Deciphering antibody fingerprints with high density peptide arrays:

From methodological development toward disease etiology

zur Erlangung des akademischen Grades eines
DOKTORS DER INGENIEURWISSENSCHAFTEN (Dr.-Ing.)

von der Fakultät für Chemieingenieurwesen und Verfahrenstechnik des
Karlsruher Instituts für Technologie (KIT)

genehmigte
DISSERTATION

von
Dipl.-Ing. Laura Weber
geboren in Pforzheim

Referent:	Prof. Dr. Jürgen Hubbuch
Korreferent:	Prof. Dr. Frank Breitling
Tag der mündlichen Prüfung:	20.10.2017

Danksagung

An dieser Stelle möchte ich mich bei allen bedanken, ohne die die Durchführung dieser Arbeit niemals möglich gewesen wäre und deren Unterstützung mir wertvolle Hilfe war und ist. Besonders danke ich Prof. Jürgen Hubbuch & Prof. Frank Breitling für die Themenstellung und die Möglichkeit, an einem so spannenden Projekt zu forschen. Danke, für die vielen Ratschläge, die Betreuung und das Vertrauen.

Allen Kollegen und Mitdoktoranden der Abteilung „Peptidarrays" und allen Mitarbeitern des IMT danke ich für die gute Zusammenarbeit in den letzten 4 Jahren. Besonders Andrea und Felix danke ich, dass sie in der Zeit am IMT zu Freunden wurden, bessere widersprechende Büro- und Dienstreisekollegen kann man sich nicht wünschen. Miriam danke ich für ihre technische Assistenz und Unterstützung, besonders in der holprigen Anfangszeit. Dario danke ich für die Überbrückung finanzieller Engpässe und mexikanische Buschtouren.

Mein Dank gilt auch der PEPperPRINT GmbH, insbesondere Thomas Felgenhauer, für den wertvollen arrayartigen Beitrag zu dieser Forschungsarbeit.

Bei Jonas Kügler, Stefan Dübel, Olivier Armant, Volker List, Richard Kneusel und Ralf Bischoff bedanke ich mich für die Zusammenarbeit in den verschiedenen Kooperationsprojekten, für ihre wissenschaftliche Expertise, die zahlreichen Experimente und Patientenseren.

Meinem TAC von BIF IGS gebührt Dank für die wissenschaftliche Begleitung meiner Arbeit: Jürgen Hubbuch, Frank Breitling, Felix Löffler, Pavel Levkin und Ralf Bischoff.

Danken möchte ich auch den Studenten, die ich während meiner Doktorarbeit betreuen durfte und die maßgeblich am Fortschritt dieser Arbeit beteiligt waren: Simone, Awale, Marina, Christoph, Ratri, Tim und Ekaterina.

Silla und Moritz danke ich für die vielen gemeinsamen Stunden auf dem Fahrrad und der Freundemensa danke ich für zahlreiche spaßige Mittags- und Kaffeepausen.

Der größte Dank gebührt meinen Freunden, meinen Eltern und Geschwistern. Danke, dass ihr immer für mich da seid und mich in allen Lebenslagen begleitet!

Zusammenfassung

Antikörper sind ein überlebenswichtiger Teil unseres Immunsystems. Sie machen Eindringlinge unschädlich und schützen uns so Tag für Tag vor Krankheitserregern. Die Antikörper unseres humoralen Immunsystems können aber auch Schaden anrichten. In Autoimmunkrankheiten oder bei Gewebetransplantationen markieren sie körpereigene Strukturen, die daraufhin vom Immunsystem angegriffen werden.

Alle diese Antikörper werden – bei Bedarf – von spezialisierten Plasmazellen in unser Blut sezerniert. Die Plasmazellen entstehen aus unreifen B Lymphozyten, die einen Reifungsprozess durchlaufen müssen, dann positiv, d.h. durch die Bindung an ein Antigen, selektiert und schließlich zur Teilung angeregt werden. Letztendlich sezerniert jede dieser Plasmazellen große Mengen identischer Antikörper mit einzigartiger Spezifität. Durch den Kontakt mit einem Antigen wird außerdem bei vielen B Lymphozyten eine somatische Hypermutation aktiviert, die letztendlich zu einer erhöhten Affinität des Antikörpers für sein Antigen führt. Bemerkenswert ist, dass die Menge der einzelnen Serumantikörper sehr variabel ist und sich so an geänderte Bedingungen anpasst (z.B. eine schwere Infektion oder ein Gewebetrauma). Dies ist darin begründet, dass in unserem Serum nur für eine begrenzte Zahl verschiedener Antikörper in therapeutischer Konzentration „Platz" ist.

Bisher war es aufgrund der großen Vielfalt einander sehr ähnlicher Antikörperspezies nicht möglich herauszufinden, genau welche Antikörperspezies in welchem Mengenverhältnis im Blut eines Patienten vorliegen und durch welches Ereignis diese entstanden sind. Wäre dies möglich, könnten durch den Vergleich von verschiedenen Patientenseren oder Seren von verschieden Zeitpunkten neu gebildete Antikörper (z.B. nach einer Infektion) identifiziert werden. Außerdem könnte untersucht werden, welche Antikörperspezies bei einem noch nicht verstandenen Krankheitsbild auftauchen oder welche Antikörper mit Schutz vor bestimmten Krankheiten korrelieren (die entsprechenden Antigene wären Kandidaten für Impfstoffe). Dafür müssen ohne Vorinformation über mögliche Krankheiten möglichst viele eindeutig voneinander unterscheidbare Antikörperspezies im Serum eines Patienten identifiziert werden.

Die vorliegende Arbeit soll die Voraussetzungen schaffen, genau dies in Zukunft erforschen zu können und beschäftigt sich daher mit der Untersuchung von Antikörper-Antigen Wechselwirkungen und die detaillierten Auswirkungen einzelner Aminosäuren auf die Bindung. Durch diese Bindungsuntersuchung können Spezifitäten von Antikörpern, die sowohl das Antigen als auch den Antikörper zu charakterisieren beziehungsweise zu identifizieren vermögen, sehr genau bestimmt werden.

Um die Spezifitäten möglichst vieler Antikörper zu bestimmen, eignen sich Hochdurchsatzverfahren wie beispielsweise Peptidarrays, die Peptide stellen dabei kurze Fragmente der Antigene nach. Peptidarrays haben den Vorteil, dass die Antikörper-Antigen Wechselwirkung in sehr großem Maßstab, mit geringem Probenverbrauch und minimaler Versuchsdauer pro Interaktion untersucht werden kann. Je nach Herstellungsmethode können bis zu mehrere Tausend Peptide auf einem Objektträger im Arrayformat präsentiert werden. Während die Untersuchung von Antikörper-Epitopen mit Peptidarrays bereits routinemäßig in der Ursachenforschung von Krankheiten und in der Impfstoffentwicklung eingesetzt werden, gab es bisher keine Vergleiche der exakten Spezifität von Antikörpern mehrerer Individuen. Um diese exakten Spezifitäten zu bestimmen, wird jede Aminosäure in einem gefundenen Epitop in einer sogenannten Substitutionsanalyse durch alle anderen Aminosäuren ersetzt und die Interaktion der neuen Peptide mit dem Antikörper untersucht. Dieser Ansatz ermöglicht es, die Rolle der einzelnen Aminosäuren in der Bindung zu analysieren und damit den „Fingerabdruck" der Antikörper zu bestimmen. Mit dem Fingerabdruck, also den bindungsrelevanten Aminosäuren, kann in öffentlichen Protein-Datenbanken nach induzierenden und kreuzreagierenden Antigenen gesucht werden und damit können erstmals neue Einblicke in dieses Gebiet gewonnen werden.

Im Fokus von drei Manuskripten stehen Antikörperspezifitäten verschiedener Patienten, die mit Substitutionsanalysen unter Verwendung von Peptidarrays bestimmt werden können. Zwei Manuskripte zeigen zum ersten Mal, dass die für eine Antikörper-Antigen Bindung essentiellen Aminosäuren krankheitsspezifisch und in verschiedenen Patienten nahezu identisch sind. Im dritten Manuskript wird evaluiert, ob eine große Peptidbibliothek in Kombination mit der Substitutionsanalyse geeignet ist, das Antikörperrepertoire eines Menschen ohne *a priori* Wissen auszulesen.

Die Analysen erfolgten mit Peptidarrays der Firma PEPperPRINT GmbH aus Heidelberg. Die Synthese der Peptide an der Oberfläche von Objektträgern, auf denen ein Polymerfilm verankert ist, erfolgt dabei mit einem partikelbasierten Festphasenansatz nach Merrifield. Fmoc-geschützte und Opfp-aktivierte Aminosäurederivate sind in Polymerpartikel eingebettet

und werden mittels einer Art Laserdrucker auf einer Oberfläche strukturiert. Nach der Strukturierung aller 20 Aminosäuren in der ersten Lage, erfolgt die chemische Synthese. Diese Entkopplung von Strukturierung und chemischer Verarbeitung beschleunigt den Herstellungsprozess enorm. Für Peptide mit einer Länge von 15 Aminosäuren sind so nur 15 chemische Syntheseschritte notwendig. Arrays, die auf diese Art hergestellt werden, weisen eine Dichte von bis zu 800 Peptiden pro cm^2 auf. Neu entwickelte Verfahren erlauben die Erhöhung dieser Dichte auf bis zu 10.000 Peptidspots pro cm^2 sowie den Einbau von Sonderbausteinen wie phosphorylierte Aminosäuren und das Zyklisieren der Peptide.

Tetanus dient in dieser Arbeit aufgrund der hohen Impfrate als Modellsystem, um die Spezifitäten von Antikörpern in verschiedenen Individuen zu vergleichen. Die Impfung gegen Tetanus löst eine Immunantwort aus, die sehr zuverlässig und langfristig das Toxin zu neutralisieren vermag. Es handelt sich um ein extrazelluläres Toxin, das vom Bakterium *Clostridium tetani* produziert wird. Da noch nicht gänzlich geklärt ist, wie neutralisierende Antikörper das Toxin unschädlich machen, ist es außerdem unerlässlich, die Immunantwort genauer zu untersuchen. Für die Untersuchung der Immunantwort auf die Impfung gegen Tetanus wurde zunächst die lineare Aminosäuresequenz des Toxins in überlappende Peptide zerschnitten. Die Peptide haben dabei eine Länge von 15 Aminosäuren mit einer Überlappung von 14 Aminosäuren. Nach der Synthese der Peptide im Arrayformat wurden diese mit den Seren von 19 geimpften Europäern inkubiert. Der gesamte Vorgang wird als „Epitope Mapping" bezeichnet und ermöglicht es, Antikörper-Epitope zu identifizieren. Die 19 Individuen zeichneten sich durch eine breite Antikörperantwort aus. Da alle Seren einen ausreichenden Impfschutz gegen Tetanus aufwiesen, ist anzunehmen, dass verschiedenen Epitope eine neutralisierende Antikörperantwort hervorzurufen vermögen. Jedoch ist nicht auszuschließen, dass die breite Antikörpervielfalt nur bei einer Detektion mit linearen Peptiden auftritt und die Untersuchung der Antikörperantwort mit proteinähnlichen Strukturen ein anderes Ergebnis aufweist. Auffallend häufig wurde ein Epitop auf der C-terminalen Domäne der schweren Kette des Tetanustoxins detektiert. Eine anschließende Substitutionsanalyse dieses Epitops zeigte, dass die Fingerabdrücke der Antikörper, in 8 von den beobachteten 9 Fällen nahezu identisch sind. Im Hinblick auf die zufällige Generierung des Antikörper-Repertoires ist dieses Ergebnis erstaunlich. Aminosäuren, die alternativ zur originalen Aminosäure ebenso eine Bindung des Antikörpers auslösten, waren in allen untersuchten Proben sehr ähnlich und wiesen häufig einen ähnlichen chemischen Charakter auf wie die originale Aminosäure. Der entsprechende Antikörper wurde aus einem Serum mittels Affinitätschromatographie im Batch-Verfahren isoliert und eine Reaktion mit dem nativen Toxoid im ELISA konnte nachgewiesen werden. Ob die identifizierte

Antikörperspezifität neutralisierende Eigenschaften in Bezug auf das Tetanustoxin aufweist, muss in weiteren Studien geklärt werden. Es konnte jedoch zum ersten Mal gezeigt werden, dass verschiedene Personen auf Sub-Antigen- und sogar Sub-Epitopebene, trotz zufälliger Bildungsprozesse, nahezu identische Antikörperspezifitäten ausbilden.

Borreliose ist seit dem ersten gehäuften Auftreten in den 1970ern in Lyme, Connecticut, eine der häufigsten von Zecken übertragenen Krankheiten und gleichzeitig ein sehr kontroverses Thema. Es ist umstritten, ob späte Manifestationen der Krankheit durch autoimmune Reaktionen oder schwer diagnostizierbare, dauerhafte Infektionen mit den verursachenden Bakterien der Gattung *Borreliella*, ausgelöst werden. Auch die Diagnose, die in den meisten Fällen auf dem Nachweis von Serumantikörpern beruht, ist problematisch auf Grund von mangelnder Spezifität und/oder Sensitivität. Das immunogene Lipoprotein VlsE (*vmp*-like sequence expressed), das an der Oberfläche der Bakterien exprimiert wird, hat sich als hilfreich in der Diagnose erwiesen. Das Protein teilt sich in 3 Domänen, von denen eine variable Domäne in der Mitte von 2 konservierten Domänen am C- und N-Terminus flankiert wird. Die variable Domäne setzt sich dabei aus 6 variablen Bereichen und 6 konservierten Regionen zusammen. Ziel der im Manuskript beschriebenen Studie war es, die Fingerabdrücke der gegen VlsE induzierten Antikörper von an Borreliose erkrankten Patienten zu vergleichen. Hierfür wurde, wie im vorherigen Abschnitt beschrieben, ein „Epitope Mapping" des VlsE Proteins mit anschließender Substitutionsanalyse der immunogenen Peptide durchgeführt. Die Peptide wurden mit den Seren von 17 VlsE positiven und 7 VlsE negativen Patienten inkubiert. Anhand ihrer Antikörperantwort, die entweder gegen Epitope auf der variablen oder konservierten Domäne gerichtet war, konnten 13 der 17 positiven Patienten in zwei Gruppen eingeteilt werden. Die Substitutionsanalyse zeigte, dass die für die Bindung essentiellen Aminosäuren in den Epitopen, auch auf der variablen Domäne, in den konservierten Bereichen liegen. Zudem konnte gezeigt werden, dass nahezu identische Aminosäuren bindungsrelevant sind, falls das gleiche Epitop erkannt wird. Die Suche nach kreuzreagierenden humanen und viralen Antigenen in Protein-Datenbanken zeigte, dass lange Fingerabdrücke spezifisch für *B. burgdorferi* sind. Die mit Substitutionsanalysen bestimmten Antikörperspezifitäten könnten als Biomarker eingesetzt werden, mit dem Vorteil der Vermeidung von kreuzreagierenden Signalen und daraus folgenden Falsch-Positiven Ergebnissen.

In einer weiteren durchgeführten Studie wurden mit einer Kombination aus Screening mit Phage Display und Next Generation Sequencing sowie Validierung und Substitutionsanalyse mit Peptidarrays ohne Vorwissen 73 Antikörperspezifitäten bestimmt. Eines der dominanten

Motive konnte mit Hilfe einer Proteindatenbank dem antigenen Hüllprotein VP1 der Enteroviren zugeordnet werden. Aufgrund der weit verbreiteten Impfung gegen das Poliovirus aus dieser Familie ist es sehr wahrscheinlich, dass ein entsprechender Antikörper im Blut vorlag und identifiziert werden konnte. Experimente mit dem Serum eines weiteren Individuums zeigten, dass die Antikörperspezifitäten nicht nur krankheits-, sondern auch patientenspezifisch sind.

Die vorliegenden Manuskripte zeigen demnach, dass Fingerabrücke von Antikörpern, wie sie mit Substitutionsanalysen auf Peptidarrays bestimmt werden können, biologisch relevant sind und die entwickelten Methoden zur Erforschung von Infektionskrankheiten und der Impfstoff- und Biomarkerentwicklung beitragen.

Peptidarrays erlauben im Allgemeinen allerdings nur eine Aussage darüber, ob eine Bindung des Antikörpers an ein Peptid stattfindet, die Art der Interaktion kann nicht beurteilt werden. Dies ist darin begründet, dass die Bindung des Antikörpers an das Peptid nach Eintreten des Gleichgewichtszustands betrachtet wird. Falls eine Interaktion stattfindet, wird diese nach dem Prinzip eines Immuno-Assays mit fluoreszenzmarkierten Sekundärantikörpern nachgewiesen. In vielen Fällen sind allerdings auch die Affinität des Antikörpers zu seinem Antigen und die vorliegende Konzentration von Interesse. Es ist daher notwendig, den Vorgang der Antikörper-Peptid Bindung bereits vor Erreichen des Gleichgewichtszustands zu beobachten. Hierfür muss die Antikörperlösung kontinuierlich mit den Peptiden im Arrayformat in Kontakt gebracht werden, was in einem mikrofluidischen Kanal realisiert werden kann. Des Weiteren muss die Antikörper-Peptid-Wechselwirkung für die Dauer des Bindungsvorgangs sichtbar gemacht werden, wofür ein optisches System notwendig ist.

Zu diesem Zweck wurde in der vorliegenden Arbeit zunächst die Basis für diese Messungen geschaffen. Eine mikrofluidische Anlage für die Inkubation des Peptidarrays wurde entwickelt und ein optischer Aufbau aufgebaut. Dieser detektiert auf Basis einer fluoreszenzanregenden LED und einer CCD Kamera zeitaufgelöst die Anlagerung von Antikörpern an Peptide. Das System wurde getestet, indem der monoklonale FLAG-M2 Antikörper im mikrofluidischen Kanal mit dem FLAG-Epitop in Kontakt gebracht und die Anlagerung mit dem optischen System detektiert wurde. Aus den Messungen mit unterschiedlichen Konzentrationen konnte eine Assoziationskonstante bestimmt werden. Dies erlaubt in Zukunft möglicherweise auch die Analyse von Affinitäten oder Konzentrationen von Antikörpern in Patientenseren, was zu einer verbesserten Diagnostik und Forschung beitragen kann.

Abstract

Antibodies against pathogenic agents shield us routinely from infections. They specifically incapacitate intruders and thus ensure our survival. At the same time, antibodies of our humoral immune system can do harm in autoimmune diseases or cause transplant rejections.

Antibodies are secreted into our blood system by plasma cells. These plasma cells develop from immature B lymphocytes, which undergo maturation, positive selection by binding to an antigen and processes that stimulate cell division. Each of these plasma cells then produces thousands of identical antibodies with the same unique specificity. After contact with an antigen, most B lymphocytes additionally undergo somatic hypermutation, which leads to increased affinity of the antibody to its antigen. Remarkably, the amount of the individual antibodies is highly flexible and adapts to modified conditions (for instance an infection or tissue trauma). The reason for this is the limited space in our serum for a maximum number of different antibodies.

Due to the large variety of very similar antibody species, it was not possible until now to find out precisely which antibody species are present in which quantity in the blood of a patient and by what events these have been induced. If this was possible, newly formed antibodies (e.g. following an infection) could be identified by comparing different patient sera or sera from different time points. In addition, it could be investigated which antibody species are present in a hitherto enigmatic disease pattern or which antibodies correlate with the protection against a disease (the corresponding antigens would be candidates for vaccines). For these investigations, as many clearly distinguishable antibody species as possible must be identified in the serum of a patient without prior information of possible diseases.

The aim of this doctoral thesis is to create the prerequisites for exploring the stated topics and, therefore, to study the effects of antibody-antigen interactions in single amino acid resolutions. Antibody specificities that have the capacity of characterizing and identifying both antibody and antigen are in the focus of the investigations.

In order to determine the specificities of as many antibodies as possible, high-throughput tools such as peptide arrays are particularly suitable. The peptides mimic fragments of antigens and the specificity of many different antibodies can be precisely determined in

parallel on a very large scale. Up to several thousand peptides can be investigated in a single experiment with low sample consumption and a minimum of time per investigated interaction. While investigations of antibody epitopes with peptide arrays are already routinely used in disease research and vaccine development, no attempts have been made to compare the exact specificities of antibodies in different individuals. To determine the exact specificity in a so-called substitution analysis, each amino acid in an identified epitope is sequentially replaced by all other amino acids, generating many different variants of the epitope. Then, each peptide variant is examined for antibody binding. With this approach, it becomes possible to analyze the role of individual amino acids regarding their importance for antibody binding to immunodominant epitopes and thus to determine the fingerprint of an antibody. The knowledge about binding relevant amino acids of the fingerprints enables queries of public protein databases for the identification of inducing or cross-reacting antigens and insights into this research area can be gained for the first time.

The three following manuscripts focus on antibody fingerprints, determined by substitution analyses with single amino acid resolution. Two manuscripts demonstrate for the first time that the amino acids, essential for the antibody-antigen interaction, are disease-specific and almost identical in different patients. The third manuscript evaluates whether a large peptide library in combination with the substitution analysis is suitable for exploring the antibody repertoire of a person without *a priori* knowledge.

The analyses were carried out using peptide arrays synthesized by the company PEPperPRINT GmbH from Heidelberg, Germany. The peptides are synthesized combinatorically on the solid surface, on which a polymer film is anchored, with a particle-based solid-phase approach according to Merrifield. In brief, Fmoc-protected and Opfp-activated amino acid derivatives are embedded in polymer particles and patterned on a surface by a device similar to a standard laser printer. After structuring all 20 amino acids in the first layer, the chemical synthesis takes place. The separation of structuring and chemical processing accelerates the manufacturing process enormously. Arrays produced in this manner have a density of up to 800 peptides per cm^2. Newly developed methods allow for an increase in density to up to 10,000 peptide spots per cm^2 as well as the incorporation of special building blocks such as phosphorylated amino acids and the cyclization of peptides.

Tetanus serves in this thesis as a model system to compare antibody fingerprints in different individuals, due to the generally high vaccination rate. The vaccination against tetanus triggers a long term immune response, which can neutralize the toxin very reliably. The toxin is extracellularly released by the bacteria *Clostridium tetani*. Since it remains

elusive how antibodies neutralize the toxin, it is essential to investigate the immune response more closely. For the investigation of the immune response to the vaccination against tetanus, the linear amino acid sequence of the toxin was first cut *in silico* into overlapping 15mer peptides with a lateral shift of one amino acid. After synthesis of these peptides in the array format, the arrays were incubated with the sera of 19 vaccinated Europeans. This process is referred to as epitope mapping and enables to mimic and identify epitopes that are targeted by antibodies. A broad antibody distribution was observed in the immune response of 19 individuals. Since all sera had protective antibody titers against tetanus, it can be assumed that different epitopes on the toxin are capable of inducing neutralizing antibody responses. However, it cannot be ruled out that the broad antibody diversity only occurs due to the detection with linear peptides and the investigation of the antibody response with protein-like structures might yield a different result. An epitope on the C-terminal domain of the heavy chain of the tetanus toxin was prominent in 9 individuals. The subsequent substitution analysis of this epitope showed that the antibody fingerprints are almost identical in 8 out of 9 cases. This result is astonishing regarding the random generation of the antibody repertoire. Amino acids, which allowed binding besides the original amino acid, were also very similar in all tested samples. They often had a similar chemical character to the original amino acid, for example, hydrophobic valine could be exchanged for isoleucine without loss of binding. The respective antibody was isolated from one serum using affinity chromatography in a batch process and its reaction with the native toxoid in an ELISA could be demonstrated. Whether this identified antibody specificity has neutralizing properties with respect to the tetanus toxin should be clarified in further studies. However, it has been shown for the first time that different individuals develop almost identical antibody specificities on the sub-antigen and even sup-epitope level.

Lyme disease has been one of the most common diseases transmitted by ticks since the first observation of clustering in Lyme, Connecticut, in the 1970s. Diagnosis and treatment, however, remain a very controversial issue. It remains elusive, whether late manifestations of the disease are triggered by autoimmune reactions or permanent infections with the causative bacteria of the species *Borreliella*. The diagnosis, which in most cases is based on the detection of serum antibodies, is problematic due to a lack of specificity or sensitivity or both. The immunogenic lipoprotein VlsE (*vmp*-like sequence expressed) is expressed on the surface of the bacteria and has proven to be a reliable diagnostic marker. The protein possesses 3 domains: one variable domain in the middle is flanked by two conserved domains at the C and N terminus. The variable domain consists of 6 variable and 6 conserved regions. The aim of the study described in the manuscript was to compare the antibody fingerprints against the

VlsE antigen of patients suffering from Lyme disease and to evaluate the disease specificity of these fingerprints. Therefore, as described in the previous section, an epitope mapping of the VlsE protein was carried out with subsequent substitution analyses of the immunogenic peptides. The peptides were incubated with the sera of 17 VlsE positive and 7 VlsE negative patients. Based on their antibody response, 13 of the 17 positive patients could be assigned to two groups. The antibodies of one group recognized preferentially epitopes at the N- and C-terminal end of the protein, antibodies of the second group were mainly directed against epitopes on the variable domain. The substitution analysis showed that the amino acids crucial for binding are located on the conserved regions of the protein, even on the variable domain. As previously investigated in the tetanus vaccination study, it was shown that if the same epitope is targeted by antibodies, the fingerprints are almost identical. The query of public protein databases revealed that the longer fingerprints are very specific for *B. burgdorferi*. The precise antibody specificities could be used in the diagnosis of Lyme disease and other infectious diseases with the advantage of enabling studies of cross-reacting (human) proteins.

In the third study, 73 antibody fingerprints were determined without *a priori* knowledge with a combination of pre-screening with phage display, next generation sequencing and validation and substitution analyses with peptide arrays. The query of one of the identified dominant motifs in a protein database led to the antigenic coat protein VP1 of enteroviruses. Due to the widespread vaccination against the poliovirus from this family, it is very likely that a corresponding antibody has been identified. Experiments in a similar manner, in which the antibody specificities in another individual were determined, show that the fingerprints are also patient specific.

Thus, the present manuscripts show the biological relevance of antibody fingerprints regarding their disease and patient specificity. A study with multiple patients and the developed methods can contribute to the investigation of etiology in infectious diseases and the development of biomarkers and vaccines.

In general, peptide arrays only allow for the determination of the binding of an antibody to a peptide. The nature of the interaction in terms of affinity cannot be assessed due to the detection of binding in equilibrium state. Interaction is visualized based on the principle of an immunoassay with fluorescently-labeled secondary antibodies. However, in many cases, the affinity of an antibody to its antigen and its concentration is of interest. Therefore, it is necessary to observe the process of antibody-antigen binding from the beginning until the equilibrium is reached. For this purpose, the antibody solution must be brought into continuous contact with the peptides in the array format, a task that can be realized in a

microfluidic channel. Furthermore, the interaction must be made visible for the duration of the binding process, which can be accomplished with an optical system.

As the basis for these measurements, an automated microfluidic system with an optical setup was designed and established. The optical set-up detects the antibody-peptide interaction in a time-resolved manner, based on a fluorescence-exciting LED and a CCD camera. The system was validated by incubating the FLAG epitope in array format with the monoclonal FLAG-M2 antibody in the microfluidic channel. The interaction was detected with the optical system for different concentrations and the association rate constant was calculated. In the future, this system could be used to determine the affinities and concentrations of antibodies in patient sera and thus lead to better diagnostics.

Contents

1 Introduction

Our immune system recognizes infectious intruders and defends us against them with a – hopefully adequate – immune response. One major part of this defense is the secretion of antibodies, which specifically target antigens and participate in the process of making them innocuous. For decades, scientists have striven to characterize the different antibody specificities that shield us from diseases or do harm in autoimmune disorders. It is of great interest, what these antibodies exactly target, and, if and how those specific antibodies correlate with disease etiology. This knowledge should help in the development of vaccines and biomarkers. The following chapters focus on principal aspects of the immune system, the humoral immune system and immune responses towards pathogens and vaccinations. The specificities of antibodies can be described with peptide arrays in unprecedented detail, and, in addition, for many different antibodies in parallel. Therefore, this work discusses high-throughput methods for the detection of antibodies in general, with a focus on peptide arrays.

1.1 Immune system and immune responses

Viruses, bacteria, fungi, and parasites constantly challenge us, and we use our innate and adaptive immune system to defend ourselves against these pathogens. First, our innate immune system uses epithelial and chemical barriers, then macrophages, the complement system and the triggering of inflammation to shield us from harm. The second important task of the immune system is the elimination of malfunctioning endogenous cells. These issues are discussed in textbooks in sufficient detail [1, 2]. Vertebrates have an additional, adaptive immune system that is presumably activated by the innate immune system and comprises the development and deployment of dendritic cells and T and B lymphocytes.

Clonal selection, random point mutations and somatic hypermutation lead to the adaption of the lymphocytes in such a way that they recognize their specific antigens with increasing affinity. The detection of antigens is primarily based on different receptors: While B-cell receptors simply bind to presented antigens, T-cell receptors demand a more specialized form of presentation: MHC (Major Histocompatibility Complex) class I and class II proteins anchor processed peptides from the inside of the cell to a groove on the MHC-molecules.

Those cells that specifically bind to presented peptides through their receptor and, in addition, receive signals through co-receptors start to proliferate in a process that is called *clonal selection*. Most B cells, improve the binding affinity of their B-cell receptor by first mutating the B cell receptor's variable genes randomly, and then selecting B cells that bind to their waning antigens with higher affinity. This process is called *somatic hypermutation*.

Many different of these positively selected B cell clones secrete antibodies into the blood serum, summed up as the humoral immune system, in which they have important roles. Some of these antibodies protect us from infections, others might be useful for diagnosis, but otherwise are irrelevant. Some antibodies can do harm, e.g. in dengue fever, or in autoimmune disorders, where autoantibodies target endogenous parts of our body, and, thereby, could be also diagnosed. Others, such as natural antibodies, have different tasks in homeostasis.

1.1.1   The humoral immune system

The humoral immune system comprises all macromolecules that are secreted into our body fluids. Major parts of the humoral immune system are complement and acute phase proteins that play an important role in the innate immunes system. The main tasks are the production and secretion of antibodies into our blood system, which are also referred to as immunoglobulins (Ig). The most abundant antibody class (IgG) has a Y-shaped structure with a molecular weight of about 150 kDa. They are secreted by short lived plasma cells that descend from B lymphocytes or B cells. For the successful defense of our body, the antibody repertoire must be versatile and at the same time avoiding auto-reactivity. The large diversity of B cell receptors (BCRs) is generated in the bone marrow, where B cells undergo V(D)J recombination. Auto-reactivity is avoided by negatively selecting those B cells that bind to autoantigens. Their BCR is edited until non-self-reactive cells can be released into the periphery [3, 4]. Another procedure to generate tolerant B-cells is *anergy*, meaning that those B-cells that are for instance generated through *somatic hypermutation* and bind to autoantigens are no longer stimulated by antigen binding. The antibodies produced by cells that have emerged without an immune response to an exogenous antigen are referred to as "natural antibodies"; they have tasks in homeostasis, e.g. targeting aging cells for removal by macrophages [5].

While the entirety of B cells produces a variety of antibodies, each cell alone produces only one type of antibody with exactly the same amino acid sequence, a process that is called *clonal expression* [6]. The mature but naïve cells wander through our body fluids and

specialized lymphatic tissues, where they are activated by dendritic cells that present antigens to T and B cells [7]. The activation takes place in germinal centers of lymphatic organs and leads to proliferation and development of the naïve cells into Ig-secreting plasma or memory cells. Whereas memory cells are part of the immunological memory and help reacting towards repeated contact with the same antigen, plasma cells secrete the bulk of 11 mg IgG per ml blood serum [8]. Depending on costimulatory signals, a *class switch* from IgM or IgD to IgG, IgA or IgE can occur [9]. This switch preserves the antigen specificity, but changes the constant regions (Fc-region) of the antibody to one of five different isotypes that are used for specialized tasks. IgG antibodies patrol through the blood and lymphatic space, IgEs are tethered to the surface of mast cells that protect epithelia via high affinity Fc-receptors and IgA antibodies are secreted into the alveolar space or the intestine to tackle pathogens in these compartments. Antibodies exert their function in different ways. Neutralizing antibodies block a toxin or a virus from entering a cell. Opsonizing antibodies tag the surface of pathogens, and, thereby, help macrophages to ingest those pathogens, which is especially important in the fight against gram-negative bacteria. Another effector function is the activation of the complement system.
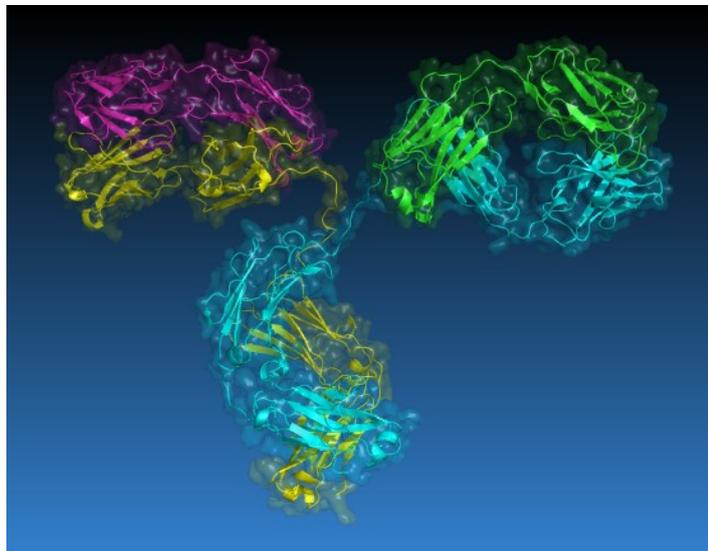
Figure 1: Structure of an antibody, depicted using the PyMol (Schrödinger, pymol.org/dsc/) molecular visualization system, based on RCSB PDB (www.rcsb.org) ID 1IGT [10].

Different theories exist on the lifespan of plasma cells from several days to several months or a conditional survival [11]. In any case, short- or long-lived plasma cells can synthesize and secrete up to 1000 antibodies per second. Therefore, and due to limited space for

antibodies in our serum, the actual repertoire, expressed at a certain time point, is most likely highly flexible. For a rough estimation of the number of different antibody species, we assume that all B cell clones together produce a bulk of 11 mg IgG per ml serum [8]. The efficient concentration of monoclonal antibodies in cancer therapies was investigated to be > 10 µg/ml [12]. Assuming that each antibody must be present in a therapeutic concentration of at least 10 µg per ml, our blood serum offers space for about 1100 different therapeutic antibody species.

1.1.2   Immune reactions towards pathogens

The vast majority of invading pathogens is neutralized at an early stage by our innate immune system before doing any harm. They are kept away by natural barriers as the epithelial layer of the skin or by chemical barriers, such as the low pH of the stomach and a diversity of antibiotic peptides that are secreted into our intestine. The second line of defense is built up by macrophages that line up all the epithelial layers of our body, e.g. underneath the skin, in the lung, or in the intestine. Whenever these macrophages sense a "pathogen associated molecular pattern" (PAMP) through their receptors, they are activated and secrete cytokines as tumor necrosis factor alpha (TNF-α) or interferon gamma (INF-γ). Upon these signals, small blood vessels nearby are sealed to block the pathogens from invading the whole body. As a result, many macrophages, granulocytes, and natural killer cells assemble around the intruders that are phagocytosed, bombarded with reactive oxygen species (ROS), marked, and eventually pierced with activated complement components.

The concept of PAMP recognizing receptors was first introduced by Janeway in 1989 [13]. Known PAMPs are bacterial or fungal cell-wall components, such as peptidoglycans, lipopeptides and lipopolysaccharides. Other PAMPs are flagellin and regularly spaced sugar moieties, viral double stranded RNA, and bacterial DNA that is differently methylated when compared to our own DNA. These PAMPs are evolutionary conserved and cannot be found in higher eukaryotes. One of the best characterized and most important family of these PAMP recognition receptors are toll-like receptors (TLRs) that are found in many different multicellular eukaryotes from the sea urchin over insects to mammals. Their discovery in the mid-1990s was awarded with the Nobel Prize in physiology or medicine in 2011 [14, 15]. The TLRs provide protection against microbial infection even in organisms without adaptive immune system. In *Drosophila*, the synthesis of anti-fungal and anti-bacterial peptides can be induced by the activation of signaling pathways via two members of the TLR family [16, 17]. Signaling *via* PAMP-receptors enhances the innate immune response by the secretion of pro-

inflammatory cytokines, which activates other immune cells. TLRs are expressed in dendritic cells and macrophages as well as in non-immune cells such as fibroblasts and epithelial cells [18].

The third task of PAMP recognizing receptors from the innate immune system is the induction of adaptive immunity [19]. Here, dendritic cells (DCs) play a critical role. Depending on the signals, these DCs induce different T cell responses: cytotoxic CD8 T cells are needed to tackle viral infections and intracellular growing bacteria; CD4 helper T cells are needed to produce neutralizing antibodies that bind a bacterial toxin or opsonizing antibodies that stimulate macrophages to phagocytose gram negative bacteria [20]. The migration of activated DCs into lymph nodes is assumed to be the first step in activating the adaptive immune response by inducing the differentiation of T cells into effector cells.

A critical step in this pathway is the mutual activation and control of lymphocytes, which has been described in various textbooks [1, 2]. Typically, a dendritic cell (DC) is first alarmed by its TLRs. This professional antigen-presenting cell (APC) then reorganizes its proteasome to load newly processed peptides onto MHC class I molecules, and, in addition, secretes cytokines to activate other cells. In parallel, the DC also loads newly processed peptides onto MHC class II molecules. Next, these DCs travel to specialized lymphatic organs, e.g. a lymph node, where T- and B-cells can specifically bind to presented antigens (T cells are restricted and bind only to peptides presented on MHC molecules). T helper cells 1 (T_H1) that recognize MHC class II molecules via their co-receptor CD4 can activate these APCs by binding to it. Thereby, they become activated themselves and turn into memory T cells and suppressor T cells. T helper cells 2 (T_{H2}), activate B cells and stimulate them to differentiate into plasma or memory B cells. Other T cells recognize MHC class I presented peptides via their co-receptor CD8. When activated, these cytotoxic CD8 T cells specifically recognize and kill those cells that present a pathogen's peptide on their MHC class I molecule.

There is growing evidence that B cells and their secreted antibodies can specialize in a way that is analogous to the different T cell subsets. It is known that different antibody classes perform specialized tasks. In addition, the glycosylation of the antibody's Fc-part decides upon the antibody's affinity towards different Fc-receptors [21].

1.1.3   Vaccines and immunity

Ever since Edward Jenner laid the foundation for vaccine-based immunization at the end of the 18th century [22], vaccines have considerably prevented morbidity and mortality. The

vaccination against smallpox, which was the vaccine that Jenner developed, successfully eradicated the disease from the planet [23].

Whereas some vaccines already elicit highly effective immune responses, capable of neutralizing the pathogen, some diseases still need be addressed. For three of the major global health issues - HIV, malaria and tuberculosis - no efficient or licensed vaccines are currently available. It still needs to be researched, how to focus the vaccine induced immune response to protective epitopes and how to elicit long-term memory responses [24]. Antibody-dependent enhancement of a disease [25] and cross-reactions induced by vaccination against viruses [26] [27, 28] are among the possible draw-backs of current vaccines.

Many successful vaccines induce neutralizing antibodies that block the binding site of a toxin or virus to prevent entry into cells [29]. It is more difficult to develop vaccines for pathogens that are extremely variable. HIV-1 for instance has several mechanisms to evade the host's immune response [30, 31]. However, after years of infection, broadly neutralizing antibody responses have been reported [32, 33]. Similarly, some patients are better protected by their immune system then others after repeated infection with *P. falciparum*, the causative agent of malaria [34].

To benefit from the information that is deposited in the adaptive immune system of these patients, we need to develop methods to readout this information. Recently, tremendous progress has been made in monitoring the human immune response [24]. Microarrays, sequencing technologies, mass spectrometry-powered proteomics and bioinformatics are among the driving forces [35, 36].

Vaccines based on antigenic peptides have several advantages. They are stable, easy and cheap to synthesize and focus the immune response on relevant epitopes; therefore, they elicit only neutralizing responses and avoid the formation of cross-reacting antibodies [37]. The identification of immunodominant epitopes and the elucidation of the combination of those for a successful neutralization of toxins or pathogens in different patients could be the first step towards a peptide based vaccine.

1.2   Studying antibody – antigen interactions

Proteins are key players in our organisms and for the understanding of fundamental biological processes, it is necessary to characterize protein-protein interactions. In particular, the antibody-antigen interaction needs to be studied for diagnostic purposes, for the treatment

and basic research of diseases and for the exploitation of the immune response in vaccinations. Antibody-binding patterns, for example, can be employed for distinguishing the health or disease condition of different individuals. Several techniques for the characterization of the humoral immune system have been developed. The following chapters focus on the characterization of antibody-antigen interaction, the principal detection of antibodies in a high-throughput manner and the antibody specificity that can be explored with peptide arrays.

1.2.1 Characterization of antibody-antigen interaction

To fully understand the molecular interaction of antibodies and antigens, it is necessary to describe and characterize the binding and dissociation processes with kinetic rate and equilibrium binding constants. The equilibrium binding constant defines the affinity of an antibody to its antigen, which can also be described as the strength of the interaction between an antibody and its epitope on an antigen or the sum of attractive and repulsive forces between them [38]. Several models have been developed to determine the equilibrium binding constants with an immobilized ligand and a ligate in solution [39, 40, 41].

If the antibody is monovalent, one epitope is targeted, the reactants have homogeneous binding sites and the reaction takes place in solution and equilibrium, all models can be derived from the law of mass action (1.1). In this equation, [Ab] is the molar antibody concentration and [X] the molar antigen concentration. The equilibrium rate constant K_{eq} can be calculated by the ratio of association rate constant k_a to dissociation rate constant k_d (1.2).

$$Ab + X \underset{k_d}{\overset{k_a}{\rightleftharpoons}} AbX \tag{1.1}$$

$$K_{eq} = \frac{k_a}{k_d} = \frac{[AbX]}{[Ab][X]} \tag{1.2}$$

For a real-time analysis of antibody-antigen interaction, which is common for biosensors and continuous microfluidic systems, the kinetics can be either limited by the reaction or be dependent on the mass transport or both [39]. True association and dissociation rate constants can only be calculated if the mass transfer is much faster than the reaction time and, therefore, can balance concentrations differences at the surface and in the bulk. If the system is mass transport limited, meaning the diffusion is limiting the interaction, the reaction obeys Fick's first law [42].

1.2.2   High-throughput technologies for antibody detection

For the detection of antibody-antigen interactions in a high-throughput manner, different methods have been established. They either concentrate on the expressed antibodies, present in the serum of an individual or on gene segments encoding the antibody molecules.

In western- or immunoblotting, proteins that have been separated according to their size or charge, are transferred from a gel to a membrane and are detected with antibodies. Today, nitrocellulose is routinely applied as membrane, which was introduced by Towbin, *et al.* in 1979 [43]. Primary and secondary antibodies are used for the specific detection of a sample. There, the secondary antibody is either labelled radioactively or conjugated to a fluorescent dye or a peroxidase. The profiling of the complete antibody repertoire and the identification of antibody reactivity patterns were first performed using immunoblotting [44]. Furthermore, different autoimmune diseases have been analyzed using immunoblotting for the detection of antibody patterns against large panels of homologous auto-antigens, such as chronic lymphocytic leukemia [45] and idiopathic thrombocytopenic purpura [46].

ELISA (Enzyme Linked Immunosorbent Assay) is an antibody based detection method with an enzymatically triggered quantitative color reaction. As the assays are simple to perform and easy to automate, they can be used in high-throughput screenings and are widely applied in analytics and diagnostics. In autoimmune research, ELISA tests in combination with bioinformatics were used for the detection of autoantibody patterns in healthy and diseased patients [47]. However, in both, ELISA and immunoblotting, the number of interaction partners that can be investigated in one experiment are limited.

For many screenings with an unknown interaction partner, a sufficient number of potential targets are necessary. Therefore, several display techniques have been developed and are commonly used. Phage Display (PD), first published in 1990 [48, 49, 50], offers the possibility to screen a library of up to 10^9 independent clones [51]. Randomly generated peptide libraries, presented on filamentous bacteriophage, have proven to be a powerful basis for the investigation of various aspects in many studies. In 2001, breast cancer specific B-cell-epitopes could be identified with PD [52]. Some of these epitopes even correlated with patient survival time and may be interesting targets for vaccine development. Peptides that are targeted by autoantibodies and indicate a growing prostate cancer, were identified in 2005 using peptide libraries displayed on phage [53]. Furthermore, PD was used to display the viral epitopes from all known human viruses to discover targets of by human serum antibodies [54]. PD has also been used to generate therapeutic antibodies as the FDA-approved human anti-TNF IgG1 against rheumatoid arthritis (Adalimumab) [55]. Yet, low

reproducibility and insufficient controllability of the displayed interaction partners (bias of biological methods) are the limiting factors of display techniques.

Next-generation sequencing or high-throughput sequencing are novel techniques for the rapid sequencing of RNA and DNA, accelerating the study of genomics and biological systems. In 2009, Joshua Weinstein *et al.* [56] used high-throughput sequencing to analyze the VDJ gene segment recombination in zebrafish to determine the immunoglobulin diversity. Besides different specific gene combinations, the majority of VDJ combinations were identified in all fish in low frequencies; only a small portion of different combinations per fish was found at similar high frequencies. Recently, high-throughput DNA sequencing of immunoglobulin genes (Ig-seq) was developed, allowing for the sensitive discovery of antibodies in autoimmunity, disease research and vaccine development [57, 58]. Yet, the gene expression at a certain time point might be different from the expressed diversity of proteins and antibodies.

Microarrays with immobilized DNA, RNA, proteins, peptides or antibody molecules, are screening tools for the simultaneous investigation of large numbers of interactions in a single experiment with minimal sample consumption. Protein chips, for example, were employed for the identification of antigens after the production of monoclonal antibodies against human liver proteins [59]. They have also shown to be useful in the identification of biomarkers in various infectious diseases [60, 61, 62]. However, the commonly applied expression of the cell lysates in *E. coli* without purification and, therefore, reduction in sensitivity of the assays, are drawbacks of protein microarrays.

1.2.3   Antibody detection with peptide arrays

Ever since B. Merrifield [63] invented the solid phase peptide synthesis and R. Frank developed the first peptide arrays with the SPOT synthesis [64], peptide arrays have evolved and versatile screenings can be performed routinely. The possibility of producing a large variety of peptides allows for the addressing of different questions regarding antibodies, such as which part of a protein (epitope) is bound by an antibody and if there is potential cross-reactivity. Thus, peptide arrays contribute to the understanding of basic immune reactions on a large scale with little sample consumption in minimal time.

