**Driving Biological Project: Developing new efficient technologies for antibody sequencing**

Funding status: funded internally at Genentech

Collaborating investigators: Jennie Lill, Wendy Sandoval

Institution: Dept. of Protein Chemistry, Genentech, South San Francisco, CA

TRDP interactions:

This project focuses on collaborative drug discovery efforts that include de novo sequencing of monoclonal and polyclonal therapeutic antibodies. It also includes development of a new approach for sequencing antibodies from complex samples like serum using a combination of next generation sequencing and mass spectrometry.

1. **Specific Aims**
2. **Development of an automated low-cost mass spectrometry based approach for sequencing monoclonal antibodies**
3. **Development of a new technology that combines NGS and MS for sequencing antibodies from complex samples**
4. **Significance**

Over the past decade MAbs have been utilized as important therapeutic agents. Traditional workflows for monoclonal antibody characterization involve isolation and identification at the genetic level for gaining full sequence information. When both the DNA and hybridomas are not available it is customary in biotechnology industry (but very time-consuming) to sequence the entire antibody de novo. Previous approaches for antibody sequencing may require months of experiments and data analysis for the determination of the complete protein sequence. As antibodies contain highly heterogeneous regions within their variable regions, the standard tandem mass spectrometric workflow followed by database search does not apply. In 2008-2012, CCMS in collaboration Jennie Lill and David Arnott at Genentech developed a new protein sequencing approach that reduced the effort for antibody sequencing by an order of magnitude.

After exposure to an antigen, the human immune response generates a diverse repertoire of antibodies with each B-cell contributing a specific monoclonal antibody to the polyclonal immune response. However, polyclonal antibodies enriched for specific properties have been his­torically difficult to produce, limiting their value for biomedical applications. The hybridoma method provided a way to obtain monoclonal antibodies and opened the door to interrogating the immune response to an antigen. However, the existing technologies to obtain antigen-specific monoclonal antibodies remain labor intensive and the antibodies they generate do not necessarily represent the actual antibody repertoire. Thus, a method capable of interrogating the complexity of the circulating antibody repertoire elicited to specific antigens is needed.

An attractive alternative to existing antibody sequencing approaches is a direct investigation of the monoclonal composition of polyclonal antibodies in circulating serum using LC-MS/MS.

This technology may aid the development of vaccines and help us better understand the immune response. However, despite the success of CCMS and Genentech joint efforts to develop efficient approach for sequencing *purified* antibodies, this approach is difficult to implement for sequencing antibodies circulating in serum. Indeed, de novo sequencing of polyclonal antibodies is very challenging because of the high complexity of the polyclonal mixture and the lack of a reference database of the constantly evolving repertoire of antibodies.

In this project, we will use affinity purification (to reduce sample complexity), LC-MS/MS analysis of the resulting sample (to generate a spectral dataset representing peptides from the antibodies of interest), and DNA sequencing (to gener­ate a reference database derived from the B-cell repertoire). We will generate both bottom-up and top-down spectral datasets for follow-up peptide/protein identification. Bottom-up analysis will incorporate multiple overlapping enzyme digestions and triplet CID/HCD/ETD fragmentation of every MS precursor, which has recently been shown to enable long (60-200 AA) and accurate (99%) de novo sequencing [cite Triples paper]. DNA sequencing reads will be assembled to generate an animal-specific database of antibodies. The set of spectra and de novo sequences will be searched against this database to identify peptides/proteins (along with PTMs) and to further extract information about the composition of antibodies. Finally, heavy and light chain sequences will be paired and expressed as recombinant monoclonal antibodies.

This new technology enables direct investigation of circulating polyclonal antibodies purified from the serum and avoids the techni­cal challenges of existing methods. While it disrupts pairing of heavy and light chains, we will be able to identify monoclonal antibody pairs as described in the Approach section. An important advantage of the proposed method is an improvement in speed to obtain functionally validated mono­clonal antibodies.

1. **Innovation**

This DBP drives technology development in the following ways.

