Immunological enrichment of low-abundance proteins for comparative proteomics

By
Vinitha Ganesan

A thesis submitted to the Department of Biological Sciences
In partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Carnegie Mellon University
Pittsburgh, Pennsylvania
2017

Thesis Advisor: Dr. Jonathan S. Minden

Acknowledgements

First and foremost, I'm eternally grateful and fortunate for all the educational opportunities the universe has provided me. I thank my advisor Dr. Jonathan S. Minden for letting me join his lab and learn from him. He has been an amazing mentor and an inspiration to keep working. I would like to thank my research advisory committee members Dr. Fred Lanni and Dr. Marcel Bruchez for their valuable opinions and advise in molding my research direction. Thanks to Dr. Olivera Finn for helpful advice during my PhD and serving as a visitor on my research advisory committee. Thanks to the department of Biological sciences and Carnegie Mellon University for providing an excellent educational atmosphere.

I thank Dr. Brigitte Schmidt for synthesis of Biotin-CDM; Dr. Dana Ascherman for the trust, guidance and support through a challenging project; Dr. Sveta Yablonska for being persistent and being a great collaborator.

I thank all the present and past members of the Minden Lab: Dr. Phu T. Van, Dr. Emily Furbee, Malachi Blundon, Amber Lucas, Brendan Redler and Ardon Shorr, Danielle Schlesinger, Tiffany Lau, Samantha Smith, Hannah Kolev, Rachel Willen. Thank you for the helpful discussions and making the lab livelier! Special thanks to all the undergraduate students I closely worked with: Raghunanda Avula, Taylor Maggiacomo, Lawton Tellin, Machika Kaku, Marisa Giglio and Anna Pyzel for being exceptional students.

I thank Dr. Tina Lee, Dr. Gordon Rule, Dr. David Hackney, Dr. John Woolford and their labs for providing support and reagents for experiments. I thank Dr. Adam Linstedt, Dr. Joel McManus and Dr. Nathan Urban for allowing me to rotate in their labs during my first year of the program, when I learned a lot. I really enjoyed working with Dr.Rita Bottino, Dr. Nick Giannoukakis, Dr.Yong Fan and Dr.Massimo Trucco in restructuring the Immunology course. Thanks to Dr. Shoba Subramanian, Ena Miceli, Matthew Salyers, Nate Frezzel for administrative support and making my life easier during this program.

I thank my Cohort for sharing the journey with me. I thank all my friends in the department: Dr. Madhumitha Ramesh, Pieter Spealman, Swati Venkat, Dr. Lina Song, Ceren Tuzmen, Tara Tappen, Dr. Ritika Tiwari and Dr. Sahil Sangani for making this department a welcoming place.

Finally, I thank my support system – my family: husband, mom, dad, Thatha and Paati. I wouldn't be here if not for you. My Pittsburgh friends Anusha, Kokila, Sushil, Ripudaman, Gaurav, Nirmala Aunty, Michelle & Dave who made this journey enjoyable. My high school teacher Dr. A.K.Murugan, who inspired me and encouraged me to apply for the PhD programs, thank you!!

ABSTRACT

In the last 20 years, proteomic studies have not yielded any FDA-approved protein antigen biomarkers for diseases. Commonly used antigen discovery method called Serological Proteome Analysis (SERPA), is not useful in identifying low abundance proteins. On the contrary, Antibody mediated identification of antigens (AMIDA) enriches low abundance protein targets and we believe could improve the discovery rate of true antigen biomarkers using proteomics. However, AMIDA has not been popular due to technical challenges. A major limitation is the contamination posed by antibodies that are used for the isolation of antigens. Antibodies being in high abundance, mask the signal of protein antigens and obstructs the mass spectrometry identification of antigens during the discovery phase of autoantigen biomarker screening.

In an effort to improve the discovery of protein antigens, we present here a solution using a reversible protein capture reagent, called Biotin-CDM. Biotin-CDM can be incorporated into AMIDA in order to remove contaminating antibodies and enrich low abundance protein antigens. We use Biotin-CDM to reversibly tag all potential target proteins in a cell lysate with biotin. The presence of biotin coupled to the target proteins allows for a secondary separation step in which antibodies are washed away from the reversibly biotinylated target proteins by binding them to an Avidin-coupled matrix. The captured target proteins are released from the Avidin matrix by reversing the Biotin-CDM link, thus releasing a pool of target proteins ready for further proteomic analysis compatible with 2DE. Here, we describe the synthesis of Biotin-CDM and optimization of conditions to label proteins without affecting antibody-antigen interactions.

We have successfully incorporated Biotin-CDM in AMIDA to identify antigens targeted by antibodies from Rheumatoid Arthritis (RA) and Interstitial lung disease associated with RA (RA-ILD) patients. Genetic predisposition and cigarette smoking have been linked to a post-translational modification called "citrullination/deimination" that is involved in generating antigens which trigger the formation of neo-epitopes (autoantibodies) in Rheumatoid Arthritis. To understand the pathogenesis of RA-ILD, we screened *in vitro* deiminated antigens that may be preferentially targeted by antibodies from RA vs RA-ILD patients. We identified several antigens including catalase and cAMP-specific 3',5'-cyclic phosphodiesterase 4D isoform PDE4D5 that immunoprecipitated with antibodies from RA & RA-ILD patients. Surprisingly, the same proteins immunoprecipitated in the treated and deiminated samples. The major changes were in the post-translational modifications (PTM) of the immunoprecipitated proteins. Further exploration into the PTM preference by RA/RA-ILD patient antibodies will need to be done in the future.

Finally, we have performed comparative proteomics study on mitochondrial proteomes from Huntington's disease cell line model. Here we show the benefits of using antibodies for the isolation of mitochondria and the critical importance of sample preparation for comparative proteomics.

Contents

List of Figures

List of Tables

Chapter 1 : Immunoproteomics technologies in the discovery of autoantigens in autoimmune diseases.

This chapter is adapted from the review article we published in Biomolecular Concepts (2016) [1].

Immunoproteomics is the study of protein targets of the immune system via high-throughput proteomic technologies [2]. The immune system ordinarily targets foreign-proteins to combat infection or prevent tumor development. However, under the influence of multiple factors such as environment, lifestyle and genetic pre-disposition, the immune system may lose self-tolerance and react against self-proteins (autoantigens), resulting in autoimmune disease. The adaptive immune response against such autoantigens causes cell death and inflammation – resulting in chronic symptoms characteristic of autoimmune disease. Autoantibodies and their cognate target antigens have been used as indicators of several auto-immune diseases. Typically, autoantibodies are used as biomarkers, rather than autoantigens – mainly because most autoantigens represent proteins that exist in normal/ healthy people, while autoantibodies generally mark disease subsets with autoreactivity against these self-antigens (which may/may not be mutated or post-translationally modified).

Autoimmune diseases can either be localized to particular organs or be systemic, with effects in multiple organs of the body. We are in critical need of complex prognostic strategies to monitor and predict the course of systemic diseases in order to institute appropriate treatment modalities. Theranostic (drug responsiveness) biomarkers are believed to be very useful in predicting drug responsiveness and determining time/cost-effective treatment plans [3]. Autoantibodies are a potent source of promising prognostic and theranostic biomarkers for systemic autoimmune diseases.

Whether autoantibodies are used as a diagnostic, prognostic or theranostic biomarkers, they are generally used in the clinical setting to probe for specific cognate autoantigens associated with particular disease states. This implies that if an autoantigen (and its

corresponding autoantibody) has to be used as some kind of biomarker, it must be identified and defined. Thus, it is crucial to discover and identify autoantigen targets in specific diseases, to develop effective diagnostic tools. Immunoproteomics includes a broad set of proteomics technologies that can be used for discovering autoantigens/autoantibodies which may serve as potential biomarkers.

The discovery of autoantigens/autoantibodies happens roughly in three phases:

i) Screening for specific autoantibody/autoantigen combinations in patients
ii) Molecular identification and characterization of the autoantigen
iii) Characterization of the candidate autoantigen's immunogenicity and the corresponding autoantibody signatures.

A common theme in the screening phase is testing the autoreactivity of circulating antibodies within bodily fluids [such as serum, cerebrospinal fluid (CSF) or synovial fluid] from patient cohorts against proteomes sourced from primary cell culture, tissue/cell culture, tissue micro-dissection, or artificially generated peptide libraries/arrays. Proteins that test positively (putative autoantigens) must then be sequenced, identified and characterized using proteomics technologies.

Here, we classify the common immunoproteomics methods used for autoantibody/autoantigen discovery into three categories, primarily based on the technology used for autoantigen screening or molecular identification:

1. Mass spectrometry (MS)-based
2. Nucleic acid-based proteomics
3. Array-based immuno-screening technologies.

In this chapter, we will analyze immunoproteomic data generated across these platforms and review strategies for improving the autoantibody/autoantigen discovery process. Once an autoantigen is identified, its antigenicity has to be validated for it (or its corresponding autoantibody) to be used as a biomarker. This biomarker validation process is essential for US Food and Drug Administration's (FDA) approval and for the

successful translation of potential biomarkers from discovery phase to clinical applications. Readers are referred to a latest review on the biomarker validation process, as it is outside of the scope of this chapter [4].

1.1. Mass spectrometry based technologies

State-of-the-art tandem MS technology is routinely used for identification of proteins in both academia and industry. Several MS-based technologies in the field of biomarker discovery have emerged over the last 20 years. Here we will look at the frequently used discovery approaches that employ MS for antigen peptide detection.

1.1.1. Serological proteome analysis (SERPA)/ PROTEOMEX

The most common method used for profiling autoreactivity of patient sera and identification of antigens in auto-immune diseases is referred to as serological proteome analysis (SERPA) [5] or PROTEOMEX [6]. In this method, whole tissue/cell extracts containing potential autoantigens are run on 2-dimensional electrophoresis (2DE) gel in triplicate (Figure 1.1).

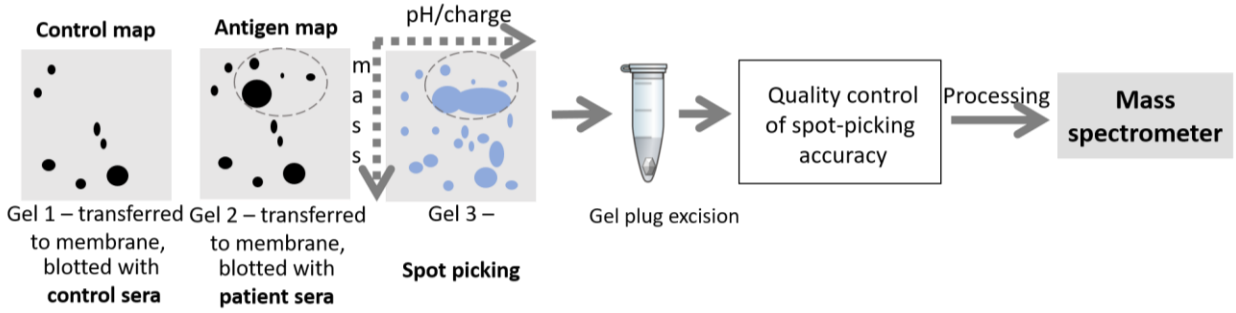


Figure 1.1 Serological Proteomics Approach (SERPA).

The experimental workflow of SERPA in identifying disease associated autoantigens. 2DE gel containing tissue extract is run in triplicates. Proteins from two of the gels are transferred to a membrane for Immunoblotting with control sera or patient sera. Control map, Antigen map are thus generated. These are compared with the third 2DE gel and antigenic protein spots are selected and excised. This gel plugs are then processed for MS peptide sequencing.

Two of these 2DE gels are used for immunoblotting: one against patient sera and the other against control sera from healthy donors. Unique protein spots that specifically

react with patient sera, but not control sera, are detected on the immunoblots and are used as guides to excise gel plugs containing the corresponding protein spots from the third 2DE gel. The gel plugs are then treated with trypsin and resulting peptides are extracted for protein identification using LC-MS/MS or matrix assisted laser desorption/ionization (MALDI)-TOF. Inherent gel-to-gel variability, however, limits the accuracy of spot picking guided by immunoblot maps, which is especially true for low-abundance protein targets. A triangulation approach involving rigorous quality control steps has been suggested in order to accurately identify the protein [7]. To confirm that the correct spots are selected, the 2DE gel from which plugs are excised is transferred to a nitrocellulose membrane and immunoblotted against patient sera. Because the diameter of the protein plug is typically smaller than the diameter of the protein spot, correctly selected spots should leave a halo of immune-reactive material surrounding the hole created by the gel plug, thus confirming that the correct spot has been excised.

Recently, a modified SERPA adapted from difference gel electrophoresis (DIGE) has been described as a fluorescence-based bi-dimensional immunoproteomics (FBIP) approach [8]. The protein mixture is labeled with Cy3 fluorescent dye and loaded on a 2DE gel. The proteins from this 2DE gel are transferred to a nitrocellulose membrane. In a co-hybridization scheme, the membrane is probed with patient sera to generate an antigenic map and with a range of monoclonal antibodies against standard proteins to generate a landmark map. The proteomic, antigenic and landmark maps are then overlaid and compared to identify potential antigenic spots on a second 2DE gel. This improvement enhances the accuracy relative to the previously described method of comparing different spot maps, and is helpful in selecting the correct protein spots across gels.

Many groups have modified SERPA by characterizing circulating antibodies from other bodily fluids such as CSF, bronchoalveolar lavage (BAL) fluid, synovial fluid etc. In some cases, other bodily fluids have also been used as a source for autoantigens [9]. The challenge in this case is depleting high abundance proteins such as albumin from CSF, BAL or synovial fluid in order to resolve the low abundance proteins [10,11].

The major advantages of SERPA are the identification of post translational modification (PTM) states of antigens and the high sensitivity afforded by immunoblots. The gel-to-gel variability in 2DE poses a number of challenges in accurately identifying potential antigenic protein spots for subsequent MS-based sequencing. Improvements to SERPA that we described above, could alleviate some of those challenges. However, the inability of 2DE to resolve hydrophobic, large, and/or basic proteins is still a concern, as a portion of the proteome cannot be screened using this approach.

Several autoimmune biomarker studies describe the discovery of putative autoantigens using SERPA. Some of these putative autoantigens were discovered in multiple autoimmune SERPA studies. These include alpha-enolase, annexin II, and actin subunits. For example, alpha-enolase alone appeared as an autoantigen in 20% of 23 autoimmune studies [9,12–33]. Surprisingly, these same proteins also appeared as antibody targets in a control study of healthy individuals employing SERPA [34]. This pattern of recurring autoantigens may be due to the common inflammatory nature of different autoimmune diseases, as well as possible autoimmune pre-disposition in seemingly healthy individuals. However, the repetitive results may also be an experimental artifact of SERPA, again highlighting limitations that have fueled the development of alternative discovery methods. We will further discuss the proteins discovered using SERPA in the meta-analysis (1.1.5) of this chapter.

1.1.2. Immuno-affinity capture technologies

1.1.2.1. Autoantibody mediated identification of antigens (AMIDA)

As opposed to using patient sera to probe 2DE gel blots, preparative-scale immunoprecipitation (IP) relies on patient/control sera to isolate and enrich autoantigens from soluble mixtures of potential target proteins. In this approach called autoantibody mediated identification of antigens (AMIDA) [35], (patient/control immunoglobulins (antibodies) are first bound to Protein-A beads; relatively large amounts of whole cell/tissue protein lysates containing potential autoantigens are then co-incubated with the antibody-beads. Unbound proteins are washed away, allowing the bound, putative

autoantigen proteins to be eluted. Eluted proteins are then resolved by gel electrophoresis or liquid chromatography (LC), processed, and sequenced via LC-MS/MS (Figure 1.2). Importantly, native or induced post-translational modifications (PTMs) of putative autoantigen targets can be characterized by using 2DE-LC-MS/MS, highlighting the versatility of this approach.

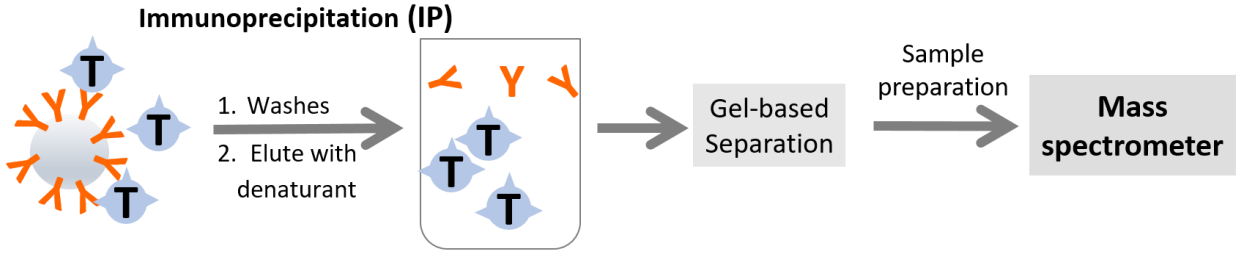


Figure 1.2 Antibody mediated identification of Autoantigens (AMIDA).

Y – antibodies from patients/healthy controls; T – target autoantigens from cell/tissue culture; . Immunoprecipitation is done using patient antibody beads on protein preparation containing potential target autoantigens. The potential target autoantigens can be processed for MS peptide sequencing.

Preparative-scale IP enriches for reactive autoantigens relative to the rest of the proteome, which in principle should be more sensitive than SERPA. However, under the extreme denaturing conditions used for eluting proteins from the antibody beads, bound antibodies often leach from the beads, contaminating the eluted samples [36] and posing a challenge for MS-identification of protein targets. The antibodies, being in high abundance, mask the true target peptide signals. This experimental barrier likely explains why so few published studies have used AMIDA in the discovery phase of autoantibody/autoantigen biomarkers over the last 10 years [37–40].

This antibody contamination problem can be solved by a two-pronged approach: at the data level and/or at the physical level of the experiments. Typically, the peptide sequences originating from antibodies are removed from the MS data during analysis. This helps in the data analysis of true antigen peptides; however, the effectiveness of this analysis strategy is dependent on the dynamic range of the MS instrument used. One of the technical solutions to remove antibody peptides physically is, covalent cross-linking

of antibodies to beads. This approach is helpful, but this process requires optimization and can become cumbersome when working with a large number of samples derived from patient cohorts. If not careful, covalent cross-linking of antibodies can result in inefficient elution of target proteins [41].

One of the major limitations of cross-linking antibodies to beads is that the bound antibodies have limited degrees of freedom to bind antigens. An alternative to this is called the free antibody approach, where unbound antibody is allowed to form immune complexes in the lysate and then the complexes are retrieved by the Protein-A beads. Using free antibody to form immune complexes is beneficial if the target protein is present in low concentrations, the antibody has a weak binding affinity for the antigen or the binding kinetics of the antibody to the antigen are slow. It is challenging to capture such target proteins that are in low abundance or may have weak affinity. We can address this challenge by tagging the proteins in lysate and then mixing them with antibodies to form immune complexes. Thereafter, the immune complexes can be captured by Protein-A beads, the unbound proteins can be washed away and bound antigens can be released. We know that these antigens are contaminated by antibodies. In order to remove these antibodies, we can make use of the tag on the protein antigens to capture the tagged proteins and the antibodies can be washed away.

Therefore, we have developed an alternative solution by synthesizing a protein tag suited for AMIDA, that allows for a free antibody approach. Biotin has extremely high affinity to avidin even under extreme conditions. Commercially available Biotin tags for proteins are either genetically encoded *in vivo* or chemically conjugated *in vitro*. Of those two, chemical conjugation is more pliable for a variety of proteomics experiments. The end goal of AMIDA is to identify protein targets using MS. When using a chemical tag for AMIDA, it is important to remove all of the tag and release the protein targets unmodified for MS identification. “Cleavable” tags leave a chemical moiety still covalently linked to proteins, which affects the charge and mass of the proteins tagged. “Reversible” tags do

not leave any moiety on the protein and thus desirable for proteomics applications such as AMIDA.

Among protein conjugation tags are two classes of coupling moieties to consider: one reacts with cysteine sulfhydryls and the other reacts with primary amine groups of the lysines & amino termini. Of the two, primary amine groups are universally present in all proteins regardless of the protein sequence. Commercially available options for Biotin conjugation via primary amines are cleavable, but not chemically reversible [42–45]. The reversible protein tags that are conjugated through sulfhydryls, are not ideal for AMIDA [46,47].

Thus, we set out to synthesize a new reversible biotin tag (Biotin-CDM) to conjugate proteins via primary amines in order to enrich low-abundance proteins and remove contaminating antibodies in AMIDA (Figure 1.3). This endeavor will be covered in detail in Chapter 2.

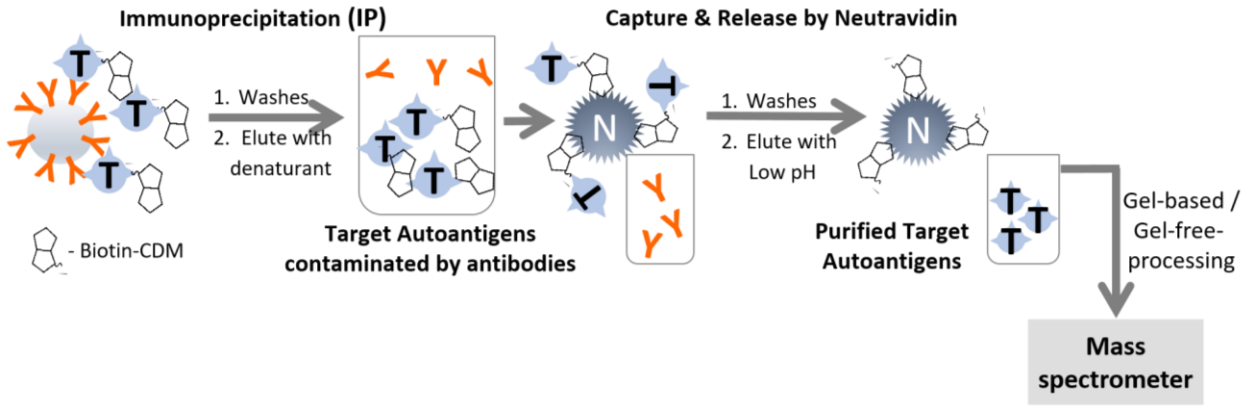


Figure 1.3 Antibody mediated identification of Autoantigens (AMIDA) using Biotin-CDM.

The experimental workflow of AMIDA enhanced by biotin-CDM to remove contaminating antibodies and identify disease associated autoantigens; Y – antibodies from patients/healthy controls; T – target autoantigens from cell/tissue culture; N – neutravidin beads. Protein preparation containing potential target autoantigens are labeled with Biotin-CDM and immunoprecipitation is done using patient antibody beads. The IP eluate is further purified via binding to neutravidin beads, washes followed by a low pH elution reversing the biotin-CDM-to-protein linkage. The purified potential target autoantigens can be processed for MS peptide sequencing.

1.1.3. Circulating immune complexome (CIC) analysis

CICs are circulating protein complexes that contain potential autoantigens, antibodies, pro-inflammatory factors and other clotting factors that occur normally in healthy individuals, but are rapidly cleared by macrophages. However, in autoimmune diseases such as rheumatoid arthritis, systemic sclerosis, and systemic lupus erythematosus, CICs accumulate in blood and can be analyzed to discover new autoantigens. Very similar to AMIDA, CIC's can be isolated from patient sera through binding to protein A/G beads. They are then eluted, trypsin digested, and directly subjected to LC-MS/MS for identification. The MS identification is expected to be obscured by peptides from immunoglobulins and various immune factors, necessitating 'subtractive' sequence analysis of non-immunoglobulin peptides. Overall, this approach – which relies on the dynamic range of protein/peptide detection in the MS instrumentation – is useful for identifying autoantigens in diseases where the presence of disease-specific CIC's is already known [48,49]. Limitations include the requirement for sophisticated, often expensive, technologies such as multiplexing samples through Orbitrap MS instruments.

1.1.4. Surface enhanced laser desorption/ionization – time of flight (SELDI-TOF)

SELDI-TOF is a simpler proteomics approach where protein signatures are compared between multiple samples. Because the identity of proteins is not defined during this comparison, this technology cannot be used for the identification of antigens. However, some studies have used SELDI-TOF for fast screening of autoantigens in several autoimmune diseases, followed by additional MS for actual peptide identification [50–56]. In this approach, antigen-antibody complexes are isolated from patient samples and immobilized on a SELDI chip prior to analysis of mass spectra. Protein peaks are semi-quantifiable and used to create protein signatures. However, a major limitation of this procedure is that only proteins <= 20 KDa size can be analyzed using this method. Moreover, antibodies can also dissociate from antigen-antibody complexes, significantly increasing the noise in the detection system.

1.1.5. Meta-analysis

In order to evaluate the quality of proteomics data generated using the above MS-based approaches, we compiled a list of protein autoantigens discovered using these methods in the last 10 years (**Figure 1.4**). We used a protein abundance database to rank the integrated cellular abundance of each of these proteins expressed in a relative quantifying unit 'parts per million (ppm)'[57]. The unit ppm is used in order to extract, combine and normalize data from several studies using various experiments and technologies. In this meta-analysis, histograms of cellular abundances of protein autoantigens discovered using the respective technologies are plotted: SERPA in Figure 1.4.A [9,12–33] and immuno-affinity capture technologies such as AMIDA, CIC analysis and SELDI-TOF in Figure 1.4.B [35, 37–41, 43–49].

In this qualitative analysis, the shape of the abundance histogram of autoantigens discovered using SERPA appears to be biased towards high abundance proteins. Countering this problem requires that low abundance proteins be enriched by using either large-scale protein preparations (increasing the loading capacity of 2DE gels) or through alternative procedures that include fractionating bodily fluids, cells or organelles (demonstrated in Chapter 4). In principle, immuno-affinity capture technologies are used for enriching autoantigens in biological samples and thus, these technologies should not be limited by protein abundance. When we compare the abundance histogram of protein autoantigens discovered using immuno-affinity capture technologies to that of SERPA, we see a marked difference between the shapes of the histogram (Figure 1.4).

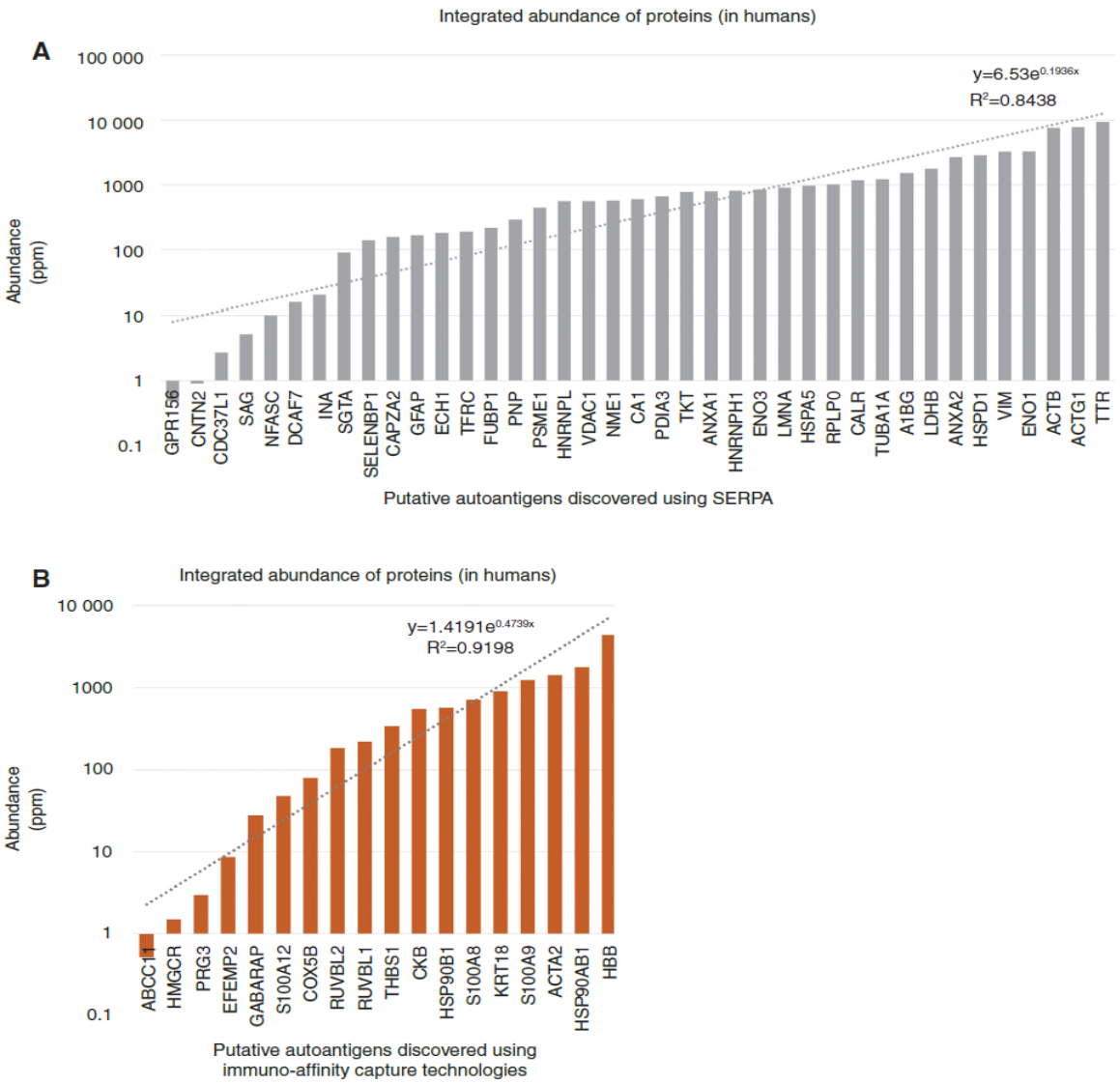


Figure 1.4 Meta-analysis of Immunoproteomic data generated using MS-based approaches.

The vertical axis represents the integrated protein abundance in ppm in a logarithmic scale. PPM (parts per million) is the unit of abundance that is used to quantify relative abundance within the proteome [57] . (A) The horizontal axis represents each protein discovered using SERPA from 24 studies [9, 12–34]. The histogram is clearly skewed towards high abundance proteins. Among these proteins, alpha-enolase, annexin II and actin appear as autoantigens in multiple autoimmune studies that employed SERPA. We notice that these 3 notorious antigens are also on the higher side of the abundance histogram. This shows that SERPA may be biased towards picking up high abundance proteins as putative autoantigens. (B) The horizontal axis represents each protein discovered using immuno-affinity capture technologies such as AMIDA, CIC and SELDI-TOF from 13 studies [35, 37–41, 43–49]. The vertical axis represents the integrated protein abundance in ppm in a logarithmic scale.

We have compared immunoproteomic data from 24 studies using SERPA [9, 12–34] and 13 studies using the other immuno-affinity capture technologies [35, 37–41, 43–49]. We observe that the abundance histogram of autoantigens discovered using immuno-affinity technologies are more equally balanced between many high-abundance proteins and low-abundance proteins, thus the immuno-affinity technologies are not as biased as SERPA towards high abundance proteins. The immuno-affinity technologies could be better for identifying low abundance target autoantigens with further technical improvements.

1.2. Nucleic acid based proteomics

While gene expression libraries do not technically fall under the category of proteomics, the following methods have been quite successful in autoantigen identification and have unique advantages to offer to the field of Immunoproteomics.