Peptide arrays are commercially available and various techniques are applied for the synthesis: they are either pre-synthesized and subsequently spotted onto a solid support or directly synthesized on functionalized surfaces (e.g. glass slides). In the first case, the quality of the peptides can be assured easily, but the procedure is complex, expensive and the spot

densities are limited. For the latter case, amino acid building blocks are deposited one by one, either with the SPOT synthesis method, photolithographic methods [65] or a laser printer [66]. With the development of the peptide laser printer, it became possible to produce customized peptide arrays that combine two important production characteristics: they can be manufactured rapidly and at low costs. The amino acid derivatives are embedded in particles consisting of a polymer material that is solid at room temperature and, therefore, can be easily structured on a solid substrate [67]. After structuring of all 20 different amino acids in the first layer, the chemical coupling steps are performed. For this purpose, the slides are heated and the particles become viscous, allowing the amino acid derivatives to diffuse and couple to free amino groups of the polymer film and to couple. Due to the temporal separation of structuring and coupling, only one coupling step needs to be performed for each layer of amino acids, which reduces the production time and increases coupling efficiency. Spot densities of up to 800 different spots per cm^2 (250 x 500 μm) can be achieved by synthesizing arrays with the described method.

Peptide libraries synthesized as microarrays have widely been used for antibody binding studies in a large variety of questions. Besides others, the have been employed for the identification of IgE epitopes of food allergens [68] and contributed to the understanding of autoimmunity in cardiovascular diseases [69]. Profiling of antitumor antibodies led to the identification of a biomarker in glioblastoma [70] and they were also used to detect epitope patterns in pulmonary tuberculosis, which are useful for the development of diagnostics and vaccines [71].

1.2.4   Antibody specificity

The specificity of an antibody allows it to discriminate between different antigens, whereas the term cross-reactivity describes the recognition of resembling antigens [72]. Historically, antibodies could be also distinguished between mono- and poly-specificity: Mono-specificity would imply that one molecule interacts with one specific other molecule much stronger than with all other molecules. Yet, the term mono-specificity has been abandoned by now, since the definition of a molecule to be mono-specific would imply that the binding character of all possible molecules has been tested, which is technically impossible. Studies investigating the binding of monoclonal antibodies towards synthetic peptides or proteins clearly showed that even these antibodies target cross-reacting antigens and, therefore, are poly- rather than mono-specific [73, 74]. Cohn and Langman [75] suggested that the selection pressure for a B

cell to preferably produce mono-specific antibodies is rooted in the necessity of distinguishing between self- and non-self-antigens.

Substitution analyses, also referred to as replacement set analyses, were introduced in 1987 by Geysen *et al.* [76]. All amino acids in one position of an antigenic peptide are replaced one by one by all other 19 amino acids and the resulting peptides are tested for their binding behavior to the antibody. For each position, one positive control with the original amino acid is synthesized. With these analyses, the role of each single amino acid in an antigenic peptide is elucidated and the amino acids that are essential for the binding of an antibody to its epitope are determined. Therefore, it is possible to determine essential and alternative amino acids that result in the binding of an antibody, which can be referred to as alternative specificities of an antibody. However, it needs to be validated, if these specificities correspond to real epitopes or mere mimotopes [77]. In both cases, the specificities allow for the identification of the originally inducing antigen and potential cross-targeted (autoimmune) proteins.

2 Research proposal

Antibodies are among the key players of our immune system, they shield us from infectious pathogens, but can also do harm in autoimmune disorders. An antibody can be sufficiently described by its specificity, which is of great value for diagnosis and etiology of infectious diseases and the development of vaccinations. The specificity of an antibody can be defined by the targeted epitope and, therefore, the specificity describes both, antigen and antibody. Several high-throughput techniques for the detection of antibodies have been developed and are commonly applied. However, due to the large variety of similar antibody species patrolling our blood, it was hitherto impossible to exactly determine all antibody species, their quantity and the inducing antigens.

High-density peptide arrays can be employed for the parallel investigation of thousands of antibody-antigen interactions with minimal sample consumption. With epitope mappings, displaying the amino acid sequence of an antigen as short overlapping peptide fragments, the antigenic protein epitopes of a pathogen can be determined. These investigations recently led to the conclusion that antibodies of different individuals target the same limited number of protein epitopes of a pathogen. However, little is known about the antibody specificity on epitope and sub-epitope level. To overcome this gap, a subsequent substitution analysis of an antigenic peptide can determine the antibody specificity in single amino acid resolution. Every amino acid in such a peptide is replaced one by one by all other amino acids and the binding of antibodies to the altered sequences is investigated. Thereby, binding and non-binding amino acids as well as alternative amino acids are pinpointed, which we call the antibody fingerprint.

Antibody fingerprints offer new possibilities in the field of disease etiology, for the development of diagnostic markers and the design of vaccines. Correlating antibody species with a hitherto enigmatic disease due to an unknown infection could identify the antigen that was inducing the formation of the antibodies without *a priori* knowledge. For the identification of population-wide conserved antibody species, the immune responses of many individuals, suffering from the same disease, must be compared. The identification of those antibody species that correlate with protection of a disease could be candidates for new vaccines. Moreover, antibody fingerprints could allow for the identification of proteins

responsible for transplant rejections or for the exploration of autoimmune disorders. Additionally, the use of fingerprints would enable the identification of cross-targeted proteins.

The aim of this doctoral thesis is to establish the foundation to comprehensively exploit antibody fingerprints: (1) The antibody fingerprints must correlate with a certain disease and be specific for it. For the evaluation of the disease specificity of antibody fingerprints, immune responses of several individuals to known disease antigens will be compared in two model studies. First, it will be investigated, whether specific antibody fingerprints against the tetanus toxin after a vaccination can be identified with epitope mappings and subsequent substitution analysis. In a similar approach, the antibody fingerprints of Lyme disease patients against antigens of the inducing pathogen will be compared. (2) It must be possible to identify inducing and cross-reacting antigens and pathogens with antibody fingerprints without *a priori* knowledge. For the evaluation of the suitability of antibody fingerprints for this task, queries of public protein databases with fingerprints against known and unknown antigens will be performed. For the identification of antibody fingerprints against unknown antigens, the humoral immune system of an individual will be screened with random peptide sequences and substitution analyses. Since many peptides are required to possibly determine all antibody species of the examined individual, a pre-screening with peptide phage display will be established. (3) Knowing the concentration and affinity of the specific antibodies could expand the application possibilities in disease etiology. To investigate the kinetics of antibody-peptide interactions, a microfluidic system with optical setup for time-resolved detection will be established.
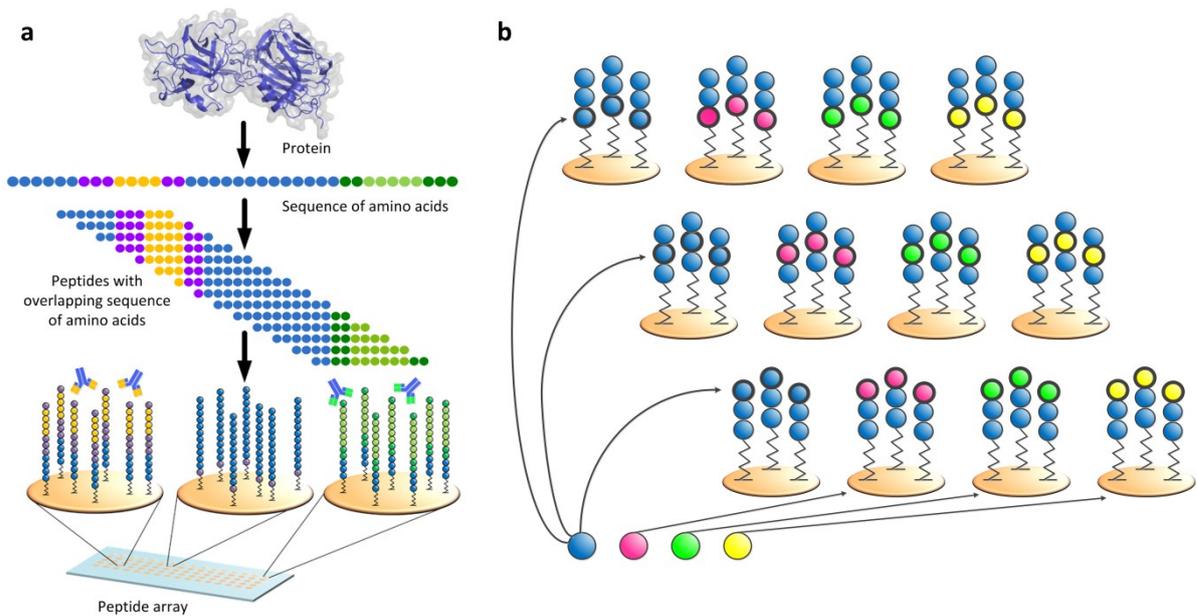
Figure 1: Principal approach of research - workflow of epitope mapping (a) and substitution analysis (b) on high density peptide arrays. For the epitope mapping, the linear sequence of an antigen is cut in silico in peptides with overlapping amino acid sequences. The array consisting of these peptides is incubated with patient serum and fluorescently labelled secondary antibodies. In the second step, antigenic peptides that have been identified in the epitope mapping are investigated in a substitution analysis. Each amino acid in such an antigenic peptide is one by one substituted by all other amino acids, while the rest of the sequence remains conserved. The array of the resulting substituted peptides is again incubated with the patient serum and secondary antibodies.

3 Published scientific material and manuscripts

Identification of a tetanus toxin specific epitope in single amino acid resolution

Andrea Palermo*, Laura K. Weber*, Simone Rentschler, Awale Isse, Martyna Sedlmayr, Karin Herbster, Volker List, Jürgen Hubbuch, Felix F. Löffler, Alexander Nesterov-Müller, Frank Breitling (*contributed equally)

This manuscript describes the combination of epitope mapping and substitution analysis for the investigation of antibody responses in 19 vaccinated Europeans towards the tetanus toxin. Besides a broad interaction of antibodies with various epitopes, one distinct epitope on the tetanus toxin was detected in 9 individuals on the C-terminal domain of the heavy chain. The substitution analysis revealed almost identical fingerprints in 8 of those 9 individuals. Furthermore, it was possible to isolate the respective antibody species from the serum of one individual with affinity chromatography using the identified peptide. The tetanus specificity of the isolated antibody was proven in an additional ELISA experiment. This study demonstrated for the first time the disease specificity of antibody fingerprints.

Antibody fingerprints in Lyme disease deciphered with high density peptide arrays

Laura K. Weber, Awale Isse, Simone Rentschler, Richard E. Kneusel, Andrea Palermo, Jürgen Hubbuch, Alexander Nesterov-Müller, Frank Breitling, Felix F. Löffler

In this paper, the results of a study with 24 Lyme disease patients are described. The antibody responses of 17 VlsE positive and 7 negative sera were investigated in an epitope mapping with subsequent substitution analysis of the immunodominant epitopes. The analysis revealed that in Lyme disease, the antibody fingerprints of different individuals towards shared epitopes of the VlsE antigen are almost identical. The antibody specificities on sub-epitope level can serve as biomarkers enabling the study of inducing and cross-reacting proteins.

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

Laura K. Weber, Andrea Palermo, Jonas Kügler, Olivier Armant, Awale Isse, Simone Rentschler, Thomas Jaenisch, Jürgen Hubbuch, Stefan Dübel, Alexander Nesterov-Müller, Frank Breitling, Felix F. Löffler

This paper describes a novel approach for the identification of causing pathogenic agents in enigmatic infectious diseases. The pipeline of prescreening with phage display, next generation sequencing and validation with peptide arrays in combination with substitution analyses revealed 73 antibody specificities in the serum of an individual without *a priori* knowledge. The query of public protein databases with the fingerprints allows for the identification of potential antigens. One of the most abundant motifs in our study led to the antigenic capsid protein of the enteroviruses. Due to common vaccination against the polio virus, a member of the enterovirus genus, the corresponding antibody species was most likely identified.

Automated microfluidic system with optical set up for the investigation of peptide-antibody interactions in an array format

Laura K. Weber*, Andrea Fischer*, Tim Schorb, Miriam Soehindrijo, Tobias C. Förtsch, Clemens Bojničić-Kninski, Daniela Althuon, Felix F. Löffler, Frank Breitling, Jürgen Hubbuch, Alexander Nesterov-Müller (*contributed equally)

In this manuscript, a microfluidic system for the automated incubation of peptide arrays is introduced. The combination with an optical set up allows for the detection of antibody-peptide interactions in array format in a time-resolved manner. The interaction of the monoclonal FLAG antibody with its epitope was investigated in different concentrations as proof of principle.

3.1 Identification of a tetanus toxin specific epitope in single amino acid resolution

A. Palermo[1]*, L.K. Weber[1]*, S. Rentschler[1], A. Isse[1], M. Sedlmayr[1], K. Herbster[1], V. List[2], J. Hubbuch[3], F.F. Loeffler[1‡], A. Nesterov-Mueller[1‡], F. Breitling[1‡]

[1] Karlsruhe Institute of Technology, Institute of Microstructure Technology, Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany

[2] Karlsruhe Institute of Technology, Medical Services, Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany

[3] Karlsruhe Institute of Technology, Institute of Engineering in Life Sciences, Section IV: Biomolecular Separation Engineering, Engler-Bunte Ring 3, 76131 Karlsruhe

* These authors contributed equally to this work.

‡ Corresponding authors:

F.F. Loeffler: felix.löffler@kit.edu

Alexander Nesterov-Müller: alexander.nesterov-mueller@kit

Frank Breitling: frank.breitling@kit.edu

Abstract

Vaccinations are among the most potent tools to fight infectious diseases. However, cross-reactions are an ongoing problem and there is an urgent need to fully understand the mechanisms of the immune response. For the development of a methodological workflow, we investigated linear epitopes in the immune response to the tetanus toxin in sera of 19 vaccinated Europeans applying epitope mapping with peptide arrays. The most prominent epitope, appearing in 9 different sera ($_{923}$IHLVNNESSEVIVHK$_{937}$), was investigated in a substitution analysis to identify the amino acids that are crucial for the binding of the corresponding antibody species– the antibody fingerprint. The antibody fingerprints of different individuals were compared and found to be strongly conserved ($_{929}$ExxEVIVxK$_{937}$), which is astonishing considering the randomness of their development. Additionally, the corresponding antibody species was isolated from one serum with batch chromatography using the amino acid sequence of the identified epitope and the tetanus specificity of the isolated antibody was verified in an ELISA experiment. Studying antibody fingerprints with peptide arrays should be transferable to any kind of humoral immune response towards protein antigens. Furthermore, antibody fingerprints have shown to be highly disease-specific and, therefore, can be employed as reliable biomarkers enabling the study of cross-reacting antigens.

Novelty Statement

Antibody fingerprints that can be deciphered with substitution analyses on peptide microarrays, define the specificity of antibodies in the resolution of single amino acids. In our study, we compared the immune response of different individuals towards the tetanus vaccination with epitope mappings and subsequent substitution analyses of a dominant epitope. We combined several steps in a pipeline and demonstrated that the identified antibody species is specific for the tetanus toxin. We believe that disease-specific antibody fingerprints will be employed for the development of biomarkers and vaccines in the future.

1. Introduction

Very effective vaccinations shield us from a lot of common infectious diseases [78]. It is known that vaccines protect us amongst others by inducing long-lived plasma cells [79, 80]. However, the precise mechanism of the immune response and the exact composition of antibodies elicited by vaccinations remain unclear. It has been shown in several studies that vaccinations against viruses can enhance other diseases by cross-reaction [27, 28, 26], or lead to antibody-dependent enhancement of the same disease [25]. The vaccination against the tetanus toxoid elicits a highly efficient neutralizing antibody response [81] and tetanus can be considered as eradicated in developed countries [82]. In 2011, a global coverage of 84 % was reached for the diphtheria-tetanus-pertussis combination vaccination, yet poor coverage was reported in low- and middle-income countries [83]. The tetanus toxin or tetanospasmin is an exotoxin, produced by the Gram-positive bacillus *Clostridium tetani*, which can normally be found in soil [82]. It consists of 1315 amino acids [84] and is structured in a heavy chain (Mw: 100 kDa) and a light chain (Mw: 50 kDa) [85]. The exact mechanism of how antibodies are able to neutralize the tetanus toxin is not fully understood, but early studies of monoclonal antibodies from mice suggest that different epitopes are targeted and also synergistic effects of different antibodies lead to its neutralization [86].

In this study, we compared the humoral immune response of 19 individuals towards the vaccination against tetanus with peptide microarrays in single amino acid resolution. In an epitope mapping, displaying the amino acid sequence of the tetanus toxin with overlapping peptides, we identified an immunodominant epitope on the heavy chain. Then, we compared the specificity of the antibodies towards this epitope with full substitution analysis. Thereby, those amino acids that are essential for the interaction of an antibody with an antigenic peptide are determined – the antibody fingerprint. We validated this specific antibody-peptide interaction on a second platform and, then, isolated the identified antibody species from the serum of one individual to prove that it interacts with the complete toxin in an ELISA experiment. The workflow of the project is depicted in Figure 1. Additionally, the antibody fingerprint was employed to query protein databases for the identification of cross-targeted antigens. Substitution (or replacement) analyses were introduced by Geysen et al. [76]. Recently, we published the potential of these fingerprints for the research in disease etiology and the read-out of the human antibody repertoire [87].

For the here described study, arrays were synthesized with a laser printer by the PEPperPRINT GmbH (Heidelberg, Germany). The solid-material-based synthesis methods

allow for the synthesis of high-density peptide arrays with over 10,000 spots per cm^2 [88, 89, 90, 91, 67].

Our experiments revealed that antibody fingerprints are highly disease-specific and can be employed to compare the immune responses of different individuals toward vaccinations or infections in unprecedented detail. Therefore, they can serve as reliable diagnostic tool and support the development of biomarkers and vaccines.
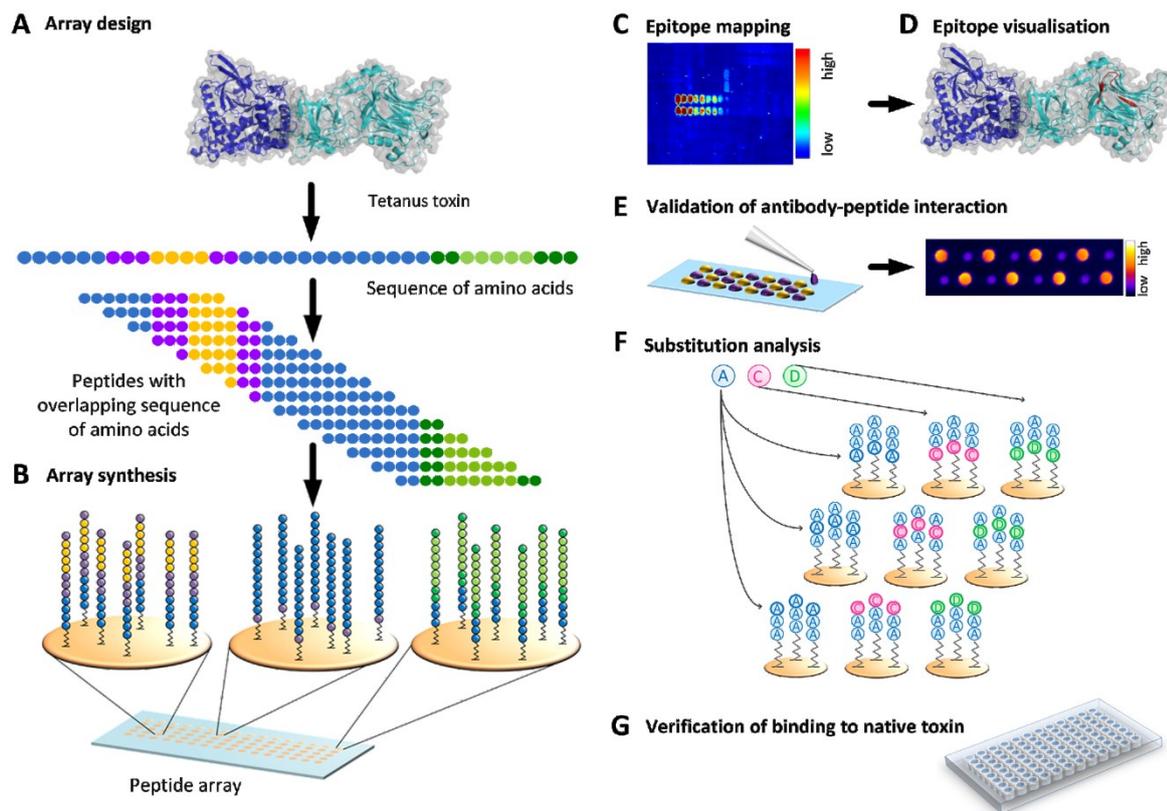
Figure 1: Workflow of the study. A: The linear amino acid sequence of a protein is cut *in silico* into peptides with overlapping sequences. B: The resulting sequences are synthesized onto a peptide array. C: The peptide array is incubated with human serum and subsequently with fluorescently labeled secondary antibodies. A fluorescence scan reveals the information which peptides are bound by serum antibodies. D: Information of bound epitopes is mapped onto the 3D protein structure. E: The antibody-peptide interaction is validated on a second platform with pre-synthesized and spotted peptides. F: In a substitution analysis, every amino acid of an antigenic peptide is systematically exchanged by all other amino acids to assess the antibody binding fingerprint. G. An antibody species is isolated with the antigenic peptide and its ability to bind to the native toxin is verified in an ELISA experiment. Parts of the figure are based on [92], reprinted with permission.

2. Material und Methods

2.1 Peptide arrays – design & synthesis

For the epitope mapping the amino acid sequence of the tetanus toxin [84] was synthesized as 1315 15mer peptides with an overlap of 14 amino acids in duplicates. 15mere peptides were chosen to cover all linear epitopes, which are reported to be up to 12 amino acids long [93], with at least 3 consecutive peptides. To enable interactions with the N- and C-terminal parts of the toxins spacer of alternating glycine and serine (GSGSGSG) were added respectively. Peptide arrays were purchased from PEPperPRINT, Heidelberg, Germany. Arrays were manufactured in a combinatorial solid-phase peptide synthesis process with a laser printer. To simplify the data analysis, a frame of Hemaggluttinin (HA) epitope peptides (YPYDVPDYAG) and FLAG epitope peptides (DYKDDDKGG) or polio peptides (KEVPALTAVETGAT) was synthesized around the array content. For the substitution analysis, each amino acid of a 15mer peptide is replaced by all other 19 proteinogenic amino acids, one after another, while the rest of the sequence is conserved. For this analysis, we selected a peptide that was targeted by antibodies from 8 of the 19 sera in the epitope mapping. To verify the interaction of serum antibodies with the identified epitope, the peptides IHLVNNESSEVIVHK-ßA-D-ßA-Peg-CCOOH, IHLVNNESSEVIVHR- ßA-D-ßA-Peg-CCOOH (negative control 1) and IHLVNNESSEVIAHK- ßA-D-ßA-Peg-CCOOH (negative control 2), purchased by Peps4LS (Heidelberg, Germany), were spotted onto a 3D Maleimid surface (PolyAn, Berlin, Germany) using a NanoPlotter 2.1 (GeSiM, Radeberg, Germany). The spotting was conducted with a Nano Tip J A070-401 20310. The peptides were diluted to 100 µM in PBS containing 10 % glycerol (99+ %) (Alfa Aeser, Karlsruhe, Germany). After spotting, the slides were dried for 2 h and subsequently blocked in PBS containing 0.4 % 2-Mercaptoethanol (Merck, Darmstadt, Germany). The washing was performed at 70 rpm on an orbital shaker (Orbital Shaker DOS-20S, Elmi Ltd., Riga, Lettland) for 3 min with PBS, 3 min deionized water, 5 min acetonitrile (VWR Chemicals, Radnor, USA) containing 0.1 % Trifluoroacetic acid (99%) (Honeywell Chemicals, Morristown, USA), 5 min Dimethylformamid (DMF) (99,8+ %) (VWR Chemicals, Radnor, USA) containing 0.5 % N,N-Diisopropylethylamine (Merck, Darmstadt, Germany), 3x5 min DMF and 2x3 min Methanol (Merck, Darmstadt, Germany). Slides were dried in an argon gas stream and stored at 4 °C until further usage.

2.2 Incubation of peptide arrays

The incubation of peptide arrays was carried out as described in Weber et al. [87]. Peptide arrays are rotated with a rotation speed of 140 RPM. Secondary antibodies (Goat anti-human IgG-Fc-Fragment DyLight®680 conjugated, Biomol, Hamburg, Germany) were diluted to 0.2 µg/mL. The secondary antibody mixture additionally contained DL800 conjugated (fluorescently labeled using a DyLight™ 800 Microscale Antibody Labeling Kit, Thermo Fisher Scientific, Rockford, USA) monoclonal anti-HA antibodies (provided by Dr. Gerd Moldenhauer, DKFZ, Heidelberg) diluted to 1 µg/mL and/or monoclonal anti-FLAG M2 antibodies (Sigma Alderich, Saint Louis, USA) diluted to 0.2 µg/mL.

Sera of 19 human donors, vaccinated against tetanus, were provided by Dr. List from the medical services at KIT. The study was approved by the state chamber of physicians of Baden-Wuerttemberg (reference number: F-2011-044 and F-2011-044#A1). Blood was extracted, centrifuged at 4000 g in BD Vacutainer R SST TM tubes (by BD Plymouth, UK) and stored at 4 °C until usage.

2.3 Image & data analysis

Fluorescence scanning with an Odyssey scanner (LI-COR Biosciences, Lincoln, USA) and evaluation with the PepSlide Analyzer (Sicasys Software GmbH, Heidelberg, Germany) was done as described in Weber et al. [87]. The foreground median intensities of spot duplicates were averaged. Dust on the surface of the arrays results in outliers which can be identified by differences in the signal intensities of spot duplicates. The average intensity of such spots is set to zero under the following two conditions: (1) the intensities of spot duplicates differ by more than 65 % in the average spot intensity and (2) the intensities of spot duplicates are higher than 50 % of the average intensity of all spots of the array. Additionally, the fluorescence scans were revised manually for outliers that are not detected by the before described criteria. For a better comparison of different sera, which show different maximum intensities, the spot intensities are normalized to the maximum spot intensity of the array for each serum. In the substitution analysis, the intensities of the peptide spots with the substituted amino acid are normalized to the intensity of the spots with the original amino acid for each row position respectively. If the substitution of an amino acid leads to a decrease in the signal of more than 40 %, the binding is considered as inhibited.

2.4 Serum purification

16 mL of serum 11 were purified using a two-step batch chromatography with protein G and pre-synthesized peptides immobilized on beads. The interaction of the isolated antibody with the native toxin was verified in a tetanus ELISA with the dilution of 1:2 and 1:10 according to the protocol. Protein G batch chromatography was conducted with a NAB Protein G Spin kit (Thermo Scientific, Rockford, USA) according to the protocol. The first four fractions were pooled and a buffer exchange to PBS was performed with a ZebaSpin Desalting kit 7K MWCO (Thermo Scientific, Rockford, USA) according to the protocol. Subsequently, an affinity purification was conducted using the SulfoLink Immobilization Kit for Peptides (Thermo Scientific, Rockford, USA) according to the protocol. The peptide used for the isolation (IHLVNNESSEVIVHK-ßA-D-ßA-Peg-spacer-CCOOH) was purchased by Peps4LS (Heidelberg, Germany). The additional peg-spacer was chosen to increase the solubility and decrease steric effects. The column was loaded 6 times with 3 mL of serum each time. The concentration of antibodies after purification (2nd and 3rd fraction pooled) was determined to be 0.053 mg/mL by optical density measurement at 280 nm in a Jenway 7305 spectrophotometer (Bibby Scientific, Staffordshire, UK). After elution, the buffer was exchanged as described before.

2.5 Tetanus titer determination by ELISA

IgG anti-tetanus antibody levels were determined by ELISA using a tetanus IgG ELISA kit (Sekisui Virotech GmbH, Rüsselsheim, Germany) according to the protocol. Evaluation was done according to the instructions; titers were calculated by the mean of three different dilutions in duplicates.

3. Results

3.1 Epitope Mapping

Peptide arrays containing 1315 peptides each, representing the tetanus toxin and synthesized as spot duplicates were incubated with sera of 19 vaccinated Europeans. All serum samples were positive by IgG tetanus ELISA; titers are shown in Supplementary Table S1. Figure 2 shows the results of the epitope mappings as heatmaps, displaying the interaction of serum antibodies with peptides representing the amino acid sequence of the tetanus toxin from left to right. The data was normalized to the maximum intensity of each array respectively. A broad distribution of interaction between antibodies of different patients and peptides was observed.
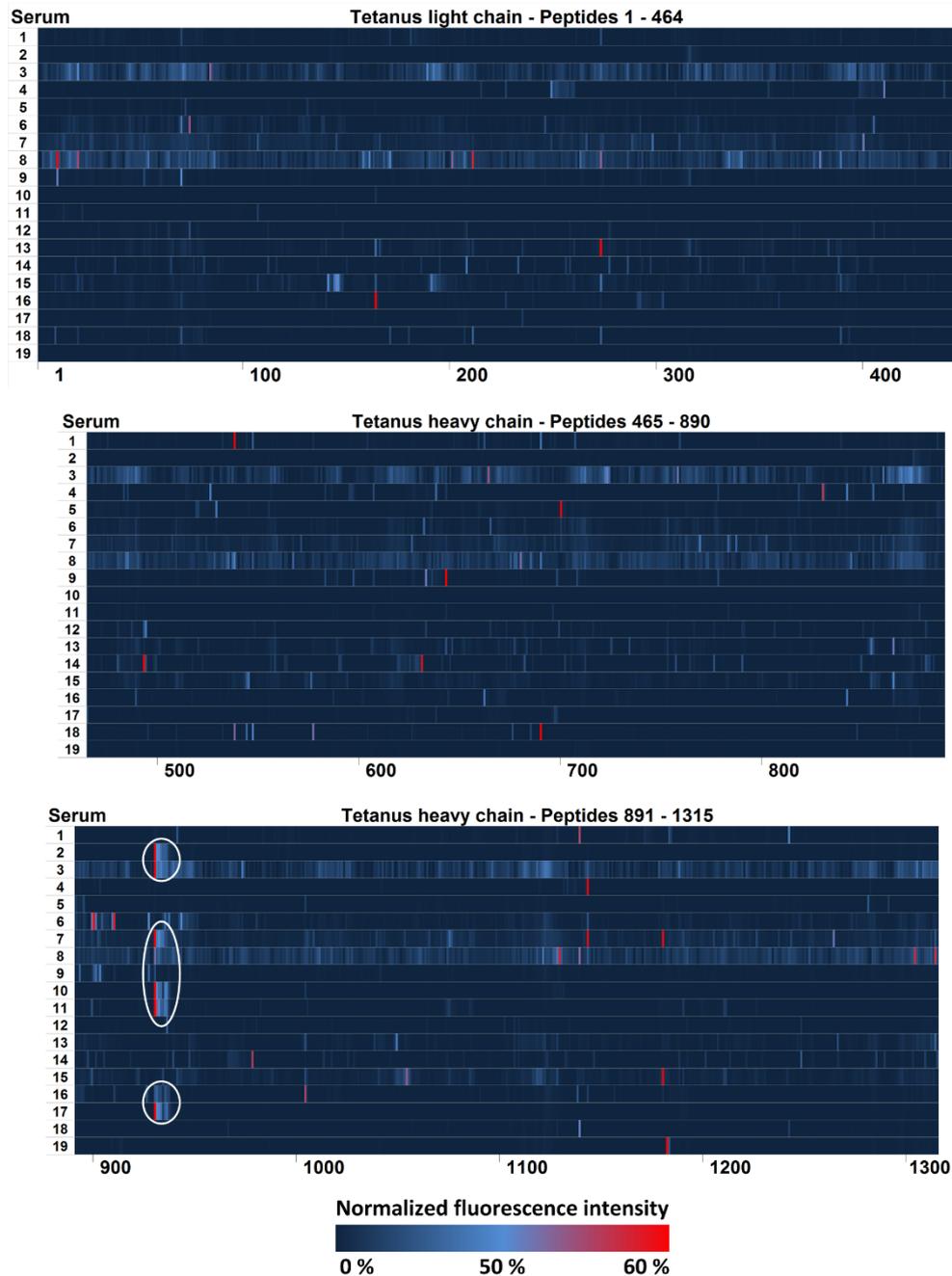
Figure 2: Heatmap of the epitope mapping of the light and heavy chain of the tetanus toxin. Sera of 19 vaccinated Europeans were incubated on peptide arrays, containing overlapping peptide fragments covering the linear amino acid sequence of the tetanus toxin. Peptides have the length of 15 amino acids and a lateral shift of one amino acid. Fluorescence intensities were normalized to the respective maximum intensity of each array. Almost no coinciding epitopes can be observed on the light chain. The white ellipses highlight the most prominent epitope of the heavy chain that can be found in several patients and was selected for the substitution analysis and isolation of the respective antibody.

One epitope is targeted in a band of successive peptide spots by antibodies of 8 individuals. Antibodies of one serum (serum 9) showed weak interaction with only the first of these successive peptide spots (see Figure 2 and Supplementary Figure S2-A). The identified epitope with the amino acid sequence $_{923}$IHLVNNESSEVIVHK$_{937}$ is located on the N-terminal domain of the C-terminal part of the heavy chain (HC-N) of the tetanus toxin. This epitope was further investigated in a substitution analysis.

3.2 Substitution analysis

In a substitution analysis, every amino acid in a known antibody-binding peptide is one by one exchanged by all other 19 amino acids (see Figure 3). For 15mer peptides, the resulting 300 peptides (15 x 20 amino acids) are synthesized on an array and incubated with the same serum as in the epitope mapping. Besides the variants, each row also contains the peptide with the original sequence of amino acids. The fluorescence intensity was normalized to the original peptide in each row, which is set to 100 %.

In the epitope mapping (see Figure 2), antibodies of 9 out of 19 sera showed interaction with the peptide $_{923}$IHLVNNESSEVIVHK$_{937}$. This peptide was substituted; the resulting arrays were synthesized in triplicates and incubated with the respective sera. Evaluation was done as described in Figure 3. The resulting antibody fingerprint of 8 sera is highly conserved and consists of the amino acids $_{929}$ExxEVIVxK$_{937}$, of which "x" represents an amino acid that is not crucial for the antibody-antigen interaction. The individual fingerprints of all sera are summed up in Table 1. Supplementary Figure S1 shows the evaluation of all substitution analyses as heatmaps. The role of the amino acid H$_{936}$ in the identified fingerprint is less stringent then the role of other positions. A broad range of interaction was observed across different sera. No conclusive correlation of the characteristics of the amino acids leading to binding or preventing the interaction in this position (hydrophobicity, charge and size) can be observed. The substitution analysis of serum 9 revealed a different fingerprint and is shown in Supplementary Figure S2. The fingerprint consists of only 3 amino acids ($_{923}$IxxVN$_{927}$) and the essential amino acids are shifted to the N-terminal part of the peptide.
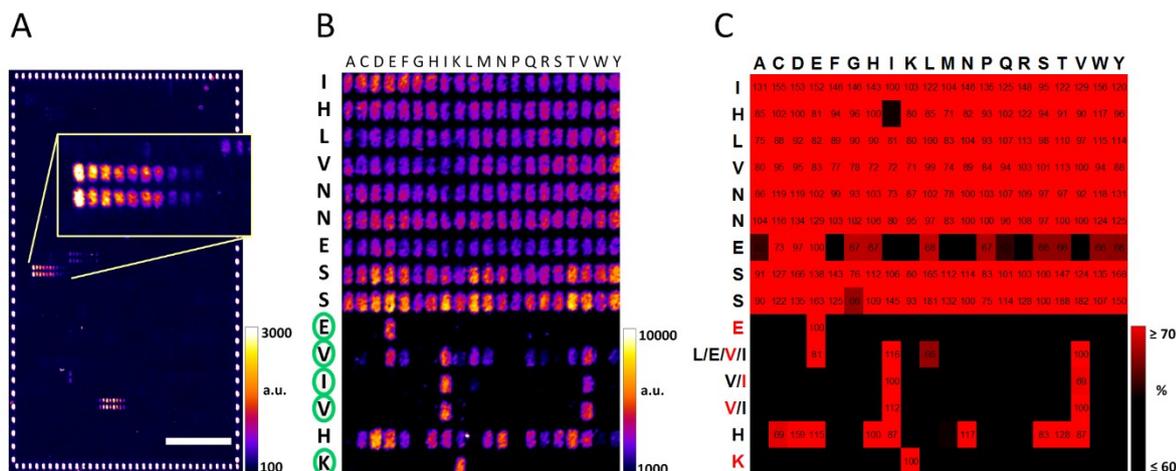
Figure 3: A: Example fluorescence scan image of an epitope mapping. A selected epitope, which was chosen for substitution analysis, is shown in detail. Scale bar 0.5 cm. B: Example fluorescence scan image of the corresponding substitution analysis of a 15mer peptide. Each amino acid position is substituted by all 19 other amino acids resulting in an array of 15 rows and 20 columns. From top to bottom left, the original sequence is shown. Those amino acid positions that are essential for binding of the antibody are highlighted with a green circle. C: Heatmap of the evaluation of the substitution analysis. The intensity of the peptide with the original amino acid in each row is set to 100 %. The relative fluorescence intensities are calculated in relation to these intensities. From top to bottom left, the original sequence is shown; letters in red indicate original amino acids that are essential for binding. In some positions two or more amino acids allow for the binding. Amino acids with similar binding effect are shown at the left side, sorted by binding intensity from left (lower) to right (higher).

3.3 Validation of antibody-peptide interaction

The interaction of serum antibodies with the identified prominent epitope of the heavy chain from the tetanus toxin was verified using a pre-synthesized peptide with the amino acid sequence IHLVNNESSEVIVHK-ßA-D-ßA-Peg-C$_{COOH}$, spotted onto functionalized surfaces. Arrays were incubated with those sera that showed interaction with the identified antigenic peptide (sera 2, 3, 7, 8, 9, 10, 11, 16, 17). The median fluorescence intensity of these peptide spots is depicted in Supplementary Figure S3. All sera showed interaction with the peptide in the validation experiment, except for serum 9. As negative control two peptides with substituted amino acids were synthesized and spotted. The substituted peptides were derived from the substitution analysis with the amino acid sequences IHLVNNESSEVIVHR-ßA-D-ßA-Peg-C$_{COOH}$ and IHLVNNESSEVIAHK-ßA-D-ßA-Peg-C$_{COOH}$. Very little to no interaction of serum antibodies with these peptides was detected. Supplementary Figure S4 shows examples of fluorescence scan images of peptide arrays incubated with serum 11 (positive) and serum 19 (negative control).

Table 1. Antibody fingerprints of 8 sera for an identified epitope. All amino acids that lead to the binding of the antibody are printed according to the relative fluorescence intensity from left to right (top to bottom) in increasing manner. If all substitution variants in a row lead to the binding of the antibody, the position is depicted with an x. If 15 or more amino acids allow for the binding in a certain position, only amino acids that prevent the binding are printed with a preceding exclamation mark. Amino acid positions that are essential for the binding of the antibodies of all individuals are highlighted in grey.

Serum							Original amino acid sequence								
	I	H	L	V	N	N	E	S	S	E	V	I	V	H	K
2	x	!I	x	x	x	x	AQSTWYGHPLCDE	x	x	E	LEVI	VI	VI	CSIVHETD	K
3	!A	x	!A	!A	!AKQ	x	ME	!AGR	!GP	E	VI	I	VI	KQYHCMEFIVSNTD	K
7	EVHCWIN	TCIEKGLRVHWFY	x	x	x	x	ILNSTVDE	x	x	E	TLVEI	VI	VI	QACFMKHNESITD	K
8	!AV	!A	!A	x	!AFGMY	ESDNIPTL	EDGL	!D	!G	E	VI	FID	V	!GP	CK
10	x	x	x	x	x	x	DE	x	x	E	VIE	I	VI	IVTHEND	K
11	!A	x	x	x	x	x	DE	x	x	E	VIE	I	VI	FQHMCKSVEINTD	K
16	FNDMRQTHLWYWVI	!C	x	!CFGK	QIHFWDSTNYL	HAYNWETGSD	E	!GNP	MAFTSLDE	E	V	I	IV	x	K
17	x	x	x	x	x	x	TNQDE	x	!G	E	EVI	I	VI	QSFMHIVTEDN	K

3.4 Isolation of epitope specific antibodies and verification of toxin specificity

To verify, that the identified antibody species was induced by the vaccination against tetanus, an antigenic peptide of the epitope (IHLVNNESSEVIVHK) was pre-synthesized, purified and immobilized on beads to isolate the corresponding antibodies of serum 11 by batch affinity chromatography. The isolated antibody species was tested for its ability to bind to the native tetanus toxin in an ELISA experiment. Therefore, an individual received a tetanus vaccination before the donation and the IgG anti-tetanus titer increased from 1.02 IU/ml before vaccination to 4.7 IU/ml six months after vaccination. The post-vaccination serum was purified by protein G batch chromatography prior to affinity chromatography. The concentration of isolated antibodies was calculated from OD (280 nm) to be 0.053 mg/mL. To control the efficiency of the affinity batch chromatography, the isolated antibodies were incubated on an array containing the peptides of the epitope mapping. The results of this epitope mapping are shown in Supplementary Figure S5. Fluorescent signals were observed only with peptides containing the essential amino acids that were identified in the substitution analysis. In the tetanus ELISA experiment, the purified antibodies resulted in an OD (450 nm)

of 0.85 at a dilution of 1:2, which is comparable to the OD (450 nm) of 0.74 for the respective serum prior to purification at a dilution of 1:1000.

3.5 Bioinformatic analysis of identified fingerprint

Antibody fingerprints allow for the query of public protein databases to identify cross-reacting antigens and organisms that might have induced the formation of the antibody in the first place. A query of the "UniProtKb/Swiss-Prot" protein database with the "ScanProsite tool" [94] for the motif "(DE)-x-x-E-(IV)-I-(IV)-(CDEHIMNSTV)-K" resulted in 47 hits (see Supplementary Table S2). This motif consists of the amino acids that are most conserved among all 8 individuals. Almost all of the resulting protein/organism hits are very unlikely to be the causative antigen, because they are either from exotic organisms or they would derive from an autoimmune reaction. However, the human rhinovirus 3 and 14 also share the identified motif, which is considered in the discussion.

3.6 Spatial position of fingerprint

Figure 4 shows the mapping of the identified fingerprints onto the three-dimensional structure of the tetanus toxin. It reveals that the amino acids of the identified epitope are either part of beta strands ($_{923}$IHLV$_{926}$ and $_{933}$VIVH$_{936}$) or coils between beta strands ($_{927}$ENNESSE$_{932}$ and K$_{936}$). The conformation of the sequence $_{923}$IHLVNNESSEVIVHK$_{937}$ is a U-turn, as can be seen in Figure 4A. The amino acid E$_{932}$ is the tip of the U-turn and $_{929}$ESSEVIVHK$_{937}$, representing the relevant part of the epitope, is fairly extended and might be the best fit for the CDR-grove of the antibody. The amino acids H$_{924}$, V$_{926}$-S$_{930}$, E$_{932}$, I$_{934}$ and H$_{936}$ are protruding from the protein. The conformation of the epitope hints to an extended or protruding surface as antigen binding site on the antibody.

4. Discussion

Clostridial neurotoxins, such as the botulinum and tetanus toxins are among the most poisonous toxins known [95]. To understand, how antibodies are able to neutralize toxins or any other infection, the mode of action of an antibody and the exact binding site is of great interest. To gain further insights into immune responses, we investigated potential binding sites of the tetanus toxin with the combination of epitope mappings and substitution analyses on peptide arrays. Furthermore, we verified the specificity of the identified epitope towards the tetanus toxin by conducting a tetanus ELISA with the isolated antibody species.

Our analyses of the antibody activity in 19 individuals towards the tetanus toxin revealed a broad distribution, ranging from the recognition of single peptides to distinct bands consisting of several consecutive peptides. These findings of a wide range of different antibodies are consistent with those of a recent study [96]. We found that the most distinct epitope ($_{923}$IHLVNNESSEVIVHK$_{937}$), recognized by antibodies of 9 different sera, is located on the N-terminal domain of the C-terminal part of the heavy chain (H_C-N). The C-terminal part of the heavy chain (H_C), consisting of two subdomains (H_C-C and H_C-N), facilitates neurospecific binding, the N-terminal part (H_N) is responsible for membrane translocation [95]. The exact role of the H_C-N domain remains elusive [97]. Even though, monoclonal antibodies, capable of neutralizing the toxin in mice, were found to be binding to the C domain of the C-terminal part of the heavy chain (H_C-C) [98, 99], antibodies binding to the identified epitope might play a synergistic role in the neutralization process. Polysialogangliosides and integral membrane protein receptors have shown to be involved in the binding process [95, 100, 101, 102]. The simultaneous binding of antibodies to ensure neutralization was described earlier [86, 103]. The role of the identified epitope should be further investigated in neutralization studies.

The homology of the identified antibody fingerprint ($_{929}$ExxEVIVxK$_{937}$), between different individuals is astonishing, especially when taking into account the randomness of the antibody development. The analysis of one serum revealed a different, less stringent antibody fingerprint, which could be explained by unspecific binding or a different HLA type of this patient. A larger pool of samples with determined HLA type should be investigated in future studies, to examine, if more individuals produce antibodies with this different fingerprint and if HLA types play a role in this context. The sequencing of the identified and isolated antibody species and a structural analysis of the antibody-antigen-complex could give further insights into the mechanism of binding, such as induced fit, and the role of less stringent amino acid positions, such as H_{936}.
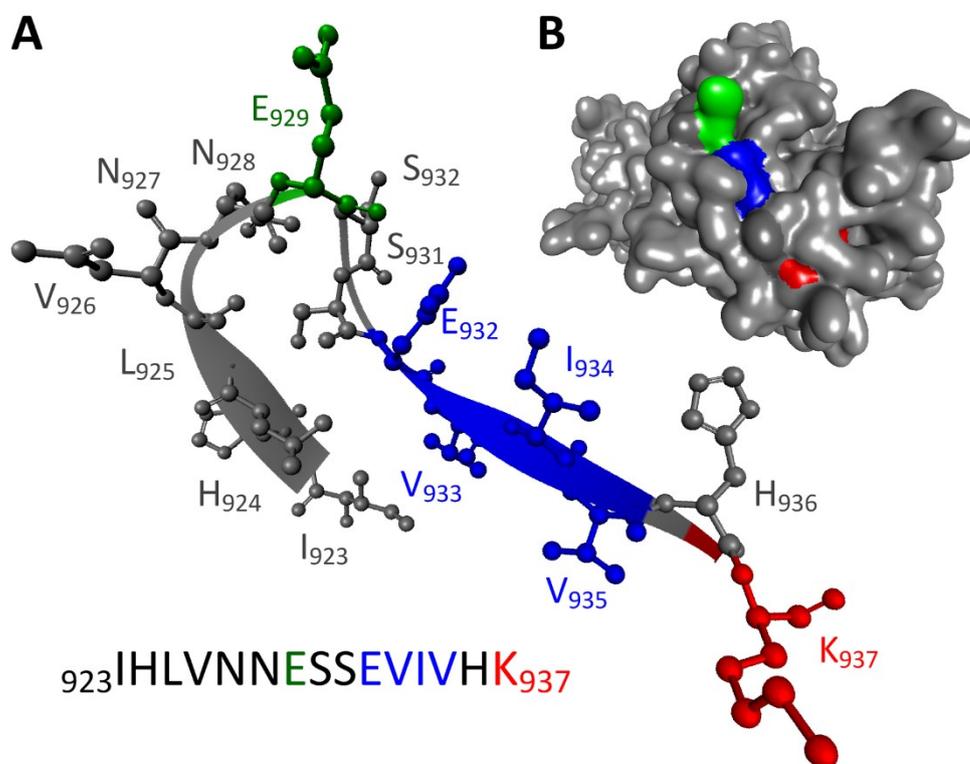
Figure 4: Three-dimensional structure of the heavy chain of the tetanus toxin. Conformation of the investigated epitope (A) and molecular surface (B) are depicted using BIOVIA Discovery Studio client [104], based on RCSB PDB (www.rcsb.org) ID 1A8D [105]. Fingerprint regions identified in the substitution analysis are highlighted in color.

