Journal of Immunological Methods 443 (2017) 45–54

Contents lists available at ScienceDirect

Journal of Immunological Methods

journal homepage: www.elsevier.com/locate/jim

Research paper

Single amino acid fingerprinting of the human antibody repertoire with high density peptide arrays

Laura K. Weber [a], Andrea Palermo [a], Jonas Kügler [b], Olivier Armant [c,1], Awale Isse [a], Simone Rentschler [a], Thomas Jaenisch [d,g,h], Jürgen Hubbuch [e], Stefan Dübel [f], Alexander Nesterov-Mueller [a,*], Frank Breitling [a,*], Felix F. Loeffler [a,h,i,**]

[a] Karlsruhe Institute of Technology, Institute of Microstructure Technology (IMT), Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany
[b] Yumab GmbH, Rebenring 33, 38106 Braunschweig, Germany
[c] Karlsruhe Institute of Technology, Institute of Toxicology and Genetics (ITG), Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany
[d] Heidelberg University Hospital, Department for Infectious Diseases, Parasitology Unit, Im Neuenheimer Feld 324, 69120 Heidelberg, Germany
[e] Karlsruhe Institute of Technology, Institute of Engineering in Life Sciences, Section IV: Biomolecular Separation Engineering, Engler-Bunte Ring 3, 76131 Karlsruhe, Germany
[f] Technische Universität Braunschweig, Department of Biotechnology, Institute for Biochemistry and Biotechnology, Spielmannstr. 7, 38106 Braunschweig, Germany
[g] German Centre for Infection Research (DZIF), partner site Heidelberg, Germany
[h] HEiKA - Heidelberg Karlsruhe Research Partnership, Heidelberg University, Karlsruhe Institute of Technology (KIT), Germany
[i] Max Planck Institute of Colloids and Interfaces, Department of Biomolecular Systems, Am Mühlenberg 1, 14476 Potsdam, Germany

ARTICLE INFO

Article history:
Received 27 October 2016
Received in revised form 14 December 2016
Accepted 3 January 2017
Available online 3 February 2017

Keywords:
Humoral immune system
Substitution analysis
Serology
Phage display
Next generation sequencing

ABSTRACT

The antibody species that patrol in a patient's blood are an invaluable part of the immune system. While most of them shield us from life-threatening infections, some of them do harm in autoimmune diseases. If we knew exactly all the antigens that elicited all the antibody species within a group of patients, we could learn which ones correlate with immune protection, are irrelevant, or do harm. Here, we demonstrate an approach to this question: First, we use a plethora of phage-displayed peptides to identify many different serum antibody binding peptides. Next, we synthesize identified peptides in the array format and rescreen the serum used for phage panning to validate antibody binding peptides. Finally, we systematically vary the sequence of validated antibody binding peptides to identify those amino acids within the peptides that are crucial for binding "their" antibody species. The resulting immune fingerprints can then be used to trace them back to potential antigens. We investigated the serum of an individual in this pipeline, which led to the identification of 73 antibody fingerprints. Some fingerprints could be traced back to their most likely antigen, for example the immunodominant capsid protein VP1 of enteroviruses, most likely elicited by the ubiquitous poliovirus vaccination. Thus, with our approach, it is possible, to pinpoint those antibody species that correlate with a certain antigen, without any pre-information. This can help to unravel hitherto enigmatic diseases.

1. Introduction

Evolution designed our antibody-based immune system to tackle infectious diseases, which is shown by the fact that we are likely to die – especially from infections with encapsulated bacteria – if our humoral immune system is compromised (Fried and Bonilla, 2009). Depending on his or her individual history of infections, every person clonally expands presumably thousands of different antibodies producing B cell clones, which secrete the bulk of the 11 mg IgG antibodies per mL blood serum (Gonzalez-Quintela et al., 2008).

These "amplified" antibody species are selected from a plethora of different antibodies. The antibody repertoire stems from the random combination of gene segments and clonal selection, which is triggered by specific binding to a pathogen's antigen. Eventually, most of these selected antibody species are further refined in their binding characteristics by random point mutations and somatic hypermutation (Murphy and Weaver, 2016).