Serological analysis of antigens by recombinant cDNA expression cloning (SEREX) is one of the oldest methods used for the identification of autoantigens in several autoimmune diseases [58–69]. Human (tissue–specific) cDNA library derived from an autoimmune patient is used to profile autoantibody repertoires from the same patient in a process called autologous typing. The proteins/epitopes that show autoreactivity are then identified through PCR-based sequencing of DNA from their respective clones. This approach is highly sensitive given the use of DNA-detection, rather than protein-detection methods (since the latter are limited by protein abundance). However, a crucial limitation of SEREX is that it lacks the ability to differentiate or detect post-translational modifications (PTMs) that are likely to play a significant role in breaching immune tolerance in autoimmune diseases such as RA [70].

Phage immunoprecipitation sequencing (PhIP-Seq) is used for profiling the autoantibody repertoires of individual patients, with the potential for 'personalized' diagnosis. In this method, a synthetic human peptidome library is screened against individual patient sera using phage display-based immuno-screening. The reactive

phages are isolated and their DNA is sequenced in a high-throughput manner, allowing peptide identification after extrapolation from the phage DNA sequence [71]. This technology has been applied to detect autoantigens in multiple autoimmune diseases such as rheumatoid arthritis and multiple sclerosis [72]. Again, however, the inability to screen for autoantibodies recognizing post-translationally modified proteins represents a significant limitation of this approach in identifying clinically useful biomarkers for various autoimmune diseases in which modified antigens are targeted.

Overall, nucleic acid based technologies are robust for screening autoantigens. In order to compensate for their limitation in characterizing the post-translation modifications, this kind of screening has to be always followed-up with orthogonal MS-based approaches to define the molecular characteristics of potential autoantigens.

1.3. Array-based immune-screening technologies

As a relatively new technology, autoantigen microarrays have been successfully used to detect and characterize autoantibody profiles for several autoimmune diseases [73–80]. These protein/peptide chips have been generated with as few as 14 proteins to as many as ~17,000 proteins [77] that can be used to screen patient sera for corresponding autoantibodies. Recently, plasmonic microarrays with fluorescent infrared enhancement have been shown to increase the dynamic range of antibody : antigen detection [81]. Because, the arrayed proteins/peptides are recombinant purified, the protein concentration range is not as variable as physiological protein concentration ranges – overcoming the limitations posed by previously described proteomics methods that can be limited by protein abundance. Moreover, detection of PTMs can be incorporated in microarray screening by using synthetic platforms such as a glycosylated peptide array[82]. An additional benefit of this technology includes profiling autoantibody signatures during disease progression [83], as has been shown in a recent review describing the use of proteomic microarrays to study autoantibody profiles in systemic lupus erythematosus [84].

Two newer methods are NAPPA and LIPS. Nucleic acid programmable protein array (NAPPA) is an in situ, cell-free protein expression microarray technology that has been used in the discovery of autoantigens in type 1 diabetes and in the detection of multiple autoantibodies in ankylosing- spondylitis [85]. This technology is at the interface of nucleic acid-based proteomics and array-based technologies. The proteins are synthesized directly on the array along with a fusion tag and captured in place using an anti-tag that is fixed to the array. This is a promising screening platform for personalized diagnosis. Luciferase immunoprecipitation systems (LIPS) is a similar technology that detects antibody: antigen binding via luciferase enzyme and has been used to profile autoantibodies [86,87]. Purified candidates are attached to beads and using the luciferase detection system, the binding events of patient antibodies are detected. This technology is robust and has been used in the characterization of autoantibody signatures and validation of autoantigens.

Though these array-based technologies could be used for screening autoantigens, these technologies are better suited for characterizing autoantibody signatures and studying disease progression. If and when these technologies are used for screening autoantigens, the candidate antigens still have to be further characterized at the molecular level, again using MS-based technologies.

1.4. Current challenges in Immunoproteomics

The proteomic search for biomarkers in the last 2 decades has resulted in a long list of candidate biomarkers for autoimmune diseases. Unfortunately, discovery efforts employing MS-based proteomics technologies have yet to yield any FDA-approved biomarkers for autoimmune diseases [88]. A number of issues may contribute to this shortcoming that is increasingly recognized in the field of proteomics [82, 83], including:

[1] technical or strategical limitations, [2] use of suboptimal statistical methods, and [3] incomplete validation of biomarker candidates.

1.4.1. Technical and strategic considerations

As the biology of autoimmunity is very complex, distinguishing true from artefactual data is critical – and highly dependent on the use of appropriate controls. Furthermore, from our experience with proteomic techniques, variability in sample preparation and handling greatly affect the quality and reliability of proteomic data. Repeated freeze-thawing of both the patient fluid samples and protein extracts from cells/tissues should be avoided because this causes protein loss and inconsistency between samples. In comparative proteomics, label-free proteomics techniques such as LC-MS/MS might produce more artifacts relative to those approaches that use intact proteins and fluorescence detection methods, such as DIGE. These considerations apply while working with any proteomics method used in biomarker discovery.

In terms of research strategies, a hypothesis-driven, targeted search may be better than an exploratory data-driven search at yielding disease-state relevant candidate biomarkers. For example, one could focus on particular PTMs implicated in a disease state during proteomics screening, profile autoreactive proteins in tissue biopsy, or use fractionated body fluids/organelles from patients as a source of autoantigens. When searching for prognostic or theranostic biomarkers, targeted immunoproteomics technologies such as glycosylated peptide array [82] or citrullination probe based MS technology [91] might be employed. In chapter 3, we describe the immunoproteomic screening of autoantigen biomarkers for Rheumatoid Arthritis (RA) patients who develop Interstitial lung disease. Smoking, a risk factor for both Interstitial lung disease and Rheumatoid Arthritis causes an increase in citrullination, a PTM in the Broncho - alveolar compartment of the lungs [92,93]. We approach this biomarker screening with a hypothesis-based approach, that citrullination leads to formation of neo-epitopes implicated in disease progression of Rheumatoid Arthritis Interstitial lung disease.

1.4.2. Use of appropriate statistical methods

As a part of meta-analysis described in this chapter, we also compared the autoantigen proteomic data for multiple sclerosis and rheumatoid arthritis, across the methods described above. No common autoantigen proteins were identified for either multiple sclerosis or Rheumatoid Arthritis when comparing different discovery methods. Yet, this observation may be favorable, indicating that orthogonal approaches improve the likelihood of establishing a more diverse set disease of biomarkers. However, many groups have opted to run fewer proteomics experiments and rely upon ANOVA or other statistical methods to pre-filter their proteomic data before validating the biomarkers [94]. This review [94] suggests that patient cohorts of at least fifty should be used and that pre-filtering of the data should be avoided, in order to make meaningful progress in the identification and validation of protein biomarkers.

1.4.3. Validation of biomarker candidates

As previously discussed, a major concern is the overlap in detected autoantigens in multiple autoimmune diseases (see 1.1 SERPA). These common proteins such as alpha-enolase, actin and annexin-II are also known for their notorious repetitiveness in 2DE based proteomic studies [95]. While autoantigen redundancy may be a general feature of systemic autoimmunity, this study[95] raises concerns related to biases in the various discovery methods employed in different studies. It is possible that these proteins could carry different post-translational modifications or express different isoforms in the disease state that have not been deduced or followed-up in the initial discovery stages – highlighting the need for further characterization. Furthermore, inter-individual differences introduce noise that may cloud the interpretation of autoantibody/autoantigen data. If proteomic data are not validated in larger patient cohorts, then the discovered autoantigens may not ultimately translate into useful biomarkers.

1.5. Conclusions

Autoimmunity is associated with self-directed, dysregulated immune responses that can negatively impact multiple organs depending on the particular disease entity. One potentially interesting use of the described immuno-proteomics methods is to monitor the changes in autoantibody profiles or patterns of autoantigen recognition in longitudinal studies of disease progression. This type of analyses should provide a deeper understanding of autoimmune disease progression and, importantly, aid in developing novel treatment strategies.

While array-based screening technologies and nucleic acid-based proteomics offer high sensitivity, and remove protein abundance bias, neither of these approaches are particularly useful for the detection of post-translational modifications. MS, on the other hand, is capable of detecting PTMs, but the use of this modality often requires targeted searches and significant amounts of patient sample. Despite these limitations, MS-based technologies are still invaluable in the protein identification phase of biomarker discovery.

Of the MS-based technologies used in immunoproteomics, AMIDA seems to have the very little bias towards high abundance proteins, yet only applied in a handful of studies due to technical challenges (see page 5). This suggests that further refinement/development of automated AMIDA could expedite progress in identification and molecular characterization of autoantigens. In Chapter 2 of this thesis, we address the technical barrier we described earlier. We have developed a secondary purification strategy using a novel reversible protein capture reagent called Biotin-CDM in order to remove the contaminating antibodies and enrich the target proteins, thus allowing for the identification of potential auto-antigens [96]. In Chapter 3, we describe the application of this strategy to screen for autoantigens that can help predict and understand the occurrence of Interstitial lung disease in Rheumatoid Arthritis. Chapter 4 deals with yet another affinity approach to isolate mitochondrial proteins to study protein trafficking

defects observed in Huntington's disease. A common theme throughout the manuscript is the use of affinity approaches and 2D-DIGE for comparing proteomes.

Chapter 2 : Immunoproteomics: Development of a novel reagent for separating antibodies from their target proteins

Chapter 2 is adapted from the research article we published in BBA Proteins and Proteomics (2015) [96].

SERPA *(*See Section 1.1*)*, the most popular immunoproteomic approach, is very much limited due to gel-to-gel variability between 2DE experiments. The same limitation of gel-to-gel variability in 2DE is what inspired Jon Minden to invent DIGE, back in 1997 [97]. Secondly, low-abundant autoantigens suffer from poor identification using a cut-map from a different 2DE gel. Although multiple strategies have been developed to address these issues in identifying autoantigens using SERPA [7,8], lack of reproducibility and inability to identify low abundant proteins persist as severe limitations to this approach. Immunoprecipitation enriches low abundant proteins and would be the ideal tool for identifying antigens. This approach using immunoprecipitation to identify proteins is called AMIDA *(*See 1.1.2.1*)* However, as previously stated in Chapter 1, proteomic analysis of immunoprecipitated samples also has significant technical limitations. Thus, we set out to address some of those limitations.

Immunoprecipitation is a ubiquitously used method in biomedical research where antibodies are first bound to a solid matrix, such as Protein-A or Protein-G beads, and then used to capture the antibodies' target proteins from cellular lysates or bodily fluids. An extension of IP is co-immunoprecipitation (Co-IP), which is commonly used to capture the binding partners of target proteins via their binding to previously characterized antibodies. Co-IP is primarily used to explore biological processes such as cell signaling and regulation by studying protein-protein interactions.

Proteomic analysis of IP or co-IP target proteins is often complicated by the presence of variable amounts of immunoglobulins and other background proteins derived from the anti-sera or the target-containing cellular extracts. The background proteins are proteins that bind non-specifically to the antibody or anti-sera beads [98]. Addition of competing

proteins such as BSA helps reduce non-specific proteins from binding the antibody beads. Using non-porous beads, Increasing the salt concentration, number of washes and adding detergent to the wash steps help in reducing the non-specific proteins from binding [99]. These are simple solutions to the non-specific protein problem.

However, the larger problem is the release of abundant immunoglobulins from the affinity beads, which can eclipse the detection of low-abundance target proteins. Generally, high concentrations of denaturants, such as urea or SDS, are used to elute proteins bound to antibodies [99]. Under these strong denaturing conditions, antibodies leach from the resin, contaminating the eluted sample [36]. This poses a serious problem for mass spectrometry (MS)-identification of low abundance target proteins. Even though it looks like the antibodies may be separated from the rest of the sample in an SDS PAGE, the proteins in high abundance leave a trail behind during separation. Immunoglobulin heavy chain runs at 50 KDa and light chain runs at 25 KDa, essentially leaving a long tail behind, covering the entire gel. This is why, even after SDS PAGE separation, we still see contaminating immunoglobulin peptides in mass spec results.

Currently, there is no simple, generic solution to eliminate antibodies from the eluate. Typical solutions to this problem are to either immobilize antibodies covalently to the polymer bead matrix or to covalently cross-link the antibodies to Protein-A or -G beads, via protein-protein crosslinking. Directly conjugating antibodies to a resin results in the random orientation of the antibodies on the matrix, potentially reducing the efficiency of the antibody-antigen interaction. Also, only purified antibodies can be directly linked to the resin, limiting one's options when working with patient anti-sera. Cross-linking of the antibodies to Protein-A/G beads can be a tedious process that requires optimization for each serum or antibody used. If not careful, could result in inefficient elution of target proteins, which is a problem for proteins in low abundance [41]. Over cross-linking risks losing antibody reactivity, while under cross-linking again can lead to variable amounts of antibody leaching from the beads during target elution. Antibodies and anti-sera are usually precious reagents. One is often loath to expend this limited material optimizing

the cross-linking protocol. As discussed previously in Chapter 1, a major limitation of cross-linking antibodies to beads is that free antibodies and antigens cannot be freely mixed together to form robust immune complexes. An alternative solution called free antibody approach can be used to form immune complexes first by mixing patient sera and lysate together. Then the immune complexes can be retrieved by the Protein-A beads. As we discussed previously, this approach allows for the enrichment of low abundance antigens or antigens with weak affinity to patient antibodies.

Seeking a way to overcome this critical limitation of Immunoproteomics, we describe here a method that allows one to efficiently separate antibodies from their target proteins and allows for the free antibody approach. Since antibodies are derived from a huge array of anti-sera, which are usually limited in supply, we decided to focus on the source of target proteins. Target proteins are typically derived from commonly used, tissue culture cell lines, which can be grown in abundance, or from tissue samples. While tissue samples may not be as abundant as cultured cell lines, they are certainly more easily obtained than specific anti-sera. To discriminate between the target proteins and the antibodies that bind them, one needs to attach a suitable affinity reagent to the target proteins. We can (1) tag the proteins in lysate; (2) mix tagged proteins (lysate) with antibodies to form immune complexes. (3) Thereafter, the immune complexes can be captured by Protein-A beads; (4) unbound proteins can be washed away and (5) bound antigens can be released for proteomic analysis. We know that these antigens are contaminated by antibodies. In order to remove these antibodies, we can make use of the tag on the protein antigens to capture the tagged proteins and the antibodies can be washed away.

Ideally, the affinity tag should be able to function properly under the harsh conditions used to elute proteins from antibody beads. Biotin is arguably the best candidate for such an affinity tag since it binds to Avidin and Avidin derivatives with exceptionally high affinity under quite harsh conditions. The ultimate goal of this method is to subject the isolated proteins for proteomic analysis, which may be affected by the presence of the affinity tag. This is particularly true for two-dimensional electrophoresis (2DE) and 2D-

DIGE [97,100]. Any tag is expected to alter the mass and charge of the targets. As discussed previously (See 1.1.2.1), due to the lack of availability of a suitable affinity tag for AMIDA [42–47], we designed a reversible biotin protein-capture reagent called Biotin-CDM. In this chapter, we show: the synthesis of Biotin-CDM and that Biotin-CDM can be used to extensively tag proteins, cellular protein lysates, that these tagged, target-proteins can be captured by their cognate antibodies, that the target proteins can be separated from their cognate antibodies and that the tag can be efficiently removed from the target proteins for 2DE, gel-based proteome analysis. This tool could help improve the identification of proteins using AMIDA.

2.1 Strategy

The goal of this effort was to create a coupling reagent with two features: one end of the reagent should be an affinity tag that could withstand relatively harsh wash conditions, while the other end should form a reversible, covalent bond with target protein lysates, thus coupling the affinity tag to the protein pool. Biotin was chosen because of its very strong affinity to Avidin and its ability to withstand denaturants, such as urea and SDS and reducing reagents, such as DTT and 2-mercaptoethanol.

There were two classes of coupling moieties to consider: one reacts with cysteine sulfhydryls and the other reacts with primary amine groups of the lysines and amino termini. We chose a primary amine-reactive moiety over sulfhydryl reactive compounds because primary amines are present on all proteins (amino terminus) regardless of the amino acid composition.

Maleic anhydride derivatives represent a class of pH reversible, amine-reactive compounds [101,102]. Maleic anhydride forms covalent bonds with amine under mildly basic conditions that can be reversed under mildly acidic conditions [103] , however the reversal requires hours of exposure to complete the reaction. The addition of alkyl groups to the 2 and 3 positions of the maleic anhydride ring greatly accelerates the amide bond reversal [104]. Thus, the overall design of the reagent was to link Biotin to 2,3 alkyl maleic anhydride, which we refer to as Biotin-CDM (see following section on synthesis). Combining CDM

with protein mixtures under mildly basic conditions couples the Biotin-CDM to the protein's lysine residues and amino termini. Lowering the pH to below 5 reverses this linkage, releasing the protein in an unmodified state, ready for further proteomic analysis (Figure 2.1).

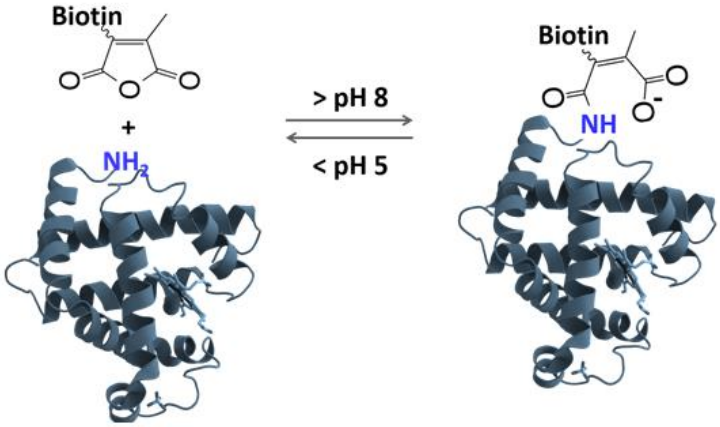


Figure 2.1 Biotin-CDM Protein Reaction Schematic

Diagrams the reversible binding reaction of Biotin-CDM with primary amines on the surface of proteins. (Relative sizes not to scale).

The general scheme for using Biotin-CDM to separate target proteins from their capturing antibodies is cartooned in Figure 2.2. (1) Antibodies found in patient sera bind to their cognate Biotin-CDM-tagged target proteins which are typically extracted from cell cultures. Immune-complexes are formed in this mixture. (2) Protein-A beads bind the these immune-complexes via the antibodies. (3) Unbound proteins are washed away. (4) A denaturant solution (such as 8M urea) strips the captured Biotin-CDM-tagged target proteins, as well as the Protein-A-bound-antibodies. Since the elution is performed under denaturing conditions, the antibodies no longer bind the target proteins. (5) This eluate is then combined with NeutrAvidin beads, capturing Biotin-CDM-tagged proteins, but not the un-tagged antibodies. (6) Any remaining non-biotinylated proteins are washed away and then (7) Low pH elution hydrolyzes Biotin-CDM link to the target proteins, thus releasing the target proteins. Biotin-CDM remains bound to the NeutrAvidin beads. Now these purified target proteins are ready to be identified and characterized using MS,

thus addressing the challenge of identifying low abundance target proteins along with the overwhelming presence of contaminating antibodies.

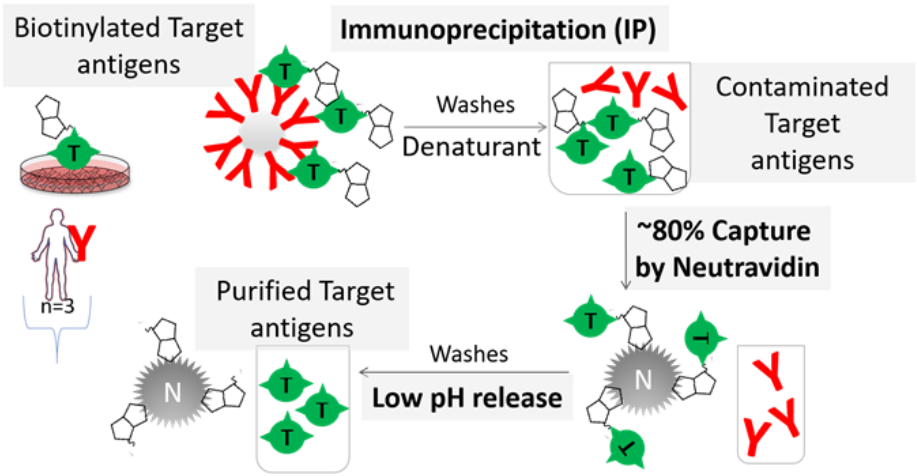


Figure 2.2 Using Biotin-CDM to capture and release target proteins in AMIDA.

Biotin-CDM tagged target proteins are shown as the letter 'T' with a bound Biotin, represented by two fused pentamer rings. Antibody is shown as the letter 'Y'. Neutravidin is shown as the letter 'N'. The flowchart depicts the elution of biotinylated target proteins from antibody beads followed by antibody removal using the reversibility of Biotin-CDM.

2.2 Materials

The hydroxy-succinimidyl ester of 2-propionic-3-methylmaleic anhydride (CDM-NHS ester) was custom synthesized by GL Synthesis Inc (Worcester, MA). Biotin was purchased from AK Scientific corporation (Union City, CA). 2,2'-(Ethylenedioxy) diethylamine was purchased from Tokyo Chemical Industry Company of America (TCI America, Portland, OR). The target compound was purified on a Buchi Corp. Sepacore (New Castle, DE), MPLC system and further analyzed by UPLC (Waters Corp., Acquity, Milford, MA). ^{1}H-NMR and COSY experiments were performed on a Bruker Corp. 300 MHz instrument (Billerica, MA). ESI-MS analysis was performed on a Finnigan LCQ (Thermo Fisher Scientific, Waltham, MA). Bovine Serum Albumin, Fraction V (BSA) was purchased from Fisher Scientific (Thermo Fisher Scientific, Waltham, MA) and Alcohol dehydrogenase (ADH) was purchased from Sigma Aldrich Corp., St. Louis, MO). Minimal DIGE dyes Cy3 and Cy5 were obtained from GE Healthcare (Pittsburgh, PA). Amicon Ultra 4 ml 10K NMWL spin filters were purchased from EMD Millipore

(Billerica, MA). High capacity NeutrAvidin beads were purchased from Pierce (Thermo Fisher Scientific, Waltham, MA). Protein-A Sepharose CL-4B was purchased from GE Healthcare (Pittsburgh, PA). Sera Mag speed bead Neutravidin was purchased from GE Healthcare. Protein A & Protein G Dynabeads were purchased from Thermo-Fisher Scientific. 18 cm pH 3-10NL IPG strips were purchased from BIO-RAD Corp. (Hercules, CA). Ac-Lys-PNA was purchased from Bachem, USA.

N-Boc-2,2'-(ethylenedioxy)diethylamine, *BocNH-2p-NH$_2$* was prepared according to the procedure of Lee et.al [105]. The starting material for this reagent, 1,2-Bis(2-aminoethoxy) ethane was purchased from TCI America Inc. Biotin-NHS ester was purchased from Chem-Impex International,Inc. Diethylisopropylamine was purchased from Aldrich Chemical Company, Inc. Ethyl acetate, methanol, acetonitrile and Dimethylformamide (DMF) were purchased from EMD Millipore. Silicagel 60 Å, Premium R_f was purchased from Sorbend Technologies. All the other reagents were purchased from Fisher Scientific. MPLC chromatography was performed on a Buchi Sepacore system using RP-18 (SMT Bulk-C18) manufactured by Separation Methods Technologies.

To prepare N-Boc-2,2'-(ethylenedioxy)diethylamine, a solution of 150 mL of MeOH with cooling at 0 °C, HCl gas (17 g) was added with stirring for 15 min. The mixture was stirred for 15 min at room temperature and was carefully added to N,N'-bis(aminoethyoxyethane) (67g, 0.466 mol) at 0 °C. The mixture was stirred for 15 min at room temperature before adding 50 ml of H_2O and stirring for another 0.5 h. To the solution $(BOC)_2O$ (101 g, 0.466 mol) in 200 ml of MeOH was added at room temperature for 10 min, and the resultant solution was stirred for 1 h. The mixture was concentrated in vacuo. Unreacted diamine was removed by diethyl ether extraction (300 mL x 2). The residue was treated with 2 N NaOH (500 ml) solution. The product in the organic layer was extracted with CH_2Cl_2 (300 ml x 3). The combined extracts were washed with 300 ml of brine, dried over anhydrous $MgSO_4$, and concentrated in vacuo to yield 64.6 g (87%) of mono-BOC product as a colorless oil with more than 97% purity by HPLC.

2.3 Chemical synthesis

Synthesis of Biotin-2p-NH_2 hydrochloride 1

N-Boc-2,2'-(ethylenedioxy)diethylamine *BocNH-2p-NH_2* (992 mg, 4 mmol) was added to biotin-NHS ester (1.36 g, 4 mmol) dissolved in 5 ml of dry DMF. Diethylisopropylamine (0.7 ml, 4 mmol) was added and the reaction mixture was stirred at 50 °C for 2 hrs. The reaction mixture was concentrated under vacuum. The residue was dissolved in ethyl acetate (70 ml) and washed with 1 M citric acid, water and 1 M bicarbonate. The organic phase was dried over sodium sulfate and concentrated to give 1.28 g (67 % yield) of an oily residue. The TLC (silica gel) showed one spot at Rf=0.2 (eluent: ethyl acetate /20% methanol) that stained positive for biotin.

Biotin-2p-NH-Boc (1 g, 2.1 mmol) was dissolved in methanol (5 ml) and 1 N HCl (5 ml) was added. The reaction was stirred overnight at room temperature. TLC showed the removal of the Boc group. The reaction mixture was concentrated and the resulting solid used as such in the next reaction step. ESI/positive M^+ 375.3 g/mol.

Synthesis of Biotin-CDM 4 (Figure 2.3**)**

Biotin-2p-NH2 hydrochloride **1** (2.2g, 5.3 mmol) was dissolved in 1M Triethylammonium bicarbonate buffer pH 8.5 (10 ml). CDM-NHS ester **2** (0.75 g, 2.65 mmol) dissolved in dry dioxane (5 ml) was added dropwise under stirring at room temperature. Stirring was continued for 1 hr. The reaction mixture was concentrated to a volume of 1 ml. 2N HCl (5 ml) was added to hydrolyze **3** and the reaction mixture was stirred for 30 minutes. The reaction mixture was separated by MPLC chromatography on a Sepacore system on RP-18 with acetonitrile/water/0.1%TFA as mobile phase using a step-gradient. The product eluted at 20% acetonitrile (monitoring at 254 nm). The product fractions were analyzed by UPLC (Waters), RP-18 column, acetonitrile/water/0.1% TFA, 0-0.5min, 0 % acetonitrile, 0.5 min-3 min 0-100% acetonitrile linear gradient, run time 5 min, monitoring at 240 nm and 256 nm. Pure fractions at Rf 1.72 were collected and concentrated to yield 486 mg (34%) of the colorless resinous product **4**.

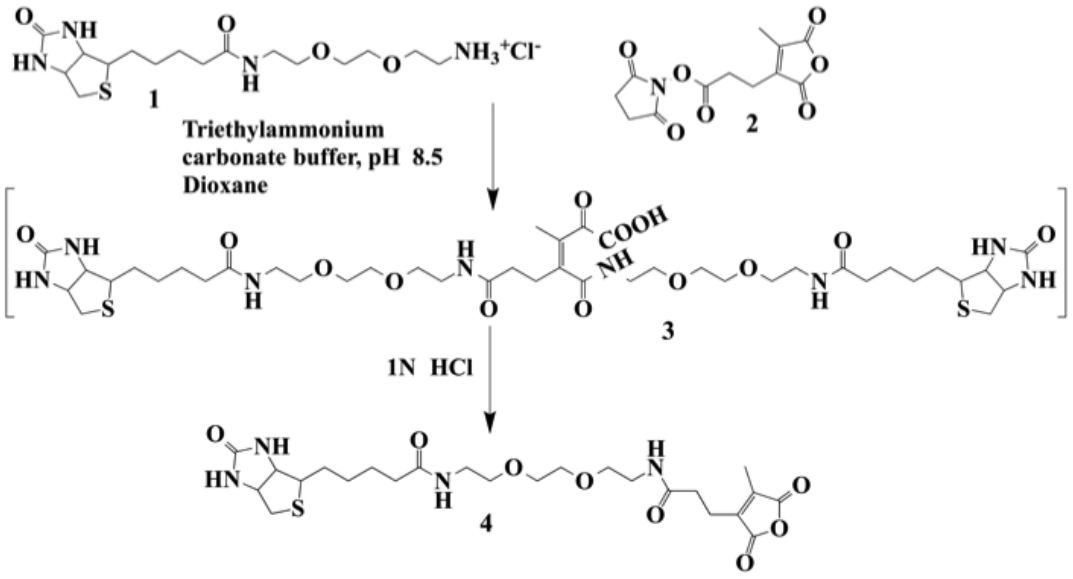


Figure 2.3 Biotin-CDM Synthesis.

Shown here is the synthetic scheme for Biotin-CDM where two molar equivalents of Biotin-2p-NH2 hydrochloride 1 are reacted with one molar equivalent of CDM-NHS ester 2. This reaction produces an intermediate compound that has Biotin-2p-amine bound to the maleic anhydride ring. This linkage is reversed in acid to produce Biotin-CDM.

^{1}H-NMR (MeOD, 300 MHz) 4.51 (1H,dd, J=7.8 Hz, 4.4 Hz), 4.32 (1H, dd, J=7.8 Hz, 4.4 Hz), 3.63 (4H,m, *2p-linker*), 3.34 (4H,m, *2p-linker*), 3.22 (1H,m), 2.94 (1H, dd, J= 7.8 Hz, 12.6 Hz), 2.78 (2H,t, J=7.2Hz, *CDM*), 2.75 (1H,d, J=12.6 Hz), 2.54 (2H,t, J=7.2 Hz), *CDM*), 2.25 (2H, t, 12.5 Hz, 2.08 (3H,s, *CDM*), 1.68 (4H,m), 1.46 (2H,m).

2.4 Methods

Note: Several rounds of method optimization were performed using different materials and methods. In this chapter, I describe the method optimization and list all the methods used. Further optimizations to reduce non-specific protein contaminants will be outlined in Chapter 3.

2.4.1 Binding and release of Biotinylated proteins

Commercially available pure proteins BSA and ADH were used to test the binding and release of biotinylated proteins from NeutrAvidin resin.

(1) Label pure BSA with Cy3: 1 mg of BSA in 200 µl of Wash Buffer (WB - 20 mM HEPES, 2 M urea, 2% CHAPS, 10 mM DTT, pH 8.0) plus 150 mM NaCl was labeled with 10 nmol of minimal Cy3-DIGE dye (in 5 µl DMF) for 15 minutes on ice.

(2) Reach fluorescently tagged BSA with Biotin-CDM in excess: The above reaction was then made up to 0.5 ml in WB and Biotin-CDM at 55 mM in DMF was added at a 1- to 5-fold mass ratio over protein and further reacted for 30 minutes on ice. Quencher (2 µl of 5 M methylamine, 10 mM HEPES-HCl pH 8 (10, 11) was added to halt the labeling reaction.

(3) Similar labeling reaction was carried out to yield Cy5-, Biotin-CDM-ADH.