Antibody fingerprints allow for the identification of potentially cross-reacting or cross-inducing antibodies and a database query with ScanProsite [94] revealed that the identified motif can also be found on the human rhinovirus strains 3 and 14. The sequence is located on the N-terminal part of the structural capsid protein VP1 [106]. The N-terminal part was shown to be immunogenic in rhinovirus strain 89, but it does not contain the specific amino acid sequence. VP1 of rhinovirus strain 14, which does contain the fingerprint, is less immunogenic [107]. Furthermore, the specific part is located on the inside of the virus, which is not part of the proposed receptor binding or neutralization site [107, 108]. Therefore, due to the great rhinovirus diversity of over 100 strains [98, 109] and the low protection against reinfection [110], we assume that the fingerprint does originate from the tetanus toxoid vaccination. It is very unlikely that 8 of our 19 individuals share the exact same antibody

specificity towards only two rhinovirus strains. Furthermore, we could verify the ability of the identified antibody species to bind to the native toxin with an additional ELISA experiment using the antibody after isolation in affinity batch chromatography.

Our study proves the suitability of peptide arrays to identify antibody species. Furthermore, these peptide targets might also serve as candidates for peptide based vaccines. The substitution analysis also revealed a new immunological insight that antibody species are highly conserved between different individuals. Antibody fingerprints enable queries of public protein databases for the identification of cross-reacting antigens and therefore, the fingerprints can serve as reliable, highly disease-specific biomarkers.

Competing interests

The authors declare competing financial interests: F.B. is shareholder of PEPperPRINT GmbH.

3.2 Antibody fingerprints in Lyme disease deciphered with high density peptide arrays

Laura K. Weber[1], Awale Isse[1], Simone Rentschler[1], Richard E. Kneusel[2], Andrea Palermo[1], Jürgen Hubbuch[3], Alexander Nesterov-Mueller[1], Frank Breitling[1]*, Felix F. Loeffler[1,4,5]*

[1]Institute of Microstructure Technology, Karlsruhe Institute of Technology, Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany

[2]DIARECT AG, Bötzinger Str. 29 B, 79111 Freiburg, Germany

[3]Institute of Process Engineering in Life Sciences, Section IV: Biomolecular Separation Engineering, Karlsruhe Institute of Technology, Engler-Bunte-Ring3, 76131 Karlsruhe, Germany

[4]HEiKA - Heidelberg Karlsruhe Research Partnership, Heidelberg University, Karlsruhe Institute of Technology (KIT), Germany

[5]Current address: Department of Biomolecular Systems, Max Planck Institute of Colloids and Interfaces, Am Mühlenberg 1, 14476 Potsdam, Germany

*Corresponding authors with equal contribution: frank.breitling@kit.edu +49 721 608-23859, felix.loeffler@mpikg.mpg.de +49 331 567-9359

Practical application

Antibody fingerprints define the specificities of antibodies in the resolution of a single amino acid and allow for the comparison of antibody responses in different individuals on sub-epitope level. They can be deciphered with substitution analyses, carried out with flexible high density peptide arrays by substituting every amino acid in an antigenic peptide with all other amino acids. As these fingerprints enable queries of public protein databases, the study of cross-targeted (human) proteins becomes possible. Therefore, the readout of antibody fingerprints can serve as a novel diagnostic tool to investigate disease etiology. As a proof-of-principle, this study evaluated the feasibility of antibody fingerprints for diagnostic purposes in Lyme disease.

Abstract

Lyme disease is the most common tick-borne infectious disease in Europe and North America. Previous studies discovered the immunogenic role of a surface-exposed lipoprotein (VlsE) of Borreliella burgdorferi. We employed high density peptide arrays to investigate the antibody response to the VlsE protein in VlsE-positive patients by mapping the protein as overlapping peptides and subsequent in-depth epitope substitution analyses. These investigations led to the identification of antibody fingerprints represented by a number of key residues that are indispensable for the binding of the respective antibody. This approach allows us to compare the antibody specificities of different patients to the resolution of single amino acids. Our study revealed that the sera of VlsE-positive patients recognize different epitopes on the protein. Remarkably, in those cases where the same epitope is targeted, the antibody fingerprint is almost identical. Furthermore, we could correlate two fingerprints with human autoantigens and an Epstein-Barr virus epitope; yet, the link to autoimmune disorders seems unlikely and must be investigated in further studies. The other three fingerprints are much more specific for B. burgdorferi. Since antibody fingerprints of longer sequences have proven to be highly disease specific, our findings suggest that the fingerprints could function as diagnostic markers that can reduce false positive test results.

1. Introduction

Lyme disease is the most common tick-born disease in Europe and North America and yet remains a problematic topic [111, 112]. The illness is caused by different species of the spirochete *Borreliella burgdorferi*, referred to as *B. burgdorferi* sensu lato. Diagnosis in absence of the one distinct clinical marker "erythema migrans" remains controversial and lacks in either sensitivity, specificity, or both [113, 114]. Due to limited success in direct detection of *B. burgdorferi* sensu lato, serological antibody-based tests are mainly applied for clinical diagnosis [115]. The Centers for Disease Control and Prevention (CDC) in the USA recommend two-tier testing [116]. This diagnosis applies enzyme immunoassays or immunofluorescent assays as a pre-test and Western immunoblotting for validation. Differential expression of proteins and observer-dependent results are among the limiting factors of western blotting [117]. The use of recombinant proteins in line immunoblot assays [118] partly overcomes those problems. Yet, cross-reacting proteins are a drawback of all serological methods [119, 120]. False positive results may occur especially when the patients suffer from other illnesses, such as syphilis [121] or viral infections [122]. Antigenic diversity of the sensu lato species [123, 124, 125]] and antigenic variation of these spirochetes, which refers to the ability of a single strain to express different variants of antigens at a much higher rate than the mutation rate [126] make diagnosis even more difficult. The Vmp-like sequence expressed (VlsE) antigen has shown to be very immunogenic and, yet, also very variable, due to genetic recombination of the vlsE cassette region [127]. The VlsE antigen (see Figure 2) consists of two invariable domains at the N- and C-terminal end and a variable domain in between, interspersed by 6 conserved invariable regions (IR 1 – IR6) [127]. Previous studies suggested that most immunodominant epitopes are located on conserved parts of VlsE, however, the variable parts influence the overall antigenicity [128]. Diagnosis, based on VlsE and a synthetic 26mer peptide (C6), which essentially represents the IR6 sequence derived from the VlsE antigen, is promising [129] especially regarding the discrimination between early and chronic Lyme disease patients [130].

Diagnosis and consequent treatment with antibiotics, especially in early stages of the illness, are necessary to prevent late manifestation and complications [131, 132]. Therefore, the development of simple and accurate tests for serodiagnosis, for example, based on a diagnostic peptide marker, is urgently needed.

In this study, we combined a peptide array-based epitope mapping of the VlsE antigen with subsequent in-depth epitope substitution analyses to determine the antibody fingerprints to immunodominant epitopes. We employed a solid-material based synthesis with a laser printer

[133] to display the whole antigen with 15mer peptides that overlap by 14 amino acids. In the second step, every position of an identified antigenic peptide is substituted by all other amino acids and the resulting peptides are incubated again with the patient serum. The evaluation of the fingerprints with the substitution analysis defines the antibody specificity with single amino acid resolution and allows for the query of public protein databases to identify potentially cross-targeted proteins. The substitution (or replacement) analysis was established in 1987 by Geysen *et al.* [76]. In a recent study, these fingerprints were used to determine the specificity of antibodies towards the viral capsid protein (VP1) of enteroviruses [134]. They have shown to be helpful in the identification of potential cross-reactivity of these antibodies, e.g. with human proteins. However, this fine mapping has not yet been applied to investigate serum antibodies of large groups of individuals.

Our experiments revealed that similar epitopes on the VlsE antigen are targeted, as shown in previous studies [130]. However, for the first time, we compared the antibody specificities on sub-epitope level. The findings that fingerprints are almost identical in different patients give new immunological insights. Besides a limited number of proteins and epitopes of a pathogen, we show that also a limited number of amino acids are commonly preferred by the immune systems of different individuals. Furthermore, the investigation of antibody fingerprints could help in elucidating the remaining question, why some Lyme disease patients show persistent symptoms despite antibiotic treatment, possibly caused by autoimmune cross-reactions. To address the latter, we performed queries of public protein databases and the human autoantigen atlas, which revealed potentially cross-targeted human proteins that play a role in autoimmune disorders. Additionally, the public protein databases were queried for typical cross-reacting viral proteins and three of five fingerprints are specific for *B. burgdorferi*. Therefore, our study serves as proof-of-principle for the investigation of antibody responses in groups of diseased individuals with epitope mappings, substitution analyses, and subsequent bioinformatics for diagnostic purposes and etiology.

2. Material & Methods

2.1 Study Design

In this study, we investigated the antibody response of 24 individuals towards 15mer peptides, representing the whole amino acid sequence of the VlsE antigen of B. burgdorferi strain B31 with a lateral shift of one amino acid. Commercially available patient sera from Europe and Northern America were provided by the company DIARECT AG (Freiburg, Germany). 17 of these sera were tested VlsE positive with protein arrays by the company DIARECT AG

(Freiburg, Germany). Sera of 7 VlsE negative patients served as negative controls. Peptide arrays were produced in spot duplicates by the company PEPperPRINT GmbH (Heidelberg, Germany) with a particle-based solid-phase peptide synthesis approach in a laser printer [133]. In the second step, immunodominant epitopes shared by several patients were fine mapped in a substitution analysis to investigate the antibody specificity in a single amino acid resolution, which we call the antibody fingerprint. The antibody fingerprints were visualized on the three-dimensional structure of the antigen with PyMol (Schrödinger, pymol.org/dsc/) to reveal the spatial position and accessibility of the epitopes.

2.2 Peptide arrays – design & synthesis

The linear amino acid sequence of the VlsE antigen of B. burgdorferi (accession number O06878, integrated into UniProtKB/TrEMBL: July 1, 1997, access February 2016) was synthesized as overlapping peptides in spot duplicates (15mer peptides with an overlap of 14 amino acids). To ensure proper interaction with the C-terminal part of the antigen, a spacer, consisting of glycine and serine (GSGSGS), was added. As controls, alternating Hemagglutinin (HA) peptide spots (YPYDVPYAG) and a peptide, derived from the polio virus (KEVPALTAVETGAT), were synthesized surrounding the array content. The peptide spot size was 200 µm x 400 µm with a spot-to-spot distance of 250 µm x 500 µm. For the substitution analysis, amino acids of the immunodominant epitopes were substituted one by one by all other amino acids and synthesized in spot duplicates. Thereby, only one amino acid at a time is exchanged, the rest of the sequence remains conserved.

2.3 Incubation of peptide arrays

For all incubation, washing, and blocking steps, glass slides with peptide arrays are mounted into PEPperCHIP incubation trays (PEPperPRINT, Heidelberg, Germany). The trays allow for the subdivision of a glass slide into separate wells. The tray is placed on an orbital shaker and rotated at 140 rpm (Orbital Shaker DOS-20S, Elmi Ltd., Riga, Lettland). Prior to the initial incubation, the slide is swelled in phosphate buffered saline with Tween 20 (PBS-T: 1 x PBS, Sigma Aldrich, Saint Louis, USA, pH 7.4; 0.05 % v/v Tween 20, Sigma Aldrich) for 10 min and blocked with blocking buffer for fluorescent western blotting (Rockland Immunochemicals Inc., Pottstown, USA) to inhibit unspecific binding. To determine the potential interaction of the fluorescently labelled secondary antibodies with any peptide, a pre-staining after blocking is performed. Secondary antibodies (Goat anti-human IgG-Fc-Fragment DyLight®680 conjugated, Biomol, Hamburg, Germany) diluted to 0.2 µg/mL in

PBS-T with 10 % v/v blocking buffer are incubated for 30 min followed by a washing and a subsequent scanning step (see 2.4). After a short washing step with PBS-T, the slides are incubated with serum diluted 1:500 for the epitope mapping and 1:100 for the substitution analysis in PBS-T with 10 % v/v blocking buffer over night at 4 °C. Afterwards, the slides are washed again in PBS-T and incubated with secondary antibodies at the same concentration as mentioned above and monoclonal Anti-HA antibodies (provided by Dr. Gerd Moldenhauer, DKFZ, Heidelberg) labelled with a DyLight™ 680 Microscale Antibody Labeling Kit (Thermo Fisher Scientific, Rockford, USA) and diluted to 1 μg/mL. Finally, the slides are washed with PBS-T, dipped into ultra-pure water and dried in a stream of argon. For storage, slides are kept under argon atmosphere at 4°C.

2.4 Image analysis & evaluation

Fluorescence scanning of the slides is performed using an odyssey scanner (LI-COR Biosciences, Lincoln, USA) with a resolution of 21 μm/pixel at a wavelength of 680 nm. The resulting 16-bit greyscale images are analyzed with the PepSlide Analyzer (Sicasys Software GmbH, Heidelberg, Germany). A grid, containing the information of sequence and position of a peptide, is positioned over the greyscale image. The median of the greyscale values of all pixels in a spot is measured and exported as raw intensity. A global background consisting of the intensity of 20 control spots above and 20 under the arrays respectively is subtracted from the raw intensity to give the foreground intensity. The intensity of spot duplicates is averaged. Outliers caused by dirt can be detected with differences in the foreground median intensity of spot duplicates. Therefore, the intensity of a peptide spot is set to twice the average spot intensity of an array if (1) the intensities of spot duplicates differ by more than 65 % and (2) the intensities of spot duplicates are higher than twice the average intensity of all spots of the array. In the substitution analysis, the intensity of the peptide with the original amino acid in each row is set to 100 %. The intensities of all other spots in a row are normalized with respect to the intensity of the peptide with the original amino acid. If the substitution leads to a decrease in intensity of more than 40 %, the binding is considered to be eliminated.

3. Results

The workflow of this study is depicted in Figure 1. First, we performed an epitope mapping of the VlsE antigen of Borreliella Burgdorferi sensu stricto with sera from 24 different individuals (17 VlsE positive, 7 VlsE negative samples). The identified immunodominant epitopes (Figure 1-A) were fine mapped in a substitution analysis (Figure 1-B). The substitution analysis was evaluated as described in the Image analysis & evaluation section and reveals the antibody fingerprint (Figure 1-C). This fingerprint defines the antibody specificity at single amino acid resolution.
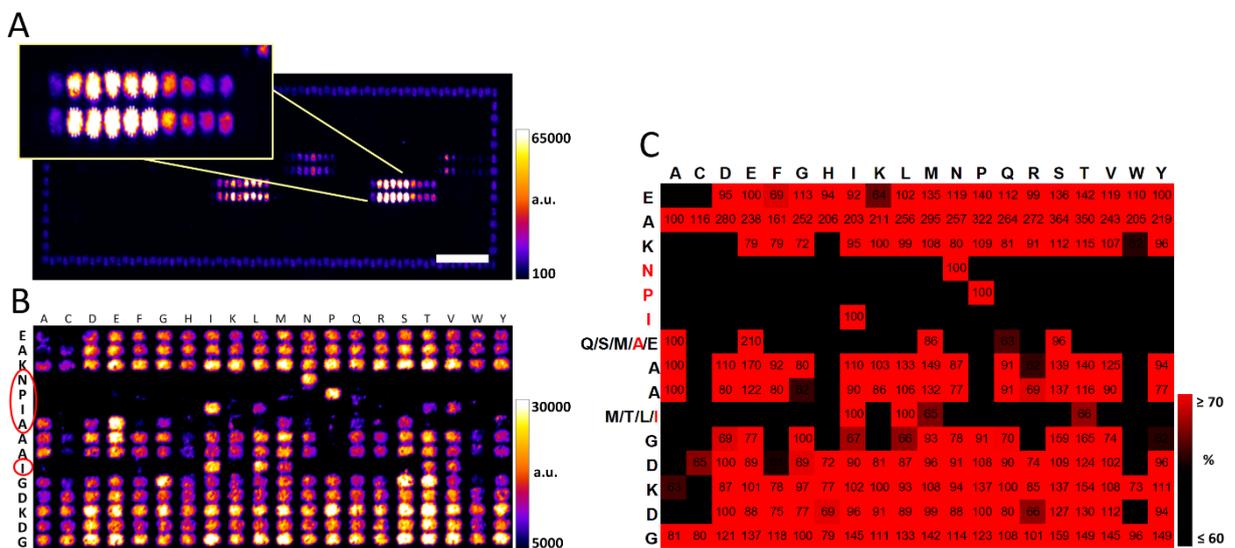
Figure 1: Workflow of the epitope mapping (A) and subsequent substitution analysis of an identified epitope (B) with evaluation (C). For the epitope mapping, the amino acid sequence of the VlsE antigen of B. burgdorferi was displayed as 15mer peptides with a lateral shift of one amino acid. (A) Fluorescence scan image of a peptide array incubated with a VlsE positive patient serum and fluorescently labelled secondary antibodies. Peptide spots were produced as duplicates. The zoomed view shows one of the immunodominant epitopes, identified in several patients. Scale bar 0.5 cm. (B) Fluorescence scan image of a peptide array for the substitution analysis of the immunodominant epitope shown in detail in (A). The original amino acids of the peptide sequence are substituted one by one by all other amino acids. The resulting peptides are printed as array duplicates (not shown) and incubated with the corresponding serum. (C) The results of (B) visualized as a heat map. Evaluation of the substitution analysis reveals the antibody fingerprint, showing the antibody specificity at single amino acid resolution. The relative fluorescence intensities of all peptides in a row are calculated according to the intensity of the peptide with the original amino acid set to 100 %. (B and C) The original sequence is shown from top to bottom left; essential amino acids are indicated in red. Substituted amino acids are shown from left to right on top of the substitution arrays.

3.1 Epitope Mapping

The antibody response of 24 different individuals towards the VlsE antigen was investigated with peptide microarrays. Figure 2 shows the results of this epitope mapping in form of a heat map. Seven VlsE negative patients (NC 1-7) served as controls. Two main groups of patients were identified based on the epitopes that were targeted by antibodies. Antibodies of group 1 (PS 1 – 7) preferably targeted N- or C-terminal parts of the antigen or both, highlighted in orange and green. These parts comprise the amino acid positions 20 to 40 with the sequence INCKSQVADKDDPTNKFYQSV and positions 338 to 357 with the sequence LRKVGDSVKAASKETPPALN. In the second group (PS 9 – 14), the antibodies preferably bound to regions on the variable domain of the antigen (pink and cyan), comprising positions 228 to 242 (VSGEQILSAIVTAAD) or positions 249 to 267 (KKPEEAKNPIAAAIGDKDG) or both. The antibody responses of PS 8 and 15 – 17 are not classifiable as part of the two main groups. Only little interaction was observed with peptides of the IR6. For a better comparison of the sera, the fluorescence intensities were normalized to the respective maximum intensity of each array. Maximum fluorescence intensities are listed in Supplementary Table S1. Figure 3 shows four examples of original fluorescence scan images of the peptide array epitope mappings. Please note the different maximum intensities. The negative controls did not show any significant fluorescent signals compared to patient samples 1 to 15. The antibody response of two patients (PS 16 and 17) were comparable with the negative controls with regard to the maximum signal intensities and the distribution of fluorescence signals over the array.

3.2 Substitution analysis

In the second step, 5 immunodominant epitope regions (highlighted in different colors in Figure 2) were fine mapped in a substitution analysis. Table 1 gives an overview of the specific peptides that were targeted by antibodies of different sera, which were investigated in the analysis to reveal the antibody fingerprints. Epitopes that were only targeted by antibodies of one serum were not substituted. Due to the very weak overall interaction of PS 16 and 17 that was similar to the negative controls, no substitution analysis was performed.

The antibody fingerprints of patients towards different epitopes on the VlsE antigen are listed in Table 2. The evaluation of all fingerprints, as shown in Figure 1, is depicted in Supplementary Figures S1 – S5.
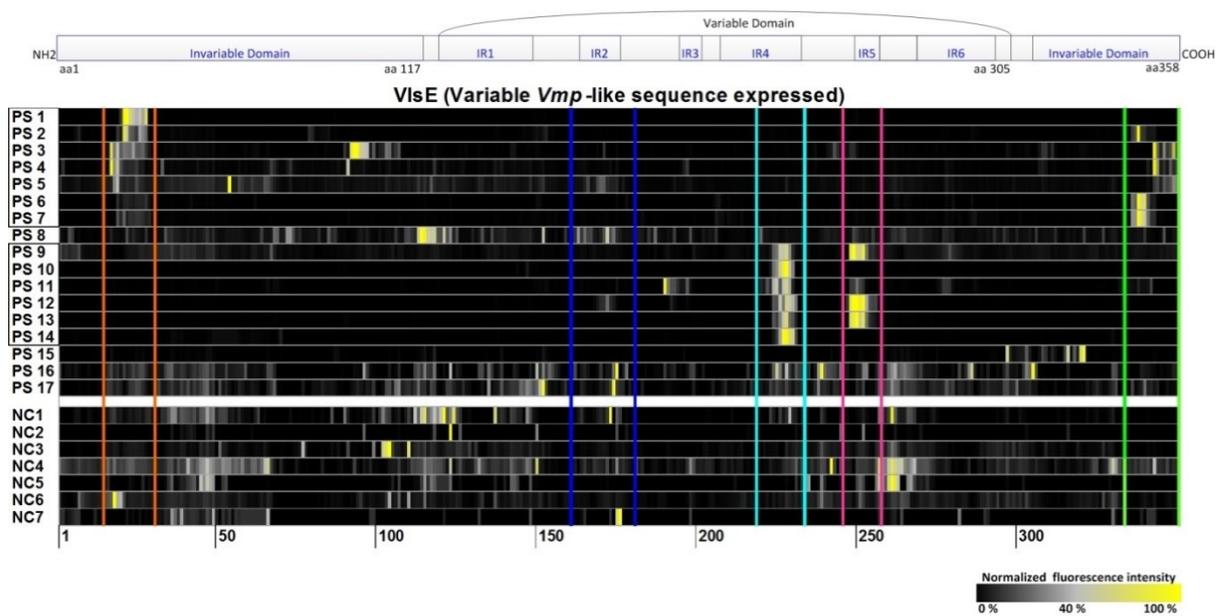
Figure 2: Epitope mapping of 17 VlsE positive patient sera (PS) and 7 negative control sera (NC). From left to right, the position for each peptide in the VlsE protein is depicted. Fluorescence intensities were normalized to the maximum of each array. Increasing relative intensities are depicted from black to yellow. Patients were grouped according to the antibody response (Group 1: PS 1-7 and Group 2: PS 9-14, PS 8 and 15 – 17 not classifiable). The five immunodominant parts of the VlsE antigen that were chosen for fingerprint analyses are highlighted in different colors. Top part of the figure is based on [135].
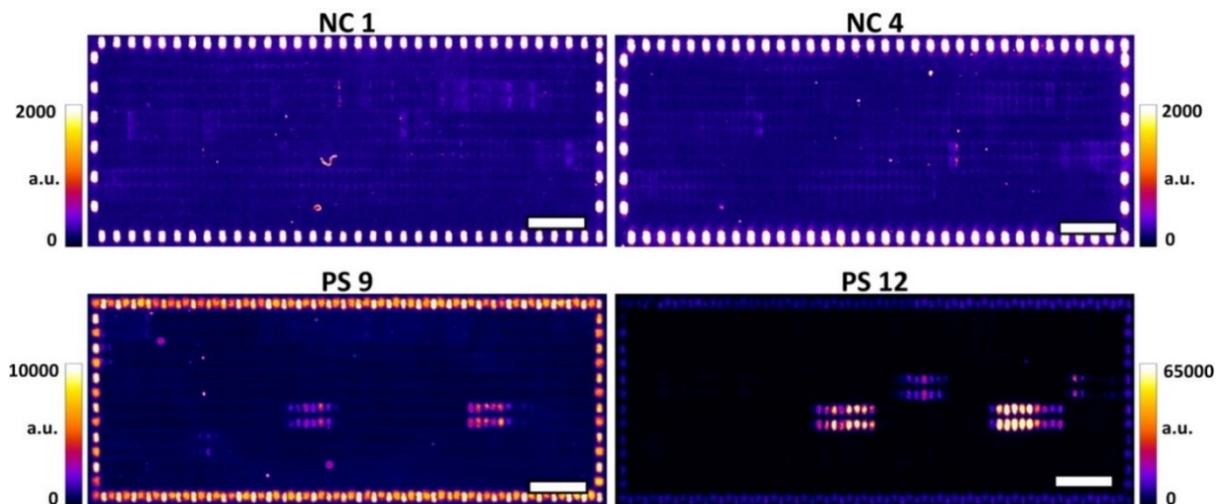
Figure 3: Original fluorescence scan images of 4 exemplary epitope mappings. Peptide microarrays were incubated with sera and fluorescently labelled secondary antibodies. Maximum intensities differ from 130 a.u. (NC 4) to 65,300 a.u. (PS 12). Control frames consisted of alternating peptide spots with an amino acid sequence derived from the polio virus. Antibodies of most individuals show strong interaction with this peptide due to the common vaccination against polio (scale bar 0.5 cm).

Table 1: Overview of peptides and sera that were investigated in substitution analyses. Peptides are grouped according to the position of the amino acids in the antigen. For each serum and epitope, the peptide with the highest fluorescence intensity was chosen for the substitution analysis.

	VlsE 20 - 40	VlsE 171–186	VlsE 228-242	VlsE 249-267	VlsE 338-357
PS 1	VADKDDPTNKFYQSV				
PS 2	VADKDDPTNKFYQSV				RKVGDSVKAASKETP
PS 3	INCKSQVADKDDPTN			KKPEEAKNPIAAAIG	DSVKAASKETPPALN
PS 4	INCKSQVADKDDPTN				DSVKAASKETPPALN
PS 5		KGIKEIVEAAGGSEK			DSVKAASKETPPALN
PS 6	INCKSQVADKDDPTN				LRKVGDSVKAASKET
PS 7	INCKSQVADKDDPTN				LRKVGDSVKAASKET
PS 8		GIKEIVEAAGGSEKL			
PS 9			VSGEQILSAIVTAAD	PEEAKNPIAAAIGDK	
PS10			VSGEQILSAIVTAAD		
PS 11			VSGEQILSAIVTAAD	PEEAKNPIAAAIGDK	LRKVGDSVKAASKET
PS 12		GIKEIVEAAGGSEKL	VSGEQILSAIVTAAD	EAKNPIAAAIGDKDG	
PS 13			VSGEQILSAIVTAAD	PEEAKNPIAAAIGDK	
PS 14			VSGEQILSAIVTAAD		

3.3 Spatial position of antibody fingerprints

Amino acids that were identified in the substitution analysis to be essential for the binding of antibodies of several patients are highlighted in color on the three-dimensional structure of two chains of the tetrameric VlsE antigen (Figure 4). These amino acids are part of the 5 epitopes that were identified in the epitope mapping. All essential amino acids are part of conserved alpha-helical structures. The structure of the N- and C-terminal parts seems to be quite flexible [135]. The N- and C-terminal epitopes (green and orange) are located on the surface at the membrane proximal site of the antigen. The other 3 epitopes (pink, cyan and blue) are oriented primarily towards the inside of the protein, partly shielded by variable loop structures. However, the molecular surface depicted in Figure 4-B and 4-D also contains some residues that are accessible from the outside. The amino acids that are essential for the binding of the antibodies are preferably located on only one side of the helical structures.

3.4 Bioinformatic analysis of fingerprints

The investigation of the exact amino acids that are essential for the binding of the antibody to its epitope, allows for a bioinformatic analysis of potential cross-reacting proteins. Therefore, the query can either help in ruling out cross-targeted organisms for diagnostic purposes or serve for the identification of cross-targeted antigens. A query of the "UniProtKb/TrEMBL" (release 18-Jan-17: 73711881 entries) and "UniProtKB/Swiss-Prot" (release 18-Jan-17: 553474 entries) protein databases with the "ScanProsite tool" [106] for the identified fingerprints resulted in numerous hits on human proteins. All potentially cross-reacting human proteins are listed in Supplementary Tables S2 – S6. Table 3 lists the fingerprints, which were used for the database query.

For the identification of human proteins that are involved in already known autoimmune disorders, all database entries of the query were cross-checked with the AAgAtlas, an online database of the National Center for Protein Sciences (Beijing) [136]. The database contains 1,126 autoantigens that were collated from the literature. Table 3 additionally lists all proteins of the database query that potentially play a role in autoimmune disease and might be connected to antibodies induced by the VlsE antigen.

It has been reported that among others, infections with the human parvovirus 19 [137], herpes simplex virus 2 [138], varicella-zoster virus [139], Epstein-Barr virus, and cytomegalovirus [140] cause similar symptoms as Lyme disease and also cause false positive Lyme disease test results due to antibody-based cross-reactions. Therefore, the "UniProtKb/TrEMBL" (release 18-Jan-17: 73711881 entries) and "UniProtKB/Swiss-Prot" (release 18-Jan-17: 553474 entries) protein databases were queried with the "ScanProsite tool" for homologies of the fingerprints with the latter viruses. The short motif D-P-T-N can be found on proteins of cytomegalovirus and varicella-zoster virus. The fingerprint I-x-(ADE)-I-x-(ADE)-A-A is shared by the EBN_A-3A protein of Epstein-Barr virus (A0A0C7T0J9), all other fingerprints are specific for B. burgdorferi, when compared to the mentioned viruses.

4. Discussion

The role of the VlsE antigen in Lyme disease remains elusive. Nonetheless, it serves as a reasonable diagnostic marker either as protein or as peptide fragments (C6 of IR6 region) [129]. With substitution analyses, we have investigated the role of single amino acids regarding their importance for antibody binding to immunodominant epitopes (also see [120]). In this present study, we demonstrated for the first time that, if the same VlsE epitope is targeted by antibodies, almost identical amino acids of the epitope are crucial for the formation of the antibody-antigen complex in different individuals. A comparison of these essential amino acids in three epitopes located on the variable domain with a sequence similarity analysis [127] revealed that these amino acids are located on conserved regions on the antigen between its variable parts.

However, our epitope mapping revealed little to no interaction with peptides covering the IR6 region. The established C6 ELISA test is based on a 26mer peptide, which could explain our lack of antibody responses with only 15mer peptides. Yet, Chandra *et al.* [130] observed interactions of almost all investigated PTLDS and post-Lyme healthy patients with 14mer peptides covering the IR6 region. These differences can be explained with (1) in our case

lower concentrations of antibodies and (2) different surface properties. It is known that surface properties such as the density of peptides influence peptide array experiments [141]. In comparison to Chandra et al., our number of positive signals was much lower, which is most likely caused by a specific surface modification.
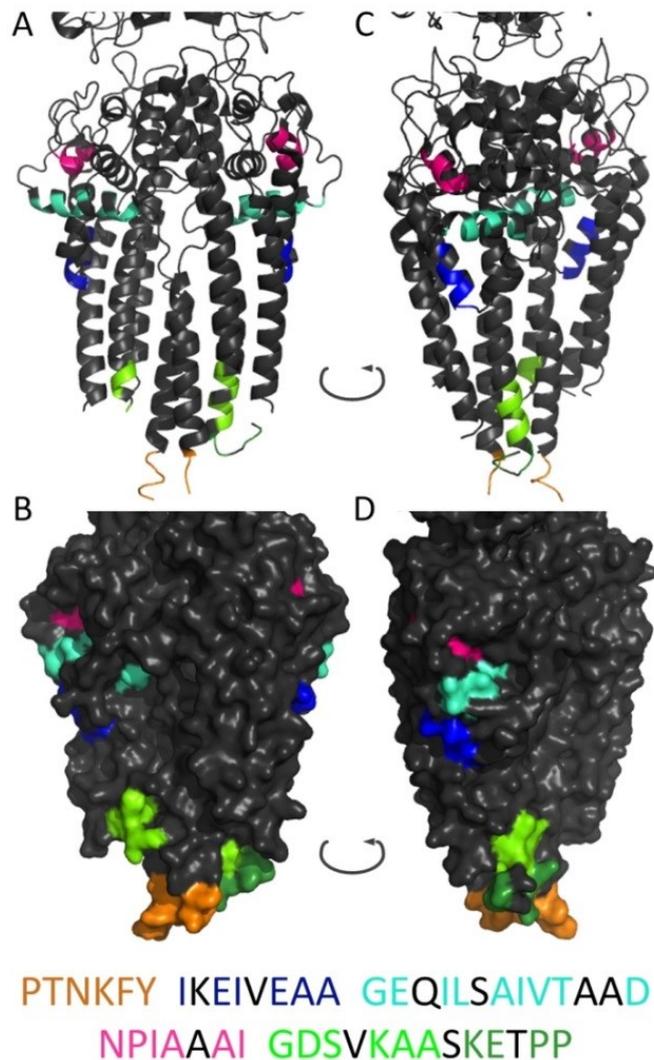
Figure 4: Three-dimensional structure of the VlsE antigen. Detailed view of the secondary structure (A and C) and molecular surface (B and D) are depicted using the PyMol (Schrödinger, pymol.org/dsc/) molecular visualization system, based on RCSB PDB (www.rcsb.org) 1L8W [135] Conserved amino acids of the fingerprint regions identified in the substitution analysis are highlighted in the specific colors from Fig.2 and face mainly to one side of the protein. The figure shows two of four subdomains of the tetrameric protein. The structure of the N- and C-terminal end has not been completely defined and differs in the subdomains. To illustrate the spatial position of the epitope at the N-terminal amino acid position 20 – 40, the fingerprint PTN (orange) was extended by the amino acids KFY.

Especially, peptides with one or more positively charged (e.g. lysine) residues tend to induce strong unspecific interaction (personal communication). To prevent unspecific antibody interaction with peptides containing positively charged residues, PEPperPRINT usually synthesizes an aspartic acid as the C-terminal amino acid to every peptide. This may cause a loss in binding events in this particular case, but it also increases the overall specificity of the assay.

In the study of Chandra *et al*. [130], the N- and C-terminal parts of VlsE were preferably targeted by one patient group, in addition to the epitopes on the variable domain, which were targeted by both groups. Based on 4 prominent epitopes, we could identify in a similar manner two main groups targeting either the N- and C-terminal parts (7 patients) or conserved regions in the variable domain (6 patients). However, due to the lack of patient disease history, no conclusion can be drawn from these findings. Two patients only recognize epitopes that cannot be classified as part of the two main groups. Another two of the 17 VlsE positive patients show similar signal intensities as the negative controls, leaving the questions open, whether these are truly positive samples or whether a linear peptide approach is applicable for these samples.

Nevertheless, our fingerprint analyses enable queries of databases for potentially cross-targeted human proteins. The query of the "UniProtKb/TrEMBL" and "UniProtKB/Swiss-Prot" protein database with the "ScanProsite tool" revealed several human proteins sharing the identified fingerprint D-P-T-N of the epitope VlsE 20 – 40, of which 5 proteins have shown to play a role in autoimmune disorders. The number of identified proteins for this epitope is large due to the short sequence of the fingerprint of only four amino acids. However, a relation between antibodies induced to this epitope and autoimmune reactions cannot be ruled out and must be investigated in further studies. Additionally, the substitution analysis allows for the identification of microorganisms and viruses inducing cross-reacting antibodies that cause false positive results in standard Lyme disease tests. As three of the identified fingerprints are specific only for *B. burgdorferi*, they can serve as reliable diagnostic marker preventing false positive results. Future studies with peptide arrays, representing all known variations of the antigen, can give further insights.

Concluding, our results clearly show that there are at least two groups of individuals targeting two different domains of the VlsE antigen. Yet, if the same domain is targeted, the essential amino acids responsible for the epitope binding are highly conserved. Therefore, the study serves as proof-of-principle for the disease specificity of fingerprints on VlsE sub-epitope level and proves their potential for diagnostic purposes and cross-reactivity studies.

Table 2: Antibody fingerprints of sera for the substituted epitopes on the VlsE antigen. The bold and colored printed rows show the original amino acid sequence of the substituted peptides. The colors (compare Figure 2) indicate different positions on the antigen. For each serum and position, the amino acids that allow for the binding of the antibody ($\geq$ 60 % relative fluorescence intensity) are listed in ascending order of the relative intensity. If 7 or more amino acids allow for the binding, the position is indicated with an x. The exact fingerprints with all amino acids allowing for the binding are shown in Supplementary Figure S1 – S5. Conserved amino acids required for detection by the sera are highlighted in grey.

Serum	I	N	C	K	S	Q	V	A	D	K	D	D	P	T	N	K	F	Y	Q	S	V
1							x	x	x	K	x	D	P	T	N	x	F	x	x	x	x
2							x	x	x	x	x	SYWRD	x	SDWYT	x	x	x	x	x	x	x
3	DFACI	x	x	x	x	x	x	x	D	x	D	D	P	T	N						
4	x	x	x	x	x	x	x	x	D	x	D	D	P	x	x						
6	x	x	x	x	x	x	x	x	x	x	x	D	P	MT	N						
7	x	x	x	x	x	x	x	x	x	x	EFYD	D	P	T	N						
	K	**G**	**I**	**K**	**E**	**I**	**V**	**E**	**A**	**A**	**G**	**G**	**S**	**E**	**K**	**L**					
5	x	NGSTED	I	x	DE	I	V	ED	AE	A	x	x	x	x	x	x					
8		x	x	x	x	x	x	x	x	x	x	x	x	E	x	IL					
12		x	VWFIL	x	AED	I	x	ADE	A	A	AG	DTGAS	TAS	x	x	x					
	V	**S**	**G**	**E**	**Q**	**I**	**L**	**S**	**A**	**I**	**V**	**T**	**A**	**A**	**D**						
9	x	x	x	ADE	x	FDAVI	LS	QYTAS	SA	AIV	x	x	x	x	ID						
10	PIYSTV	RVSPT	ATGS	SDE	x	IV	L	x	x	LVI	IVST	SVT	x	x	x						
11	x	S	G	DE	SMQED	VI	L	x	x	x	ITEV	SVET	x	x	VETYD						
12	x	x	QGP	E	x	I	L	x	A	VLI	MSIVLT	EST	x	x	YVEDST						
13	x	EDST	AGPS	DE	x	LIV	L	AMTDSE	A	I	VT	DVTSE	x	x	STYED						
14	x	x	G	TSNED	x	I	L	x	LDAME	I	VI	ST	MTS	x	x						
	K	**K**	**P**	**E**	**E**	**A**	**K**	**N**	**P**	**I**	**A**	**A**	**A**	**I**	**G**	**D**	**K**	**D**	**G**		
3	x	x	AIQP	AE	E	x	K	x	P	x	x	x	x	VPTI	x						
9			x	x	x	x	x	N	APT	I	DQTMAE	x	A	IL	x	ESTD	x				
11			x	x	x	x	MK	N	P	I	SAMQE	ADE	A	LI	x	MIED	x				
13			x	x	x	x	x	N	P	I	MDQSAE	VTDSAE	A	MTIVL	x	x	x				
12			x	x	x	x	x	N	P	I	QSMAE	x	x	MTLI	x	x	x	x	x		
	L	**R**	**K**	**V**	**G**	**D**	**S**	**V**	**K**	**A**	**A**	**S**	**K**	**E**	**T**	**P**	**P**	**A**	**L**	**N**	
2		x	x	x	x	x	x	x	x	WYCFDA	A	x	x	x	x	x					
6	x	x	x	x	G	ED	SGA	x	RMK	A	A	IVTS	K	x	x						
7	x	x	x	x	G	ED	SGA	x	VSRTKM	A	EA	VEST	K	x	x						
11	x	x	x	x	G	DTSN	S	x	DK	x	X	x	x	KLFCDE	x						
3						x	x	x	x	x	x	x	K	E	x	P	P	x	L	N	
4						x	x	x	x	x	x	x	x	E	x	P	ASP	x	PL	N	
5						x	x	x	x	x	x	x	K	x	TDS	P	QANSPD	GDTAPS	FYL	N	

Table 3: Fingerprints used for the query of the "UniProtKb/TrEMBL" and "UniProtKb/Swiss-Prot" protein databases with the "ScanProsite tool" on all human proteins. The fingerprints with the respective amino acid sequences that were used for the queries are shown in column 2. All resulting hits are listed in Supplementary Tables S2 – S6. Human autoantigens that potentially cross-react with antibodies assumed to be originally induced by the VlsE antigen of B. burgdorferi are listed in column 3. Column 5 shows exemplary autoimmune disorders that are presumably connected with the respective autoantigen.

Epitope	Fingerprint	Protein (Gene Code)	Exemplary disease
VlsE 20 – 40	D-P-T-N	Epididymis secretory sperm binding protein Li94n, Stress-induced-phosphoprotein 1 (Hsp70/Hsp90-organizing protein) (STIP1)	Autism spectrum disorder [142]
		Glycoprotein receptor gp330, Megalin, (LRP2)	Thyroiditis [143]
		Proprotein convertase subtilisin/kexin type 1 (NEC1) (PCSK1)	Endocrine system disease [144]
		Ubiquitin conjugation factor E4A (UBE4A)	Crohn's disease [145]
		Ubiquitin Specific Peptidase 4, Proto-oncogene (USP4)	Systemic lupus erythematosus and Sjogren's syndrome [146]
VlsE 171 – 186	I-x-(ADE)-I-x-(ADE)-A-A		
VlsE 228 – 242	(GP)-(ADE)-x-(LIV)-L-x-A-(LAVI)-(ISVT)-(SVET)-x-x-(DTSVE)		
VlsE 249 – 267	N-P-I-(SAMQE)-x-A-(TLI)	Carboxypeptidase E (CPE)	Acquired metabolic disease [147]
VlsE 338 – 357 Part 1	G-(ED)-(SGA)-x-K-A-A-x-K		
VlsE 338 – 357 Part 2	K-E-x-P-(ASP)-x-L-N		

Competing interests

The authors declare competing financial interests: Frank Breitling is shareholder of PEPperPRINT GmbH. Alexander Nesterov-Mueller, Felix F. Loeffler, and Frank Breitling are named on pending patent applications relating to molecule array synthesis (application no. PCT/EP2013/001141 and PCT/EP2014/001046).

3.3 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,*]

[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

[1] Novel address: Institut de Radioprotection et de Sureté Nucléaire (IRSN), PRP-ENV/SERIS/LECO, Cadarache, Saint-Paul-lez-Durance 13115, France

*Corresponding authors:
F.F. Loeffler: felix.löffler@kit.edu
Frank Breitling: frank.breitling@kit.edu
Alexander Nesterov-Müller: alexander.nesterov-mueller@kit.edu

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 [148]. 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 [8].

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 [1].

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 [149]. 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 are targeted [44, 150]. 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. [56] 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 [57, 58]. 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 [151]. By focusing antigenic attention upon irrelevant or highly variable epitopes, also referred to as deceptive imprinting [152], 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 [153, 154, 155]. In a similar approach, Wang et al. used phage displayed peptide libraries [156, 49, 50] to identify a pattern of targeted peptides that indicated a growing prostate cancer [53]. 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 [54]. Screening random peptide libraries with mouse or human serum antibodies also allowed to distinguish healthy mice from mice that were infected with helminth parasites [157] or to pinpoint rheumatoid arthritis specific antigens [158]. Moreover, Reineke et al. used the SPOT method from Ronald Frank [64] to synthesize arrays with some 5500 random peptides to identify linear peptides,

which served as surrogate binders ("mimotopes") for known conformational antibody epitopes [73].
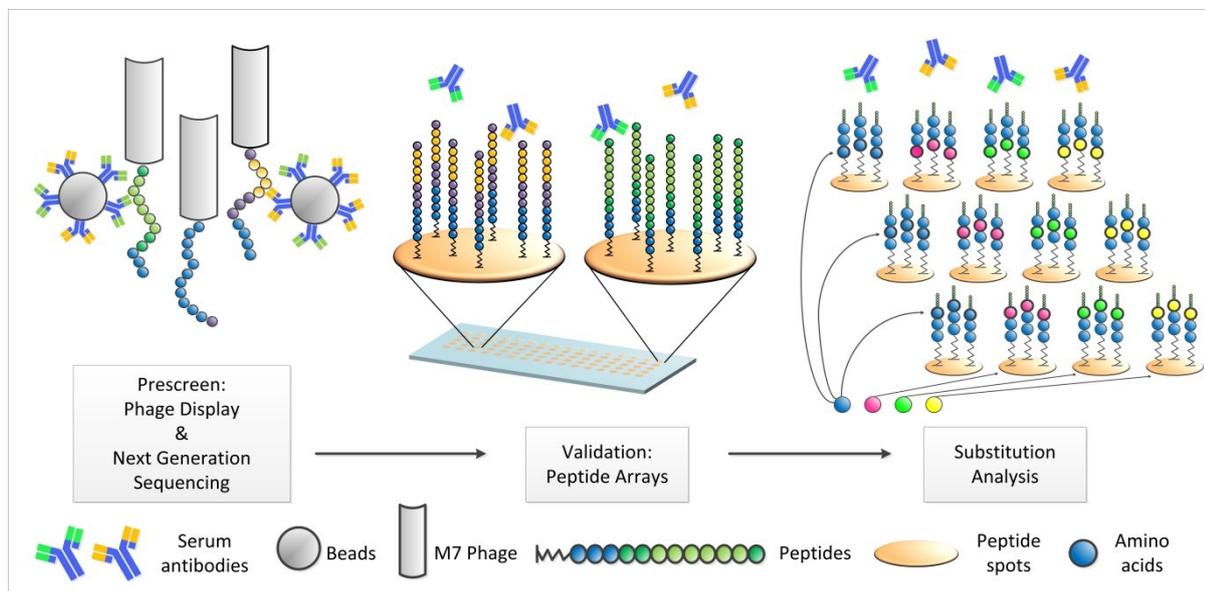
Figure 1: Workflow. In an initial pre-screen, up to 10^9 random peptides displayed on phage were screened for their binding to serum antibodies, immobilized on beads. Next, the identified epitope peptides were validated with solid material-based peptide microarray technology. Finally, the validated epitopes were fine mapped by comprehensive substitution analysis. The resulting "epitope fingerprints" enable the identification of those proteins that match the antibody specificity, and, eventually, the correlation to disease causing agents.

