Development of an on-chip in vitro Proteome Analysis System

Khaled Mahmoud

Faculty of Sciences, Ain Shams University. Current address; Faculty of Engineering, University of Leeds, United Kingdom.

Abstract: I created a protein in vitro expression and characterization system on a solid support platform. The system has been produced in a scalable format that suits high throughput applications as well as on-chip diagnosis purposes. This system is an innovative combination of a number of complementary technologies for analysis and interrogation of protein-encoding sequences populations. It enables efficient identification and characterization of the protein components in complex protein-encoding populations. The system uses a unique nucleic acid tagging mechanism, in combination with solid phase cloning, antibody tagging and mass spectrometric analysis to identify protein species in a complex environment. This novel technology generates mRNA species (the “Transcriptome”) from the target biological source and uses a system of tagging to label individual mRNA species uniquely. This allows individual transcripts, corresponding to encoded proteins, to be separated to individual physical addresses on micro-fabricated arrays, such as a series of nano-wells, in which DNA capture sequences, complementary to the labelling tags, have been chemically attached. Treatment of individually captured transcripts with in vitro expression lysates, enables production and subsequent analysis of the expressed proteins.

Key words:

INTRODUCTION

It has been relatively easy for the scientific community to participate in molecular biology because techniques for the handling and manipulation of nucleic acids tend to be generic. This reflects the fact that nucleic acids are made of only four simple building blocks and the physical chemical properties of nucleic acids are only marginally influenced by their primary sequence. In contrast, the behaviour of proteins is entirely dependent on their primary sequence. The very different physical and chemical properties of the twenty amino acid building blocks of proteins mean therefore that different proteins behave in very different ways in vitro or in vivo. It is difficult to define a generic range of techniques that will enable any individual protein to be analysed. This means that tools to facilitate the handling and manipulation of proteins are crucial, particularly for less experienced investigators (Harry et al., 2000, Kellermayer, 2005).

Completion of the human, mouse and other genome projects had multiple impacts on research in the life sciences (Mouse Genome Sequencing Consortium, 2002). On the one hand, the rate of identification of genes responsible for a variety of monogenic and polygenic diseases has increased enormously (Bond et al., 2002). Moreover, it became increasingly clear that the behaviour of a gene product is difficult or impossible to predict merely from its DNA sequence (Dove, 1999; Graveley, 2001). Definition of the complete complement of genes, and therefore proteins encoded in these organisms, means that functional analysis of the proteome on a scale previously impossible, matured into the well established field of proteomics.

Proteomic analysis is largely a multidisciplinary art of wide range biomedical applications. The combination of chemical, physical as well as computational techniques is helping the development of more efficient proteome analysis tools.

It is relatively straightforward to prepare recombinant proteins in bacterial systems but insolubility is often a problem. Protein/protein interactions can be studied in yeast-2 hybrid systems, but again this technology is not trivial to use and becomes particularly difficult to handle, for example, where proteins are themselves transcriptional activators or where complex multi-protein interactions are involved. Protein/protein interactions can also be identified using immunoprecipitation techniques. Again, this technology can be difficult to use.
in practice. Firstly, it requires the availability of a satisfactory antibody to the protein in question, which may be difficult to generate. It will also require an antibody which is active in Western blotting and again the availability of these cannot be guaranteed. Alternatively, immunoprecipitation can be undertaken after transfecting cells with the protein of interest expressed fused to an epitope tag. Reliable antibodies to a range of such tags are commercially available. However, it may not always be possible to efficiently transfect a specific cell line of interest to establish the role of a particular protein in that tissue.

The currently available protein high throughput analysis techniques such as ribosome display are difficult and cannot easily be automated (Mattheakis et al., 1994; Heyman et al., 1999, Weng et al., 2002, Veenstra & Zhou 2007). In addition, technologies for making protein arrays (MacBeath & Schreiber, 2000) require availability of purified protein samples which is a formidable task when high throughput analysis is required. Separation of complex mixtures of proteins such as those found in the cell is also possible by 2D polyacrylamide gel electrophoresis. It is possible to detect quantitative changes in levels of the individual proteins by this approach, either by labelling the proteins directly or by blotting. 2D gel electrophoresis can be interfaced with mass spectrometry to allow the characterisation of individual protein spots. Again, this is non-trivial technology for the non-expert and it has its limitations (Gygi et al., 2000). Thus, there is scope for new and improved approaches to protein analysis.