Excess dye and Biotin-CDM were removed using Amicon 10K NMWL spin filters with five 10-fold washes of WB plus 150 mM NaCl. Cy3/5-, Biotin-CDM-labeled proteins (5 µg BSA or 100 µg ADH) were bound to 200 µl of 50% slurry of high capacity NeutrAvidin slurry in WB plus 150 mM NaCl for 30 minutes at 4 °C by rotating end-over-end. The flow through was collected by centrifugation at 100 X g for 30 s. The NeutrAvidin beads were washed with a sequence of 0.5 mL solutions containing WB plus 500 mM, 150 mM, 50 mM and 5 mM NaCl. This sequence was used to first reduce non-specific binding and to prepare the beads for a low-salt elution step, which is required for subsequent proteome analysis. The bound proteins were released from the linked Biotin-CDM and eluted from the NeutrAvidin beads using 300 µl of low pH Elution Buffer pH 3.7 (EB - 20 mM citrate, 8 M urea, 2% CHAPS, 10 mM DTT, 2 mM NaCl) by rotating end-over-end at 4 °C for 30 minutes. This elution step was repeated once more to recover maximum protein.

To assess the efficacy of binding and release, a 2% volume of each fraction was loaded on to a 12% SDS-PAGE gel. Fluorescence quantification was done with ImageJ. Care was taken to subtract the background fluorescence signal in order to determine the fraction of input protein bound and released from the beads.

2.4.2 AMIDA

i. HeLa protein lysate preparation

Two different lysis methods were used for extracting proteins from HeLa cells. Lysis method 1 using Lysis Buffer-1 (LB-1) was originally used and then we switched to Lysis method 2 using Lysis Buffer-2 (LB-2) as it yielded higher concentration of proteins. This is because lysis method 1 uses mechanical disruption of cells and method 2 uses a mild detergent. The method used will be mentioned in the corresponding results section.

In both the methods, Protein extract (1 mg in 1ml Lysis Buffer LB) was labeled with 10 nmol minimal Cy3-DIGE dye (resuspended in 4 µl DMF) for 15 minutes on ice. For biotinylated samples, 2 mg Biotin-CDM (from 55 mM stock in DMF) was added to 1 mg of protein extract for 30 minutes on ice and then 10 µl of Quencher was added to stop the labeling reaction.

Lysis method 1: T2-HeLa cells lysates were prepared by first rinsing the 80% confluent 10 cm culture plates with 10 ml cold Dulbecco's phosphate-buffered saline (DPBS) twice. The cells were scraped in 10 ml buffer containing 250 mM Sucrose, 20 mM HEPES and 1 mM EDTA pH 8.0. The cells were pelleted at 930 X g for 5 minutes. The cell pellet was washed in the same sucrose buffer and pelleted at 930 X g for 5 minutes. The pellet volume was estimated and resuspended in 3 volumes of lysis buffer (LB-1) containing 20 mM HEPES pH 8.0, 0.3 M NaCl, 1 mM EDTA, 10 mM DTT, 5 mM CaCl2, protease inhibitor (10 µl/ml) and disrupted using a 3-ml ball homogenizer. The suspension was pelleted at 13000 X g for 15 minutes. Protein concentration of the supernatant was measured using Bradford reagent (Sigma).

Lysis method 2 (Preferred method): T2-HeLa cells lysates were prepared by first rinsing the 80% confluent 10 cm culture plates with 10 ml cold (DPBS) twice. The cells were scraped in 10 ml buffer containing 250 mM Sucrose, 20 mM HEPES and 1 mM EDTA pH 8.0. The cells were pelleted at 930 X g for 5 minutes. The cell pellet was washed in the same sucrose buffer and pelleted at 930 X g for 5 minutes. The pellet was resuspended in IP Lysis buffer (LB-2): 100 mM HEPES pH 8.0, 150 mM NaCl, 1% IgePal CA 630, 1 mM

EDTA, 5% glycerol, 1 mM PMSF and 10 µg/mL Leupeptin + Pepstatin mixture. The cell suspension was pelleted at 13000 X g for 5 minutes. Protein concentration of the supernatant was measured using BCA kit (Pierce). For large scale lysate preparation, we placed a custom order of suspension cell culture from Cell Culture Company, MN.

ii. Antisera/Antibody beads preparation

As discussed previously, protein-fluorophore labeling was done similar to labeling BSA and ADH in the Section 2.4.1 of this chapter. Again, over several rounds of optimization, we have used both porous (Protein A agarose beads from GE healthcare) and non-porous (magnetic beads – Dynabeads from Thermo-Fisher Scientific). For Immunoprecipitation: Protein-G beads were used for binding monoclonal antibodies. Protein-A beads were used for binding polyclonal antibodies and anti-sera.

Over the course of optimization, for binding biotinylated target proteins: we used two types of beads. High capacity Neutravidin agarose beads (Pierce) were used originally. We switched to Sera-mag Neutravidin magnetic beads (GE Healthcare) to prevent loss of protein due to proteins getting trapped in porous beads. The type of beads will be mentioned in the corresponding results section.

Cy3-Tubulin (a gift from Dr. David Hackney) diluted in lysis buffer was immunoprecipitated using Cy5-labeled-antibody "12G10 anti-alpha-tubulin" (Developmental Studies Hybridoma Bank) attached to Protein-G Dynabeads (magnetic) (Thermo Fisher Scientific Life Technologies). Rabbit anti-P115 polyclonal antibody (a gift from Dr. Adam Linstedt) was bound to Protein-A Dynabeads (magnetic) (Thermo Fisher Scientific Life Technologies). Patient Antisera was provided by Dr. Dana Ascherman, U Miami, FL. Antisera was bound to Protein-A agarose beads and then labeled with Cy5.

iii. **Immunoprecipitation**

IP was carried out by end-over-end mixing the Cy-3 labeled protein extract with the Cy5-labeled antibody / anti-sera beads for 2 hours at 4 °C. The beads were washed 5 times with 500 µl of PBS pH 7.4, 0.05% Triton X-100, 0.5 mM PMSF. To elute the proteins from

immune-complexes bound to the beads, 100 µl of a solution containing: 10 mM HEPES pH 8.0, 7 M urea, 2M thiourea, 4% CHAPS, and 10 mM DTT, was added to the beads with vigorous shaking (1000 rpm) for 30 minutes at 4 °C in a mixer block (Bulldog Bio). The beads were then magnetically separated to elute the bound proteins.

iv. Separation of biotinylated target proteins from antibodies

The eluate from the Protein-A beads was diluted with three volumes of a solution containing 100 mM HEPES pH 8 and 150 mM NaCl, lowering the urea concentration to 2 M in the binding reaction. The diluted eluate was incubated with 15 µl SERA Mag Speedbead Neutravidin magnetic beads (GE Healthcare Lifesciences). The Neutravidin beads were washed with a sequence of 1 mL solutions listed in Table 2.1.

Components (stocks used)	WB 2	WB 3	WB 4	EB (pH 3.7)
Urea	4.33 g	4.33 g	4.33 g	4.33 g
CHAPS	0.18 g	0.18 g	0.18 g	0.18 g
1M DTT	90 µL	90 µL	90 µL	90 µL
2 M NaCl	675 µL	225 µL	22.5 µL	9 µL
1M HEPES pH 8.0	900 µL	900 µL	45 µL	---
1M Citric acid pH 2.5	---	---	---	180 µL
Make up to	9 mL	9 mL	9 mL	9 mL

Table 2.1 Buffers used for Sera-Mag speed bead Neutravidin

This sequence was used to first reduce non-specific binding and to prepare the beads for a low-salt elution step, which is required for subsequent proteome analysis. The bound proteins were released from the linked Biotin-CDM and eluted from the Neutravidin beads using 30 µl of low pH elution buffer (EB 20 mM citrate, 8 M urea, 2% CHAPS, 10 mM DTT, 2 mM NaCl pH 3.7) by vigorous shaking (1000 rpm) at 4 °C for 1 hour in a mixer block (Bulldog Bio).

v. DIGE and Two-dimensional electrophoresis

2DE was performed using 18 cm pH 3-10NL IPG strips (Bio-Rad) as previously described[106]. The strips were rehydrated with a rehydration buffer containing 0.2% pH 3-10NL Bio-lyte (Bio-Rad) for 18 hours and IEF was carried out on a Protean IEF instrument (Bio-Rad). The second dimension, SDS-PAGE gel was composed of a 12% resolving gel

run on a Protein II xi electrophoresis apparatus (Bio-Rad). Transfer from IEF to the second dimension was carried out using in-gel equilibration protocol[107]. The gel was imaged and quantified using an in-house-built fluorescent gel imager[108,109]. Gel images were analyzed in ImageJ. Dust specks and "hot" pixels were removed from the images by using the "Process>Noise>Remove Outlier" function of ImageJ. Image contrast for "normal" contrast images was set with a single round of the "Image>Adjust>Brightness/Contrast>Auto" function, which sets the minimum and maximum intensity to the 5th and 95th percentile, respectively. "High" contrast images were set manually to visually enhance the contrast as to view low intensity signals. The intensity ranges of these "normal" and "high" contrast images will be discussed in the results section. Protein spot detection and analysis was performed using DeCyder software (GE Healthcare).

2.5 Results and discussion

2.5.1 Synthesis of Biotin-CDM

The synthesis of Biotin-CDM was carried out in two steps Figure 2.3. First, two equivalents of Biotin-2p-NH_2, which was derived from coupling of N-Boc-2,2'-(ethylenedioxy) diethylamine[105] and biotin-OSu[110] followed by Boc-deprotection, were reacted with one equivalent of the hydroxysuccinimidyl ester of 2-propionic-3-methylmaleic anhydride (CDM-NHS ester) in triethylammonium carbonate buffer, pH 8.5. Two equivalents of Biotin-2p-NH_2 was used in this first step because both the NHS and maleic anhydride moieties of CDM-NHS were capable of reacting with Biotin-2p-NH_2. Since the CDM amide is reversible, while the amide formed through the NHS ester is stable, the resulting bis-biotinylated intermediate was treated with 1N HCl to release the reversible CDM-bound Biotin-2p-NH_2 moiety. The target compound was purified by RP-MPLC on a preparative C-18 column using a step acetonitrile gradient in an aqueous solution (0.1% TFA) and characterized by ^{1}H-NMR.

2.5.2 Optimizing the Biotin-CDM labeling of proteins

An enzyme-kinetics based colorimetric assay was developed to measure the labeling efficiency of Biotin-CDM. Ac-PNA-Lys is a substrate for Trypsin. When digested by Trypsin, a chromophore with peak absorbance at 405 nm is released and thus absorbance @405 nm increases as the reaction proceeds further. Any amine blocking compound such as Biotin-CDM or Acetic anhydride can block Ac-PNA-Lys at the lysine side chain, thus inhibiting trypsin's digestion of this substrate. The assay parameters were originally set up using acetic anhydride as the control (data not shown).

Ac-PNA-Lys was labeled with 150-fold molar excess Biotin-CDM for 30 minutes on ice. Excess Biotin-CDM was quenched using 10-fold molar excess of methylamine over Biotin-CDM in the reaction. Trypsin was added. Reaction kinetics was measured for 15 minutes after adding Trypsin, via change in absorbance at 405 nm (Tecan Safire Plate reader was used) as shown in Figure 2.4.

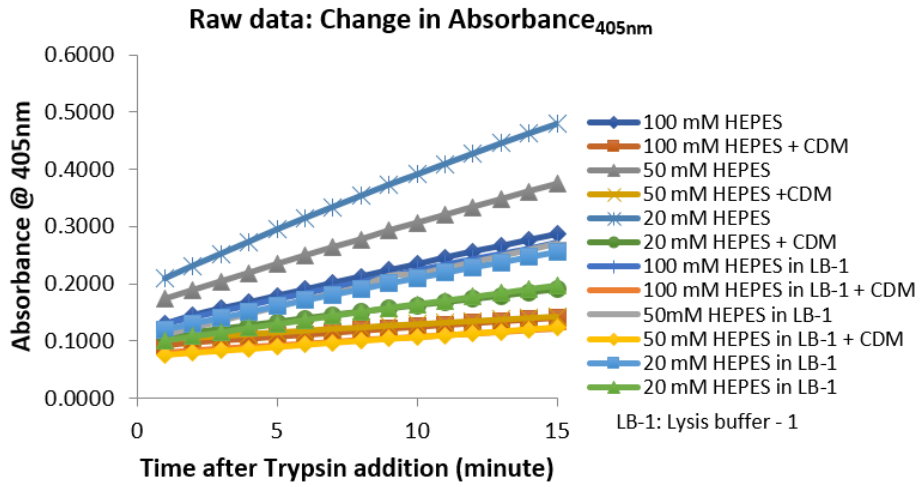


Figure 2.4 Colorimetric assay to measure Biotin-CDM's labeling efficiency.

Raw data from TECAN Safire: Absorbance at 405 nm measured in kinetics mode for 15 minutes. The amine blocking activity of CDM is tested in various HEPES molarity in assay buffer (Simple HEPES buffer) and Lysis buffer-1 (LB-1).

The uninhibited trypsin hydrolysis reactions have the higher slopes. When CDM is added across different conditions, the trypsin lysis reaction is inhibited and the reaction rate is

reduced as can be seen by the lower slopes at the bottom of the graph. The amine blocking effect of Biotin-CDM inhibits the release of chromophore from trypsin catalyzed lysis. The slope for each condition was then calculated from this graph. The labeling efficiency of Biotin-CDM directly correlates with the extent of inhibition of Trypsin activity.

For each condition to be tested, there are two slopes taken into account. For example, to test the labeling efficiency of CDM in 100mM HEPES buffer, $Slope_{100mM\ HEPES}$ and $Slope_{100mM\ HEPES + CDM}$ is calculated. The labeling activity of Biotin-CDM in 100mM HEPES can then be calculated using the formula:

*Labeling activity = [($Slope_{100mM\ HEPES}$ - $Slope_{100mM\ HEPES + CDM}$)/ $Slope_{100mM\ HEPES}$] * 100 %*

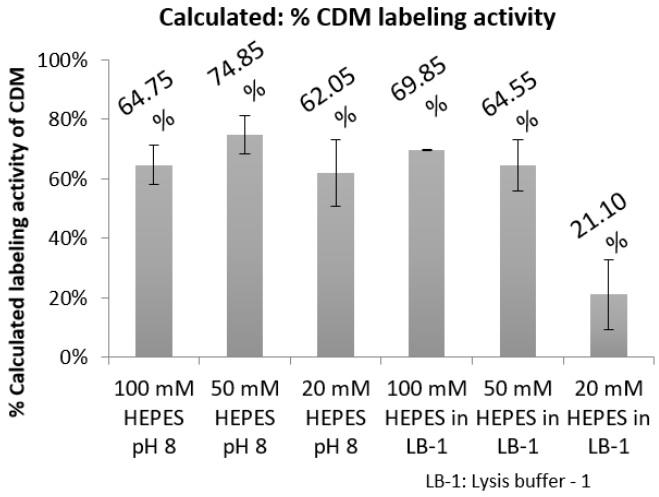


Figure 2.5 Labeling activity of Biotin-CDM in lysis buffer.

Mean values from experimental duplicates are plotted with standard deviation bars to show the range.

We tested the effect of buffer composition on Biotin-CDM's amine labeling efficiency. In simpler HEPES buffers of ionic strength 100 mM, 50 mM and 20 mM, the labeling activity of CDM did not vary (Figure 2.5). However, the activity of CDM was markedly reduced in the lower ionic strength tested for HEPES in the LB-1 buffer (LB-1: 20 mM HEPES, 300 mM NaCl, 1 mM EDTA, 2 mM $CaCl_2$,) *(see section (i) on page 29)*. This effect is likely due to the presence of the other buffer components such as EDTA or $CaCl_2$ in LB-1. Thus, 100 mM HEPES was used in all buffers that were used for extracting HeLa cells to overcome

any inhibitory effects from other buffer components. Subsequently, LB-2 (discussed earlier in section (i) on page 29) was found to be the most efficient for extracting proteins for Immunoprecipitation. Biotin-CDM's amine blocking reaction reaches equilibrium at the point when 68.5% ± 6.5% of the available primary amines are labeled. It is important to note that, in this assay, the substrate contained only one primary amine per molecule. Most proteins contain more than 1 primary amine (due to the presence of multiple lysines), thus increasing the likelihood of attaching multiple Biotin-CDM moieties per protein.

We then tested the activity of CDM in LB-2 to make sure the labeling activity of CDM is acceptable. This assay has since been used as a quality control test to make sure each batch of Biotin-CDM has an acceptable level of activity in preferred buffer conditions (*See appendix I on page 87*). Parameters such as incubation times for labeling and storage conditions for Biotin-CDM were also tested using the same assay. Incubation times between 30 minutes and 1 hour did not affect the labeling efficiency of Biotin-CDM. Different conditions of storage (-80°C, -20°C) of unreacted Biotin-CDM did not have any effect on the Biotin-CDM's labeling efficiency either.

2.5.3 Optimizing the capture and release of Biotin-CDM-labeled proteins

Next, we wished to optimize the capture and release of Biotin-CDM-tagged proteins on NeutrAvidin beads. NeutrAvidin was chosen because it has the highest biotin binding capacity of commercially available Avidin-derivative matrices. There were four optimization criteria: (1) determine the ratio of Biotin-CDM to protein for an excess of NeutrAvidin beads to capture the majority of protein in a solution; (2) determine the binding capacity of NeutrAvidin for Biotin-CDM-tagged protein; (3) determine the optimal conditions for maximum reversal of the CDM linkage and release of protein from NeutrAvidin beads; and (4) assess the fraction of protein species captured and released from NeutrAvidin beads relative to the starting protein lysate.

To optimize the extent of Biotin-CDM labeling for binding to NeutrAvidin beads, our goal was to obtain the maximum labeling the proteins possible without any adverse

effects on protein solubility and accessibility to antibody binding. We now know that with 150-fold molar excess of Biotin-CDM (over primary amines): 68.5% ± 6.5% available primary amines are blocked. We therefore aimed for the amount of Biotin-CDM labeling that allowed for 60-80% biotinylated protein binding on neutravidin beads.

Purified BSA and ADH were first fluorescently tagged with either Cy3-NHS or Cy5-NHS minimal DIGE dyes to aid in quantification. These proteins were then tagged with Biotin-CDM. It was essential to remove as much free Biotin-CDM as possible so that the free Biotin-CDM would not compete with Biotin-CDM-tagged protein for NeutrAvidin binding. Therefore, unbound fluorescent dye and biotin-CDM were removed by spin-dialysis over five rounds of dilution and concentration with binding buffer. The resulting proteins were used for method optimization as follows.

Since each protein contains varying moles of primary amines, we used mass fold excess of Biotin-CDM over proteins for optimization. Coupling ratio of mass fold excess of Biotin-CDM over proteins is denoted as "Biotin-CDM:Protein" throughout this thesis. The weight-to-weight ratio of Biotin-CDM to 5 µg of Cy3- labeled BSA was varied from 0.5 to 2 (Figure 2.6A). All Biotin-CDM labeling ratios of 0.5 to 2 produced capture efficiencies within the target range. The ratios between 1:1 and 2:1 gave the best capture extents for BSA. Increasing the Biotin-CDM to protein ratio beyond 2 led to a decrease in protein solubility (especially when spin-dialysis is performed to remove excess Biotin-CDM). Note that spin-dialysis is not necessary when Biotin-CDM is used for AMIDA, as the excess Biotin-CDM gets removed automatically during the immunoprecipitation washes. Spin-dialysis step was used only for the optimization experiments.

To determine the binding capacity of NeutrAvidin for Biotin-CDM-tagged protein, we titrated Biotin-CDM-tagged BSA (1:1 labeling ratio of Biotin-CDM:protein) relative to 100 µl of packed NeutrAvidin beads. The upper limit of protein binding for this given amount of NeutrAvidin, was found to be 100 µg (Figure 2.6B). This amount of protein is also typical of the amount used for 2D-DIGE experiments.

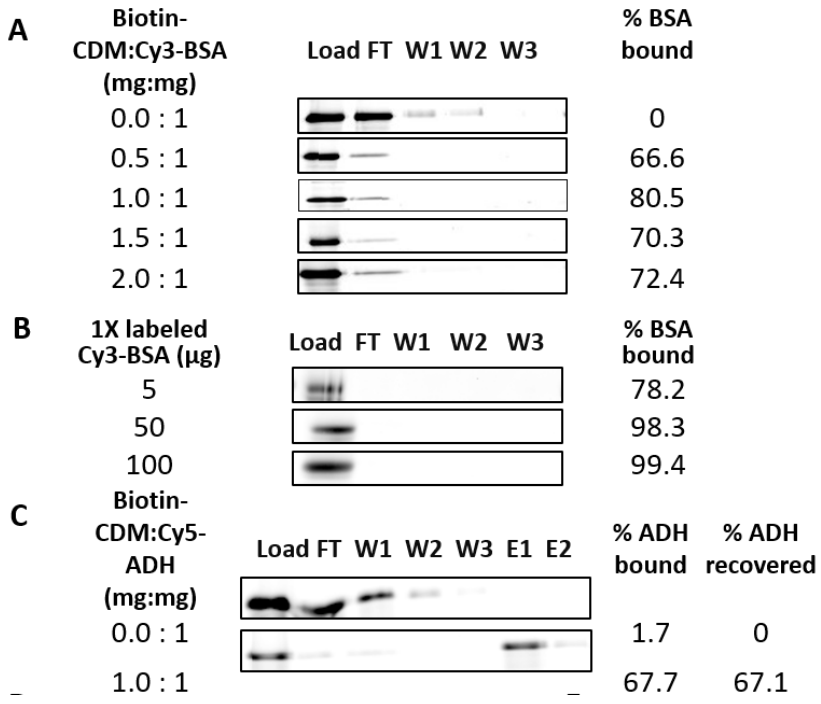


Figure 2.6 Optimization of Capture & Release of Biotinylated Proteins.

(A) SDS-PAGE gel at MW 66 KD, shows the titration of increasing Biotin-CDM to BSA labeling ratio and its impact on binding of Biotinylated-BSA to NeutrAvidin beads. (B) SDS-PAGE gel at MW 66 KD, shows the titration of increasing Biotin-CDM tagged BSA relative to NeutrAvidin beads. (C) SDS-PAGE gel at MW 40 KD, Shows the efficacy of Biotin-CDM tagged ADH release from NeutrAvidin beads.

A variety of conditions were tested for optimizing the scission of the Biotin-CDM-protein bond and the release of protein from the NeutrAvidin beads. The most important considerations were to: determine the optimal pH required to reverse the CDM linkage without adversely affecting protein solubility and finding the balance of salts, buffers, denaturants and detergents that are compatible with follow-on proteomics analysis. For samples that will be analyzed by 2D DIGE, the optimal reversal of CDM linkage was obtained using a low-salt containing elution buffer EB (pH 3.7): 20 mM citrate, 8 M urea, 2% CHAPS, 10 mM DTT, 2 mM NaCl. Incubation of the Biotin-CDM tagged ADH bound to NeutrAvidin beads for 30 minutes at 4 °C provided nearly 100% recovery of the bound protein Figure 2.6C. Figure 2.6C also demonstrates that this eluted material can be

analyzed by conventional SDS-PAGE by mixing the eluate 1:1 with 2 X Laemmli sample buffer.

Finally, we used 2D-DIGE to assess the spectrum of biotin-CDM-tagged HeLa lysate proteins captured and released by NeutrAvidin beads relative to the input protein lysate. Cy3-labeled, Biotin-CDM-tagged HeLa cell lysate (~50 µg extracted using lysis method 1; *see (i) on page 29*) that had been captured and released from NeutrAvidin beads was compared to an equal amount of Cy5-labeled HeLa cell lysate. The 2D-DIGE analysis demonstrated that the majority of protein species were retained throughout the Biotin-CDM tagging, NeutrAvidin binding and acidic reversal of the Biotin-CDM bond.

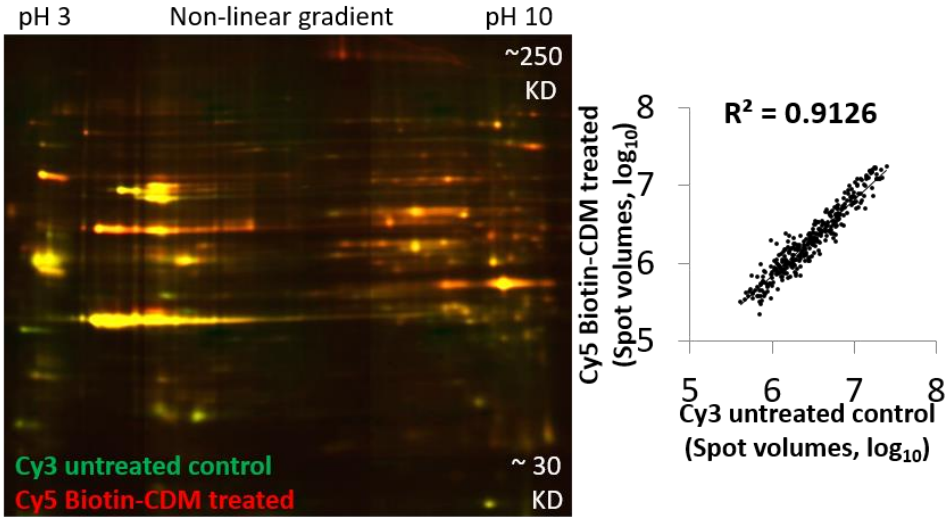


Figure 2.7 HeLa lysate captured and released using Biotin-CDM

Shows a 2D-DIGE of whole HeLa cell lysate labeled with Cy3 (green); whole HeLa cell lysate labeled with Cy5 (red) & Biotin-CDM, excess Biotin-CDM removed via spin-filtration, then captured & released from neutravidin beads. The graph shows the linear correlation between spot volumes of Cy3 and Cy5 channels from the accompanying gel image (Spot identification and quantification was done using a software called DeCyder).

Figure 2.7 shows a 2D-DIGE gel of 2:1 Biotin-CDM labeling of Cy5-labeled HeLa lysate versus Cy3-labeled HeLa control lysate that was not treated with Biotin-CDM. The overlay of the Cy3 (green) and Cy5 (red) channels shows that a very good representation of the majority of HeLa protein spots in the biotin-CDM tagged proteins that were captured by NeutrAvidin beads and released by reversing the Biotin-CDM linkage

versus untreated Cy3-labeled proteins. Automated image analysis (DeCyder) of this 2D-DIGE gel showed that the 328 detected protein spots that ranged over two-orders of magnitude in abundance were represented in both images with a correlation value (R^2) of 0.91. The maximum fold-difference between treated and untreated for individual proteins within these samples was about ±2.5; none of the protein spots detected in the untreated, control sample were absent from the Biotin-CDM captured sample. It is important to note that to perform this analysis the samples were extensively spin dialyzed with 10 kD NMWL devices to remove unbound Biotin-CDM, the consequence of which was a significant loss of protein, resulting in relatively under-loaded 2D gels, particularly in the lower molecular weight range < 20-30 kD. In spite of sub-optimal protein loads, the 2D-DIGE gels showed an excellent recovery of Biotin-CDM labeled proteins from the NeutrAvidin beads. Since even a single Biotin-CDM bound to a protein will still be captured on the NeutrAvidin beads, all of the released proteins should be free of the Biotin-CDM tag. These results demonstrate that the Biotin-CDM is completely released from the target proteins as evidenced by the perfect overlap between control and Biotin-CDM captured proteins.

All of the experiments described were repeated at least three times, all with consistent results. These optimization data show that tagging protein lysates with Biotin-CDM yielded a representative pool of target proteins that can be efficiently captured and released under conditions that will allow for further proteome analysis. These conditions also allow for the removal of unwanted reagents such as salts, denaturants, detergents, and buffers which is important for the efficiency of isoelectric focusing, the first dimension of 2DE.

2.5.4 IP and Biotin-CDM cleanup of known protein antigens using antibodies

a) Polyclonal antibody against P115

To test the efficacy of Biotin-CDM as a method for removing antibody contaminants from an immunoprecipitation of a known antigen using a well-characterized antibody, we

used anti-p115 antibody. p115 is a 115 kD protein with a single transmembrane domain near its C-terminus that resides in the Golgi membrane and is required for proper sorting of Golgi vesicles[111]. HeLa cell lysates, which are known to contain p115, were prepared (using lysis method 1) in LB-1 augmented with 0.5% Triton X-100 to solubilize membrane proteins.

Biotin-CDM treated: The HeLa cell lysate (1 mg) was fluorescently labeled with Cy3-NHS and tagged with Biotin-CDM (2:1 labeling ratio of Biotin-CDM:protein) (Figure 2.8 C and D).

Control: HeLa cell lysate was fluorescently labeled with Cy3-NHS, but not coupled to Biotin-CDM (Figure 2.8 A and B).

Antibodies: Rabbit anti-p115 anti-serum was mixed with Protein-A beads to immobilize the anti-p115 antibodies. After mild washing, the bound proteins were labeled with Cy5-NHS to monitor antibody abundance throughout the IP procedure.

The control experiment, which lacked Biotin-CDM tagging, showed that most of the cell lysate passed directly through the anti-p115: Protein-A beads (Figure 2.8.A, -bead flow thru). Elution of the anti-p115: Protein-A beads with a solution containing 8M urea released a large amount of Cy5-labeled antibody that hinders mass-spec identification of the target proteins (Figure 2.8.A, -p115 eluate). Figure 2.8.A shows the gel image set at "normal" contrast (intensity range – 425 to 19631 counts), which prevents one form visualizing low abundance proteins. To visualize the presence HeLa proteins in the antibody bead eluate, "high" contrast image (intensity range – 425 to 966 counts) of the same gel is shown in Figure 2.8.B.