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 [159] with our novel solid-material-based synthesis method for high-density peptide arrays, enabling densities of over 10,000 peptide spots per cm^2 [88, 67, 89, 91, 90]. For the latter, either a laser printer [67] or a laser scanning system [90] 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 Figure 1). It reveals the original epitopes of a patient's antibodies or the mimotopes, and can also uncover potential cross-reactivity [73]. 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 pre-screening 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.
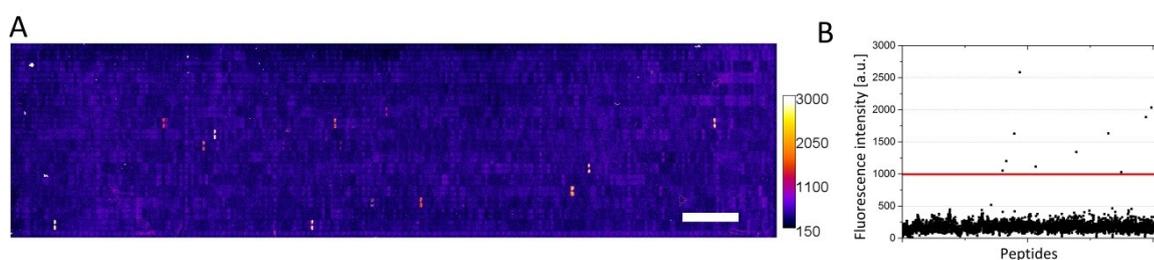
Figure 2: Example peptide array staining and intensity distribution plot. (A) Fluorescence scan image of a validation peptide array with 4128 peptide spots duplicates, which were identified by phage display on human serum. The array was stained with the same serum DL680 conjugated secondary antibodies. Bright double spots correspond to peptides that were strongly bound by antibodies. Contrast and brightness were optimized, scale bar 0.5 cm. (B) Intensity distribution plot of the peptide array with the threshold (red line). The mean fluorescence intensity of the spot duplicates was plotted over the peptides. Peptides were sorted according to the row and column on the slide. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

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 in the one letter abbreviation. Thus, the impact of individual amino acids on the binding of the antibody can be assessed and the crucial amino acids identified. In several cases, minor amino acid variations had no impact on the binding activity. Peptide variants that showed a drop in intensity of > 50% in comparison to the original binding intensity were not considered as

binders. If 5 or more substituted amino acids were found to allow for antibody binding, the position was considered as non-essential.
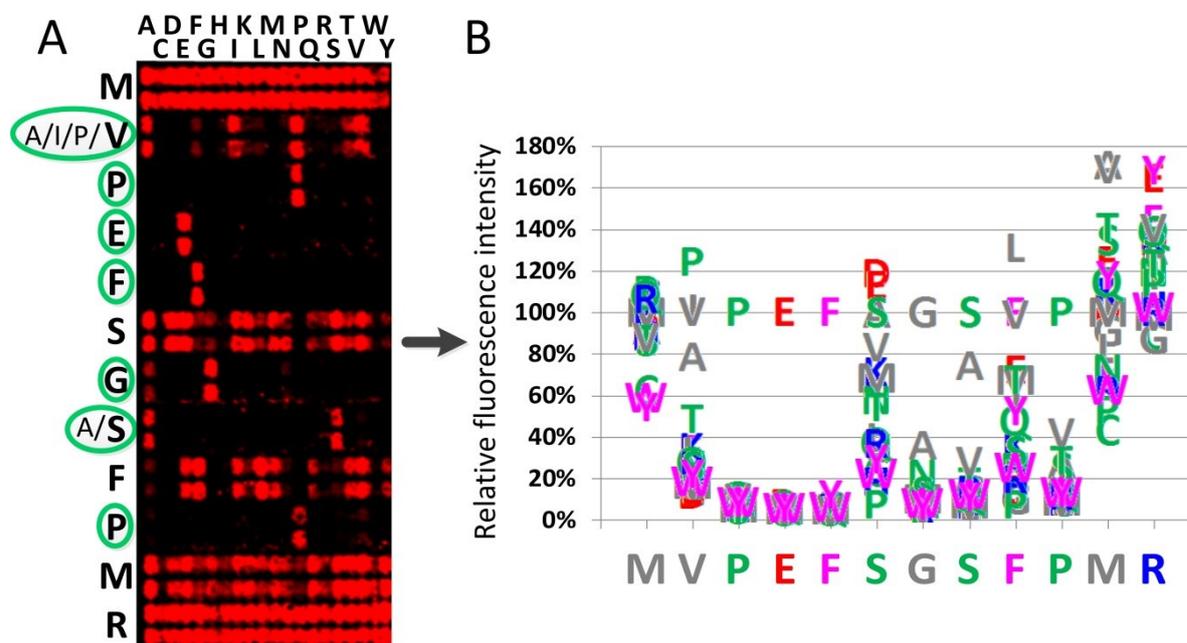
Figure 3: Substitution analysis with each amino acid position substituted by all 19 other amino acids. (A) Example fluorescence scan image of a substitution analysis of a 12mer peptide, consisting of 12 spot duplicate rows and 20 amino acid columns. From top to bottom left, the original sequence is shown in bold next to the scan image. Amino acids that were essential for the binding of the antibody are highlighted in green. Contrast and brightness were adjusted. (B) Relative fluorescence intensities in reference to the original amino acid in the substituted peptide sequence. Each peptide variant is represented by the letter of the amino acid that was synthesized at the respective position. The original sequence is depicted on the x-axis. The intensity of the peptide with the original amino acid was set to 100%. Intensities of peptide variants were correlated to the intensity of the original peptide in each row respectively.

2.3 Identified peptide binders

All 73 peptides that showed distinct signals in the substitution analysis were re-validated on another microarray. The obtained intensity values were transformed with the inverse hyperbolic sine (arsinh) [160].

Table1: Identified peptide binders. Essential amino acids at each position of the peptide are depicted and relative binding strength is indicated by the transformed fluorescence intensities. Amino acids of the original peptides are depicted in red. Peptides with similar binding motifs are grouped and separated; the last group is heterogeneous with no similarities.

Original peptide sequence	Amino acid position in epitope												Transformed original intensity
	1	2	3	4	5	6	7	8	9	10	11	12	
NPVEXXX motif													
NPVEDYLDYSVI	N	P	V,I	E	X	X	X	X	X	X	X	X	8.371
ESYYMNPVEMFV	X	X	X	X	X	N	P	V	E	X	F,Y	I,V	7.312
LPTESNPVEDWI	X	X	X	X	S,T,V,R	N	P	V	E	E,D	!P	I,V	7.091
NPIESYIASVFS	N,D	P	X	E	X	Y	X	X	X	X	X	X	6.774
NPVELLLKTSSD	A,N	P	X	V,A	E	X	X	X	X	X	X	X	6.366
NPVETQIVLSLV	N	P,N,Y	V,P	E,Y	!P	!P	!P	!P	X	X	X	X	5.937
NPVEKWISMRTM	N	P,Q,Y	V	E	!P	!P	I	X	X	X	X	X	5.896
NPVELLLLMGIS	N	P	V	E	X	X	I,V,L,T	X	X	X	X	X	5.733
NGIERLLEEPVS	X	X	I	E	!P	L	I,L,V	X	X	X	X	X	5.729
NPVENWIDPKSI	N	N,P,A	V,A,P	E,V,A	!P	!P	X	D,I,E	X	X	X	X	5.614
VNPVEYYLDTMR	X	N	T,P,S	V	E	!P	X	L	E,D	X	X	X	4.912
NPVEALLSKFHM	N	P	D,E,V	E	E,D,A,F	X	X	X	X	X	X	X	3.567
NPVESYLANYTS	N	T,N,P	X	E,V	X	X	X	X	X	X	X	X	2.521
VTNPYESLVQEK	X	X	N,D,S,G	P,Q	Y,I	X	X	L,F	X	X	E	K	2.285
XPEFXGSXX motif													
SPPEFSGSTVGL	X	V,I,P	P	E	F	X	G	S	X	P,V	X	X	9.334
MVPEFSGSFPMR	X	P,I,V,A	P	E	F	X	G	S,A	X	P	X	X	9.173
RMPPEFMGSLPQ	X	X	X	P	E	F	X	G	S,T	X	P	X	9.107
QLFVPEFAGSSP	X	X	X	V,P,I	P	E	F	X	G	S	X	X	8.373
LVPEFTGSTPFR	X	I,P,V	P	E	F	X	G	S	X	P,V,A	X	X	7.461
SHGAIEFDGAFP	X	X	X	X	X	E	F	D	G	A,S	F	X	7.05
VPEFAGHVPSTA	V,I,P	P	E	F	X	G	X	X	A,V,P,I	X	X	X	6.803
VPEMNGSLIARK	V,P	P	P,E,Q	F,Y,L,M	X	X	S	S,L,T	X	X	X	X	5.954
KXXFPQXT motif													
HPIVKTYFAQTT	X	X	X	X	K	X	Y,F	F	X	Q	T,K,M,V	T,V,I	6.234
APAKTYFGQTTD	X	X	X	K	E,T,V	F,Y	F	X	Q	T,F	T	X	6.223
QMKAWFPQTTYD	X	X	K	!P	X	F	P	Q	X	T,V	X	X	3.911
ATKVMFPQRIYV	X	X	K,E	X	X	F	P	Q	X	T,I,V	X	X	3.249
QPAKTYFNQVTL	X	X	X	K	T,E	F,Y	F,Y	X	Q,Y	X	T,I	X	3.21
QAAAKTMFPQNT	X	X	X	X	K	T,E,Q	X	F	P	Q	!P	T	2.336
LXAXETX motif													
SPSIDAFETSIF	X	P,Q	X	L,M,I,W	D	A	X	E,S	T	S	X	X	9.261
SWVLTATETGSS	X	X	X	L,F	T	A	X	E	T	G,S	X	X	8.425
EMRFPSLSASDT	X	X	X	X	P	S,G,A,P	L	T,S,N	A	!P	E,D	T	7.349
QAPTLDAQETAL	X	X	P	X	L	T,S,D,N	A	X	S,E,T	T	S,G,A,N	X	7.248
TALDAVSTGFSW	X	P,A,E,S	L	D	A	X	E,T,S	T	G,S	X	X	X	6.61
GAIGSLTADSTS	X	X	I,V	P,G	P,S,A	L	S,T	A	X	X	T	N,S,R	6.573
SLGSLSAYETGR	X	X	X	X	L,F	X	A	X	E,D	T	X	X	6.53
TLYRPPLTSAET	X	X	X	X	X	L	T	A,S	A	E	X	X	5.963
LIADLNAESTSR	X	X	X	D	F,I,L,Y	N,S,D,T	A	X	S,T,D,E	T	S,T,N,G	X	5.934
SLGAAGARVTTV	X	X	X	X	X	X	X	X	V,W	X	T,Y,W	V,W,Y	5.808
ASGPLHAGATGL	X	X	X	P,A	L	X	A	X	X	T	N,S,G	X	5.778
GHKPVLTALSTA	X	X	X	P	A,V,P	L	T	A	X	X	T	S,G,A	3.738
VMPAIHAGVTGA	X	X	P	A	L,I	X	A	G,A	A,V,E	T	S,G	X	1.942
Heterogeneous motifs													
VHPTLTVTSKEI	I,V	D,E,H,N	I,K,P,V	T	L,V	K,M,T,L	V	X	X	K	X	X	7.237
ELDVSLRVMPKV	X	X	D	X	T,S	V,T,L,I	X	V	X	X	K	X	7.028
SKLAMEIMSGPV	X	X	X	X	!P	!P	I	!P	S	G	P	X	6.995
SNAVTSSKSPRM	D,T,S,N	D,N	A	I,V,W	T	D,S,E	X	K	X	X	X	X	6.935
QADATVLTKPKT	X	X	D	X	T	V,L	X	V,T	K	P,A,S	K	X	6.798
KAVSDASRGTVF	K	X	V,I,T,L	S,T	X	X	S,T	R	X	X	X	X	6.716
DYPKIANFQEYA	X	Y	P	X	I	X	N	F,Y	X	X	X	X	6.683
GGQVRSIHSGPT	X	X	X	X	X	X	I	!P	S	G	P	T	6.589
QHWPTNVDSVTV	X	X	X	X	X	X	V	D,E	X	V	X	X	6.548
ETKSDDMLLSNV	X	X	K	X	D	D	X	L,I	X	T,S	X	X	6.407
MTVDRTVRVASK	X	X	X	D	X	T,S	L,V,T	X	V	X	X	K	6.366
DLLIRDAITDTK	X	X	X	X	X	N,D	A	I,V,T	T	D,S	X	K	6.359
AYHVDTVSDAGW	X	X	X	V	X	T,S,E	V,I	X	X	X	X	X	6.324
VDTINLPQNTIQ	V,I	D,E	X	V,I,T	X	X	X	X	X	X	X	X	6.095
VMSVNASTTAAN	V	X	X	V	N,D,E	X	X	X	X	X	X	X	6.067
TLHAPMTIRSGP	X	X	X	X	X	X	T	C,I	X	S,T	G	P	5.692
TNLHRVMTVVNM	X	X	X	X	X	V	X	T	V,R,I,S	X	X	X	5.372
LRPNAVQTDTLA	X	X	P	D,N	A	V	Q	X	X	X	X	X	5.319
QMRQLTEYGSEK	X	X	X	X	X	X	X	F,Y	X	X	E	K	4.917
VLSSTAIKVDSV	X	X	X	X	X	X	X	X	V,L	D,E	T,S	V	4.832
VHTVHDVFTAFG	V	D,H,E,N	T,D,S	X	X	X	X	X	X	X	X	X	4.538
AKIRMFLDTDYK	X	X	X	X	X	F	L	D	T	D	!P	K	4.315
ASWGPIAIDRVN	X	X	X	X	X	X	X	V,I	D	!P	X	N,F	4.054
SQQYALTNSTTN	X	X	X	X	X	X	X	N,A	X	T	T,N	N	2.89
QPQTKSFYPQYV	X	X	X	X	K	T,S,E	F	F,Y	P	Q	X	T,I,V	2.878
NVVDRVNRTGVV	X	X	A,V	D	!P	V	A,N	X	X	X	X	X	2.693
WSALPERTSLPV	X	X	X	X	T,P	N,E	X	T,W,Y	S,W,Y	X	P,W,Y	X	2.315
RPAIVDQVSSSP	X	X	X	X	V	D	X	V,T	S,N	X	X	X	2.217
QWNWRVRSVANV	X	X	X	X	Q,R,N,H	V	X	T,S	V	X	X	X	2.097
AQLHPTTLVKHK	X	X	X	D,E,H,N	X	T	V,T,L	X	V	K,Q	X	K	0.925
TSYRPPLNVCQD	X	X	X	X	P,Q	X	L	N,C	X	X	X	X	0.884
QSHSLFYPHPYG	X	X	X	X	X	X	F,Y	P,D,E,A	H	P	X	G	0.545

Table 1 shows a list of the original amino acid sequences of all successfully substituted peptides with the essential amino acid positions. Peptides which exhibited similar antibody fingerprints are grouped, separated by breaks. The last group is heterogeneous without significant similarities in binding motifs. The list is ordered according to the staining intensities on the control slide. For a detailed view on each peptide binder (analogous to Fig. 3B), we refer to the Supplementary materials (Figs. S3–S7).

2.4 Statistical analysis of epitope properties

For the dataset shown in Table 1, we compared the length of the epitopes (i.e. the number of amino acid positions from the first to the last crucial position) with the number of essential amino acids per epitope (Fig. 4A), and we also analyzed the relative frequency of the number of possible amino acids at the crucial positions according to the length of the epitope (Fig. 4B). Epitopes with a length of 8 amino acids and 5, 6, or 7 essential positions had the highest frequency (19 of 73 peptides). Epitopes of 9 amino acids with 7 essential positions, 8 amino acids with 7 essential positions, and 4 amino acids with 3 essential positions were identified 7 times each.
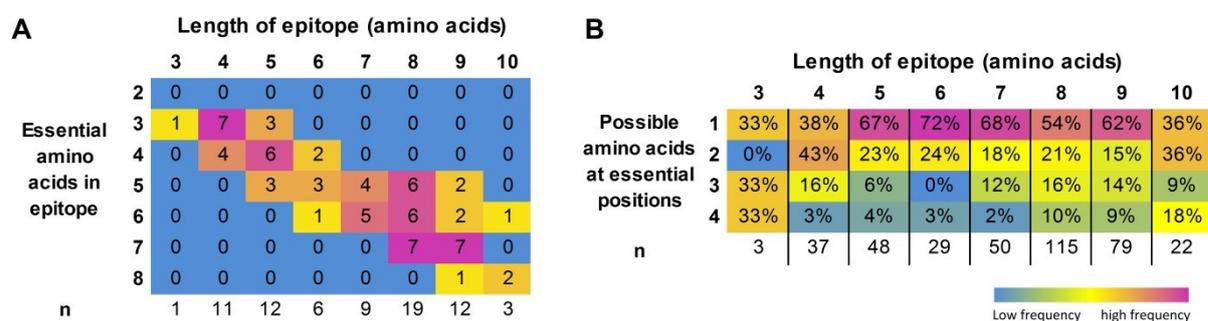
Figure 4: Statistical epitope properties. (A) Heat map of the number of peptides with a specific epitope length and the number of essential amino acids in the respective epitope. (B) Heat map of the relative frequency of possible amino acids at essential positions of the epitopes.

Fig. 4B shows that the relative frequency of possible amino acids at essential positions in the epitopes with lengths between 5 and 9 is low. Thus, epitopes with the length of 5–9 amino acids seem to have more conserved binding fingerprints, with mostly one possible amino acid at one position.

2.5 Meme analysis

A meme analysis [161] compares different sequences in regard to correlating amino acids. Thereby, it is possible to pinpoint amino acids that potentially play a role in the antibody binding. Yet, the meme analysis lacks the ability to identify essential amino acids if there is only one sequence of one type. In Fig. 5, meme analyses for selected peptides are shown. Motif A "NPVEXXX" was found in 14, motif B "XPEFXGSXX" in 8, motif C "KXXFPQXT" in 6, and motif D "LXAXETX" in 13 sequences. This implies that 41 of the 73 binders might be caused by only 4 sequences. However, motif D shows a rather low E-value, which can be interpreted as a coincidental finding. Thus, we assume that we found 48 unique fingerprints or antibody binders respectively.
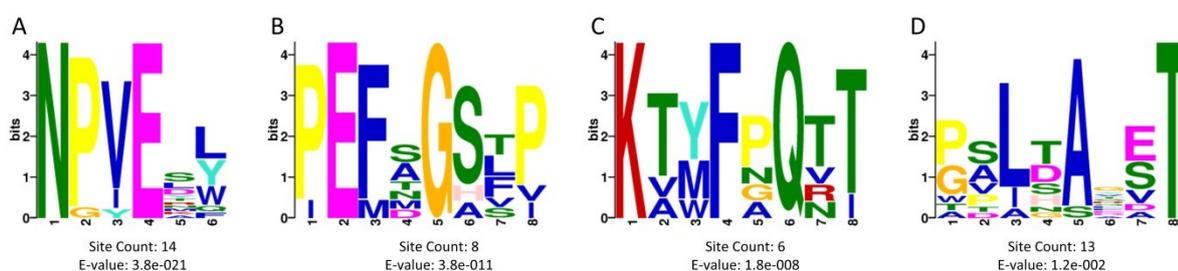
Figure 5: Meme analysis of substituted peptide binders. Only motifs with a prevalence of > 3 and an E-value < 0.1 are shown, which results in 4 sequences. The sequences cover 41 of the 73 binders.

2.6 Epitope binding variance over time

Next, we evaluated the variance of the antibody response of a serum donor at different points in time. For the previous screens and analyses, the serum from the time point t = 24 months was used. Now, identical laser-generated high diversity peptide arrays, containing the above identified epitopes, were incubated with three different serum samples collected at 3 different time points, each about 24 months apart. Each serum was incubated on 9 arrays containing the previously identified epitopes as spot duplicates, resulting in 18 data points per epitope and serum. The intensity values were transformed with the arsinh function [160] and the average transformed intensity values were calculated for each serum. Fig. 6 and Fig. 7 show the average transformed intensity values of two groups: in the first group of peptides (Fig. 6) the average intensities at different time points do not differ significantly, in the second group (Fig. 7) at least two of the intensities of different time points show a significant difference.

Figure 6: Epitope profile of an individual donor at three different time points – group 1. Results are shown as transformed intensities for each peptide. Serum staining intensities at different time points do not differ significantly (Student's *t*-test two-tailed), 18 data points were acquired for each peptide and serum, mean values and standard deviations are shown.

2.7 Validation of biological significance

To validate that the fingerprint analysis can be performed with other patient samples, and also reveal different peptide binders, we have analyzed a second patient serum in a similar approach (see Supplementary materials Figs. S1, S2, and Table S1). We generated a peptide array with random amino acid sequences, incubated this array with the donor sample from the previous experiments, and stained with secondary antibodies. Afterwards, we incubated this array with a sample from a different patient, and also stained with secondary antibodies (Fig. S1). We selected those peptide binders, which showed a high reactivity only with the serum of the second patient. Finally, these peptides were screened in a substitution analysis, resulting in patient specific antibody fingerprints (Fig. S2, Table S1).
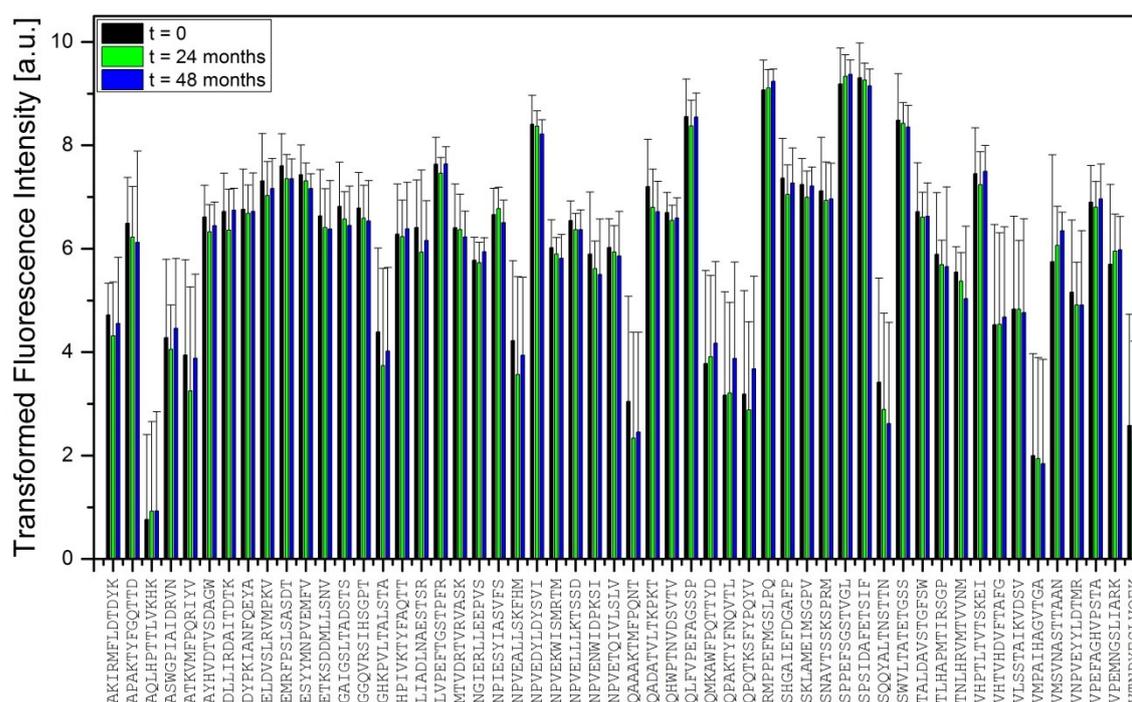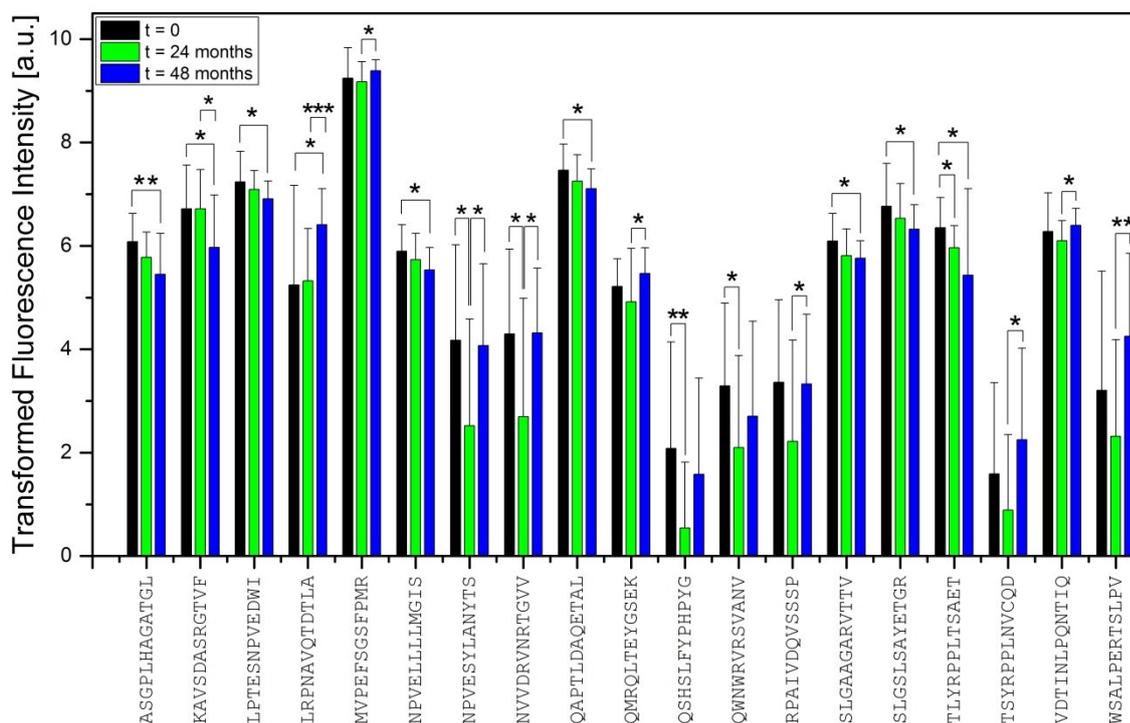
Figure 7: Epitope profile of an individual donor at three different time points – group 2. Results are shown as transformed intensities for each peptide. 18 data points were acquired for each peptide and serum. For at least two time points per peptide the differences in intensities were significant (Student's t-test two-tailed) with different p-values: ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$. 18 data points were acquired for each peptide and serum, mean values and standard deviations are shown.

2.8 Bioinformatic analysis of fingerprints

The four most abundant fingerprints, shown in Fig. 5, were queried in the "UniProtKb/Swiss-Prot" protein database (release 2016_11 of 30-Nov-16: 553231 entries) with the "ScanProsite tool" [106]. The query included splice variants (Swiss-Prot) and excluded fragments. All results of the query are listed in Supplementary tables S2–S5. However, most proteins are very unlikely to be the causative agent of the antibody species since they are either part of exotic organisms or would probably cause a noticeable autoimmune reaction. Therefore, the results that are most likely to be the causative antigen are listed in Table 2.

Table 2: Results of the query of the "UniProtKb/Swiss-Prot" protein database with the "ScanProsite tool" for the identified most prominent motifs. The amino acid sequences that were used for the queries are shown in column 2. The resulting number of hits and the respective Supplementary table, including all results, are listed in columns 3 and 4. Furthermore, the table shows the most likely causative agent and antigen, the amino acids of the epitope, and the position on the antigen.

Motif	Sequence for query	Hits	Table	Potential causative agent	Potential antigen	Aa residues	Aa position
NPVEXXX	A,N,D-N,P-V-E-E,D,A,F-F,Y-I,V,L	32	S2	Human cytomegalovirus	Protein HHLF1	ANVEDYL	575 - 581
					Protein TRS1	ANVEDYL	576 - 582
XPEFXGSXX	V,I,P-P-E-F-x-G-A,S-x-V,P	14	S3	Staphylococcus aureus	Extracellular matrix protein-binding protein emp.	VPEFkGSlP	328 - 336
KXXFPQXT	K,E-x-F,Y-F,Y-P-Q,Y-T,K,M,V,F-T,V,I	31	S4	-	-	-	-
LXAXETX	L-D,T-A-x-E-T-G,S	102	S5	Coxsackievirus	Capsid protein VP1	LTAvETG	620-626
				Echovirus	Capsid protein VP1	LTAvETG	600-606
				Human rhinovirus	Capsid protein VP1	LTAnETG	601-607
				Poliovirus	Capsid protein VP1	LTAvETG	623-629

3. Discussion

Antibodies protect us from life-threatening infections, but may also do harm in autoimmune diseases or, in an unfortunate genetic setting, lead to transplant rejection. Moreover, some antibody species have homeostatic tasks, for example band 3 specific IgG antibodies that mark old erythrocytes for phagocytosis [162] or antibodies that mark those cells for destruction that endured an oxygen deficiency in a stroke [163] or in a growing tumor. In any case it would be interesting to know the exact epitopes of all the antibody species that patrol in the blood of a patient, especially if the patient suffers from a clinical picture caused by an unknown antigen. We present a pipeline of methods that allows for a comprehensive read out of the human antibody repertoire and their specificities without a priori knowledge. We identified 73 antibody fingerprints from one patient serum by sequentially employing a pre-screen with a phage displayed peptide library (38,533 binders), validating these potential

peptide binders in the array format, and finally identifying those amino acids that are crucial for antibody binding (97 binders in substitution analysis as 23,280 different peptides, 73 validated and 48 unique fingerprints). The serum antibody specificities of the same donor at three different points in time over 4 years revealed that all found peptides were similarly targeted by serum antibodies at all time points. This might have two reasons: (1) Our method only allows for the identification of very stable antibody species, or, (2) our serum is mainly composed of a limited number of stable antibody species, whereas in times of infection, the repertoire is rapidly expanded.

To identify the specificity of every antibody species in a patient's serum, we hypothesize that each type of antibody has at least one linear peptide binding epitope, whether or not it belongs to the original antigen or is a mimotope which binds due to a cross-reaction [73]. We also hypothesize that only a limited number of important epitopes per pathogen are shared by most patients. Previous studies could demonstrate that most of these epitopes consist of short motifs with a limited number of key residues [93, 164, 165]. Thus, if the diversity of a peptide library is large enough, it should be possible to characterize each antibody by at least one binding event. Recently, we have introduced a new technology which enables us to synthesize peptide arrays with a very high density of > 17,000 spots per cm^2 [90] and thus makes studies with very high numbers of peptides affordable. Since this new flexible production method also allows for the synthesis of modified amino acids, together with recent developments in the synthesis of cyclic peptides, the search scope will even be expanded further in the future.

We queried a public protein database for the four most abundant motifs that were found in several peptides in the substitution analyses, which were additionally identified in a MEME analysis. The resulting hits pointed to potential causing antigens. Most antigens are very unlikely to be the cause: They would either cause autoimmune reactions or originate from exotic organisms, which usually do not come into contact with the immune system. However, for three motifs, a potential causing antigen could be identified. Sophisticated bioinformatics should improve the interpretation of the database results. The capsid proteins VP1 of four viruses share the LXAXETX motif with the amino acids L-T-A-n/v-E-T-G. The epitope LTAVETG, part of the poliovirus, can be also found in the human epitope database (iedb.org). From personal communication with the company PEPperPRINT, we learned that the peptide KEVPA**LTAVETG**AT shows interaction with > 90% of all European individuals, most likely due to the common polio vaccination. Furthermore, Cello et al. showed that all their investigated serum samples reacted preferably with the peptide <u>PALTA</u><u>VETGATNPL</u> (underlined amino acids are conserved between different enterovirus types) in a peptide

ELISA [166]. Therefore, we are capable of identifying an antibody species that is very likely to be present in the investigated serum. For one of the four queried motifs, no obvious antigen could be detected.

Conclusion

We developed an approach which theoretically allows us to read out the complete humoral immune response of a patient without any a priori knowledge. Thereby, we can determine the specificity of an antibody to the single amino acid level, which we call the antibody fingerprint. We discovered 73 of these antibody fingerprints in a patient's serum. The bioinformatical analysis of one of the most abundant fingerprints led us to the antigenic capsid protein VP1 of the poliovirus. Due to the ubiquitous vaccination, it is very likely for the patient to have an antibody of this specificity.

A study with a large group of healthy versus diseased patients and appropriate bioinformatics should allow for the identification of specific fingerprints that are characteristic for many different diseases or antigens and could help to elucidate enigmatic diseases. These studies might also answer the question if our approach can identify antibody species with low affinity and/or low concentration. Considering the diversity of still enigmatic diseases human mankind is facing, our approach could contribute to a better understanding of their symptoms or etiology.

4. Material and methods

4.1 Study design

In this study, we investigated the potential of a novel screening pipeline approach to comprehensively read out the antibody specificities in the serum of a patient: We combine phage display pre-screening with a subsequent two-step peptide microarray analysis. The serum of a donor (approval by the ethics committee was acquired and informed consent was obtained) is analyzed in a phage display panning together with DNA sequencing. The resulting binders are then synthesized on peptide arrays and incubated with the original serum for validation of binding. Subsequently, these validated epitopes are synthesized again onto a peptide array, but this time in a substitution analysis mapping, by exchanging every amino acid at every position, to fine map the antibody binding. The validated peptide set was screened with two additional sera from the same donor from two other points in time, spanning a total of 48 months. The study was approved by the state chamber of physicians of

Baden-Wuerttemberg (reference number: F-2011-044 and F-2011-044#A1). The serum of a second individual (ethical clearance was obtained from the institutional review board (IRB) of the Medical Faculty of Heidelberg University and informed consent was obtained) was investigated in a similar approach to show biological significance.

4.2 Phage display

Panning round 1: Supernatant of 20 μL 30 mg/mL ProteinA Dynabeads (0.6 mg beads bind ~ 5 μg hIgG) was removed in a magnetic base and beads were washed twice with 500 μL Phosphate buffered saline (PBS). 5 μL of serum was diluted in 200 μL PBS, added to the beads and incubated with end over end shaking for 30 min to allow the antibodies to bind to the protein A. 500 μL of blocking buffer (10 mg/mL BSA (Fraction V pH 7) in $1 \times$ PBS with 0.1% Tween20) was added with 5 μL of serum to the beads and again incubated with end over end shaking for 30 min at room temperature. Afterwards beads were washed 3 times with 500 μL PBS-T (0.1% Tween20). 1×10^{11} phage of a PhD12 library (NEB) (10 μL 1×10^{13} pfu/mL, $100 \times$ overrepresentation) in 200 μL PBS-T (0.1% Tween20) were added to the beads and incubated for 1 h at room temperature under shaking. Supernatant was removed in a magnetic base and beads were washed 10 times with 500 μL PBS-T (0.1% Tween20). Elution of phage was done with 500 μL 0.2 M GlycinHCl at pH 2.2 for 10 min at room temperature. Eluate was collected and neutralized with 75 μL of 1 M TrisHCl, pH 9.1. For amplification of eluted phage, 30 mL E. coli (ER2738 OD600 ~ 0.05) were infected with these and grown for 4.5 h at 37 °C and 250 rpm. Bacteria were removed by centrifugation for 20 min at 4000 rpm and 4 °C, supernatant was added to 1/5 volume of PEG/NaCl (20% (w/v) PEG 6000, 2.5 M NaCl) and incubated over night at 4 °C. Phage-PEG/NaCl suspension was centrifuged at 4000 rpm and 4 °C for 1 h. Supernatant was discarded and phage pellet was suspended in 500 μL PBS, centrifuged again for 2 min at 14,000 rpm and stored at 4 °C. For titration, 10 μL of two dilutions (10^{-1} and 10^{-3}) of the eluted phage, resulting in 10^{-3} and 10^{-5}, were plated in top agar. 200 μL of E. coli (ER2738 OD600 ~ 0.5) were added to 10 μL of the dilution and incubated for 2 min. Culture was added to 3 mL of top agar (~ 45–50 °C, 0.7% Agar) and poured on pre-heated LB/IPTG/Xgal plates (1.5% Agar, 50 mg/L IPTG, 40 mg/L Xgal). After cooling of agar the plates were incubated at 37 °C over night. Titration of the eluted phage of the first panning round resulted in 2.6×10^6 pfu/mL.

Panning round 2: ProteinA Dynabeads were washed blocked and incubated with serum as in panning round 1. For pre-incubation of amplified phage 10 μL of beads were washed twice with PBS-T, blocked with blocking buffer for 20 min, washed again three times, added to

$\sim 1 \times 10^{11}$ amplified phage of the last panning round in 200 µL PBS-T (0.5% Tween20) and incubated for 20 min at room temperature under shaking. Pre-incubated phage were added to the antibody beads and incubated for 1 h at room temperature under shaking. Washing, elution and neutralization was again done as in panning round 1. Amplification of eluted phage was performed as done in panning round 1. For titration, 10 µL of two dilutions (10^{-1} and 10^{-3}) of the eluted phage and 10 µL of the 10^{-9} dilution (resulting in 10^{-11}) of amplified phage were plated in top agar. Titration was then performed as in panning round 1. Titration of the eluted phage of the second panning round resulted in 4.64×10^{7} pfu/mL, titration of the amplified phage of the first panning round resulted in 2.36×10^{13} pfu/mL.

Panning round 3: Panning, amplification and titration were performed as in panning round 2. Titration of the eluted phage of the third panning round resulted in 1.112×10^{9} pfu/mL, titration of the amplified phage of the second panning round resulted in 3.4×10^{13} pfu/mL.

4.3 Next generation sequencing

Preparation of phage DNA for NGS: 10^{12}–10^{13} phage particles were mixed with 1/5 volume of PEG/NaCl (20% (w/v) PEG 6000, 2.5 M NaCl) and incubated on ice for two hours. Phage-PEG/NaCl suspension was centrifuged at 14,000 rpm and 4 °C for 15 min. Supernatant was discarded and pellet was thoroughly dissolved in 63 µL NaI solution (10 mM Tris-HCl pH 8, 1 mM EDTA, 4 M NaI). 156 µL ethanol (100%) was added and the solution was incubated on ice for two hours to precipitate DNA. Solution was centrifuged at 14,000 rpm and 4 °C for 15 min and DNA pellet was washed in 200 µL 70% ethanol to remove residual salt. Solution was centrifuged again at 14,000 rpm and 4 °C for 15 min. Supernatant was discarded and DNA pellet was air dried for 15–20 min at room temperature. The precipitated DNA was further purified using phenol–chloroform extraction. The DNA pellet was resuspended in 300 µL water and transferred to a 2 mL PhaseLock tube. An equivalent amount of phenol–chloroform (1:1 v/v) was added, shook thoroughly and centrifuged at 14,000 rpm for 1 min. An additional equivalent amount of phenol–chloroform (1:1 v/v) was added, shook and centrifuged again at 14,000 rpm for 1 min. The aqueous layer (~ 300 µL) was transferred into a new microfuge tube and 30 µL NaAc solution (3 M, 30 µL) and 600 µL 100% ethanol was added. The solution was incubated at -20 °C for 2 h to precipitate DNA. Solution was centrifuged at 14,000 rpm and 4 °C for 15 min and DNA pellet was washed in 400 µL 70% ethanol to remove residual salt. Solution was centrifuged again at 14,000 rpm and 4 °C for 15 min. Supernatant was discarded and DNA pellet was air dried for 15–20 min at room temperature. DNA was resuspended in 30 µL RNAse free water and stored at -20 °C.

Peptide DNA amplification for NGS: DNA isolated from the amplified phage was subjected to PCR amplification with primers flanking the variable region (For-primer: TATTCTCACTCT, Rev.-primer: CGAACCTCCACC). For each panning round, unique nucleotide-barcodes were added to the 5` end of the primer. The PCR mixture contained 1 × Phusion HF buffer, 1 U Hot Start Phusion Polymerase, 200 μM dNTPs (each), 500 nM For- and Rev.-primer and 100 ng phage DNA template. For amplification the following temperature regimen was employed: 30 s incubation at 98 °C, followed by 35 cycles of PCR amplification at 98 °C (10 s), 54 °C (10 s), and 72 °C (30 s) and a final step at 72 °C (5 min). The PCR products were run on a 2% (w/v) agarose gel in TBE buffer. Peptide DNA bands (~ 75 bp) were excited from the gel and DNA was extract from the gel slices using the NucleoSpin Gel and PCR Clean-up Kit (Macherey-Nagel). DNA was eluted in 20 μL Buffer NE (5 mM TrisHCl pH 7.5) and stored at − 20 °C.

NGS: Next generation sequencing libraries were made with the TruSeq ChIP library prep kit (Illumina). Briefly, PCR amplified fragments, 75 bp long, from the second and third panning rounds were pooled to reach a final concentration of 20 ng/μL in a final volume of 60 μL. End repair and ends adenylation were made following manufacturer's instruction (Illumina) and DNA purified with the Minilute kit (Qiagen) following recommendations. Adapters with specific barcodes were ligated, libraries amplified by 10 cycles of PCR and finally purified on Minilute columns followings recommendations. Quality of the libraries was assessed on DNA1000 Chip (Bianalyser 2100, Agilent) and DNA concentration determined by fluorometric method (QBit HS kit, Life Technologies). Up to 3 libraries were sequenced on a single lane at 2 × 100 bp on a HiSeq1500 (Illumina). Demultiplexing and generation of the fastq files were performed with CASAVA v1.7 (Illumina). Bioinformatics processing of the reads were made using an in-house bash script. Briefly, reads (R1) passing the Illumina chastity filters were selected for subsequent analysis. First, reads with Phred score > 25 and perfect match to the generic phage display adapter (CGAACCTCCACC) in forward or reverse orientation were selected using fastx_toolkit v0.0.13 using the options -q 25 -p 70. Barcodes were then trimmed and insert translated into peptide sequence using EMBOSS transeq. Finally, occurrence of distinct peptide sequences was computed in bash.

4.4 Peptide arrays

Validation arrays: The list of peptides obtained from sequencing was first filtered for peptides with the length of exact 12 amino acids. Peptide sequences of shorter length were discarded;

those with longer length were trimmed to the correct length by cutting the protruding amino acids at the C-terminus. Peptides containing a stop (*) were substituted by peptides with a glutamine at the respective position since the "*" was generated by amber (TAG) stop codons, which should have been translated as a glutamine since the employed *E. coli* strain was an amber suppressor line. Randomly appearing "X" was substituted by alanine. The remaining 38,533 peptides were synthesized on 10 peptide arrays for validation. Each array contained a maximum of 4128 different 12mer peptide sequences as spot duplicates. Peptide arrays were obtained from PEPperPRINT GmbH, Germany, produced by combinatorial synthesis with a laser printer.

Substitution analysis arrays: peptides showing a significant higher signal then the background on the validation arrays were analyzed in a substitution analysis. Every amino acid in the sequence of a peptide was substituted by all other amino acids while the rest of the sequence was conserved. Arrays were manufactured by PEPperPRINT GmbH, Germany by combinatorial synthesis with a laser printer [67].

Incubation of peptide arrays with human serum: All incubation and washing steps were performed in PEPperCHIP incubation trays (PEPperPRINT GmbH, Germany), which allow for a minimized sample incubation volume and, in different variants, also the subdivision of the array substrate glass slide ($1'' \times 3''$) into 2 to 16 separate incubation wells for each slide. The assay for the incubation of peptide arrays consists of two steps, each followed by a fluorescence scan based on an immune assay: the pre-incubation for revealing false positive signals by binding of the fluorescently labelled secondary antibody followed by the main incubation with serum and the secondary antibodies. Each step starts with a pre swelling of the array with PBS-T ($1 \times$ PBS, Sigma Aldrich, pH 7.4, 0.05% v/v Tween 20, Sigma Aldrich) for 10 min. Each array was blocked when incubated the first time with blocking buffer (Rockland Blocking Buffer, Rockland) for 30 min and washed afterwards shortly with PBS-T. For the pre incubation step the array was incubated with secondary DL680 conjugated anti-human antibodies (Rockland), diluted to 0.2 μg/mL in PBS-T with 10% v/v blocking buffer for 30 min. In the main step the array was incubated with serum, diluted 1:500 (1:00 for substitution analyses) in PBS-T with 10% v/v blocking buffer over night at 4 °C. After washing shortly in PBS-T it was incubated with secondary antibody as described above. After each incubation step the array was washed shortly with PBS-T and di-ionized water, dried with Argon and scanned using an Odyssey scanner (LI-COR) at 680 nm with a resolution of 21 μm/pixel.

Evaluation of peptide arrays: Fluorescence intensities were evaluated with the software PepSlide Analyzer. The median fluorescence intensity of each peptide spot was calculated and intensities of spot duplicates were averaged. Global background intensities were deduced of the spot intensity to obtain the foreground intensity of each peptide. The intensity values of the peptide spots on the control slide were transformed with the inverse hyperbolic sine (arsinh) [160].

For evaluating the influence of individual amino acids in the substitution analysis, the fluorescence intensity of each peptide with a substituted amino acid was related to the fluorescence intensity of the peptide with the original amino acid in each row.

Competing interests

The authors declare competing financial interests: F.B. is shareholder of PEPperPRINT GmbH. A.N.-M., F.F.L., and F.B. are named on pending patent applications relating to molecule array synthesis (application no. PCT/EP2013/001141 and PCT/EP2014/001046).

3.4   Automated microfluidic system with optical set up for the investigation of peptide-antibody interactions in an array format

Laura K. Weber[1][†], Andrea Fischer[1][†], Tim Schorb[1], Miriam Soehindrijo[1], Tobias C. Förtsch[1], Clemens von Bojničić-Kninski[1], Daniela Althuon[1], Felix F. Loeffler[1], Frank Breitling[1], Jürgen Hubbuch[2], Alexander Nesterov-Müller[1]

[1]Institute of Microstructure Technology, Karlsruhe Institute of Technology, Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen

[2]Biomolecular Separation Engineering, Karlsruhe Institute of Technology, Engler-Bunte-Ring 1, 76131 Karlsruhe

[†]these authors contributed equally

Abstract

Peptide arrays enable high throughput screening of antibody-protein interactions. Thousands of peptides can be screened simultaneously with a minimum of sample volume. Besides the information whether an antibody binds specifically to a peptide, it is of interest to characterize the peptide-antibody interaction. Therefore, we developed a microfluidic system for the automated incubation of peptide arrays with biological samples and for the optical characterization of antibody-peptide interactions in the array format.