Here, I used biochemical tools to invent a novel proteome analysis system with scalable capability to suite a wide range of applications. The scheme provides a more efficient alternative to available proteomic analysis approaches; by overcoming major drawbacks and presenting a generic scalable platform. It also opens the door for a wide range of applications ranging from on-chip high density diagnosis tools to large scale high throughput analysis platforms.

MATERIALS AND METHODS

All chemicals were purchased from Sigma at Molecular Biology Grade, unless otherwise stated. Oligonucleotides were synthesized using phosphoramidite chemistry, by GeneLink, New York, USA. They were synthesized at 1 micro mole scale, PAGE purified, desalted and lyophilized. Resuspension of oligoes has always been done in deionized water (Millipore). Corning® DNA-BIND® surface multiwell plates, were purchased from Corning Europe. Plasmid vectors were purchased from Invitrogen, Clontech, Qiagene and Promega. MALDI-TOF GOLD PSII protein chip was purchased from Ciphergen, USA. Analysis of captured protein species on the chip was conducted on Ciphergen Protein Chip reader, while the analysis of the MS spectrum was done via Ciphergen Protein Chip Software. Pictures for fluorescent proteins were taken with Nikon Eclipse C1si Confocal Microscope System, Nikon, USA. DNA manipulation enzymes, as well as competent cells, were purchased from Stratagene, UK. ZeroBlunt® TOPO® Cloning Kits was purchased from Invitrogen, USA, while the coupled transcription/translation system, the Protein Script-PRO system was purchased from Ambion, USA. All restriction enzymes were purchased from New England Biolabs, UK.

RESULTS AND DISCUSSION

Key System Design:

I reasoned it might be possible to use solid-phase cloning to capture individual mRNA species from complex mixtures into specific locations in micro-wells. I also articulated that if this was possible, I would use in vitro coupled transcription/translation reagents to express the corresponding protein in the same micro-well to create a miniature protein microarray system. It would then be possible to analyse such proteins individually by, for example, mass spectrometric methods. Alternatively, if these proteins were all epitope tagged, then it should be possible to retain them within individual micro-wells for the study. It would also be possible, for example, to take a labelled protein of interest, apply it to all the micro-wells in a large array and address the question of which mRNAs in a complex mixture generate proteins that interact with the labelled protein. This technology will be particularly useful in identifying unknown proteins that interact with multi-protein complexes, by labelling the latter and then determining binding to specific micro-well addresses.

There is an extensive literature pertaining to all of the different components of the system described above, but no completely satisfactory fusion of technologies to deliver the complete package. Microfabrication and high throughput analysis are established in bio-analysis (Bosse et al., 1996; McCreedy, 2000; Kricka, 2001). Brenner et al., (2000) have developed novel approaches to tag individual molecules in complex mixtures of nucleic acids and used nucleic acids complimentary to these tags to physically separate different cDNA species. It is possible to generate extremely complex mixtures of tags for this purpose. Brenner (1994), Shchepinov...
et al., (1997) and Southern et al., (1999) have discussed theoretical aspects of the design of capture oligonucleotides and the ways in which to choose complimentary sequences so as to avoid individually tagged cDNA species from interacting one with another. Highly efficient transcription/translation systems are now commercially available to generate proteins from immobilised DNA fragments (Andreadis & Chrissey, 2000), and many efficient in vitro protein expression systems are widely used (Nakano & Yamane 1998; Kim et al., 1996; Makeyev et al., 1999; Alimov et al., 2000; Madin et al., 2000; Yamamoto et al., 2000 & Shimizu et al., 2001).

Microanalysis of proteins physically separated into high density arrays is also well developed (Zhu et al., 2000). Protein mass spectrometric identification instruments were adapted to the high throughput scale (Ekstrom et al., 2000) There is an enormous literature on the production of DNA microarrays (Chee et al., 1996; Cheung et al., 1999 & Lipshutz et al., 1999), which can be achieved by robotic spotting of individual PCR products (Hegde et al., 2000). Oligonucleotides are synthesised as arrays in situ by local application of pre-synthesised oligonucleotides to coated surfaces e.g. to polystyrene surfaces, using 5'-amino-functionalised oligonucleotides, (Hacia et al., 1998), or by combinatorial photolithographic technologies (Whitesides et al., 2001).