1. New MS-based de novo sequencing algorithms to handle mixtures of highly homologous proteins (e.g. proteins with single-AA polymorphisms and/or branching CDRs).
2. Combining NGS and MS for antibody sequencing opened a new era in antibody sequencing by providing an ability to sequence antibodies directly from complex samples.
3. Bringing top-down MS to antibody sequencing efforts provides an efficient way to close gaps in sequencing monoclonal antibodies and to move towards polyclonal antibody sequencing. This project will represent the first application of top-down mass spectrometry to full-length antibody sequencing.
4. **Approach**

We will first perform affinity purification of polyclonal material from a serum sample to enrich for monoclonal antibodies with desired functional properties, e.g., for high-affinity interactions.

Once the desired properties have been enriched, LC-MS/MS will be used to identify the monoclonal components of the purified fraction by matching to a custom reference database of antibody V regions produced by next-generation sequencing of the B cells. To generate cDNA library, we will isolate memory B cells from peripheral blood mononuclear cells to generate a cDNA library. The cDNA libraries from these cell pools is amplified by PCR (using primers selected from constant regions) followed by sequencing of heavy and light chain V regions using a Roche 454 and Illumina instruments. The NGS reads will be assembled into a custom protein database using software developed at CCMS. In contrast to hyperimmunized laboratory animals, translating this approach to the isolation of circulating human antibodies faces several challenges since accessing immunological organs that are highly enriched in B cells, such as the spleen, is not feasible in humans.

Bottom-up analysis will incorporate separate Trypsin, Chymotrypsin, AspN, GluC, ArgC, LysC, LysN, and Pepsin digestions to yield a wide distribution of overlapping peptides for MS/MS. Every MS precursor will then be subjected to CID, HCD, and ETD MS/MS fragmentation, which will provide corroborating evidence of peptide fragmentation and aid in the assembly of spectra from highly charged precursors by SPS and MetaSPS [cite MetaSPS]. Since MetaSPS was designed to sequence purified proteins or those from small un-related mixtures, we do not expect it to be fully sensitive in assembling antibody sequences in the presence of highly homologous mixtures. De novo sequences longer than 50 AA would still be useful for spanning unknown CDRs and connecting them to known V-regions via error-tolerant database matching. But if established approaches prove inadequate for generating long antibody sequences, we will orderly address the problem by (1) updating MetaSPS sequencing algorithms to process highly related mixtures, (2) incorporating top-down spectra to correctly order and extend de novo sequences, and (3) utilizing prior knowledge of known antibodies and NGS to develop a database-assisted assembly approach that can tolerate homologous mixtures.

We will next identify high-confidence V-region peptides overlapping with unique CDRs using newly developed software that integrates NGS and MS approaches to antibody analysis. To produce recombinant monoclonal antibodies, heavy and light chains will be combinatorially paired and expressed. The recombinant antibodies will be further screened to identify unique heavy and light chain pairs. We will further analyze how the obtained monoclonal antibodies react against diverse epitopes and evaluate their relative binding activity.

After testing this technology in rabbits, we will next identify neutralizing antibodies elicited against a pathogen from the plasma of an infected individual. We will initially focus on infections that have high prevalence in healthy adults. This will require designing a purification procedure for enriching specifically for neutralizing antibodies.

The antibodies we isolated using stringent purification method could represent only a fraction of

the total antigen-specific antibodies found in circulation due to a variety of reasons; (i) too stringent purification conditions for lower-affinity antibodies; (ii) limited B cell source for generation of NGS reads, and (ii) the limitations of detection by MS instruments. Functionally validated antibodies identified using the proposed approach can be used for further mining of additional clonally related antibody chains from the NGS database. These additional

antibody chains could have been missed owing to their low affinity or low abundance in serum.

1. **Summary**

In vaccine research, antibody titers measured by ELISA are often used to identify neutralizing epitopes. However, this techniques does not allow direct interrogation of the circulating antibody population. The proposed research will integrate our NGS-based and proteomics-based antibody discovery approaches to sequence antiviral monoclonal antibodies directly from serum. The proposed method could be used to track changes in the vaccine-specific circulating antibody repertoire over time and examine their correlation with the evolving individual antibodies. The combination of NGS-based and MS-based approaches also allows one to reconstitute a combined pool of neutralizing human monoclonal antibodies. A neutralizing mixture made by recombinant human monoclonal antibodies would provide an improved clinical tool for antiviral immunotherapy.

1. **Impact and synergy with TRD projects**