of protein array technology has been the short-comings of two-dimensional gel electrophoresis. We have already reviewed the substantial contribution of 2DGE gels to our understanding of the protein changes associated with heart failure in both man and animal models of human heart failure. The principal bottleneck with 2DGE is the identification of the proteins where more than one protein may be present per spot. MALDI mass spectrometry greatly improved this problem, and narrow range first dimension *pI* range strips also helped.

New generation seven tesla Fourier transform ion cyclotron resonance mass spectrometers are now available that can identify protein isoforms that differ by only a few Daltons (Lee *et al.*, 2002). These instruments may be expensive but they bring immense power to the problem of identifying protein differences between failing and non-failing hearts. They may even overcome a fundamental failing in the analysis of the proteome, namely absolute rather than relative quantitative evaluation of individual proteins.

Bioinformatics

Making sense of complexity

The goal of proteomics is to provide a mechanistic understanding of the exact proteome and of functional changes associated with heart failure, regardless of the triggering etiology. This understanding, when combined with insights obtained from genomic data, will help to identify overlapping and unique pathways that ultimately render the heart incapable of maintaining cardiac output. Although the exact relation between mRNA and protein quantity is not clear due to the effect of protein processing and turnover, these two processes are ultimately linked.

A collective understanding of the proteins involved in the genesis of heart failure will enable the design of new treatments and clinical interventions. To achieve this goal, additional information about protein localization, function and binding partners must be obtained. Proteomic experiments can reveal changes in protein expression (increase, decrease, de novo and isoform changes) and PTMs that may participate in a particular disease or process. This information will certainly lead to new hypothesis that can be tested experimentally.

The continuous cycling of discovery/proteomic science and hypothesis-driven experiments will eventually lead to a more complete understanding of heart disease. To handle the complex information that must be collected requires the application of bioinformatic techniques – the objective being to make sense of the accumulated data from the proteomic studies, the literature and other experiments carried out on a specific set of samples (review: Winslow and Boguski, in press).

The rapid pace of development of proteomics technologies and the resulting diversity and complexity of proteomics data poses special challenges. In particular, methods for structuring and searching proteomics databases to retrieve groups of proteins based upon well-known pathways, functional classifications and specific PTMs must be developed. Methods for annotating and differentiating PTMs predicted from protein motifs using computational algorithms versus those for which there is direct experimental evidence are required.

Protein concentrations and other measured attributes should be compared with values determined in reference samples to enhance data quantification. With regard to protein identification based on MS, annotations must provide meaningful statistical measures of the quality of match.

Data representation and dissemination must be facilitated by the adoption of standards for data description. As an example of such an effort, the Human Proteome Organization is promoting the development of standard formats for the representation and exchange of MS and protein–protein interaction data and annotations. These formats are derivatives of extensible markup language (XML), a language that originated as a standard for document formatting, but which is now used as a format to transfer structured data of any kind over the World Wide Web.

Finally, Web services, a technology building on the ability of simple object access protocol (SOAP) to support distributed network communication, have great potential as a tool for making both data and computational algorithms transparently available to other software applications. This will facilitate the machine discovery, communication, and analyses of proteomic as well as genomic data.

In addition, cell behaviour is regulated in a complex manner through a diversity of interacting gene expression, signal transduction, metabolic, and electrophysiological pathways. Pathway properties are themselves determined by factors such as the specific nature of molecular interactions, formation of multi-molecular complexes, and by subcellular localization.

Representation of information on biological pathways in a form that supports complex querying and modelling is an important goal of post-genomics biology.

Finally, there is currently no single software platform with which data obtained using different proteomic methods (i.e. MS data, ICAT, 2DGE and 2D/LC) can be analysed simultaneously with physiological (i.e. echocardiographs) and genomic data. Clearly there is a need for software tools and database federation systems, which can track and analyse data samples collected using multiple means using statistical methods.

Today, we only have data obtained from different heart failure cohorts done under different circumstances and using different tools. As mentioned earlier, the inherent difficulty in dealing with such disparate data is those experimental limitations and heterogeneity in human samples due to 'haziness' around clinical diagnosis and the underlying affect of drugs and treatments.

In the best scenario, myocardial samples from the same large cohort of heart failure patients will be analysed by multiple broad based approaches. Eventually, given sufficient molecular information from a variety of proteomic techniques as well as genomics, will reveal why particular patient subpopulations have a better prognosis than others, or respond differently to therapeutic treatments. In the future, it will be possible to individualize diagnosis and treatment depending on in-depth knowledge of disease processes that produce given disease phenotype.

Conclusion

Never before have scientists conceptually or technically been able to perform large-scale protein gene analyses. Even given the limitations with current proteomic techniques there is still the ability to see deeply into the genome and proteome. There is a drive to develop more compressive algorithms that either stand-alone or can be used to enhance existing techniques. With these new developments, and with the goodwill that comes from collaboration, the future holds much promise.