Selection of Protein Expression Constructs:

I selected six well characterized proteins; green fluorescent protein (GFP), yellow fluorescent protein (YFP), red fluorescent proteins (RFP), dihydroxyolate reductase (DHFR), b-lactamase (b-lac) and Chloramphenicol AcetylTransferase (CAT). These proteins are easy to express as well as characterize via their well documented physical and biochemical properties (Wachter 2006, Koguma et al., 1995, Medeiros, 1984, Shaw and Leslie 1991). I created six constructs for in vitro expression of these proteins; constructs were designed upon commercially available plasmids (Living Colors Fluorescent Proteins Vectors, Clonetech), (pCAT3 Reporter Vectors, Promega), (pEXP5-NT/CALML 3, Invitrogen) and (The QIaexpressionist pQ-40 Expression Vector, Qiagen).

Selection of Solid-support Cloning Surfaces:

With regard to solid-phase cloning, I designed my experiment in two stages, a microwell-stage and Ciphergen chip phase. In both cases, I selected an activated surface suitable for coupling to aminated oligonucleotide. This surface was that of the DNA-BIND 96-well strip plates, from Corning (Figure 3, a, b, & c)) and the Ciphergen GOLD PS20 Protein Chip (Figure 3, d). Both surfaces are capable of binding aminated oligonucleotides as they are based upon a layer of reactive N-oxysuccinimide esters, referred to as NOS groups, which react with nucleophiles, such as primary amines. These NOS groups are covalently linked to the polystyrene surface, in case of the Corning DNA-BIND plates, and the metal surface, in case of the Ciphergen Protein Chip. The PSII is a solid-phase chip, consisting of a rectangle 7.8 cm long, 3 mm thick stainless steel slide. The surface is coated with titanium oxide and derivatized with a bi-functional linker. Once binding event happens, it cannot be washed off the surface. NOS is a widely applied active ester commonly used to couple NOS activated molecules to protein. DNA with primary amine, added synthetically or by in vitro manipulation, can be directly coupled to the NOS surfaces. This coupling is specific and not affected by the amines attached to the adenine guanine, and cytosine rings. Within the bases, the electron resonance created by the ring stabilizes the amino groups making them relatively unreactive, as compared to the attached primary amines. In fact, nucleic acids lack groups that will react with NOS thus DNA and RNA can be considered inert biomolecules during standard coupling procedures.

Construction of Tagged Expression Oligonucleotides:

In order to create the solid-phase cloning and in vitro expression system, it was essential for the population of the proteins investigated to be generated in a uniquely tagged in vitro expression format. In order to do this, I created a set of 12 oligonucleotides, for the amplification of the six constructs from the source vectors using the polymerase chain reaction (Figure 1 and Figure 2). The sex constructs are; b-lactamase construct, DHFR construct, CAT construct, GFP construct, RFP construct & YFP construct. Oligonucleotides were designed as forward and reverse primers for PCR amplification, where in the forward primer series, oligomers were developed with a 5’ distal flanking sequence to allow cloning or subsequent reamplification.

In each primer pair, and in addition to the expression cassette, oligonucleotides also included unique restriction sites for bacterial cloning and expression, prior to the in vitro analysis. Forward oligonucleotides in all of the constructs included a FLAG peptide sequence for subsequent anti-FLAG antibody capture. Each of the reverse oligonucleotides included a unique tag sequence that works as a selection and assortment...
element for solid-phase cloning. The exact sequence of this part of the oligonucleotide was determined by a C++-based computer algorithm; that generates non-complementary and specifically hybridizing DNA sequences. I designed this programme on statistical as well as biochemical principles of DNA hybridization properties as described by Brenner (1994) and Southern et al., (1999 and references therein). In brief, these sequences are designed to have particular G-C contents but have less than 10 potential interacting sequences with any of the other forty-mers in the overall set. Whilst it is possible to generate highly complex mixtures of shorter oligomers (e.g. a collection of randomly synthesised twenty-mer oligonucleotides will contain 420 individual sequences). Many of these will have relatively high levels of self-complimentarity and therefore the potential to cause spurious hybridisation. Moving to longer oligonucleotides and setting a cut-off level with a maximum of 10 potentially interacting residues within any given sequence causes massive reductions in background. The computer algorithm I have developed addressed this problem successfully, data not shown.