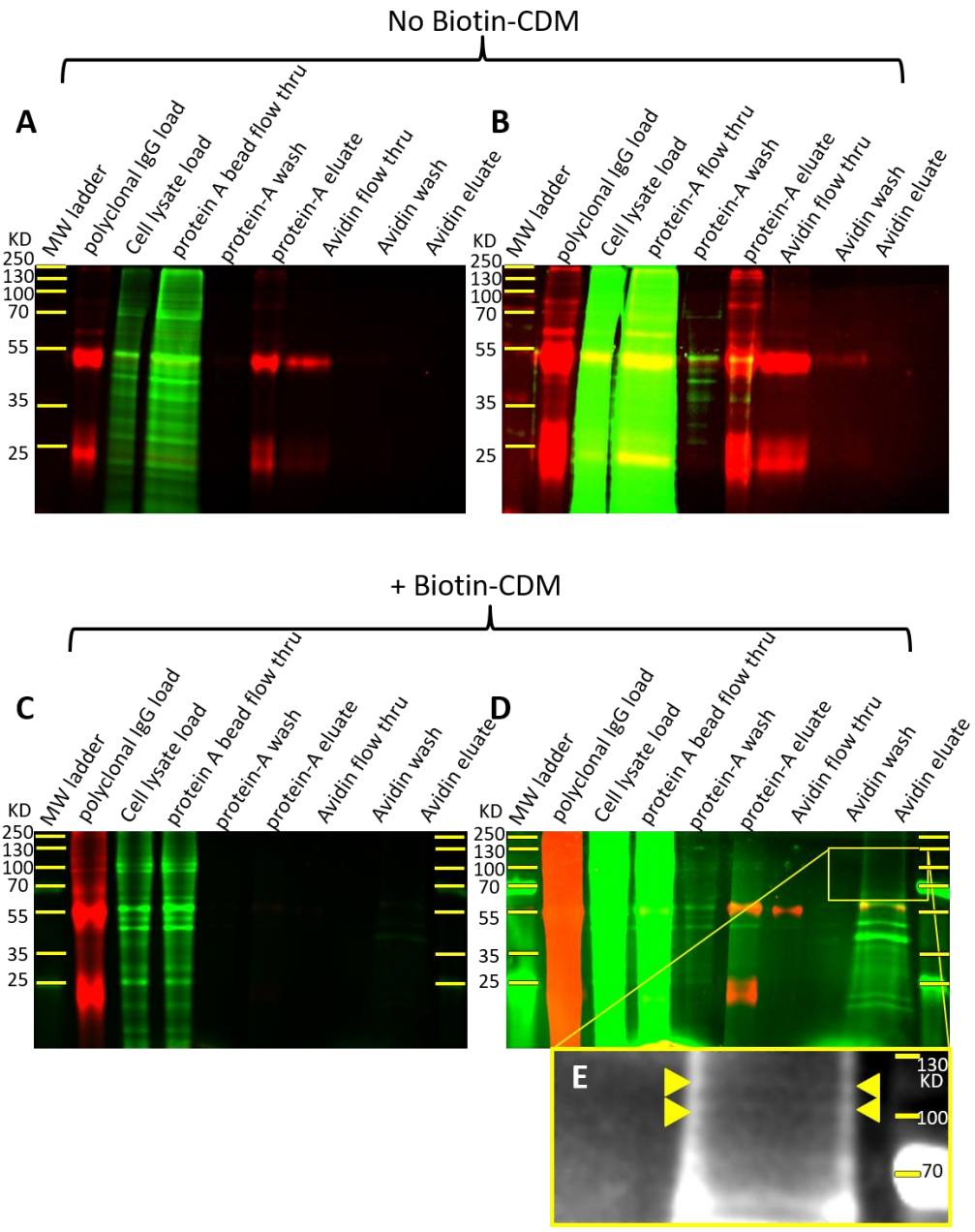


Figure 2.8 Biotin-CDM Capture of a Known Antigen using polyclonal antibody.

(A, normal contrast, and B, high contrast) images of an SDS-PAGE gel containing samples Cy5-anti-p115 antibody and Cy3-tagged HeLa lysate in a control IP experiment without Biotin-CDM. (C, normal contrast, and D, high contrast) images of an SDS-PAGE gel containing samples Cy5-anti-p115 antibody and Cy3-and Biotin-CDM tagged HeLa lysate in an IP experiment. The molecular weight ladder contained 25, 35, 55, 70, 100, 130 and 250 kD standards. (E) Shows a magnified and contrast enhanced view of the 70-130 kD mass range of the NeutrAvidin eluate, the mass standards shown here are 70, 100 and 130 kD. The arrowheads indicate the putative p115 band.

At this enhanced intensity range, one can detect a small amount of released Cy3-HeLa protein (green) and an overwhelming amount of Cy5-anti-p115 antibody (red) (Figure 2.8.B, compare red and green signals in the anti-p115 eluate lane). This material was combined with NeutrAvidin beads. When the proteins lacked Biotin-CDM, all of the Cy5-antibody and Cy3-HeLa protein was found in the bead flow through and wash. As expected, virtually no protein was detected in the eluate of this control experiment, indicating that without Biotin-CDM no protein binds to the NeutrAvidin beads (Figure 2.8.B, Avidin eluate).

As expected, coupling of Biotin-CDM to the HeLa lysates gave a different result (Figure 2.8. C and D). "Normal" contrast image (intensity range – 220 to 20091 counts) of the gel showed only the contents of the antibody beads and the cell lysate load and flow through (Figure. Figure 2.8.B, Cell lysate load and -bead flow thru lanes). The antibody that leaked from the Protein-A beads in the eluate was not visible in this "normal" contrast image due to slight variation between experiments, but is seen in the "high" contrast image (intensity range - 220 to 1060 counts) of the same gel (Figure 2.8.C, -p115 eluate). Increasing the image contrast allowed one to visualize that many Cy3-HeLa (green) proteins were released from both the anti-p115:Protein-A beads and the NeutrAvidin beads (Figure 2.8.C, protein-A eluate and Avidin eluate lanes). On the other hand, Cy5-antibody (red) was only detected in the Protein-A bead eluate, but only slightly detectable in the NeutrAvidin eluate (Figure 2.8.D, protein-A eluate and Avidin eluate lanes). Thus Biotin-CDM effectively removed the vast majority of antibody from the anti-p115 : protein-A eluate.

This experiment using a known antigen, was designed to perform both IP and co-IP. The anti-p115: Protein-A beads were only washed in low salt buffer to detect proteins bound to the target protein and its binding partners. Thus, both specific and non-specific proteins may be bound to the antibody beads. Further IP experiments were performed with more stringent washes to assess the specificity of the bound proteins. One protein that was expected to bind throughout the experiment was p115. As opposed to boiling

all of the antibody beads to load onto the SDS PAGE gel, only 20% fraction of the cleaned-up IP eluate was analyzed in the gel shown. This fraction of eluate was below the detection limit of the antibody using an immunoblot. However, magnifying the section of the 100 to 130 kDa range of the gel and enhancing the contrast revealed 2 protein bands at ~115 kDa (Figure 2.8.E).

In order to improve upon these results and remove the non-specific proteins to visualize p115, we replaced the porous Protein-A agarose beads with non-porous Protein-A Dynabeads (magnetic beads). (*For method: See section (iv) on page 31*). As seen in Figure 2.9, it is now easy to see the unique band at 115K that is present in IP with the antibody. The same band is not present in the no-antibody control (100% eluate was loaded in this gel). The background bands are greatly reduced with the use of non-porous magnetic beads for immunoprecipitation. Figure 2.9 also shows that Biotin-CDM labeling of antigens allows for antibody-antigen interaction and does not negatively affect the detection of antigens.

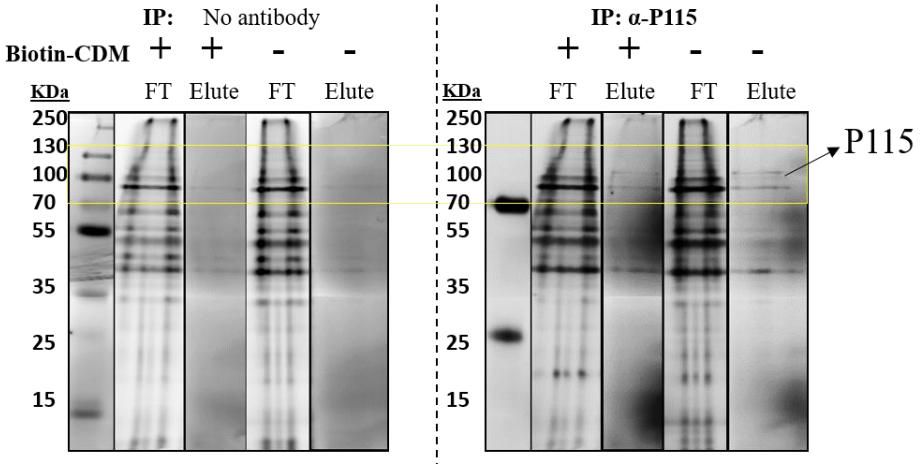


Figure 2.9 Immunoprecipitation of P115 using Dynabeads.

Shows the improvements after replacing Protein A agarose beads with Protein-A Dynabeads magnetic beads. The FT (flow through) and eluates are from the Protein-A Dynabeads. All of the elute fraction was loaded on 12% SDS gels. Only the fluorescence channel showing Cy3-HeLa is imaged. Contrast has been adjusted separately for the individual lanes, so the lanes are not comparable quantitatively.

b) Monoclonal Antibody against Tubulin

Using Tubulin IP as a simplest model of AMIDA, we optimized the parameters for immunoprecipitation using Dynabeads and capture & release of immunoprecipitated proteins using Sera Mag Neutravidin speed beads respectively *(See methods: (iii) & (iv) on page 31)*. While optimizing capture & release using Biotinylated-ADH & Biotinylated-BSA, we were limited by the need for an additional step of spin dialysis in order to remove excess Biotin-CDM. Hence, we used immunoprecipitated Tubulin to further optimize the capture & release of biotinylated proteins where spin-dialysis is not necessary because excess Biotin-CDM gets washed away in the washes performed during immunoprecipitation. Cy3-Tubulin monomers (in green) were tagged with Biotin-CDM (1:1, 2:1, 6:1 labeling ratio of Biotin-CDM: protein). Monoclonal antibody against Tubulin was labeled with Cy5 (in red) and bound to Protein-G Dynabeads (Figure 2.10).

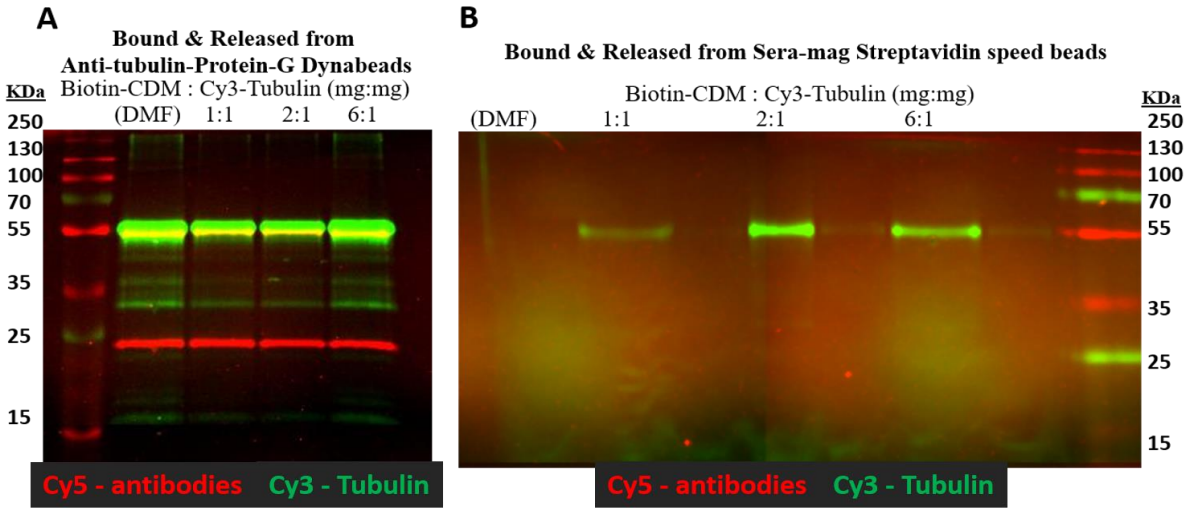


Figure 2.10 Immunoprecipitation of purified Tubulin and the clean-up of the mono-clonal antibody used.

(A) Cy5-monoclonal antibody (in red) against tubulin was bound to Protein-G Dynabeads and biotinylated (Coupling ratio of Biotin-CDM:Protein used were 1:1, 2:1, 6:1) Cy3-Tubulin monomers (in green) were added to the Antibody beads. Immunoprecipitation was carried out with multiple washes and eluates were run on a 12% SDS PAGE gel. Antibodies (in red) co-elute with tubulin (in green). (B) Immunoprecipitated (Coupling ratio of Biotin-CDM:Protein used were 1:1, 2:1, 6:1) Tubulin was run through Sera-mag Neutravidin speed beads and the antibodies were washed away. Antibodies (in red) have been washed away and only tubulin (in green) was eluted.

The eluates from the immunoprecipitation (Figure 2.10) show that Tubulin (green) was successfully immunoprecipitated with the various ratios of Biotin-CDM labeling. The bands in red are immunoglobulins from the monoclonal antibody used for immunoprecipitation. The overlap in tubulin protein band and immunoglobin heavy chain is seen as yellow in the image. The fluorescence level of immunoprecipitated tubulin is slightly lower in Biotin-CDM tagged conditions compared to the untagged control (which had DMF added instead of Biotin-CDM) (Figure 2.10.A). However, 6:1 coupling ratio immunoprecipitated the same level of tubulin as the "no Biotin-CDM" control (DMF lane). This shows that at the labeling ratio of 6:1, we are able to recover almost all the protein at the immunoprecipitation step. While 6:1 ratio seems to give the best results, the magnetic beads (Protein-A Dynabeads) precipitated at such high levels of DMF. At 6:1 ratio, there is considerably more DMF in the reaction. This is because Biotin-CDM is dissolved in DMF after synthesis and purification.

Because precipitation of magnetic beads could introduce more variability, this is not an optimal condition. Therefore, coupling ratios of 2:1, 3:1 or even 4:1 are preferred. Note that these ratios are higher compared to the first round of optimization where spin-dialysis was used for removing excess Biotin-CDM.

As seen in Figure 2.10.B, when the eluates from immunoprecipitation are run through Sera mag Neutravidin speed beads, the biotinylated proteins are captured and released (in green). The contaminating antibodies (in red) that were seen in Figure 2.10.A were washed away.

These data using polyclonal antibodies and monoclonal antibodies to immunoprecipitate P115 and tubulin respectively, demonstrate that this approach is effective at removing antibodies from IP and co-IP experiments. The target proteins are effectively captured and retained by the NeutrAvidin beads and released under mildly acidic conditions. These data demonstrate that Biotin-CDM labeling of target proteins still allows for antibody-antigen binding.

2.5.5 IP and Biotin-CDM cleanup of unknown antigens using patient sera

To evaluate the suitability for the use of Biotin-CDM to cleanup target protein samples immunoprecipitated using patient anti-sera, we tested anti-serum from a patient suffering from Rheumatoid Arthritis. Understanding the range and identity of autoantigens targeted by a patient's immune system will be helpful in furthering our understanding of this common autoimmune disease. Antibodies from the serum of an anonymous Rheumatoid Arthritis (RA) patient were captured by Protein-A beads and labeled with Cy5-NHS so that the antibody proteins could be tracked through this procedure. The RA Sera:Protein-A beads were challenged with Cy3-labeled, Biotin-CDM-tagged HeLa cell lysate (extracted through lysis method–1, *See (i) on page 29*).

As was the case with anti-p115 antibodies, both Cy5-RA-antibodies and Cy3- cell lysate proteins were eluted from the Protein-A agarose beads (Figure 2.11A). This eluate was applied to NeutrAvidin agarose beads. Also, as shown previously, an abundance of Cy5-antibody flowed through the Protein-A beads, which was easily detected in the "normal" contrast SDS-PAGE gel image (intensity range—848 to 21402) (Figure 2.11A). "High" contrast image (intensity range—163 to 2666) was required to observe the abundance of Cy3- labeled HeLa lysate proteins eluted from the NeutrAvidin beads, with a very small amount of Cy5-RA-antibody (Figure 2.11B). Quantification of Cy5 fluorescence from antibody shows that, on an average 98.4% antibodies present in the eluate are removed following this Biotin-CDM cleanup procedure.

An important aspect of the reversibility of Biotin-CDM is that it releases proteins without the Biotin tag regenerating amines, which is essential for 2DE gel analysis. The NeutrAvidin eluate, which contained Cy3-HeLa protein and Cy5-antibodies, was analyzed by 2D DIGE (Figure 2.11 C-E). Figure 2.11E shows a wide variety of HeLa proteins were captured by the RA-anti- serum:Protein-A beads and the NeutrAvidin beads. Automated image analysis (DeCyder) of the cy3 gel (Figure 2.11C) image showed that there are 88 detected protein spots.

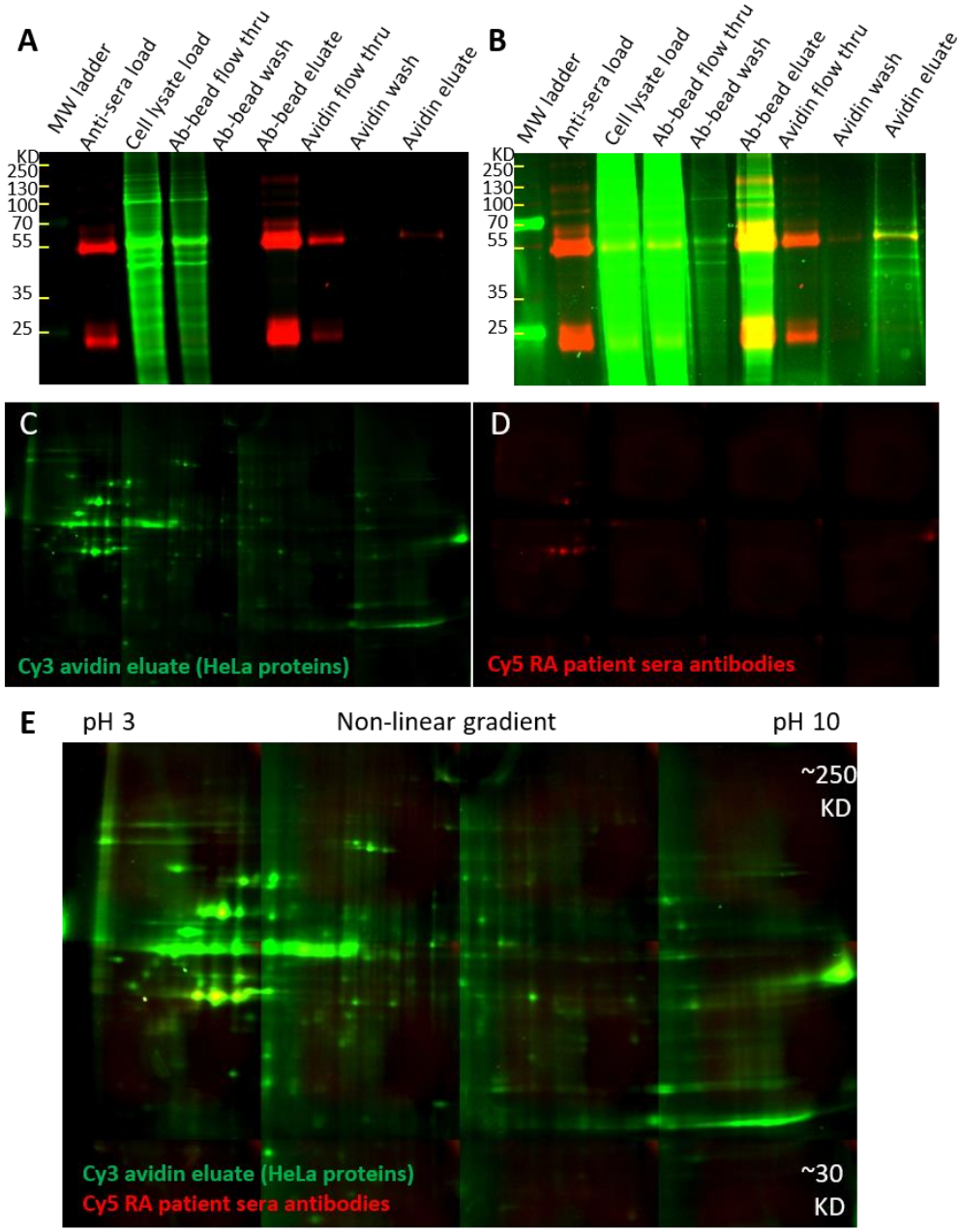


Figure 2.11 Biotin-CDM Capture of Immunoprecipitated unknown Antigens using patient sera.

(A – "normal" contrast, and B – "high" contrast) images of an SDS-PAGE gel containing samples Cy5-RA-anti-serum and Cy3-and Biotin-CDM tagged HeLa lysate in an IP experiment. (C) Cy3 image of a 2D DIGE gel containing Cy5-RA-anti-serum and Cy3, Biotin-CDM-NeutrAvidin eluate. (D) Cy5 image of a 2D DIGE gel containing Cy5-RA-anti-serum and Cy3, Biotin-CDM-NeutrAvidin eluate. (E) Overlay of the C3 and Cy5 images of a 2D DIGE gel containing Cy5-RA-anti-serum and Cy3, Biotin-CDM-NeutrAvidin eluate.

These 88 protein spots covered a wide range of molecular weights, isoelectric points (pI) and abundances. A few spots appeared to coincide with fluorescent spots in the Cy5 image of the DIGE gel (Figure 2.11D) and the overlaid Cy3:Cy5 image (Figure 2.11E). The vast majority of the protein spots were Cy3-labeled HeLa target proteins, demonstrating that antibodies were successfully removed and will no longer interfere with 2DE separation or mass spectrometric identification.

There were few common Cy3- and Cy5-protein spots that appeared to be present in both the Cy3-HeLa lysate and Cy5-RA-anti-serum. We speculate that these may be abundant serum proteins that are present in the RA-anti-serum and may be serum contaminants from the HeLa lysate. Mass spectrometric identification of these proteins and further experimentation will be required to determine the source of these back- ground proteins. Note that agarose beads were used in this experiment for immunoprecipitation and antibody removal. However, as described in the previous section (*See (*b) on page 44b), replacing non-porous agarose beads with magnetic beads should eliminate most of the background proteins. Further steps to eliminate background/non-specific bound proteins will be described in Chapter 3.

2.6 Conclusions

These data demonstrate that the reagent Biotin-CDM is capable of tagging proteins from whole cell lysates to be used as targets for IP and co-IP experiments. The Biotin-CDM tag allows for secondary isolation step whereby highly abundant antibodies can be separated from the target protein pool. Through this approach, we are able to recover 60-80% of the target proteins while removing 95-98% of the antibody contamination. On average, we enrich the target proteins by 26-fold (target protein/antibody retained) from the IP eluates, using Biotin-CDM approach. Thus, providing a general means to cleaning up IP and co-IP products for further proteomics analysis. The key benefit from this procedure are: (1) efficient removal of interfering antibodies from the IP target pool, (2) release of the target proteins without the complication of carrying a variable number of biotin moieties, and (3) this method provides a means to produce standard preparations of

target lysates that do not require antibody-to-antibody optimization for IP experiments. The approach we describe is a combination of immunoprecipitation and 2D-DIGE that was not feasible before. As a proof of principle, we will describe the autoantigen screening we performed using the new and improved version of AMIDA to predict Interstitial lung disease in Rheumatoid Arthritis patients in the next chapter (Chapter 3). In chapter 3, a free antibody approach is taken to allow for enrichment of antigens that may be in low abundance or have weak affinity to patient antibodies. This was not possible previously with the cross-linking of antibodies.

Chapter 3 Screening for autoantigens to predict Interstitial lung disease in Rheumatoid Arthritis patients

This study was done in collaboration with Dr.Dana Ascherman, MD, Univ. of Miami, FL.

Rheumatoid Arthritis (RA) is one of the most common autoimmune diseases, affecting ~1% of the adult US population (Figure 3.1). The most common symptoms of RA are chronic inflammation of joints. The risk factors that affect the outcome of RA fall into 3 categories: genetics, environment and lifestyle factors [112]. RA is heterogenic condition where several single nucleotide polymorphisms (SNPs) are implicated. This systemic disease affects not only joints, but also extra-articular systems involving the heart and lungs. There are 2 antibody-based tests available to diagnose Rheumatoid Arthritis: Anti-citrullinated peptide antibody (ACPA) test and Rheumatoid factor (RF) test. Combined testing could be used for a more accurate prediction [113]. The disease could progress systemically with extra-articular manifestations (EAM) in the lungs, heart or vascular tissue. Of these manifestations, lung disease is a major contributor to morbidity and mortality [114].

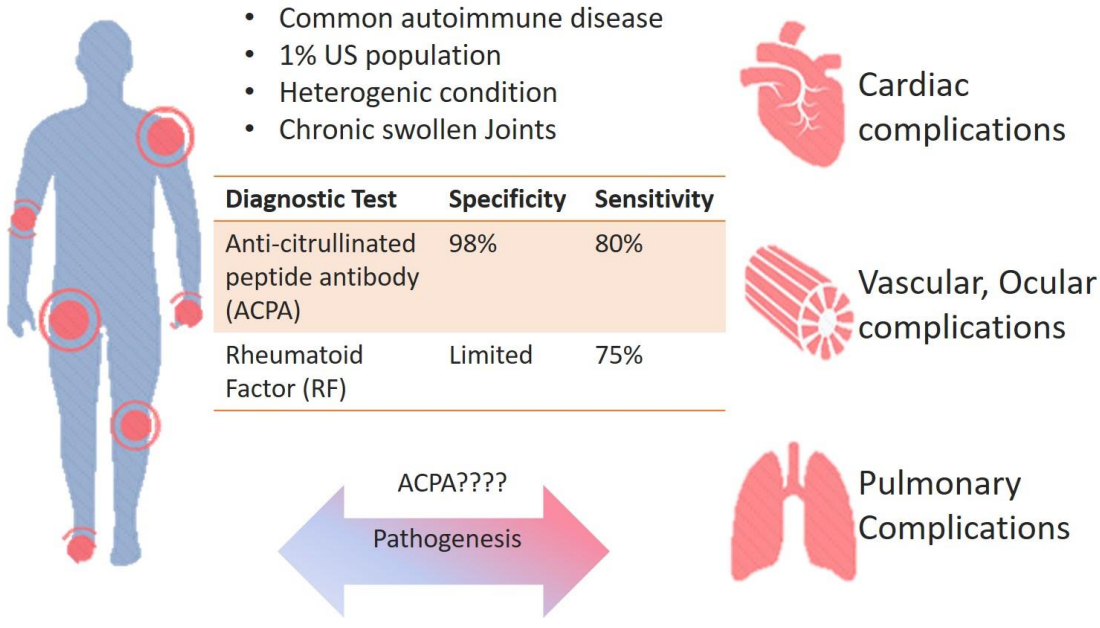


Figure 3.1 Extra-articular manifestations in Rheumatoid Arthritis

Smoking is a well-known risk factor involved in the development of RA [92]. Breathing has the lungs sampling the environment constantly. Thus lungs may be the initial site of disease development [115]. This is also supported by the fact that 40-67% of RA patients have some degree of lung involvement [116]. Lung involvement in RA patients is associated with the SNPs in HLADRB1 (HLA class II histocompatibility antigen) and PAD14 (enzyme that catalyzes citrullination) alleles [117,118]. Particularly, patients with high titers of ACPA have a higher chance of extra-articular manifestations in the lung [114]. In this chapter, we will focus on Interstitial lung disease (ILD) - for developing which, RA patients suffer an increased risk compared to the general population. RA-ILD is an important and early feature of RA and has a poor prognosis [119]. Further preventing the effects of ILD could reduce the mortality of RA patients by ~13% [120].

Drugs used for treating RA—in particular, methotrexate and the tumor necrosis factor-alpha inhibitors—have been associated with RA-ILD in numerous case reports and case series, although it is difficult to distinguish association from causality [121]. An association between RA and lung cancer has also been described, speculating the effects of cytotoxic drugs used for treatment of RA [122]. Currently, High resolution CT Scan (HRCT) is the method for definite diagnosis of ILD associated with RA. There is an urgent need for serological biomarkers for less invasive diagnosis of RA-ILD so that appropriate treatment can be instituted.

Predictive models incorporating biomarkers may allow us to identify individuals at high risk for developing progressive RA-ILD, thereby facilitating earlier diagnosis as well as monitoring of disease progression and response to therapy. A limited number of RA-ILD biomarkers have been formally examined. These include RF, Krebs von den Lungen-6 (KL-6), and ACPA (including ACPA [HSP90] antibodies) [38,123,124]. Other potential biomarkers include levels of several chemokines and cytokines. Since patients with high titers of ACPA have a higher chance of ILD, we wish to screen the target protein autoantigens of Anti-citrullinated peptide antibodies in RA-ILD.

There are two evolving paradigms for RA pathogenesis. According to the earlier model, the autoimmune disease begins in the joints and then spreads to other organs such as heart or lungs (Figure 3.2). For example: a minor trauma or infection could activate fibroblasts and cause inflammation in the joints. In patients with genetic pre-disposition, the inflammation is chronic because of a mis-regulated immune response. Cytokine storm is a predominant feature of the RA pathogenesis which could lead to the transfer of pro-inflammatory cells and cytokines into various organs such as heart or lungs. This could lead to extra-articular manifestations. While this model may explain some types of extra-articular manifestations, the involvement of environmental factors in disease development is less clear in this model.

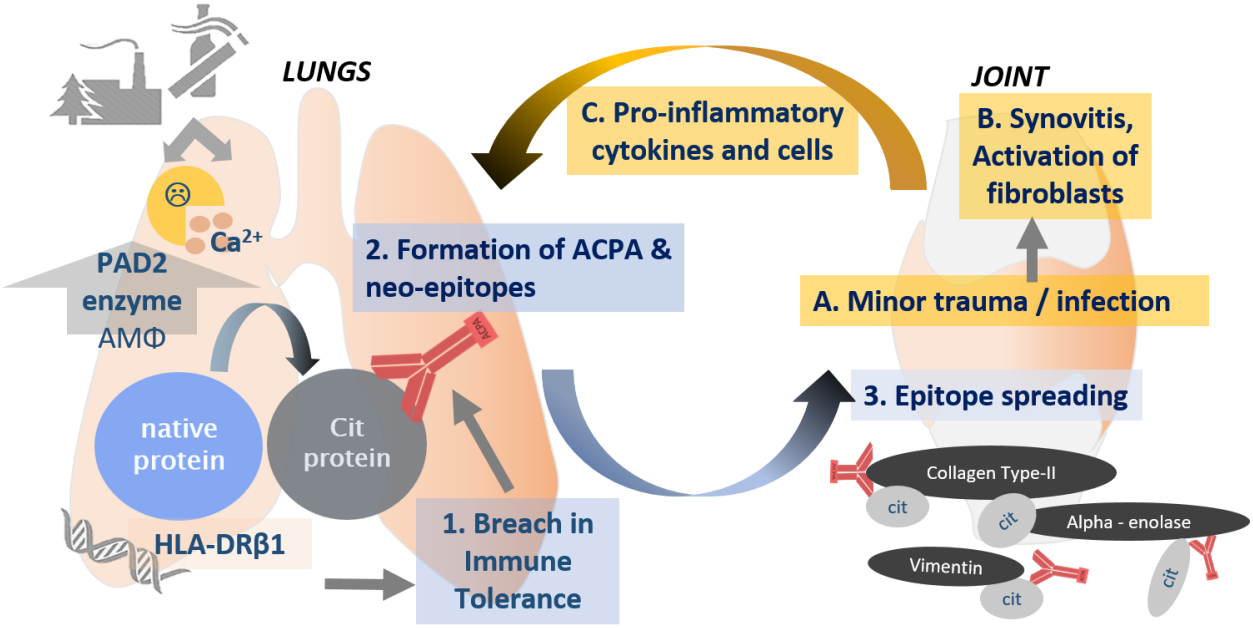


Figure 3.2 Two evolving paradigms for RA-ILD pathogenesis.

Joints-to-lungs model: Steps in yellow numbered A, B, C show the outline of disease progression from joint to lungs in RA-ILD. Lungs-to-joints model: Steps in blue numbered 1,2,3 show the outline of diseases progression from lungs to jins in RA-ILD.

In the newer model of RA pathogenesis, immune tolerance is breached in the lungs and then spreads to the joints [125]. While 98% of Anti-Citrullinated peptide antibodies (ACPA)-positive patients have RA, this model is supported by the key observation that a sub-

group of ACPA positive individuals have lung disease in the absence of articular manifestations[114,126–129]. Epidemiologic studies have demonstrated associations between smoking, lung disease, protein citrullination (a post-translational modification), possession of the shared epitope, and development of RA [114]. Smoking induces an increase in intracellular Peptidyl arginine deiminase 2 (PAD2) activity in the alveolar compartment of the lungs [93] (Figure 3.2). This causes an increase in the citrullination of proteins. In patients with genetic predisposition, citrullination in lungs create cryptic/neo-epitopes that promote the breakdown of tolerance as well as subsequent autoimmune responses against citrullinated proteins in relevant target tissues. Consistent with this model, immunohistochemical studies have demonstrated the presence of citrullinated proteins in lung explant tissue derived from patients with RA-ILD [130].