References

- Arrell DK, Neverova I, Van Eyk JE. Cardiovascular proteomics: evolution and Potential. *Circ Res* 2001;88:763–773. [PubMed: 11325867]
- Barrans JD, Allen PD, Stamatou D, Dzau VJ, Liew CC. Construction of a human cardiovascular cDNA microarray: portrait of the failing heart. *Biochem Biophys Res Commun* 2001;280:964–969. [PubMed: 11162619]
- Barrans JD, Allen PD, Stamatou D, Dzau VJ, Liew CC. Global gene expression profiling of end-stage dilated cardiomyopathy using a human cardiovascular-based cDNA microarray. *Am J Pathol* 2002;160:2035–2043. [PubMed: 12057908]
- Bennetts BH, Burnett L, dos Remedios CG. Differential co-expression of a-actin genes within the human heart. *J Mol Cell Cardiol* 1986;18:993–996. [PubMed: 3783733]
- Boey W, Everett AW, Sleep J, Kendrick-Jones J, dos Remedios CG. Uncoupling of actin activated myosin ATPase activity from actin binding by a monoclonal antibody directed against the N-terminus of myosin light chain 1. *Biochemistry* 1992;31:4090–4095. [PubMed: 1533157]
- Collins FS, Green ED, Guttmacher AE, Guyer MS. A vision for the future of genomics research. *Nature* 2003;422:835–847. [PubMed: 12695777]
- Corbett JM, Why HJ, Wheeler CH, Richardson PJ, Archard LC, Yacoub MH, Dunn MJ. Cardiac protein abnormalities in dilated cardiomyopathy detected by two-dimensional polyacrylamide gel electrophoresis. *Electrophoresis* 1998;19:2031–2042. [PubMed: 9740065]
- Coumans JVF, Yeoh T, Seeto RK, Keogh A, Brennan K, Gunning P, Hardeman E, dos Remedios CG. Actin-binding proteins in mouse C2 myoblasts and myotubes: a combination of affinity chromatography and two-dimensional gel electrophoresis. *J Mol Cell Cardiol* 1997;29:895–905. [PubMed: 9152850]

- Dowsey AW, Dunn MJ, Yang GZ. The role of bioinformatics in two-dimensional gel electrophoresis. *Proteomics* 3in press
- Edmondson RD, Vondriska TM, Biederman KJ, Zhang J, Jones RC, Zheng Y, Allen DL, Xiu JX, Cardwell EM, Pisano MR, Ping P. Protein kinase C epsilon signaling complexes include metabolism- and transcription/translation-related proteins: complimentary separation techniques with LC/MS/MS. *Mol Cell Proteomics* 2002;1:421–433. [PubMed: 12169683]
- Heinke MY, Wheeler CH, Chang D, Einstein R, Drake-Holland A, Dunn MJ, dos Remedios CG. Protein changes observed in pacing-induced heart failure using two-dimensional electrophoresis. *Electrophoresis* 1998;19:2021–2030. [PubMed: 9740064]
- Heinke MY, Wheeler CH, Yan JX, Amin V, Chang D, Einstein R, Dunn MJ, dos Remedios CG. Altered myocardial protein expression in pacing-induced canine heart failure: an update. *Electrophoresis* 1999;20:2086–2093. [PubMed: 10451120]
- Hwang JJ, Allen PD, Tseng GC, Lam CW, Fananapazir L, Dzau VJ, Liew CC. Microarray gene expression profiles in dilated and hypertrophic cardiomyopathic end-stage heart failure. *Physiol Genomics* 2002;10:31–44. [PubMed: 12118103]
- Jager D, Holtz J, Redpath NT, Muller SP, Ponicke K, Heinroth-Hoffmann I, Werdan K, Muller-Werdan U. The aging heart: influence of cellular and tissue aging on total content and distribution of the variants of elongation factor-2. *Mech Ageing Dev* 2002;123:1305–1319. [PubMed: 12297334]
- Labugger R, McDonough JL, Neverova I, Van Eyk JE. Proteomic analysis of cardiac myofilament protein: 2DE focusing and detection of troponin T. *Proteomics* 2002;2:673–678. [PubMed: 12112847]
- Lal S, Christopherson RI, dos Remedios CG. Antibody arrays: an embryonic but rapidly growing technology. *Drug Discov Today* 2002;7:S143–S149. [PubMed: 12546881]
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, et al. Initial sequencing and analysis of the human genome. *Nature* 2001;409:860–921. [PubMed: 11237011]
- Lee S-W, Berger SJ, Martinovic S, Pasa-Tolie L, Anderson GA, Shen Y, Zhao R, Smith RD. *Proc Natl Acad Sci* 2002;99:5942–5947. [PubMed: 11983894]
- Le Naour F, Hohenkirk L, Grolleau A, Misek DE, Lescure P, Geiger JD, Hanash S, Beretta L. Profiling changes in gene expression during differentiation and maturation of monocyte-derived dendritic cells using both oligonucleotide microarrays and proteomics. *J Biol Chem* 2001;276:17920–17931. [PubMed: 11279020]
- McDonald WH, Yates JR 3rd. Shotgun proteomics and biomarker discovery. *Dis Markers* 2002;18:99–105. [PubMed: 12364816]
- Mitchell P. A perspective on protein microarrays. *Nat Biotechnol* 2002;20:225–229. [PubMed: 11875416]
- Murphy, AJ.; Van Eyk, JE. The role of troponin abnormalities as a definitive cause for stunned myocardium. Review in *Depth Series on Stunned and Hibernating Myocardium*. In: Kloner, Robert, editor. *Coron Artery Dis.* 12. 2001. p. 343-347.
- Neverova I, Van Eyk JE. Application of reversed phase high performance liquid chromatography for subproteomic analysis of cardiac muscle. *Proteomics* 2002;2:22–31. [PubMed: 11788988]
- Stanley B, Neverova I, Van Eyk JE. Effect of detergents on two-dimensional gel electrophoresis of human cardiac tissue. *Proteomics*. in press
- Taylor SW, Fahy E, Zhang B, Glenn GM, Warnock DE, Wiley S, Murphy AN, Gaucher SP, Capaldi RA, Gibson BW, Ghosh SS. Characterization of the human heart mitochondrial proteome. *Nat Biotechnol* 2003;21:281–286. [PubMed: 12592411]
- Trahair T, Yeoh T, Cartmill T, Keogh A, Spratt P, Chang V, dos Remedios CG, Gunning P. Myosin light chain gene expression associated with disease states of the human heart. *J Mol Cell Cardiol* 1993;26:577–585. [PubMed: 8377217]
- Turecek F. Mass spectrometry in coupling with affinity capture-release and isotope-coded affinity tags for quantitative protein analysis. *J Mass Spectrom* 2002;37:1–14. [PubMed: 11813306]
- Venter C, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, et al. The sequence of the human genome. *Science* 2001;291:1304–1351. [PubMed: 11181995]

