Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Chemical Engineering

Mapping serum antibody repertoires using peptide libraries

Michael L Paull and Patrick S Daugherty

Antibodies in blood provide a rich source of immunological information. Antibody repertoire analysis seeks to decode this information to empower the development of vaccines, diagnostics, and therapeutics. To this end, various approaches have been developed to determine epitopes using peptide libraries. Approaches have used random or proteome-derived peptide libraries in a microarray or surface display format. For methods using random libraries, motif discovery software has been developed to identify common binding signatures. The analysis of thousands of samples and dozens of diseases has shown that there are often disease-specific epitopes, even though individual antibody repertoires are unique. The recent developments in antibody repertoire analysis hold the potential to enable comprehensive immune evaluations.

Address
Department of Chemical Engineering, University of California, Santa Barbara, CA 93106, USA

Corresponding author: Daugherty, Patrick S (psdaug@gmail.com)

Current Opinion in Chemical Engineering 2018, **19**:21–26

This review comes from a themed issue on **Biological engineering**

Edited by **Sharon Gerecht** and **Jennifer A Maynard**

https://doi.org/10.1016/j.coche.2017.12.001

Introduction

Antibodies bind specifically to their targets and are therefore relevant to numerous scientific, medical, and industrial applications. The immune system continuously makes antibodies even after an infection has been resolved. Therefore, the antibodies in serum constitute an immunological record. The ability to access and interpret this record could impact many areas of biotechnology and healthcare. In particular, the identification of disease-associated antigens has enabled the development of numerous diagnostic tests for infectious [1], autoimmune [2], and allergic conditions [3]. Furthermore, epitope information can be used to inform the development of more efficacious vaccines [4]. With the growing number of antibody-based therapeutics, there is a need to characterize antibody binding to measure specificity and avoid undesired cross-reactivity [5]. Additionally, knowledge of antibody binding sites can enable the design of more effective affinity reagents for diverse applications [6]. Finally, antibody repertoire analysis methods will augment efforts to characterize the 'healthy' antibody repertoire which could be useful for detecting the onset of disease [7•]. With the recent development of high-density peptide microarrays, high-throughput sequencing, and increased computational power, there is increasing interest in antibody repertoire analysis.

Although the expression 'antibody repertoire' is frequently used, it is informative to divide its usage into methods focusing on paratopes or epitopes. The terms paratope and epitope refer to the binding regions of the antibody and antigen, respectively (Figure 1). Paratope-focused methods analyze antibody CDR regions through B-cell DNA sequencing and LC–MS/MS [8,9]. These methods have proven useful for monitoring the evolution of an immune response, determining which antibody clonotypes are most abundant, and investigating class switching [10]. Alternatively, methods that identify protein epitopes have the distinct advantages of requiring minimal serum, rather than B cells, and allowing for the identification and analysis of antigens. This review focuses on approaches that use peptide libraries to determine protein epitopes for the antibody repertoire.

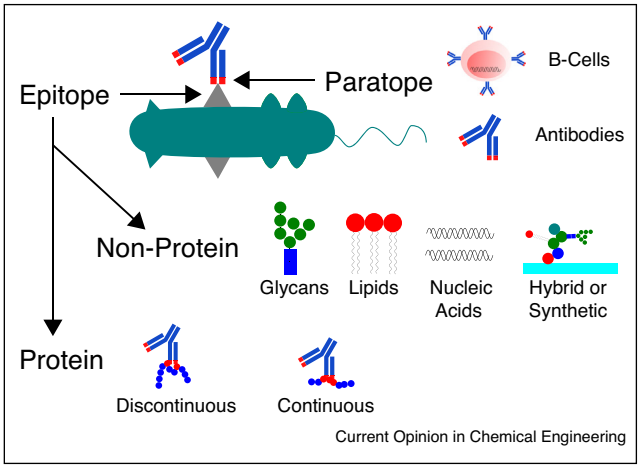
Protein epitope characteristics

Protein epitopes are typically considered to be continuous ('linear') or discontinuous ('conformational') (Figure 1). Continuous epitopes are comprised of a single sequence whereas discontinuous epitopes are comprised of amino acids distant in sequence, but close in the folded protein. For the cases in which the epitopes of interest are discontinuous, several methods have been developed [11–13]. It has been suggested that because >90% of epitopes are discontinuous, searching for continuous epitopes may be fruitless [14]. However, from an analysis of PDB antibody/antigen structures, it was determined that epitopes are generally composed of around 15 residues and that 85% of epitopes contain at least one five amino acid contiguous stretch [15]. Strictly speaking, nearly all epitopes are discontinuous, however, the frequent occurrence of linear segments suggests that there is utility in identifying linear epitopes [16]. And importantly, linear protein epitope discovery remains bioinformatically tractable.