PAD2 activity requires high conc. of Ca2+ which could be provided by the induction of apoptosis. Alternatively, PAD2 could be released into the extracellular matrix. Active PAD2 catalyzes the conversion of arginine to citrulline, called citrullination/deimination of proteins non-specifically (we will refer to this modification as deimination from now). These modified intracellular proteins may then be presented to T- helper cells as a result of apoptosis or be released outside the cell - due to other forms of cell death such as necrosis or NETosis. Moreover, SNPs in HLADRB1 (HLA class II histocompatibility antigen) have been associated with increased affinity for deiminated peptides. HLA Class II molecules present antigens to T-helper cells that activate B-cells in the production of antibodies. Abundance of deiminated proteins and their affinity to HLADRB1 explains a link to activation of deiminated antigen-specific T helper cells [118]. Since T-cells are not used to "seeing" deiminated peptides during their development in the thymus, this event could result in a breach of immune tolerance locally (Figure 3.2). Once deiminated autoantigen specific B cells are activated, ACPA could be formed in the lungs. These autoantibodies and autoantibody generating lymphocytes can then circulate through the body and act as sentinels waiting for a second insult such as trauma, injury or bacterial

infection, leading to "epitope spreading" and an expanded repertoire of ACPA capable of recognizing different proteins.

There are a 2 major sub-types of ILD associated with RA: non-UIP (nonspecific/cellular interstitial pneumonia) and UIP (Usual interstitial pneumonia).

RA-ILD with a non-UIP (nonspecific/cellular interstitial pneumonia) pattern may come about when an immune response against citrullinated peptides taking place in another site (e.g. the joints) subsequently affects the lungs [131]. This is evidenced by the infiltration of cells into the lungs from outside the organ. This phenomenon could be explained by the earlier model of RA-ILD progression from joints to lungs. Non-UIP RA-ILD is responsive to immunosuppression if early treatment is provided.

Whereas, RA-ILD with a UIP (Usual interstitial pneumonia) pattern may represent a disease process in which an immune response against citrullinated proteins in lungs may then promote an articular disease indicative of RA [131]. This sub-type could be explained by the later model of RA-ILD progression from lungs to joints. Of the two sub-types, UIP is the most common form of ILD associated with RA [132]. UIP RA-ILD has poor prognosis and may require lung transplantation [131].

Currently, there are no tests for early detection of RA-ILD. Typically, tests involving HRCT scan are performed when the disease has already progressed and is causing respiratory problems. Availability of serological biomarkers will improve early detection and facilitate appropriate monitoring/treatment for patients who are at high risk for developing progressive RA-ILD. Currently testing for ACPA is a robust diagnostic test for RA; discovering specific autoantigen targets of ACPA that are linked to ILD will be key to establishing specific predictive biomarkers for lung complications in RA.

From our collaborator Dr. Dana Ascherman MD (Univ. of Miami), we have access to serum samples from RA/RA-ILD patients. We hypothesize that we will find unique autoantigen targets in RA patients with ILD relative to those without lung disease. This will also help us understand the complex mechanisms of RA-ILD pathogenesis. To do

this, we are using affinity proteomics to identify the differential targets of ACPA in RA with and without ILD. Based on differential immunoprecipitation of proteins from cell extracts subjected to in vitro deimination, this approach enables us to isolate deiminated proteins preferentially recognized by sera from patients with RA-ILD. Studying deimination requires appropriate methods such as 2D-DIGE and mass spectrometry[133]. Thus, we use our enhanced AMIDA using Biotin-CDM followed by 2D-DIGE and Mass spectrometric analysis of electrophoretically separated protein spots to identify preferential antigen targets of ACPA in RA-ILD.

3.1 Method Optimization

I outline the various steps used for the isolation of deiminated autoantigens from HeLa lysate using pooled patient sera in Figure 3.3. As mentioned in Chapter 2, I will further describe optimizations we performed and the choice of materials used.

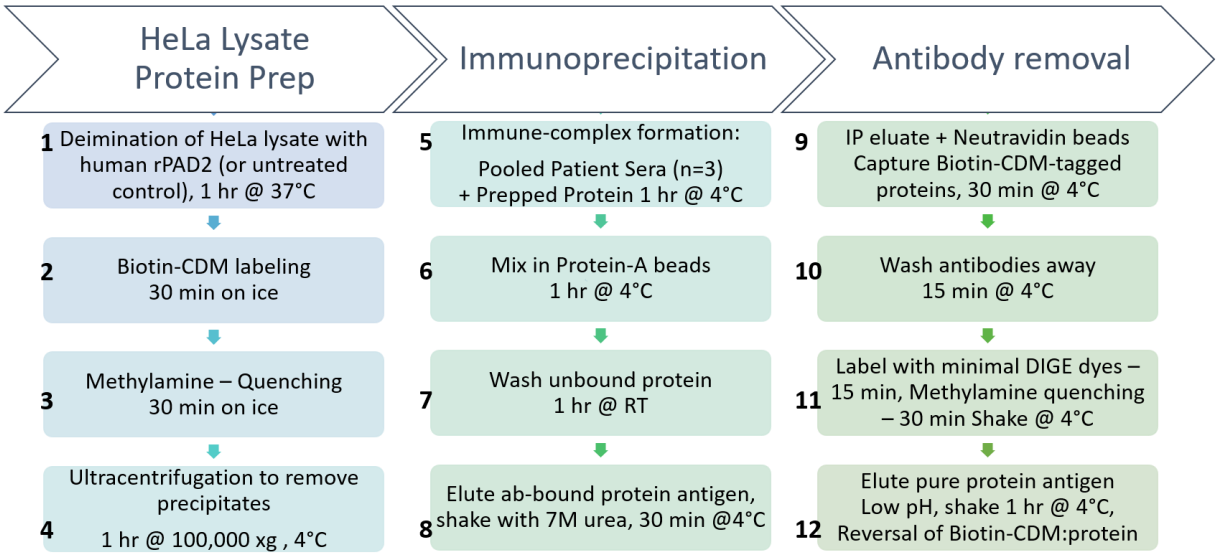


Figure 3.3 Antibody mediated isolation of antigen from *in vitro* deiminated-HeLa lysate using pooled patient sera.

HeLa Lysate Protein Prep includes Deimination, Biotin-CDM labeling, Methylamine-Quenching & Ultracentrifugation. Immunoprecipitation involves patient sera + lysate mixing, addition of Protein-A beads, followed by washing unbound protein and elution of ab-bound antigens. Antibody removal involves capturing the immunoprecipitated proteins on Neutravidin beads and washing away the antibodies. Then, the bound proteins are labeled with Fluorophores Cy3 or Cy5. Finally, the captured proteins are released by lowering the pH. These samples are now ready to be run on 2D-DIGE and mass spectrometric analysis.

Three major changes we made to the protocol are: An ultracentrifugation step was added to reduce background proteins. Free antibody approach was followed for immunoprecipitation to enrich for antigens in low abundance or have weak affinity to patient antibodies. CyDye labeling of proteins were done on neutravidin beads to increase the visibility of antigens in low abundance. I will now describe the optimizations we did, to arrive at this refined protocol.

i. Biotin-CDM labeling & Quenching of HeLa lysate

Target proteins are typically derived from commonly used, tissue culture cell lines, which can be grown in abundance, or from tissue samples. We did not have access to tissue samples or primary cell cultures from patient lungs. HeLa cell lysate was chosen as a source of autoantigens. As we will see soon, our choice to use HeLa cell lysate is also justified by the fact that we are generating disease- specific modification of proteins *in vitro*. To validate the data from this study, future experiments will need to be performed in primary cell cultures from patient lung tissue. HeLa cell lysates were extracted using lysis method-2 described in Chapter *2 (See (i) on page 29).* Biotin-CDM labeling was done using the mass-fold coupling ratio of 3:1 (Biotin-CDM: protein) as described previously. To prevent Biotinylation of antibodies, any remaining active CDM was quenched using 10-fold molar excess of methylamine (primary amines) over Biotin-CDM. (See Appendix II for a detailed protocol).

ii. Deimination

PAD is an enzyme that catalyzes the conversion of arginine to citrulline in the presence of high concentrations of calcium as seen in Figure 3.4. Several tissue–specific PAD isoforms are implicated in human disease and health (Figure 3.4) [134]. For example, normal functioning of PAD in the skin is essential for deimination of filaggrin. Deiminated filaggrin dissociates from Keratin and is degraded into peptides. These peptides help keep the skin moisturized [134]. PAD6 is implicated in infertility in women. As a part of

normal immune response against pathogens, PAD4 deiminates histones in neutrophils, thus releasing the neutrophil's chromosomes to form traps called NETs. PAD2 and PAD4 are the most likely candidate PAD isotypes for the deimination of synovial proteins in RA [135].

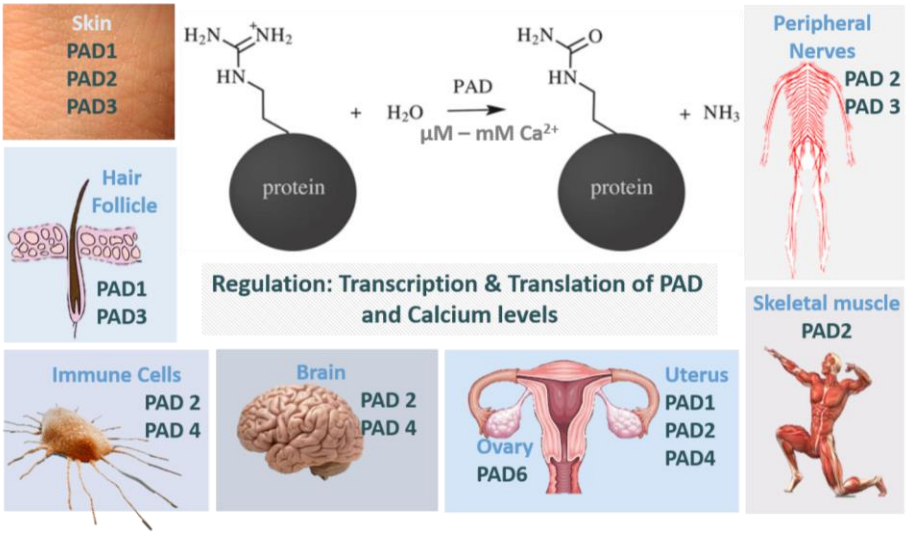


Figure 3.4 Tissue-specific PAD enzyme isoforms.

PAD2 has been shown to be upregulated in the lungs as an effect of smoking [93]. It is important to note that, while deimination is a common post translational modification throughout the body, the formation of antibodies against deiminated proteins occurs only in patients with genetic predisposition (Rheumatoid Arthritis & some cases of cancer) [136].

As seen in Figure 3.5.A, we tested human recombinant PAD2 & PAD4 enzymes (Cayman chemical) for *in vitro* deimination of fibrinogen (a typical substrate for testing PAD enzymes) in the presence of 5 mM $CaCl_2$ at 30 minutes and 1-hour incubation time. Since divalent calcium ions are required for the activity of PAD, the reaction was stopped using 5 mM EDTA as a chelating agent to sequester the divalent calcium ions. (See Appendix II for a detailed protocol). Both PAD2 & PAD4 generated in vitro deiminated proteins to a similar extent. We then treated HeLa lysate with PAD4 for 1 hour and labeled untreated lysates with Cy3 and PAD4-treated lysates with Cy5. Treated and untreated lysates were

mixed and run on a 2DE gel Figure 3.5.B. . Protein spots in green are untreated and those in red are PAD4 treated; spots in yellow are common between the two conditions.

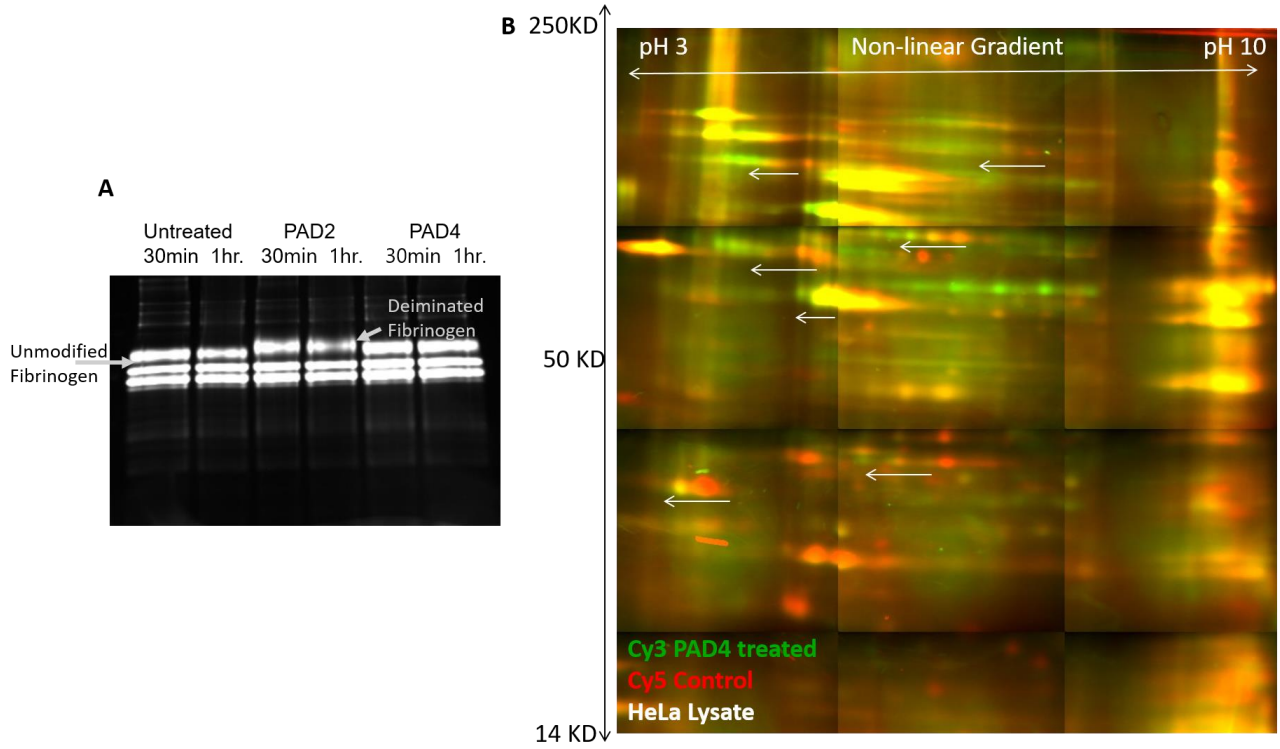


Figure 3.5 In vitro deimination using human rPAD2 and rPAD4.

A. Cy3-Fibrinogen was deiminated *in vitro,* using human recombinant PAD2 and PAD4 enzymes for 30 minutes and 1 hour. When deiminated, there is a visible shift in electrophoretic mobility of one of the fibrinogen subunits in a 12% SDS gel. B. HeLa lysate was treated with PAD4 for 1 hour and resulted in a train of spots in a 2DE gel. Protein spots in red are untreated and those in green are PAD4 treated; spots in yellow are common between the two conditions. There is an overall shift of red to green from basic to acidic pH. As discussed before, when deiminated using PAD enzymes, arginines on proteins are converted to citrullines. Proteins lose positive charges that were on arginines. This causes an overall shift in the isoelectric point of treated proteins from basic end towards the acidic end, as can be seen by the white arrows on the image.

In Figure 3.5.B, there is an overall shift of green to red from basic to acidic pH. As discussed before, when deiminated using PAD enzymes, arginines on proteins are converted to citrullines. Proteins lose positive charges that were on arginines. This causes an overall shift in the isoelectric point of treated proteins from basic end towards the acidic end, as can be seen by the white arrows on the image.

Therefore as seen in both panels of Figure 3.5, depending on protein composition (prevalence of arginine) there could be an indirect electrophoretic mobility shift accompanying the shift in isoelectric point of proteins.

iii. Ultracentrifugation

During our preliminary experiments, we observed a lot of protein precipitation after deimination. As seen in Figure 3.6, protein precipitation is a function of incubation duration at 37 °C. Incubation at 37 °C is necessary for *in vitro* deimination.

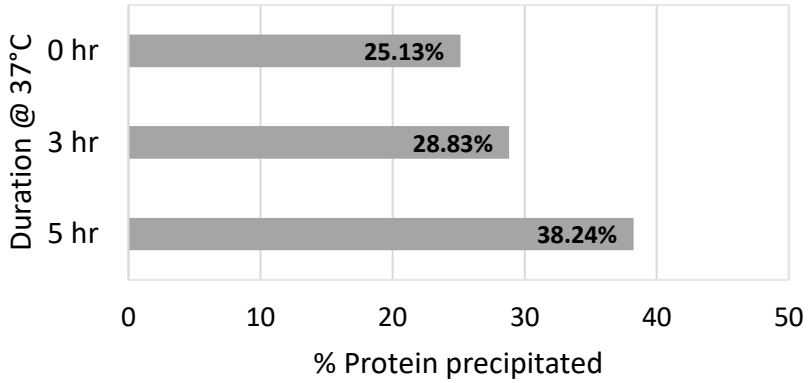


Figure 3.6 Protein precipitation as a function of incubation time at 37 °C

Because we are performing deimination *in vitro*, chances are the reaction is not as controlled as it is inside a cell. As seen in Figure 3.5.B, we can expect hyper-deimination of proteins, that is modification on variable number of arginines. This could contribute to protein precipitation as well. Different ratios of Biotin-CDM labeling did not cause any protein precipitation (data not shown).

We solved this problem by removing the precipitate from the reactions. To remove any insoluble complexes before we perform immunoprecipitation, ultracentrifugation at 100,000 x g was done. This step was key to removing background proteins while performing IP on deiminated HeLa lysate. Typical AMIDA experiments use 5 mg protein [35]. To compensate for the loss of protein in precipitation caused by the *in vitro* deimination step, we set the starting amount of protein to be 10 mg.

iv. Immunoprecipitation & Antibody removal

As discussed in Chapter 1 & 2, immunoprecipitation using a free antibody approach was not possible using cross-linking antibodies to the beads. We decided to use a free antibody approach here to increase the likelihood of discovery of low abundance antigens or antigens that have weak affinity to patient antibodies. Patient sera and biotinylated HeLa lysate were mixed together to form immune complexes. The immune complexes were captured on Protein-A Dynabeads. The unbound proteins were washed and the antigens were eluted using denaturing buffer containing urea. The eluate was mixed with Sera mag speed bead Neutravidin and ~98% contaminating antibodies were removed as described previously in Chapter 2. (Also see Appendix II for a detailed protocol). As we discussed in Chapter 2, use of non-porous magnetic beads greatly reduced the co-elution non-specific background proteins.

v. CyDye labeling & Elution

The preferentially bound biotinylated-antigens on to circulating antibodies from Patient RA and Patient RA-ILD pools were eluted from protein-A beads and captured by Neutravidin beads. Non-biotinylated proteins were then washed away. While the biotinylated proteins remained bound on Neutravidin beads, CyDyes were added for labeling. (See detailed protocol in Appendix II). This is opposite from the protocol described in Chapter 2. The reasoning is as follows: Suppose, we did fluorescence labeling at the same time as Biotinylation, we would end up using 100 times the CyDye that we normally use for a regular DIGE experiment in the lab. Here, we start with 10 mg protein from HeLa lysate per condition whereas typical DIGE experiments use a starting amount of 100 µg protein. At the end of immunoprecipitation and antibody removal, we were left with protein amounts less than 50 µg. In order to increase the visualization of low amount of proteins, we labeled the target proteins while they were still captured on the Neutravidin beads. Then, pH is lowered for CDM hydrolysis and elution of target proteins.

vi. 2D-DIGE

The samples were then run on 2D-DIGE using standard protocol. We used the mini-2DE gels. In the mini-2DE gel setup, there is considerably less protein loss. Therefore, it is better for visualization of low abundance proteins. (See Appendix III for detailed protocol).

vii. Mass spectrometry

The gel plugs of detected proteins were cut from the gel. In-gel digestion of the protein spots was done with trypsin using a standard manual extraction protocol [137]. Samples were sent to Penn State Proteomics Facility for mass spectrometric analysis. MS spectra taken from 30-minute gradient from an Eksigent NanoLC-Ultra-2D Plus and Eksigent LC through a 200 µm x 0.5 mm Chrom XP C18-CL 3 µm 120 Å and elution through a 75 µm x 15 cm C18-CL 3 µm 120 Å both Eksigent LC Column. ABSciex 5600 TripleTOF settings used: Parent scan acquired for 250 msec, then up to 50 MS/MS spectra acquired over 2.5 seconds for a total cycle time of 2.8 seconds. Gas 1(Nitrogen)= 7, Gas 3(Nitrogen)= 25.

The peptide sequences obtained were analyzed against the Ref Human database. The basic scoring system in Protein-Pilot *(*Sciex*)* analysis was used. Based on hundreds of thousands of Mass spectra and associated peptide and protein IDs, is called the Unique Score, assigned to each Protein identified based on the number and strength of the associated peptide IDs. Any protein with an Unique Score above 1.3 is confidently identified at better than 95% confidence (analogous to a p-value <0.05); an Unique Score above 2.0 is confidently ID'd at better than 99% confidence (analogous to p<0.01); above 3.0 is confidently ID'd at better than 99.9% confidence (analogous to p<0.001). Therefore, anything above 1.3 is considered confident.

3.2 Screening for RA/RA-ILD antigens

Patient samples used were obtained from the replication cohort previously described by Chen et al [138]. HRCT of the chest and pulmonary function testing tests (PFTs) were

performed to assess radiographic and functional abnormalities indicative of ILD. All patient samples were scored based on the severity of ILD: where 0 = no ILD, 1 = indeterminate ILD, 2 = mild/moderate ILD, and 3 = severe ILD, 5 = Definite ILD but not sub-classified as 2 or 3).

We pooled a group of patient sera based on the availability of patient sera samples and severity of ILD in that cohort. Using pooled antibodies can increase the chance of discovering a common antigen [139]. Although pooling of patient sera is limited by inter-individual variability, we chose this approach as a first step in the process of autoantigen screening to conserve resources. Future validations will be performed using individual patient samples.

Due to the limited availability of each category of patient sample: In each pool, we chose two patient samples with mild/moderate ILD (Score = 2) and one patient sample with definite ILD (Score = 5). Thus, we performed the study with a total of N = 18 samples with 9 in RA & RA-ILD category each.

As we discussed before, there are sub-types of ILD with different histopathologies such as UIP and non-UIP. These sub-types were not distinguished in this cohort.

#	RA (reference)	RA-ILD (reference)
Pool# 1	16, 24, 36	14, 83, 94
Pool# 2	22, 90, 54	27, 76, 49
Pool# 3	23, 37, 55	70, 29, 25

Table 3.1 List of patient samples used.

The reference numbers denoted here represent the patient identification used in the cohort.

Patient Pool 1

Patient pools were made according to Table 3.1. Immunoprecipitation and antibody removal were performed as described previously using pool 1. For Pool 1, the samples that were immunoprecipitated with RA patient sera were labeled with Cy3 (green) and

those that were immunoprecipitated with RA-ILD were labeled with Cy5 (red) fluorescent dyes. The fluorescently labeled proteins were then released in a low pH buffer and run on DIGE. As seen in Figure 3.7, we obtained a fingerprint of the antigens preferentially immunoprecipitated using RA Vs RA-ILD patient sera.

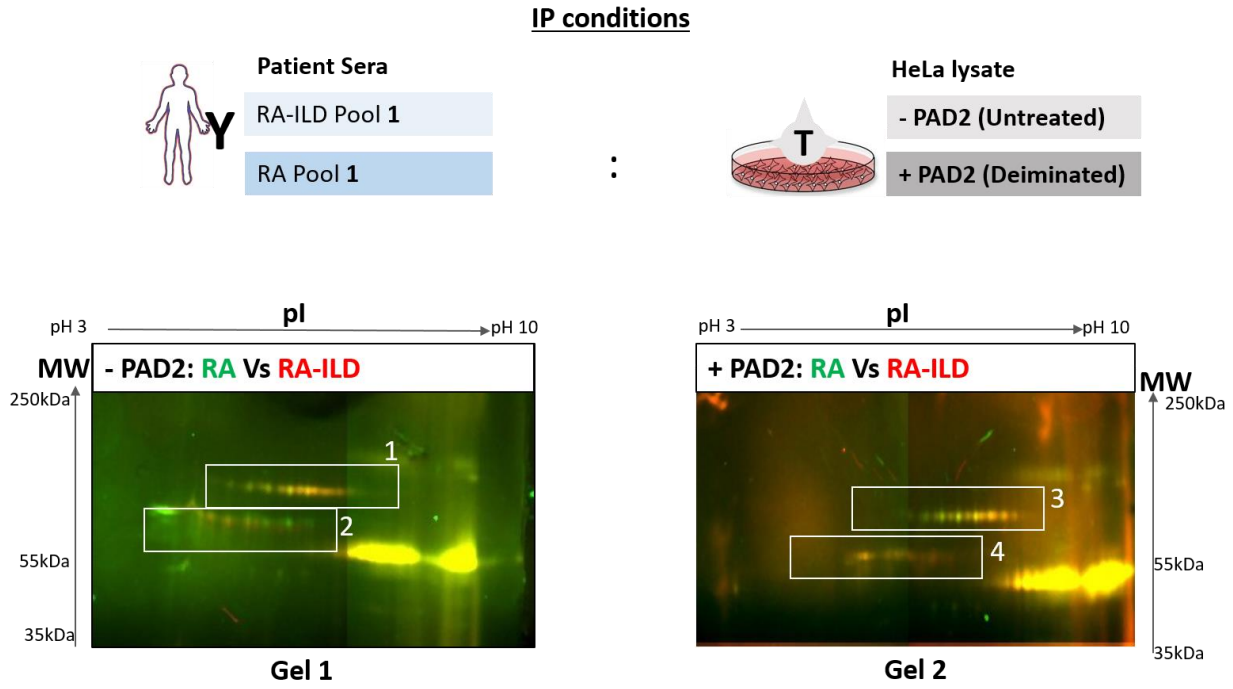


Figure 3.7 Patient Pool 1: Difference gel electrophoresis of immunoprecipitated antigens.

Gel 1 shows the comparison between preferentially immunoprecipitated HeLa lysate between RA & RA-ILD patient pool 1. Gel 2 shows the comparison between preferentially immunoprecipitated *in vitro* deiminated HeLa lysate between RA & RA-ILD patient pool 1. In both images, HeLa proteins in green were immunoprecipitated by RA sera antibodies and those in red were immunoprecipitated by RA-ILD sera antibodies. Protein spots in yellow are heavy chain immunoglobulins (antibodies). Regions of interested are numbered 1-4.

We observed that the untreated and PAD-treated DIGE gels looked very similar. From here, I will refer to the series of horizontally separated protein spots as charge train. Two charge trains were observed and they are numbered as regions 1-4 in Figure 3.7. In both regions 1 & 3, the proteins preferentially immunoprecipitated with RA are on the acidic

side and that of RA-ILD are on the basic side of the charge trains i.e. red spots on the right and green spots on the left. This shows that PAD treatment did not change the pattern of region 1. However, regions 2 & 4 look very different with PAD treatment. Region 1 probably has a second post translational modification that might contribute to RA/RA-ILD preference. Once we identify the proteins, then we could further study the modifications closely. Unfortunately, characterizing the second modification is outside of the scope of this thesis, due to limited resources available.

The yellow swaths below the antigen protein spots are likely immunoglobulin heavy chain from the patient sera. As shown in Chapter 2, we were able to remove 98.5% contaminating antibodies using the Biotin-CDM – Neutravidin beads scheme. However, in this chapter, to visualize low abundance proteins, we labeled the immunoprecipitated proteins on the beads. Whereas in Chapter 2, we labeled antibodies and antigens separately at the beginning of immunoprecipitation. The key difference is, when labeling biotinylated proteins on beads, they have some of their primary amines already blocked. CyDyes are conjugated through primary amines as well. Thus, the availability of primary amines for labeling any remaining antibodies (less than 2%) vs that of biotinylated bound antigens is higher. Therefore, what we see as the yellow common proteins, are lower in abundance but show up very bright in the fluorescence images. This is confirmed by the fact that we do not see any immunoglobulin peptides during mass spectrometry analysis, which was the original goal of our Biotin-CDM effort as described in Chapter 2.

As is visible in Figure 3.7, the spots are very close to each other. We were unable to cut single spots that were either red or green i.e. preferentially immunoprecipitated with RA or RA-ILD sera. We cut each region (Regions: 1-4 in Figure 3.7) and processed for mass spectrometry. The protein ID's of each region were analyzed across technical replicates of pool 1. Statistically significant protein ID's (p value < 0.05) were short-listed based on the approximate molecular weight range they fall into (with reference to DIGE). The results are tabulated in Table 3.2. Proteins that repeated across the different gel regions are in bold. We expect a certain level of overlap in the protein identification of the

different spots, because modified versions of the same protein could have different electrophoretic mobilities in the horizontal dimension and a shift in isoelectric point along the horizontal dimension.

Gel region	UNIPROT ID	Name	p-value	pI	MW
2	Q76B58	BMP/retinoic acid-inducible neural-specific protein 3 isoform-2	< 0.01	6.61	76211.84
2	Q08499	**cAMP-specific 3',5'-cyclic phosphodiesterase 4D isoform PDE4D5**	< 0.01	5.31	91114.89
2	P04040	**catalase**	< 0.001	6.95	59624.98
2	P05060	secretogranin-1 precursor	< 0.01	5.02	76325.89
3	V9HWI4	**lactotransferrin isoform 1 precursor**	< 0.001	8.47	76165.29
3	A0A0C4DG33	peroxisome biogenesis factor 1 isoform 2	< 0.01	6.04	136584.88
3	A6NKG5	retrotransposon-like protein 1	< 0.01	5.09	155047.51
4	P04040	**catalase**	< 0.01	6.95	59624.98
4	P02788	**lactotransferrin isoform 1 precursor**	< 0.01	8.47	76165.29
4	Q9H4E7	REVERSED differentially expressed in FDCP 6 homolog	< 0.01	5.78	73910.33
4	Q08499	**cAMP-specific 3',5'-cyclic phosphodiesterase 4D isoform PDE4D5**	< 0.01	5.31	91114.89

Table 3.2 Mass spectrometry data from Patient Pool 1 gels.

Mass spectrometry data from gel regions 2-4 are shown here. Region 1 was not identified due to limited material. Proteins in bold repeated across different gel regions, and are potential candidates for future investigations.

Patient Pool 2

The antigen fingerprints in the patient pool 2 had a very similar pattern to patient pool 1 (Figure 3.8). Here, the untreated samples that were immunoprecipitated with RA patient sera were labeled with Cy3 (green) and those that were immunoprecipitated with RA-ILD were labeled with Cy5 (red) fluorescent dyes. The colors were reversed for PAD-treated samples. Region 4 from pool 2 (RA-ILD in green) looks very similar to that of pool 1 (RA-ILD in red). We are not able to see region-1 in gel-3. Region 3 of pool 2 looks very similar to that of pool 1. Acidic side is preferred by RA and basic by RA-ILD antibodies. We will look at this closely later in this chapter (Figure 3.10).