- Weekes J, Watson RR, Dunn MJ. Murine retrovirus infection and the effect of chronic alcohol consumption: proteomic analysis of cardiac protein expression. *Alcohol Alcoholism* 2003;38:103–108.
- Westbrook JA, Yan JX, Wait R, Weslon SY, Dunn MJ. Zooming-in on the proteome: very narrow-range immobilised pH gradients reveal more protein species and isoforms. *Electrophoresis* 2001;22:2865–2871. [PubMed: 11565781]
- Winslow R, Boguski MS. Genomic informatics: current status and future prospects. *Circ Res* 92in press

THE "CARDIOCHIP":

Fig. 1.
cy3: Normal adult heart, cy5: failing heart.

Fig. 2. Silver stained 2-D gel of whole tissue extracts (7M urea, 2M thiourea, 4% ASB-14) from normal human left ventricle. Approximately 450 μ g of whole tissue extract was separated by pH 3–10 linear gradient IPG strips followed by 10% acrylamide SDS-PAGE (Stanley and Van Eyk, unpublished).

Table 1
A comparison of CardioChip and AffyChip technology

Type	cDNA-based CardioChip	Oligonucleotide AffyChip
Custom-made arrays	Yes	No
Cost per chip	Cheaper	Expensive
Identification of cardiac genes	52k ESTs customized for cardiovascular system	Generalized transcripts. So far no custom made arrays
Cost of development of the array	About \$35,000 for whole human genome	Nil
Ability to work on unknown sequences	Yes	No – oligonucleotide sequence must be known (not provided)
Ability to discover new genes	Yes	No
Density of genes per slide	12,000-elements/chip: 2761 known genes, 3406 matched ESTs and 4489 novel ESTs, 192 bacterial controls/slide	Higher density – 22,000 transcripts. One gene, 20 probe pairs, 25 nucleotides/pr = 500 nucleotides per gene
Robustness of controls	Cross-hybridization not as robust as AffyChip Serial	More robust, gene-specific cross-hybridization subtraction
Manufacturing process	PCR and spotting increase slide-to-slide variability	Parallel synthesis of the array with little slide-to-slide variability
Hybridization	Competitive, 2 dye system	Non-competitive, 1 dye system
Dynamic range, linearity	Lower	100-fold higher
Precision at high/low ends	Higher imprecision	Lower imprecision
Ease of use	Difficult for beginners	Easy for beginners
Detection method	Flexible open system	Convenient integrated system
Troubleshooting	Open system	Closed system