1. Introduction

The study of protein binding is the basis of fundamental research regarding cancer related signaling cascades as well as for diagnostics regarding detection of disease-related antibodies [167]. For the interrogation of protein binding events, peptide arrays are a well-established method [66]. They enable multiplexed high-throughput assays and require minimal sample volume. At the Institute of Microstructure Technology at KIT, high-density peptide arrays and their applications are developed and optimized [88, 91]. Peptide arrays can be used to search for linear epitopes in the immune response to diseases or vaccines: displaying the whole amino acid sequence of a toxoid as overlapping peptides, the toxoid specific antibodies in a patient serum can be analyzed and characterized [168] (see fig. 1 a-c). We were especially interested to set up a system that – in addition to mere antibody binding – would also deliver

kinetic data in an array format for many different anti-body-peptide pairs. Therefore, our experimental setup allows incubating 120 different peptides with different biological samples (e.g. sera or cell culture supernatant), while binding events are detected by time resolved fluorescence imaging.
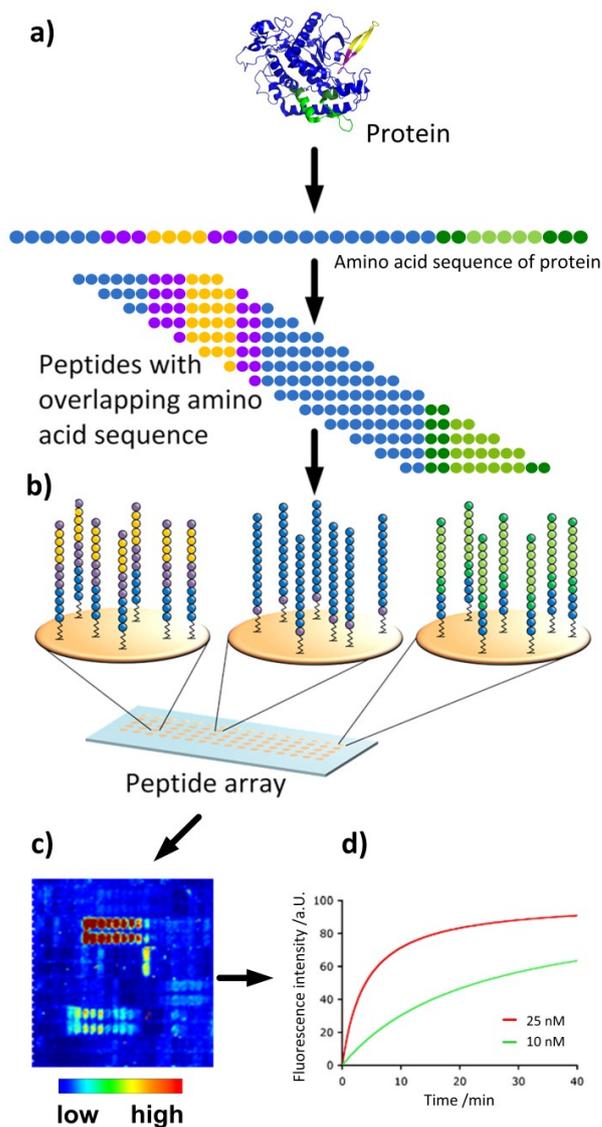
Figure 1: a) The amino acid sequence of a protein is cut in silico into peptides with overlapping sequences. b) These sequences are synthesized onto a peptide array. c) The peptide array is incubated with fluorescently la-belled antibodies. d) The fluorescence signal is detected with a camera over time for different antibody concentrations. Example protein structure from [169] modeled with PyMOL.

1.1 Microfluidic system

All incubation steps are automatically performed in a microfluidic system. A continuous, laminar flow allows the contact and interaction of the sample antibodies with the peptides, which are immobilized in array format.

The core part of the experimental setup is a microfluidic channel of the MicCell system by GeSiM (see fig. 2). The walls and the top of the channel are molded into Polydi-methylsiloxane (PDMS). The actual channel is formed when the peptide array is attached to the PDMS. A cover plate contains threads to connect the tubing. By screwing the cover plate onto the frame, a defined pressure is applied to the PDMS and the peptide array to seal the microfluidic channel. There are two parallel channels molded into the PDMS, each 150 µm in height and 3.5 mm in width. A continuous flow is realized by an automated peristaltic pump (Reglo ICC, Ismatec). All process relevant buffers are connected to the system via a motorized valve.
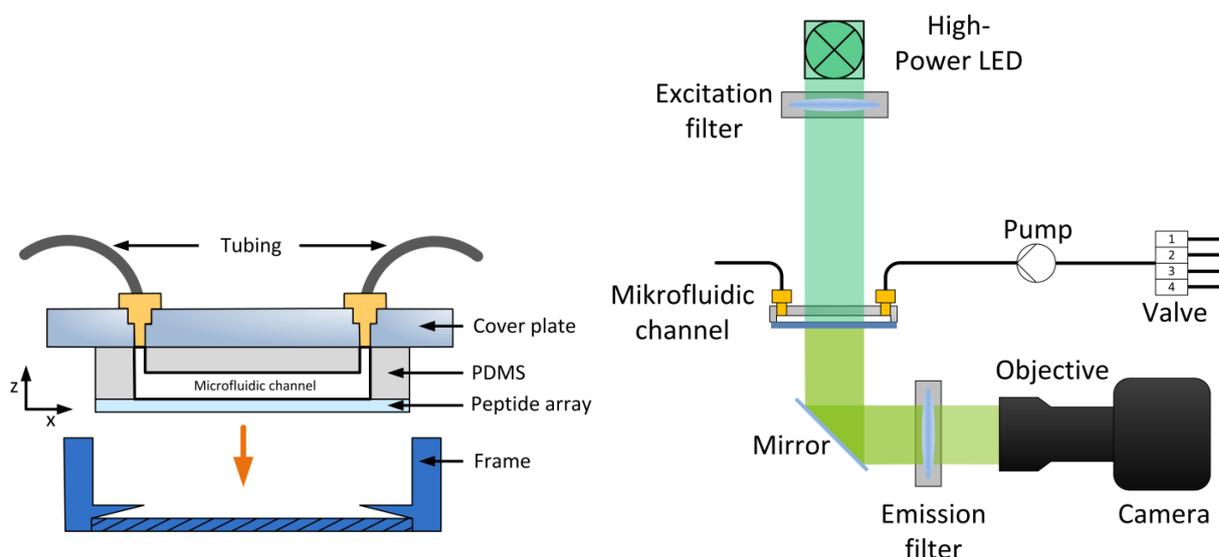
Figure 2: MicCell system by GeSiM (left) and schematic of experimental setup (right).

1.2 Optical setup

For the detection of binding events, we chose fluorescence, due to the high sensitivity. Fluorescence on the peptide array is excited by a high-power LED at 530 nm. A filter set (F41-007 HQ-Set, AHF Analysetechnik) separates excitation and emission. The signal is detected with a CCD-Camera (ProgRes C5, Jenoptik) (see fig. 2). Exposure times of up to 10

seconds are possible. The fluorescence signal increases over time as more anti-bodies bind to the peptides.

1.3 Peptide arrays and antibodies

To establish the system, a monoclonal Anti-FLAG M1 antibody (Sigma-Aldrich) binding to the FLAG peptide (amino acid sequence DYKDDDDK) is used. The anti-body is fluorescently labelled with a Lightning-Link™ Rapid conjugation kit (Innova Biosciences) with the fluorophore Dylight550 according to the instructions. Peptides were ordered from Peps4LS and spotted with a Spotting Roboter (MicroGrid II, Genomic Solutions) onto 3D-maleimide functionalized surfaces (PolyAn).

The peptides are functionalized with a C-terminal cysteine, which contains a thiol in the side chain that binds to the maleimide. After the completion of the spotting procedure, unreacted maleimides on the surface are blocked by washing the surface with mercaptoethanol. Unbound peptides from the spotting process are removed by a washing step with 0.1% trifluoroacetic acid in acetonitrile.

2. Experiments

All buffers used for washing, blocking and incubation are degassed and filtered sterile. In the first step the microfluidic channel is filled with phosphate buffered saline with 0.05 % Tween20 (PBS-T, Sigma Aldrich) and the array is preswollen for 10 minutes under flow with a volume flow rate of 600 µl/min. Next, the system is blocked with Rockland buffer (Biotrend) for 30 minutes to prevent any unspecific binding of antibodies in the system. The incubation of the array with antibodies is realized at a flow velocity of 16.4 mm/s, resulting in a Reynolds number of 2.1 for the microfluidic channel.

Every minute, the CCD camera takes an image with an exposure time of 8 seconds and a digital gain of 8. Bias and flat field correction is performed with ImageJ. The intensity of the signal is calculated as the average grey scale value of all pixels of one peptide spot. Finally, the intensity of designated background control areas is subtracted from the intensity of the peptide spots. For the calculation of the equilibrium dissociation rate constant K_D, a pseudo-first-order model according to O'Shannassay and Winzor [170] is used.

3. Results and Outlook

With the above described system, we detected kinetic profiles of a monoclonal antibody (see fig. 4). We calculated an equilibrium dissociation rate constant K_D of $6.1 \cdot 10^{-8}$M.

In the future, we plan to investigate different peptide-antibody combinations with our system and establish reference measurements for validation. Potentially, our system will allow for obtaining dissociation rate constants for 120 different peptide spots simultaneously.
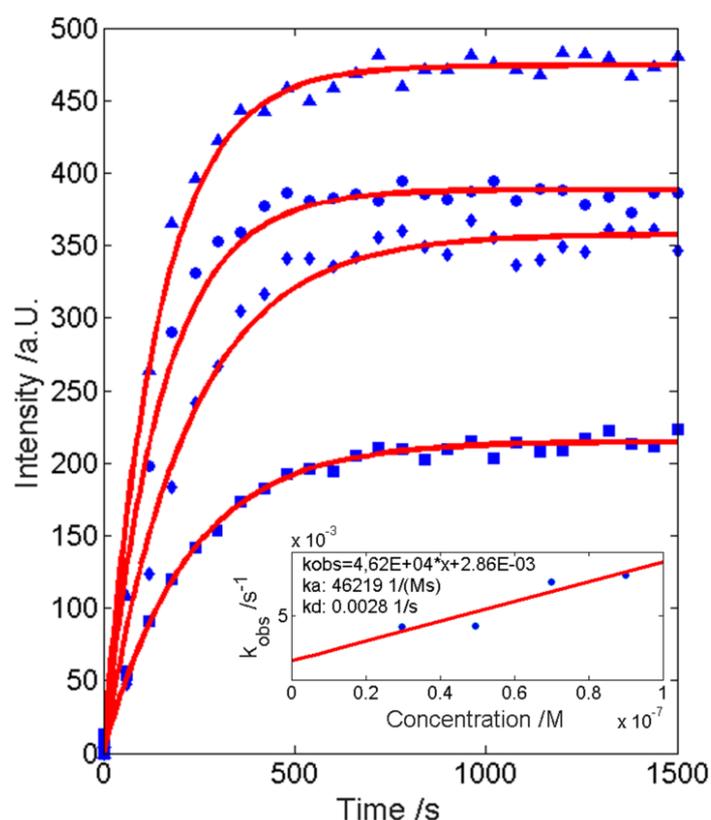
Figure 4: Fluorescence intensity detected over time for anti-FLAG antibodies labelled with Cy3 incubated in different concentrations on peptides: ■ 30 nM, ♦ 50 nM, ● 70 nM, ▲ 90 nM. Fitting according to [170]. $I = I_{max} (1 - e^{-kobst})$ with I_{max} representing the fluorescence intensity at equilibrium and $k_{obs} = k_a c + k_d$. k_a and k_d are the respective association and dissociation rate constants.

4   Conclusion and outlook

In this doctoral thesis, it was evaluated, whether antibody fingerprints are suitable for the investigation of the etiology of infectious diseases and, whether this knowledge can be employed for diagnostics. Antibody fingerprints, which are deciphered in high density peptide array screenings, describe the key residues that are indispensable for the binding of an antibody to its epitope. They define the specificity of antibodies in a very detailed manner. In this work, the basis to exploit antibody fingerprints was established.

In two model studies, it was evaluated, whether antibody fingerprints are disease specific. In these studies, the immune responses of different individuals were examined with high density peptide arrays. In a first step, immunogenic peptides were identified in epitope mappings. Subsequently, the fingerprints of identified antigenic peptides were determined with substitution analyses.

In the first model study, the antibody response of vaccinated Europeans toward the tetanus toxin was examined. The antibodies of different individuals targeted various epitopes on the toxin. According to a reference diagnostic experiment, all individuals were sufficiently immunized. Thus, either different antibody species provide protection against the toxin, or the decisive antibodies recognize conformational or discontinuous epitopes that cannot be detected with linear peptide arrays. The substitution analysis of an epitope that was targeted by multiple individuals revealed that the antibody fingerprints of different individuals are almost identical. The corresponding epitope is located on a domain of the toxin with yet unknown function. The neutralizing character of this antibody species should be subject of future studies. In the second model study, the immune response of Lyme disease patients toward the immunodominant antigen VlsE of the disease-causing spirochete *B. burgdorferi* was investigated. Similar to the first study, different combinations of epitopes were targeted, but, the antibody fingerprints for shared epitopes were again almost identical. For both studies, queries of protein databases confirmed that the study of potential cross-reacting proteins is possible, whereas especially fingerprints consisting of several amino acids have shown to be highly disease specific.

To exploit antibody fingerprints to pinpoint antigens in enigmatic infectious diseases, the antibody specificities of an individual were screened in a proof-of-principle with a random

peptide library. First, antibody-peptide interactions were investigated with a large phage display peptide library. Then, the identified binders from the error-prone phage display approach were validated with peptide arrays. Afterwards, the validated binders were studied in substitution analyses to determine the antibody fingerprints. As a result, 73 antibody species fingerprints could be identified, which might represent significant parts of the antibody diversity of the examined individual. Potential causing antigens of the antibody species could be identified in bioinformatical analyses by querying public protein databases. Most likely, one of the identified prominent motifs correlates with an antibody species induced by the vaccination against the polio virus.

Furthermore, a microfluidic system with an optical set-up was established and validated. The binding curves of a monoclonal antibody in different concentrations towards its epitope were detected in a time-resolved manner.

In conclusion, the antibody fingerprints deciphered with high density peptide arrays were disease-specific in both investigated model studies. Therefore, they have a great potential in diagnostics, enabling studies of inducing and cross-reacting antigens. The specificity to other diseases needs to be subject of future studies. For the identification of pathogens in hitherto enigmatic infectious diseases, the immune responses of large groups of diseased vs. healthy patients should be compared in the future. It should be investigated, whether it is possible to identify shared antibody species in these groups without *a priori* knowledge. Then, the targeted antigens and pathogens could be pinpointed with bioinformatic analyses. A prerequisite for the latter is the presence of a correlating antibody species that targets a linear epitope, which can be detected with current peptide array technology. The examination of the kinetics of the antibody-antigen interaction will even increase the knowledge of antibody-antigen interactions in the future.

5 Bibliography

[1]   Murphy, K. and C. Weaver, Janeway's immunobiology. 2016: Garland Science.

[2]   Abbas, A.K., A.H. Lichtman, and S. Pillai, Cellular and molecular immunology. 2014: Elsevier Health Sciences.

[3]   Nemazee, D., Receptor editing in lymphocyte development and central tolerance. *Nature Reviews Immunology*, 2006. 6(10): p. 728-740.

[4]   Pieper, K., B. Grimbacher, and H. Eibel, B-cell biology and development. *Journal of Allergy and Clinical Immunology*, 2013. 131(4): p. 959-971.

[5]   Jerne, N.K., The natural-selection theory of antibody formation. *Proceedings of the National Academy of Sciences*, 1955. 41(11): p. 849-857.

[6]   Nossal, G.J., Antibody production by single cells. *British journal of experimental pathology*, 1958. 39(5): p. 544.

[7]   Parker, D.C., T cell-dependent B cell activation. *Annual review of immunology*, 1993. 11(1): p. 331-360.

[8]   Gonzalez- Quintela, A., R. Alende, F. Gude, J. Campos, et al., Serum levels of immunoglobulins (IgG, IgA, IgM) in a general adult population and their relationship with alcohol consumption, smoking and common metabolic abnormalities. *Clinical & Experimental Immunology*, 2008. 151(1): p. 42-50.

[9]   Lebman, D.A. and R.L. Coffman, Interleukin 4 causes isotype switching to IgE in T cell-stimulated clonal B cell cultures. *The Journal of experimental medicine*, 1988. 168(3): p. 853-862.

[10]   Harris, L.J., S.B. Larson, K.W. Hasel, and A. McPherson, Refined structure of an intact IgG2a monoclonal antibody. *Biochemistry*, 1997. 36(7): p. 1581-1597.

[11]   Radbruch, A., G. Muehlinghaus, E.O. Luger, A. Inamine, et al., Competence and competition: the challenge of becoming a long-lived plasma cell. *Nature Reviews Immunology*, 2006. 6(10): p. 741-750.

[12]   Werther, W.A., T.N. Gonzalez, S.J. O'Connor, S. McCabe, et al., Humanization
       of an anti-lymphocyte function-associated antigen (LFA)-1 monoclonal
       antibody and reengineering of the humanized antibody for binding to rhesus
       LFA-1. *The Journal of Immunology*, 1996. 157(11): p. 4986-4995.

[13]   Janeway, C.A. Approaching the asymptote? Evolution and revolution in
       immunology. in *Cold Spring Harbor symposia on quantitative biology*. 1989.
       Cold Spring Harbor Laboratory Press.

[14]   O'Neill, L.A., D. Golenbock, and A.G. Bowie, The history of Toll-like
       receptors [mdash] redefining innate immunity. *Nature Reviews Immunology*,
       2013. 13(6): p. 453-460.

[15]   Aderem, A. and R.J. Ulevitch, Toll-like receptors in the induction of the innate
       immune response. *Nature*, 2000. 406(6797): p. 782-787.

[16]   Lemaitre, B., E. Nicolas, L. Michaut, J.-M. Reichhart, et al., The dorsoventral
       regulatory gene cassette spätzle/Toll/cactus controls the potent antifungal
       response in Drosophila adults. *Cell*, 1996. 86(6): p. 973-983.

[17]   Williams, M.J., A. Rodriguez, D.A. Kimbrell, and E.D. Eldon, The 18- wheeler
       mutation reveals complex antibacterial gene regulation in Drosophila host
       defense. *The EMBO Journal*, 1997. 16(20): p. 6120-6130.

[18]   Kawasaki, T. and T. Kawai, Toll-like receptor signaling pathways. *Frontiers in
       immunology*, 2014. 5.

[19]   Iwasaki, A. and R. Medzhitov, Control of adaptive immunity by the innate
       immune system. *Nature immunology*, 2015. 16(4): p. 343-353.

[20]   Merad, M., P. Sathe, J. Helft, J. Miller, et al., The dendritic cell lineage:
       ontogeny and function of dendritic cells and their subsets in the steady state and
       the inflamed setting. *Annual review of immunology*, 2013. 31.

[21]   Nimmerjahn, F. and J.V. Ravetch, Divergent immunoglobulin g subclass
       activity through selective Fc receptor binding. *Science*, 2005. 310(5753): p.
       1510-1512.

[22]   Jenner, E., On the origin of the vaccine inoculation. 1863: G. Elsick.

[23]   Riedel, S. Edward Jenner and the history of smallpox and vaccination. in
       Baylor University Medical Center. Proceedings. 2005. Baylor University
       Medical Center.

[24]   Koff, W.C., D.R. Burton, P.R. Johnson, B.D. Walker, et al., Accelerating next-generation vaccine development for global disease prevention. *Science*, 2013. 340(6136): p. 1232910.

[25]   Halstead, S.B., Pathogenesis of dengue: challenges to molecular biology. *Science*, 1988. 239(4839): p. 476.

[26]   Openshaw, P.J. and J.S. Tregoning, Immune responses and disease enhancement during respiratory syncytial virus infection. *Clinical microbiology reviews*, 2005. 18(3): p. 541-555.

[27]   Huisman, W., B. Martina, G. Rimmelzwaan, R. Gruters, et al., Vaccine-induced enhancement of viral infections. *Vaccine*, 2009. 27(4): p. 505-512.

[28]   Khurana, S., C.L. Loving, J. Manischewitz, L.R. King, et al., Vaccine-induced anti-HA2 antibodies promote virus fusion and enhance influenza virus respiratory disease. *Science Translational Medicine*, 2013. 5(200): p. 200ra114-200ra114.

[29]   Plotkin, S.A., Correlates of protection induced by vaccination. *Clinical and Vaccine Immunology*, 2010. 17(7): p. 1055-1065.

[30]   Burton, D.R., R.C. Desrosiers, R.W. Doms, W.C. Koff, et al., HIV vaccine design and the neutralizing antibody problem. *Nature immunology*, 2004. 5(3): p. 233-236.

[31]   Kwong, P.D., M.L. Doyle, D.J. Casper, C. Cicala, et al., HIV-1 evades antibody-mediated neutralization through conformational masking of receptor-binding sites. *Nature*, 2002. 420(6916): p. 678-682.

[32]   Stamatatos, L., L. Morris, D.R. Burton, and J.R. Mascola, Neutralizing antibodies generated during natural HIV-1 infection: good news for an HIV-1 vaccine? *Nature medicine*, 2009. 15(8): p. 866-870.

[33]   Sather, D.N., J. Armann, L.K. Ching, A. Mavrantoni, et al., Factors associated with the development of cross-reactive neutralizing antibodies during human immunodeficiency virus type 1 infection. *Journal of virology*, 2009. 83(2): p. 757-769.

[34]   Hafalla, J.C., O. Silvie, and K. Matuschewski, Cell biology and immunology of malaria. *Immunological reviews*, 2011. 240(1): p. 297-316.

[35]   Wu, X., T. Zhou, J. Zhu, B. Zhang, et al., Focused evolution of HIV-1 neutralizing antibodies revealed by structures and deep sequencing. *Science*, 2011. 333(6049): p. 1593-1602.

[36]   Davies, D.H., X. Liang, J.E. Hernandez, A. Randall, et al., Profiling the humoral immune response to infection by using proteome microarrays: high-throughput vaccine and diagnostic antigen discovery. *Proceedings of the National Academy of Sciences of the United States of America*, 2005. 102(3): p. 547-552.

[37]   Reche, P.A., E. Fernandez-Caldas, D.R. Flower, M. Fridkis-Hareli, et al., Peptide-based immunotherapeutics and vaccines. *Journal of immunology research*, 2014. 2014.

[38]   Steward, M.W. and A.M. Lew, The importance of antibody affinity in the performance of immunoassays for antibody. *Journal of immunological methods*, 1985. 78(2): p. 173-190.

[39]   Glaser, R.W., Antigen-antibody binding and mass transport by convection and diffusion to a surface: a two-dimensional computer model of binding and dissociation kinetics. *Analytical biochemistry*, 1993. 213(1): p. 152-161.

[40]   O'Shannessy, D.J., Determination of kinetic rate and equilibrium binding constants for macromolecular interactions: a critique of the surface plasmon resonance literature. *Current opinion in biotechnology*, 1994. 5(1): p. 65-71.

[41]   Sadana, A. and D. Sii, Binding kinetics of antigen by immobilized antibody: influence of reaction order and external diffusional limitations. *Biosensors and Bioelectronics*, 1992. 7(8): p. 559-568.

[42]   Atkins, P., Physical Chemistry, 492. 1990, Oxford University Press, Fourth Edition, Oxford, UK.

[43]   Towbin, H., T. Staehelin, and J. Gordon, Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proceedings of the National Academy of Sciences*, 1979. 76(9): p. 4350-4354.

[44]   Haury, M., A. Grandien, A. Sundblad, A. Coutinho, et al., Global analysis of antibody repertoires. 1. An immunoblot method for the quantitative screening of a large number of reactivities. *Scandinavian journal of immunology*, 1994. 39(1): p. 79-87.

[45]   Stahl, D., S. Lacroix-Desmazes, W. Sibrowski, M.D. Kazatchkine, et al., Broad alterations of self-reactive antibody-repertoires of plasma IgM and IgG in B-cell chronic lymphocytic leukemia (B-CLL) and B-CLL related target-restricted autoimmunity. *Leukemia & lymphoma*, 2001. 42(1-2): p. 163-176.

[46]   Sundblad, A., C. Ferreira, A. Nobrega, M. Haury, et al., Characteristic generated alterations of autoantibody patterns in idiopathic thrombocytopenic purpura. *Journal of autoimmunity*, 1997. 10(2): p. 193-201.

[47]   Quintana, F.J., G. Getz, G. Hed, E. Domany, et al., Cluster analysis of human autoantibody reactivities in health and in type 1 diabetes mellitus: a bio-informatic approach to immune complexity. *Journal of autoimmunity*, 2003. 21(1): p. 65-75.

[48]   Cwirla, S.E., E.A. Peters, R.W. Barrett, and W.J. Dower, Peptides on phage: a vast library of peptides for identifying ligands. *Proceedings of the National Academy of Sciences*, 1990. 87(16): p. 6378-6382.

[49]   Scott, J.K. and G.P. Smith, Searching for peptide ligands with an epitope library. *Science*, 1990. 249(4967): p. 386-390.

[50]   Devlin, J.J., L.C. Panganiban, and P.E. Devlin, Random peptide libraries: a source of specific protein binding molecules. *Science*, 1990. 249(4967): p. 404-406.

[51]   Smith, G.P. and V.A. Petrenko, Phage display. *Chemical reviews*, 1997. 97(2): p. 391-410.

[52]   Hansen, M.H., B. Ostenstad, and M. Sioud, Antigen-specific IgG antibodies in stage IV long-time survival breast cancer patients. *Molecular Medicine*, 2001. 7(4): p. 230.

[53]   Wang, X., J. Yu, A. Sreekumar, S. Varambally, et al., Autoantibody signatures in prostate cancer. *New England Journal of Medicine*, 2005. 353(12): p. 1224-1235.

[54]   Xu, G.J., T. Kula, Q. Xu, M.Z. Li, et al., Comprehensive serological profiling of human populations using a synthetic human virome. *Science*, 2015. 348(6239): p. aaa0698.

[55]   Bongartz, T., A.J. Sutton, M.J. Sweeting, I. Buchan, et al., Anti-TNF antibody therapy in rheumatoid arthritis and the risk of serious infections and

malignancies: systematic review and meta-analysis of rare harmful effects in randomized controlled trials. *Jama*, 2006. 295(19): p. 2275-2285.

[56]   Weinstein, J.A., N. Jiang, R.A. White, D.S. Fisher, et al., High-throughput sequencing of the zebrafish antibody repertoire. *Science*, 2009. 324(5928): p. 807-810.

[57]   Georgiou, G., G.C. Ippolito, J. Beausang, C.E. Busse, et al., The promise and challenge of high-throughput sequencing of the antibody repertoire. *Nature biotechnology*, 2014. 32(2): p. 158-168.

[58]   Khan, T.A., S. Friedensohn, A.R.G. de Vries, J. Straszewski, et al., Accurate and predictive antibody repertoire profiling by molecular amplification fingerprinting. *Science Advances*, 2016. 2(3): p. e1501371.

[59]   Hu, S., Y. Li, G. Liu, Q. Song, et al., A protein chip approach for high-throughput antigen identification and characterization. *Proteomics*, 2007. 7(13): p. 2151-2161.

[60]   Doolan, D.L., Y. Mu, B. Unal, S. Sundaresh, et al., Profiling humoral immune responses to P. falciparum infection with protein microarrays. *Proteomics*, 2008. 8(22): p. 4680-4694.

[61]   Vigil, A., C. Chen, A. Jain, R. Nakajima-Sasaki, et al., Profiling the humoral immune response of acute and chronic Q fever by protein microarray. *Molecular & Cellular Proteomics*, 2011. 10(10): p. M110. 006304.

[62]   Liang, L., X. Tan, S. Juarez, H. Villaverde, et al., Systems biology approach predicts antibody signature associated with Brucella melitensis infection in humans. *Journal of proteome research*, 2011. 10(10): p. 4813-4824.

[63]   Merrifield, R., Automated synthesis of peptides. *Science*, 1965. 150(3693): p. 178-185.

[64]   Frank, R., Spot-synthesis: an easy technique for the positionally addressable, parallel chemical synthesis on a membrane support. *Tetrahedron*, 1992. 48(42): p. 9217-9232.

[65]   Fodor, S.P., J.L. Read, M.C. Pirrung, L. Stryer, et al., Light-directed, spatially addressable parallel chemical synthesis. *science*, 1991: p. 767-773.

[66]   Katz, C., L. Levy-Beladev, S. Rotem-Bamberger, T. Rito, et al., Studying protein–protein interactions using peptide arrays. *Chemical Society Reviews*, 2011. 40(5): p. 2131-2145.

[67]   Stadler, V., T. Felgenhauer, M. Beyer, S. Fernandez, et al., Combinatorial synthesis of peptide arrays with a laser printer. *Angewandte Chemie International Edition*, 2008. 47(37): p. 7132-7135.

[68]   Kühne, Y., G. Reese, B.K. Ballmer-Weber, B. Niggemann, et al., A novel multipeptide microarray for the specific and sensitive mapping of linear IgE-binding epitopes of food allergens. *International archives of allergy and immunology*, 2015. 166(3): p. 213-224.

[69]   Müller, A.-M., M. Bockstahler, G. Hristov, C. Weiß, et al., Identification of novel antigens contributing to autoimmunity in cardiovascular diseases. *Clinical Immunology*, 2016. 173: p. 64-75.

[70]   Mock, A. and C. Herold-Mende, Non-invasive glioblastoma immunoprofiling by printed peptide arrays. *OncoImmunology*, 2016. 5(2): p. e1069941.

[71]   Gaseitsiwe, S., D. Valentini, S. Mahdavifar, I. Magalhaes, et al., Pattern recognition in pulmonary tuberculosis defined by high content peptide microarray chip analysis representing 61 proteins from M. tuberculosis. *PLoS One*, 2008. 3(12): p. e3840.

[72]   Frank, S.A., Immunology and evolution of infectious disease. 2002: Princeton University Press.

[73]   Reineke, U., C. Ivascu, M. Schlief, C. Landgraf, et al., Identification of distinct antibody epitopes and mimotopes from a peptide array of 5520 randomly generated sequences. *Journal of immunological methods*, 2002. 267(1): p. 37-51.

[74]   Michaud, G.A., M. Salcius, F. Zhou, R. Bangham, et al., Analyzing antibody specificity with whole proteome microarrays. *Nature biotechnology*, 2003. 21(12): p. 1509-1512.

[75]   Cohn, M. and R.E. Langman, The protection: the unit of humoral immunity selected by evolution. *Immunological reviews*, 1990. 115(1): p. 7-147.

[76]   Geysen, H.M., S.J. Rodda, T.J. Mason, G. Tribbick, et al., Strategies for epitope analysis using peptide synthesis. *Journal of immunological methods*, 1987. 102(2): p. 259-274.

[77]   Kramer, A., T. Keitel, K. Winkler, W. Stöcklein, et al., Molecular basis for the binding promiscuity of an anti-p24 (HIV-1) monoclonal antibody. *Cell*, 1997. 91(6): p. 799-809.

[78]   Ehreth, J., The global value of vaccination. *Vaccine*, 2003. 21(7): p. 596-600.

[79]   Germain, R.N., Vaccines and the future of human immunology. *Immunity*, 2010. 33(4): p. 441-450.

[80]   Pulendran, B. and R. Ahmed, Immunological mechanisms of vaccination. *Nature immunology*, 2011. 12(6): p. 509-517.

[81]   Cooper, P., I. Espinel, W. Paredes, R. Guderian, et al., Impaired tetanus-specific cellular and humoral responses following tetanus vaccination in human onchocerciasis: a possible role for interleukin-10. *Journal of Infectious Diseases*, 1998. 178(4): p. 1133-1138.

[82]   Cook, T., R. Protheroe, and J. Handel, Tetanus: a review of the literature. *British Journal of Anaesthesia*, 2001. 87(3): p. 477-487.

[83]   Machingaidze, S., E. Rehfuess, R. von Kries, G.D. Hussey, et al., Understanding interventions for improving routine immunization coverage in children in low-and middle-income countries: a systematic review protocol. *Systematic reviews*, 2013. 2(1): p. 1.

[84]   Fairweather, N.F. and V.A. Lyness, The complete nucleotide sequence of tetanus toxin. *Nucleic acids research*, 1986. 14(19): p. 7809-7812.

[85]   Schiavo, G.G., F. Benfenati, B. Poulain, O. Rossetto, et al., Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolytic cleavage of synaptobrevin. *Nature*, 1992. 359(6398): p. 832-835.

[86]   Volk, W., B. Bizzini, R. Snyder, E. Bernhard, et al., Neutralization of tetanus toxin by distinct monoclonal antibodies binding to multiple epitopes on the toxin molecule. *Infection and immunity*, 1984. 45(3): p. 604-609.

[87]   Weber, L.K., A. Palermo, J. Kügler, O. Armant, et al., Single amino acid fingerprinting of the human antibody repertoire with high density peptide arrays. *Journal of Immunological Methods*, 2017. 443: p. 45-54.

[88]   Beyer, M., A. Nesterov, I. Block, K. König, et al., Combinatorial synthesis of peptide arrays onto a microchip. *Science*, 2007. 318(5858): p. 1888-1888.

[89] Loeffler, F., C. Schirwitz, J. Wagner, K. Koenig, et al., Biomolecule arrays using functional combinatorial particle patterning on microchips. *Advanced Functional Materials*, 2012. 22(12): p. 2503-2508.

[90] Loeffler, F.F., T.C. Foertsch, R. Popov, D.S. Mattes, et al., High-flexibility combinatorial peptide synthesis with laser-based transfer of monomers in solid matrix material. *Nature communications*, 2016. 7.

[91] Maerkle, F., F.F. Loeffler, S. Schillo, T. Foertsch, et al., High- Density Peptide Arrays with Combinatorial Laser Fusing. *Advanced Materials*, 2014. 26(22): p. 3730-3734.

[92] Weber, L.K., A. Fischer, T. Schorb, M. Soehindrijo, et al., Automated microfluidic system with optical set up for the investigation of peptide-antibody interactions in an array format. *Microsystems Technology in Germany 2016*, ISSN 2191-7183

[93] Buus, S., J. Rockberg, B. Forsström, P. Nilsson, et al., High-resolution mapping of linear antibody epitopes using ultrahigh-density peptide microarrays. *Molecular & Cellular Proteomics*, 2012. 11(12): p. 1790-1800.

[94] De Castro, E., C.J. Sigrist, A. Gattiker, V. Bulliard, et al., ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic acids research*, 2006. 34(suppl 2): p. W362-W365.

[95] Schiavo, G., M. Matteoli, and C. Montecucco, Neurotoxins affecting neuroexocytosis. *Physiological reviews*, 2000. 80(2): p. 717-766.

[96] Lavinder, J.J., Y. Wine, C. Giesecke, G.C. Ippolito, et al., Identification and characterization of the constituent human serum antibodies elicited by vaccination. *Proceedings of the National Academy of Sciences*, 2014. 111(6): p. 2259-2264.

[97] Binz, T. and A. Rummel, Cell entry strategy of clostridial neurotoxins. *Journal of neurochemistry*, 2009. 109(6): p. 1584-1595.

[98] Gustafsson, B., E. Whitmore, and M. Tiru, Neutralization of tetanus toxin by human monoclonal antibodies directed against tetanus toxin fragment C. *Hybridoma*, 1993. 12(6): p. 699-708.

[99] Yousefi, M., R. Khosravi-Eghbal, A. Reza Mahmoudi, M. Jeddi-Tehrani, et al., Comparative in vitro and in vivo assessment of toxin neutralization by anti-

tetanus toxin monoclonal antibodies. *Human vaccines & immunotherapeutics*, 2014. 10(2): p. 344-351.

[100]   Parton, R., C. Ockleford, and D. Critchley, Tetanus toxin binding to mouse spinal cord cells: an evaluation of the role of gangliosides in toxin internalization. *Brain research*, 1988. 475(1): p. 118-127.

[101]   Pierce, E.J., M.D. Davison, R. Parton, W.H. Habig, et al., Characterization of tetanus toxin binding to rat brain membranes. Evidence for a high-affinity proteinase-sensitive receptor. *Biochemical Journal*, 1986. 236(3): p. 845-852.

[102]   Schiavo, G., G. Ferrari, O. Rossetto, and C. Montecucco, Tetanus toxin receptor Specific cross- linking of tetanus toxin to a protein of NGF- differentiated PC 12 cells. *FEBS letters*, 1991. 290(1-2): p. 227-230.

[103]   Wang, H., R. Yu, T. Fang, T. Yu, et al., Tetanus Neurotoxin Neutralizing Antibodies Screened from a Human Immune scFv Antibody Phage Display Library. *Toxins*, 2016. 8(9): p. 266.

[104]   Ref. Dassault Systèmes BIOVIA, Discovery Studio Modeling Environment, Release 2017, San Diego: Dassault Systèmes, 2016.

[105]   Knapp, M., B. Segelke, and B. Rupp, The 1.61 Angstrom structure of the tetanus toxin. Ganglioside binding region: solved by MAD and MIR phase combination. *Am. Crystallogr. Assoc*, 1998. 25: p. 90.

[106]   Rossmann, M.G., E. Arnold, J.W. Erickson, E.A. Frankenberger, et al., Structure of a human common cold virus and functional relationship to other picornaviruses. *Nature*, 1985. 317(6033): p. 145.

[107]   Niespodziana, K., K. Napora, C. Cabauatan, M. Focke-Tejkl, et al., Misdirected antibody responses against an N-terminal epitope on human rhinovirus VP1 as explanation for recurrent RV infections. *The FASEB Journal*, 2012. 26(3): p. 1001-1008.

[108]   Smith, T.J., M.J. Kremer, M. Luo, G. Vriend, et al., The site of attachment in human rhinovirus 14 for antiviral agents that inhibit uncoating. *Science*, 1986. 233(4770): p. 1286-1293.

[109]   Palmenberg, A.C., D. Spiro, R. Kuzmickas, S. Wang, et al., Sequencing and analyses of all known human rhinovirus genomes reveal structure and evolution. *Science*, 2009. 324(5923): p. 55-59.

[110] Mäkelä, M.J., T. Puhakka, O. Ruuskanen, M. Leinonen, et al., Viruses and bacteria in the etiology of the common cold. *Journal of clinical microbiology*, 1998. 36(2): p. 539-542.

[111] Stricker, R.B. and L. Johnson, Lyme disease: the next decade. *Infect Drug Resist*, 2011. 4: p. 1-9.

[112] Stricker, R.B., A. Lautin, and J.J. Burrascano, Lyme disease: point/counterpoint. *Expert review of anti-infective therapy*, 2005. 3(2): p. 155-165.

[113] Donta, S.T., Issues in the diagnosis and treatment of lyme disease. *The open neurology journal*, 2012. 6(1).

[114] Fister, R., L. Weymouth, J. McLaughlin, R. Ryan, et al., Comparative evaluation of three products for the detection of Borrelia burgdorferi antibody in human serum. *Journal of clinical microbiology*, 1989. 27(12): p. 2834-2837.

[115] Aguero-Rosenfeld, M.E., G. Wang, I. Schwartz, and G.P. Wormser, Diagnosis of Lyme borreliosis. *Clinical microbiology reviews*, 2005. 18(3): p. 484-509.

[116] Centers for Disease Control and Prevention, Recommendations for test performance and interpretation from the Second National Conference on Serologic Diagnosis of Lyme Disease. *MMWR. Morbidity and mortality weekly report*, 1995. 44(31): p. 590-591

[117] Dressler, F., J.A. Whalen, B.N. Reinhardt, and A.C. Steere, Western blotting in the serodiagnosis of Lyme disease. *Journal of Infectious Diseases*, 1993. 167(2): p. 392-400.

[118] Goettner, G., U. Schulte-Spechtel, R. Hillermann, G. Liegl, et al., Improvement of Lyme borreliosis serodiagnosis by a newly developed recombinant immunoglobulin G (IgG) and IgM line immunoblot assay and addition of VlsE and DbpA homologues. *Journal of clinical microbiology*, 2005. 43(8): p. 3602-3609.

[119] Bruckbauer, H., V. Preac-Mursic, R. Fuchs, and B. Wilske, Cross-reactive proteins ofBorrelia burgdorferi. *European Journal of Clinical Microbiology and Infectious Diseases*, 1992. 11(3): p. 224-232.

[120] Coleman, J.L. and J.L. Benach, Characterization of antigenic determinants of Borrelia burgdorferi shared by other bacteria. *Journal of Infectious Diseases*, 1992. 165(4): p. 658-666.

[121] Magnarelli, L.A., J.N. Miller, J.F. Anderson, and G. Riviere, Cross-reactivity of nonspecific treponemal antibody in serologic tests for Lyme disease. *Journal of clinical microbiology*, 1990. 28(6): p. 1276-1279.

[122] Mayo, D.R. and D.W. Vance Jr, Parvovirus B19 as the cause of a syndrome resembling Lyme arthritis in adults. *N Engl J Med*, 1991. 324: p. 419-420.

[123] Jauris-Heipke, S., R. Fuchs, M. Motz, V. Preac-Mursic, et al., Genetic heterogenity of the genes coding for the outer surface protein C (OspC) and the flagellin of Borrelia burgdorferi. *Medical microbiology and immunology*, 1993. 182(1): p. 37-50.

[124] Roessler, D., U. Hauser, and B. Wilske, Heterogeneity of BmpA (P39) among European isolates of Borrelia burgdorferi sensu lato and influence of interspecies variability on serodiagnosis. *Journal of clinical microbiology*, 1997. 35(11): p. 2752-2758.

[125] Theisen, M., B. Frederiksen, A. Lebech, J. Vuust, et al., Polymorphism in ospC gene of Borrelia burgdorferi and immunoreactivity of OspC protein: implications for taxonomy and for use of OspC protein as a diagnostic antigen. *Journal of clinical microbiology*, 1993. 31(10): p. 2570-2576.

[126] Seifert, H.S. and M. So, Genetic mechanisms of bacterial antigenic variation. *Microbiological reviews*, 1988. 52(3): p. 327-336.

[127] Zhang, J.-R., J.M. Hardham, A.G. Barbour, and S.J. Norris, Antigenic variation in Lyme disease borreliae by promiscuous recombination of VMP-like sequence cassettes. *Cell*, 1997. 89(2): p. 275-285.

[128] McDowell, J.V., S.-Y. Sung, L.T. Hu, and R.T. Marconi, Evidence that the variable regions of the central domain of VlsE are antigenic during infection with Lyme disease spirochetes. *Infection and immunity*, 2002. 70(8): p. 4196-4203.

[129] Liang, F.T., A.C. Steere, A.R. Marques, B.J. Johnson, et al., Sensitive and specific serodiagnosis of Lyme disease by enzyme-linked immunosorbent assay with a peptide based on an immunodominant conserved region of Borrelia burgdorferi VlsE. *Journal of clinical microbiology*, 1999. 37(12): p. 3990-3996.

[130] Chandra, A., N. Latov, G.P. Wormser, A.R. Marques, et al., Epitope mapping of antibodies to VlsE protein of Borrelia burgdorferi in post-Lyme disease syndrome. *Clinical immunology*, 2011. 141(1): p. 103-110.

[131] Steere, A.C., G.J. Hutchinson, D.W. Rahn, L.H. SIGAL, et al., Treatment of the
       early manifestations of Lyme disease. *Annals of Internal Medicine*, 1983. 99(1):
       p. 22-26.

[132] Wormser, G.P., R.J. Dattwyler, E.D. Shapiro, J.J. Halperin, et al., The clinical
       assessment, treatment, and prevention of Lyme disease, human granulocytic
       anaplasmosis, and babesiosis: clinical practice guidelines by the Infectious
       Diseases Society of America. *Clinical Infectious Diseases*, 2006. 43(9): p.
       1089-1134.

[133] Stadler, V., T. Felgenhauer, M. Beyer, S. Fernandez, et al., Combinatorial
       Synthesis of Peptide Arrays with a Laser Printer. *Angewandte Chemie
       International Edition*, 2008. 47(37): p. 7132-7135.

[134] Maccari, G., A. Genoni, S. Sansonno, and A. Toniolo, Properties of Two
       Enterovirus Antibodies that are Utilized in Diabetes Research. *Scientific
       reports*, 2016. 6.

[135] Eicken, C., V. Sharma, T. Klabunde, M.B. Lawrenz, et al., Crystal structure of
       Lyme disease variable surface antigen VlsE of Borrelia burgdorferi. *Journal of
       Biological Chemistry*, 2002. 277(24): p. 21691-21696.

[136] Wang, D., L. Yang, P. Zhang, J. LaBaer, et al., AAgAtlas 1.0: a human
       autoantigen database. *Nucleic Acids Research*, 2017. 45(D1): p. D769-D776.

[137] Emery, R., S.C. Eppes, J.D. Klein, and C.D. Rose, Misdiagnosis of Parvovirus
       B19 Infection as Lyme Disease: A Series of Patients with Falsely Positive
       Lyme Serology. *Infectious Diseases in Clinical Practice*, 1997. 6(5): p. 342-
       344.

[138] Strasfeld, L., L. Romanzi, R.H. Seder, and V.P. Berardi, False-positive
       serological test results for Lyme disease in a patient with acute herpes simplex
       virus type 2 infection. *Clinical infectious diseases*, 2005. 41(12): p. 1826-1827.

[139] Feder Jr, H.M., M.A. Gerber, S.W. Luger, and R.W. Ryan, False positive
       serologic tests for Lyme disease after varicella infection. *The New England
       journal of medicine*, 1991. 325(26): p. 1886-1887.

[140] Goossens, H., A. Van Den Bogaard, and M. Nohlmans, Evaluation of fifteen
       commercially available serological tests for diagnosis of Lyme borreliosis.
       European Journal of Clinical Microbiology & Infectious Diseases, 1999. 18(8):
       p. 551-560.

[141]   Sykes, K.F., J.B. Legutki, and P. Stafford, Immunosignaturing: a critical
        review. *Trends in biotechnology*, 2013. 31(1): p. 45-51.

[142]   Braunschweig, D., P. Krakowiak, P. Duncanson, R. Boyce, et al., Autism-
        specific maternal autoantibodies recognize critical proteins in developing brain.
        Translational psychiatry, 2013. 3(7): e277.