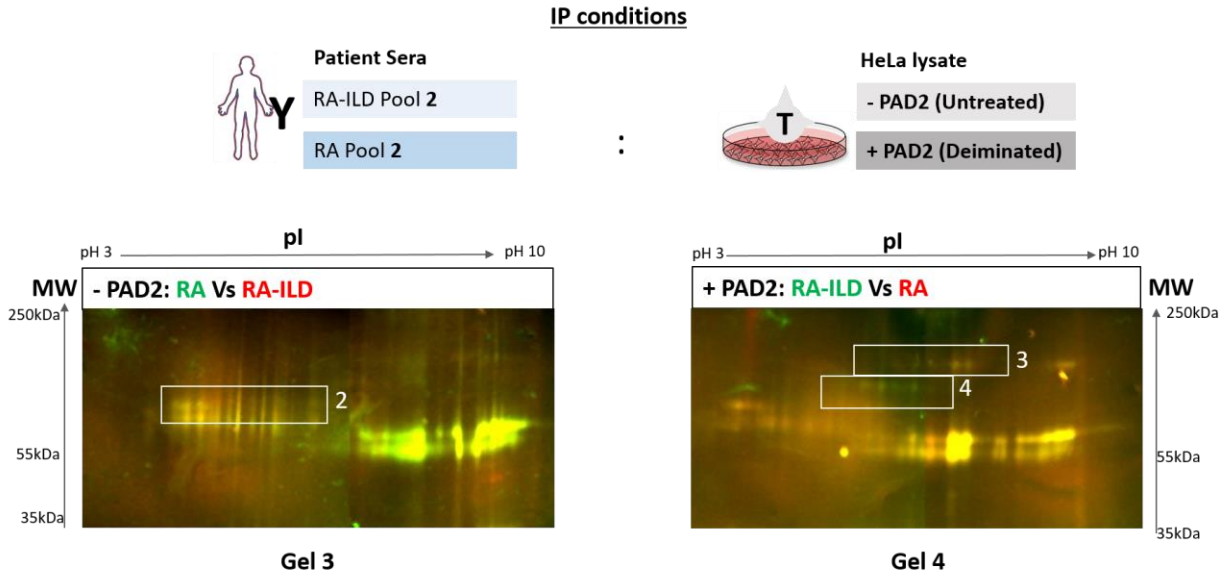


Figure 3.8 Patient Pool 2: Difference gel electrophoresis of immunoprecipitated antigens.

Gel 3 shows the comparison between preferentially immunoprecipitated HeLa lysate between RA & RA-ILD patient pool 2. Gel 4 shows the comparison between preferentially immunoprecipitated *in vitro* deiminated HeLa lysate between RA & RA-ILD patient pool 2. Regions of interested that replicated between pool 1 and pool 2 are numbered 2,3 & 4.

Gel region	UNIPROT ID	Name	p-value	pI	MW
3	Q08499	**cAMP-specific 3',5'-cyclic phosphodiesterase 4D isoform PDE4D5**	< 0.05	5.31	91114.89
4	A0A024R098, Q01433	AMP deaminase 2	<0.01	6.46	100687.86
4	P10909	PREDICTED: clusterin isoform X1	<0.01	5.89	50062.56
4	Q5TCI8, P02545	PREDICTED: lamin isoform X2	<0.01	6.55	55762.35
4	Q9UKU6	PREDICTED: thyrotropin-releasing hormone-degrading ectoenzyme isoform X2	<0.01	6.5	116999.63

Table 3.3 Mass spectrometry data from Patient Pool 2 gels.

Again, we cut each region (Regions: 1-4 in Figure 3.8) and processed for mass spectrometry. The results are tabulated in Table 3.3. Even though, the finger prints on DIGE looked similar, cAMP-specific 3',5'-cyclic phosphodiesterase 4D isoform PDE4D5 is the only protein that repeated between patient pools 1 & 2.

Patient Pool 3

Here, the untreated samples that were immunoprecipitated with RA patient sera were labeled with Cy3 (green) and those that were immunoprecipitated with RA-ILD were labeled with Cy5 (red) fluorescent dyes. The colors were reversed for PAD-treated samples.

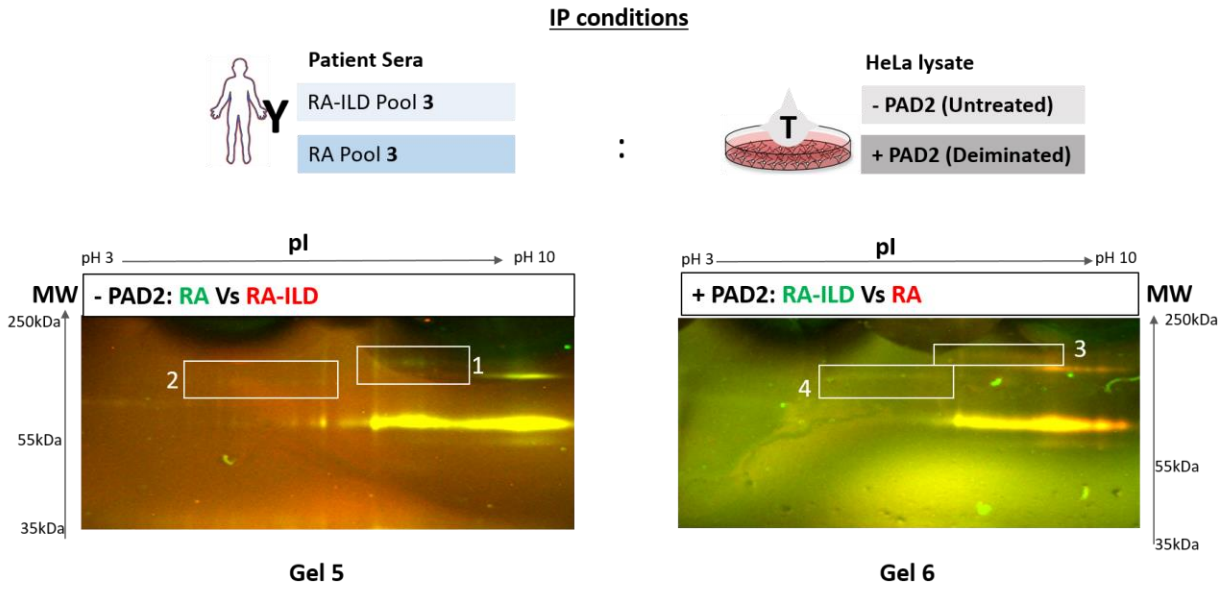


Figure 3.9 Patient Pool 3: Difference gel electrophoresis of immunoprecipitated antigens.

Gel 3 shows the comparison between preferentially immunoprecipitated HeLa lysate between RA & RA-ILD patient pool 2. Gel 4 shows the comparison between preferentially immunoprecipitated *in vitro* deiminated HeLa lysate between RA & RA-ILD patient pool 2. Regions of interested that replicated between pool 1 and pool 2 are numbered 2,3 & 4.

Gel region	UNIPROT ID	Name	p-value	pI	MW
2	O75150	PREDICTED: E3 ubiquitin-protein ligase BRE1B	<0.05	5.93	113650.4
2	P48594	PREDICTED: serpin B4	<0.05	5.86	44853.93

Table 3.4 Mass spectrometry data from Patient Pool 3 gels.

In pool 3, the spots did not resolve well and in general, the proteins were low in concentration. It is important to note that, while we are able to see common autoantigens

using our method, there is inter-individual patient variability. Antibodies from different patients may have varying affinities to the same antigen.

Nevertheless, we cut the visible 1-4 regions and did mass spectrometry analysis (Table 3.4). We could only identify region 2 and it was very different from the proteins identified previously in pools 1 & 2.

3.3 Discussion

First, we have successfully removed contaminating antibodies such that they do not interfere with mass spectrometry analysis anymore. The background proteins and non-specific binding have been greatly reduced due to the use of magnetic beads and removing insoluble precipitates using ultracentrifugation. In Chapter 2, Figure 2.11 we detected 88 protein spots with RA patient sera. In this chapter, we observed a handful number of protein spots, which are possibly true antigens and not background proteins. The CyDye labeling performed on captured target proteins help us visualize proteins in low concentrations better.

We assumed that each charge train seen in DIGE, is a result of the *in vitro* deimination performed using PAD2 treatment. Variable number of arginine were modified giving rise to proteins of different isoelectric point (pI), as we saw earlier in Figure 3.5. However, we do see a charge train also in the untreated samples, which we speculate could be an effect of another post-translational modification that we haven't characterized. It is also possible that PAD expressed in HeLa cells deiminated certain proteins expressed in HeLa cells [140]. Based on the change in only pI dimension, the second modification is probably either deimination, phosphorylation, acetylation, succinylation or a combination of these. Since HeLa cells are not physiologically relevant to Interstitial lung disease, these experiments have to be repeated in lung tissue extract. We used antibodies from patient sera. It is possible that these antibodies were present in the form of immune-complexes, already bound to their target proteins. In future studies, the immune complexes have to be disrupted before performing immunoprecipitation experiments.

When we take a closer look at regions 1 & 3 (Figure 3.10), we see a similar trend in all 3 panels i.e a "leftward" shift of given protein in DIGE. The acidic end of the charge train is generally preferred by RA antibodies and basic by RA-ILD antibodies. We initially hypothesized that antibodies from RA-ILD patients may preferentially bind to certain deiminated antigens. However, looking at the pattern here, we speculate that antibodies from RA and RA-ILD patients recognize the same antigen, but different modified versions of the same antigen. In all panels of Figure 3.10: proteins/protein isoforms that immunoprecipitated preferentially with RA-ILD, tend to focus towards the basic pH end of the charge train. A possible explanation for this observation denotes epitope spreading. According to the latter model of RA-ILD pathogenesis, autoimmune response is triggered in the lungs and then the disease spread through epitope spreading. Our results show that while antibodies in RA-ILD patients may target the protein, the antibodies in RA patients specifically target the modified form of the same target proteins.

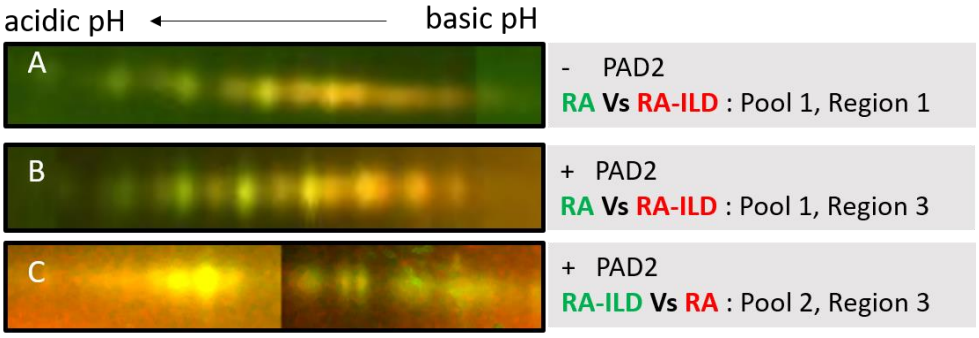


Figure 3.10 A closer look at Regions 1 & 3 from Pools 1 & 2 DIGE.

A) Shows the charge train in region 1 from Pool 1, untreated samples. B) Shows the charge train in region 3 from Pool 1, PAD2-treated samples. C) Shows the charge train in region 3 from Pool 2, PAD2-treated samples.

In Figure 3.10.B, a second pattern observed is green spots alternating with red/yellow spots. This again strengthens our hypothesis that there is a second post-translational modification involved, other than deimination. If all the modifications are in fact, deimination, then it is possible that antibodies from RA patients prefer the deiminated version of the same antigen compared to antibodies from RA-ILD patients. This is an

interesting observation that needs more exploration. Further experiments using advanced mass spectrometry will need to be performed to pinpoint the exact modification that is preferred by RA-ILD Vs RA patient samples. Alternative explanation is that the charge train is made of not one but several proteins. In any case, the proteins identified in this study can be characterized further in future studies.

We have listed the potential antigens identified in Table 3.2, Table 3.3 and Table 3.4. Of these, the most likely candidates for further investigation are Catalase and cAMP-specific 3',5'-cyclic phosphodiesterase 4D isoform PDE4D5. Increased levels of antibodies against catalase has been seen before in Rheumatoid Arthritis and systemic lupus erythematous [141]. We saw catalase twice in technical replicates of pool 1. Deimination of catalase could trigger the formation of antibodies against catalase. This could in turn play a role in the pathogenesis of RA-ILD.

cAMP-specific 3',5'-cyclic phosphodiesterase 4D isoform PDE4D5 has not been identified as an autoantigen before. We saw this protein identified in both pools 1 & 2. This protein hydrolyzes the second messenger cAMP, which is a key regulator of many important physiological processes [142]. This protein is present throughout the body especially in epithelial cells and skeletal muscle. Deiminated version of the protein could be a target of antibodies in RA-ILD.

This screen for autoantibody/autoantigen profiles has identified potential autoantigens that are targeted in both RA and RA-ILD. The next step is characterizing these proteins in lung tissue extracts. Identifying the modification sites on an individual autoantigen protein will enable us to make recombinant modified proteins. These recombinant modified proteins could then be used for biomarker validation using ELISA performed on a different RA-ILD patient cohort. The recombinant modified autoantigens could also be used as bait for co-immunoprecipitation in primary lung cell cultures exposed to smoke. This will help us understand the complex networks involved in the molecular pathogenesis of this autoimmune disease and the lung complications that develop due to smoking.

Thus, we have successfully removed contaminating antibodies and identified novel autoantigens of RA/RA-ILD using our new method.

Chapter 4 Studying mitochondrial import defect in Huntington disease

This study was done in collaboration with Dr. Sveta Yablonska and Dr. Robert Friedlander, Department of Neurosurgery, UPMC.

Huntington's disease (HD) is an autosomal dominant neurodegenerative disorder caused by expression of huntingtin protein (HTT) with an expanded polyglutamine (polyQ) stretch caused by CAG repeat expansion beyond 35 on the 5'-end of the coding reagion of HTT gene [143]. There is no effective treatment for HD which affects 30,000 people in the USA. Although the neuropathological mechanisms of HD are not clear, the mutant HTT protein is known to effect calcium regulation, decrease energetic function, lead to impaired mitochondrial protein trafficking, and disruption of mitochondrial dynamics [144–149].

The vast majority of mitochondrial proteins are encoded in the nucleus, translated on cytoplasmic ribosomes and exported to the mitochondria in an immature form carrying a N-terminal mitochondrial targeting sequence [150]. These proteins are transported though the multi-subunit mitochondrial TOM (translocase of outer membrane) and TIM (translocase of inner membrane) complexes. Previously published data from our collaborator demonstrated that, mutant HTT fragments directly interact with the TIM complex proteins and inhibit mitochondrial import of the mitochondrial matrix protein ornithine trans-carbamylase (OTC) [151]. We hypothesized that due to this interaction between mutant HTT and the TIM complex, the levels of nuclear-encoded mitochondrial proteins will be reduced in mitochondria. Alterations in the mitochondrial proteome were hypothesized to induce changes in mitochondrial proteostasis, leading to multiple cellular pathologies [152].

To test this hypothesis, we chose ST-Hdh-Q7/Q7 and -Q111/Q111 mouse striatal cell lines that express full-length wild type (Q7) and mutant HTT (Q111) respectively. Q111 is a well-established cell line model of HD derived from an HTT-knock-in murine embryo [153]. Mitochondria isolated from Q111 cells demonstrate reduction of import of OTC compared to mitochondria from wild type HTT expressing cells [151].

Several proteomic studies have been done to explore mitochondrial proteome disturbances in multiple pathological conditions [154–157]. However, none of the studies differentiated between mitochondrial and cytoplasmic protein levels. This is because, those studies were not done on isolated mitochondria from cells or tissues expressing mutant HTT or mutant HTT fragments. To date, all our knowledge about mitochondrial proteome damage in HD came from studies performed on total brain lysates of HD mice models [158–160] and postmortem brain of HD patients [161]. Proteomic analysis of such a complex biological mixture significantly limits the number of potentially identified mitochondrial proteins which may be altered in this neuropathology.

In order to assess the downstream effects of the mutant HD associated mitochondrial protein defect [151], we performed Two dimensional Difference gel electrophoretic (2D-DIGE) analysis of Q7 mitochondrial protein lysate Vs Q111 mitochondrial protein lysate from mouse cell lines. Here, I will describe the results from two different mitochondrial isolation methods that were used to assess mitochondrial proteome changes.

4.1. Materials and Methods

i. Cell Culture

ST-Hdh-Q111/Q111 and ST-Hdh-Q7/Q7 cells were obtained from Marcy McDonald {reference}. Cells were cultured in DMEM media supplemented with 5% FBS, 1% Sodium pyruvate at 33˚C in the presence of 5% CO2.

ii. Mitochondria fractionation using Percoll gradient

Mitochondria were isolated from Q7 and Q111 cell lines as described previously using Percoll gradient centrifugation [162].

iii. Mitochondria isolation using anti-TOM20 antibody

An Alternative scheme for isolating mitochondria, is the use of an antibody against a mitochondrial outer membrane protein to pull down mitochondria from cell lysate. Immediately prior to each experiment, cells were collected and subjected to the

Mitochondria Isolation MACS Kit (Miltenyi Biotec) according to the manufacturer protocol. Briefly, cells were homogenized in ice-cold lysis buffer (from MACS kit), in ratio 1 ml Lysis buffer per $10x10^6$ cells and incubated with magnetic nano-beads conjugated with anti-TOM20 antibody for 1 hr in a cold room. The suspension was applied to a magnetic stand which retains bound mitochondria. Washes were performed and mitochondria were eluted with mitochondria isolation buffer provided in the kit.

iv. 2D-DIGE

Mitochondria fractions were re-suspended in lysis buffer containing 7 M Urea, 2 M Thiourea, 10 mM HEPES pH 8.0, 10 mM DTT and 4% CHAPS. The mitochondrial fractions in amount of 100 µg of protein from Q7 were labeled with Cy3-NHS mono-reactive dye (GE Healthcare), from Q111 – with Cy5-NHS (GE Healthcare), and then combined. For isoelectric focusing (1st dimension separation), mixed samples were applied to 18 cm IPG strips pH 3-10NL (GE Healthcare). IPG strip were transferred onto a 12% SDS-PAGE using in-gel equilibration [107]. Spots with relative high intensity of Cy3 or Cy5 fluorescence were excised using our homemade fluorescence gel imager/spot picker platform [108,109].

v. Mass spectrometry analysis

Excised gel plugs from 2D-DIGE were submitted for LC-MS/MS at University of Pittsburgh. Gel spots were reductively alkylated with DTT & IAA and digested with trypsin according to standard protocol [137]. Peptides were analyzed by nanoreverse phase HPLC interfaced with an LTQ XL linear ion trap mass spectrometer (Thermo Fisher Scientific). The tandem mass spectra (MS/MS) obtained, were analyzed by the MASCOT (Matrix Science) search engine against mouse protein database. Identified peptides and proteins were further statistically validated with the Scaffold software. Only proteins with high confidence identifications were considered: 1) Protein identification probability of 99% or above, 2) Peptide identification probability of 95% or above and 3) minimum

two peptides. The proteins were also cross-referenced against DIGE gel images to verify their approximate molecular weight (MW) and isoelectric point (pI).

vi. Western Blotting

2D-DIGE Gel was assembled with PVDF-FL membrane into a transferring sandwich and was transferred in 1X TG buffer with 10% methanol at a voltage of 45 V at 4 °C overnight. After transfer, the PVDF-FL membrane was blocked in Odyssey Blocking Buffer (Li-Cor Biosciences). The PVDF-FL membrane was immunostained with specific primary antibody (Abcam) for mitochondrial proteins. After that, the membrane was probed with infrared labeled secondary antibodies (Li-Cor Biosciences). Following several washes in 1X PBST buffer, membrane was scanned on Odyssey CLx Li-Cor.

4.2. Results

Mitochondria from Q7 and Q111 cell lines were fractionated using the Percoll gradient centrifugation. They were labeled with CyDyes and run together on a 2DE gel (Figure 4.1). Protein spots in green are from wild-type and those in red are from mutant HTT cell lines.

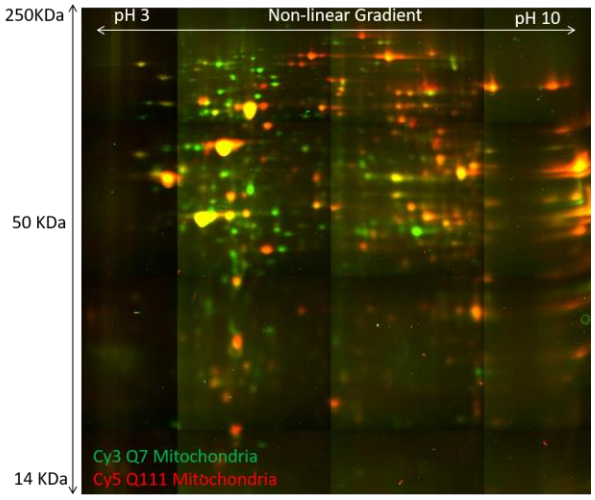


Figure 4.1 2D-DIGE analysis of mitochondria from Q7 Vs Q111, fractionated using Percoll gradient centrifugation.

Protein spots in green are from wild-type and those in red are from mutant HTT cell-lines. Protein spots in yellow overlap between the two conditions.

The gel is very well resolved, but there are very few overlapping protein spots between the two conditions (protein spots in yellow). Instead the majority of proteins from Q7 appeared at the acidic side of the gel. This was a very unusual observation. It turned out, these results were an artifact caused due to the fractionation method used for isolating mitochondria. Mitochondria in mutant HTT cell lines show a different phenotype from that of the wild-type. Mitochondria in HTT mutant cell line are in general, smaller in size compared to that of the wild-type. Mutant Huntingtin disrupts mitochondrial import of proteins like OTC [151]. In addition, there is nothing in the literature about differential import of acidic vs basic proteins. We suspect, using a centrifugation-based method biased for mitochondria of comparable in size and density to that of wild-type. This caused a bias in the DIGE comparison of mitochondrial proteomes between mutant and wild-type conditions.

Therefore, we chose a mitochondrial isolation method that is not dependent on the size or density of the organelle. We addressed this challenge by choosing an antibody-based method to affinity purify mitochondria. Anti-TOM20 antibody bound to magnetic beads were used for the isolation of mutant and wild-type mitochondria. Using this method, we isolated mitochondria from Q7 and Q111 cells and ran DIGE comparisons (Figure 4.2).

We observed several changes between Q7 and Q111. Many of these proteins are either red or green, meaning they are present in abundance in one condition and not the other. In order to merge the Cy3 and Cy5 channels, we needed a reference protein that would remain unchanged between wild-type and mutant. For this purpose, we chose mitochondrial Cytochrome C oxidase 1 (mt cox1) as our internal standard. This protein is encoded in the mitochondria and synthesized inside the mitochondria. Therefore, we assumed that any impairment to the mitochondria caused due to mutant HTT, shouldn't affect the protein concentration of mt cox1 in mitochondria. The gene expression of mt cox 1 could be indirectly affected by mutant HTT, this assumption will be further explored in our collaborators' lab.

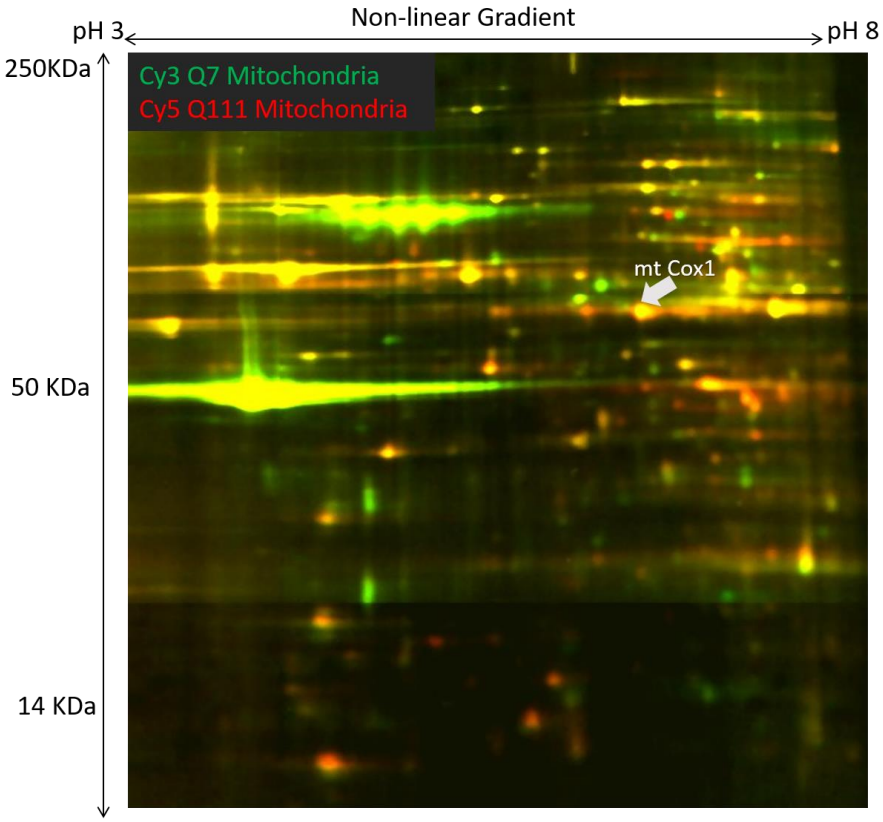


Figure 4.2 DIGE comparison of Q7 Vs Q111 mitochondria isolated using anti-TOM22 antibody beads.

Images in Cy3 and Cy5 channels were merged so that mt cox1 protein spots are yellow (grayscale adjustment), resulting in this DIGE comparison. Protein spots in green are from Q7 mitochondria and those in red are from Q111 mitochondria. Protein spots in yellow are common to both Q7 and Q111.

To identify the mt cox1 protein spot, we performed a 2D-western blot using antibody against mt cox1 (Figure 4.3). Cy5-Q7 was run on a 2DE gel and then transferred to a PVDF-FL membrane for western blotting. The 2° antibody signal was captured using fluorescence as well. Therefore, we merged Cy5 image and antibody fluorescence (Infrared) to detect the position of mt cox1 on the 2DE gel. We compared this against the 2D-DIGE gel and identified the position of mt cox1. It correlated with the molecular weight (~56 KDa) and isoelectric point (6.1) of the protein.

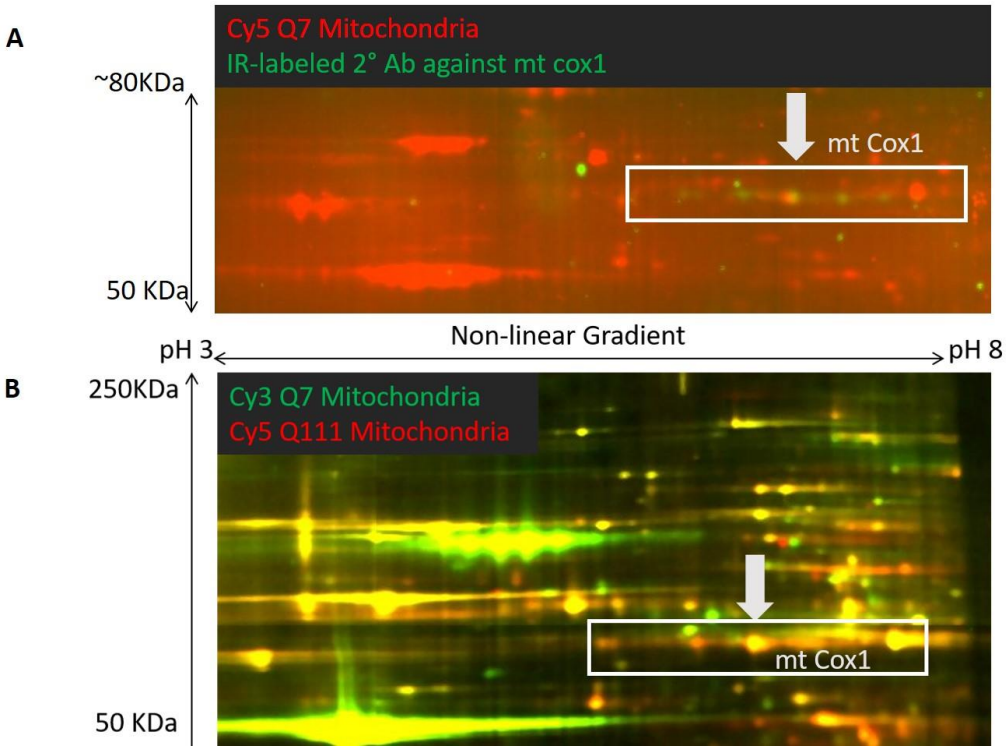


Figure 4.3 2DE Western blot to identify reference protein mt Cox1.

(A) Shows the merged image of the western blotting membrane in Cy5 (Q7 mitochondria) and IR (fluorescent 2° antibody) channels, when probed for mitochondrial Cox1. The bright green speckles on the membrane are artifacts from antibody clumping. The overlapping protein spots in yellow are mt Cox1. (B) Using the surrounding proteins as landmarks, we located the position of mt cox1 in the 2D-DIGE gel.

Using mt cox1 as the reference protein, we visually normalized the Cy3 and Cy5 channels in the DIGE gel shown before (Figure 4.2). Proteins of interest that were present only in Q7 (green) or in Q111 (red), were cut and processed for identification using mass spectrometry.

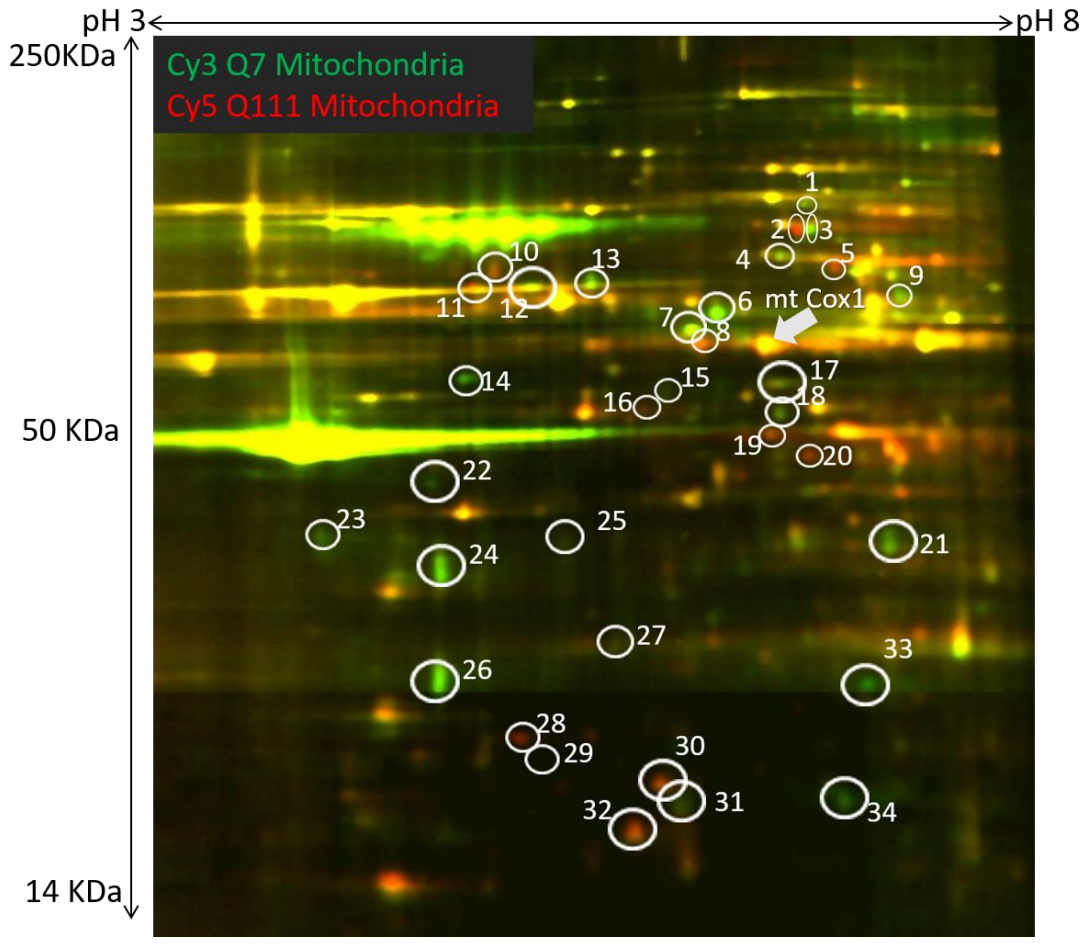


Figure 4.4 Proteins of interest from DIGE comparison: Q7 Vs Q111 mitochondria.