Although the programme was used here for the design of six unique oligonucleotide sequences, it can be used to generate any number of unique DNA sequences of variable lengths. The limitations to the number of unique tags that can be generated is determined by the length of the oligomer and the number of nucleotides or their variables that are used. In addition, the overall length of the reverse primer which is synthesized is limited by the phosphoramidite chemistry used in the synthesis process. The length of the DNA tag sequence segment, used in the tagging process here, was 40 bp, but for other implementations of the technology, this can be longer or shorter depending on the overall complexity of the constructs to be assorted.

**Construction of anti-tag Capture Oligomers:**

To solid-phase clone each of the selected six constructs, there will have to be a unique mechanism to capture individually tagged constructs on a solid surface. In order to do this, I created a set of six oligonucleotides each contains, in a reverse phase, a segment of DNA sequence with perfect complement to the tag section in the reverse construct amplification primers (Figure 1 and Figure 2). I called these
Fig. 2: Polyacrylamide gel electrophoresis of forward, reverse and capture oligonucleotides, details in Figure 1. For each oligomer, the left lane shows the crude primer prior to purification, while the right lane is showing 1 micro litre of the PAGE-purified oligo.
oligonucleotides, capture oligoes, and they were synthesized with a special carbon linker. The carbon linker was used in order to spatially separate the DNA capturing segment of the oligomer from the solid support to which the capturing tags are linked. I created anti-tag capture oligonucleotides with a 5’ amino C12 linker (Figure 3, f) for use on both the DNA-BIND plate wells and the Ciphergen Protein Chip. The C12 carbon chain linker (Amino C12) was used to provide necessary steric and charge freedom via a greater distance between the oligo and the solid surface (Steel et al., 2000)

Fig. 3: Solid supports used for solid-phase cloning and coupled transcription/translation and the C12 carbon linker. A: Corning DNA-BIND 8 well stripes in their holder tray. B: Corning 8 well strip side view. C: close up of Corning DNA BIND wells. D: Ciphergen PSII protein chip. F: C12 carbon linker used for immobilization of anti-tag capture oligoes.

Amplification & Cloning of Constructs:
Forward & reverse oligomers were used to amplify constructs from the commercially available plasmids; (Living Colors Flourescent Proteins Vectors, Clonetech), (pCAT3 Reporter, Promega), (pEXP5-NT/CALML 3, Invitrogen) and (The QIAexpressionist pQ-40, Qiagen). Amplification was done in a polymerase chain reaction using *Pfu* DNA polymerase (Stratagene) according to manufacturers recommendations. Constructs were cloned in ZeroBlunt® TOPO® Cloning vector, Invitrogen, according to manufacturer’s recommendations, and
Fig. 3: Agarose gel electrophoresis for digested plasmid constructs (a), and gel purified fragments of the six proteins. b-lac; b-lactamase, CAT; Chloramphenicol AcetylTransferase, DHFR; dihydroxyolate b-lac; b-lactamase, CAT; Chloramphenicol AcetylTransferase, DHFR; dihydroxyolate reductase, RFP; red fluourescent protein, GFP; green fluourescent protein, YFP; yellow fluourescent protein. M: Invitrogen ready load DNA Plus molecular weight marker.

subsequently released with flanking restriction sites showed in Figure 1. Fragments were separated on 1% agarose gel, purified using Qiagene - QIAquick Gel Extraction Kit, and then resuspended in deionised water. Purified constructs were ready for in vitro coupled transcription/translation; Figure 3.