[143]   Marinò, M., L. Chiovato, J.A. Friedlander, F. Latrofa, et al., Serum Antibodies
        against Megalin (GP330) in Patients with Autoimmune Thyroiditis 1. *The
        Journal of Clinical Endocrinology & Metabolism*, 1999. 84(7): p. 2468-2474.

[144]   Tatsumi, K.-i., S. Tanaka, T. Takano, S. Tahara, et al., Frequent appearance of
        autoantibodies against prohormone convertase 1/3 and neuroendocrine protein
        7B2 in patients with nonfunctioning pituitary macroadenoma. *Endocrine*, 2003.
        22(3): p. 335-340.

[145]   Sakiyama, T., H. Fujita, and H. Tsubouchi, Autoantibodies against
        ubiquitination factor E4A (UBE4A) are associated with severity of Crohn's
        disease. *Inflammatory bowel diseases*, 2008. 14(3): p. 310-317.

[146]   Chan, E.K. and E. Tan, Human autoantibody-reactive epitopes of SS-B/La are
        highly conserved in comparison with epitopes recognized by murine
        monoclonal antibodies. *Journal of Experimental Medicine* 1987. 166(6): p.
        1627-1640.

[147]   Bonifacio, E., S. Genovese, S. Braghi, E. Bazzigaluppi, et al., Islet autoantibody
        markers in IDDM: risk assessment strategies yielding high sensitivity.
        Diabetologia, 1995. 38(7): p. 816-822.

[148]   Fried, A.J. and F.A. Bonilla, Pathogenesis, diagnosis, and management of
        primary antibody deficiencies and infections. *Clinical microbiology reviews*,
        2009. 22(3): p. 396-414.

[149]   Borrebaeck, C.A., Antibodies in diagnostics–from immunoassays to protein
        chips. *Immunology today*, 2000. 21(8): p. 379-382.

[150]   Stahl, D., S. Lacroix-Desmazes, L. Mouthon, S.V. Kaveri, et al., Analysis of
        human self-reactive antibody repertoires by quantitative immunoblotting.
        Journal of Immunological Methods, 2000. 240(1–2): p. 1-14.

[151]   Crompton, P.D., M.A. Kayala, B. Traore, K. Kayentao, et al., A prospective
        analysis of the Ab response to Plasmodium falciparum before and after a

malaria season by protein microarray. *Proceedings of the National Academy of Sciences*, 2010. 107(15): p. 6958-6963.

[152]   Tobin, G.J., J.D. Trujillo, R.V. Bushnell, G. Lin, et al., Deceptive imprinting and immune refocusing in vaccine design. *Vaccine*, 2008. 26(49): p. 6189-6199.

[153]   Legutki, J.B., D.M. Magee, P. Stafford, and S.A. Johnston, A general method for characterization of humoral immunity induced by a vaccine or infection. *Vaccine*, 2010. 28(28): p. 4529-4537.

[154]   Legutki, J.B. and S.A. Johnston, Immunosignatures can predict vaccine efficacy. *Proceedings of the National Academy of Sciences*, 2013. 110(46): p. 18614-18619.

[155]   Stafford, P., Z. Cichacz, N.W. Woodbury, and S.A. Johnston, Immunosignature system for diagnosis of cancer. *Proceedings of the National Academy of Sciences*, 2014. 111(30): p. E3072-E3080.

[156]   Smith, G.P., Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science*, 1985. 228(4705): p. 1315-1317.

[157]   Bongartz, J., N. Bruni, and M. Or-Guil, Epitope mapping using randomly generated peptide libraries. *Epitope Mapping Protocols: Second Edition*, 2009: p. 237-246.

[158]   Dybwad, A., Ø. Førre, J. Kjeldsen- Kragh, J.B. Natvig, et al., Identification of new B cell epitopes in the sera of rheumatoid arthritis patients using a random nanopeptide phage library. *European journal of immunology*, 1993. 23(12): p. 3189-3193.

[159]   Matochko, W.L., K. Chu, B. Jin, S.W. Lee, et al., Deep sequencing analysis of phage libraries using Illumina platform. *Methods*, 2012. 58(1): p. 47-55.

[160]   Sundaresh, S., D.L. Doolan, S. Hirst, Y. Mu, et al., Identification of humoral immune responses in protein microarrays using DNA microarray data analysis techniques. *Bioinformatics*, 2006. 22(14): p. 1760-1766.

[161]   Bailey, T.L. and C. Elkan, Fitting a mixture model by expectation maximization to discover motifs in bipolymers. *Proceedings of the International Conference on Intelligent Systems for Molecular Biology*, 1994. 2(2): p. 28-36.

[162] Lutz, H.U., F. Bussolino, R. Flepp, S. Fasler, et al., Naturally occurring anti-band-3 antibodies and complement together mediate phagocytosis of oxidatively stressed human erythrocytes. *Proceedings of the National Academy of Sciences*, 1987. 84(21): p. 7368-7372.

[163] Fleming, S.D., M.R. Pope, S.M. Hoffman, T. Moses, et al., Domain V Peptides Inhibit β2-Glycoprotein I-Mediated Mesenteric Ischemia/Reperfusion-Induced Tissue Damage and Inflammation. *The Journal of Immunology*, 2010. 185(10): p. 6168-6178.

[164] Sivalingam, G.N. and A.J. Shepherd, An analysis of B-cell epitope discontinuity. *Molecular immunology*, 2012. 51(3): p. 304-309.

[165] Richer, J., S.A. Johnston, and P. Stafford, Epitope identification from fixed-complexity random-sequence peptide microarrays. *Molecular & Cellular Proteomics*, 2015. 14(1): p. 136-147.

[166] Cello, J., A. Samuelson, P. Stålhandske, B. Svennerholm, et al., Identification of group-common linear epitopes in structural and nonstructural proteins of enteroviruses by using synthetic peptides. *Journal of clinical microbiology*, 1993. 31(4): p. 911-916.

[167] Breitling, F., A. Nesterov, V. Stadler, T. Felgenhauer, et al., High-density peptide arrays. *Molecular BioSystems*, 2009. 5(3): p. 224-234.

[168] Schirwitz, C., F.F. Loeffler, T. Felgenhauer, V. Stadler, et al., Sensing immune responses with customized peptide microarrays. *Biointerphases*, 2012. 7(1): p. 47.

[169] Breidenbach, M.A. and A.T. Brunger, 2.3 Å Crystal Structure of Tetanus Neurotoxin Light Chain†. *Biochemistry*, 2005. 44(20): p. 7450-7457.

[170] O'Shannessy, D.J. and D.J. Winzor, Interpretation of deviations from pseudo-first-order kinetic behavior in the characterization of ligand binding by biosensor technology. *Analytical biochemistry*, 1996. 236(2): p. 275-283.

A Supporting Information

A.1 Supporting Information for: Identification of a tetanus toxin specific epitope in single amino acid resolution

Anti-tetanus titer

According to the WHO standard, tetanus antitoxin levels are expressed in international units (IU) and are classified according to international standards for standard ELISA: <0.01 IU/ml, no protection; 0.01–0.1 IU/ml, uncertain, short-term protection and ≥0.1 IU/ml, full protection.

.

Table S1: IgG anti-tetanus antibody levels

Serum	IU/ml Tetanus
1	4.07
2	5.88
3	1.76
4	2.53
5	2.02
6	0.80
7	1.59
8	7.88
9	2.85
10	0.75
11	1.02
12	0.23
13	2.21
14	1.99
15	0.93
16	0.95
17	4.80
18	3.34
19	0.75

Substitution analysis
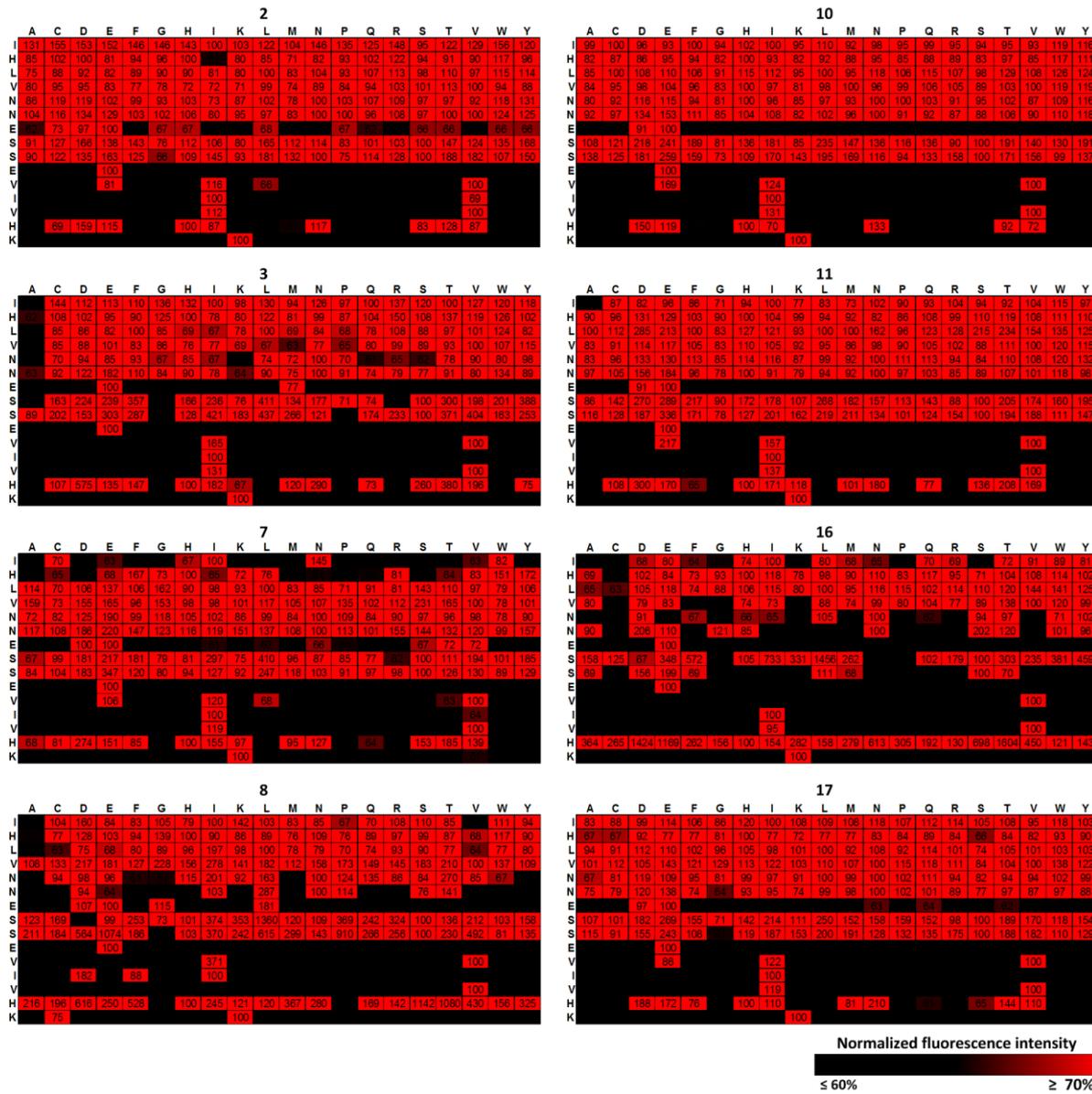
Figure S1: Heatmaps of the evaluation of the substitution analysis for 8 different sera. From top to bottom left, the original sequence is shown. Substituted amino acids are shown on top of each heatmap. The intensity of the peptide with the original amino acid in each row is set to 100 %. The relative fluorescence intensities are calculated in relation to these intensities.
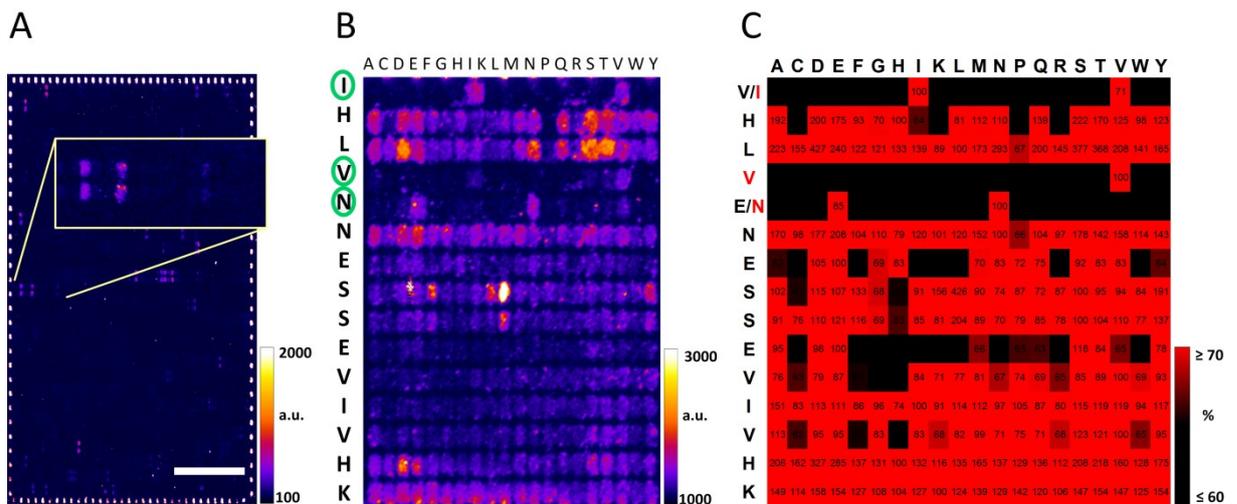
Figure S2: (A) Fluorescence scan image of the epitope mapping of serum 9. The peptide chosen for substitution analysis corresponds to the second visible spot in the enlarged area. Scale bar 0.5 cm. (B) Fluorescence scan image of the corresponding substitution analysis of a 15mer peptide. Each amino acid position is substituted by all 19 other amino acids resulting in an array of 15 rows and 20 columns. From top to bottom left, the original sequence is shown. (C) Heatmap of the evaluation of the substitution analysis. The intensity of the peptide with the original amino acid in each row is set to 100 %. The relative fluorescence intensities are calculated in relation to these intensities. From top to bottom left, the original sequence is shown; letters in red indicate original amino acids that are essential for binding. In some positions two or more amino acids allow for the binding, alternative amino acids are depicted left of the original amino acid if the intensity is lower and right of the original amino acid if the intensity is higher.

Validation of antibody-peptide interaction
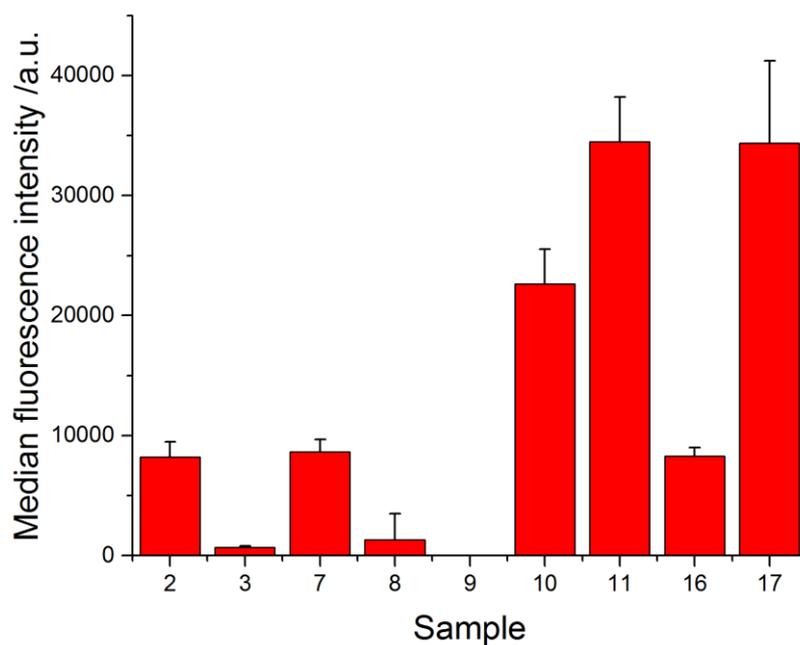
Figure S3: Results of validation experiments to prove that the antibody-peptide interaction is independent from the individual surfaces. Sera of 8 individuals that showed interaction with the most prominent epitope identified in the epitope mapping were incubated on arrays containing peptides with the sequence IHLVNNESSEVIVHK. The peptide was pre-synthesized, purified and spotted onto functionalized glass slides. For each serum, 36 peptide spots were incubated and the median fluorescence intensity was calculated.
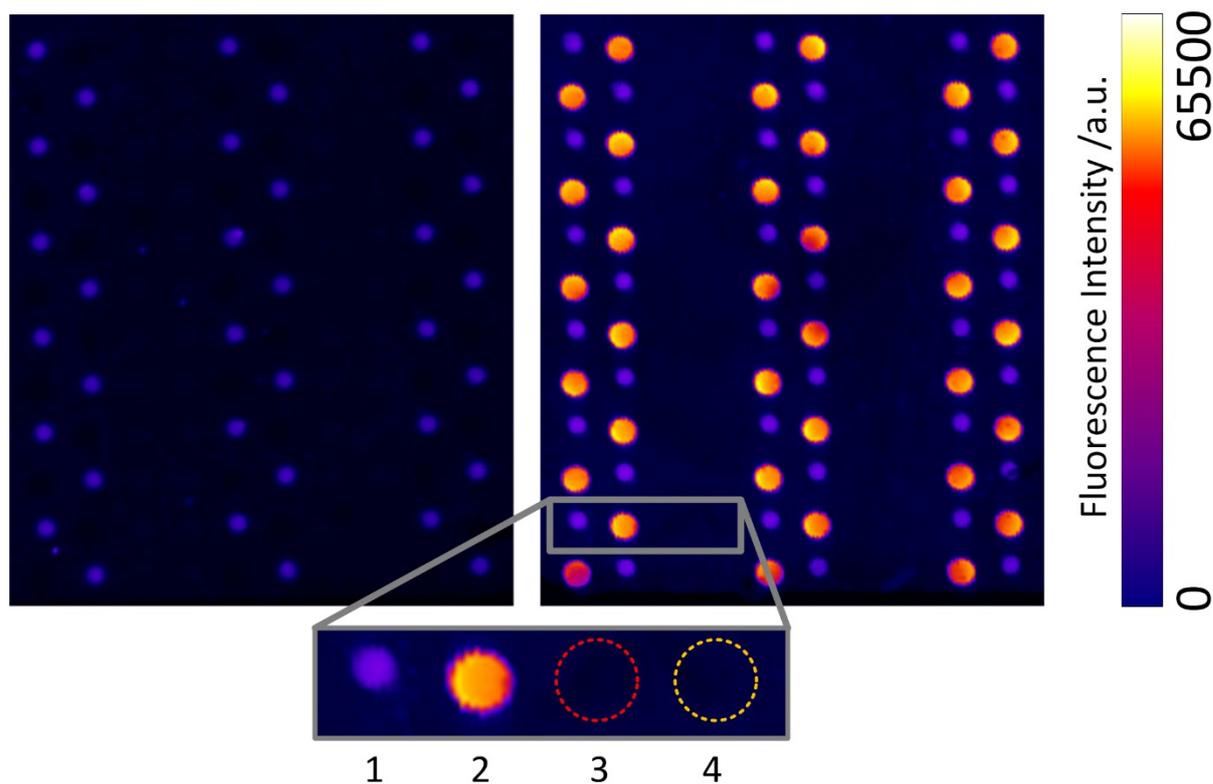
Figure S4: Exemplary fluorescence scans of validation of antibody peptide interaction with 4 different peptides: 1 (MVPEFSGSFPMR): positive control, the peptide sequence was identified in a separate study, 2 (IHLVNNESSEVIVHK): antigenic peptide used for isolation of antibody species, 3 (IHLVNNESSEVIVHR): negative control 1 and 4 (IHLVNNESSEVIAHK): negative control 2. Both negative controls were identified in the substitution analysis to prevent the binding of the corresponding antibody species. Arrays consisted of 36 spots per peptide; peptides were pre-synthesized and spotted onto functionalized surfaces. The array on the left was incubated with serum 19 as negative control; the array on the right was incubated with serum 11.

Isolation of epitope specific antibodies
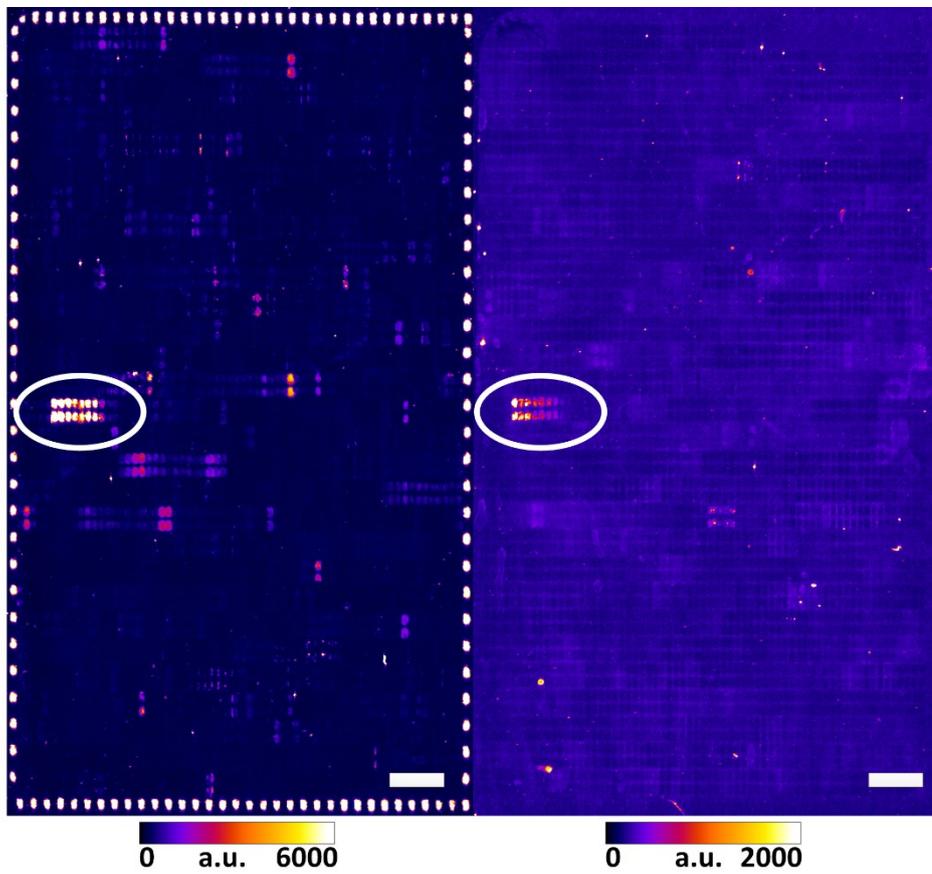
Figure S5: Fluorescence scan images of the epitope mapping of serum 11 before and upon isolation of epitope-specific antibodies. Scale bar 0.5 cm. The white ellipse highlights the respective epitope including the antigenic peptide used for isolation of the antibody species.

Bioinformatic analysis of identified fingerprint

Table S2: Hits of a protein database query for the identified fingerprint. The table shows the protein on which the motif is located, the organism in which the protein was found, the amino acid position and the exact sequence. The query was conducted with the following settings: include splice variants (Swiss-Prot); exclude fragments; Hits for motif on all UniProtKB/Swiss-Prot (release 2016_11 of 30-Nov-16: 553231 entries) database sequences found 47 hits in 47 sequences. All hits within the same protein of the same organism are shown in the same row. Hits in different versions of the same protein were deleted.

Organism	Sequence	Query result	Protein	Aa position
Aeropyrum pernix	EtgEIIISK	>sp|Q9YC05|RRP42_AERPE (276 aa)	Exosome complex component	177 - 185
Alteromonas mediterranea	DvlEVIVNK	>sp|B4RVA8|LEPA_ALTMD (598 aa)	Elongation factor 4 (EF-4)	173 - 181
Archaeoglobus fulgidus	EhdEVIVEK	>sp|O28387|RS12_ARCFU (142 aa)	30S ribosomal protein S12	94 - 102
Caenorhabditis briggsae	EndEIIVEK	>sp|Q60MF5|RN207_CAEBR (836 aa)	Probable RING finger protein 207 homolog	745 - 753
Caenorhabditis elegans	EkeEIIIVK	>sp|O44199|RAD50_CAEEL (1298 aa)	DNA repair protein rad-50	708 - 716
Caenorhabditis elegans	EenEVIIEK	>sp|Q09353|SENP_CAEEL (697 aa)	Sentrin-specific protease	114 - 122
Caenorhabditis elegans	EirEIIVSK	>sp|Q09372|YS44_CAEEL (655 aa)	Uncharacterized protein ZK177.4	64 - 72
Clostridium acetobutylicum	EtgEIIVEK	>sp|Q97EG9|RPOB_CLOAB (1241 aa)	DNA-directed RNA polymerase subunit beta	298 - 306
Clostridium tetani	EssEVIVHK	>sp|P04958|TETX_CLOTE (1315 aa)	Tetanus toxin	929 - 937
Deinococcus geothermalis	EvkEVIVDK	>sp|Q1J195|ACP_DEIGD (76 aa)	Acyl carrier protein (ACP)	6 – 14
Dictyostelium discoideum	EkiEIIVDK	>sp|Q54NW7|VAM7A_DICDI (216 aa)	Vesicle-associated membrane protein 7A	153 - 161
Emericella nidulans	EhdEIIVEK	>sp|Q5BB57|RRF2M_EMENI (921 aa)	Ribosome-releasing factor 2, mitochondrial	299 - 307
Enterococcus faecium	EtgEIIVEK	>sp|Q8GCR6|RPOB1_ENTFC (1208 aa)	DNA-directed RNA polymerase subunit beta	306 - 314
Finegoldia magna	DtgEIIVHK	>sp|B0S2E5|RPOC_FINM2 (1202 aa)	DNA-directed RNA polymerase subunit beta	831 - 839
Flavobacterium psychrophilum	DtcEVIVEK	>sp|A6GX63|SECA_FLAPJ (1116 aa)	Protein translocase subunit SecA	833 - 841
Homo sapiens	EekEIIVIK	>sp|Q5JQC9|AKAP4_HUMAN (854 aa)	A-kinase anchor protein 4 (AKAP-4)	66 – 74 75 - 83
Homo sapiens	EseEVIIEK	>sp|A4UGR9|XIRP2_HUMAN (3374 aa)	Xin actin-binding repeat-containing protein 2	647 - 655
Human rhinovirus 14	EleEVIVEK	>sp|P03303|POLG_HRV14 (2179 aa)	Genome polyprotein	572 - 580
Human rhinovirus 3	EleEVIVEK	>sp|Q82081|POLG_HRV3 (2178 aa)	Genome polyprotein	572 - 580
Ignicoccus hospitalis	EgvEVIISK	>sp|A8AA20|RS3_IGNH4 (241 aa)	30S ribosomal protein S3	181 - 189
Lactobacillus salivarius	DtgEIIVNK	>sp|Q1WVA5|RPOB_LACS1 (1199 aa)	DNA-directed RNA polymerase subunit beta	307 - 315
Lecanicillium sp.	DleEVIINK	>sp|A0A024F910|VLMS_LECSP (8903 aa)	Nonribosomal peptide synthetase vlms	1760 - 1768
Methanocaldococcus jannaschii	DrdEIIINK	>sp|Q57958|Y538_METJA (260 aa)	Uncharacterized protein MJ0538	176 - 184
Mus musculus	EekEIIVIK	>sp|Q60662|AKAP4_MOUSE (849 aa)	A-kinase anchor protein 4 (AKAP-4)	66 – 74 75 - 83
Mus musculus	EgtEIIITK	>sp|Q8C0Q2|ZHX3_MOUSE (951 aa)	Zinc fingers and homeoboxes protein 3	171 - 179

Mycoplasma capricolum subsp. capricolum	DqaEIIIDK	>sp|Q2SSW4|SYA_MYCCT (896 aa)	Alanine--tRNA ligase	728 - 736
Neosartorya fumigata	DnaEVIVEK	>sp|Q4WAW3|FTMA_ASPFU (2211 aa)	Nonribosomal peptide synthetase 13	477 - 485
Oceanobacillus iheyensis	EtgEVIVEK	>sp|Q8ETY7|RPOC_OCEIH (1203 aa)	DNA-directed RNA polymerase subunit beta	856 - 864
Petrotoga mobilis	EfkEIIIDK	>sp|A9BF51|MOAA_PETMO (323 aa)	GTP 3',8-cyclase	207 - 215
Physcomitrella patens subsp. patens	DemEVIIIK	>sp|A9SVH7|RBR_PHYPA (1116 aa)	Retinoblastoma-related protein	1013 - 1021
Rattus norvegicus	EekEIIVIK	>sp|O35774|AKAP4_RAT (847 aa)	A-kinase anchor protein 4 (AKAP-4)	74 - 82
Rattus norvegicus	EgtEIIITK	>sp|Q80Z36|ZHX3_RAT (951 aa)	Zinc fingers and homeoboxes protein 3	170 - 178
Saccharomyces cerevisiae	DteEVIVIK	>sp|Q08951|AP3D_YEAST (932 aa)	Adaptor-related protein complex 3 subunit delta	889 - 897
Schizosaccharomyces pombe	EqnEIIITK	>sp|Q7Z992|SST6_SCHPO (487 aa)	ESCRT-I complex subunit vps23	260 - 268
Xenopus laevis	DsrEIIVEK	>sp|Q9PTD7|CING_XENLA (1360 aa)	Cingulin	304 - 312
Yarrowia lipolytica	EvfEVIITK	>sp|P41925|RYL2_YARLI (209 aa)	Ras-like GTP-binding protein RYL2	164 - 172

A.2 Supporting Information for: Antibody fingerprints in Lyme disease deciphered with high density peptide arrays

Table S1: Maximum fluorescence intensities of the epitope mappings

Serum	Maximum fluorescence intensity a.u.
1	65284
2	8337
3	1328
4	5912
5	1462
6	10135
7	14885
8	664
9	3122
10	47726
11	26523
12	65300
13	37291
14	24483
15	1750
16	394
17	395
NC 1	134
NC 2	409
NC 3	583
NC 4	130
NC 5	277
NC 6	590
NC 7	678

The following Supplementary Figures S1 – S5 show the antibody fingerprints of all investigated individuals, deciphered with substitution analyses on high density peptide arrays for immunodominant epitopes of the VlsE antigen.
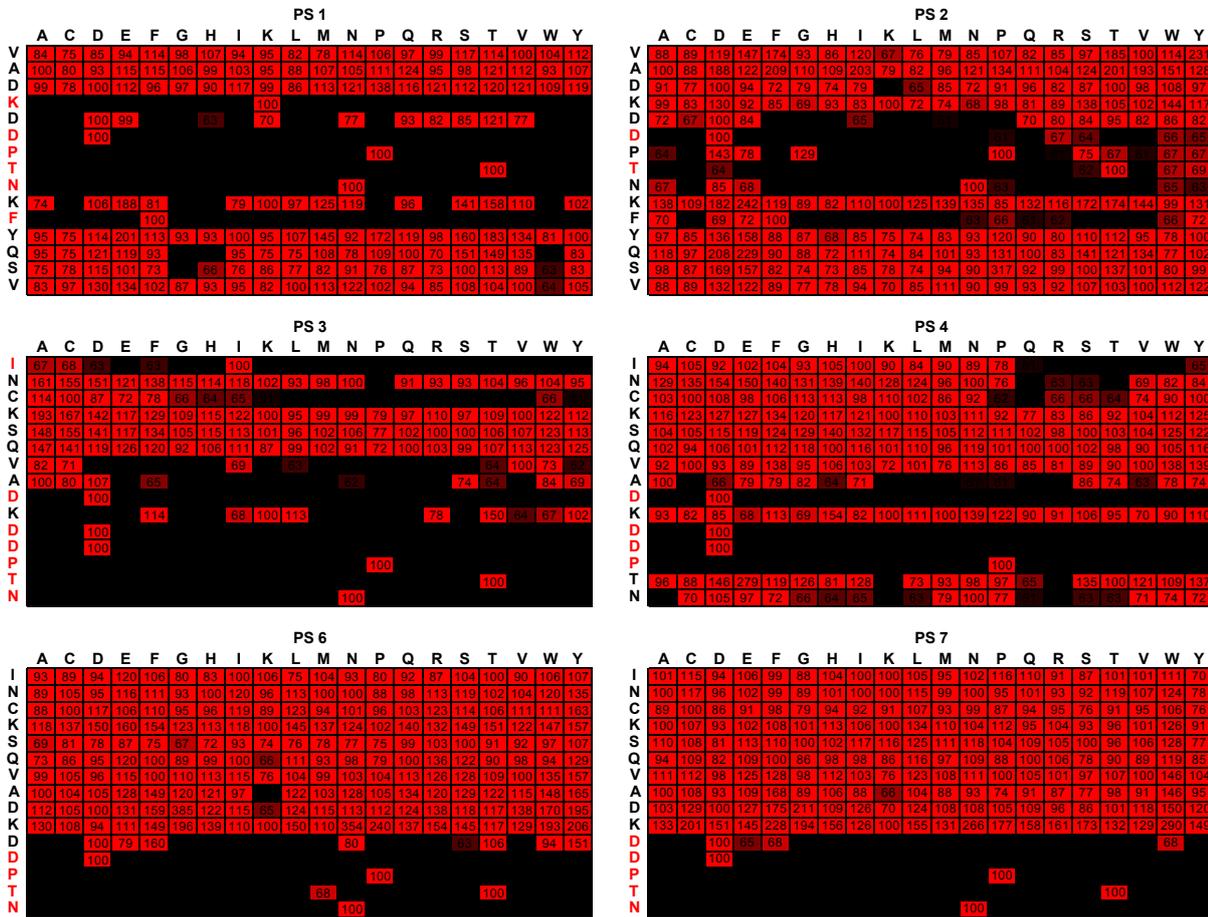
Figure S1: Antibody fingerprints of 6 sera for the epitope at position 20 – 40 on the VlsE antigen. The original amino acids of the peptide sequence were one by one substituted by all other amino acids. The resulting peptides were printed as array duplicates (not shown) and incubated with the corresponding serum. The results are visualized as a heat map. The relative fluorescence intensities of all peptides in a row are calculated according to the intensity of the peptide with the original amino acid set to 100 %. The original sequence is shown from top to bottom left; essential amino acids are indicated in red. Substituted amino acids are shown from left to right on top of the substitution arrays.

Figure S2: Antibody fingerprints of 3 sera for the epitope at position 171 – 186 on the VlsE antigen. The original amino acids of the peptide sequence were one by one substituted by all other amino acids. The resulting peptides were printed as array duplicates (not shown) and incubated with the corresponding serum. The results are visualized as a heat map. The relative fluorescence intensities of all peptides in a row are calculated according to the intensity of the peptide with the original amino acid set to 100 %. The original sequence is shown from top to bottom left; essential amino acids are indicated in red. Substituted amino acids are shown from left to right on top of the substitution arrays.
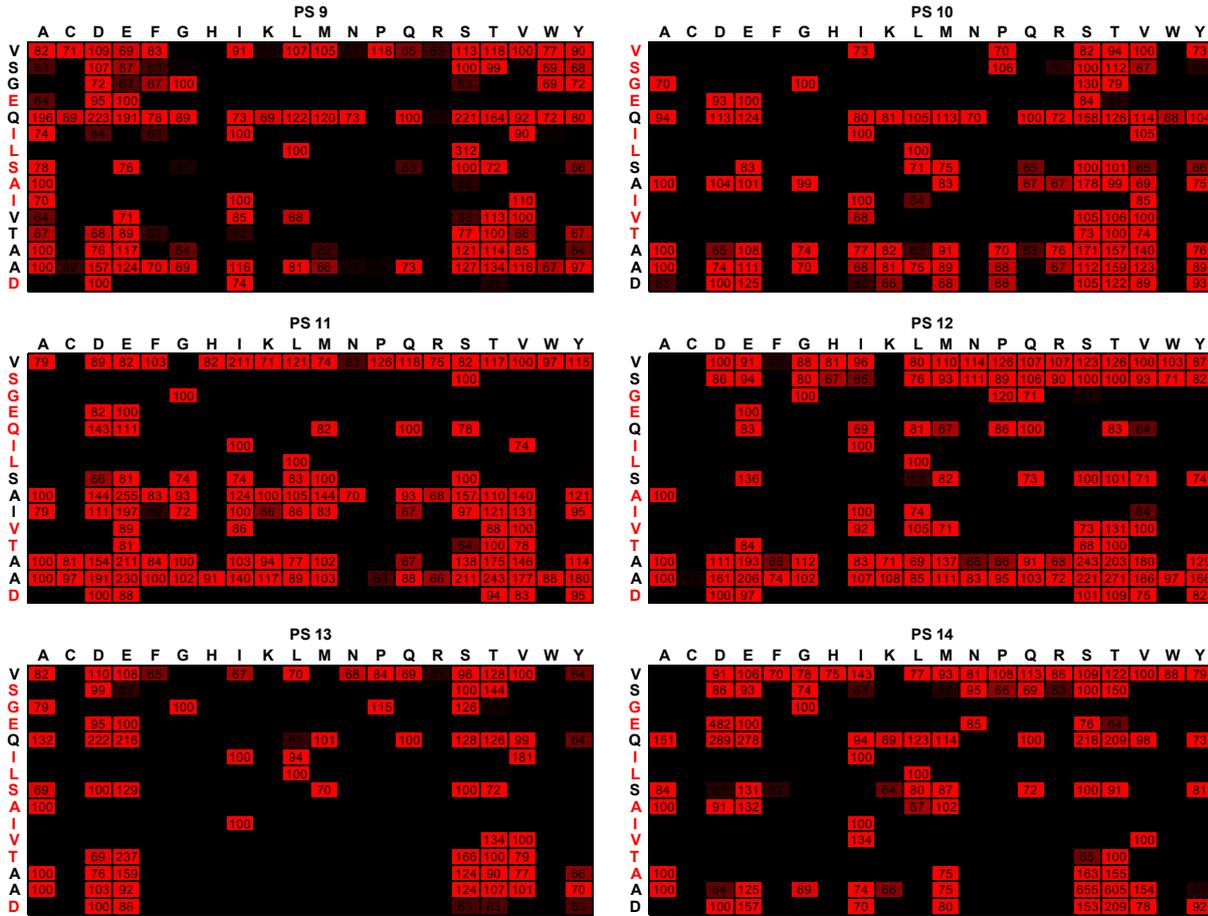
Figure S3: Antibody fingerprints of 6 sera for the epitope at position 228 – 242 on the VlsE antigen. The original amino acids of the peptide sequence were one by one substituted by all other amino acids. The resulting peptides were printed as array duplicates (not shown) and incubated with the corresponding serum. The results are visualized as a heat map. The relative fluorescence intensities of all peptides in a row are calculated according to the intensity of the peptide with the original amino acid set to 100 %. The original sequence is shown from top to bottom left; essential amino acids are indicated in red. Substituted amino acids are shown from left to right on top of the substitution arrays.

Figure S4: Antibody fingerprints of 5 sera for the epitope at position 249 – 267 on the VlsE antigen. The original amino acids of the peptide sequence were one by one substituted by all other amino acids. The resulting peptides were printed as array duplicates (not shown) and incubated with the corresponding serum. The results are visualized as a heat map. The relative fluorescence intensities of all peptides in a row are calculated according to the intensity of the peptide with the original amino acid set to 100 %. The original sequence is shown from top to bottom left; essential amino acids are indicated in red. Substituted amino acids are shown from left to right on top of the substitution arrays.

Figure S5: Antibody fingerprints of 7 sera for the epitope at position 338 – 357 on the VlsE antigen. The original amino acids of the peptide sequence were one by one substituted by all other amino acids. The resulting peptides were printed as array duplicates (not shown) and incubated with the corresponding serum. The results are visualized as a heat map. The relative fluorescence intensities of all peptides in a row are calculated according to the intensity of the peptide with the original amino acid set to 100 %. The original sequence is shown from top to bottom left; essential amino acids are indicated in red. Substituted amino acids are shown from left to right on top of the substitution arrays.

Supplementary Tables S2 – S6 list all potentially cross-reacting human proteins that have been identified with queries of the antibody fingerprints in protein databases.

Table S2: Results of the query of all human proteins of the "UniProtKb/TrEMBL" and "UniProtKb/Swiss-Prot" protein databases with the "ScanProsite tool" for the epitope VlsE 20 – 40 and the fingerprint D-P-T-N. The table lists the proteins on which the motif is located, the amino acid position and the exact sequence. The query included splice variants and fragments. Similar entries, corresponding to the same protein, were deleted. The proteins printed in bold are linked to autoimmune disorders.

Protein	Database entry	Aa sequence	Position
2'-5'-oligoadenylate synthetase 2	>sp|P29728|OAS2_HUMAN	DPTN	296-299
Activating signal cointegrator 1 complex subunit3	>sp|Q8N3C0|ASCC3_HUMAN	DPTN	1226-1229
ADAM10 protein	>tr|A0AV88|A0AV88_HUMAN	DPTN	285-288
ADAM metallopeptidase domain 10	>tr|A0A024R5U5|A0A024R5U5_HUMAN	DPTN	285-288
alpha-1,2- Mannosidase	>tr|B4DEE1|B4DEE1_HUMAN	DPTN	221-224
Alternative protein CFP,	>tr|L0R836|L0R836_HUMAN	DPTN	105-108
ATPase family AAA domain-containing protein 5, Chromosome fragility-associated gene 1 protein	>sp|Q96QE3|ATAD5_HUMAN	DPTN	1740-1743
Attractin, DPPT-L	>sp|O75882|ATRN_HUMAN	DPTN	1099-1102
Attractin-like protein 1	>sp|Q5VV63|ATRN1_HUMAN	DPTN	1050-1053
Bardet-Biedl Syndrome 4 isoform 1, Bardet-Biedl Syndrome 4	>tr|A0A0S2Z3A9|A0A0S2Z3A9_HUMAN	DPTN	232-235
Coiled-coil domain-containing protein 60	>tr|F5H2N8|F5H2N8_HUMAN	DPTN	94-97
Constitutive coactivator of peroxisome proliferator-activated receptor gamma, Constitutive coactivator of PPAR-gamma	>sp|Q96EK7|F120B_HUMAN	DPTN	556-559
DEAD box polypeptide 5, ets variant protein 4 fusion protein	>tr|C1IK54|C1IK54_HUMAN	DPTN	388-391
Death-associated protein 6	>tr|A0A024RCS3|A0A024RCS3_HUMAN	DPTN	167-170
Death-associated protein 6 variant	>tr|Q53F85|Q53F85_HUMAN	DPTN	167-170
Disintegrin and metalloproteinase domain-containing protein 10, ADAM10	>sp|O14672|ADA10_HUMAN	DPTN	285-288
DNA polymerase delta subunit 2, DNA polymerase delta subunit p50	>sp|P49005|DPOD2_HUMAN	DPTN	300-303
Dynein heavy chain 8	>sp|Q96JB1|DYH8_HUMAN	DPTN	3919-3922
Epididymis secretory sperm binding protein Li94n, Stress-induced-phosphoprotein 1(Hsp70/Hsp90-organizing protein	>tr|V9HW72|V9HW72_HUMAN	DPTN	255-258
ER degradation-enhancing alpha-mannosidase-like protein 2	>sp|Q9BV94|EDEM2_HUMAN	DPTN	442-445
ER membrane protein complex subunit 2, Tetratricopeptide repeat protein 35	>sp|Q15006|EMC2_HUMAN	DPTN	117-120
ETS translocation variant 4, Adenovirus E1A enhancer-binding protein	>sp|P43268|ETV4_HUMAN	DPTN	353-356
Ets variant gene 4, E1A enhancer binding protein	>tr|D3DX44|D3DX44_HUMAN	DPTN	179-182
EWS protein, E1A enhancer binding protein chimera	>tr|A2NIV0|A2NIV0_HUMAN	DPTN	347-350
Glycoprotein receptor gp330, Megalin, LRP-2	>tr|Q7Z5C0|Q7Z5C0_HUMAN	DPTN	1016-1019
GPRIN family member 3	>tr|A0AVI5|A0AVI5_HUMAN	DPTN	528-531
G protein-regulated inducer of neurite outgrowth 3, GRIN3	>sp|Q6ZVF9|GRIN3_HUMAN	DPTN	528-531
Homeobox protein DLX-3	>sp|O60479|DLX3_HUMAN	DPTN	246-249
Intraflagellar transport protein 74 homolog, Capillary morphogenesis gene 1 protein	>sp|Q96LB3-2|IFT74_HUMAN	DPTN	352-355
KIAA0103 variant (ER membrane protein complex subunit 2)	>tr|Q53HG5|Q53HG5_HUMAN	DPTN	117-120
Kin of IRRE-like protein 1, Kin of irregular chiasm-like	>sp|Q96J84|KIRR1_HUMAN	DPTN	600-603

protein 1			
Kin of IRRE-like protein 2, Kin of irregular chiasm-like protein 2	>sp|Q6UWL6|KIRR2_HUMAN	DPTN	598-601
Kin of IRRE-like protein 3, Kin of irregular chiasm-like protein 3	>sp|Q8IZU9|KIRR3_HUMAN	DPTN	632-635
Laforin	>sp|B3EWF7|EP2A2_HUMAN	DPTN	327-330
Mitogen-activated protein kinase-binding protein 1, JNK-binding protein 1	>sp|O60336|MABP1_HUMAN	DPTN	353-356
MSSP-2	>tr|Q5CZ65|Q5CZ65_HUMAN	DPTN	5-8
NCF4 protein	>tr|Q6FGM9|Q6FGM9_HUMAN	DPTN	235-238
Neutrophil cytosol factor 4, NCF-4	>sp|Q15080|NCF4_HUMAN	DPTN	235-238
Non-receptor tyrosine-protein kinase TYK2, EC2.7.10.2	>sp|P29597|TYK2_HUMAN	DPTN	917-920
Peptidyl-prolyl cis-trans isomerase-like 2, PPIase	>sp|Q13356|PPIL2_HUMAN	DPTN	160-163
Peptidylprolyl isomerase (Cyclophilin)-like 2	>tr|A0A024R1C1|A0A024R1C1_HUMAN	DPTN	160-163
PGAP2-interacting protein, Cell wall biogenesis protein 43 C-terminal homolog	>sp|Q9H720|PG2IP_HUMAN	DPTN	648-651
Phosphatidylcholine:ceramide choline phosphotransferase 2, EC2.7.8.27	>sp|Q8NHU3|SMS2_HUMAN	DPTN	19-22
Probable maltase-glucoamylase 2, Maltase-glucoamylase (alpha-glucosidase)	>sp|Q2M2H8|MGAL_HUMAN	DPTN	1012-1015
Proprotein convertase subtilisin/kexin type 1 (NEC1)	>tr|A1L3V6|A1L3V6_HUMAN	DPTN	201-204
ProteinFAM214A,	>sp|Q32MH5|F214A_HUMAN	DPTN	751-754
ProteinkinaseC-bindingprotein1, CutaneousT-celllymphoma-associatedantigense14-3	>sp|Q9ULU4|PKCB1_HUMAN	DPTN	380-383
Proteinkinase C-binding protein 1	>tr|A0A087WV57|A0A087WV57_HUMAN	DPTN	355-358
Putative stereocilin-like protein, Stereocilin Pseudogene 1	>sp|A6NGW2|STRCL_HUMAN	DPTN	199-202
RNA-binding motif, single-stranded-interacting protein 1	>tr|A0A0S2Z499|A0A0S2Z499_HUMAN	DPTN	139-142
RNA-binding motif, single-stranded-interacting protein 2	>tr|A0A0S2Z4B3|A0A0S2Z4B3_HUMAN	DPTN	138-141
SAM domain-containing protein SAMSN-1 (Hematopoietic adaptor containing SH3 and SAM domains 1)	>sp|Q9NSI8|SAMN1_HUMAN	DPTN	48-51
SET-binding protein SEB	>sp|Q9Y6X0|SETBP_HUMAN	DPTN	412-415
Sideroflexin-4, Breastcancer resistance marker 1	>sp|Q6P4A7|SFXN4_HUMAN	DPTN	51-54
Solute Carrier Family 12 Member 1, Bumetanide-sensitivesodium-potassium	>sp|Q13621|S12A1_HUMAN	DPTN	296-299
Spectrin alpha chain, erythrocytic 1	>sp|P02549|SPTA1_HUMAN	DPTN	83-86
Stereocilin	>sp|Q7RTU9|STRC_HUMAN	DPTN	199-202
STRC protein,	>tr|A1L379|A1L379_HUMAN	DPTN	199-202
Stress-induced-phosphoprotein1, STI1	>sp|P31948|STIP1_HUMAN	DPTN	255-258
Tubby-related protein 1, Tubby-likeprotein1	>sp|O00294|TULP1_HUMAN	DPTN	355-358
Tyrosine-protein kinase	>tr|A0A024R7E4|A0A024R7E4_HUMAN	DPTN	917-920
Ubiquitin Carboxyl-terminal hydrolase 4	>sp|Q13107|UBP4_HUMAN	DPTN	147-150
Ubiquitin conjugation factor E4A	>sp|Q14139|UBE4A_HUMAN	DPTN	318-321
Ubiquitin Specific Peptidase 4, Proto-oncogene	>tr|A0A024R2T0|A0A024R2T0_HUMAN	DPTN	147-150
Ubiquitin Specific Protease	>tr|Q53H56|Q53H56_HUMAN	DPTN	147-150
ZMYND8 protein	>tr|A6H8Y8|A6H8Y8_HUMAN	DPTN	375-378

Table S3: Results of the query of all human proteins of the "UniProtKb/TrEMBL" and "UniProtKb/Swiss-Prot" protein databases with the "ScanProsite tool" for the epitope VlsE 171–186 and the fingerprint I-x-(ADE)-I-x-(ADE)-A-A. The table lists proteins on which the motif is located, the amino acid position and the exact sequence. The query included splice variants and fragments. Similar entries, corresponding to the same protein, were deleted.