Protein spots of interested are circled and labeled 1-34 here. mt Cox1 is indicated using a white arrow.

Five biological replicates of Q7 Vs Q111 comparison were performed. Each DIGE experiment was performed in duplicate, with reciprocal labeling so as to avoid dye dependent changes [163]. Several changes between Q7 and Q111 mitochondrial proteomes were observed. Of those, we chose the changes that were reproduced in at least 3 of the 5 replicates we performed. Mass spectrometry identification of these protein spots were done as described previously. Protein identifications were short-listed based on the total number of unique peptides identified for each protein in all of the 5 replicates. We then compared the molecular weight and isoelectric point of identified proteins to their

position in the 2D-DIGE gels. Protein IDs were thus prioritized and assigned to each spot (Table 4.1). Some spots contained more than one protein (example: Spots #2 & #3). We were unable to assign IDs to 17 protein spots due to their low abundance. Future experiments will need to be performed to identify those.

4.1. Discussion

There were several observed scenarios of spot pattern occurring in the DIGE gels. We have addressed a few themes as follows.

i. Glycerol-3-Phosphate dehydrogenase

See glycerol-3-phosphate dehydrogenase (Figure 4.4, spot #1). Green color indicates it was enriched in Q7 mitochondria and is reduced in Q111 mitochondria. This is in line with our initial hypothesis of impaired import of mitochondrial proteins in the mutant HTT cell line (i.e. Q111).

ii. Phosphoenolpyruvate carboxylase & Succinate dehydrogenase

See spots #2 & #3 (Figure 4.4). Spot #2 is green and #3 is red. According to the MS data (Table 4.1), both protein spots contained Phosphoenolpyruvate carboxylase & Succinate dehydrogenase. It is unclear which protein contributes to the green to red (left to right) shift we observed in the gel image. The isoforms of one or both of these proteins are different between Q7 and Q111. This shift is only in the horizontal direction, i.e. there is change in the isoelectric point of the protein. This effect could be due to a post-translational modification caused by an indirect effect of the import defect of a modifying enzyme such as kinase or a phosphatase.

iii. Aldehyde dehydrogenase

See Aldehyde dehydrogenase (spot #8) in Figure 4.4. It is enriched in Q111 and reduced in Q7. This is an opposite effect compared to our original hypothesis. A possible explanation for the enrichment of Aldehyde dehydrogenase in Q111 mitochondria is,

somehow the cell compensates for the effects of mutant HTT and overexpresses certain proteins.

iv. Ornithine aminotransferase

The last theme is a diagonal shift between red and green as seen in spots #15 and #16, which are two different isoforms of Ornithine aminotransferase. Again, this could be an indirect effect due to a modifying enzyme being affected by mutant HTT. The observed changes could be caused due to N-terminal cleavage or localization to the wrong mitochondrial compartment. It would be interesting to characterize the modification on such proteins and explore the mechanism of mitochondrial protein import defect.

Spot #	Uniprot ID	Protein name	Molecular weight	pI	# of MS datasets seen in	Total # of unique peptides	Sub-cellular Location
1	P48678	LMNA_MOUSE Prelamin-A/C	74,238.90	6.41	2	43	Nucleus
1	A2AQR0, Q64521	Glycerol-3-phosphate dehydrogenase, mitochondrial	82,830.80	6.01	2	8	Mitochondrial
2	Q8BH04	Phosphoenolpyruvate carboxykinase [GTP], mitochondrial	70,528.70	6.24	4	90	Mitochondrial
2	Q8K2B3	Succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial	72,585.60	6.34	3	54	Mitochondrial
3	Q8BH04	Phosphoenolpyruvate carboxykinase [GTP], mitochondrial	70,528.70	6.24	4	77	Mitochondrial
3	Q8K2B3	Succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial	72,585.60	6.34	3	35	Mitochondrial
4	P80318	T-complex protein 1 subunit gamma	60,631.10	6.28	2	34	Cytosol
5	Q99KE1	NAD-dependent malic enzyme, mitochondrial	65801.4	6.47	3	70	Mitochondrial
5	O08749	DLDH_MOUSE Dihydrolipoyl dehydrogenase, mitochondrial	54,272.40	6.43	2	24	Mitochondrial
8	P47738	Aldehyde dehydrogenase, mitochondrial	56,537.60	6.05	4	102	Mitochondrial
9	P50544	Very long-chain specific acyl-CoA dehydrogenase, mitochondrial	70877	7.72	2	9	Mitochondrial
10	Q5SX75	Procollagen-proline, 2-oxoglutarate 4-dioxygenase	61,016.10	5.55	2	29	cytosol, ER
10	P63038	60 kDa heat shock protein, mitochondrial	60956.8	5.35	2	15	Mitochondrial
11	Q5SX75	Procollagen-proline, 2-oxoglutarate 4-dioxygenase	61,016.10	5.55	3	66	cytosol, ER
11	P27773	PDIA3_MOUSE Protein disulfide-isomerase A3	56,680.40	5.69	2	36	ER
11	Q60715	P4HA1_MOUSE Prolyl 4-hydroxylase subunit alpha-1	60,912.00	5.62	2	21	Mitochondrion
13	P11983	TCPA_MOUSE T-complex protein 1 subunit alpha	60,449.90	5.82	2	73	Cytoplasm
14	Q99JY9	ARP3_MOUSE Actin-related protein 3	47,357.70	5.61	2	19	Cytoplasm
14	P56480	ATPB_MOUSE ATP synthase subunit beta, mitochondrial	56,301.20	4.99	2	17	Mitochondrial
15	P29758	Ornithine aminotransferase, mitochondrial	48,355.90	5.73	1	5	Mitochondrial
16	P29758	Ornithine aminotransferase, mitochondrial	48,355.90	5.73	2	13	Mitochondrial
17	P17182	ENOA_MOUSE Alpha-enolase	47,141.70	6.36	2	37	Cytoplasm
17	Q9D8N0	EF1G_MOUSE Elongation factor 1-gamma	50,061.30	6.33	2	16	Cytoplasm, ER
19	Q8BMD8	SCMC1_MOUSE Calcium-binding mitochondrial carrier protein SCaMC-1	52,903.80	7.02	2	22	Inner mitochondrial membrane
19	P35486	ODPA_MOUSE Pyruvate dehydrogenase E1 component subunit alpha, somatic form, mitochondrial	43,232.50	6.78	2	16	Mitochondrial
19	Q9QYR9	ACOT2_MOUSE Acyl-coenzyme A thioesterase 2, mitochondrial	49,658.00	6.09	2	10	Mitochondrial
20	Q9JHI5	IVD_MOUSE Isovaleryl-CoA dehydrogenase, mitochondrial	46,326.50	6.29	2	46	Mitochondrial
20	Q8K248	HPDL_MOUSE 4-hydroxyphenylpyruvate dioxygenase-like protein	39,909.80	6.35	2	15	
20	O88986	KBL_MOUSE 2-amino-3-ketobutyrate coenzyme A ligase, mitochondrial	44,931.90	6.46	2	10	Mitochondrial
20	Q8BMD8	SCMC1_MOUSE Calcium-binding mitochondrial carrier protein SCaMC-1	52,903.80	7.02	2	6	Inner mitochondrial membrane
28	P67778	PHB_MOUSE Prohibitin	29,820.60	5.57	3	19	Inner mitochondrial membrane
32	P20108	PRDX3_MOUSE Thioredoxin-dependent peroxide reductase, mitochondrial	28,127.00	5.71	2	16	Mitochondrial

Table 4.1. Mass spectrometry data analysis of Q7 vs Q111 mitochondria DIGE.

Therefore, we observe several changes caused by mutant HTT in the mitochondrial proteome. Some observations are in line with our initial hypothesis. However, the changes we observed are not unidirectional. Some mitochondrial proteins may experience an impairment in import into mitochondria. A close look at the structural and functional properties of such proteins could further our understanding of molecular mechanisms involved in Huntington's disease.

In order to explore their role in physiology, these proteins that have been identified will have to be further quantified in brain tissue of HTT mutant mice or patient brain tissue. mt-cox1 could be used to normalize western blots when quantifying these newly identified proteins. Thus, we have shown the DIGE comparison of mitochondrial proteome using two different methods. It is interesting to note that, based on the method chosen for mitochondrial isolation, we could end up with biased data depending on mitochondrial morphology. Using an antibody to isolate organelles and further perform DIGE appears to be a fruitful approach to study molecular mechanisms such as protein trafficking within cellular compartments. This study also shows that sample preparation is very important when performing comparative proteomics.

Chapter 5 Concluding Remarks

I've thus described the synthesis of a novel reagent and method development using Biotin-CDM. As a proof of principle, I've also shown that the method developed can be applied in autoantigen screening. The data we gathered to study Interstitial lung disease in the context of Rheumatoid Arthritis would open up further studies investigating the role of the identified autoantigens in the pathogenesis of RA-ILD. I have also shown that affinity proteomics is very useful for studying mitochondrial proteomes using DIGE. The data from this effort will be useful for studying the underlying molecular mechanism involving mitochondrial protein trafficking in Huntington's disease.

Moving forward, I see Biotin-CDM being useful for autoantigen screening and beyond. It is interesting to note that, CDM's reversibility and amine-blocking abilities were first studied in 1960's. CDM is so versatile that we are still coming up with new variations of the molecule. Before we made Biotin-CDM, Minden lab had made a previous version "Histidine-CDM", which was used for sample preparation for DIGE. Biotin-CDM could replace Histidine-CDM owing to the excellent binding properties of Biotin-Avidin. Secondly, what we learned in synthesis of- and method development using- Biotin-CDM will be useful when developing other proteomics reagents using CDM as a parent compound. Finally, as a pH sensitive-reversible-primary amine labeling reagent there could be other useful applications for Biotin-CDM in other areas of proteomics / biochemistry to enrich and purify proteins.

The last 5 years have been a great learning experience for me and gave me a lot of perspective. Above all, the 5 years has given me confidence and proved that "Where there is a will, there is a way". There were several humbling moments that made me aware of my limited experience and helped me learn from the people around me. On the other hand, because no one else has the exact same research knowledge or experience, there were instances where I was considered the "expert" on certain things. Those moments made me aware of the responsibility I owed to the scientific community, to perform well-

controlled experiments and being truthful in reporting the scientific data. Synthesizing data and testing hypotheses is at the core of the doctoral study. I'm grateful for the opportunity to hone this skill throughout my time as a PhD candidate. I'm immensely happy that my work will be useful in the future for Minden Lab and the larger scientific community.

Learning and teaching are two sides of the same coin. I also got the opportunity to directly mentor and teach 5 undergraduate students in the lab. Mentoring was not as simple as being a teaching assistant on courses. Mentoring for me, was a leadership experience. It was challenging to focus on my research projects and design projects & experiments for undergraduate students to perform. I admire my advisor Jonathan Minden, who is so good at multi-tasking: writing grants, advising graduate and undergraduate students, performing individual research among many other things. Being passionate about this is contagious. Witnessing such motivated undergraduate students work with me lifted my spirit on multiple occasions. In my opinion, willing to learn from anyone and anywhere is what being a scientist is all about.

During my PhD, I worked with several collaborators. Some have been fruitful, while others have not. Taking initiative and willing to solve challenging problems takes a lot of time and energy. Before investing much time and energy, I learnt that it is important to evaluate the nature of the collaboration. Being a proteomics research lab, the Minden lab gets a lot of requests for collaboration. Most of those invitations are for us to perform DIGE experiments on the collaborators' samples. All successfully completed/ongoing collaborative research projects have had something in common and that is, these projects went beyond just performing DIGE experiments. In my observation with 6 DIGE project collaborators, the project moved forward smoothly only when (1) the preliminary data was worthwhile to pursue; (2) Samples are readily available either locally or can be easily obtained by us; and (3) There was enthusiasm to continue the project from both sides of the collaboration by participating in grant proposals together. Of the 6 DIGE collaborations, only 2 have been successful. Communication is the key to working

together in any setting. The collaborators should be committed to seeing the research move forward and willing to communicate at all times, whatever the status of the project may be. Especially in this modern age with several technologies for communication, it is easy to both co-ordinate with collaborators as well as avoid each other. I learnt that it is important to evaluate the nature of the collaboration before investing oneself in collaborative projects.

That said, working in collaboration brings a lot of soft skills into action. I worked on projects ranging from chemical synthesis to tissue culture and biomarker screening. Such a wide range of projects and the interdisciplinary nature of the work wouldn't have been possible without forming and maintaining successful collaborations. I joined the Minden lab to work on proteomics method development. I'm also happy that I got to work in another field - Immunology. Immunology has been never done in the Minden lab before and I am happy to have stepped into this field through our collaborations. Working on an autoimmune disease in a proteomics method development lab is a rare opportunity, which I'm very grateful for. Being able to develop a method and further applying that to a study that could impact human health/disease is fulfilling. One of the undergraduate students who worked with me on the RA-ILD project, was even inspired to continue her graduate studies in Immunology. I hope the lessons we learnt during my PhD will be useful for the Minden lab, as much as it is for me.

APPENDIX I - CDM labeling activity assay

Notes:

- Reserve MBIC TECAN plate reader for at least 20 minutes of measurements
- Do the following in 0.6mL tubes
- Make sure to perform DMF control for each test condition

Components:

1. Substrate: 2.5mM Ac-PNA-Lys (stored in -20)
2. Enzyme: Trypsin - tissue culture grade (stored in -20)
3. Assay buffer: 100mM HEPES pH 8.0
4. Biotin-CDM (56mM or 30mg/mL): 150-fold molar excess over substrate in the assay reaction.
5. Methylamine (1M): 10-fold molar excess over Biotin-CDM to quench any active CDM in the assay reaction.

Assay Set up:

	H_2O µL	Substrate µL	1M HEPES pH 8.0 µL	Biotin - CDM µL	DMF µL	On ice for 30 min	Methyl-amine µL	On ice for 30 min	100mM HEPES pH 8.0 µL	Trypsin µL
No substrate	58.6	--	8.8	--	13.2		7.4		122	10
No enzyme	68.6	2	8.8	--	13.2		7.4		122	--
DMF control	56.6	2	8.8	--	13.2		7.4		122	10
CDM Reaction	56.6	2	8.8	13.2	--		7.4		122	10
Test Condition	56.6 (Test buffer)	2	--	13.2	--		7.4		122	10
DMF control for test condition	56.6 (Test buffer)	2	--	--	13.2		7.4		122	10

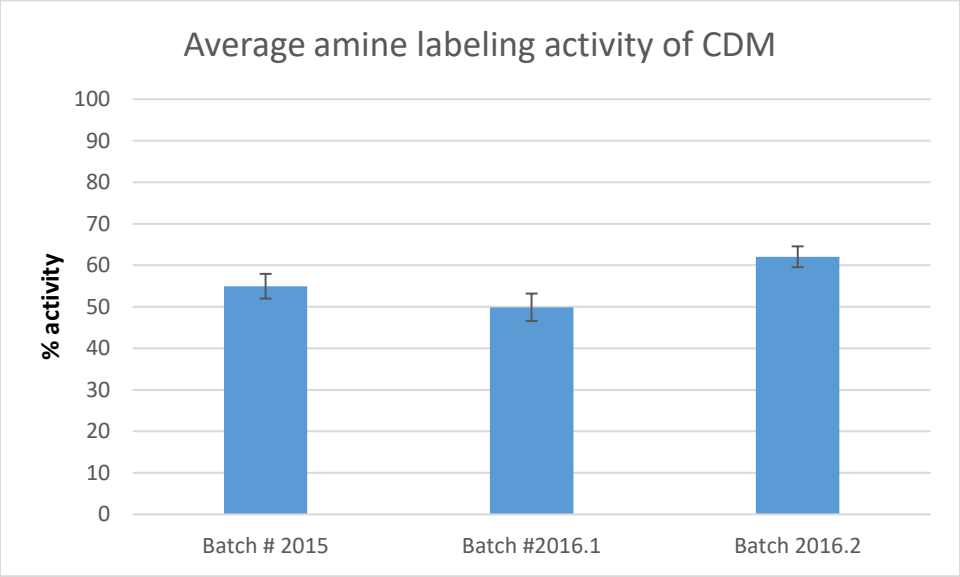
- Mix everything, follow incubation times and then immediately after adding trypsin, mix with pipette, transfer to 96 well-plate
- Make sure there are no bubbles.
- Measure selected wells using LysC protocol in TECAN (absorbance @405nm for 15 minutes).
- Plot Absorbance (y-axis) Vs time(x-axis) and calculate slope for each condition

Calculate for each condition:

% amines blocked by CDM = $[(Slope_{DMF\ control} - Slope_{BiotinCDM\ rxn}) / Slope_{DMF\ control}]*100$

See results from Chapter 2 Section 2.5.2 on page 33

Sample quality control data:



APPENDIX II – AMIDA

From Chapter 3, This is the refined protocol used for Antibody meditated identification of antigens (AMIDA) incorporating Biotin-CDM.

Step 1, Deimination followed by Biotin-CDM labeling:

Turn on 37°C Water bath in Jon's side of the lab (across from the balance).
Bring lysates and PAD2 from -80 and place on ice before beginning the experiment.
Use Eppendorf lo-bind tubes from the packaging on bench for the citrullination step.

#	1 Untreated	2 Deiminated	3 Untreated	4 Deiminated	Final
HeLa 1.2 mg/mL	10 mL	10 mL	10 mL	10 mL	Total 6 mg each This volume is "V_{HeLa}"
0.5M CaCl2	100 µL	100 µL	100 µL	100 µL	~5mM (in HeLa extract volume, not total) (V_{HeLa})/100
PAD2 10 units each	-	32 µL	-	32 µL	PAD2 alliquots in -80 HeLa PAD box
DO NOT VORTEX, Spin down quickly.					
Incubate on 37°C for 1 hour					
0.5 M EDTA To stop citrullination	100 µL	100 µL	100 µL	100 µL	5mM -> same volume as $CaCl_2$
Transfer each to a 15mL conical					
Make-up Buffer 100mM HEPES pH 8.0 150mM NaCl 5% Glycerol 1% IgePal CA 630 1mM PMSF 0.01mg/ml Leu+Pep	480 µL	438 µL	480 µL	438 µL	Make up to 12mL with this buffer This buffer is stored in 4°C Fridge
Biotin-CDM (30 mg/mL)	1200 µL	1200 µL	1200 µL	1200 µL	**3X mass fold excess over protein extract**
On ice for 30 minutes					
Quencher	120 µL	120 µL	120 µL	120 µL	~10-fold molar excess over Biotin-CDM
On ice for 30 minutes					

Step 2: Ultracentrifugation

Transfer this prep to cleaned ultracentrifuge tubes (polycarbonate tubes capacity 35mL), weigh all of them. If not the same weight, add make-up buffer so the weights are same for all the tubes. Proceed with Ultracentrifugation for 1hr @30000rpm (or 100000 xg) in common room at 4°C. Save the supernatant in a 15mL conical. Rotor used: Ti 70, which is stored in the shared space in our cold room.

Cleaning: Once supernatant is removed, to clean the polycarbonate tubes, pour approximately 500uL of 20% SDS to cover the pellet, add water up to brim, leave it for 15 minutes soaking in SDS. Clean it out. This is a gentle cleaning step so as not to scratch the ultracentrifuge tubes. Thaw patient sera while waiting for the centrifuge to finish

Step 3: Free antibody approach: Immune complex formation

Sera used in this experiment, stored in -80 Sera box:

#	RA	RA-ILD
Pool 2	22, 90, 54	27, 76, 49
Pool 3	23, 37, 55	70, 29, 25

Mix 120 µL each of the sera for this experiment.

Mix Ab with Ag as follows:

1 (uncit)	180uL RA pooled Patient sera Pool 2
2 (cit)	180uL RA pooled Patient Sera Pool 2
3 (uncit)	180uL RA-ILD pooled Patient Sera Pool 2
4 (cit)	180uL RA-ILD pooled Patient Sera Pool 2
5 (uncit)	180uL RA pooled Patient sera Pool 3
6 (cit)	180uL RA pooled Patient Sera Pool 3
7 (uncit)	180uL RA-ILD pooled Patient Sera Pool 3
8 (cit)	180uL RA-ILD pooled Patient Sera Pool 3

Mix in cold room end-over in lo-bind 5mL tubes for 1 hour

From this point onwards, DO NOT CENTRIFUGE, IMMUNE COMPLEX COULD BREAK.

Step 4: Immune complex binding to Dynabeads Protein-A magnetic beads

1. Carefully vortex the bottle of dynabeads before transferring to each tube.
2. Note: If the beads settle down, vortex the bottle again!!
3. For 8 reactions: Transfer 3.6mL beads to a 5mL tube
4. Place on magnetic stand, once the beads stick to the side of the tube, remove supernatant completely and discard. Take a beaker and label it "discard". (to be tossed at the end of the day into biohazard)
5. Go to cold room
6. Re-suspend in 3.6mL wash buffer (PBS with TX-100 & PMSF) gently by pipetting up and down.
7. Make sure the beads are all mixed and homogenous.
8. Now, transfer 450 µL of beads to each reaction (the reactions are already in 15 mL tubes, at this point)

9. Avoid foam while mixing, do not spin immune complexes
10. Mix end-over for 1 hour in cold room

Thaw following before the washing step begins : DIGE LB & DTT
Turn on cooling mixer in 1D room, set temperature= 0 deg C

Step 5: Washing unbound lysate & Elution

1. Bring the rotating thingy from the cold room to the bench.
2. Place 15mL tubes in magnetic rack.
3. Wait for the beads to stick to the side of the tube
4. Remove supernatant and discard.
5. Remove the tube and place it on the bench, before adding any buffer *This is important- so the beads don't clump together*
6. Add 3mL of wash buffer to each tube: PBS with 0.1% Triton X-100, 0.5mM PMSF (To make this buffer, add: 500 µL from PMSF stock, 500 µL from 10% Tx-100 clear bottle, make up to 50 mL with PBS) and mix end-over for 5 minutes at RT.
   Keep the buffer on ice
7. Place the tubes on the magnetic rack
8. Make sure to get the liquid/beads that is stuck inside of the cap
9. Wait for the magnetic beads to stick to the side of the tube.
10. Remove wash on magnetic stand and discard.
11. Place the tube on the bench
12. Repeat the wash steps twice: steps 10-15
13. For the 4th wash, rotate for 10 minutes.
14. Remove wash on magnetic stand, place tubes on the bench.
    Add the following for Elution

Tube	DIGE LB µL (in -80 fridge)	1M DTT µL
1	150	2
2	150	2
3	150	2
4	150	2
5	150	2
6	150	2
7	150	2
8	150	2

15. Incubate for 30 minutes on cooling mixer at 1000 rpm and lowest temp set at 0 deg C(The temperature may only go down to ~4°C-7°C, that is fine, it varies according to room temperature)
16. Collect Elute on magnetic stand

Step 6: Capture by neutravidin beads

Continue with magnetic neutravidin (speed beads - 10ug/uL binding capacity -35 picomoles of bsa/ul beads). These beads are in the little fridge in 1D room.

Make following buffers: WB2, WB3, WB4
EB should be in the -80 buffers box

Binding elutes with Magnetic neutravidin

1. Vortex magnetic neutravidin beads
2. For 8 reactions: Add **240 µL magnetic neutravidin beads** to one 1.5mL **low-binding tube**
3. Separate on magnet and discard supernatant
4. Add **1 mL Binding buffer**. (100mM HEPES pH 8.0, 150mM NaCl, 0.5mM PMSF + 2% CHAPS)
5. Mix the beads by pipetting up and down. Make sure they are homogenous.
6. Label 8 low-binding tubes: 1-8
7. Add **125 µL beads** to each tube
8. Separate supernatant on magnetic stand and discard supernatant
9. Now, carefully add corresponding protein elutes to beads
10. Add 500 µL **Binding buffer**. (100mM HEPES pH 8.0, 150mM NaCl, 0.5mM PMSF + 2% CHAPS)
11. Rotate end-over-end at 4°C for 30 minutes

Make following buffers: WB2, WB3, WB4
EB should be in the -80 buffers box

Step 7: Washes, Cy-dye labeling and elution

12. Separate on magnetic stand and discard wash.
13. Wash buffers with NaCl concentration WB2=150mM, WB3=50mM, WB4=5mM are used next.
14. Add 1mL WB2, mix by rotating the beads away from the magnet in the tube, separate on magnetic stand
15. Repeat with (1 mL) WB3 and (1mL) WB4. Discard: washes in the waste 50mL conical.
16. Transfer beads + last wash to a new tube, Separate on magnetic stand and discard WB4.
17. Cy3 and Cy5 tubes re-suspended in fresh DMF: 2.5ul and 3ul respectively.
18. Add 30uL WB4 and 1uL Cy3, Cy5 to respective tubes according to this table:

Tube	Cy3 µL	Cy5 µL
1 (RA: Untreated)	1	-
2 (RA: Deiminated)	-	1
3 (RA-ILD: Untreated)	-	1
4 (RA-ILD: Deiminated)	1	-
5(RA: Untreated)	1	-
6 (RA: Deiminated)	-	1
7 (RA-ILD: Untreated)	-	1
8 (RA-ILD: Deiminated)	1	-

19. Incubate at lowest temperature (close to 0degc) with shaking for 15 min in the mixing block
20. Add 1uL quencher to each
21. Incubate at lowest temperature (close to 0degC) with shaking for 15 min in the mixing block.
22. Separate and remove the liquids on magnetic stand
23. Add 30uL EB to each tube

24. 1 hour on cooling mixer at 4-5degC (lowest temp)
25. Separate on magnetic stand, save Elute
26. Run on Mini-2D-DIGE set up.

MAKE WASH BUFFERS FRESH
Make sure there are enough of these buffers before starting experiment

components	**WB 2**	**WB 3**	**WB 4**	**EB (pH 3.7)**	
Urea	4.33	4.33	4.33	4.33	g
CHAPS	0.18	0.18	0.18	0.18	g
1M DTT	90	90	90	90	µL
2 M NaCl	675	225	22.5	9	µL
1M Hepes pH 8.0	900	900	45	------	------
1M Citric acid pH 2.5	-	-	-	180	µL
Make up each to 9 mL with water					

Stored in fridge:

1. Wash buffer for Protein A beads after lysate addition:
   PBS with 0.05% Triton X-100, 0.5mM PMSF
2. Binding buffer for Neutravidin:
   100mM HEPES pH 8.0, 150mM NaCl, 0.5mM PMSF + 2% CHAPS

Components	Wash buffer for IP (100 mL)	Make up buffer (2 mL)	Binding buffer for Neutravidin beads (50 mL)
PBS	100 mL	-	-
1M HEPES	-	200 µL	5 mL
10% Triton x-100	500 µL	-	-
IgePal CA-630	-	20 µL	-
Glycerol	-	100 µL	-
50mM PMSF	1 mL	40 µL	500 µL
10mg/mL Leu + Pep		2 µL	-
2M NaCl	-	150 µL	3.75 mL
CHAPS	-	-	1 g

APPENDIX III - Mini 2DE Set up

Day 1- Rehydration:
Rehydrate two 7cm strips with 125uL + 2.5uL IPG buffer each overnight

- Total rehydration buffer for 2 strips - 250uL
- Total IPG buffer for 2 strips - 5uL

Day 2- IEF, transfer and SDS PAGE:

- Recommended protein load:

	Suitable sample load (µg of protein)		
Immobiline DryStrip	Silver /Deep Purple™	Coomassie (preparative)	CyDye
7cm 3–11 NL	3–6	25–60	10
7cm 3–10	3–6	25–60	10
7cm 3–10 NL	3–6	25–60	10
7cm 4–7	4–8	25–150	13
7cm 3–5.6 NL	8–16	40–240	26
7cm 5.3–6.5	8–16	40–240	26
7cm 6.2–7.5	8–16	40–240	26
7cm 6–11	8–16	40–240	26
7cm 7–11NL	8–16	40–240	26

- Maximum protein volume: ~60uL
- For every 50uL sample, add 1uL appropriate IPG buffer
- Start IEF in the morning
- **Protocol#10 on GE machine: VG 6hr 7cm strip**
  1 step and hold 100V 1hr
  2 step and hold 300V 2hr
  3 gradient 1000V 30 minutes
  4 gradient 5000V 1hr:30 minutes
  5 step and hold 5000V 30 minutes
- Normal Volt hour range: 5000 Vhr - 8000 Vhr

Transfer:

- Pre-cast gels from Jule, Inc. stored in 4degC fridge in 2D room
- Transfer to mini gel using regular protocol: 1mL agarose stack, 0.5mL agarose seal

SDS-PAGE:

- Use the tank setup with maroon gaskets, regular tank buffers.
- Run through stack at 60 V and then increase voltage to XX (usually ~80V) where the starting current is 35 milli-amps per gel. The current goes down over time though. Takes 2.5 hrs. to finish.