**Functionalizing the DNA-BIND Plates & the Ciphergen Chip Surfaces:**

Generally, I bound individual 5'-amino forty-mers to individual wells, as in this way each well became a specific physical address capable of capturing the specific transcription/translation construct with the specific complementary tag, since the well contained anti-tag oligonucleotides linked at their 5 prime ends via an amido-linkage. In addition to the anti-tag capture oligonucleotides, I bound anti-FLAG antibody (Sigma) to individual wells of the Corning DNA-BIND microwell stripe as well as the Ciphergen PSII Chip. I used variable oligonucleotide and antibody concentrations on the surfaces, until optimum levels for subsequent protein expression and capture stages were established, data not shown. I found that capture oligoes were best used at 0.1 mg/ml while the anti-FLAG antibody was best at 0.25 mg/ml. Capture oligonucleotides were prepared in deionized water and the actual coupling of the aminated oligonucleotide to DNA-BIND™ was accomplished by mixing the DNA in a phosphate buffer, pH 8.5, and then adding the solution to a DNA-BIND™ well. Anti-FLAG antibody was coupled to the surface in phosphate buffer at the same pH. The coupling event, with the DNA-BIND™ surface, was accomplished by incubating the plates for 10 minutes at 37°C, while the antibody, was incubated after woods overnight at 4°C. This coupling is specific and is not affected either by the amines attached to the adenine guanine, and cytosine rings or the subsequent anti-FLAG binding. Non-covalently bound oligoes and anti-bodies, were removed by 10 successive post-coupling washes with 1% SDS at room temperature. The Ciphergen Chip was functionalized with the same oligonucleotide and antibody concentrations although all treatments were preceded with treating the active chip spots with 5 micro litre phosphate buffer, pH 7.0. In addition all Ciphergen chip processing was carried out in a humidity chamber, to prevent drying of the chip active surfaces.

**In vitro Coupled-expression Cloning:**

The in vitro on-chip proteome analysis system, described here, was designed to use the T7 transcription/translation for producing proteins from their encoding constructs. In order to do this, solid-support cloning had to be accomplished first. I divided the experiment into two stages; the first is a large scale 100
Fig. 5: Fluorescent imaging of coupled transcription/translation of the \textit{in vitro} cloned and expressed constructs after solid-phase cloning. Images were taken at the three excitation wavelengths zero power objective of Nikon Eclipse C1si Confocal Microscope System, NikonRFP, RFP; red fluorescent protein, GFP; green fluorescent protein, YFP; yellow fluorescent protein. indicated, using micro litre DNA-BIND-well stage, and the second is a 10 micro litres PSII Ciphergen Chip stage. In each case, I created an expression construct with a single-stranded 3' terminus by digesting the gel purified expression constructs with T4 DNA polymerase in presence of 1mM dATP at 37°C for 60 minutes then reaction was stopped by heating at 75°C for 10 minutes, followed with gel extraction and purification with the Qiaquick gel extraction kit (Qiagen). This treatment exposed the 3'-tag end of each expression construct and made it ready for capture by the specific anti-tag immobilized on the solid surface. Purified capture-ready constructs were mixed in 100 microgram equimolar mix inside individual wells, each with only one anti-tag capture oligo. The equimolar mix was incubated for 3 hours at room temperature followed with 3 washes with 1% SDS in PBS buffer pH 7.0. Individual wells were then treated with 20 micro litres coupled transcription/translation mix (PROTEINScript II T7 kit, Ambion), according to manufacturers recommendations, and then incubated at 30°C for 60 minutes. Corning DNA-BIND wells were washed 5 times with 1% SDS in 10 mM PBS pH 7.0, while Ciphergen Chip spots were washed with 10 mM PBS pH 7.0, while kept in the humidity chamber.

In case of the Corning DNA-BIND wells, I used one row of four wells to immobilize four different concentrations of the anti-tag capture oligomers; 0.2, 0.1, 0.05 and 0.025 mg. In case of the Ciphergen Chip, the same anti-tag capture oligoes concentrations and incubation were used but on the active spots.
Fig. 6: Ciphergen SELDI analysis of in vitro expression products captured on Ciphergen chip, following solid-phase cloning using anti-DHFR capture oligoes (a), anti-GFP capture oligoes (b), anti-RFP capture oligoes (c), anti-YFP capture oligoes (d), anti-CAT capture oligoes (e), anti-b-lac capture oligoes (f).