Protein	Database entry	Aa sequence	Position
Calpain small subunit 1	>sp|P04632|CPNS1_HUMAN	IsAIsEAA	66-73
Erlin-1, Endoplasmic reticulum lipid raft-associated protein 1	>sp|O75477|ERLN1_HUMAN	IsEIeDAA	242-249
Erlin-2, Endoplasmic reticulum lipid raft-associated protein 2	>sp|O94905|ERLN2_HUMAN	IsEIeDAA	242-249
Helicase SRCAP	>sp|Q6ZRS2|SRCAP_HUMAN	ItDIaAAA	584-591
Pentatricopeptide repeat domain-containing protein 3, mitochondrial, 28S ribosomal protein S39, mitochondrial	>sp|Q96EY7|PTCD3_HUMAN	IkDIsEAA	149-156
SPFH domain family, member 1, isoform CRA_a.	>tr|D3DR65|D3DR65_HUMAN	IsEIeDAA	244-251

Table S4: Results of the query of all human proteins of the "UniProtKb/TrEMBL" and "UniProtKb/Swiss-Prot" protein databases with the "ScanProsite tool" for the epitope VlsE 249-267 and the fingerprint N-P-I-(SAMQE)-x-A-(TLI). The table lists proteins on which the motif is located, the amino acid position and the exact sequence. The query included splice variants and fragments. Similar entries, corresponding to the same protein, were deleted. The proteins printed in bold are linked to autoimmune disorders.

Protein	Database entry	Aa sequence	Position
Carboxypeptidase E , CPE	>sp|P16870|CBPE_HUMAN	NPIAnAT	386-392
Dedicator of cytokinesis protein 10	>tr|H0YFC5|H0YFC5_HUMAN	NPIEvAI	177-183
Dedicator of cytokinesis protein 9	>tr|H0Y3S1|H0Y3S1_HUMAN	NPIEvAI	508-514
DOCK9 protein	>tr|Q7Z6H4|Q7Z6H4_HUMAN	NPIEvAI	415-421
Dopamine receptor interacting protein 2	>tr|Q4W4X9|Q4W4X9_HUMAN	NPIEvAI	532-538
PRAME family member 20	>sp|Q5VT98|PRA20_HUMAN	NPIStAT	387-393
PRAME family member 25	>sp|A6NGN4|PRA25_HUMAN	NPISmAT	390-396
PRAME family member 27	>sp|A3QJZ7|PRA27_HUMAN	NPISmAT	390-396
PRAME family member 5	>sp|Q5TYX0|PRAM5_HUMAN	NPISmAT	388-394
PRAME family member 6	>sp|Q5VXH4|PRAM6_HUMAN	NPISmAT	388-394
PRAME family member 9/15	>tr|A0A096LNW4|A0A096LNW4_HUMAN	NPISmAT	390-396
Putative PRAME family member 26	>sp|H0Y7S4|PRA26_HUMAN	NPISmAT	294-300
Transmembrane protein 131 , Protein RW1	>sp|Q92545|TM131_HUMAN	NPIElAI	573-579

Table S5: Results of the query of all human proteins of the "UniProtKb/TrEMBL" and "UniProtKb/Swiss-Prot" protein databases with the "ScanProsite tool" for the epitope VlsE 338-357 Part 1 and the fingerprint G-(ED)-(SGA)-x-K-A-A-x-K. The table lists proteins on which the motif is located, the amino acid position and the exact sequence. The query included splice variants and fragments. Similar entries, corresponding to the same protein, were deleted.

Protein	Database entry	Aa sequence	Position
GON-4-like protein	>sp|Q3T8J9|GON4L_HUMAN	GESiKAAgK	1576-1584
YY1AP-related protein1	>tr|A4PB67|A4PB67_HUMAN	GESiKAAgK	1576-1584

Table S6: Results of the query of all human proteins of the "UniProtKb/TrEMBL" and "UniProtKb/Swiss-Prot" protein databases with the "ScanProsite tool" for the epitope VlsE 338-357 Part 2 and the fingerprint K-E-x-P-(ASP)-x-L-N. The table lists proteins on which the motif is located, the amino acid position and the exact sequence. The query included splice variants and fragments. Similar entries, corresponding to the same protein, were deleted.

Protein	Database entry	Aa sequence	Position
Ataxin-1-like (Brother of ataxin-1)	>sp|P0C7T5|ATX1L_HUMAN	KEePSpLN	357-364
Structure-specific endonuclease subunit SLX4 (BTB/POZ domain-containing protein 12)	>sp|Q8IY92|SLX4_HUMAN	KEaPPgLN	1684-1691

A.3 Supporting Information for: Single amino acid fingerprinting of the human antibody repertoire with high density peptide arrays

Validation of biological significance

To validate that the fingerprint analysis can be performed with other patient samples, and also reveal different peptide binders, we have analyzed the serum of a patient in a similar approach. Results are shown in Supplementary Figures S1 and S2 and Supplementary Table S1.
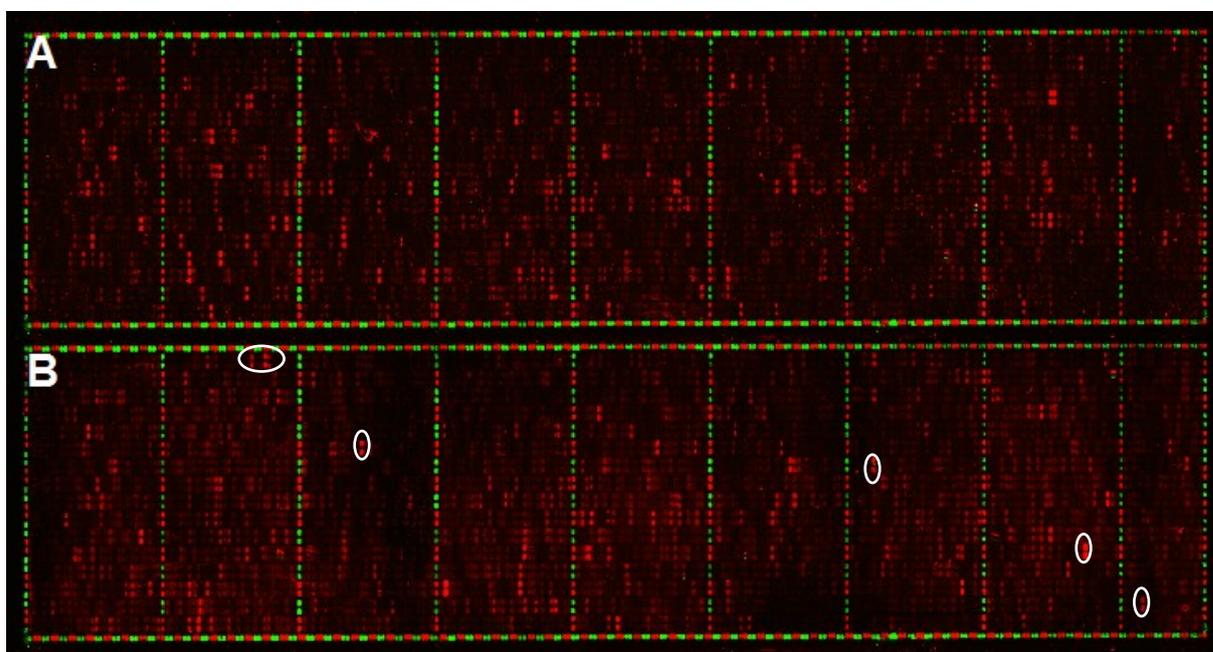

Figure S1: Peptide array with 4,128 different random peptides. (A) Pre-staining with the European control serum, and (B) subsequent staining with a patient serum ID 102. Some peptide binders that only show interactions with the patient serum are highlighted in an exemplary way.
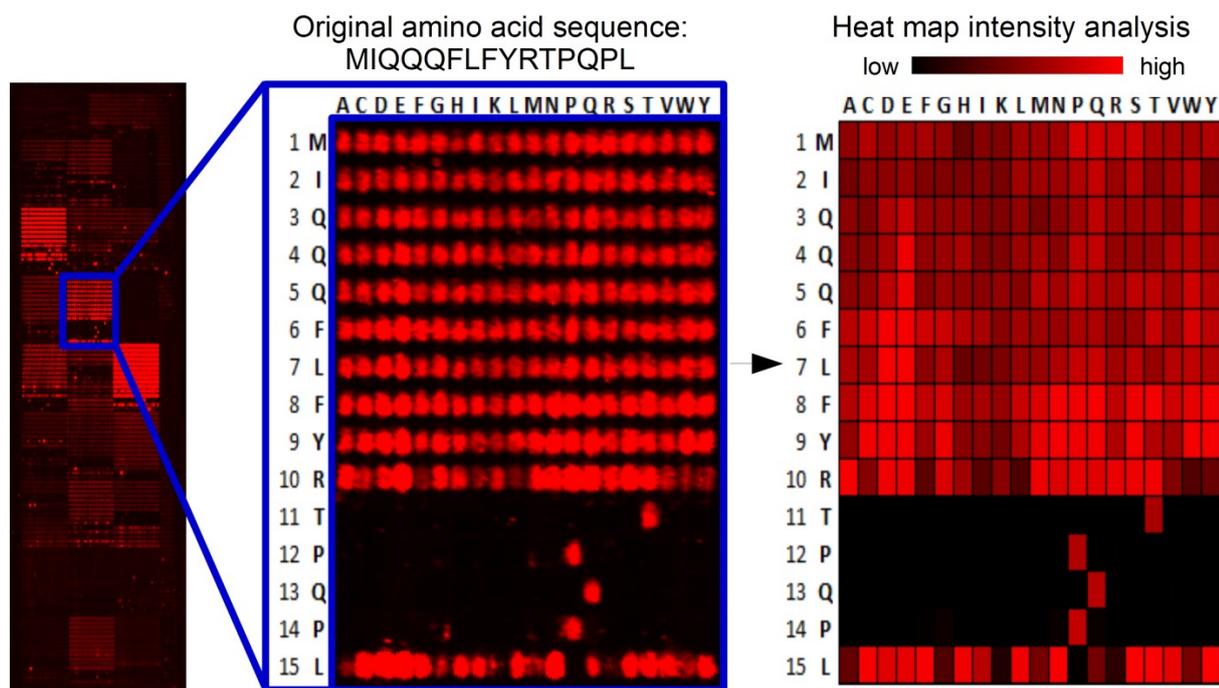
Original amino acid sequence: MIQQQFLFYRTPQPL

Heat map intensity analysis

Figure S2: Amino acid substitution analysis of the top 27 peptide binders derived from the random peptide array staining (Figure S1). Each of the 27 peptides was synthesized in a subarray of 300 different substitution variants (15 x 20 spots) on one big array (left) of 8,100 peptide spots (27 x 300 spots). The array was stained with the patient serum. A magnified example of a subarray (center) and the corresponding heat map of the staining intensities (right) is shown. Each row in these fields represents the substitution analysis in one position of the sequence. On the left of each row, the original amino acid and its position in the peptide are shown. Each column is headed by an amino acid letter, which substitutes the current amino acid (row) position. Thus, in each spot, a single amino acid is substituted; the rest of the sequence is conserved.

Table S1: Identified peptide binders for the patient serum. 22 of 27 substituted peptide binders resulted in fingerprints. Original amino acid sequences and essential amino acids at each position of the respective peptide are depicted. Essential amino acids of the original peptides are depicted in red.

Original peptide sequence	Amino acid position in epitope														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
DLMLAEMHMGEVMSN	X	X	X	X	X	X	X	X	X	F,G	E	X	X	X	X
ENMWKPTLNTTTEPV	X	X	X	X	X	X	X	X	X	X	X	X	T,E	P	X
FRTGEYEEDLDKLGG	X	X	X	G	E	X	I,M,E,G	I,D,L,E	L,F,E,H,D	X	X	X	X	X	X
FWELPPSPETDQKRW	X	X	X	X	X	X	X	X	X	X	D	X	K	X	C,W,Y,L,F
GHVISAVGEERKLGK	X	X	X	X	X	X	X	G	E	E	X	X	X	X	X
GPIQNPGHNGSVEDT	X	X	X	X	X	X	X	X	N	G	X	X	D,E	D,E	X
HIRDYPHYYWGELRL	X	X	X	X	X	X	X	X	X	X	G	E	X	X	X
HKQPEMPWVYMFGGE	X	X	X	X	X	X	X	X	X	X	X	X	X	G,E	E
IMAIGINATFAKENK	X	X	X	X	X	X	X	X	X	X	A,G	X	E	X	K
INAMEGELQMQETVH	X	X	X	X	X	G	A,E	I,D,L,E	M,E,I,Q	X	D,E,Q	X	X	X	X
IRMYEKSKPPVIQPS	X	X	X	X	X	X	X	X	X	X	X	X	Q	X	S
MFTLFSSWTVDEQQV	X	X	X	X	X	X	X	X	X	X	G,E,D	E	M,Q,T,V,I	Q	V
MIQQQFLFYRTPQPL	X	X	X	X	X	X	X	X	X	X	T	P	Q	P	X
MTKYWGHLETHYQPS	X	X	X	X	X	X	X	X	X	S,P,T	X	X	Q	P	S
NAPRHEEWLVATEQP	X	X	X	X	X	S,W,Y,D,E	W,Y,D,E	E,W,D	X	X	Q,A,E,T,P	T	X	D,Q,E	P
NMHNIEETVDQKFRD	X	X	X	X	X	X	X	X	P,V,I	D	X	K	I,L,F	X	F,D
SHSQSLEHTTMNALH	X	X	X	X	X	L,I	X	H	I,T,V	T	X	N	X	X	X
SNGSIYSKPHNGTMM	X	X	X	X	X	X	X	X	P	X	N	X	X	X	X
TAGGKPPFGTDGSGS	X	X	X	X	X	X	X	X	X	X	X	S,G	C,G,D,S,E	Q,G	X
TENYQKFTDIMPAFV	X	X	X	X	X	X	F,Q	X	D,E	X	X	X	X	F,P	X
WKYSISWGFTRGFFN	X	X	X	X	X	X	X	X	X	X	X	F,E,G,Y	F,E	F	E,N,D
WLITEGREQATTAAI	X	X	X	X	X	X	X	D,E	X	X	M,T,Q	T	X	X	F,Y,K,I,R

Bioinformatic analysis of fingerprints

The four most abundant fingerprints, shown in Figure 5, were queried in the "UniProtKb/Swiss-Prot" protein database (release 2016_11 of 30-Nov-16: 553231 entries) with the "ScanProsite tool". All results of the query are listed in Supplementary Tables S2 – S5.

Table S2: Results of the query of the "UniProtKb/Swiss-Prot" protein database with the "ScanProsite tool" for the NPVEXXX motif with the amino acids A/N/D-N/P-V-E-E/D/A/F-F/Y-I/V/L. The table lists the organism and protein on which the motif is located, the amino acid position and the exact sequence. The query included splice variants and excludes fragments.

Organism	Antigen	Database entry	Position & Epitope
Acanthamoeba polyphaga mimivirus (APMV)	Uncharacterized protein R261.	sp|Q5UP22|YR261_MIMIV (181 aa)	32-38 DNVEAYI
Archaeoglobus fulgidus (strain ATCC 49558 / VC-16 / DSM 4304 / JCM 9628 / NBRC 100126)	Sulfite reductase, dissimilatory-type subunit beta (EC 1.8.99.5) (Hydrogensulfite reductase subunit beta).	sp|Q59110|DSRB_ARCFU (366 aa)	88-94 NNVEFFV
Autographa californica nuclear polyhedrosis virus (AcMNPV)	Late expression factor 7.	sp|P41677|LEF7_NPVAC (226 aa)	50-56 NNVEDYL
Bacillus phage B103 (Bacteriophage B103)	Proximal tail tube connector protein (Gene product 11) (gp11) (Lower collar protein) (Protein p11).	sp|Q37892|TUB11_BPB03 (293 aa)	245-251 NNVEDYI
Bacillus thuringiensis subsp. morrisoni	Pesticidal crystal protein Cry3Aa (73 kDa crystal protein) (Crystaline entomocidal protoxin) (Insecticidal delta-endotoxin CryIIIA(a)).	sp|P0A380|CR3AA_BACTM (644 aa)	138-144 NNVEDYV
Bacillus thuringiensis subsp. san diego	Pesticidal crystal protein Cry3Aa (73 kDa crystal protein) (Crystaline entomocidal protoxin) (Insecticidal delta-endotoxin CryIIIA(a)).	sp|P0A381|CR3AA_BACTD (644 aa)	138-144 NNVEDYV
Bacillus thuringiensis subsp. tenebrionis	Pesticidal crystal protein Cry3Aa (73 kDa crystal protein) (Crystaline entomocidal protoxin) (Insecticidal delta-endotoxin CryIIIA(a)).	sp|P0A379|CR3AA_BACTT (644 aa)	138-144 NNVEDYV
Bos taurus (Bovine)	Dihydroxyacetone phosphate acyltransferase (DAP-AT) (DHAP-AT) (EC 2.3.1.42) (Acyl-CoA:dihydroxyacetonephosphateacyltransferase) (Glycerone-phosphate O-acyltransferase).	sp|A4IF87|GNPAT_BOVIN (680 aa)	236-242 APVEFFL
Cyanothece sp. (strain PCC 7424) (Synechococcus sp. (strain ATCC 29155))	RNA 3'-terminal phosphate cyclase (RNA cyclase) (RNA-3'-phosphate cyclase) (EC 6.5.1.4).	sp|B7KCF3|RTCA_CYAP7 (350 aa)	285-291 APVEEFL
Debaryomyces hansenii (strain ATCC 36239 / CBS 767 / JCM 1990 / NBRC 0083 / IGC 2968) (Yeast) (Torulaspora hansenii)	Protein STU1.	sp|Q6BK07|STU1_DEBHA (1529 aa)	1106-1112 DNVEEYV
Homo sapiens (Human)	Dihydroxyacetone phosphate acyltransferase (DAP-AT) (DHAP-AT) (EC 2.3.1.42) (Acyl-CoA:dihydroxyacetonephosphateacyltransferase) (Glycerone-phosphate O-acyltransferase).	sp|O15228|GNPAT_HUMAN (680 aa)	236-242 APVEFFL
Homo sapiens (Human)	Dihydroxyacetone phosphate acyltransferase (DAP-AT) (DHAP-AT) (EC 2.3.1.42) (Acyl-CoA:dihydroxyacetonephosphateacyltransferase) (Glycerone-phosphate O-acyltransferase).	sp|O15228-2|GNPAT_HUMAN (619 aa)	175-181 APVEFFL
Human cytomegalovirus (strain AD169) (HHV-5) (Human herpesvirus 5)	Protein HHLF1.	sp|P09695|TRS1_HCMVA (788 aa)	575-581 ANVEDYL
Human cytomegalovirus (strain Merlin) (HHV-5) (Human herpesvirus 5)	Protein TRS1.	sp|Q6SVX2|TRS1_HCMVM (788 aa)	576-582 ANVEDYL

Methylobacterium extorquens (strain ATCC 14718 / DSM 1338 / JCM 2805 / NCIMB 9133 / AM1)	Glycerate dehydrogenase (GDH) (EC 1.1.1.29) (Glyoxylate reductase) (Hydroxypyruvate dehydrogenase) (NADH-dependent hydroxypyruvate reductase) (HPR) (HPR-A).	sp|Q59516|DHGY_METEA (314 aa)	298-304 DNVEAFV
Mus musculus (Mouse)	Dihydroxyacetone phosphate acyltransferase (DAP-AT) (DHAP-AT) (EC 2.3.1.42) (Acyl-CoA:dihydroxyacetonephosphateacyltransferase) (Glycerone-phosphate O-acyltransferase).	sp|P98192|GNPAT_MOUSE (678 aa)	235-241 APVEFFL
Mus musculus (Mouse)	Ras-GEF domain-containing family member 1C.	sp|Q9D300|RGF1C_MOUSE (466 aa)	246-252 NNVEAYV
Mus musculus (Mouse)	Ras-GEF domain-containing family member 1C.	sp|Q9D300-2|RGF1C_MOUSE (469 aa)	249-255 NNVEAYV
Mus musculus (Mouse)	von Willebrand factor A domain-containing protein 3A.	sp|Q3UVV9|VWA3A_MOUSE (1148 aa)	406-412 NPVEEFV
Petrotoga mobilis (strain DSM 10674 / SJ95)	GTPase Obg (EC 3.6.5.-) (GTP-binding protein Obg).	sp|A9BK05|OBG_PETMO (440 aa)	256-262 DPVEDYI
Pseudoalteromonas haloplanktis (strain TAC 125)	Methionine--tRNA ligase (EC 6.1.1.10) (Methionyl-tRNA synthetase) (MetRS).	sp|Q3IL10|SYM_PSEHT (673 aa)	497-503 DNVEAFL
Rattus norvegicus (Rat)	Dihydroxyacetone phosphate acyltransferase (DAP-AT) (DHAP-AT) (EC 2.3.1.42) (Acyl-CoA:dihydroxyacetonephosphateacyltransferase) (Glycerone-phosphate O-acyltransferase).	sp|Q9ES71|GNPAT_RAT (678 aa)	235-241 APVEFFL
Rattus norvegicus (Rat)	von Willebrand factor A domain-containing protein 3A.	sp|A1A5Q7|VWA3A_RAT (737 aa)	406-412 NPVEEFV
Rhizopus delemar (strain RA 99-880 / ATCC MYA-4621 / FGSC 9543 / NRRL 43880) (Mucormycosis agent) (Rhizopus arrhizus var. delemar)	FK506-binding protein 4 (EC 5.2.1.8) (Histone proline isomerase) (Peptidyl-prolyl cis-trans isomerase) (PPIase) (Rotamase).	sp|P0C1J6|FKBP4_RHIO9 (382 aa)	115-121 DNVEDFL
Rhodobacter sphaeroides (strain ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158)	Coproporphyrinogen III oxidase, anaerobic 1 (Coprogen oxidase) (Coproporphyrinogenase) (EC 1.3.98.3).	sp|P33770|HEMF_RHOS4 (452 aa)	78-84 APVEAYV
Rhodobacter sphaeroides (strain ATCC 17025 / ATH 2.4.3)	Oxygen-independent coproporphyrinogen III oxidase (CPO) (EC 1.3.98.3) (Coproporphyrinogen III dehydrogenase) (CPDH).	sp|P95651|HEMN_RHOS5 (452 aa)	78-84 APVEAYV
Rhodococcus erythropolis (strain PR4 / NBRC 100887)	4-hydroxy-2-oxovalerate aldolase 1 (HOA 1) (EC 4.1.3.39) (4-hydroxy-2-keto-pentanoic acid aldolase 1) (4-hydroxy-2-oxopentanoate aldolase 1).	sp|C0ZPX1|HOA1_RHOE4 (347 aa)	241-247 APVEAFV
Rhodococcus jostii (strain RHA1)	4-hydroxy-2-oxovalerate aldolase 3 (HOA 3) (EC 4.1.3.39) (4-hydroxy-2-keto-pentanoic acid aldolase 3) (4-hydroxy-2-oxopentanoate aldolase 3).	sp|Q0S815|HOA3_RHOJR (339 aa)	235-241 APVEAFV
Rhodococcus opacus (strain B4)	4-hydroxy-2-oxovalerate aldolase 3 (HOA 3) (EC 4.1.3.39) (4-hydroxy-2-keto-pentanoic acid aldolase 3) (4-hydroxy-2-oxopentanoate aldolase 3).	sp|C1BAJ4|HOA3_RHOOB (339 aa)	235-241 APVEAFV
Staphylococcus epidermidis (strain ATCC 12228)	CTP synthase (EC 6.3.4.2) (Cytidine 5'-triphosphate synthase) (Cytidine triphosphate synthetase) (CTP synthetase) (CTPS) (UTP--ammonia ligase).	sp|Q8CNI2|PYRG_STAES (535 aa)	338-344 DNVEAYL
Staphylococcus epidermidis (strain ATCC 35984 / RP62A)	CTP synthase (EC 6.3.4.2) (Cytidine 5'-triphosphate synthase) (Cytidine triphosphate synthetase) (CTP synthetase) (CTPS) (UTP--ammonia ligase).	sp|Q5HM95|PYRG_STAEQ (535 aa)	338-344 DNVEAYL
Staphylococcus haemolyticus (strain JCSC1435)	CTP synthase (EC 6.3.4.2) (Cytidine 5'-triphosphate synthase) (Cytidine triphosphate synthetase) (CTP synthetase) (CTPS) (UTP--ammonia ligase).	sp|Q4L808|PYRG_STAHJ (535 aa)	338-344 NNVEDYL

Table S3: Results of the query of the "UniProtKb/Swiss-Prot" protein database with the "ScanProsite tool" for the XPEFXGSXX motif with the amino acids V/I/P-P-E-F-x-G-A/S-x-V/P. The table lists the organism and protein on which the motif is located, the amino acid position and the exact sequence. The query included splice variants and excludes fragments.

Organism	Antigen	Database entry	Position & Epitope
Mus musculus (Mouse)	Lambda-crystallin homolog (EC 1.1.1.45) (L-gulonate 3-dehydrogenase) (Gul3DH).	sp|Q99KP3|CRYL1_MOUSE (319 aa)	269-277 VPEFsGAtV
Oryctolagus cuniculus (Rabbit)	Lambda-crystallin (EC 1.1.1.45) (L-gulonate 3-dehydrogenase) (Gul3DH).	sp|P14755|CRYL1_RABIT (319 aa)	269-277 IPEFsGAtV
Staphylococcus aureus	Extracellular matrix protein-binding protein emp (40 kDa vitronectin-binding cell surface protein)	sp|P0C6P1|EMP_STAAU (340 aa)	328-336 VPEFkGSlP
Staphylococcus aureus (strain bovine RF122 / ET3-1)	Extracellular matrix protein-binding protein emp.	sp|Q2YWL4|EMP_STAAB (339 aa)	327-335 VPEFkGSiP
Staphylococcus aureus (strain COL)	Extracellular matrix protein-binding protein emp.	sp|Q5HHM6|EMP_STAAC (340 aa)	328-336 VPEFkGSlP
Staphylococcus aureus (strain MSSA476)	Extracellular matrix protein-binding protein emp.	sp|Q6GB43|EMP_STAAS (340 aa)	328-336 VPEFkGSlP
Staphylococcus aureus (strain Mu50 / ATCC 700699)	Extracellular matrix protein-binding protein emp.	sp|Q99VJ2|EMP_STAAM (340 aa)	328-336 VPEFkGSlP
Staphylococcus aureus (strain MW2)	Extracellular matrix protein-binding protein emp.	sp|Q8NXI8|EMP_STAAW (340 aa)	328-336 VPEFkGSlP
Staphylococcus aureus (strain N315)	Extracellular matrix protein-binding protein emp.	sp|Q7A6P4|EMP_STAAN (340 aa)	328-336 VPEFkGSlP
Staphylococcus aureus (strain NCTC 8325)	Extracellular matrix protein-binding protein emp.	sp|Q2G0I2|EMP_STAA8 (340 aa)	328-336 VPEFkGSlP
Staphylococcus aureus (strain Newman)	Extracellular matrix protein-binding protein emp.	sp|A6QF98|EMP_STAAE (340 aa)	328-336 VPEFkGSlP
Staphylococcus aureus (strain USA300)	Extracellular matrix protein-binding protein emp.	sp|Q2FIK4|EMP_STAA3 (340 aa)	328-336 VPEFkGSlP

Table S4: Results of the query of the "UniProtKb/Swiss-Prot" protein database with the "ScanProsite tool" for the KXXFPQXT motif with the amino acids KE/-x-F/Y-F/Y-P-Q/Y-T/K/M/F/V-T/V/I. The table lists the organism and protein on which the motif is located, the amino acid position and the exact sequence. The query included splice variants and excludes fragments.

Organism	Antigen	Database entry	Position & Epitope
Acholeplasma laidlawii (strain PG-8A)]	tRNA-specific 2-thiouridylase MnmA (EC 2.8.1.13).	sp|A9NFP7|MNMA_ACHLI (375 aa)	332-339 KvFYPQKI
Aster yellows witches'-broom phytoplasma (strain AYWB)]	tRNA-specific 2-thiouridylase MnmA (EC 2.8.1.13).	sp|Q2NIM9|MNMA_AYWBP (378 aa)	335-342 EiYYPQTI
Bacteroides thetaiotaomicron (strain ATCC 29148 / DSM 2079 / NCTC 10582 / E50 / VPI-5482)]	UDP-N-acetylglucosamine--N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase (EC 2.4.1.227) (Undecaprenyl-PP-MurNAc-pentapeptide-UDPGlcNAc GlcNAc transferase).	sp|Q8A258|MURG_BACTN (372 aa)	232-239 KyYYPQVT
Campylobacter jejuni (strain RM1221)]	Thymidylate kinase (EC 2.7.4.9) (dTMP kinase).	sp|Q5HV26|KTHY_CAMJR (192 aa)	112-119 EnFFPQKI
Campylobacter jejuni subsp. doylei (strain ATCC BAA-1458 / RM4099 / 269.97)]	Thymidylate kinase (EC 2.7.4.9) (dTMP kinase).	sp|A7H476|KTHY_CAMJD (192 aa)	112-119 EdFFPQKI
Campylobacter jejuni subsp. jejuni serotype O:2 (strain ATCC 700819 / NCTC 11168)]	Thymidylate kinase (EC 2.7.4.9) (dTMP kinase).	sp|Q9PPF3|KTHY_CAMJE (192 aa)	112-119 EnFFPQKI
Campylobacter jejuni subsp. jejuni serotype O:23/36 (strain 81-176)]	Thymidylate kinase (EC 2.7.4.9) (dTMP kinase).	sp|A1VZB4|KTHY_CAMJJ (192 aa)	112-119 EnFFPQKI
Campylobacter jejuni subsp. jejuni serotype O:6 (strain 81116 / NCTC 11828)]	Thymidylate kinase (EC 2.7.4.9) (dTMP kinase).	sp|A8FLH9|KTHY_CAMJ8 (192 aa)	112-119 EnFFPQKI
Chlorobium tepidum (strain ATCC 49652 / DSM 12025 / NBRC 103806 / TLS)]	Uncharacterized N-acetyltransferase CT2212 (EC 2.3.1.-).	sp|P59599|Y2212_CHLTE (155 aa)	126-133 KeYFPQKI

Homo sapiens (Human)]	Gametocyte-specific factor 1-like (Protein FAM112A).	sp|Q9H1H1|GTSFL_HUMAN (148 aa)	118-125 KtFFPQKV
Homo sapiens (Human)]	Nebulin.	sp|P20929|NEBU_HUMAN (6669 aa)	468-475 KgFFPQTI
Homo sapiens (Human)]	Nebulin.	sp|P20929-2|NEBU_HUMAN (8525 aa)	468-475 KgFFPQTI
Homo sapiens (Human)]	Nebulin.	sp|P20929-3|NEBU_HUMAN (8525 aa)	468-475 KgFFPQTI
Homo sapiens (Human)]	Nebulin.	sp|P20929-4|NEBU_HUMAN (6669 aa)	468-475 KgFFPQTI
Homo sapiens (Human)]	Reelin (EC 3.4.21.-).	sp|P78509|RELN_HUMAN (3460 aa)	1897-1904 EfYFPQTT
Homo sapiens (Human)]	Reelin (EC 3.4.21.-).	sp|P78509-2|RELN_HUMAN (3458 aa)	1897-1904 EfYFPQTT
Homo sapiens (Human)]	Reelin (EC 3.4.21.-).	sp|P78509-3|RELN_HUMAN (3427 aa)	1897-1904 EfYFPQTT
Meyerozyma guilliermondii (strain ATCC 6260 / CBS 566 / DSM 6381 / JCM 1539 / NBRC 10279 / NRRL Y-324) (Yeast) (Candida guilliermondii)]	Pro-apoptotic serine protease NMA111 (EC 3.4.21.-).	sp|A5DAL3|NM111_PICGU (991 aa)	59-66 EnYFPQTT
Mus musculus (Mouse)]	Reelin (EC 3.4.21.-) (Reeler protein).	sp|Q60841|RELN_MOUSE (3461 aa)	1898-1905 EfYFPQTT
Mus musculus (Mouse)]	Reelin (EC 3.4.21.-) (Reeler protein).	sp|Q60841-2|RELN_MOUSE (3459 aa)	1898-1905 EfYFPQTT
Mus musculus (Mouse)]	Reelin (EC 3.4.21.-) (Reeler protein).	sp|Q60841-3|RELN_MOUSE (3428 aa)	1898-1905 EfYFPQTT
Ochrobactrum anthropi (strain ATCC 49188 / DSM 6882 / JCM 21032 / NBRC 15819 / NCTC 12168)]	Biotin synthase (EC 2.8.1.6).	sp|A6X2S8|BIOB_OCHA4 (336 aa)	175-182 ErFYPQVI
Rattus norvegicus (Rat)]	Inactive serine protease 45 (Inactive testis serine protease 5).	sp|Q6IE62|PRS45_RAT (330 aa)	214-221 KtFYPQVI
Rattus norvegicus (Rat)]	Reelin (EC 3.4.21.-).	sp|P58751|RELN_RAT (3462 aa)	1899-1906 EfYFPQTT
Rattus norvegicus (Rat)]	Reelin (EC 3.4.21.-).	sp|P58751-2|RELN_RAT (3460 aa)	1899-1906 EfYFPQTT
Rattus norvegicus (Rat)]	Reelin (EC 3.4.21.-).	sp|P58751-3|RELN_RAT (3429 aa)	1899-1906 EfYFPQTT
Rhesus cytomegalovirus (strain 68-1) (RhCMV)]	Envelope glycoprotein B (gB).	sp|P89053|GB_RHCM6 (854 aa)	756-763 EhFFPYVV

Table S5: Results of the query of the "UniProtKb/Swiss-Prot" protein database with the "ScanProsite tool" for the LXAXETX motif with the amino acids L-D/T-A-x-E-T-G/S. The table lists the organism and protein on which the motif is located, the amino acid position and the exact sequence. The query included splice variants and excludes fragments.

Organism	Antigen	Database entry	Position & Epitope
Agathobacter rectalis (strain ATCC 33656 / DSM 3377 / JCM 17463 / KCTC 5835 / VPI 0990) (Eubacterium rectale)	Peptide chain release factor 1 (RF-1).	sp|C4Z911|RF1_AGARV (358 aa)	148-154 LDAdETG
Aquifex aeolicus (strain VF5)	ATP-dependent 6-phosphofructokinase (ATP-PFK) (Phosphofructokinase) (EC 2.7.1.11) (Phosphohexokinase).	sp|O67605|PFKA_AQUAE (321 aa)	110-116 LTAeETG
Arabidopsis thaliana (Mouse-ear cress)	Cysteine-rich repeat secretory protein 39 (Plasmodesmata-located protein 4) (PDLP4).	sp|Q6E263|CRR39_ARATH (319 aa)	196-202 LTAaETG
Arabidopsis thaliana (Mouse-ear cress)	Cysteine-rich repeat secretory protein 39 (Plasmodesmata-located protein 4) (PDLP4).	sp|Q6E263-2|CRR39_ARATH (318 aa)	195-201 LTAaETG
Arabidopsis thaliana (Mouse-ear cress)	Bifunctional bis(5'-adenosyl)-triphosphatase/adenylylsulfatase FHIT (EC 3.6.1.29) (EC 3.6.2.1) (Fragile histidine triad protein) (Purine nucleoside phosphoramidase FHIT) (EC 3.9.1.-).	sp|F4KEV7|FHIT_ARATH (180 aa)	75-81 LTAdETS
Arabidopsis thaliana (Mouse-ear cress)	Bifunctional bis(5'-adenosyl)-triphosphatase/adenylylsulfatase FHIT (EC	sp|F4KEV7-2|FHIT_ARATH (160 aa)	55-61 LTAdETS

	3.6.1.29) (EC 3.6.2.1) (Fragile histidine triad protein) (Purine nucleoside phosphoramidase FHIT) (EC 3.9.1.-).		
Bos taurus (Bovine)	Calcium/calmodulin-dependent 3',5'-cyclic nucleotide phosphodiesterase 1B (Cam-PDE 1B) (EC 3.1.4.17) (63 kDa Cam-PDE).	sp|Q01061|PDE1B_BOVIN (534 aa)	206-212 LDAlETG
Chlorobium phaeovibrioides (strain DSM 265 / 1930) (Prosthecochloris vibrioformis (strain DSM 265))	Serine--tRNA ligase (EC 6.1.1.11) (Seryl-tRNA synthetase) (SerRS) (Seryl-tRNA(Ser/Sec) synthetase).	sp|A4SFD5|SYS_CHLPM (430 aa)	93-99 LTAlETS
Coxsackievirus A21 (strain Coe)	Genome polyprotein	sp|P22055|POLG_CXA21 (2206 aa)	620-626 LTAvETG
Coxsackievirus A24 (strain EH24/70)	Genome polyprotein	sp|P36290|POLG_CXA24 (2214 aa)	624-630 LTAvETG
Coxsackievirus A9 (strain Griggs)	Genome polyprotein	sp|P21404|POLG_CXA9 (2201 aa)	599-605 LTAvETG
Coxsackievirus B1 (strain Japan)	Genome polyprotein	sp|P08291|POLG_CXB1J (2182 aa)	601-607 LTAaETG
Coxsackievirus B2 (strain Ohio-1)	Genome polyprotein	sp|Q9YLG5|POLG_CXB2O (2187 aa)	602-608 LTAvETG
Coxsackievirus B3 (strain Nancy)	Genome polyprotein	sp|P03313|POLG_CXB3N (2185 aa)	601-607 LTAaETG
Coxsackievirus B3 (strain Woodruff)	Genome polyprotein	sp|Q66282|POLG_CXB3W (2185 aa)	601-607 LTAaETG
Coxsackievirus B4 (strain E2)	Genome polyprotein	sp|Q86887|POLG_CXB4E (2183 aa)	599-605 LTAvETG
Coxsackievirus B4 (strain JVB / Benschoten / New York/51)	Genome polyprotein	sp|P08292|POLG_CXB4J (2183 aa)	599-605 LTAvETG
Coxsackievirus B5 (strain Peterborough / 1954/UK/85)	Genome polyprotein	sp|Q03053|POLG_CXB5P (2185 aa)	599-605 LTAaETG
Coxsackievirus B6 (strain Schmitt)	Genome polyprotein	sp|Q9QL88|POLG_CXB6S (2184 aa)	599-605 LTAvETG
Cyanothece sp. (strain PCC 7425 / ATCC 29141)	ATP synthase epsilon chain (ATP synthase F1 sector epsilon subunit) (F-ATPase epsilon subunit).	sp|B8HP54|ATPE_CYAP4 (138 aa)	41-47 LTAlETG
Echovirus 1 (strain Human/Egypt/Farouk/1951) (E-1)	Genome polyprotein	sp|O91734|POLG_EC01F (2184 aa)	600-606 LTAvETG
Echovirus 11 (strain Gregory)	Genome polyprotein	sp|P29813|POLG_EC11G (2195 aa)	600-606 LTAvETG
Echovirus 12 (strain Travis)	Genome polyprotein	sp|Q66575|POLG_EC12T (2193 aa)	599-605 LTAaETG
Echovirus 30 (strain Bastianni)	Genome polyprotein	sp|Q9WN78|POLG_EC30B (2194 aa)	599-605 LTAvETG
Echovirus 5 (strain Noyce)	Genome polyprotein	sp|Q9YLJ1|POLG_EC05N (2196 aa)	601-607 LTAvETG
Echovirus 6 (strain Charles)	Genome polyprotein	sp|Q66474|POLG_EC06C (2191 aa)	599-605 LTAaETG
Echovirus 9 (strain Barty)	Genome polyprotein	sp|Q66577|POLG_EC09B (2203 aa)	600-606 LTAaETG
Echovirus 9 (strain Hill)	Genome polyprotein	sp|Q66849|POLG_EC09H (2193 aa)	600-606 LTAaETG
Emericella nidulans (strain FGSC A4 / ATCC 38163 / CBS 112.46 / NRRL 194 / M139) (Aspergillus nidulans)	Uncharacterized protein AN0679.	sp|P0C154|Y0679_EMENI (251 aa)	48-54 LDApETS
Escherichia coli (strain 55989 / EAEC)	Hydroxylamine reductase (EC 1.7.99.1) (Hybrid-cluster protein) (HCP) (Prismane protein).	sp|B7LD66|HCP_ECO55 (550 aa)	218-224 LDAgETG
Escherichia coli (strain ATCC 8739 / DSM 1576 / Crooks)	Hydroxylamine reductase (EC 1.7.99.1) (Hybrid-cluster protein) (HCP) (Prismane protein).	sp|B1IWP9|HCP_ECOLC (550 aa)	218-224 LDAgETG
Escherichia coli (strain K12 / DH10B)	Hydroxylamine reductase (EC 1.7.99.1) (Hybrid-cluster protein) (HCP) (Prismane protein).	sp|B1X814|HCP_ECODH (550 aa)	218-224 LDAgETG
Escherichia coli (strain K12 / MC4100 / BW2952)	Hydroxylamine reductase (EC 1.7.99.1) (Hybrid-cluster protein) (HCP) (Prismane protein).	sp|C4ZY45|HCP_ECOBW (550 aa)	218-224 LDAgETG
Escherichia coli (strain K12)	Hydroxylamine reductase (EC 1.7.99.1) (Hybrid-cluster protein) (HCP) (Prismane protein).	sp|P75825|HCP_ECOLI (550 aa)	218-224 LDAgETG
Escherichia coli (strain K12)	Uncharacterized protein YggR.	sp|P52052|YGGR_ECOLI (326 aa)	212-218 LTAaETG
Escherichia coli (strain SE11)	Hydroxylamine reductase (EC 1.7.99.1)	sp|B6I8U4|HCP_ECOSE (550	218-224

|			
	(Hybrid-cluster protein) (HCP) (Prismane protein).	aa)	LDAgETG
Escherichia coli (strain SMS-3-5 / SECEC)	Hydroxylamine reductase (EC 1.7.99.1) (Hybrid-cluster protein) (HCP) (Prismane protein).	sp|B1LN27|HCP_ECOSM (550 aa)	218-224 LDAgETS
Escherichia coli (strain UTI89 / UPEC)	Hydroxylamine reductase (EC 1.7.99.1) (Hybrid-cluster protein) (HCP) (Prismane protein).	sp|Q1RE52|HCP_ECOUT (552 aa)	220-226 LDAgETG
Escherichia coli O1:K1 / APEC	Hydroxylamine reductase (EC 1.7.99.1) (Hybrid-cluster protein) (HCP) (Prismane protein).	sp|A1A9B1|HCP_ECOK1 (552 aa)	220-226 LDAgETG
Escherichia coli O127:H6 (strain E2348/69 / EPEC)	Hydroxylamine reductase (EC 1.7.99.1) (Hybrid-cluster protein) (HCP) (Prismane protein).	sp|B7UMW5|HCP_ECO27 (550 aa)	218-224 LDAgETG
Escherichia coli O139:H28 (strain E24377A / ETEC)	Hydroxylamine reductase (EC 1.7.99.1) (Hybrid-cluster protein) (HCP) (Prismane protein).	sp|A7ZJU2|HCP_ECO24 (550 aa)	218-224 LDAgETG
Escherichia coli O157:H7 (strain EC4115 / EHEC)	Hydroxylamine reductase (EC 1.7.99.1) (Hybrid-cluster protein) (HCP) (Prismane protein).	sp|B5YSG9|HCP_ECO5E (550 aa)	218-224 LDAgETG
Escherichia coli O157:H7	Hydroxylamine reductase (EC 1.7.99.1) (Hybrid-cluster protein) (HCP) (Prismane protein).	sp|Q8X6L0|HCP_ECO57 (550 aa)	218-224 LDAgETG
Escherichia coli O17:K52:H18 (strain UMN026 / ExPEC)	Hydroxylamine reductase (EC 1.7.99.1) (Hybrid-cluster protein) (HCP) (Prismane protein).	sp|B7NAM4|HCP_ECOLU (550 aa)	218-224 LDAgETG
Escherichia coli O45:K1 (strain S88 / ExPEC)	Hydroxylamine reductase (EC 1.7.99.1) (Hybrid-cluster protein) (HCP) (Prismane protein).	sp|B7MHH8|HCP_ECO45 (550 aa)	218-224 LDAgETG
Escherichia coli O6:H1 (strain CFT073 / ATCC 700928 / UPEC)	Hydroxylamine reductase (EC 1.7.99.1) (Hybrid-cluster protein) (HCP) (Prismane protein).	sp|Q8FJE0|HCP_ECOL6 (550 aa)	218-224 LDAgETG
Escherichia coli O6:K15:H31 (strain 536 / UPEC)	Hydroxylamine reductase (EC 1.7.99.1) (Hybrid-cluster protein) (HCP) (Prismane protein).	sp|Q0TJH6|HCP_ECOL5 (550 aa)	218-224 LDAgETG
Escherichia coli O7:K1 (strain IAI39 / ExPEC)	Hydroxylamine reductase (EC 1.7.99.1) (Hybrid-cluster protein) (HCP) (Prismane protein).	sp|B7NPH0|HCP_ECO7I (550 aa)	218-224 LDAgETS
Escherichia coli O8 (strain IAI1)	Hydroxylamine reductase (EC 1.7.99.1) (Hybrid-cluster protein) (HCP) (Prismane protein).	sp|B7M801|HCP_ECO8A (550 aa)	218-224 LDAgETG
Escherichia coli O81 (strain ED1a)	Hydroxylamine reductase (EC 1.7.99.1) (Hybrid-cluster protein) (HCP) (Prismane protein).	sp|B7MQX7|HCP_ECO81 (550 aa)	218-224 LDAgETG
Escherichia coli O9:H4 (strain HS)	Hydroxylamine reductase (EC 1.7.99.1) (Hybrid-cluster protein) (HCP) (Prismane protein).	sp|A7ZYH7|HCP_ECOHS (550 aa)	218-224 LDAgETG
Escherichia fergusonii (strain ATCC 35469 / DSM 13698 / CDC 0568-73)	Hydroxylamine reductase (EC 1.7.99.1) (Hybrid-cluster protein) (HCP) (Prismane protein).	sp|B7LN38|HCP_ESCF3 (550 aa)	218-224 LDAgETG
Eubacterium eligens (strain ATCC 27750 / VPI C15-48)	Methionyl-tRNA formyltransferase (EC 2.1.2.9).	sp|C4Z520|FMT_EUBE2 (315 aa)	155-161 LDAkETG
Guillardia theta (Cryptomonas phi)	Chaperone protein dnaK (HSP70) (Heat shock 70 kDa protein) (Heat shock protein 70).	sp|P29215|DNAK_GUITH (627 aa)	281-287 LTAtETG
Homo sapiens (Human)	Putative butyrophilin subfamily 2 member A3.	sp|Q96KV6|BT2A3_HUMAN (586 aa)	502-508 LTAmETS
Homo sapiens (Human)	C3 and PZP-like alpha-2-macroglobulin domain-containing protein 8.	sp|Q8IZJ3|CPMD8_HUMAN (1885 aa)	242-248 LDAcETG
Homo sapiens (Human)	C3 and PZP-like alpha-2-macroglobulin domain-containing protein 8.	sp|Q8IZJ3-2|CPMD8_HUMAN (1815 aa)	242-248 LDAcETG
Homo sapiens (Human)	Calcium/calmodulin-dependent 3',5'-cyclic nucleotide phosphodiesterase 1B (Cam-PDE 1B) (EC 3.1.4.17) (63 kDa Cam-PDE).	sp|Q01064|PDE1B_HUMAN (536 aa)	208-214 LDAlETG
Homo sapiens (Human)	Calcium/calmodulin-dependent 3',5'-cyclic nucleotide phosphodiesterase 1B (Cam-PDE 1B) (EC 3.1.4.17) (63 kDa Cam-PDE).	sp|Q01064-2|PDE1B_HUMAN (516 aa)	188-194 LDAlETG
Human rhinovirus 14 (HRV-14)	Genome polyprotein	sp|P03303|POLG_HRV14 (2179 aa)	601-607 LTAnETG
Human rhinovirus 16 (HRV-16)	Genome polyprotein	sp|Q82122|POLG_HRV16 (2153 aa)	602-608 LDAaETG
Human rhinovirus 1A (HRV-1A)	Genome polyprotein	sp|P23008|POLG_HRV1A	604-610