Bibliography

1. Ganesan, V., Ascherman, D. P. & Minden, J. S. Immunoproteomics technologies in the discovery of autoantigens in autoimmune diseases. *Biomol. Concepts* **7,** (2016).
2. Purcell, A. W. & Gorman, J. J. Immunoproteomics: Mass spectrometry-based methods to study the targets of the immune response. *Mol Cell Proteomics* **3,** 193–208 (2004).
3. Tassinari, O. W., Caiazzo Jr., R. J., Ehrlich, J. R. & Liu, B. C. Identifying autoantigens as theranostic targets: antigen arrays and immunoproteomics approaches. *Curr Opin Mol Ther* **10,** 107–115 (2008).
4. Kohler, K. & Seitz, H. Validation processes of protein biomarkers in serum--a cross platform comparison. *Sensors (Basel)* **12,** 12710–12728 (2012).
5. Klade, C. S. *et al.* Identification of tumor antigens in renal cell carcinoma by serological proteome analysis. *Proteomics* **1,** 890–898 (2001).
6. Kellner, R. *et al.* Targeting of tumor associated antigens in renal cell carcinoma using proteome-based analysis and their clinical significance. *Proteomics* **2,** 1743–1751 (2002).
7. Cottrell, T. R., Hall, J. C., Rosen, A. & Casciola-Rosen, L. Identification of novel autoantigens by a triangulation approach. *J Immunol Methods* **385,** 35–44 (2012).
8. Dutoit-Lefevre, V. *et al.* An Optimized Fluorescence-Based Bidimensional Immunoproteomic Approach for Accurate Screening of Autoantibodies. *PLoS One* **10,** e0132142 (2015).
9. Colomba, P. *et al.* Identification of biomarkers in cerebrospinal fluid and serum of multiple sclerosis patients by immunoproteomics approach. *Int J Mol Sci* **15,** 23269–23282 (2014).
10. Adkins, J. N. *et al.* Toward a human blood serum proteome: analysis by multidimensional separation coupled with mass spectrometry. *Mol Cell Proteomics* **1,** 947–955 (2002).
11. Linke S. Doraiswamy and E. H. Harrison, T., Linke, T., Doraiswamy, S. & Harrison, E. H. Rat plasma proteomics: effects of abundant protein depletion on proteomic analysis. *J Chromatogr B Anal. Technol Biomed Life Sci 849(1-2) 273- 281* **849,** 273–281 (2007).
12. Chen, P. *et al.* Annexin A2 as a target endothelial cell membrane autoantigen in Behcet's disease. *Sci Rep* **5,** 8162 (2015).
13. Sharma, S. *et al.* Identification of autoantibodies against transthyretin for the screening and diagnosis of rheumatoid arthritis. *PLoS One* **9,** e93905 (2014).
14. Regent, A. *et al.* Identification of target antigens of anti-endothelial cell antibodies in patients with anti-neutrophil cytoplasmic antibody-associated vasculitides: a proteomic approach. *Clin Immunol* **153,** 123–135 (2014).
15. Bruschi, M. *et al.* Glomerular autoimmune multicomponents of human lupus nephritis in vivo: alpha-enolase and annexin AI. *J Am Soc Nephrol* **25,** 2483–2498 (2014).

16. Liu, H. *et al.* Detection and analysis of autoantigens targeted by autoantibodies in immunorelated pancytopenia. *Clin Dev Immunol* **2013,** 297678 (2013).
17. Kapustian, L. L. *et al.* Hsp90 and its co-chaperone, Sgt1, as autoantigens in dilated cardiomyopathy. *Hear. Vessel.* **28,** 114–119 (2013).
18. Biswas, S. *et al.* Identification of novel autoantigen in the synovial fluid of rheumatoid arthritis patients using an immunoproteomics approach. *PLoS One* **8,** e56246 (2013).
19. Van den Bergh, K. *et al.* Heterogeneous nuclear ribonucleoprotein h1, a novel nuclear autoantigen. *Clin Chem* **55,** 946–954 (2009).
20. Liu, J. *et al.* Identification of antigenic proteins associated with trichloroethylene-induced autoimmune disease by serological proteome analysis. *Toxicol Appl Pharmacol* **240,** 393–400 (2009).
21. Derfuss, T. *et al.* Contactin-2/TAG-1-directed autoimmunity is identified in multiple sclerosis patients and mediates gray matter pathology in animals. *Proc Natl Acad Sci U S A* **106,** 8302–8307 (2009).
22. Vermeulen, N. *et al.* Anti-alpha-enolase antibodies in patients with inflammatory Bowel disease. *Clin Chem* **54,** 534–541 (2008).
23. Lovato, L. *et al.* Transketolase and 2′,3′-cyclic-nucleotide 3′-phosphodiesterase type I isoforms are specifically recognized by IgG autoantibodies in multiple sclerosis patients. *Mol Cell Proteomics* **7,** 2337–2349 (2008).
24. Kimura, A. *et al.* Proteomic analysis of autoantibodies in neuropsychiatric systemic lupus erythematosus patient with white matter hyperintensities on brain MRI. *Lupus* **17,** 16–20 (2008).
25. Kikuchi, T., Yoshida, Y., Morioka, T., Gejyo, F. & Oite, T. Human voltage-dependent anion selective channel 1 is a target antigen for antiglomerular endothelial cell antibody in mixed connective tissue disease. *Mod Rheumatol* **18,** 570–577 (2008).
26. Gini, B. *et al.* Novel autoantigens recognized by CSF IgG from Hashimoto's encephalitis revealed by a proteomic approach. *J Neuroimmunol* **196,** 153–158 (2008).
27. Fae, K. C. *et al.* PDIA3, HSPA5 and vimentin, proteins identified by 2-DE in the valvular tissue, are the target antigens of peripheral and heart infiltrating T cells from chronic rheumatic heart disease patients. *J Autoimmun* **31,** 136–141 (2008).
28. Tilleman, K. *et al.* In pursuit of B-cell synovial autoantigens in rheumatoid arthritis: Confirmation of citrullinated fibrinogen, detection of vimentin, and introducing carbonic anhydrase as a possible new synovial autoantigen. *Proteomics Clin Appl* **1,** 32–46 (2007).
29. Okunuki, Y. *et al.* Proteomic surveillance of autoimmunity in Behcet's disease with uveitis: selenium binding protein is a novel autoantigen in Behcet's disease. *Exp Eye Res* **84,** 823–831 (2007).
30. Mathey, E. K. *et al.* Neurofascin as a novel target for autoantibody-mediated axonal injury. *J Exp Med* **204,** 2363–2372 (2007).
31. Tanaka, Y. *et al.* Proteomic surveillance of autoantigens in relapsing

polychondritis. *Microbiol Immunol* **50,** 117–126 (2006).
32. Matsuo, K. *et al.* Identification of novel citrullinated autoantigens of synovium in rheumatoid arthritis using a proteomic approach. *Arthritis Res Ther* **8,** R175 (2006).
33. Kim, C. W. *et al.* Disease-specific proteins from rheumatoid arthritis patients. *J Korean Med Sci* **21,** 478–484 (2006).
34. Li, W. H. *et al.* Proteomics-based identification of autoantibodies in the sera of healthy Chinese individuals from Beijing. *Proteomics* **6,** 4781–4789 (2006).
35. Gires, O. *et al.* Profile identification of disease-associated humoral antigens using AMIDA, a novel proteomics-based technology. *Cell Mol Life Sci* **61,** 1198–1207 (2004).
36. Yang, W., Steen, H. & Freeman, M. R. Proteomic approaches to the analysis of multiprotein signaling complexes. *Proteomics* **8,** 832–851 (2008).
37. Kaji, K. *et al.* Autoantibodies to RuvBL1 and RuvBL2: a novel systemic sclerosis-related antibody associated with diffuse cutaneous and skeletal muscle involvement. *Arthritis Care Res* **66,** 575–584 (2014).
38. Harlow, L. *et al.* Identification of citrullinated hsp90 isoforms as novel autoantigens in rheumatoid arthritis-associated interstitial lung disease. *Arthritis Rheum* **65,** 869–879 (2013).
39. Kuo, Y. B. *et al.* Identification and clinical association of anti-cytokeratin 18 autoantibody in COPD. *Immunol Lett* **128,** 131–136 (2010).
40. Stea, E. A. *et al.* Analysis of parotid glands of primary Sjogren's syndrome patients using proteomic technology reveals altered autoantigen composition and novel antigenic targets. *Clin Exp Immunol* **147,** 81–89 (2007).
41. Sousa, M. M. *et al.* Antibody cross-linking and target elution protocols used for immunoprecipitation significantly modulate signal-to noise ratio in downstream 2D-PAGE analysis. *Proteome Sci. 2011 91* **75,** 663–70 (2011).
42. Kang, S. *et al.* Synthesis of biotin-tagged chemical cross-linkers and their applications for mass spectrometry. *Rapid Commun. Mass Spectrom.* **23,** 1719–1726 (2009).
43. Thermo Scientific. Avidin-Biotin Technical Handbook. 51 (2009).
44. Van Hoang, M. *et al.* Quantitative proteomics employing primary amine affinity tags. *J. Biomol. Tech.* **14,** 216–223 (2003).
45. Lee, S., Wang, W., Lee, Y. & Sampson, N. S. Cyclic acetals as cleavable linkers for affinity capture. *Org. Biomol. Chem.* **13,** 8445–52 (2015).
46. Gartner, C. A., Elias, J. E., Bakalarski, C. E. & Gygi, S. P. Catch-and-release reagents for broadscale quantitative proteomics analyses. *J. Proteome Res.* **6,** 1482–1491 (2007).
47. Nathani, R. I. *et al.* Reversible protein affinity-labelling using bromomaleimide-based reagents. *Org. Biomol. Chem.* **11,** 2408–11 (2013).
48. Ohyama, K. *et al.* Proteomic profiling of antigens in circulating immune complexes associated with each of seven autoimmune diseases. *Clin Biochem* **48,** 181–185 (2015).
49. Ohyama, K. & Kuroda, N. Immune complexome analysis. *Adv Clin Chem* **60,** 129–

141 (2013).
50. Gallo, A. *et al.* Gross Cystic Disease Fluid Protein-15(GCDFP-15)/Prolactin-Inducible Protein (PIP) as Functional Salivary Biomarker for Primary Sjogren's Syndrome. *J Genet Syndr Gene Ther* **4,** (2013).
51. Broadwater, L. *et al.* Analysis of the mitochondrial proteome in multiple sclerosis cortex. *Biochim Biophys Acta* **1812,** 630–641 (2011).
52. Yamamoto, M. *et al.* Proteomics analysis in 28 patients with systemic IgG4-related plasmacytic syndrome. *Rheumatol Int* **30,** 565–568 (2010).
53. Baillet, A. *et al.* Synovial fluid proteomic fingerprint: S100A8, S100A9 and S100A12 proteins discriminate rheumatoid arthritis from other inflammatory joint diseases. *Rheumatol.* **49,** 671–682 (2010).
54. Okrojek, R., Grus, F. H., Matheis, N. & Kahaly, G. J. Proteomics in autoimmune thyroid eye disease. *Horm Metab Res* **41,** 465–470 (2009).
55. Raju, R. *et al.* Autoimmunity to GABAA-receptor-associated protein in stiff-person syndrome. *Brain* **129,** 3270–3276 (2006).
56. de Seny, D. *et al.* Discovery of new rheumatoid arthritis biomarkers using the surface-enhanced laser desorption/ionization time-of-flight mass spectrometry ProteinChip approach. *Arthritis Rheum* **52,** 3801–3812 (2005).
57. Wang, M., Herrmann, C. J., Simonovic, M., Szklarczyk, D. & von Mering, C. Version 4.0 of PaxDb: Protein abundance data, integrated across model organisms, tissues, and cell-lines. *Proteomics* **15,** 3163–3168 (2015).
58. Alahmad, A. *et al.* Desmoplakin and KIF20B as target antigens in patients with paroxysmal nocturnal haemoglobinuria. *Br J Haematol* **151,** 273–280 (2010).
59. Reimand, K. *et al.* Testis-expressed protein TSGA10 an auto-antigen in autoimmune polyendocrine syndrome type I. *Int Immunol* **20,** 39–44 (2008).
60. Avila, J. *et al.* Autoantigenic nuclear proteins of a clinically atypical renal vasculitis. *J Autoimmune Dis* **5,** 3 (2008).
61. Bensing, S. *et al.* Pituitary autoantibodies in autoimmune polyendocrine syndrome type 1. *Proc Natl Acad Sci U S A* **104,** 949–954 (2007).
62. Uchida, K. *et al.* Identification of specific autoantigens in Sjogren's syndrome by SEREX. *Immunology* **116,** 53–63 (2005).
63. Ou, Y., Enarson, P., Rattner, J. B., Barr, S. G. & Fritzler, M. J. The nuclear pore complex protein Tpr is a common autoantigen in sera that demonstrate nuclear envelope staining by indirect immunofluorescence. *Clin Exp Immunol* **136,** 379–387 (2004).
64. Dunphy, E. J., Eickhoff, J. C., Muller, C. H., Berger, R. E. & McNeel, D. G. Identification of antigen-specific IgG in sera from patients with chronic prostatitis. *J Clin Immunol* **24,** 492–502 (2004).
65. Skoldberg, F. *et al.* Identification of AHNAK as a novel autoantigen in systemic lupus erythematosus. *Biochem Biophys Res Commun* **291,** 951–958 (2002).
66. Schmits, R., Kubuschok, B., Schuster, S., Preuss, K. D. & Pfreundschuh, M. Analysis of the B cell repertoire against autoantigens in patients with giant cell arteritis and polymyalgia rheumatica. *Clin Exp Immunol* **127,** 379–385 (2002).

67. Lim, Y. *et al.* Identification of autoantibodies associated with systemic lupus erythematosus. *Biochem Biophys Res Commun* **295,** 119–124 (2002).
68. Kira, J. [Molecular immunogenetic approach to the pathogenesis of multiple sclerosis]. *Rinsho Shinkeigaku* **42,** 1198–1200 (2002).
69. Jeoung, D. I. *et al.* Autoantibody to DNA binding protein B as a novel serologic marker in systemic sclerosis. *Biochem Biophys Res Commun* **299,** 549–554 (2002).
70. Bhat, S., Mary, S., Banarjee, R., Giri, A. P. & Kulkarni, M. J. Immune response to chemically modified proteome. *Proteomics Clin Appl* **8,** 19–34 (2014).
71. Larman, H. B. *et al.* Autoantigen discovery with a synthetic human peptidome. *Nat Biotechnol* **29,** 535–541 (2011).
72. Larman, H. B. *et al.* PhIP-Seq characterization of autoantibodies from patients with multiple sclerosis, type 1 diabetes and rheumatoid arthritis. *J Autoimmun* **43,** 1–9 (2013).
73. Haggmark, A. *et al.* Proteomic profiling reveals autoimmune targets in sarcoidosis. *Am J Respir Crit Care Med* **191,** 574–583 (2015).
74. Zingaretti, C. *et al.* Identification of new autoantigens by protein array indicates a role for IL4 neutralization in autoimmune hepatitis. *Mol Cell Proteomics* **11,** 1885–1897 (2012).
75. Wu, L. & Song, G. Identification of new autoimmune hepatitis-specific autoantigens by using protein microarray technology. *Methods Mol Biol* **909,** 227–239 (2012).
76. Huang, W. *et al.* Novel systemic lupus erythematosus autoantigens identified by human protein microarray technology. *Biochem Biophys Res Commun* **418,** 241–246 (2012).
77. Hu, C. J. *et al.* Identification of new autoantigens for primary biliary cirrhosis using human proteome microarrays. *Mol Cell Proteomics* **11,** 669–680 (2012).
78. Song, Q. *et al.* Novel autoimmune hepatitis-specific autoantigens identified using protein microarray technology. *J Proteome Res* **9,** 30–39 (2010).
79. Leidinger, P. *et al.* Novel autoantigens immunogenic in COPD patients. *Respir Res* **10,** 20 (2009).
80. Song, G. *et al.* New centromere autoantigens identified in systemic sclerosis using centromere protein microarrays. *J Rheumatol* **40,** 461–468 (2013).
81. Tabakman, S. M. *et al.* Plasmonic substrates for multiplexed protein microarrays with femtomolar sensitivity and broad dynamic range. *Nat Commun* **2,** 466 (2011).
82. Blixt, O. *et al.* A high-throughput O-glycopeptide discovery platform for seromic profiling. *J Proteome Res* **9,** 5250–5261 (2010).
83. Silverman, G. J. *et al.* Genetic imprinting of autoantibody repertoires in systemic lupus erythematosus patients. *Clin Exp Immunol* **153,** 102–116 (2008).
84. Zhu, H., Luo, H., Yan, M., Zuo, X. & Li, Q. Z. Autoantigen Microarray for High-throughput Autoantibody Profiling in Systemic Lupus Erythematosus. *Genomics Proteomics Bioinforma.* **13,** 210–218 (2015).
85. Wright, C. *et al.* Detection of multiple autoantibodies in patients with ankylosing spondylitis using nucleic acid programmable protein arrays. *Mol Cell Proteomics*

11, M9 00384 (2012).
86. Burbelo, P. D. *et al.* Extrapancreatic autoantibody profiles in type I diabetes. *PLoS One* **7,** e45216 (2012).
87. Burbelo, P. D. *et al.* Rapid serological detection of autoantibodies associated with Sjogren's syndrome. *J Transl Med* **7,** 83 (2009).
88. Anderson, N. L., Ptolemy, A. S. & Rifai, N. The riddle of protein diagnostics: future bleak or bright? *Clin Chem* **59,** 194–197 (2013).
89. Anderson, N. L. & Anderson, N. G. The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics* **1,** 845–867 (2002).
90. Hortin, G. L., Carr, S. A. & Anderson, N. L. Introduction: Advances in protein analysis for the clinical laboratory. *Clin Chem* **56,** 149–151 (2010).
91. Lewallen, D. M. *et al.* Chemical Proteomic Platform To Identify Citrullinated Proteins. *ACS Chem Biol* **10,** 2520–2528 (2015).
92. Bang, S. Y. *et al.* Smoking increases rheumatoid arthritis susceptibility in individuals carrying the HLA-DRB1 shared epitope, regardless of rheumatoid factor or anti-cyclic citrullinated peptide antibody status. *Arthritis Rheum.* **62,** 369–377 (2010).
93. Makrygiannakis, D. *et al.* Smoking increases peptidylarginine deiminase 2 enzyme expression in human lungs and increases citrullination in BAL cells. *Ann. Rheum. Dis.* **67,** 1488–92 (2008).
94. Hernández, B., Parnell, A. & Pennington, S. R. Why have so few proteomic biomarkers 'survived' validation? (Sample size and independent validation considerations). *Proteomics* **14,** 1587–1592 (2014).
95. Petrak, J. *et al.* Deja vu in proteomics. A hit parade of repeatedly identified differentially expressed proteins. *Proteomics* **8,** 1744–1749 (2008).
96. Ganesan, V. *et al.* Immuno-proteomics: Development of a novel reagent for separating antibodies from their target proteins. *Biochim. Biophys. Acta - Proteins Proteomics* **1854,** 592–600 (2015).
97. Unlu, M. Difference gel electrophoresis. *Biochem Soc Trans* **27,** 547–549 (1999).
98. Howell, J. M., Winstone, T. L., Coorssen, J. R. & Turner, R. J. An evaluation of in vitro protein-protein interaction techniques: assessing contaminating background proteins. *Proteomics* **6,** 2050–2069 (2006).
99. Harlow, E. & Lane, D. C. N.-J. or A. B. R. R. Q. . . H. 1999. *Using antibodies : a laboratory manual*. (Cold Spring Harbor Laboratory Press, 1999).
100. O'Farrell, P. H. High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* **250,** 4007–4021 (1975).
101. Atassi, M. Z. & Habeeb, A. F. [49] Reaction of proteins with citraconic anhydride. *Methods Enzym.* **25,** 546–553 (1972).
102. Perham, H. B. F. D. and R. N. Reversible Blocking of Amino Groups with Citraconic Anhydride,. *Biochem. J.* **109,** 312 (1968).
103. Nieto, M., Palacián, E., Angela Nieto, M. & Palacian, E. Effects of temperature and pH on the regeneration of the amino groups of ovalbumin after modification with citraconic and dimethylmaleic anhydrides. *Biochim. Biophys. Acta* **749,** 204–210

(1983).
104. J, K. A. & W, L. P. Structure and efficiency in intramolecular and enzymic catalysis. Catalysis of amide hydrolysis by the carboxy-group of substituted maleamic acids. *Perkin, J C S II* 1206–1214 (1971).
105. Lee, D. W. & Ha, H.-J. Selective Mono-BOC Protection of Diamines. *Synth. Commun.* **37,** 737–742 (2007).
106. Viswanathan, S., Unlu, M. & Minden, J. S. Two-dimensional difference gel electrophoresis. *Nat Protoc* **1,** 1351–1358 (2006).
107. Van, P. T. *et al.* In-gel equilibration for improved protein retention in 2DE-based proteomic workflows. *Electrophoresis* **In press,** (2014).
108. Van, P. T., Bass, V., Shiwarski, D., Lanni, F. & Minden, J. High dynamic range proteome imaging with the structured illumination gel imager. *Electrophoresis* **35,** 2642–2655 (2014).
109. Sellers, K. F., Miecznikowski, J., Viswanathan, S., Minden, J. S. & Eddy, W. F. Lights, Camera, Action! systematic variation in 2-D difference gel electrophoresis images. *Electrophoresis* **28,** 3324–3332 (2007).
110. von Grünigen, R. *et al.* Immunoassay with Captured Hapten. A Sensitive Gastrin Assay with Biotinyl-Gastrin Derivatives. *Biol. Chem. Hoppe & Seyler* **373,** 163–172 (1991).
111. Guo, Y., Punj, V., Sengupta, D. & Linstedt, A. D. Coat-tether interaction in Golgi organization. *Mol Biol Cell* **19,** 2830–2843 (2008).
112. Szodoray, P. *et al.* Anti-citrullinated protein/peptide autoantibodies in association with genetic and environmental factors as indicators of disease outcome in rheumatoid arthritis. *Autoimmun. Rev.* **9,** 140–143 (2010).
113. Vander Cruyssen, B. *et al.* Prediction models for rheumatoid arthritis during diagnostic investigation: evaluation of combinations of rheumatoid factor, anti-citrullinated protein/peptide antibodies and the human leucocyte antigen-shared epitope. *Ann. Rheum. Dis.* **66,** 364–9 (2007).
114. Shaw, M., Collins, B. F., Ho, L. A. & Raghu, G. Rheumatoid arthritis-associated lung disease. *Eur. Respir. Rev.* **24,** 1–16 (2015).
115. McInnes, I. B. & Schett, G. The pathogenesis of rheumatoid arthritis. *N Engl J Med* **365,** 2205–2219 (2011).
116. Yunt, Z. X. & Solomon, J. J. Lung Disease in Rheumatoid Arthritis. *Rheum Dis Clin North Am* **41,** 225–236 (2015).
117. Song, S. T. *et al.* Association of Single Nucleotide Polymorphisms of PADI4 and HLA-DRB1 Alleles with Susceptibility to Rheumatoid Arthritis-Related Lung Diseases. *Lung* **194,** 745–53 (2016).
118. Scally, S. W. *et al.* A molecular basis for the association of the HLA-DRB1 locus, citrullination, and rheumatoid arthritis. *J Exp Med* **210,** 2569–2582 (2013).
119. Koduri, G. *et al.* Interstitial lung disease has a poor prognosis in rheumatoid arthritis: Results from an inception cohort. *Rheumatology* **49,** 1483–1489 (2010).
120. Bongartz, T. *et al.* Incidence and mortality of interstitial lung disease in rheumatoid arthritis: a population-based study. *Arthritis Rheum.* **62,** 1583–91

(2010).
121. Hallowell, R. W. & Horton, M. R. Interstitial lung disease in patients with rheumatoid arthritis: spontaneous and drug induced. *Drugs* **74,** 443–50 (2014).
122. Archontogeorgis, K., Steiropoulos, P., Tzouvelekis, A., Nena, E. & Bouros, D. Lung cancer and interstitial lung diseases: A systematic review. *Pulm. Med.* **2012,** (2012).
123. Doyle, T. J. *et al.* A roadmap to promote clinical and translational research in rheumatoid arthritis-associated interstitial lung disease. *Chest* **145,** 454–463 (2014).
124. Harlow, L. *et al.* Anti-citrullinated heat shock protein 90 antibodies identified in bronchoalveolar lavage fluid are a marker of lung-specific immune responses. *Clin Immunol* **155,** 60–70 (2014).
125. Cavagna, L. *et al.* The multifaceted aspects of interstitial lung disease in rheumatoid arthritis. *Biomed Res. Int.* **2013,** (2013).
126. Demoruelle, M. K. *et al.* Brief report: airways abnormalities and rheumatoid arthritis-related autoantibodies in subjects without arthritis: early injury or initiating site of autoimmunity? *Arthritis Rheum* **64,** 1756–1761 (2012).
127. Lin, Y. T. & Tsai, T. H. Anti-CCP antibody in a patient with generalized discoid lupus erythematosus, vasculitis and acute interstitial lung disease but not rheumatoid arthritis. *J Eur Acad Dermatol Venereol* **28,** 826–828 (2014).
128. Fischer, A. *et al.* Lung disease with anti-CCP antibodies but not rheumatoid arthritis or connective tissue disease. *Respir Med* **106,** 1040–1047 (2012).
129. Tomioka, H. *et al.* Case of interstitial lung disease with anti-EJ and anti-CCP antibodies preceding rheumatoid arthritis. *Respir Investig* **50,** 66–69 (2012).
130. Bongartz, T. *et al.* Citrullination in extra-articular manifestations of rheumatoid arthritis. *Rheumatol.* **46,** 70–75 (2007).
131. Paulin, F., Doyle, T. J., Fletcher, E. A., Ascherman, D. P. & Rosas, I. O. Rheumatoid Arthritis-Associated Interstitial Lung Disease and Idiopathic Pulmonary Fibrosis: Shared Mechanistic and Phenotypic Traits Suggest Overlapping Disease Mechanisms. *Rev Invest Clin* **67,** 280–286 (2015).
132. Lee, H. K. *et al.* Histopathologic pattern and clinical features of rheumatoid arthritis-associated interstitial lung disease. *Chest* **127,** 2019–2027 (2005).
133. De Ceuleneer, M., Van Steendam, K., Dhaenens, M. & Deforce, D. In vivo relevance of citrullinated proteins and the challenges in their detection. *Proteomics* **12,** 752–60 (2012).
134. Nicholas, A. P. & Bhattacharya Editors, S. K. *Protein Deimination in Human Health and Disease*.
135. Vossenaar, E. R. Expression and activity of citrullinating peptidylarginine deiminase enzymes in monocytes and macrophages. *Ann. Rheum. Dis.* **63,** 373–381 (2004).
136. Kumar, S., Sethi, S., Irani, F. & Bode, B. Y. Anticyclic Citrullinated Peptide Antibody-Positive Paraneoplastic Polyarthritis in a Patient With Metastatic Pancreatic Cancer. *Am. J. Med. Sci.* **338,** 511–512 (2009).
137. Speicher, K. D., Kolbas, O., Harper, S. & Speicher, D. W. Systematic analysis of

peptide recoveries from in-gel digestions for protein identifications in proteome studies. *J. Biomol. Tech.* **11,** 74–86 (2000).
138. Chen, J. *et al.* Biomarkers of rheumatoid arthritis-associated interstitial lung disease. *Arthritis Rheumatol.* **67,** 28–38 (2015).
139. Diz, A. P., Truebano, M. & Skibinski, D. O. F. The consequences of sample pooling in proteomics: An empirical study. *Electrophoresis* **30,** 2967–2975 (2009).
140. Jiang, Z. *et al.* Investigating citrullinated proteins in tumour cell lines. *World J. Surg. Oncol.* **11,** 260 (2013).
141. Mansour, R. Ben *et al.* Increased levels of autoantibodies against catalase and superoxide dismutase associated with oxidative stress in patients with rheumatoid arthritis and systemic lupus erythematosus. *Scand. J. Rheumatol.* **37,** 103–108 (2008).
142. Zhang, K. Y. . *et al.* A Glutamine Switch Mechanism for Nucleotide Selectivity by Phosphodiesterases. *Mol. Cell* **15,** 279–286 (2004).
143. Zuccato, C., Valenza, M. & Cattaneo, E. Molecular mechanisms and potential therapeutical targets in Huntington's disease. *Physiol Rev* **90,** 905–981 (2010).
144. Milakovic, T., Quintanilla, R. A. & Johnson, G. V. Mutant huntingtin expression induces mitochondrial calcium handling defects in clonal striatal cells: functional consequences. *J Biol Chem* **281,** 34785–34795 (2006).
145. Giacomello, M., Oliveros, J. C., Naranjo, J. R. & Carafoli, E. Neuronal Ca(2+) dyshomeostasis in Huntington disease. *Prion* **7,** 76–84 (2013).
146. Choo, Y. S., Johnson, G. V, MacDonald, M., Detloff, P. J. & Lesort, M. Mutant huntingtin directly increases susceptibility of mitochondria to the calcium-induced permeability transition and cytochrome c release. *Hum Mol Genet* **13,** 1407–1420 (2004).
147. Bossy-Wetzel, E., Petrilli, A. & Knott, A. B. Mutant huntingtin and mitochondrial dysfunction. *Trends Neurosci* **31,** 609–616 (2008).
148. Trushina, E. *et al.* Mutant huntingtin impairs axonal trafficking in mammalian neurons in vivo and in vitro. *Mol Cell Biol* **24,** 8195–8209 (2004).
149. Chang, D. T., Rintoul, G. L., Pandipati, S. & Reynolds, I. J. Mutant huntingtin aggregates impair mitochondrial movement and trafficking in cortical neurons. *Neurobiol Dis* **22,** 388–400 (2006).
150. Dudek, J., Rehling, P. & van der Laan, M. Mitochondrial protein import: common principles and physiological networks. *Biochim Biophys Acta* **1833,** 274–285 (2013).
151. Yano, H. *et al.* Inhibition of mitochondrial protein import by mutant huntingtin. *Nat Neurosci* **17,** 822–831 (2014).
152. Landles, C. & Bates, G. P. Huntingtin and the molecular pathogenesis of Huntington's disease. Fourth in molecular medicine review series. *EMBO Rep* **5,** 958–963 (2004).
153. Trettel, F. *et al.* Dominant phenotypes produced by the HD mutation in STHdh(Q111) striatal cells. *Hum Mol Genet* **9,** 2799–2809 (2000).
154. Xie, J. *et al.* A two-dimensional electrophoretic map of human mitochondrial proteins from immortalized lymphoblastoid cell lines: a prerequisite to study

mitochondrial disorders in patients. *Proteomics* **5,** 2981–2999 (2005).
155. Li, Q. *et al.* ALS-linked mutant superoxide dismutase 1 (SOD1) alters mitochondrial protein composition and decreases protein import. *Proc Natl Acad Sci U S A* **107,** 21146–21151 (2010).
156. James, R. *et al.* Proteomic analysis of mitochondria in APOE transgenic mice and in response to an ischemic challenge. *J Cereb Blood Flow Metab* **32,** 164–176 (2012).
157. Fiorini, A. *et al.* Lack of p53 affects the expression of several brain mitochondrial proteins: insights from proteomics into important pathways regulated by p53. *PLoS One* **7,** e49846 (2012).
158. Zabel, C. *et al.* A large number of protein expression changes occur early in life and precede phenotype onset in a mouse model for huntington disease. *Mol Cell Proteomics* **8,** 720–734 (2009).
159. Perluigi, M. *et al.* Proteomic analysis of protein expression and oxidative modification in r6/2 transgenic mice: a model of Huntington disease. *Mol Cell Proteomics* **4,** 1849–1861 (2005).
160. Liu, X., Miller, B. R., Rebec, G. V & Clemmer, D. E. Protein expression in the striatum and cortex regions of the brain for a mouse model of Huntington's disease. *J Proteome Res* **6,** 3134–3142 (2007).
161. Sorolla, M. A. *et al.* Proteomic and oxidative stress analysis in human brain samples of Huntington disease. *Free Radic Biol Med* **45,** 667–678 (2008).
162. Sims, N. R. & Anderson, M. F. Isolation of mitochondria from rat brain using Percoll density gradient centrifugation. *Nat. Protoc.* **3,** 1228–1239 (2008).
163. Viswanathan, S., Unlu, M. & Minden, J. S. Two-dimensional difference gel electrophoresis. *Nat Protoc* **1,** 1351–1358 (2006).