Analysis of Captured Proteins:
Taking advantage of the physical properties of the fluorescent proteins, I used Nikon Eclipse C1si Confocal Microscope System, to analyse individual well stripes in which the protein constructs were expressed. Figure 5, shows that on using illumination at 475 nm, only wells with anti-GFP protein capture tags showed GFP excitation, while at 525 nm wells with anti-YFP tags are lit and wells with anti-RFP tag were only visible when 560 nm was used. No fluorescence was detected for wells with anti-b-lac, anti-DHFR and anti-CAT capture oligomers. For all the proteins captured on the Ciphergen PSII Chip, the mass spectral analysis was performed directly onto each spot on the chip according to manufacturer’s recommendations. After coupled transcription/translation and washing, each spot was treated with 1 ìl of the Ciphergen energy-absorbing matrix, 4-hydroxy-3,5-dimethoxycinnamic acid, in 50% acetonitrile and 0.05% trifluoroacetic acid. The chip was then dried on a 95°C heat block and then transferred to the Ciphergen Protein Chip Reader for analysis. The high voltage detector sensitivity and the laser intensity settings were determined empirically to be optimal for efficient mass spectral analysis of the protein or peptide being studied. Specifically, a high voltage detector sensitivity of 10 along with a laser intensity setting of 260 was determined to be optimal for efficient mass spectral analysis of HSA- and protein A-containing samples. A high voltage detector setting of 10 and a laser intensity of 180 proved optimal for analysis of samples containing the synthetic peptide. Figure 6, shows analysis results which confirm the capture of specific protein species on each chip active spot, corresponding to the immobilized anti-tag capture oligoes present. Signals obtained were clear and loud which indicated the suitability of the system for protein analysis. All data were generated by averaging 65 laser shots on different positions of each protein spot. The Ciphergen Protein Chip Software, was used to identify protein peaks, assign masses, and calculate intensities.

Conclusion:
I have used this pilot system to optimise both levels of oligonucleotide immobilised on surfaces and quantities of PCR amplicons that need to be applied for effective capture, and solid-phase cloning of protein encoding sequences. Having established this, I used an E. coli –based in vitro expression lysate (the Ambion
Protein Script system) to transcribe and translate the captured constructs into proteins at individual wells and spots. Results showed a quantitative correlation between the levels of protein expression achieved and the amounts of amplicons captured, Figure 5. In particular, I amplified constructs encoding green fluorescent protein, yellow fluorescent protein and red fluorescent protein spiked into complex mixtures. I was also able to show that these could be selectively captured and that the appropriate fluorescent protein was easily detectable, after in situ expression. I also expressed three other proteins in this fashion; dihydroxyfolate reductase (DHFR), b-lactamase (b-lac) and Chloramphenicol AcetylTransferase (CAT). Detection and characterization depended upon mass spec as well as characteristic excitation fluorescence of the relevant protein product. I have explored a number of methods to capture these FLAG-tagged proteins at their presence sites of synthesis for further analysis. Again, in terms of optimisation of the process, I have mainly used Ciphergen PSII chips, labelled with anti-FLAG antibody to capture products. These have given the expected results when analysed on SELDI-mass spectrometry. In summary, I have demonstrated that I can use available protein encoding information to amplify protein mixtures and physically clone them in vitro, without using cellular hosts. This process was efficiently coupled to expressing these proteins in situ, from these addressed constructs, followed by efficient characterization.

This system has a great potential to automate the process of proteome analysis. Generation of comprehensive construct libraries, from almost any biological source, can be interrogated relatively easy, especially when mass spec analysis is used to do the initial identification. Once specific proteins of interest are identified, their physical as well as other biochemical characteristics can be studied. In addition, this system allows direct retrieval of the corresponding encoding sequences without the hassles usually involved in cell-based cloning and expression. The volume of the reaction can be reduced to accommodate the specific method of detection adopted in the screening process. This technology is particularly suitable for high throughput applications as all steps starting from design of amplification oligoes to final identification and characterization are suitable for full automation. I anticipate that this technology would represent a more user friendly approach than currently available approaches for protein analysis, such as 2D gel electrophoresis, immuno-precipitation, or even yeast2-hybrid.

ACKNOWLEDGMENTS

Khaled Mahmoud would like to express gratitude for the financial support received from West Riding Medical Trust, UK. Many thanks for the unlimited support received from Professor Sir Alex Markham, Chief Executive of UK Cancer Research, and Dr. Jack Leek, Leeds Institute of Molecular Medicine.

REFERENCES