		(2157 aa)	LDAaETG
Human rhinovirus 1B (HRV-1B)	Genome polyprotein	sp|P12916|POLG_HRV1B (2157 aa)	604-610 LDAaETG
Human rhinovirus 2 (HRV-2)	Genome polyprotein	sp|P04936|POLG_HRV2 (2150 aa)	601-607 LDAaETG
Human rhinovirus 3 (HRV-3)	Genome polyprotein	sp|Q82081|POLG_HRV3 (2178 aa)	601-607 LTAnETG
Human rhinovirus A serotype 89 (strain 41467-Gallo) (HRV-89)	Genome polyprotein	sp|P07210|POLG_HRV8A (2164 aa)	608-614 LDAaETG
Klebsiella pneumoniae (strain 342)	Chaperone protein DnaK (HSP70) (Heat shock 70 kDa protein) (Heat shock protein 70).	sp|B5Y242|DNAK_KLEP3 (638 aa)	569-575 LTAlETS
Lecanicillium sp.	Nonribosomal peptide synthetase vlms (NRPS vlmS) (EC 6.3.2.-) (Verlamelin biosynthesis protein S).	sp|A0A024F910|VLMS_LECSP (8903 aa)	711-717 LDAgETG
Moorella thermoacetica (strain ATCC 39073 / JCM 9320)	Phosphopentomutase (EC 5.4.2.7) (Phosphodeoxyribomutase).	sp|Q2RID0|DEOB_MOOTA (389 aa)	310-316 LDAlETS
Mus musculus (Mouse)	Interphotoreceptor matrix proteoglycan 2 (Sialoprotein associated with cones and rods proteoglycan) (Spacrcan).	sp|Q80XH2|IMPG2_MOUSE (1243 aa)	59-65 LDAtETG
Mus musculus (Mouse)	Interphotoreceptor matrix proteoglycan 2 (Sialoprotein associated with cones and rods proteoglycan) (Spacrcan).	sp|Q80XH2-2|IMPG2_MOUSE (1134 aa)	59-65 LDAtETG
Mus musculus (Mouse)	Photoreceptor-specific nuclear receptor (Nuclear receptor subfamily 2 group E member 3) (Retina-specific nuclear receptor).	sp|Q9QXZ7|NR2E3_MOUSE (395 aa)	125-131 LDAmETG
Mycobacterium sp. (strain JLS)	Histidinol-phosphate aminotransferase (EC 2.6.1.9) (Imidazole acetol-phosphate transaminase).	sp|A3Q130|HIS8_MYCSJ (377 aa)	190-196 LDAmETG
Mycobacterium sp. (strain KMS)	Histidinol-phosphate aminotransferase (EC 2.6.1.9) (Imidazole acetol-phosphate transaminase).	sp|A1UHK7|HIS8_MYCSK (377 aa)	190-196 LDAmETG
Mycobacterium sp. (strain MCS)	Histidinol-phosphate aminotransferase (EC 2.6.1.9) (Imidazole acetol-phosphate transaminase).	sp|Q1B7G5|HIS8_MYCSS (377 aa)	190-196 LDAmETG
Neisseria gonorrhoeae	Twitching motility protein.	sp|Q06581|PILT_NEIGO (374 aa)	242-248 LTAaETG
Neosartorya fumigata (strain ATCC MYA-4609 / Af293 / CBS 101355 / FGSC A1100) (Aspergillus fumigatus)	Mediator of RNA polymerase II transcription subunit 21 (Mediator complex subunit 21).	sp|Q4X0I5|MED21_ASPFU (207 aa)	189-195 LTAvETG
Poliovirus type 1 (strain Mahoney)	Genome polyprotein	sp|P03300|POLG_POL1M (2209 aa)	623-629 LTAvETG
Poliovirus type 1 (strain Sabin)	Genome polyprotein	sp|P03301|POLG_POL1S (2209 aa)	623-629 LTAvETG
Poliovirus type 2 (strain Lansing)	Genome polyprotein	sp|P06210|POLG_POL2L (2207 aa)	622-628 LTAvETG
Poliovirus type 2 (strain W-2)	Genome polyprotein	sp|P23069|POLG_POL2W (2205 aa)	622-628 LTAvETG
Poliovirus type 3 (strain 23127)	Genome polyprotein	sp|P06209|POLG_POL32 (2206 aa)	620-626 LTAvETG
Poliovirus type 3 (strains P3/Leon/37 and P3/Leon 12A	Genome polyprotein	sp|P03302|POLG_POL3L (2206 aa)	620-626 LTAvETG
Pseudomonas aeruginosa (strain ATCC 15692 / DSM 22644 / CIP 104116 / JCM 14847 / LMG 12228 / 1C / PRS 101 / PAO1)	Twitching mobility protein.	sp|P24559|PILT_PSEAE (344 aa)	215-221 LTAaETG
Rattus norvegicus (Rat)	Interphotoreceptor matrix proteoglycan 2 (PG10.2) (Sialoprotein associated with cones and rods proteoglycan) (Spacrcan).	sp|P70628|IMPG2_RAT (1241 aa)	59-65 LDAtETG
Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's yeast)	Protein BRE4 (Brefeldin A-sensitivity protein 4).	sp|Q07660|BRE4_YEAST (1125 aa)	389-395 LTAmETS
Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's yeast)	Ubiquitin fusion degradation protein 1 (UB fusion protein 1) (Polymerase-interacting protein 3).	sp|P53044|UFD1_YEAST (361 aa)	71-77 LTAnETG
Salinibacter ruber (strain DSM 13855 / M31)	Glutamyl-tRNA(Gln) amidotransferase subunit A (Glu-ADT subunit A) (EC 6.3.5.7).	sp|Q2S4S2|GATA_SALRD (514 aa)	13-19 LDAgETS
Schizosaccharomyces pombe (strain 972 / ATCC 24843) (Fission yeast)	eIF-2-alpha kinase activator gcn1 (Translational activator gcn1).	sp|Q10105|GCN1_SCHPO (2670 aa)	1128-1134 LDAtETS
Schizosaccharomyces pombe (strain	Nucleoporin nup184 (Nuclear pore protein	sp|Q9P7M8|NU184_SCHPO	12-18

972 / ATCC 24843) (Fission yeast)	nup184).	(1564 aa)	LDAfETS
Shigella boydii serotype 18 (strain CDC 3083-94 / BS512)	Hydroxylamine reductase (EC 1.7.99.1) (Hybrid-cluster protein) (HCP) (Prismane protein).	sp|B2TUK6|HCP_SHIB3 (550 aa)	218-224 LDAgETG
Shigella boydii serotype 4 (strain Sb227)	Hydroxylamine reductase (EC 1.7.99.1) (Hybrid-cluster protein) (HCP) (Prismane protein).	sp|Q323M9|HCP_SHIBS (552 aa)	220-226 LDAgETG
Shigella dysenteriae serotype 1 (strain Sd197)	Hydroxylamine reductase (EC 1.7.99.1) (Hybrid-cluster protein) (HCP) (Prismane protein).	sp|Q32DZ1|HCP_SHIDS (552 aa)	220-226 LDAgETG
Shigella flexneri serotype 5b (strain 8401)	Hydroxylamine reductase (EC 1.7.99.1) (Hybrid-cluster protein) (HCP) (Prismane protein).	sp|Q0T8K8|HCP_SHIF8 (550 aa)	218-224 LDAgETG
Shigella flexneri	Hydroxylamine reductase (EC 1.7.99.1) (Hybrid-cluster protein) (HCP) (Prismane protein).	sp|Q83S05|HCP_SHIFL (552 aa)	220-226 LDAgETG
Shigella sonnei (strain Ss046)	Hydroxylamine reductase (EC 1.7.99.1) (Hybrid-cluster protein) (HCP) (Prismane protein).	sp|Q3Z3R0|HCP_SHISS (552 aa)	220-226 LDAgETG
Swine vesicular disease virus (strain H/3 '76) (SVDV)	Genome polyprotein	sp|P16604|POLG_SVDVH (2185 aa)	599-605 LTAaETG
Swine vesicular disease virus (strain UKG/27/72) (SVDV)	Genome polyprotein	sp|P13900|POLG_SVDVU (2185 aa)	599-605 LTAaETG
Synechococcus sp. (strain PCC 6716)	ATP synthase epsilon chain (ATP synthase F1 sector epsilon subunit) (F-ATPase epsilon subunit).	sp|Q05375|ATPE_SYNP1 (138 aa)	41-47 LTAlETG
Synechocystis sp. (strain PCC 6803 / Kazusa)	Chromosome partition protein Smc.	sp|P73340|SMC_SYNY3 (1200 aa)	806-812 LTAlETS
Thermosynechococcus elongatus (strain BP-1)	ATP synthase epsilon chain (ATP synthase F1 sector epsilon subunit) (F-ATPase epsilon subunit)	sp|Q8DLG7|ATPE_THEEB (138 aa)	41-47 LTAlETG
Xenopus laevis (African clawed frog)	DDB1- and CUL4-associated factor 8 (WD repeat-containing protein 42A).	sp|Q6NRH1|DCAF8_XENLA (601 aa)	50-56 LTAdETG

Fingerprints of 73 peptide binders

The Supplementary Figures S3 – S7 show the substitution analysis graphs of all 73 identified peptide binders. If the scale on the y-axis does not allow the read-out of the essential amino acids, the maximum was set to 200 % and an additional diagram is depicted next to the original diagram. The peptides are grouped according to Table 1.
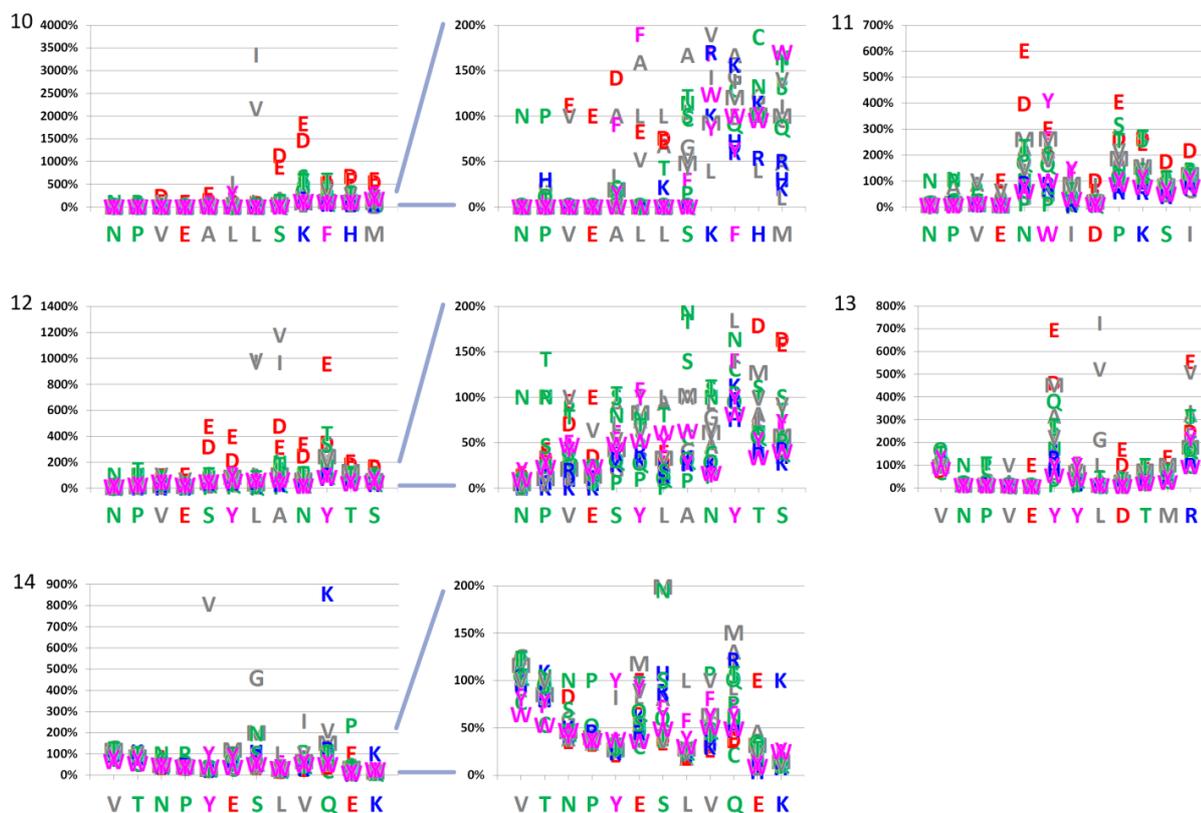
NPVEXXX motif
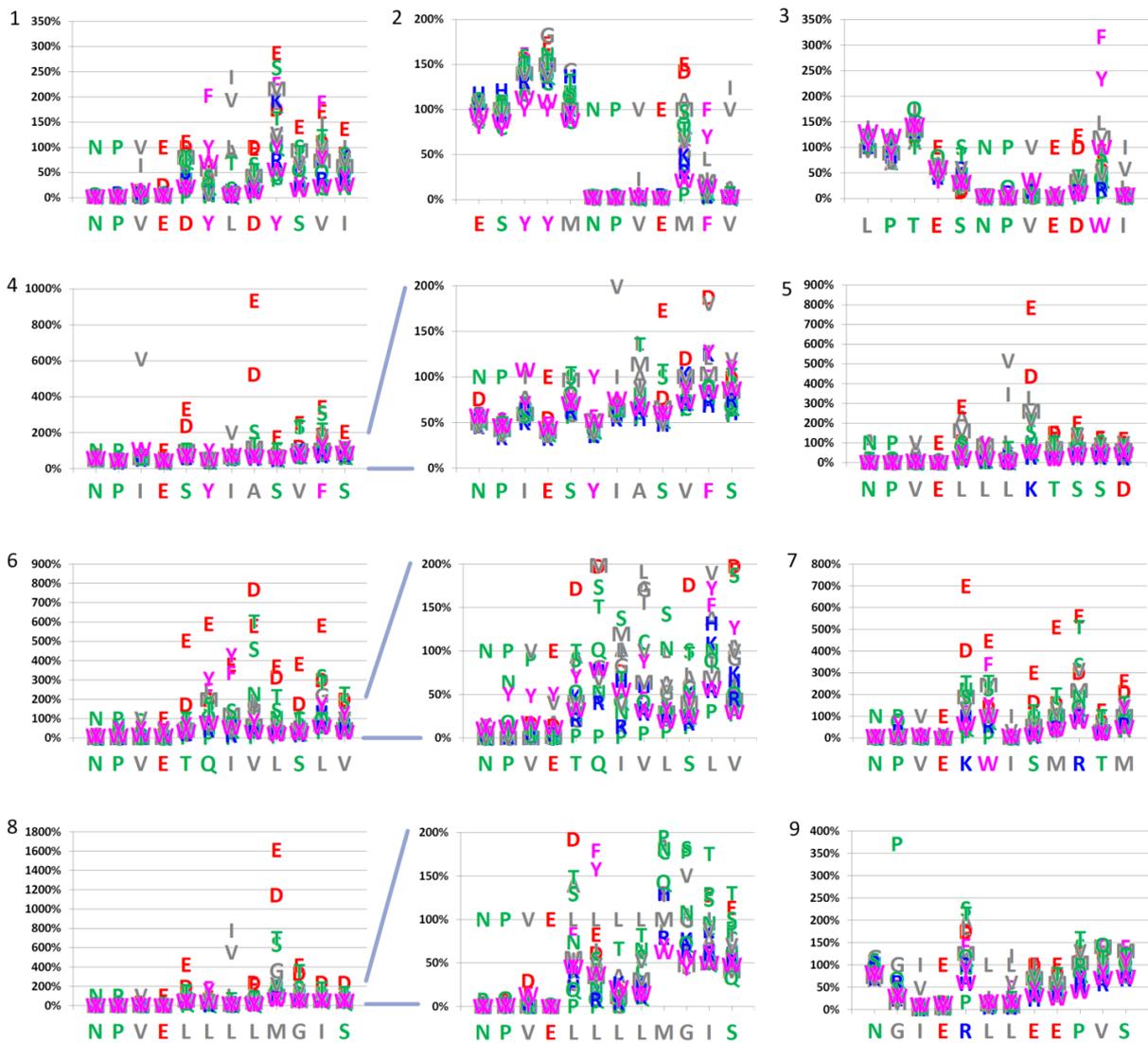
NPVEXXX motif (continued)

Figure S3: Substitution analyses of 14 peptide binders containing the "NPVEXXX" motif. Relative fluorescence intensities in reference to the original amino acid in the substituted peptide sequence. Each peptide variant is represented by the letter of the amino acid that was synthesized at the respective position. The original sequence is depicted on the x-axis. The intensity of the peptide with the original amino acid was set to 100 %. Intensities of peptide variants were correlated to the intensity of the original peptide in each row respectively.
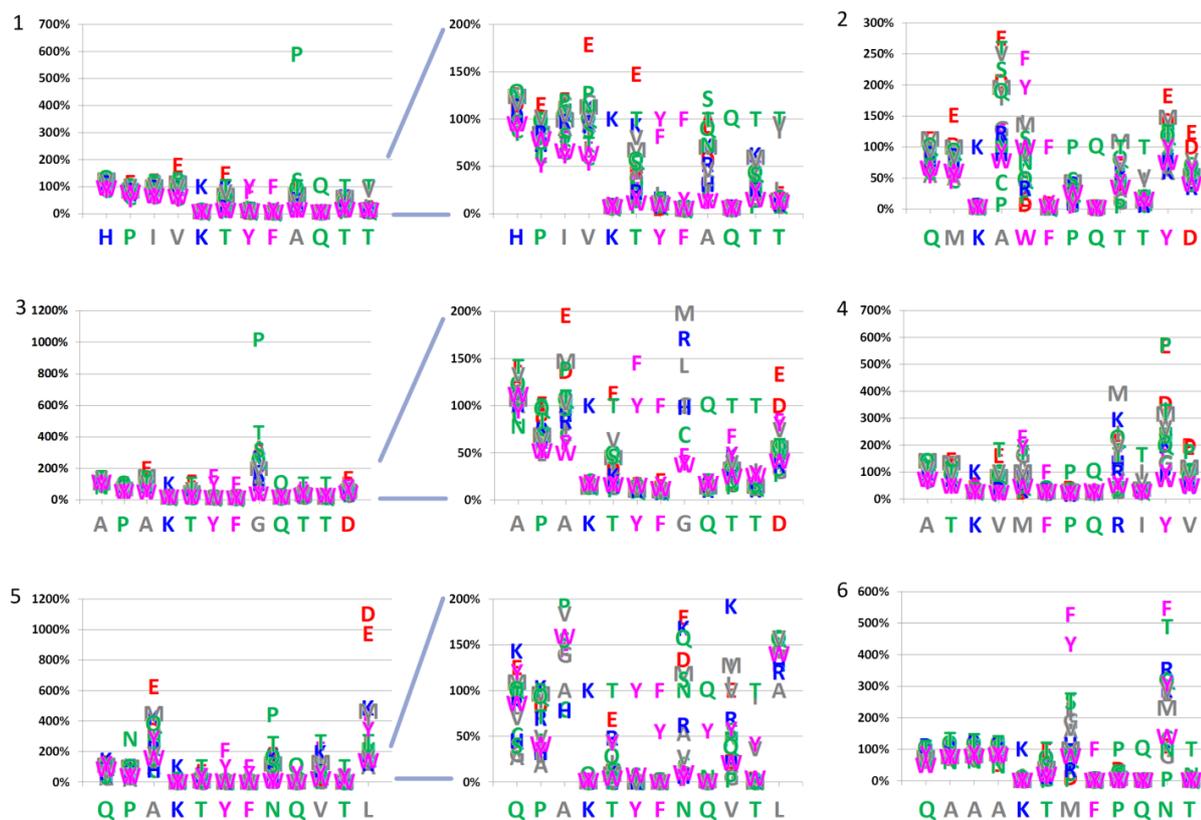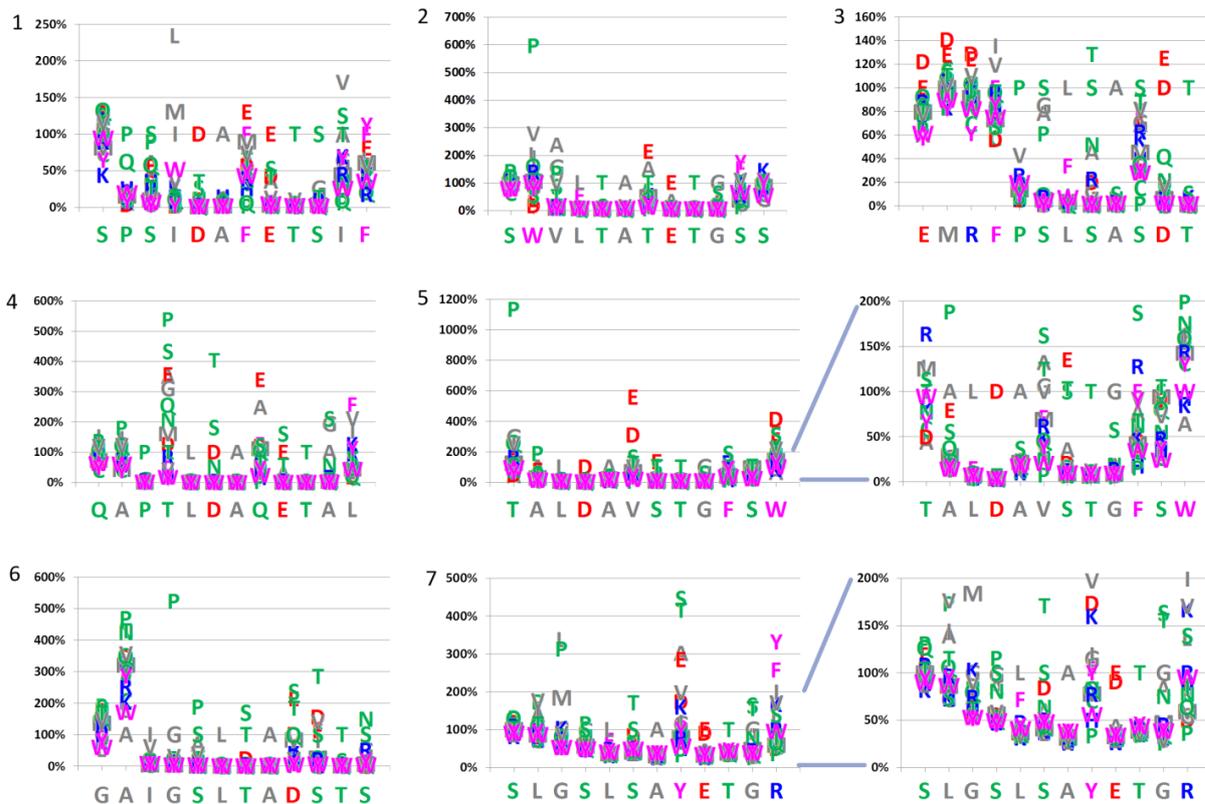
XPEFXGSXX motif
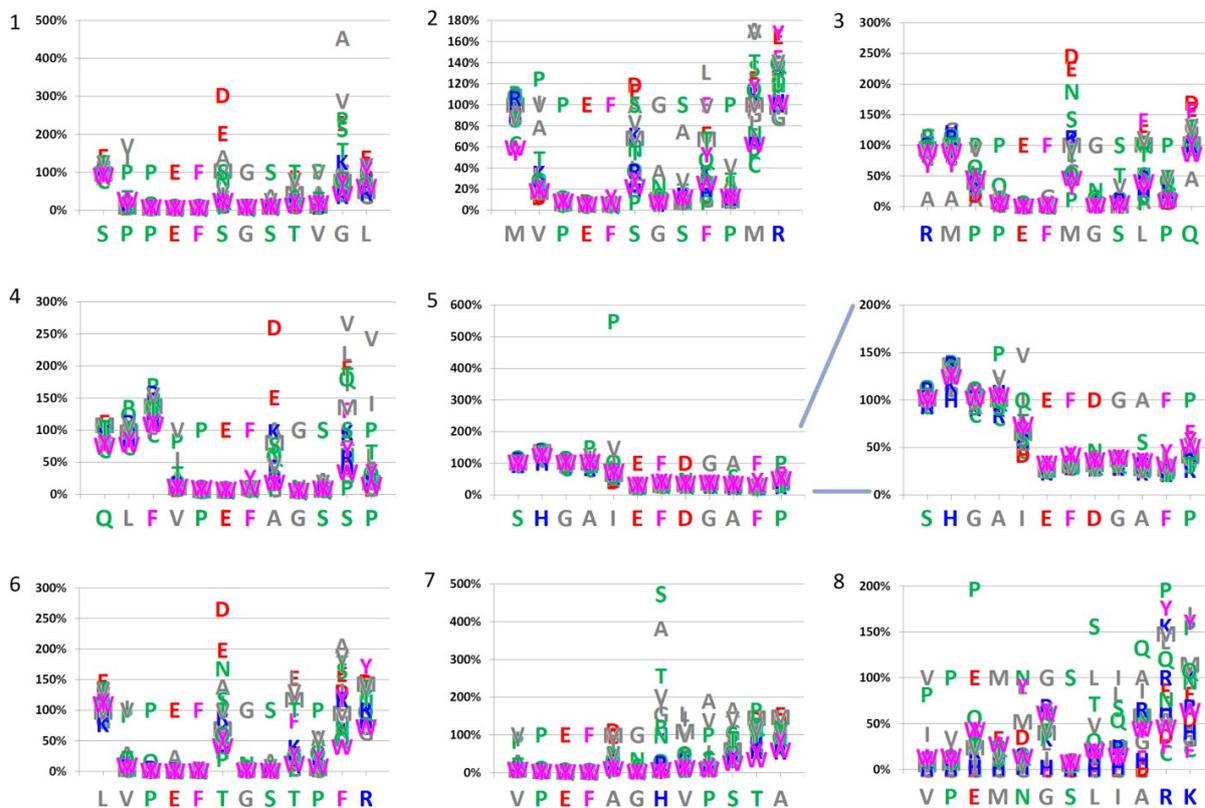
Figure S4: Substitution analyses of 8 peptide binders containing the "XPEFXGSXX" motif. Relative fluorescence intensities in reference to the original amino acid in the substituted peptide sequence. Each peptide variant is represented by the letter of the amino acid that was synthesized at the respective position. The original sequence is depicted on the x-axis. The intensity of the peptide with the original amino acid was set to 100 %. Intensities of peptide variants were correlated to the intensity of the original peptide in each row respectively.
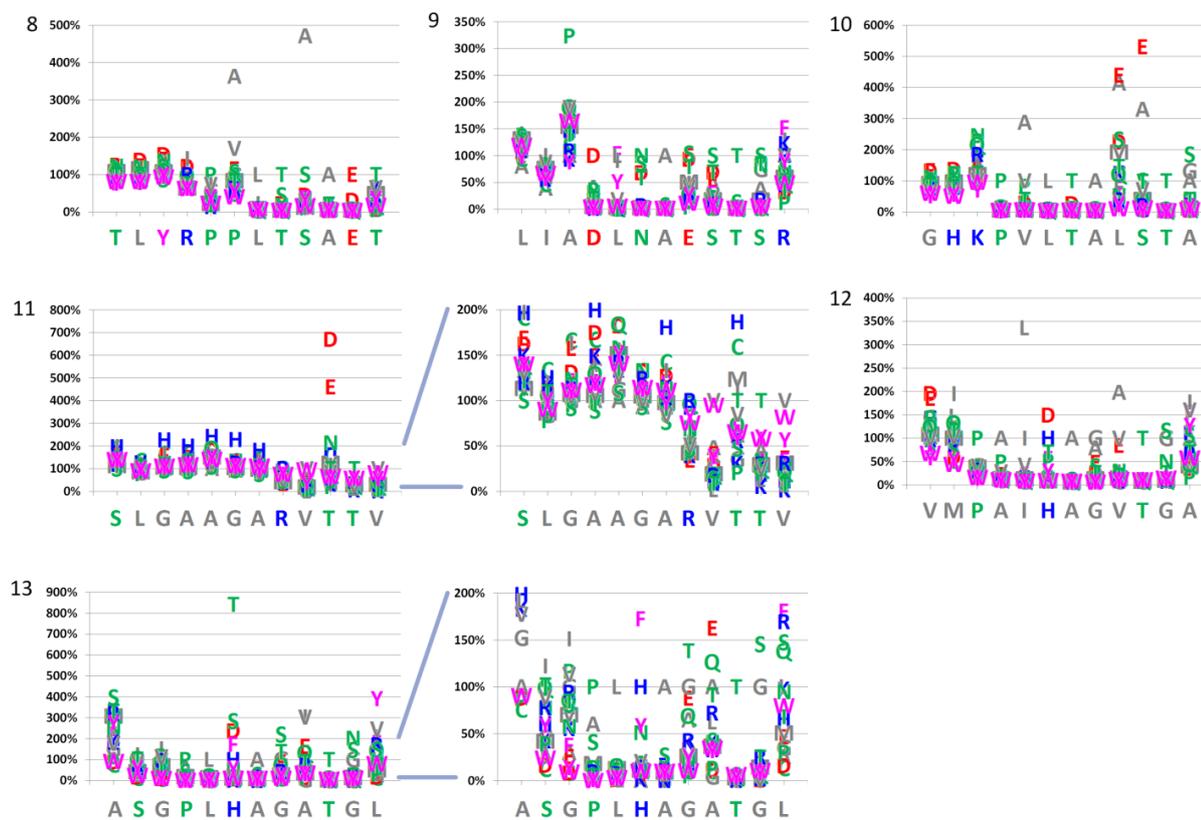
KXXFPQXT motif

Figure S5: Substitution analyses of 6 peptide binders containing the "KXXFPQXT" motif. Relative fluorescence intensities in reference to the original amino acid in the substituted peptide sequence. Each peptide variant is represented by the letter of the amino acid that was synthesized at the respective position. The original sequence is depicted on the x-axis. The intensity of the peptide with the original amino acid was set to 100 %. Intensities of peptide variants were correlated to the intensity of the original peptide in each row respectively.
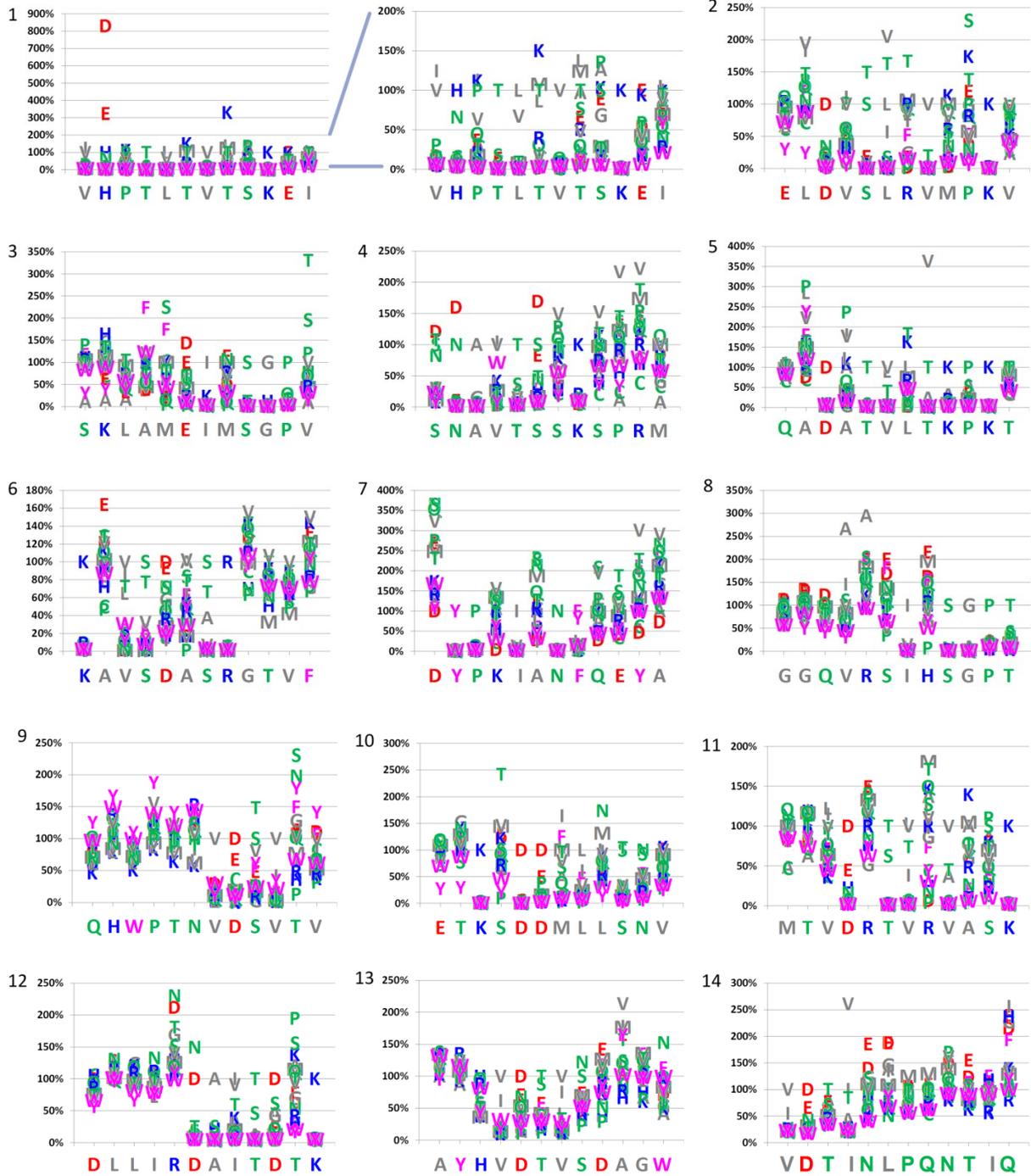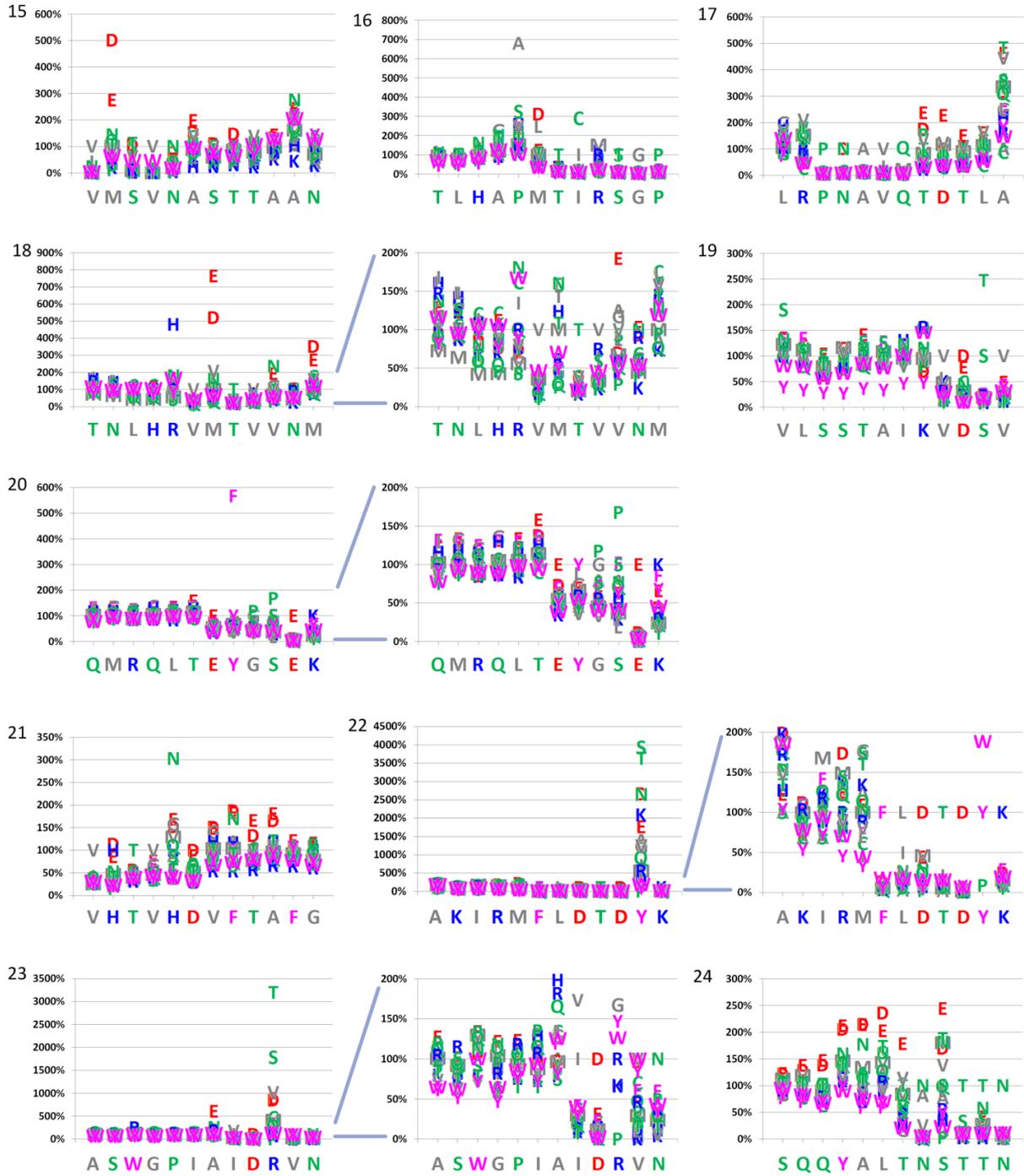
LXAXETX motif

LXAXETX motif (continued)

Figure S6: Substitution analyses of 13 peptide binders containing the "LXAXETX" motif. Relative fluorescence intensities in reference to the original amino acid in the substituted peptide sequence. Each peptide variant is represented by the letter of the amino acid that was synthesized at the respective position. The original sequence is depicted on the x-axis. The intensity of the peptide with the original amino acid was set to 100 %. Intensities of peptide variants were correlated to the intensity of the original peptide in each row respectively.

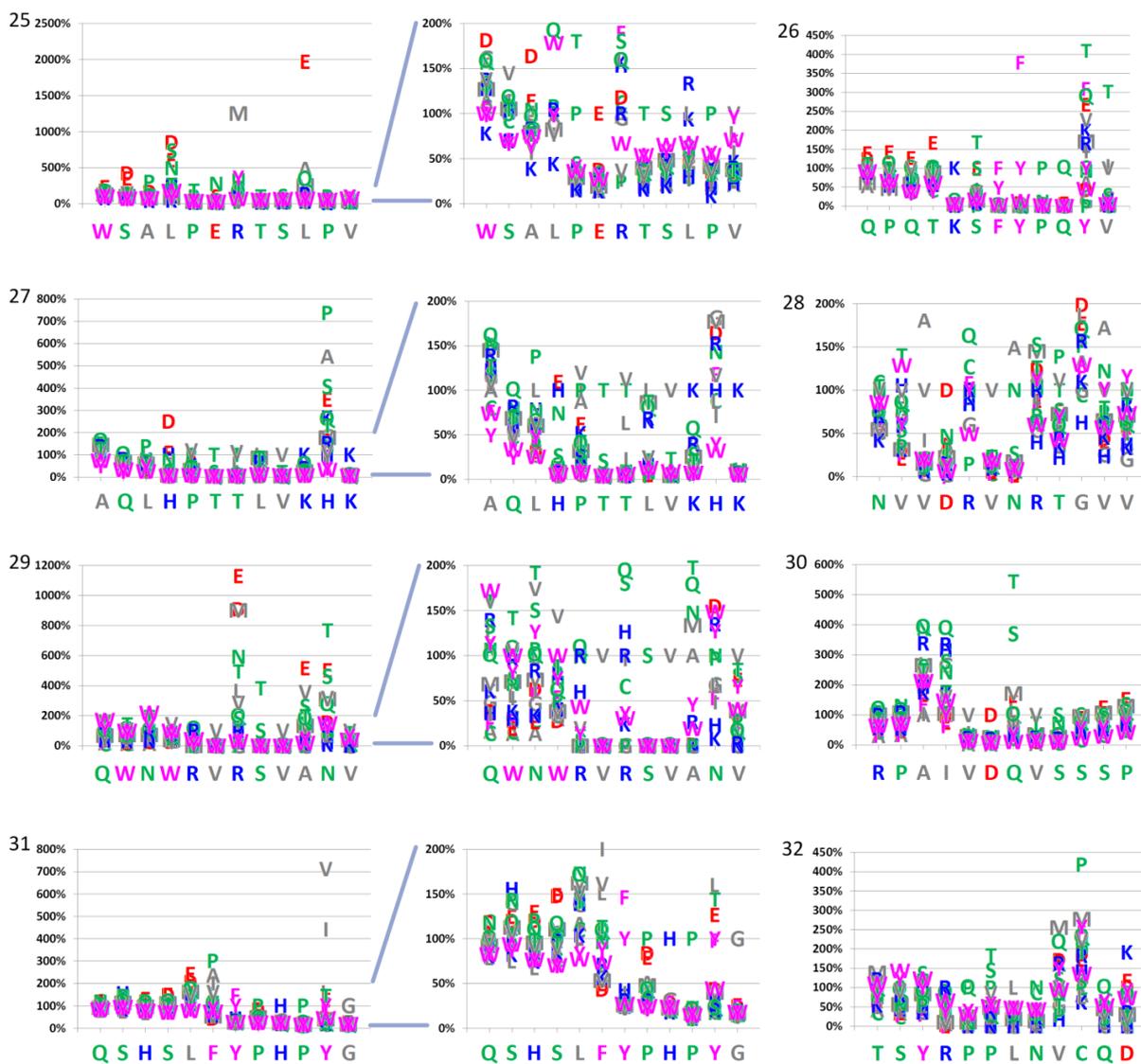*Heterogeneous motifs*

Heterogeneous motifs (continued)

Heterogeneous motifs (continued)

Figure S7: Substitution analyses of 32 peptide binders containing heterogeneous motifs. Relative fluorescence intensities in reference to the original amino acid in the substituted peptide sequence. Each peptide variant is represented by the letter of the amino acid that was synthesized at the respective position. The original sequence is depicted on the x-axis. The intensity of the peptide with the original amino acid was set to 100 %. Intensities of peptide variants were correlated to the intensity of the original peptide in each row respectively.