Classification of antibody repertoire mapping approaches

Approaches for mapping the antibody repertoire use random or non-random peptide libraries in a microarray

Figure 1



An overview of the classification of antibody/antigen binding interactions. The paratope and epitope refer to the binding regions of the antibody and antigen, respectively. Paratope-focused methods focus on serum antibodies or antibody-producing B-cells. Epitopes can be comprised of non-protein molecules such as glycans, lipids, nucleic acids, combination structures, or synthetic molecules. Continuous protein epitopes are comprised of a single sequence whereas discontinuous protein epitopes are comprised of amino acids distant in sequence, but close in the folded protein.

or surface display format [17]. For microarrays, peptides are typically printed or synthesized on glass slides, whereas for surface display, peptides are most often displayed on bacteriophages, bacteria, or yeast. Relevant information for each approach referenced in this review can be found in Table 1.

Non-random methods often use libraries constructed by tiling target protein sequences into overlapping peptides. While the concept of determining epitopes by tiling antigens seems straightforward, there are important limitations. There is reason to believe that, paradoxically, random peptides may be able to capture binding specificities that the tiled antigens themselves cannot [18]. One potential explanation for this observation could be that an antibody binds to a peptide from a random library with higher affinity than the corresponding fragment of the antigen. Further, proteome-derived peptide libraries are typically several orders of magnitude smaller than random libraries (e.g. 10^5 [19••] versus 10^{11} [20]).

The use of random libraries allows for a less biased experimental approach to epitope determination. However, with random libraries, the burden shifts to computational motif discovery since it is necessary to resolve randomness into a coherent signal. Additionally, associating an epitope with an antigen becomes a challenge.

Microarrays are potentially more reproducible, less laborious, and more quantitative than surface display systems. However, microarrays usually display 3–5 orders of magnitude fewer peptides than surface display libraries, which can attain library diversities up to 10^{11}. A consequence is that microarrays may not contain enough information for certain applications. Also, surface display peptide libraries can be propagated by growth which reduces cost.

Non-random microarray methods

For non-random microarrays, antigens from a pathogen of interest or the human proteome are tiled into overlapping peptides. Multiple sclerosis (MS) autoantibodies have been examined in depth using a microarray with presumed MS autoantigens and Epstein–Barr Virus (EBV) antigens [21]. This analysis discovered peptides that were bound by the serum antibodies of MS subjects, but not by matched controls. A peptide containing 'RRPFF' from EBV Nuclear Antigen 1 (EBNA1) was stated to be disease-specific, however a study with a larger sample size showed that this epitope was prevalent in the general population [19••]. Antigens of the parasitic protozoan *Trypanosoma cruzi* have been tiled to analyze serum samples from subjects infected with Chagas disease [22•]. In this case, the use of high-density microarrays identified multiple new disease-specific peptides and antigens. However, only a fraction of *T. cruzi* antigens could be examined because of the large size of this parasite's proteome. These examples demonstrate that microarray size can restrict the number of serum samples and antigens that can be analyzed.

Another method used a human proteome microarray with six amino acid lateral shifts followed by a targeted microarray with only single amino acid shifts [23•]. This method identified two potential novel autoantigens for narcolepsy and multiple sclerosis. An approach for determining the fine specificity of epitopes used a non-random microarray, followed by an exhaustive mutagenesis scheme on selected epitopes [24]. The scheme was later refined and made available through the online server ArrayPitope [25].

Non-random surface display methods

An exemplar surface displayed non-random library is T7-Pep, in which peptides representing the human proteome are incorporated into a phage library [26]. Human proteome libraries are useful for probing autoantibodies using phage immunoprecipitation sequencing (PhIP-Seq). The initial application of this method identified candidate autoantigens in subjects with paraneoplastic syndromes. T7-Pep was also used for a large-scale PhIP-Seq screen of nearly 300 antibody repertoires from subjects with type 1 diabetes, multiple sclerosis, rheumatoid arthritis, and healthy controls [19••]. Most antibody-binding peptides were unique to individuals, suggesting that each antibody repertoire is unique. Even with large sample sizes, disease-specific peptides with high sensitivity were not found. Their absence may be due to disease heterogeneity or the inherent stochasticity of the humoral response.