In today's typical antibody-based diagnostics, the disease status is concluded from the binding of a patient's serum antibodies to one or a few proteins (Borrebaeck, 2000). This protein-specific binding is possible despite the randomness in antibody generation: Immunoblotting techniques validate that only a limited number of proteins of a pathogen

* Corresponding authors.
** Correspondence to: F.F. Loeffler, Karlsruhe Institute of Technology, Institute of Microstructure Technology (IMT), Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany.
E-mail addresses: alexander.nesterov-mueller@kit.edu (A. Nesterov-Mueller), frank.breitling@kit.edu (F. Breitling), felix.loeffler@mpikg.mpg.de (F.F. Loeffler).
[1] Novel address: Institut de Radioprotection et de Sureté Nucléaire (IRSN), PRP-ENV/SERIS/LECO, Cadarache, Saint-Paul-lez-Durance 13115, France.

http://dx.doi.org/10.1016/j.jim.2017.01.012
0022-1759/

are targeted (Haury et al., 1994; Stahl et al., 2000). Thus, the immune systems of different individuals seem to commonly prefer only a few proteins out of thousands of potential antigens of a bacterial pathogen or in autoimmune disease. However, quite a few basic scientific questions are still unanswered: Do the immune systems of different patients not only target the same proteins, but also similar epitopes? And, more importantly, could we get clues about the cause of an enigmatic disease from a determination of all epitopes of the "amplified" antibodies?

There have been many different approaches to analyze the diversity of a patient's individual antibody repertoire. In 2009, Weinstein et al. (Weinstein et al., 2009) used high-throughput sequencing to find out that the majority (>50%) of all possible V, D, J gene segment combination is indeed present in zebrafish to constitute its immunoglobulin diversity, adding support to the idea of an essentially random generation of antibody species. Recently, high-throughput DNA sequencing of immunoglobulin genes (Ig-seq) has been introduced, which allows for the quantitative read out of molecular information of the humoral immunity (Georgiou et al., 2014; Khan et al., 2016). However, both techniques rely on the conclusion that the full antibody repertoire can be deduced from the gene expression at a certain time point. This expression might be somewhat different to the actual diversity and specificities of antibodies.

Functional screens have been used to directly determine the antigen targets. Crompton et al. used protein arrays to find out that up to 41% of the displayed proteins from the malaria pathogen *Plasmodium falciparum* are targeted by the serum antibodies of patients (Crompton et al., 2010). By focusing antigenic attention upon irrelevant or highly variable epitopes, also referred to as deceptive imprinting (Tobin et al., 2008), the malaria pathogen makes it extraordinarily challenging to develop an efficient and targeted vaccine. The Johnston group used spotted arrays with some 10,000 random peptides that were stained with patient sera to discover patterns of stained peptides that indicated a growing cancer, a vaccination, or an infection (Legutki et al., 2010; Legutki and Johnston, 2013; Stafford et al., 2014). In a similar approach, Wang et al. used phage displayed peptide libraries (Smith, 1985; Devlin et al., 1990; Scott and Smith, 1990) to identify a pattern of targeted peptides that indicated a growing prostate cancer (Wang et al., 2005). Xu et al. generated a phage library, displaying proteome-wide peptides from all known human viruses, to discover viral epitopes that are targeted by serum antibodies from human sera (Xu et al., 2015). Screening random peptide libraries with mouse or human serum antibodies also allowed to distinguish healthy mice from mice that were infected with helminth parasites (Bongartz et al., 2009) or to pinpoint rheumatoid arthritis specific antigens (Dybwad et al., 1993). Moreover, Reineke et al. used the SPOT method from Ronald Frank (Frank, 1992) to synthesize arrays with some 5500 random peptides to identify linear peptides, which served as surrogate binders ("mimotopes") for known conformational antibody epitopes (Reineke et al., 2002).

However, two technical bottlenecks still have to be overcome before we can access the complete information embodied in antibodies patrolling in a patient: (i) the synthesis of peptide arrays is still expensive, and (ii) we would need to know exactly which amino acids within a binding peptide are crucial for the binding of many different antibody species. The second point is necessary to query data bases for potential target proteins of serum antibodies, which might play a role in an enigmatic disease.

Here, we combined next generation sequencing of a phage display peptide library (Matochko et al., 2012) with our novel solid-material-based synthesis method for high-density peptide arrays, enabling densities of over 10,000 peptide spots per cm^2 (Beyer et al., 2007; Stadler et al., 2008; Loeffler et al., 2012; Maerkle et al., 2014; Loeffler et al., 2016). For the latter, either a laser printer (Stadler et al., 2008) or a laser scanning system (Loeffler et al., 2016) is used to structure a synthesis surface with ≥20 different amino acid building blocks, which are embedded in a solid matrix material. Then, a simple heating step starts the coupling reaction by melting the matrix material. Parallel synthesis on the same array surface is achieved by repeating these steps for the different amino acid building blocks, until the desired peptide length is reached. This large number of peptides not only allows for an almost comprehensive coverage of entire genomes but also for substitution analyses: Systematically substituting every amino acid position in a sequence of a specific antibody-binding peptide with every other amino acid, it is possible to identify the key residues for each epitope. This approach allows us to evaluate the specificity of an antibody down to the individual amino acid level and refines the profiling of the antibody repertoire (see workflow in Fig. 1). It reveals the original epitopes of a patient's antibodies or the mimotopes, and can also uncover potential cross-reactivity (Reineke et al., 2002). Thus, besides gaining knowledge on the antibody repertoire, our principle may be useful for elucidating clinical phenotypes with unknown disease antigens, without any *a priori* knowledge.

2. Results

In this study, we investigated the potential of a novel three step screening pipeline to comprehensively read out the antibody specificities in the serum of a donor subject. We combined phage display prescreening with a subsequent two-step peptide microarray analysis. First, a serum was analyzed using an epitope phage display panning followed by DNA sequencing. Then, the resulting binders were synthesized on laser-generated high diversity peptide arrays and incubated with the original serum. Subsequently, epitopes of identified binders were synthesized again on laser-generated high diversity peptide arrays, but this time with variants representing a complete substitution analysis set: We generated $12 \times 19 = 228$ individual variants of each 12mer peptide, where each amino acid position of the original sequence is substituted one by one with all other 19 amino acids. These analyses resulted in 73 antibody fingerprints. We could identify four different motifs that were present in several peptide binders. These four motifs were used to query public protein databases.

2.1. Validation arrays

Phage display with 12mer peptide presenting phage was carried out in three panning rounds. The peptide encoding DNA fragments of bound phage of panning rounds two and three were sequenced and *in silico* translated into amino acid sequences. For both panning rounds together, the sequencing resulted in a total of 38,533 different 12mer peptides. These peptides were synthesized on 10 different arrays with up to 4128 peptides as spot duplicates per array and incubated with the serum. Fig. 2A shows the fluorescence scan of an array that was stained with the serum.

The fluorescence intensity of each peptide was analyzed by calculating the average of spot duplicates. For each array, a threshold was defined. Peptides showing higher intensities than this threshold were further analyzed in the following substitution analysis. Fig. 2B shows the intensity distribution plot of the array in Fig. 2A. The threshold was set to an intensity value of 1000 a.u. and at least 0.15% of the peptides of each slide (4128 peptides) with the highest fluorescence intensity were substituted. The screens resulted in 97 peptides which exceeded the threshold.

2.2. Substitution analysis

For the substitution analysis, each of the 97 validated 12mer binders was synthesized in $12 \times 19 = 228$ peptide variants and 12 original sequences as 240 duplicate spots and incubated again with the serum. We screened a total of 23,280 different peptide double spots, resulting in 73 distinct fingerprints. Fig. 3 shows the fluorescence scan of a comprehensive substitution analysis of a typical peptide in spot duplicates. Each row represents the substitution at one position of the original peptide with 19 different amino acids (plus the original one), whereas each column corresponds to one of the 20 amino acids in alphabetical order